Estimation and identifiability of kinetic parameters in dynamical models of biochemical reaction networks

Håvard G. Frøysa

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2020

UNIVERSITY OF BERGEN

Estimation and identifiability of kinetic parameters in dynamical models of biochemical reaction networks

Håvard G. Frøysa

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 18.02.2020

Preface

This dissertation is submitted as a partial fulfillment of the requirements for the degree Doctor of Philosophy (PhD) at the Department of Mathematics, University of Bergen. The subject of the thesis is estimation of parameters in dynamical models of biochemical reaction networks. Parts of the work in the thesis have been performed in collaboration with the Department of Biological Sciences, University of Bergen.

Paper A, B and E are mathematical papers where the candidate is the first author. Paper C and D are biological papers where the candidate is the second author and performed modelling and data analysis, respectively. The candidate did not take part in the lab work of paper C and D.

Advisory committee:

- Guttorm Alendal (University of Bergen, Department of Mathematics)

- Hans Julius Skaug (University of Bergen, Department of Mathematics)

- Frank Nilsen (University of Bergen, Department of Biological Sciences)

Outline

The thesis is organised in two parts. Part I gives the background and motivation for the five papers included in part II.

Ch. 1 in part I contains the motivation for the thesis and gives an introduction to the biological context. The structure of a general reaction network is defined and studied in Ch. 2, while the dynamical properties of such networks are discussed in Ch. 3. Functions for the kinetics of the reactions in the network are introduced in Ch. 4, giving a system of ordinary differential equations for the concentrations in the network. This leads to the problem of parameter estimation and identifiability of the kinetic parameters considered in Ch. 5, which is the main focus of the papers. Ch. 6 is about data analysis of concentrations from an exposure experiment using statistical methods. Finally in part I, Ch. 7 contains an overview of the papers.

List of papers

Paper A:
Håvard G. Frøysa, Hans J. Skaug and Guttorm Alendal. *Experimental design for parameter estimation in steady-state linear models of metabolic networks.* Revision submitted to: Mathematical Biosciences, October 2019.

Paper B:
Håvard G. Frøysa, Shirin Fallahi and Nello Blaser. *Evaluating model reduction under parameter uncertainty.* BMC Systems Biology, **12**:79 (2018). doi:10.1186/s12918-018-0602-x

Paper C:
Kareem Eldin Mohammed Ahmed, **Håvard G. Frøysa**, Odd André Karlsen, Jørn V. Sagen, Gunnar Mellgren, Steven Verhaegen, Erik Ropstad, Anders Goksøyr and Ralf Kellmann. *LC-MS/MS based profiling and dynamic modelling of the steroidogenesis pathway in adrenocarcinoma H295R cells.* Toxicology in Vitro, **52**, 332-341 (2018). doi:10.1016/j.tiv.2018.07.002

Paper D:
Kareem Eldin Mohammed Ahmed, **Håvard G. Frøysa**, Odd André Karlsen, Nello Blaser, Karin Elisabeth Zimmer, Hanne Friis Berntsen, Steven Verhaegen, Erik Ropstad, Ralf Kellmann and Anders Goksøyr. *Effects of defined mixtures of POPs and endocrine disruptors on the steroid metabolome of the human H295R adrenocortical cell line.* Chemosphere, **218**, 328-339 (2019). doi:10.1016/j.chemosphere.2018.11.057

Paper E:
Håvard G. Frøysa. *Experimental design for parameter estimation in a dynamical model of the steroidogenesis pathway with a binary tree structure.*

Acknowledgements

It is a strange feeling to write this section. After four years as a PhD candidate I am now ready to submit my thesis. Doing a PhD in mathematics has been a goal of mine from long before I became a PhD candidate, and to be honest I believe that it has been expected from several of my friends and family too. For this reason, I believed that doing a PhD would be easy. Well, it was not. Doing a PhD *should* not be easy, but I made it with what I feel is a rewarding final result.

There are several people that should be thanked on this day. First of all, I would like to thank my supervisors Guttorm Alendal, Hans J. Skaug and Frank Nilsen. A special thanks to Guttorm and Hans for your guidance, support and valuable comments on my papers sent from all over the globe! It is highly appreciated, and I should probably have asked for your advice more often. Unfortunately the sea lice escaped my mathematical models this time Frank, but I am thankful for your help in the beginning of the project. You provided the inspiration for what turned out to be paper A in the thesis.

Next, I want to thank my co-authors of the published papers, in particular Kareem Eldin Mohammed Ahmed, Shirin Fallahi and Nello Blaser. I really enjoyed working together with you on the papers. A special thanks also to Anders Goksøyr for introducing me to Kareem and the steroidogenesis which became a major part of my thesis.

A big thanks goes out to the whole dCod 1.0 project. It has been nice getting to know all of you, both in a scientific and social setting. I am also extremely thankful for all the events that I have gotten the opportunity to be a part of through dCod.

I would further express my thanks to the whole Department of Mathematics, it has been a really nice place to work. A special thanks to Sondre for numerous lunches, and to Kristian for getting me to enjoy coffee when I needed it the most. I would also thank my dad, Sondre and Håvard for reading through parts of my thesis, thank you!

Finally, a big thanks to all my family and friends outside the university. Your support and interest in my project is much appreciated! A special thanks to my parents for your advice and kindness, I am extremely grateful.

Håvard G. Frøysa
November 2019

Give thanks to the Lord, for he is good;
his love endures forever.
Psalm 107,1

Contents

Part I

Background

Chapter 1

Introduction and biological motivation

The main theme throughout this thesis is mathematical modelling of biochemical reaction networks. In this chapter we give some motivation and biological background, while the remaining chapters will be of mathematical character.

1.1 Motivation

There are several applications where mathematical modelling of biochemical reaction networks could be valuable. Before we give an introduction to reaction networks and their models, we briefly present the three applications that have motivated this thesis.

1.1.1 Salmon louse

The original motivation of my PhD project was to study growth of the salmon louse (Lepeophtheirus salmonis) [24, 25, 35]. The salmon louse is a parasite that infects salmonids, and is a major threat to both farmed and wild salmon [77]. Indeed, sea lice in general are responsible for large commercial losses for the salmon farming industry due to control requirements [19, 56]. For this reason, large research projects like the Sea Lice Research Centre [5] have been initiated to improve future sea lice control.
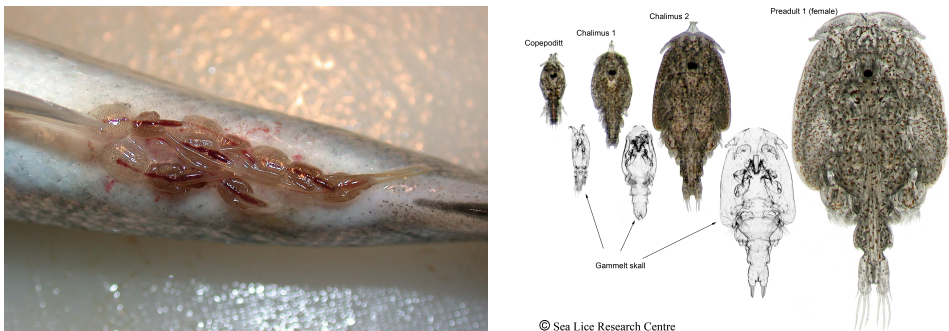
Figure 1.1: *Lice on a salmon (left) and the salmon louse life cycle (right). Courtesy of the Sea Lice Research Centre [5].*

The growth of salmon louse consists of several stages. Between some stages, a new exoskeleton has to be formed to enter the next stage. A major component in the formation of this exoskeleton is chitin [36, 37]. This can be produced either from the chitin in the old exoskeleton or from nutrients, where the chitin pathway is a part of the amino sugar and nucleotide sugar metabolism [3]. It would then be of great interest if we were able to make a good model for the amino sugar and nucleotide sugar metabolism, and potentially use it to predict the behaviour of the salmon louse under different treatments designed to limit the growth process by inhibiting the chitin production. Unfortunately, I did not end up making any such models for the salmon louse in the thesis.

1.1.2   The dCod 1.0 project

Early in my PhD period, I joined the dCod 1.0 project [7], running from 2016 to 2020, funded by the Research Council of Norway within the Centre for Digital Life Norway [1]. The goal of the project is *to create a deeper understanding of cods' adaptations and reactions to stressors in the environment* by combining knowledge within environmental toxicology, biology, bioinformatics and mathematics.

In particular is the behaviour of the cod liver and the fatty acid metabolism [2] of special interest to model and understand, where precision cut liver slices are extensively used in the studies [14, 26, 27, 85] in addition to field studies [20].

Unfortunately, the thesis does not contain any modelling directly related to cod. However, most of the co-authors of the papers included in the thesis are members of the dCod 1.0 project. This way, the dCod 1.0 project has been the most important scientific community during the work on the thesis.

1.1.3   Steroidogenesis

The third and final motivation for the thesis is the study of steroidogenesis [6]. The steroid hormone biosynthesis is an important part of the metabolism for both animals and humans. However, more than 800 chemicals are referred to as endocrine disrupting chemicals that may interfere with this system and potentially cause harm [12, 52]. For this reason, it is of great interest to build a model of steroidogenesis that can be used to understand the response of the system to various endocrine disrupting chemicals. The first step in this process is to develop an assay for performing measurements of the steroids, which is done in paper C and D of this thesis by our biological collaborators [8, 9]. Based on this, I have created a mathematical model of steroidogenesis in paper C which is studied in more details in paper E of the thesis.

1.2 Biochemical reaction networks

To study the phenomena listed above, we need to model the chemical processes that take place inside the cells of a living organism. The whole set of such chemical reactions in an organism is called the *metabolism* of the organism, and may be divided into several metabolic *pathways*. In each of these pathways, one chemical is transformed into another chemical, where these chemicals are referred to as a *metabolites*. An overview of all the metabolic pathways in the general metabolism from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [46] is shown in Fig. 1.2.
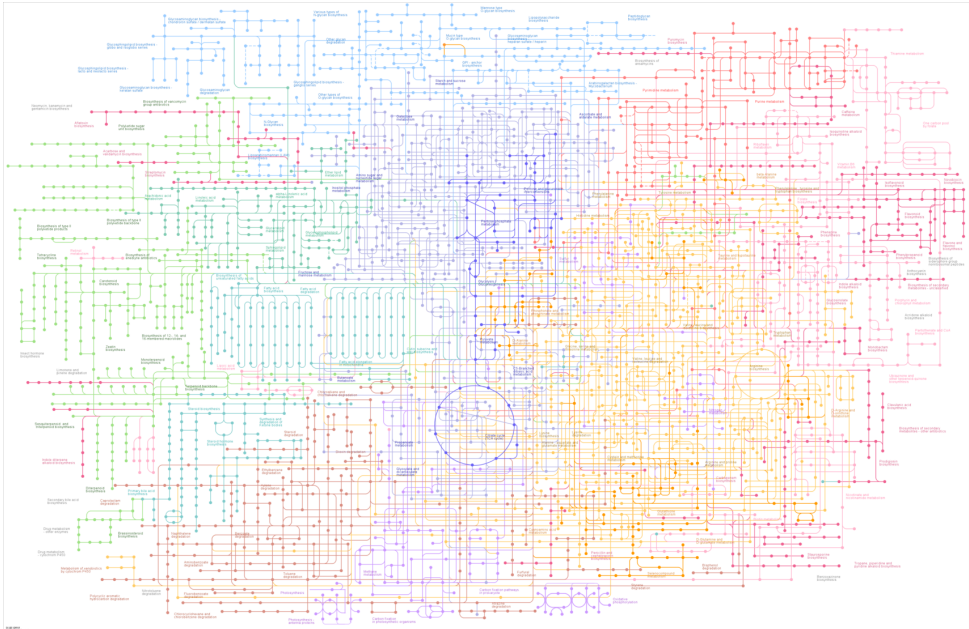
Figure 1.2: *Overview of all the metabolic pathways in the metabolism from KEGG [4].*

1.2.1 Reconstruction

We see from Fig. 1.2 that there is an overwhelming number of metabolic pathways in the metabolism, where each pathway consists of several chemical reactions. However, which pathways and reactions that are active vary between different biological species and Fig. 1.2 shows all the possible known pathways.

In a controlled experiment are the sets of active reactions and metabolites in the metabolism of an organism given by the genes, called the *genome* of the organism. An attempt at obtaining these sets is called a *reconstruction* of the organism, where a reconstruction of the whole metabolism is said to be a *genome scale reconstruction*.

Reconstructions were first published in the 1990s, where the E. coli bacteria was the main model organism [63]. The first genome scale reconstruction of E. coli was published in 2000 by Edwards and Palsson [23], consisting of 627 reactions and 438 metabolites. However, the improved reconstruction of E. coli in 2011 by Orth et al.

[62] consists of 2251 reactions and 1136 metabolites. Where the first reconstruction only models parts of the metabolism, the later reconstructions of E. coli now attempt to model the whole metabolism.

A good reconstruction for human is of course an important goal within the field, and the Recon 2 published in 2013 by Thiele et al. [76] consists of 7440 reactions and 5063 metabolites. In total, more and more organisms are now getting reconstructed with 6239 organisms having been reconstructed as of February 2019 [32].

To illustrate what the result of a reconstruction may look like, we return to the amino sugar and nucleotide metabolic pathway discussed in Sec. 1.1.1 for the salmon louse. This pathway is shown in Fig. 1.3 with green coloring of the relevant reactions for Drosophila melanogaster (fruit fly), which is the available reconstruction assumed to be most similar to the salmon louse. This implies that only the green parts should be included in a model of the pathway for the salmon louse.
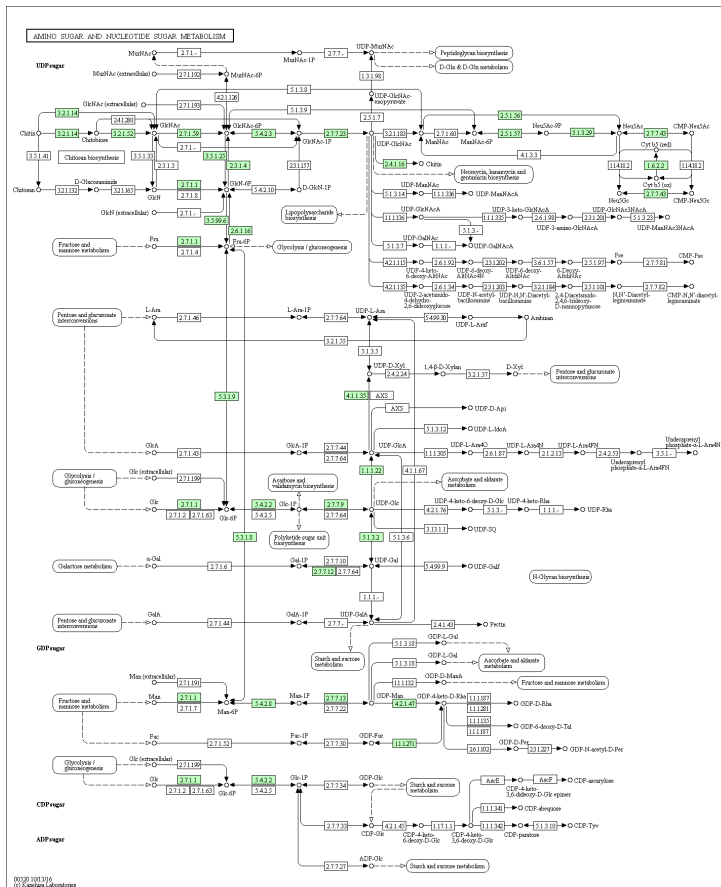
Figure 1.3: *The amino sugar and nucleotide sugar metabolic pathway discussed in relation to the salmon louse in Sec. 1.1.1. Illustration from KEGG with green colouring of the relevant reactions for Drosophila melanogaster (fruit fly) [3].*

We will in this thesis assume the reaction networks to be given and not focus any

more on the reconstruction process even though this is an important part in the modelling of biochemical reaction networks. Our models will also only be for smaller networks, sub-networks or pathways like the steroid hormone biosynthesis pathway discussed in Sec. 1.1.3, i.e. not for complete genome-scale networks.

1.2.2 Metabolites

Metabolites are made of only six different chemical elements [63, p. 155]. These six are carbon (C), oxygen (O), nitrogen (N), hydrogen (H), phosphorous (P) and sulfur (S). Note that an individual metabolite typically does not contain all six elements. It is also worth mentioning that different metabolites may have the same elemental composition. Some metabolites from the pathways mentioned in Sec. 1.1 are listed below.

- Amino sugar and nucleotide sugar metabolism [3]:

  - Chitin, $(C_8H_{13}O_5N)_n$
  - Chitobiose, $C_{16}H_{28}N_2O_{11}$

- Steroid hormone biosynthesis [6]:

  - Pregnenolone, $C_{21}H_{32}O_2$
  - Testosterone, $C_{19}H_{28}O_2$
  - Cortisone, $C_{21}H_{28}O_5$

The chemical formulas are of less interest for our mathematical modelling, and we use a generic mathematical notation X_i for the metabolites in the later chapters.

1.2.3 Chemical reactions

Chemical reactions in metabolism are of only three different categories [63, p. 154]. To illustrate the three categories, let C be a primary metabolite, P a phosphate group and A a co-factor such as ATP. The first category is a *reversible conversion* given by

$$CP \quad \Longleftrightarrow \quad PC \tag{1.1}$$

which is a rearrangement in the molecule without changes in the elemental composition. The second category of reactions is a *bi-molecular association* such as

$$C+P \quad \Longleftrightarrow \quad CP \tag{1.2}$$

where a new compound is formed. The final category is a *co-factor-coupled reaction*

$$C+AP \quad \Longleftrightarrow \quad CP+A \tag{1.3}$$

where P is donated from AP to C through an intermediate not included in the chemical reaction, and the co-factor A is denoted a carrier.

In our models, however, most of the reactions will be modelled as

$$X_1 \longrightarrow X_2 \tag{1.4}$$

where X_1 and X_2 are different metabolites. This will of course affect the models, but is a common assumption to obtain models of manageable complexity. For more about this, see e.g. Feliu and Wiuf [29] about simplifying models with intermediate species.

1.3   Systems biology

To model the large networks described above, we are led to the relatively new field of *systems biology* that has emerged in full strength in the 21st century together with the genome-scale reconstructions [48, 63]. The traditional approach in cell biology has been reductionism where the components of the biological system are modelled individually. With systems biology, however, the system is modelled as a whole since there may be interesting emergent properties only found on a systems level [13].

The transition from so-called components biology to systems biology can be regarded as a paradigm shift within cell and molecular biology that has taken place within the last 20 years. One of the main drivers of this shift has been the ability of new high-throughput omics technologies to produce large data sets with extensive information about the cellular functions [21, 81]. Hand in hand with this, there has also been a computational revolution where the computational power needed to handle the large amount of data has become available. The data in systems biology typically need a lot of processing and analysis, which leads to the field of bioinformatics [55]. To help with these analyses, there also exist toolboxes like COBRA 2.0 [72] and RAVEN 2.0 [82].

The focus in this thesis is on the mathematical modelling of biochemical reaction networks and parameter estimation in such models. For this reason, we do not go in further depth about the underlying biology and bioinformatics that lead to our models. It is, however, important to remember where our models originate from. One thing that could be of specific interest to remember is the dual causality that makes modelling in biology more complicated than in e.g. mechanics, as described by the quote below.

While biological functions obey the physical laws,
their functions are not predictable by the physical laws alone.

                                                        -Bernhard Ø. Palsson [63, p. 251]

1.4   Balancing of chemical reactions

The reactions will later be assumed given and typically modelled by $X_1 \rightarrow X_2$ as in Eq. 1.4. Before we move to the network formulation in the next chapter, however, we show how to make sure that a general chemical reaction obeys conservation of mass, which later will be an underlying assumption for the reactions.

A reaction is said to be *balanced* and satisfies conservation of mass if the number of atoms for each element is the same on both sides. Reactions that do not satisfy this are said to be *unbalanced* and the process of making them balanced is called *balancing*. This is done by adding integer values called *stoichiometric constants* in front of the different compounds in the reactions. For simple unbalanced reactions such as

$$H_2 + O_2 \quad \longrightarrow \quad H_2O \tag{1.5}$$

this can be done by inspection and we see that

$$2H_2 + O_2 \quad \longrightarrow \quad 2H_2O \tag{1.6}$$

is the balanced reaction. However, for more complicated chemical reactions such as

$$Pb + Na + C_2H_5Cl \quad \longrightarrow \quad Pb(C_2H_5)_2 + NaCl \tag{1.7}$$
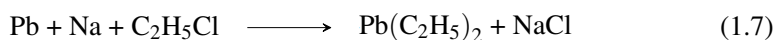
it is not as straight-forward to find the balanced reaction. Such a problem is typically solved by trial and failure in high school chemistry until a solution is obtained, and in many cases including Eq. 1.7 this could potentially be a satisfactory approach. However, the balancing of a chemical reaction can be done systematically by a linear equation system where the stoichiometric constants are the unknowns.

Let $\mathbf{s}$ be the vector of unknown stoichiometric constants to be found, where each of the entries in $\mathbf{s}$ is associated with a compound. A substrate compound on the left hand side of the reaction will have a negative value in $\mathbf{s}$ as it is consumed, while a product compound on the right hand side of the reaction will have a positive value. For the example of Eq. 1.5, we have $\mathbf{s} = \begin{bmatrix} s_1 & s_2 & s_3 \end{bmatrix}^T$ where the interpretation of $\mathbf{s}$ is

$$(-s_1)H_2 + (-s_2)O_2 \longrightarrow (s_3)H_2O \tag{1.8}$$
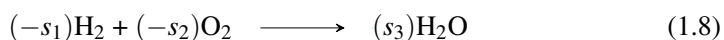
such that s_1 is associated with H_2, s_2 with O_2 and s_3 with H_2O. Note that the negative values of s_1 and s_2 are used in front of their compounds to obtain a positive value.

Now, let $\mathbf{E}$ be the *elemental matrix* of a reaction that is used to find $\mathbf{s}$. The rows of $\mathbf{E}$ correspond to the different elements that participate in the reaction and the columns to the different compounds of the reaction. Each entry of the matrix is then the number of a given element in one molecule of a given compound, such that all the entries of the matrix are non-negative integers. If an entry is zero, the corresponding element is not part of the corresponding compound. For the example of Eq. 1.5, we have

$$\mathbf{E} = \begin{array}{c} \\ H \\ O \end{array} \begin{array}{c} H_2 \ \ O_2 \ \ H_2O \\ \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix} \end{array} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{1}{2} \end{bmatrix} \tag{1.9}$$

where the row reduced echelon form is calculated on the right hand side.

Having the matrix $\mathbf{E}$ and the vector $\mathbf{s}$, conservation of mass is given by

$$\mathbf{Es} = \mathbf{0}. \tag{1.10}$$

which is a homogeneous linear equation system where we are interested in the smallest integer solution. For the example of Eq. 1.5 we obtain the solution

$$\mathbf{s} = \gamma \begin{bmatrix} -1 & -\frac{1}{2} & 1 \end{bmatrix}^T \tag{1.11}$$

from Eq. 1.9 where $\gamma \in \mathbb{R}$ is a free parameter. The smallest integer solution is given by $\gamma = 2$ which gives $\mathbf{s} = \begin{bmatrix} -2 & -1 & 2 \end{bmatrix}^T$ and the same balanced reaction as proposed in Eq. 1.6 if the obtained values for $\mathbf{s}$ are substituted into Eq. 1.8.

For the more advanced example of Eq. 1.7 we have the elemental matrix

$$\mathbf{E} = \begin{array}{c} \\ Pb \\ Na \\ C \\ H \\ Cl \end{array} \begin{array}{c} Pb \ \ Na \ \ C_2H_5Cl \ \ Pb(C_2H_5)_2 \ \ NaCl \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 2 & 4 & 0 \\ 0 & 0 & 5 & 10 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \end{array} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.12}$$

where the row reduced echelon form again is on the right. This gives the solution

$$\mathbf{s} = \gamma \begin{bmatrix} -\frac{1}{2} & -1 & -1 & \frac{1}{2} & 1 \end{bmatrix}^{\mathrm{T}} \qquad (1.13)$$

of Eq. 1.10 where $\gamma \in \mathbb{R}$ is a free parameter, and the the entries in $\mathbf{s}$ correspond to the compounds as seen in $\mathbf{E}$. The smallest integer solution given by $\gamma = 2$ is

$$\mathbf{s} = \begin{bmatrix} -1 & -2 & -2 & 1 & 2 \end{bmatrix}^{\mathrm{T}}, \qquad (1.14)$$

which gives the desired balanced chemical reaction

$$\mathrm{Pb} + 2\mathrm{Na} + 2\mathrm{C_2H_5Cl} \longrightarrow \mathrm{Pb(C_2H_5)_2} + 2\mathrm{NaCl}. \qquad (1.15)$$

Note that there is always a free parameter in the solution for $\mathbf{s}$ if a non-trivial solution exists, as Eq. 1.10 is a homogeneous equation. This free parameter is used to find an integer solution for $\mathbf{s}$, and is only a scaling of all the stoichiometric constants. If there, however, is more than one free parameter in the solution this implies that several different reactions are possible for the given compounds which will be of relevance in the next chapter. Also note that it is not specified in $\mathbf{E}$ which compounds that belong to the left and right hand side of the reaction as all the entries are non-negative. For this reason, the sign of $\mathbf{s}$ is impossible to determine mathematically and additional information is needed to decide the direction of the reaction.

Chapter 2

Structure of reaction networks

In this chapter we define the *stoichiometric matrix* **S** for a reaction network, and explore some of its features. This matrix is a mathematical representation of the network structure, and is important for the dynamical models of reaction networks in the next chapters. The main reference for this and the next chapter is the textbook "Systems biology" by Palsson [63], and we adopt much of the notation from there.

2.1 Stoichiometric matrix

Assume that we have a network consisting of r irreversible reactions R_j involving n metabolites with chemical names X_i. The network can then be written as

$$R_j: \quad \sum_{i=1}^{n} \alpha_{ij} X_i \rightarrow \sum_{i=1}^{n} \beta_{ij} X_i \qquad j = 1, \dots, r \tag{2.1}$$
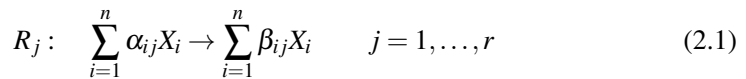
where the stoichiometric coefficients α_{ij} and β_{ij} are non-negative integers [18].

The stoichiometric matrix $\mathbf{S} = \{S_{ij}\} \in \mathbb{Z}^{n \times r}$ is given by the unitless entries

$$S_{ij} = \beta_{ij} - \alpha_{ij} \tag{2.2}$$

which gives the net production of metabolite X_i in reaction R_j.

The columns of **S** correspond to reactions, and the rows to metabolites such that

$$\mathbf{S} = \begin{array}{c} \text{Metabolites } X_i \end{array} \overset{\text{Reactions } R_j}{\begin{bmatrix} S_{11} & \cdots & S_{1r} \\ \vdots & \ddots & \vdots \\ S_{n1} & \cdots & S_{nr} \end{bmatrix}} \tag{2.3}$$

where the labels X_i and R_j may be omitted later for simplicity. Note that the network of Eq. 2.1 could be visualized as a directed graph with the metabolites as the nodes and the reactions as edges, which we will make use of for the examples following later.

2.1.1   Relation to balancing of reactions

Before we proceed with the exploration of $\mathbf{S}$, we make one last visit to the balancing of chemical reactions in Sec. 1.4. The stoichiometric matrix $\mathbf{S}$ of Eq. 2.3 can be written

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 & \dots & \mathbf{s}_r \end{bmatrix} \tag{2.4}$$

where the vectors $\mathbf{s}_j$ are the columns of $\mathbf{S}$. Each of these vectors correspond to a reaction and is equivalent to the vector $\mathbf{s}$ in Sec. 1.4. Let now $\mathbf{E}$ be the elemental matrix for all the n metabolites that participate in the network of Eq. 2.1. If all of the reactions in Eq. 2.1 are balanced, then all $\mathbf{s}_j$ satisfy the balance equation in Eq. 1.10 such that

$$\mathbf{E}\mathbf{s}_j = \mathbf{0} \quad \text{for } j = 1, \dots, r \tag{2.5}$$

which can be collected into the matrix equation

$$\mathbf{E} \begin{bmatrix} \mathbf{s}_1 & \dots & \mathbf{s}_r \end{bmatrix} = \mathbf{E}\mathbf{S} = \mathbf{0}. \tag{2.6}$$

From now on, however, we are no longer concerned with the balancing of the reactions. The main reason for this is that we may omit some compounds from the models for simplicity as discussed briefly in Sec. 1.2.3 on chemical reactions.

2.1.2   Some properties of the stoichiometric matrix S

In the typical case where X_i is present only at one side of the reaction R_j, we have $S_{ij} = \beta_{ij}$ or $S_{ij} = -\alpha_{ij}$ from Eq. 2.2. However, Eq. 2.2 in the general case allows a metabolite X_i to be present at both sides of a reaction.

When an open system is modelled, we have $\alpha_{ij} = 0$ or $\beta_{ij} = 0$ for all i if the reaction R_j is entering or leaving the system, respectively. A reversible reaction can be represented in Eq. 2.1 by two irreversible reaction, or one may represent it directly as a reversible reaction. Some parts of the thesis, however, require all the reactions to be irreversible. For large networks, most of the α_{ij} and β_{ij} are zero making $\mathbf{S}$ a sparse matrix. This is because only a few metabolites are involved in each reaction independently of the number of reactions. For genome-scale models the most common number of compounds participating in a reaction is four including co-factors [63, p. 165].

Linear maps are networks where the reactions only have one input and one output like $X_1 \rightarrow X_2$ in Eq. 1.4 such that $\mathbf{S}$ has only two non-zeros entries in each column. Most of the networks in the papers of this thesis are assumed to be such linear maps.

2.2   Example networks

To illustrate some properties of the stoichiometric matrix $\mathbf{S}$, we create some simple toy networks in this section. These are much smaller than real networks, but are useful to understand the properties of $\mathbf{S}$.
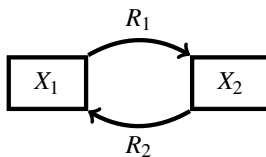
Figure 2.1: *Example network consisting of a loop. Each node is a metabolite, and each arrow a reaction. The reactions are given in Eq. 2.7.*

2.2.1 Loop network

Our first toy example is shown in Fig. 2.1, and is a loop consisting of two metabolites and two irreversible reactions that together form a reversible reaction.

We assume for simplicity that all non-zero α_{ij} and β_{ij} are equal to one. The reactions of the network in Fig. 2.1 can then be written on the form of Eq. 2.1 as

$$
\begin{aligned}
R_1 : \quad X_1 &\longrightarrow X_2 \\
R_2 : \quad X_2 &\longrightarrow X_1
\end{aligned}
\tag{2.7}
$$

which gives the 2×2 stoichiometric matrix

$$
\mathbf{S} = \begin{array}{c} \\ X_1 \\ X_2 \end{array} \overset{\begin{array}{cc} R_1 & R_2 \end{array}}{\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}} \quad \text{with } \mathrm{Rank}\,(\mathbf{S}) = 1.
\tag{2.8}
$$

2.2.2 Split network

Our second example consists of four metabolites and six reactions. It is shown in Fig. 2.2, and the most prominent feature is the split into two branches. The network is open since it has the entering reaction R_1, and also the leaving reactions R_5 and R_6.
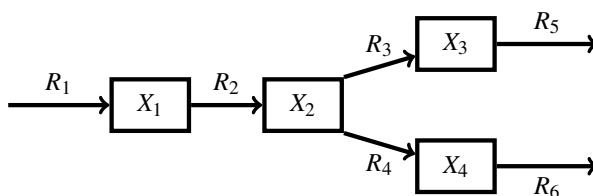
Figure 2.2: *Example network consisting of a split. Each node is a metabolite, and each arrow a reaction. The reactions are given in Eq. 2.9.*

By assuming once again that all non-zero α_{ij} and β_{ij} are one as for the previous example, the reactions of the network can be written on the form of Eq. 2.1 as

$$
\begin{aligned}
R_1 : \quad &\longrightarrow X_1 \\
R_2 : \quad X_1 &\longrightarrow X_2 \\
R_3 : \quad X_2 &\longrightarrow X_3 \\
R_4 : \quad X_2 &\longrightarrow X_4 \\
R_5 : \quad X_3 &\longrightarrow \\
R_6 : \quad X_4 &\longrightarrow
\end{aligned}
\tag{2.9}
$$
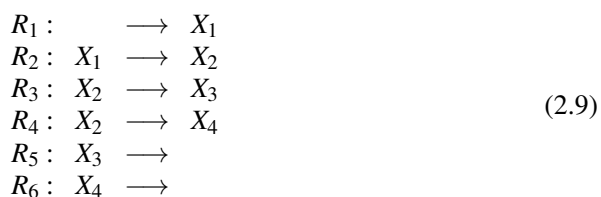
where R_1 is an entering reaction, R_2, R_3 and R_4 are internal reactions, and finally are R_5 and R_6 leaving reactions. The 4×6 stoichiometric matrix becomes

$$\mathbf{S} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \overset{\begin{array}{cccccc} R_1 & R_2 & R_3 & R_4 & R_5 & R_6 \end{array}}{\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}} \quad \text{with Rank}(\mathbf{S}) = 4. \qquad (2.10)$$

2.2.3 Network of complexes

The two example networks this far have at most one metabolite on each side of all the reactions. For this reason, we want a network with interactions such that Eq. 2.1 may have more than one non-zero α_{ij} and β_{ij} per reaction. Also, we want a network where all the non-zero α_{ij} and β_{ij} not necessarily are one. To meet these requirements, we introduce the example of Fig. 2.3 which is a modified version of a network in Rao et al. [67]. It consists of four metabolites and five reactions that interact through four different complexes, which will be discussed in the next section.
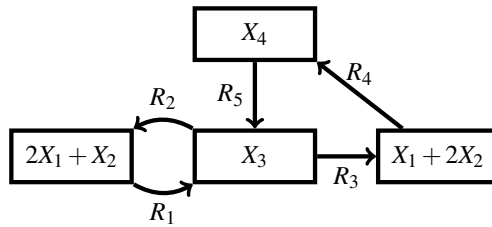
Figure 2.3: *Network based on Rao et al. [67]. The nodes are complexes consisting of potentially several metabolites and the arrows are reactions. The reactions are given in Eq. 2.11.*

The reactions of the network on the form of Eq. 2.1 are given by

$$\begin{array}{rlcl} R_1: & 2X_1 + X_2 & \longrightarrow & X_3 \\ R_2: & X_3 & \longrightarrow & 2X_1 + X_2 \\ R_3: & X_3 & \longrightarrow & X_1 + 2X_2 \\ R_4: & X_1 + 2X_2 & \longrightarrow & X_4 \\ R_5: & X_4 & \longrightarrow & X_3 \end{array} \qquad (2.11)$$
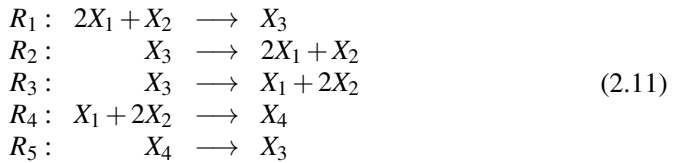
where we see that R_1 and R_2 together form a reversible reaction, and that the network is a closed system. The resulting 4×5 stoichiometric matrix for the network is

$$\mathbf{S} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \overset{\begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \end{array}}{\begin{bmatrix} -2 & 2 & 1 & -1 & 0 \\ -1 & 1 & 2 & -2 & 0 \\ 1 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}} \quad \text{with Rank}(\mathbf{S}) = 3. \qquad (2.12)$$

2.3 Complex factorization of stoichiometric matrix S

Metabolites interact in different combinations for the reactions of a network as seen in the last example of the previous section. One of the reactions in Eq. 2.11 is for instance

$$2X_1 + X_2 \longrightarrow X_3 \tag{2.13}$$

where we see that two units of X_1 and one unit of X_2 is transformed to one unit of X_3. This implies that the combination $2X_1 + X_2$ of metabolites can be thought of as a unit in this reaction. We will in this section describe the reactions of a network in terms of such units that will be called complexes. Based on this, we also do a factorization of the stoichiometric matrix S from Rao et al. [67], which is fundamental in paper B.

A *complex* $\mathcal{C}_i$ is any combination of metabolites that occur on either side of the reactions for a network. The set of complexes for a network is then the union of all left and right hand sides of the reactions. For the reactions in Eq. 2.11 we e.g. have

$$
\begin{array}{|c|c|c|c|}
\hline
\mathcal{C}_1 & \mathcal{C}_2 & \mathcal{C}_3 & \mathcal{C}_4 \\
\hline
2X_1 + X_2 & X_3 & X_1 + 2X_2 & X_4 \\
\hline
\end{array}
\tag{2.14}
$$

that can be recognized as the four different nodes of Fig. 2.3. Note that e.g. X_1 and $2X_1$ would be different complexes if they both are the left or right hand side of an equation. Also note that in the case where all the left and right hand sides of the reactions are single metabolites, the set of complexes are simply the metabolites such that $\mathcal{C}_i = X_i$. It could also be mentioned that a complex has nothing to do with complex numbers.

Assume that there are n_c complexes. These can then be represented by a *complex matrix* $\mathbf{Z} \in \mathbb{N}_0^{n \times n_c}$ where n is the number of metabolites as before. The rows of $\mathbf{Z}$ are associated with the metabolites and the columns with the complexes. Each entry is then the number of molecules for the corresponding metabolite in one unit of the corresponding complex. For the complexes of Eq. 2.14 we get the complex matrix

$$
\mathbf{Z} =
\begin{array}{c}
\\
X_1 \\
X_2 \\
X_3 \\
X_4
\end{array}
\begin{array}{c}
\mathcal{C}_1 \ \ \mathcal{C}_2 \ \ \mathcal{C}_3 \ \ \mathcal{C}_4 \\
\begin{bmatrix}
2 & 0 & 1 & 0 \\
1 & 0 & 2 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
\tag{2.15}
$$

where all the entries are non-negative since a complex can not contain a negative amount of a metabolite. Note that there is no information about the reactions of the network in $\mathbf{Z}$ other than the candidate complexes to be used in reactions. Note that in the case where $\mathcal{C}_i = X_i$ we get the identity matrix for $\mathbf{Z}$.

All the reactions of a network can now be rewritten such that the left and right hand sides are individual complexes. For the reactions in Eq. 2.11 we get

$$
\begin{array}{llll}
R_1: & \mathcal{C}_1 & \longrightarrow & \mathcal{C}_2 \\
R_2: & \mathcal{C}_2 & \longrightarrow & \mathcal{C}_1 \\
R_3: & \mathcal{C}_2 & \longrightarrow & \mathcal{C}_3 \\
R_4: & \mathcal{C}_3 & \longrightarrow & \mathcal{C}_4 \\
R_5: & \mathcal{C}_4 & \longrightarrow & \mathcal{C}_2
\end{array}
\tag{2.16}
$$

where the complexes are given by Eq. 2.14. This information can be represented by a *linkage matrix* $\mathbf{B} \in \mathbb{Z}^{n_c \times r}$ where n_c is the number of complexes and r the number of reactions as before. The linkage matrix is similar to the stoichiometric matrix $\mathbf{S}$, but the rows are associated with the complexes instead of the metabolites. The entries of $\mathbf{B}$ are then -1 if the corresponding complex is the substrate of the reaction and 1 if the complex is the product. All other entries of $\mathbf{B}$ are zero, such that there are only three possible values for an entry in $\mathbf{B}$. For the reactions in Eq. 2.16 we get

$$\mathbf{B} = \begin{array}{c} \\ \mathcal{C}_1 \\ \mathcal{C}_2 \\ \mathcal{C}_3 \\ \mathcal{C}_4 \end{array} \begin{array}{ccccc} R_1 & R_2 & R_3 & R_4 & R_5 \\ \left[\begin{array}{ccccc} -1 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right] \end{array}. \tag{2.17}$$

Note that $\mathbf{B}$ does not contain any information about which metabolites that are involved in each reaction, only which complexes. Also note that in Rao et al. [67] and paper B, only internal reactions are included in $\mathbf{B}$. It is, however, possible to also include exchange reactions in $\mathbf{B}$ as we will do for the split example of Fig. 2.2 below.

Having defined $\mathbf{Z}$ and $\mathbf{B}$, the stoichiometric matrix $\mathbf{S}$ of Eq. 2.3 can be factorized as

$$\mathbf{S} = \mathbf{ZB} \tag{2.18}$$

according to Rao et al. [67]. Be aware, however, that potential exchange reactions in $\mathbf{S}$ must also be included in $\mathbf{B}$ for Eq. 2.18 to hold. For the example of Fig. 2.3 we get

$$\mathbf{S} = \mathbf{ZB} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \tag{2.19}$$

which is easy to verify that gives the same value of the matrix $\mathbf{S}$ as given in Eq. 2.12.

For completeness, we perform the factorization of Eq. 2.18 also for the two other examples. For the loop network in Fig. 2.1 we get

$$\mathbf{S} = \mathbf{ZB} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \tag{2.20}$$

where $\mathbf{S}$ is given in Eq. 2.8. For the split network in Fig. 2.2 we get

$$\mathbf{S} = \mathbf{ZB} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix} \tag{2.21}$$

where $\mathbf{S}$ is given in Eq. 2.10. Both of the two latter examples have the identity matrix as the complex matrix $\mathbf{Z}$ since the complexes are the individual metabolites. Also note that the split example contains exchange reactions included in $\mathbf{S}$ and $\mathbf{B}$.

2.4 Topological network properties and binary stoichiometric matrix

In this section we consider $\mathbf{S}$ as a connectivity matrix and explore the elementary topological properties of a reaction network as in the textbook of Palsson [63, ch. 10].

To do so, we define the *binary stoichiometric matrix* $\hat{\mathbf{S}} = \{\hat{S}_{ij}\}$ of $\mathbf{S}$. It is given by

$$\hat{S}_{ij} = \begin{cases} 0 & \text{if } S_{ij} = 0 \\ 1 & \text{if } S_{ij} \neq 0 \end{cases} \tag{2.22}$$

such that $\hat{S}_{ij} = 1$ if and only if metabolite X_i is involved in reaction R_j. Note that the exact value of the stoichiometric coefficient S_{ij} is not important as we now define quantities to analyse elementary topological properties of reaction networks based on $\hat{\mathbf{S}}$.

The first quantity is the *participation number* for the reactions given by

$$\pi_j = \sum_{i=1}^{n} \hat{S}_{ij} \tag{2.23}$$

which gives the number of metabolites X_i that are involved in reaction R_j. The number of metabolites involved in a reaction is not affected by adding more reactions, so π_j is small for all reactions independently of the number n of metabolites in the network.

The second quantity is the *connectivity number* for the metabolites given by

$$\rho_i = \sum_{j=1}^{r} \hat{S}_{ij} \tag{2.24}$$

which gives the number of reactions R_j that the metabolite X_i is involved in. This is a measure of how connected a metabolite is, or how many links it has. Since a metabolite can participate in many reactions and appear at different positions in a network, the number ρ_i can increase with the number r of reactions in the network.

The participation numbers π_j and the connectivity numbers ρ_i can be computed directly or by adjacency matrices of $\hat{\mathbf{S}}$ that also give some additional information. By pre-multiplication of $\hat{\mathbf{S}}$ by its transpose we obtain the *reaction adjacency matrix*

$$\mathbf{A}_v = \hat{\mathbf{S}}^{\mathrm{T}}\hat{\mathbf{S}} \tag{2.25}$$

which is a symmetrical matrix where the entries are inner products of the columns of $\hat{\mathbf{S}}$. The main diagonal of $\mathbf{A}_v$ will be the participation numbers π_j, while the off-diagonal elements count how many metabolites two reactions have in common.

By post-multiplication of $\hat{\mathbf{S}}$ by its transpose we get the *compound adjacency matrix*

$$\mathbf{A}_x = \hat{\mathbf{S}}\hat{\mathbf{S}}^{\mathrm{T}} \tag{2.26}$$

which also is a symmetrical matrix, but the entries are now inner products of the rows of $\hat{\mathbf{S}}$. The main diagonal will be the connectivity numbers ρ_i, while the off-diagonal elements count how many reactions two metabolites both are involved in.

We now calculate the quantities above for the three example networks of Sec. 2.2.

Loop example

The matrix $\mathbf{S}$ for the loop network in Fig. 2.1 is given by Eq. 2.8 such that

$$\hat{\mathbf{S}} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \Rightarrow \quad \mathbf{A}_v = \hat{\mathbf{S}}^T\hat{\mathbf{S}} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \hat{\mathbf{S}}\hat{\mathbf{S}}^T = \mathbf{A}_x \tag{2.27}$$

where the equality of $\mathbf{A}_v$ and $\mathbf{A}_x$ is due to $\hat{\mathbf{S}}$ being symmetric. The main diagonals then give the participation and connectivity numbers

$$\begin{array}{|c|c|} \hline \pi_1 & \pi_2 \\ \hline 2 & 2 \\ \hline \end{array} \quad \text{and} \quad \begin{array}{|c|c|} \hline \rho_1 & \rho_2 \\ \hline 2 & 2 \\ \hline \end{array} \tag{2.28}$$

where we see that both reactions involve two metabolites, and that both metabolites are involved in two reactions. This can easily be verified visually from Fig. 2.1.

Split example

The split network in Fig. 2.2 has the matrix $\mathbf{S}$ of Eq. 2.10 with binary form

$$\hat{\mathbf{S}} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \begin{array}{c} R_1 \quad R_2 \quad R_3 \quad R_4 \quad R_5 \quad R_6 \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \end{array}. \tag{2.29}$$

The compound adjacency matrix for the metabolites then becomes

$$\mathbf{A}_x = \hat{\mathbf{S}}\hat{\mathbf{S}}^T = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \begin{array}{c} X_1 \quad X_2 \quad X_3 \quad X_4 \\ \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 3 & 1 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix} \end{array} \tag{2.30}$$

where the main diagonal gives the following connectivity numbers:

$$\begin{array}{|c|c|c|c|} \hline \rho_1 & \rho_2 & \rho_3 & \rho_4 \\ \hline 2 & 3 & 2 & 2 \\ \hline \end{array} \tag{2.31}$$

We see that X_1, X_3 and X_4 are involved in two reactions, while X_2 is involved in three. By the off-diagonal entries of $\mathbf{A}_x$ we see that X_2 has one reaction in common with each of the other X_i's, and that there is no pair of metabolites without X_2 where both are involved in the same reaction. This is easily verified by a visual inspection of Fig. 2.2.

The reaction adjacency matrix for the example will be

$$\mathbf{A}_v = \hat{\mathbf{S}}^T\hat{\mathbf{S}} = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \end{array} \begin{array}{c} R_1 \quad R_2 \quad R_3 \quad R_4 \quad R_5 \quad R_6 \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 1 & 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \end{array} \tag{2.32}$$

where the main diagonal gives the following participation numbers:

$$
\begin{array}{|c|c|c|c|c|c|}
\pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5 & \pi_6 \\
\hline
1 & 2 & 2 & 2 & 1 & 1
\end{array}
\tag{2.33}
$$

We see that R_1, R_5 and R_6 only involve one metabolite each, while the three remaining reactions involve two each. The non-zero off-diagonal entries in $\mathbf{A}_v$ are all one, and correspond to the reactions that have a metabolite in common. As for the matrix $\mathbf{A}_x$, the values of the matrix $\mathbf{A}_v$ can also easily be verified by a visual inspection of Fig. 2.2.

Complex example

The last example is the one of Fig. 2.3 with the matrix $\mathbf{S}$ of Eq. 2.12 giving

$$
\hat{\mathbf{S}} = \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \end{array}
\begin{array}{cccccc}
 & R_1 & R_2 & R_3 & R_4 & R_5 \\
\left[\begin{array}{ccccc}
1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 1
\end{array}\right].
\end{array}
\tag{2.34}
$$

This gives the compound adjacency matrix

$$
\mathbf{A}_x = \hat{\mathbf{S}}\hat{\mathbf{S}}^{\mathrm{T}} = \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \end{array}
\begin{array}{cccc}
 X_1 & X_2 & X_3 & X_4 \\
\left[\begin{array}{cccc}
4 & 4 & 3 & 1 \\
4 & 4 & 3 & 1 \\
3 & 3 & 4 & 1 \\
1 & 1 & 1 & 2
\end{array}\right]
\end{array}
\tag{2.35}
$$

where the main diagonal once again gives the connectivity numbers:

$$
\begin{array}{|c|c|c|c|}
\rho_1 & \rho_2 & \rho_3 & \rho_4 \\
\hline
4 & 4 & 4 & 2
\end{array}
\tag{2.36}
$$

We see that X_4 is involved in two reactions, while the remaining three metabolites are involved in four reactions each. From the off-diagonal entries of $\mathbf{A}_x$ we also see that all pairs of metabolites have at least one reaction in common.

The reaction adjacency matrix for the example will be

$$
\mathbf{A}_v = \hat{\mathbf{S}}^{\mathrm{T}}\hat{\mathbf{S}} = \begin{array}{c} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{array}
\begin{array}{ccccc}
 R_1 & R_2 & R_3 & R_4 & R_5 \\
\left[\begin{array}{ccccc}
3 & 3 & 3 & 2 & 1 \\
3 & 3 & 3 & 2 & 1 \\
3 & 3 & 3 & 2 & 1 \\
2 & 2 & 2 & 3 & 1 \\
1 & 1 & 1 & 1 & 2
\end{array}\right]
\end{array}
\tag{2.37}
$$

where the participation numbers from the main diagonal are the following:

$$
\begin{array}{|c|c|c|c|c|}
\pi_1 & \pi_2 & \pi_3 & \pi_4 & \pi_5 \\
\hline
3 & 3 & 3 & 3 & 2
\end{array}
\tag{2.38}
$$

 We see that R_5 involves two metabolites, and that the remaining four reactions involve three metabolites. Also, from the off-diagonal entries of $\mathbf{A}_v$ we see that all pairs of reactions have at least one metabolite in common that is involved in both reactions.

Chapter 3

Dynamical properties of reaction networks

This far we have considered the stoichiometric matrix $\mathbf{S}$ of Eq. 2.3 only as a topological map giving the network structure of a reaction network. In this chapter, however, we start to explore the dynamical properties of a reaction network. To do this, we implicitly assume a continuum hypothesis such that we can work with the concentrations of the metabolites instead of molecule counts. An assumption like this is standard, and at least valid for molar quantities as the Avogadro number for molecules per mole is $6.02 \cdot 10^{23}$ [44, p. 30]. For instance are the concentrations in paper C of this thesis in the range from 1 to 2000 nM, or in the range from $6 \cdot 10^{14}$ to $1.2 \cdot 10^{18}$ molecules per liter.

Assume that we have a network on the form of Eq. 2.1 with n compounds and r reactions. Let the vector $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^n_{\geq 0}$ be the concentrations of the compounds in the network such that x_i is the concentration of X_i. Note that all entries of $\mathbf{x}$ must be non-negative since they are concentrations. Similarly, let $\mathbf{v} = \begin{bmatrix} v_1 & \ldots & v_r \end{bmatrix} \in \mathbb{R}^r$ be the fluxes (reaction rates) of the network such that v_j is the flux of reaction R_j. If the reaction R_j is irreversible, the value of the flux v_j must be non-negative.

3.1   Dynamical equations

The concentrations $\mathbf{x}$ and the fluxes $\mathbf{v}$ are now considered dynamical quantities. It is assumed that $\mathbf{x}$ only can change through the fluxes $\mathbf{v}$, and that these changes satisfy the mass balance defined by $\mathbf{S}$. The rate of change for $\mathbf{x}$ is then given by

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v} \tag{3.1}$$

which is the fundamental equation for dynamical modelling of a reaction network [63]. Note that there are no assumptions or restrictions on the fluxes $\mathbf{v}$ in Eq. 3.1. In the next chapter we discuss various kinetic functions $\mathbf{v} = \mathbf{v}(\mathbf{x})$ such that the fluxes are given by the concentrations. This will make Eq. 3.1 into a potentially non-linear system

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) \tag{3.2}$$

of ordinary differential equations where $\mathbf{f}(\mathbf{x})$ is given by $\mathbf{S}$ and $\mathbf{v}(\mathbf{x})$. We will, however, first look on the dynamical system in the form of Eq. 3.1 and evaluate the mathematically possible dynamical behaviours. This is determined purely by the matrix $\mathbf{S}$, and in the next section we evaluate the relevant properties of $\mathbf{S}$ using standard linear algebra.

3.2   Dynamical properties of the stoichiometric matrix

In this section we analyse the stoichiometric matrix $\mathbf{S}$ from Eq. 2.3 in light of Eq. 3.1. This is done by using standard linear algebra, for an introduction see e.g. Lay [53].

The matrix $\mathbf{S}$ has four fundamental linear spaces with the following interpretations:

- $\mathrm{Null}\,(\mathbf{S})$, null space, fluxes giving steady-state

- $\mathrm{Row}\,(\mathbf{S})$, row space, fluxes giving dynamic behaviour

- $\mathrm{Null}\,(\mathbf{S}^{\mathrm{T}})$, left null space, conservation laws

- $\mathrm{Col}\,(\mathbf{S})$, column space, dynamic quantities

Before we explore the individual spaces further, remember that

$$\dim(\mathrm{Row}\,(\mathbf{S})) = \dim(\mathrm{Col}\,(\mathbf{S})) = \mathrm{Rank}\,(\mathbf{S}) \tag{3.3}$$

which motivates the calculation of the ranks for the examples in Sec. 2.2. We also have

$$\dim(\mathrm{Row}\,(\mathbf{S})) + \dim(\mathrm{Null}\,(\mathbf{S})) = r \tag{3.4}$$

where r is the number of reactions and thus also columns in $\mathbf{S}$. Similarly, we have

$$\dim(\mathrm{Col}\,(\mathbf{S})) + \dim\left(\mathrm{Null}\,(\mathbf{S}^{\mathrm{T}})\right) = n \tag{3.5}$$

where n is the number of metabolites and thus also rows in $\mathbf{S}$. To find bases for the four linear spaces, we will later make use of the row reduced echelon forms of $\mathbf{S}$ and $\mathbf{S}^{\mathrm{T}}$.

3.2.1 $\mathrm{Null}\,(\mathbf{S})$

The null space of $\mathbf{S}$ is given by the fluxes $\mathbf{v}_{\mathrm{ss}}$ that satisfy $\mathbf{S}\mathbf{v}_{\mathrm{ss}} = \mathbf{0}$, making them steady-states of the dynamical system in Eq. 3.1. This space is of particular interest since networks often are assumed to be in steady-state as in paper A and E of this thesis. The much used method of Flux balance analysis also explores this space, see Sec. 3.4.

3.2.2 $\mathrm{Row}\,(\mathbf{S})$

The row space of $\mathbf{S}$ is given by the fluxes $\mathbf{v}_{\mathrm{dyn}}$ that are linear combinations of the rows in $\mathbf{S}$, and a basis is given by the non-zero rows of the row reduced echelon form of $\mathbf{S}$.

The non-trivial vectors $\mathbf{v}_{\mathrm{dyn}}$ satisfy $\mathbf{S}\mathbf{v}_{\mathrm{dyn}} \neq \mathbf{0}$, implying that they are non-steady-states of the dynamical system in Eq. 3.1. Note that all flux vectors $\mathbf{v} \in \mathbb{R}^r$ have a unique decomposition $\mathbf{v} = \mathbf{v}_{\mathrm{dyn}} + \mathbf{v}_{\mathrm{ss}}$ where $\mathbf{v}_{\mathrm{dyn}} \in \mathrm{Row}\,(\mathbf{S})$ and $\mathbf{v}_{\mathrm{ss}} \in \mathrm{Null}\,(\mathbf{S})$ such that the fluxes can be decomposed into a dynamical part and a steady-state part. Also note that $\mathbf{S}\mathbf{v}_{\mathrm{dyn}} = \mathbf{S}\left(\mathbf{v}_{\mathrm{dyn}} + \mathbf{v}_{\mathrm{ss}}\right)$ where $\mathbf{v}_{\mathrm{ss}} \in \mathrm{Null}\,(\mathbf{S})$, such that a non-trivial $\mathbf{v}_{\mathrm{dyn}}$ is still a non-steady-state if a steady-state flux vector $\mathbf{v}_{\mathrm{ss}}$ is added to it.

3.2.3   Null $\left(\mathbf{S}^{\mathsf{T}}\right)$

The left null space of $\mathbf{S}$ is the $\mathbf{u}$ that satisfy $\mathbf{S}^{\mathsf{T}}\mathbf{u} = \mathbf{0}$. From Eq. 3.1 we then have

$$\mathbf{u}^{\mathsf{T}} \begin{bmatrix} \frac{dx_1}{dt} \\ \vdots \\ \frac{dx_n}{dt} \end{bmatrix} = 0 \quad \text{for all} \quad \frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v} \tag{3.6}$$

such that $\mathbf{u}^{\mathsf{T}}\mathbf{x} = c$ is a conservation law of Eq. 3.1 for some $c \in \mathbb{R}$. Note that it is a conservation law for concentrations, and may or may not be a conservation law for mass.

3.2.4   Col $(\mathbf{S})$

The column space of $\mathbf{S}$ is given by the $\mathbf{u}$ that are linear combinations of the columns in $\mathbf{S}$, and a basis is given by the non-zero rows of the row reduced echelon form of $\mathbf{S}^{\mathsf{T}}$.

The non-trivial $\mathbf{u} \in \mathrm{Col}\,(\mathbf{S})$ satisfy $\mathbf{S}^{\mathsf{T}}\mathbf{u} \neq \mathbf{0}$, which in contrast to Eq. 3.6 means that

$$\mathbf{u}^{\mathsf{T}} \begin{bmatrix} \frac{dx_1}{dt} \\ \vdots \\ \frac{dx_n}{dt} \end{bmatrix} \neq 0 \quad \text{for some} \quad \frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v} \tag{3.7}$$

such that $\mathbf{u}^{\mathsf{T}}\mathbf{x}$ in general is a dynamical quantity and not a conservation law of Eq. 3.1.

3.2.5   Singular value decomposition (SVD)

The Singular Value Decomposition (SVD) of the matrix $\mathbf{S}$ is a much used factorization

$$\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}} \tag{3.8}$$

of a matrix that gives bases for all the four linear spaces discussed above [78]. The Rank $(\mathbf{S})$ first columns in the matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ is a basis for Col $(\mathbf{S})$, and the remaining columns a basis for Null $\left(\mathbf{S}^{\mathsf{T}}\right)$. Similarly is the Rank $(\mathbf{S})$ first columns in $\mathbf{V} \in \mathbb{R}^{r \times r}$ a basis for Row $(\mathbf{S})$, and the remaining columns a basis for Null $(\mathbf{S})$. The matrix $\mathbf{\Sigma} \in \mathbb{R}_{\geq 0}^{n \times r}$ is diagonal with the non-negative so-called singular values in descending order on the diagonal where the number of non-zero singular values is equal to Rank $(\mathbf{S})$.

The matrix $\mathbf{S}$ of Eq. 2.8 for the loop network in Fig. 2.1 and 3.1 has SVD

$$\mathbf{S} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \right) \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \right)^{\mathsf{T}}$$

$$= \left(\frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} \right) \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} \right)^{\mathsf{T}} \tag{3.9}$$

where we see that there is only Rank $(\mathbf{S}) = 1$ non-zero singular value. Then $\mathbf{u}_1$ is a basis for Col $(\mathbf{S})$, $\mathbf{u}_2$ a basis for Null $\left(\mathbf{S}^{\mathsf{T}}\right)$, $\mathbf{v}_1$ a basis for Row $(\mathbf{S})$ and $\mathbf{v}_2$ a basis for Null $(\mathbf{S})$. See the next section for a visual interpretation of all these vectors.

It is possible to calculate the SVD for large networks as efficient algorithms exist, but the resulting factorization may not be so easy to interpret by manual inspection.

3.3   Dynamical properties of the example networks

We now calculate the linear spaces of $\mathbf{S}$ discussed above for the examples in Sec. 2.2 and interpret the resulting basis vectors visually.

3.3.1   Loop example

We start with the loop example of Fig. 2.1 that is redrawn in Fig. 3.1 with flux labels.
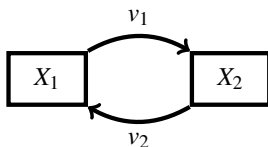
Figure 3.1: *The example network in Fig. 2.1 redrawn with labels v_j for the fluxes.*

From the SVD in Eq. 3.9 we have for the flux vectors $\mathbf{v}$ that

$$\text{Row}\,(\mathbf{S}) = \text{Span}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) \quad \text{and} \quad \text{Null}\,(\mathbf{S}) = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) \tag{3.10}$$

such that dynamic vectors $\mathbf{v}_{\text{dyn}}$ have $v_1 = -v_2$ resulting in a net flux from X_1 to X_2 or vice versa. Steady-state vectors $\mathbf{v}_{\text{ss}}$ in contrast have $v_1 = v_2$ resulting in no net flux.

For the vectors $\mathbf{u}$ we see from Eq. 3.9 that the spaces are

$$\text{Col}\,(\mathbf{S}) = \text{Span}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) \quad \text{and} \quad \text{Null}\,(\mathbf{S}^{\text{T}}) = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) \tag{3.11}$$

such that the sum $x_1 + x_2$ of the concentrations is conserved, which is an attribute of the network being closed without entering or leaving reactions as seen in Fig. 3.1. The difference $x_1 - x_2$, however, is a dynamical quantity and changes if there is a net flux.

3.3.2   Split example

As for the loop, we redraw Fig. 2.2 for the split example in Fig. 3.2 with flux labels.
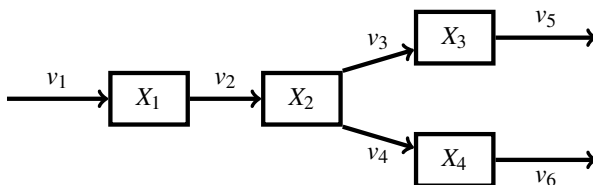
Figure 3.2: *The example network in Fig. 2.2 redrawn with labels v_j for the fluxes.*

Since we do not have the SVD of $\mathbf{S}$ we calculate the row reduced echelon forms

$$\mathbf{S} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{S}^{\mathrm{T}} \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.12}$$

to find the four linear spaces. For the flux vectors $\mathbf{v}$ this gives us the spaces

$$\mathrm{Row}\,(\mathbf{S}) = \mathrm{Span} \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} \right) \quad \text{and} \quad \mathrm{Null}\,(\mathbf{S}) = \mathrm{Span} \left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right) \tag{3.13}$$

where the steady-state basis vectors correspond to a flux through the upper or lower branch of Fig. 3.2. The row space does not have an equally obvious interpretation, other than that it gives all the non-steady-state flux vectors. For the vectors $\mathbf{u}$ we have

$$\mathrm{Col}\,(\mathbf{S}) = \mathrm{Span} \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right) \quad \text{and} \quad \mathrm{Null}\,\left(\mathbf{S}^{\mathrm{T}}\right) = \{\mathbf{0}\} \tag{3.14}$$

which implies that there are no conservation laws, and that all the concentrations are independent dynamical quantities without any restrictions.

3.3.3 Complex example

Once again, we start by redrawing Fig. 2.3 with flux labels in Fig. 3.3.
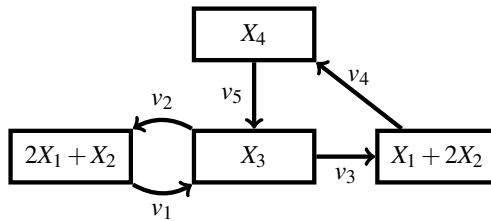
Figure 3.3: *The example network in Fig. 2.3 redrawn with labels v_j for the fluxes.*

As for the split example, we continue by calculating the row reduced echelon forms

$$\mathbf{S} \sim \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{S}^{\mathrm{T}} \sim \begin{bmatrix} 1 & 0 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & -\frac{1}{3} \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.15}$$

to find the four linear spaces. For the flux vectors $\mathbf{v}$ we get the spaces

$$\text{Row}\,(\mathbf{S}) = \text{Span}\left(\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}\right) \quad \text{and} \quad \text{Null}\,(\mathbf{S}) = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}\right) \quad (3.16)$$

where the steady-state vectors correspond to the two loops in Fig. 3.3, and the row space gives the remaining possible fluxes. For the vectors $\mathbf{u}$ we get the spaces

$$\text{Col}\,(\mathbf{S}) = \text{Span}\left(\begin{bmatrix} 3 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}\right) \quad \text{and} \quad \text{Null}\,(\mathbf{S}^\text{T}) = \text{Span}\left(\begin{bmatrix} 1 \\ 1 \\ 3 \\ 3 \end{bmatrix}\right) \quad (3.17)$$

such that the quantity $x_1 + x_2 + 3x_3 + 3x_4$ is conserved, which is a feature of the system being closed as for the loop example of Fig. 3.1. Since there are no other conservation laws, the interpretation of the column space is that any set of values for three of the concentrations may be obtained from Eq. 3.1. The fourth concentration, however, will then by given by the conservation law.

3.4 Flux balance analysis (FBA)

We continue to study the dynamical system $\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v}$ of mass balance from Eq. 3.1. In the previous section we found all the possible states and behaviours of the system. Some of these states are, however, biologically unlikely and we now want to find a biologically feasible flux vector. This leads to *flux balance analysis* (FBA) where the first papers leading up to the method were published in the 1980s [30, 64, 65, 83]. In the last 20 years, FBA has become a popular method to analyse the fluxes in biochemical networks that works well also on genome-scale networks [47, 61]. One reason for this is that FBA only requires a limited amount of information about the network, but still gives valuable information where more complicated approaches may fail.

The main assumption of FBA is that the network is assumed to be in steady-state

$$\mathbf{S}\mathbf{v}_{ss} = \mathbf{0} \tag{3.18}$$

such that the flux vector $\mathbf{v}_{ss} \in \text{Null}\,(\mathbf{S})$ has to satisfy the system of linear equations in Eq. 3.18. However, the number r of columns (reactions) in $\mathbf{S}$ is typically larger than the number n of rows (metabolites), i.e. $r > n$. For this reason, Eq. 3.18 will be underdetermined and $\text{Null}\,(\mathbf{S})$ a non-trivial space with dimension

$$\dim\,(\text{Null}\,(\mathbf{S})) = r - \text{Rank}\,(\mathbf{S}) \geq r - n \tag{3.19}$$

as we saw for the examples in Sec. 3.3.

To decide a single vector $\mathbf{v}_{ss} \in \text{Null}\,(\mathbf{S})$, more constraints then Eq. 3.18 are needed. First of all, we assume the fluxes to be bounded such that $\mathbf{v}_{ss}$ has to satisfy a set

$$\mathbf{v}_{\min} \leq \mathbf{v}_{ss} \leq \mathbf{v}_{\max} \tag{3.20}$$

of inequality constraints. In the case of irreversible reactions, $\mathbf{v}_{min} = \mathbf{0}$ which will be relevant in the next section. The upper bound $\mathbf{v}_{max}$ corresponds to capacity or saturation limits, and in the lack up such limits may be set to infinity.

The constraints of Eq. 3.20 reduce the set of feasible $\mathbf{v}_{ss}$ to a subset of Null $(\mathbf{S})$. One could then impose further constraints to further reduce the set of feasible $\mathbf{v}_{ss}$. However, this is in general still not sufficient to give a single vector. To do so, a criterion to choose between the various $\mathbf{v}_{ss}$ in the feasible set must be specified. This criterion will be based on the assumption that the cells where the reactions take place in some way are optimized. This can be to maximize the biomass yield, the ATP-yield, end-product yield or some other cellular function [63].

It is assumed that the optimality condition can be expressed as a linear objective function $\mathbf{c}^T\mathbf{v}_{ss}$ of the fluxes to be maximized where the weights $\mathbf{c} \in \mathbb{R}^r$ are known. Note that without the constraints of Eq. 3.20, $\mathbf{c}^T\mathbf{v}_{ss}$ will in general be unbounded.

With $\mathbf{v}_{min}$, $\mathbf{v}_{max}$ and $\mathbf{c}$ specified, FBA reduces to the linear program

$$
\begin{aligned}
&\max & &\mathbf{c}^T\mathbf{v}_{ss} \\
&\text{s.t.} & &\mathbf{S}\mathbf{v}_{ss} = \mathbf{0} \\
& & &\mathbf{v}_{min} \leq \mathbf{v}_{ss} \leq \mathbf{v}_{max}
\end{aligned}
\tag{3.21}
$$

that is a well studied standard optimization problem and may be solved by the Simplex method, see e.g. Vanderbei et al. [79]. Solutions are guaranteed to exist if $\mathbf{v}_{min} \leq \mathbf{0}$, and the solution is guaranteed to be bounded if $\mathbf{v}_{max}$ is finite.

In the case of degeneracy, Eq. 3.21 will not have a unique solution and we obtain Alternative optimal solutions (AOS). This is a common phenomena for genome-scale networks, and there are different approaches to address this. One of these is to apply a secondary optimization, e.g. to minimize the total flux of the solution [63].

Note that FBA does not give any information about the concentration vector $\mathbf{x}$ other than that it is constant since the network is assumed to be in steady-state. Since we are interested in modelling the concentration vector $\mathbf{x}$, the method of FBA has not been applied in the papers of the thesis. However, both paper A and E assume the network to be in steady-state such that the first constraint of Eq. 3.18 is satisfied.

As a final comment, we mention that there exist several extensions or modifications of FBA. Some of these methods are flux variability analysis (FVA) [33], dynamical flux balance [40] and RAMP [58].

3.5 Extreme pathways

In the previous section we saw that the nullspace of $\mathbf{S}$ discussed in Sec. 3.2.1 is of special interest since it gives the steady-states of the network. A basis for this space is therefore valuable and may be calculated by standard linear algebra as for the examples in Sec. 3.3. Such a basis contains $\dim(\text{Null}(\mathbf{S})) = r - \text{Rank}(\mathbf{S})$ basis vectors, and any steady-state vector $\mathbf{v}_{ss}$ may be expressed as a unique linear combination of these vectors. Each vector is a steady-state in itself that satisfies Eq. 3.18, but the individual flux vectors of the basis may not be biologically meaningful.

For irreversible reactions, we have the condition $v_j \geq 0$ such that a steady-state should satisfy $\mathbf{v}_{ss} \in \mathbb{R}^r_{\geq 0}$ if all the reactions are irreversible which is a common assump-

tion. The set of feasible steady-state flux vectors $\mathbf{v}_{ss}$ then becomes the convex space

$$\mathbf{S}\mathbf{v}_{ss} = 0 \quad \text{and} \quad \mathbf{v}_{ss} \in \mathbb{R}^r_{\geq 0} \, . \tag{3.22}$$

However, the basis vectors for Null$(\mathbf{S})$ may have negative entries. This violates the non-negativity condition such that a basis vector in itself may be infeasible as discussed above. For this reason is it beneficial to use a set of convex basis vectors $\mathbf{p}_k$ when all the reactions are irreversible. Each of these fulfils the non-negativity condition $\mathbf{p}_k \in \mathbb{R}^r_{\geq 0}$ and is a feasible vector. All steady-states $\mathbf{v}_{ss}$ can then be written as a linear combination

$$\mathbf{v}_{ss} = \sum_k \gamma_k \mathbf{p}_k \quad \text{where } \gamma_k \geq 0 \text{ for all } k. \tag{3.23}$$

The convex basis vectors $\mathbf{p}_k$ are called *extreme pathways*, and are unique up to scaling [63]. The different $\mathbf{p}_k$ give the the possible balanced flux vectors in a network, and are useful for the analysis of possible steady-states. However, the number of $\mathbf{p}_k$ is typically much greater than the dimension of Null$(\mathbf{S})$, and grows much faster than the network size due to a combinatorial nature. A general algorithm for calculation of the extreme pathways is presented in Schilling et al. [74], but scaling issues make it hard to calculate the full set for large networks as the computation is N-P hard.

We now calculate the extreme pathways for the examples in Sec. 3.3 and paper A.

Loop example

For the loop example of Fig. 3.1 we only get one extreme pathway

$$\mathbf{p}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{3.24}$$

illustrated in Fig. 3.4 which is the same as the basis vector for the null space.
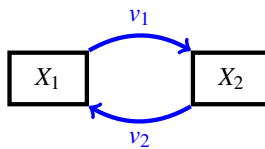
Figure 3.4: *The only extreme pathway for the example network in Fig. 3.1. Both fluxes have dark blue arrows indicating that they are non-zero.*

Split example

The split example of Fig. 3.2 results in the two extreme pathways

$$\begin{bmatrix} \mathbf{p}_1^{\mathrm{T}} \\ \mathbf{p}_2^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{3.25}$$

illustrated in Fig. 3.5. We see that the extreme pathways once again are the same as the regular basis and correspond to the upper or lower branch of the network, respectively.
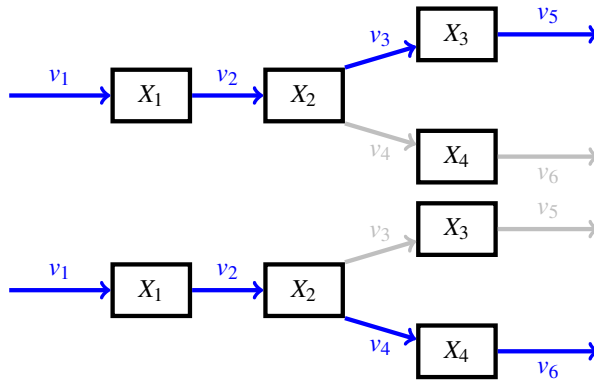
Figure 3.5: *The two extreme pathways for the example network in Fig. 3.2. The dark blue arrows indicate non-zero fluxes, and the light gray arrows indicate fluxes with value zero.*

Complex example

For the complex example of Fig. 3.3 we once again get two extreme pathways

$$\begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{3.26}$$

illustrated in Fig. 3.6. The two extreme pathways are also here the same as the regular basis for the null space, and correspond to the two loops of the network.
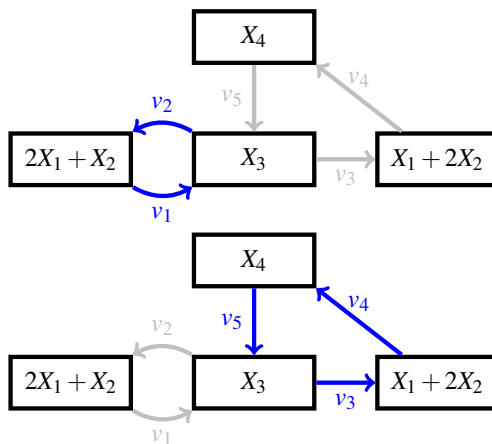
Figure 3.6: *The two extreme pathways for the complex example in Fig. 3.3. The dark blue arrows indicate non-zero fluxes, and the light gray arrows indicate fluxes with value zero.*

Example from paper A

For each of the three examples considered above we saw that the extreme pathways simply were the regular basis vectors for the null space as they did not contain any

negative values. To illustrate the proposed combinatorial behaviour of the extreme pathways on a small scale example, we use the example of paper A shown in Fig. 3.7.
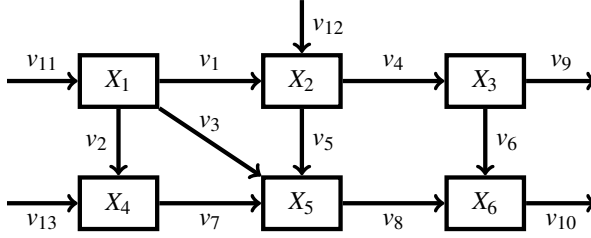
Figure 3.7: *Example network from paper A. The nodes are metabolites with names X_j and the directed edges are irreversible reactions with reaction rates (fluxes) v_j.*

This example has the 6×13 stoichiometric matrix

$$
\mathbf{S} = \begin{bmatrix}
-1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0
\end{bmatrix} \tag{3.27}
$$

with $\mathrm{Rank}\,(\mathbf{S}) = 6$ such that $\dim\,(\mathrm{Null}\,(\mathbf{S})) = 7$. A regular basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_7\}$ for $\mathrm{Null}\,(\mathbf{S})$ is then given by seven vectors where the natural choice is the rows

$$
\begin{bmatrix} \mathbf{v}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{v}_7^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix}
1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{bmatrix} \tag{3.28}
$$

of this matrix where we see that there are several negative entries. By the application of Schilling et al. [74], we obtain the nine extreme pathways $\mathbf{p}_k$ given by

$$
\begin{bmatrix} \mathbf{p}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{p}_9^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1
\end{bmatrix} \tag{3.29}
$$

where each row of the matrix is an extreme pathway. All the extreme pathways are shown in Fig. 3.8 where blue is a non-zero flux and light gray a flux with value zero.
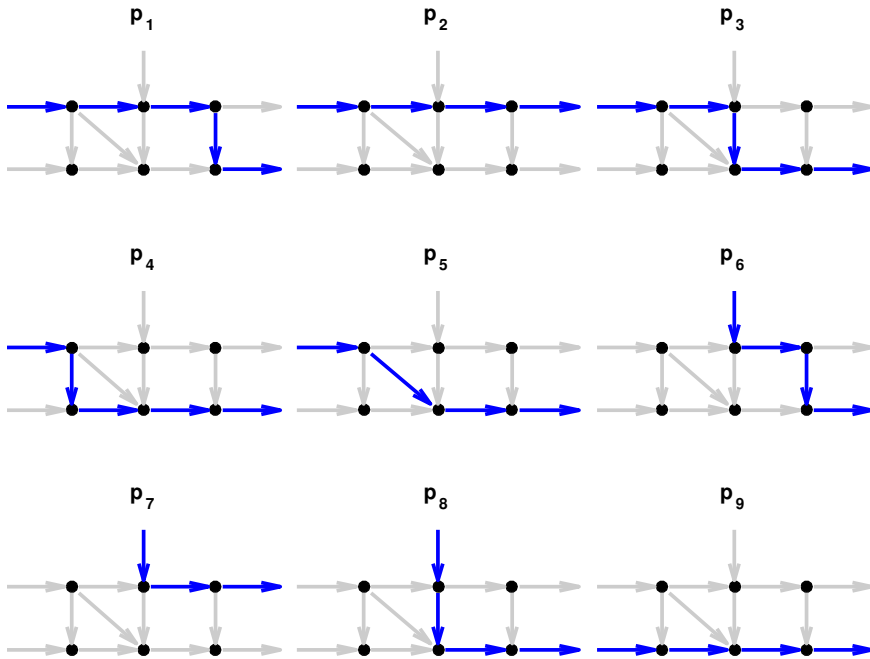
Figure 3.8: *The nine extreme pathways* $\mathbf{p}_k$ *in Eq. 3.29 for the example in Fig. 3.7. The dark blue arrows are non-zero fluxes and the light gray arrows fluxes with value zero.*

We see that there are nine extreme pathways in contrast to the dimension seven of the null space. Since there are no loops in the network of Fig. 3.7, we see in Fig. 3.8 that each of the extreme pathways is a route through the network.

3.5.1 Relation to active reactions in paper A and E

In paper A and E we classify the reactions as active or non-active under the different input fluxes, which can be done by the use of extreme pathways. To illustrate this, we continue to use the example of paper A in Fig. 3.7.

The last three fluxes in Fig. 3.7 are input fluxes denoted by b_j in paper A such that

$$\mathbf{v} = \begin{bmatrix} v_1 & \ldots & v_{10} & b_1 & b_2 & b_3 \end{bmatrix}^{\mathrm{T}} \tag{3.30}$$

is the full flux vector. Since there are no loops in the network, all extreme pathways $\mathbf{p}_k$ have exactly one non-zero input flux and can be associated with a unique input b_j. Let then P_j be the set of extreme pathways $\mathbf{p}_k$ where the j'th input b_j is non-zero. For the network in Fig. 3.7 we have the extreme pathways of Eq. 3.29 shown in Fig. 3.8 and see that $P_1 = \{\mathbf{p}_1, \ldots, \mathbf{p}_5\}$, $P_2 = \{\mathbf{p}_6, \ldots, \mathbf{p}_8\}$ and $P_3 = \{\mathbf{p}_9\}$. If we represent each set P_j

with a matrix $\mathbf{P}_j$ containing its extreme pathways $\mathbf{p}_k$'s we get

$$
\begin{aligned}
\mathbf{P}_1 &= \begin{bmatrix} \mathbf{p}_1^T \\ \vdots \\ \mathbf{p}_5^T \end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0
\end{bmatrix}, \\[4pt]
\mathbf{P}_2 &= \begin{bmatrix} \mathbf{p}_6^T \\ \vdots \\ \mathbf{p}_8^T \end{bmatrix} = \begin{bmatrix}
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0
\end{bmatrix} \text{ and} \\[4pt]
\mathbf{P}_3 &= \mathbf{p}_9^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}
\end{aligned}
\tag{3.31}
$$

where the rows of $\mathbf{P}_j$ are the vectors in the set P_j. The last three columns correspond to the input fluxes $\mathbf{b} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}^T$. Note that all $b_j = 1$ in $\mathbf{P}_j$.

The set of the active reactions for a given input b_j can now be found as the non-zero columns of $\mathbf{P}_j$ in Eq. 3.31 with black color. All the resulting sets are shown in Fig. 3.9, and are in accordance with the active reactions found in paper A.
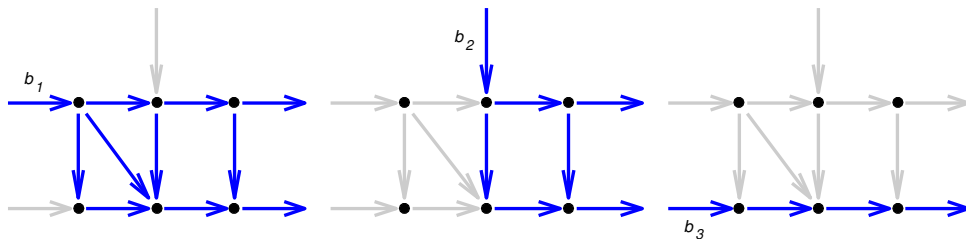
Figure 3.9: *Fig. 3 of paper A. The active reactions are shown in dark blue and the non-active reactions in light gray for each of the input fluxes b_j.*

Chapter 4

Kinetics and model reduction

In this chapter we leave the formulation $\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v}$ of Eq. 3.1 for the dynamical system used to model the concentrations and fluxes in the network. This is done by assuming that the fluxes are given by the concentrations by some function $\mathbf{v}(\mathbf{x})$ called *kinetics*. The dynamical system of Eq. 3.1 for the network then becomes

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}) \quad \text{where} \quad \mathbf{f}(\mathbf{x}) = \mathbf{S}\mathbf{v} \quad \text{and} \quad \mathbf{v} = \mathbf{v}(\mathbf{x}) , \tag{4.1}$$

which is a system of ordinary differential equations. Even though $\mathbf{v}$ is not explicitly visible in the differential equations of Eq. 4.1, the system is still $\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v}$ from Eq. 3.1 such that the discussion in the previous chapter about the dynamical properties of the network is still valid. A function $\mathbf{v}(\mathbf{x})$ may, however, limit which states of the network that could be attained by the dynamical system of Eq. 4.1.

The study of kinetics dates back to 1864 and the publication "Studies concerning affinity" by the Norwegian scientists Cato M. Guldberg and Peter Waage [34]. They proposed what today is known as the law of mass action that is used to model the reaction rate of an elementary reaction [57]. Several of the kinetics proposed in the next section can be derived from this principle, but the law of mass action may not be applied directly as we do not model all intermediates and co-factors of the chemical reactions. It is also worth mentioning that the model equations of Eq. 4.1 have much in common with compartmental models, see e.g. Anderson [10] for an introduction.

4.1   Kinetics

We now discuss some of the most common functions $\mathbf{v}(\mathbf{x})$ for the kinetics of reactions. A dynamical model for a network may use different kinetics for the various reactions, and we have used several of them in the papers of this thesis.

4.1.1   Michaelis-Menten kinetics

The first kinetics we look at is the Michaelis-Menten [45] for enzymatic reactions

$$\text{E+S} \quad \underset{k_b}{\overset{k_f}{\rightleftharpoons}} \quad \text{ES} \quad \xrightarrow{k_{\text{cat}}} \quad \text{E+P} \tag{4.2}$$
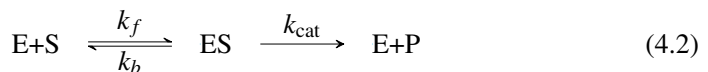
where S is a substrate, E an enzyme, ES an intermediate and P a product. The obtained reaction rate will essentially simplify the chemical reaction of Eq. 4.2 to the reaction

$$S \longrightarrow P \tag{4.3}$$

without intermediates and enzymes such that only the concentrations of S and P are explicit variables in the model.

Let [E], [S], [P] and [ES] denote the concentrations of E, S, P and ES, respectively. By the law of mass action [57] for the elementary reactions of Eq. 4.2, we obtain

$$\frac{d[\text{E}]}{dt} = -k_f[\text{E}][\text{S}] + k_b[\text{ES}] + k_{\text{cat}}[\text{ES}] \tag{4.4}$$

$$\frac{d[\text{S}]}{dt} = -k_f[\text{E}][\text{S}] + k_b[\text{ES}] \tag{4.5}$$

$$\frac{d[\text{P}]}{dt} = k_{\text{cat}}[\text{ES}] \tag{4.6}$$

$$\frac{d[\text{ES}]}{dt} = k_f[\text{E}][\text{S}] - k_b[\text{ES}] - k_{\text{cat}}[\text{ES}] \tag{4.7}$$

for the concentrations where k_f, k_b and k_{cat} are constants. By adding the derivatives for the concentrations [E] and [ES] in Eq. 4.4 and 4.7 together, we get the derivative

$$\frac{d}{dt}([\text{E}] + [\text{ES}]) = 0 \tag{4.8}$$

such that the total amount of enzyme is constant. This gives a conservation law

$$[\text{E}] + [\text{ES}] = E_0 \tag{4.9}$$

for some constant $E_0 \in \mathbb{R}_{\geq 0}$. It is further assumed that the substrate S and the complex ES is in instantaneous chemical equilibrium. This gives the relation

$$k_f[\text{E}][\text{S}] = k_b[\text{ES}] \tag{4.10}$$

and by substituting for [E] from Eq. 4.9 into Eq. 4.10 we obtain

$$[\text{ES}] = \frac{E_0[\text{S}]}{\frac{k_b}{k_f} + [\text{S}]} \tag{4.11}$$

which inserted into Eq. 4.6 gives the production rate

$$\frac{d[\text{P}]}{dt} = k_{\text{cat}}[\text{ES}] = \frac{k_{\text{cat}}E_0[\text{S}]}{\frac{k_b}{k_f} + [\text{S}]} \tag{4.12}$$

for the product P of the chemical reaction in Eq. 4.2, which will be the flux v of the simplified chemical equation in Eq. 4.3.

Let $V_{\text{max}} = k_{\text{cat}}E_0$ and $K_M = \frac{k_b}{k_f}$ be constants given by the previously used constants. The flux v from Eq. 4.12 is then a function of the substrate concentration [S] given by

$$v = \frac{d[\text{P}]}{dt} = \frac{V_{\text{max}}[\text{S}]}{K_M + [\text{S}]} \tag{4.13}$$
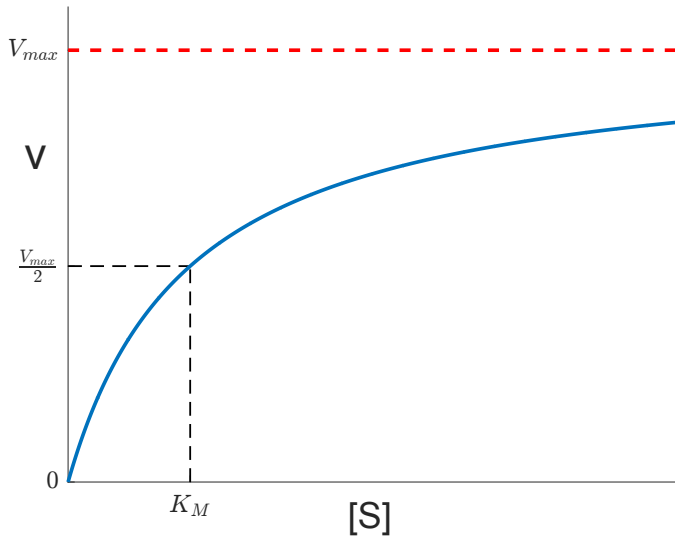
Figure 4.1: *The flux value v as a function of the substrate concentration* [S] *for the Michaelis-Menten kinetics given by Eq. 4.13. For* [S] = K_M *is the flux* $v = \frac{V_{max}}{2}$.

where the constants V_{max} and K_M must be positive.

V_{max} is the maximum flux, and K_M is the concentration [S] that gives a flux

$$v\left(K_M\right) = \frac{V_{\text{max}}}{2} \tag{4.14}$$

of half the maximal flux value such that K_M could be regarded a typical concentration e.g. if the equations were to be made non-dimensional. Also note the properties

$$v\left(0\right) = 0 \quad \text{and} \quad \lim_{[S] \to \infty} v\left([S]\right) = V_{\text{max}} \tag{4.15}$$

that can also be seen in Fig. 4.1. The function for *v* is a linear curve for sufficiently small values of [S] where there is enough enzyme and the amount of substrate is limiting for the reaction. For large values of [S], however, the limiting factor is the amount of enzyme and not the amount of substrate such that the function for *v* becomes flat. The information about the enzyme is intrinsic in the constants K_M and V_{max}, and we see that the transition from the linear to the constant regime is happening close to the substrate concentration [S] = K_M. If the Michaelis-Menten kinetics is used for the reaction S → P of Eq. 4.3 with only a substrate S and a product P, the model will have the rate

$$\frac{d\left[\text{S}\right]}{dt} = -\frac{V_{\text{max}}\left[\text{S}\right]}{K_M + \left[\text{S}\right]} \tag{4.16}$$

for the consumption of S, in addition to Eq. 4.13 for the production rate of P.

Several of the reactions in the yeast glycosis example model of paper B use Michealis-Menten kinetics or a slightly modified version of it. As a final comment on Michealis-Menten, we also mention that V_{max} and K_M values for some reactions can be found in databases such as BRENDA [71] and SABIO-RK [84].

4.1.2   Hill kinetics

A generalization of the Michaelis-Menten kinetics is the Hill equation [39] given by

$$v = \frac{d\,[\text{P}]}{dt} = \frac{V_{\max}\,[\text{S}]^n}{K_D + [\text{S}]^n} \tag{4.17}$$

where $V_{\max}$, K_D and n are positive constants. The Hill kinetics shares the properties

$$v(0) = 0 \quad \text{and} \quad \lim_{[\text{S}] \to \infty} v\,([\text{S}]) = V_{\max} \tag{4.18}$$

with the Michaelis-Menten kinetics for all n. It can be non-dimensionalized as

$$\frac{v}{V_{\max}} = \frac{\left(\frac{[\text{S}]}{K_A} \right)^n}{1 + \left(\frac{[\text{S}]}{K_A} \right)^n} \tag{4.19}$$

where $(K_A)^n = K_D$. We see that for $[\text{S}] = K_A$ is the flux

$$v(K_A) = \frac{V_{\max}}{2} \quad \text{for all } n \tag{4.20}$$

such that K_A takes the role of K_M in the Michaelis-Menten kinetics.
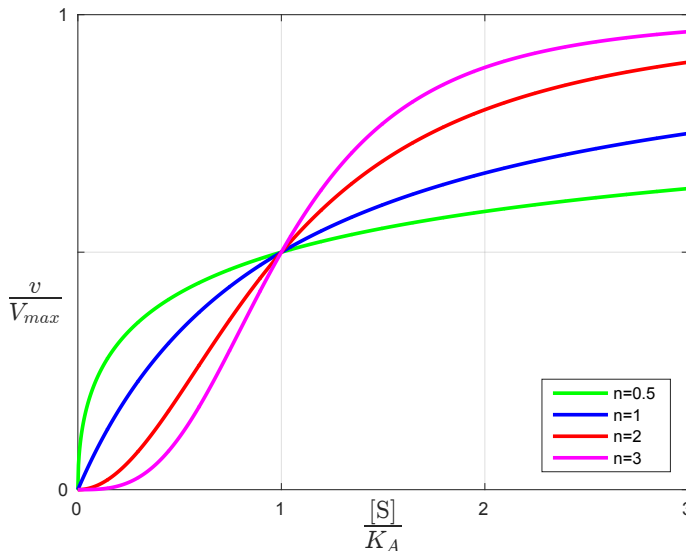
Figure 4.2: *The flux v as a function of the substrate concentration* $[S]$ *for the non-dimensional Hill kinetics in Eq. 4.19 and different n. For* $[S] = K_A$ *is the flux* $v = \frac{V_{max}}{2}$ *for all n.*

The non-dimensional Hill kinetics of Eq. 4.19 is plotted in Fig. 4.2 for different values of n where $n = 1$ gives the Michaelis-Menten kinetics of Fig. 4.1.

4.1.3 Zero and first order kinetics

The last kinetics functions that we will discuss are zero and first order kinetics, which are used for the fluxes in paper A, C and E. They are also used for the small example in paper B. A zero order flux v is assumed to be constant such that

$$v = k \qquad (4.21)$$

for some constant k. This is a reasonable assumption for fluxes that are saturated, and also for fluxes entering a network from the outside. A first order flux is assumed to be

$$v = kx \qquad (4.22)$$

where k is a constant and x is the substrate concentration in accordance with the notation of the previous chapters. The assumption of a first order flux is reasonable if the substrate is the limiting factor of the reaction such that it is far from being saturated. This is the case for the Michealis-Menten kinetics of Fig. 4.1 for low substrate concentrations, and a first order flux may be used to approximate it in the linear regime.

Remember that the Michaelis-Menten kinetics of Eq. 4.13 has two parameters, while the first order kinetics of Eq. 4.22 only has one. For this reason, it may be desirable to use first order kinetics to decrease the number of parameters in a model.

The parameters k in Eq. 4.21 and Eq. 4.22 are called *kinetic parameters*, and are the ones that we want to estimate in this thesis. In paper C we do this for first order kinetic parameters in a model of steroidogenesis using real data from an experiment [8]. In paper A and E we study the identifiability of first order kinetic parameters denoted by θ instead of k under given zero order input fluxes denoted by b instead of k. See the next chapter for more about parameter estimation.

4.2 Model reduction

The model equations on the form of Eq. 4.1 for a biochemical reaction network may have a high degree of complexity, especially if more advanced kinetics than the ones mention above are applied. For this reason, the model may possess complicated dynamics and be difficult to analyse. One particularly challenging situation is if the model is highly sensitive to parameter values and at the same time contains a large number of parameters that need to be estimated. This could e.g. lead to an overfitted model [38]. If the model is too complicated for this or other reasons, it could be beneficial to apply some kind of model reduction to reduce its complexity. A dynamical model of a reaction network should, however, be of sufficient complexity to be able to represent important dynamical features of the network. To guarantee this, a potential model reduction should be performed in a controlled manner that does not reduce too much. This is typically done by requiring that the output of the reduced model is sufficiently close to the output of the original model by some defined measure.

One common way to reduce a dynamical model of high complexity is by model order reduction, for a general introduction see e.g. Schilders et al. [73]. A model order reduction reduces the dimension of the state space, in our case the dimension of the concentration vector **x**. This implies that some of the assumptions previously made in this thesis could actually be viewed as a model order reduction. In Sec. 1.2.3 about

chemical reactions, we discussed the various types of chemical reactions that we all model by the irreversible reaction $X_1 \longrightarrow X_2$ of Eq. 1.4. By leaving the intermediates and co-factors of the reactions out of the model, this represents a model order reduction discussed in e.g. Feliu and Wiuf [29] and Sáez et al. [70]. The second part of removing intermediates in the chemical reactions from the model, is to obtain rate laws for the simplified reactions used in the model. All of the kinetics discussed in Sec. 4.1 are mathematically suitable for this, and the derivation of the Michaelis-Menten kinetics in Sec. 4.1.1 justifies its biological applicability. It could also be viewed as a model reduction to apply a simplified kinetics like the zero or first order kinetics in Sec. 4.1.3 to decrease the number of parameters in the model.

Assume now that we have a given network and a dynamical model on the form of Eq. 4.1 that we somehow want to reduce. To perform the model reduction, we need a method to reduce the model and a method to compare the reduced model to the original model. There are several ways to do this with the most common being lumping, sensitivity analysis and time-scale analysis [60, 66, 75]. However, we focus on the reduction method presented in Rao et al. [67] that paper B is an extension of. The method applies to a large class of kinetics, including the ones discussed in Sec. 4.1.

The reduction method of Rao et al. [67] starts by specifying a set $\mathcal{M}_I$ of important compounds, e.g. the compounds that are possible to measure the concentration of. The concentrations of these important compounds will be used to evaluate the reduced models. Next, all the complexes (see Sec. 2.3) of the network are divided into two categories. The first category will be all the complexes containing at least one of the important compounds $\mathcal{M}_I$, and these complexes are not considered for reduction. The second category is the remaining complexes not containing any of the important compounds $\mathcal{M}_I$, and are the complexes considered for reduction.

The method assumes that the model reaches an asymptotically stable steady-state that can be found be integrating the system for a long enough time. A complex is then reduced by setting the concentration constant equal to its steady-state value, which can be done simultaneously for any given set of complexes. For a discussion about the possibility of multiple steady-states in a model such that the method can not be applied, see e.g. Conradi et al. [18]. Being able to reduce the model for any given set of complexes, Rao et al. [67] presents an iterative procedure to select which complexes to reduce. At each step, the complex yielding the smallest error between the reduced and original model is added to the set of reduced complexes until a certain value for the error is reached and the process stopped. The error is computed by the integral

$$I_T\left(\mathbf{x}_r, \mathbf{x}_f\right) = \sum_{i \in \mathcal{M}_I} \frac{1}{T n\left(\mathcal{M}_I\right)} \int_0^T \left|1 - \frac{x_{ir}(t)}{x_{if}(t)}\right| dt \qquad (4.23)$$

where T is length of the time interval considered, $n\left(\mathcal{M}_I\right)$ is the number of compounds in $\mathcal{M}_I$, $x_{ir}(t)$ is the concentration of compound number i in the reduced model with vector notation $\mathbf{x}_r$ and $x_{if}(t)$ with vector notation $\mathbf{x}_f$ is the same for the full model. This gives the relative difference between the concentrations of the important compounds $\mathcal{M}_I$ in the reduced and original model over the given time interval $[0, T]$.

One drawback of the method of Rao et al. [67] is that it does not consider parameter uncertainty as the error in Eq. 4.23 only is computed for a single parameter set. In

paper B [31] we address this, and start by defining the symmetric error measure

$$E_T\left(\mathbf{x}_1, \mathbf{x}_2\right) = \frac{1}{2}\left(I_T\left(\mathbf{x}_1, \mathbf{x}_2\right) + I_T\left(\mathbf{x}_2, \mathbf{x}_1\right)\right) \tag{4.24}$$

based on Eq. 4.23 that can be computed for any two models without favouring one of them. Our method assumes that we have a given number of parameter sets, and that all possible reduced models are calculated for all parameter sets. By using single linkage clustering and Eq. 4.24 as dissimilarity measure, we cluster the original and reduced models for all the parameter sets which can be visualized in a dendrogram [43] like the ones shown in Fig. 4.4. A given reduction will then have some distribution throughout the dendrogram for the various parameter sets that can be visualized by coloring.

To evaluate a reduction, we use a Kolmogorov-Smirnov (KS) test [17] for the distributions of the reduction and the original model. Reductions with a test score below some threshold are said to be consistent with the original model. The best reduced model is then chosen to be the consistent model that is reduced the most.
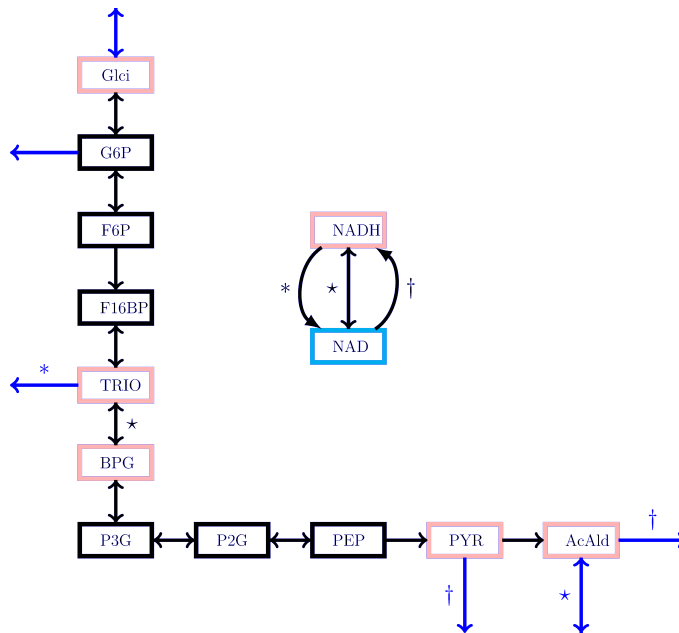
Figure 4.3: *Fig. 2 of paper B, the yeast glycosis example network from Rao et al. [67]. The nodes are compounds and the arrows reactions. The pink compounds are the important compounds not considered for reduction, while the black ones are candidates for reduction. For a more detailed description of the figure, see the original figure in paper B.*

To test the method, the yeast glycosis example network of Fig. 4.3 from Rao et al. [67] is considered in paper B with the resulting dendrograms in Fig. 4.4. Each panel of Fig. 4.4 is the result for a sample of 100 parameter sets from a log-normal distribution. The expectation of each parameter is equal to its reference value, and the standard deviation of each parameter is equal to its reference value divided by a scaling factor. We see that for high parameter uncertainty (scaling 3) the original models are distributed

throughout the whole dendrogram such that all models are equally good and the model may be maximally reduced. For low parameter uncertainty (scaling 100) the original models are clustered closely together and the model should not be reduced much. This illustrates the need to consider parameter uncertainty when doing model reduction to obtain a suitable model complexity compared to the parameter uncertainty.

It is also shown for a toy example in paper B that it could sometimes give a lower error to reduce two complexes than only one, such that the iterative method of Rao et al. [67] may not reduce the model even though an acceptable reduction exists.
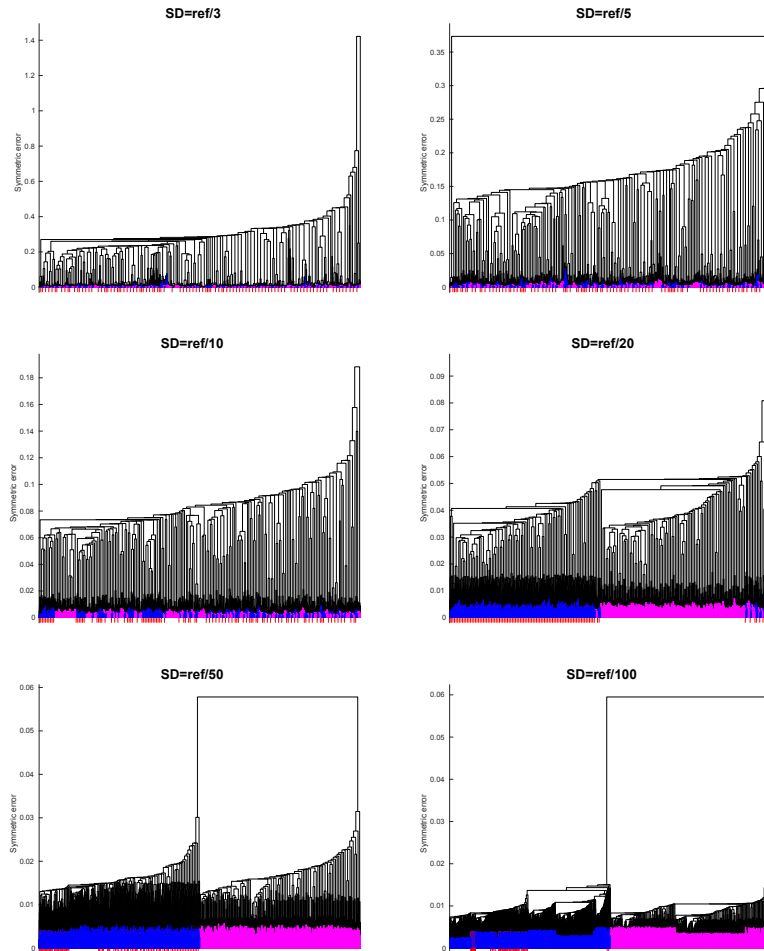
Figure 4.4: *Fig. 5 of paper B. Dendrograms for the example in Fig. 4.3 where parameters were sampled from a log-normal distribution with standard deviation equal the reference value of the parameter divided by the scaling factor given in the title of each panel. The original models are colored in red, models where F16BP is reduced in purple and all other models in blue.*

Chapter 5

Parameter estimation and identifiability

A dynamical model for the concentrations in a reaction network contains several kinetic parameters defined by the kinetics. If $\boldsymbol{\theta}$ is the parameter vector, Eq. 4.1 can be written

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \tag{5.1}$$

to make the dependence on the parameters explicit. For some reactions, the parameter values from a controlled experiment can be found in databases such as BRENDA [71] and SABIO-RK [84]. However, the parameters must in general be estimated if a dynamical model for a reaction network is to be made. This is done in paper C for a model of steroidogenesis, and we start the chapter by looking at this example.

5.1 Parameter estimation in paper C

The steroidogenesis pathway [6] discussed in Sec. 1.1.3 is studied in paper C [8] for forskolin stimulated cells. It consists of the 19 steroids shown as boxes in Fig. 5.1. In the experiment of paper C, cells were exposed to 1.5μM forskolin. The concentrations of the steroids were then measured at the nine different time points t_k of Tab. 5.1 after the forskolin exposure. We see that the time span of the whole experiment is 72 hours, and that the measurements are more frequent in the first part of the experiment. The concentrations were measured for three independent biological replicates with two technical replicates each such that we have six observations for each concentration. However, some values were undetected and are considered as missing in the following.

k	0	1	2	3	4	5	6	7	8
t_k [h]	0	2	4	6	12	24	36	48	72

Table 5.1: *The nine time points t_k for the measurements of the steroids with unit hours.*

For each of the steroids, the average concentration at each time point is calculated where missing values are omitted from the calculations. The resulting average concentrations for all the steroids except Pregnenolone are shown as the circles in Fig. 5.2. We see that some of the early data points e.g. for Estriol are missing, which implies that all the observations are missing for the steroid at that time point.
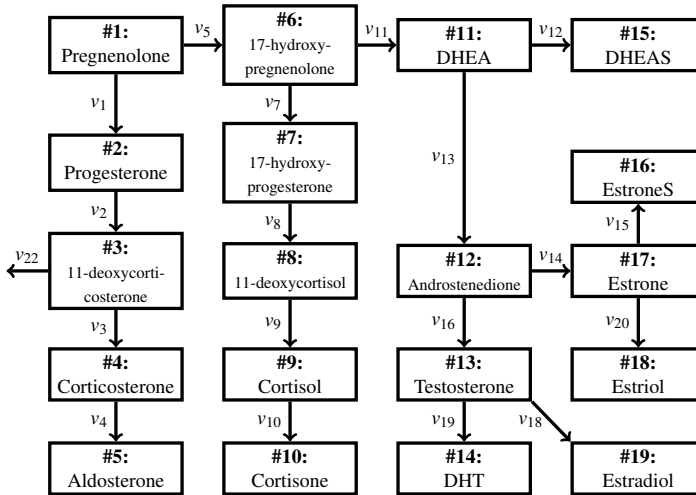
Figure 5.1: *A redrawing of Fig. 4 in paper C showing the dynamical model for steroidogenesis. The boxes are steroids and the arrows reactions with reaction rates (fluxes) v_j. Some reaction indices are not used since some fluxes are removed in paper C, see the manuscript for details.*
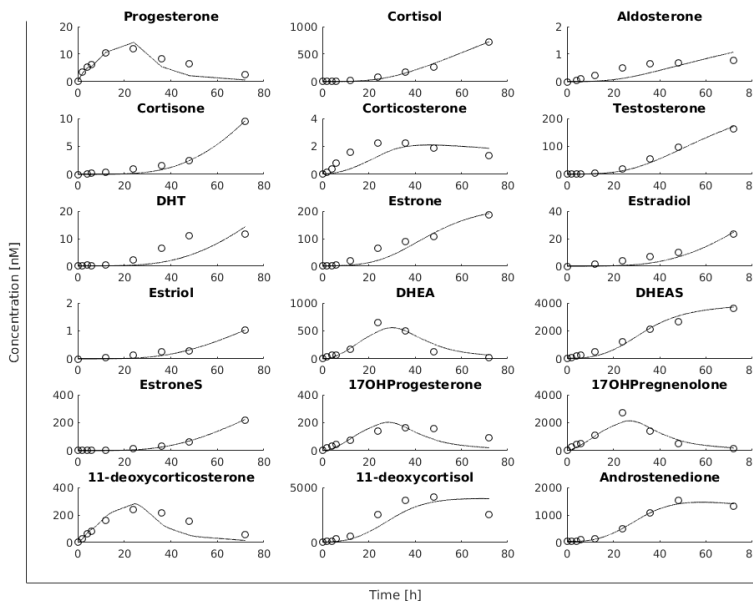
Figure 5.2: *Fig. 5 in paper C. The circles are the average measured concentrations at the time points of Tab. 5.1 and the lines the values from the dynamical model of Eq. 5.2 with the parameter values of Tab. 5.2. The unit of the time is hours and the unit of the concentrations is nanomolar, i.e. $nM = 10^{-9}$ mol/L.*

The dynamical model of Fig. 5.1 for steroidogenesis is proposed in paper C, and studied further in paper E using the results of paper A.

Since there are no inputs to the system, the concentration of Pregnenolone is modelled as given and interpolated from the average concentrations. The remaining concentrations are modelled by a dynamical system on the form discussed in the previous chapters, where the fluxes v_j are given by the first order kinetics of Eq. 4.22. This gives

$$
\begin{aligned}
\frac{dx_2}{dt} &= v_1 - v_2 & &= \theta_1 x_1 - \theta_2 x_2 \\
\frac{dx_3}{dt} &= v_2 - v_3 - v_{22} &= \theta_2 x_2 - \theta_3 x_3 - \theta_{22} x_3 \\
\frac{dx_4}{dt} &= v_3 - v_4 & &= \theta_3 x_3 - \theta_4 x_4 \\
\frac{dx_5}{dt} &= v_4 & &= \theta_4 x_4 \\
\frac{dx_6}{dt} &= v_5 - v_7 - v_{11} &= \theta_5 x_1 - \theta_7 x_6 - \theta_{11} x_6 \\
\frac{dx_7}{dt} &= v_7 - v_8 & &= \theta_7 x_6 - \theta_8 x_7 \\
\frac{dx_8}{dt} &= v_8 - v_9 & &= \theta_8 x_7 - \theta_9 x_8 \\
\frac{dx_9}{dt} &= v_9 - v_{10} & &= \theta_9 x_8 - \theta_{10} x_9 \\
\frac{dx_{10}}{dt} &= v_{10} & &= \theta_{10} x_9 \\
\frac{dx_{11}}{dt} &= v_{11} - v_{12} - v_{13} &= \theta_{11} x_6 - \theta_{12} x_{11} - \theta_{13} x_{11} \\
\frac{dx_{12}}{dt} &= v_{13} - v_{14} - v_{16} &= \theta_{13} x_{11} - \theta_{14} x_{12} - \theta_{16} x_{12} \\
\frac{dx_{13}}{dt} &= v_{16} - v_{18} - v_{19} &= \theta_{16} x_{12} - \theta_{18} x_{13} - \theta_{19} x_{13} \\
\frac{dx_{14}}{dt} &= v_{19} & &= \theta_{19} x_{13} \\
\frac{dx_{15}}{dt} &= v_{12} & &= \theta_{12} x_{11} \\
\frac{dx_{16}}{dt} &= v_{15} & &= \theta_{15} x_{17} \\
\frac{dx_{17}}{dt} &= v_{14} - v_{15} - v_{20} &= \theta_{14} x_{12} - \theta_{15} x_{17} - \theta_{20} x_{17} \\
\frac{dx_{18}}{dt} &= v_{20} & &= \theta_{20} x_{17} \\
\frac{dx_{19}}{dt} &= v_{18} & &= \theta_{18} x_{13}
\end{aligned}
\tag{5.2}
$$

as model equations where x_i is the concentration of steroid i, and the various θ_j are the kinetic parameters to be estimated. Note that k is used instead of θ in paper C.

Let now $\bar{x}_i(t_k)$ be the average measured concentration of steroid i at time t_k. To run the dynamical model of Eq. 5.2 we use the initial values

$$
x_i(t_0) = \bar{x}_i(t_0)
\tag{5.3}
$$

where the initial value for a steroid is set to zero if all the measurements were missing. Further, let $\hat{x}_i$ be the maximal average concentration for steroid i given by

$$
\hat{x}_i = \max_k \{\bar{x}_i(t_k)\}
\tag{5.4}
$$

which is used as normalization constant in paper C. Note that paper E also uses similar normalization constants, but these are given by the final dynamical model of paper C.

The estimated values $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ in paper C are by the minimization

$$
\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\arg\min} \sum_{\substack{k=1 \\ \bar{x}_i(t_k) \text{ exists}}}^{8} \sum_{i=2}^{19} \left(\frac{x_i(t_k) - \bar{x}_i(t_k)}{\hat{x}_i} \right)^2
\tag{5.5}
$$

of a weighted least squares objective function. The minimization was performed numerically by using the package TMB in R [51] which calculates point estimates and standard errors for the estimates. The resulting values are shown in Tab. 5.2.

	Estimate [h^{-1}]	Error [h^{-1}]
θ_1	$2.94 \cdot 10^{-1}$	$1.02 \cdot 10^{0}$
θ_2	$1.24 \cdot 10^{1}$	$4.36 \cdot 10^{1}$
θ_3	$3.43 \cdot 10^{-4}$	$2.15 \cdot 10^{-4}$
θ_4	$1.02 \cdot 10^{-2}$	$6.12 \cdot 10^{-3}$
θ_5	$5.99 \cdot 10^{-1}$	$1.43 \cdot 10^{-1}$
θ_7	$6.49 \cdot 10^{-2}$	$4.24 \cdot 10^{-2}$
θ_8	$6.78 \cdot 10^{-1}$	$4.10 \cdot 10^{-1}$

	Estimate [h^{-1}]	Error [h^{-1}]
θ_9	$4.29 \cdot 10^{-3}$	$3.19 \cdot 10^{-3}$
θ_{10}	$5.67 \cdot 10^{-4}$	$5.12 \cdot 10^{-4}$
θ_{11}	$7.94 \cdot 10^{-2}$	$4.74 \cdot 10^{-2}$
θ_{12}	$1.89 \cdot 10^{-1}$	$1.32 \cdot 10^{-1}$
θ_{13}	$1.03 \cdot 10^{-1}$	$6.91 \cdot 10^{-2}$
θ_{14}	$6.54 \cdot 10^{-3}$	$3.94 \cdot 10^{-3}$

	Estimate [h^{-1}]	Error [h^{-1}]
θ_{15}	$3.98 \cdot 10^{-2}$	$3.00 \cdot 10^{-2}$
θ_{16}	$3.31 \cdot 10^{-3}$	$2.19 \cdot 10^{-3}$
θ_{18}	$5.79 \cdot 10^{-3}$	$4.89 \cdot 10^{-3}$
θ_{19}	$3.33 \cdot 10^{-3}$	$2.66 \cdot 10^{-3}$
θ_{20}	$1.84 \cdot 10^{-4}$	$1.49 \cdot 10^{-4}$
θ_{22}	$6.12 \cdot 10^{-1}$	$2.32 \cdot 10^{0}$

Table 5.2: *Parameter estimates from Tab. 4 in paper C including point estimate and estimated standard error of the estimate. The parameters are denoted by k_j instead of θ_j in paper C.*

We see that the standard errors of the estimates are large compared to the estimates, which is reasonable due to the limited amount of data. In particular are the errors for the parameters θ_1, θ_2 and θ_{22} large, which are associated with the leaving flux v_{22}. However, we see from Fig. 5.2 that the model captures the main trends of the data.

Even though the quality of the resulting dynamical model may be questionable, it illustrates how data can be used to fit a dynamical data. This motivates the need to develop experimental assays that are able to obtain time-course data of concentrations, which was the main objective of paper C. It could also be noted that the dynamical model could only be valid for a limited time interval as it contains seven terminal nodes where the concentrations would accumulate. For this reason, an adjusted model is presented in paper E to allow for interesting steady-states.

5.2   Identifiability

An important question when doing statistical inference to estimate model parameters is whether or not the true parameter values could be obtained from data. Assume some model is given where $\mathbf{g}(\boldsymbol{\theta})$ is the model output for parameter values $\boldsymbol{\theta}$. The output could be either a dynamical function of some other variable or a vector of fixed values. The model parameters $\boldsymbol{\theta}$ are said to be *identifiable* if the output $\mathbf{g}(\boldsymbol{\theta})$ satisfies

$$\mathbf{g}(\boldsymbol{\theta}_1) = \mathbf{g}(\boldsymbol{\theta}_2) \quad \Rightarrow \quad \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \tag{5.6}$$

where the equality for $\mathbf{g}(\boldsymbol{\theta})$ is an identity in the case of $\mathbf{g}(\boldsymbol{\theta})$ being a function of some other variable. The definition of identifiability in Eq. 5.6 is often referred to as *structural identifiability* as it is purely a property of the model structure. In paper A, we say that the parameters $\boldsymbol{\theta}$ are *globally identifiable* if Eq. 5.6 holds for all $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. If Eq. 5.6 holds for all $\boldsymbol{\theta}_2$ for a fixed $\boldsymbol{\theta}_1$, we say that the parameters $\boldsymbol{\theta}$ are *locally identifiable* at $\boldsymbol{\theta}_1$. Note that with these definitions, the parameters are globally identifiable if they are locally identifiable for all values.

There are several publications about identifiability of dynamical models relevant to reaction networks [11, 16, 49, 50, 59, 68, 69, 80]. Most of these assume a general dynamical function $\mathbf{g}(\boldsymbol{\theta})$, and do not assume any special structure of the model. Some

of the papers also have slightly different definitions of identifiability than the ones used in paper A, at least for the local identifiability.

In paper A, we study the identifiability of first order kinetic parameters in a reaction network at steady-state. We assume that the input fluxes to network are known design parameters, and that the model output $\mathbf{g}(\boldsymbol{\theta})$ is the steady-state concentrations of the metabolites for different values of the input fluxes. We present an identifiability criterion using the stoichiometric matrix and the active reactions discussed in Sec. 3.5.1 that is easy to check. We also provide a criterion to check if there exists a set of values for the input fluxes such that identifiability is attained.

In paper E, the method of paper A is applied to the dynamical model of paper C in the previous section. To do so, the method is extended and generalized to networks with a tree structure like the model of paper C. See the manuscript for more details.

5.3 Fisher information matrix and experimental design

The structural identifiability of a model is considered in the previous section, which implicitly assumes that the output $\mathbf{g}(\boldsymbol{\theta})$ from the model is known without errors. This is, however, rarely the case as $\mathbf{g}(\boldsymbol{\theta})$ must be observed and the observations will contain errors. For this reason, one should also consider the *practical identifiability* of the parameters even though the model is structurally identifiable. One method to explore the practical identifiability of a model is the profile likelihood, see Raue et al. [68].

In paper A, the kinetic parameters $\boldsymbol{\theta}$ are estimated using maximum likelihood estimation. The observations are assumed to be the steady-states for m different sets of known input fluxes given by a vector $\mathbf{b}^l$ for each case where $l \in \{1, \dots, m\}$. For each input vector $\mathbf{b}^l$, the true steady-state is denoted $\mathbf{z}^l$ and the observed steady-state $\mathbf{y}^l$. It is assumed that the vector $\mathbf{y}^l$ is a realization of stochastic variable $\mathbf{Y}^l$ with a normal distribution $\mathbf{Y}^l \sim \mathrm{N}\left(\mathbf{z}^l, \sigma^2 \mathbb{I}\right)$ where σ^2 is the variance of each observation and $\mathbb{I}$ the identity matrix. In addition are all the different $\mathbf{Y}^l$ assumed to be mutually independent. This gives a joint density $f\left(\mathbf{y}^1, \dots, \mathbf{y}^m | \boldsymbol{\theta}, \mathbf{b}^1, \dots, \mathbf{b}^m, \sigma^2\right)$ for all the observations $\mathbf{Y}^l$.

The likelihood function L for given observations $\mathbf{y}^1, \dots, \mathbf{y}^m$ is then given by

$$L\left(\boldsymbol{\theta} | \mathbf{y}^1, \dots, \mathbf{y}^m, \mathbf{b}^1, \dots, \mathbf{b}^m, \sigma^2\right) = f\left(\mathbf{y}^1, \dots, \mathbf{y}^m | \boldsymbol{\theta}, \mathbf{b}^1, \dots, \mathbf{b}^m, \sigma^2\right) \qquad (5.7)$$

which is a function of $\boldsymbol{\theta}$. The maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ is then

$$\hat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} \left[L\left(\boldsymbol{\theta} | \mathbf{x}\right)\right]. \qquad (5.8)$$

Having defined the joint density, we compute the Fisher information matrix (FIM) $\mathbf{I}_{\boldsymbol{\theta}} = \{I_{ij}\}$ analytically in paper A [15, 54]. The entries I_{ij} of the matrix are given by

$$I_{ij} = \mathbb{E}_Y \left[\left(\frac{\partial \ln f}{\partial \theta_i}\right) \left(\frac{\partial \ln f}{\partial \theta_j}\right)\right] \qquad (5.9)$$

where $\boldsymbol{\theta}$ is given and the expectation is taken with respect to the density f of the $\mathbf{Y}^l$'s.

The FIM is a measure of the parameter information in the observations, and is computed for a given parameter value $\boldsymbol{\theta}$. If $\mathbf{I}_{\boldsymbol{\theta}}$ is non-singular for a given value of $\boldsymbol{\theta}$, the

parameters are locally identifiable at this value. This implies that $\mathbf{I}_{\boldsymbol{\theta}}$ is non-singular for all $\boldsymbol{\theta}$ if the parameters are globally identifiable. The asymptotic covariance matrix for the estimate $\hat{\boldsymbol{\theta}}$ of Eq. 5.8 is given by $\mathbf{I}_{\boldsymbol{\theta}}^{-1}$. In particular is the asymptotic estimation variance $\sum_j \text{Var}\left(\hat{\theta}_j\right)$ given by $\text{Tr}\left(\mathbf{I}_{\boldsymbol{\theta}}^{-1}\right)$ which we calculate in paper A for an example.

Having an analytical expression for the FIM is uncommon, as it usually must be calculated numerically. In paper A we use it to find D-optimal [28] inputs $\mathbf{b}_*^l$ given by

$$\left[\mathbf{b}_*^1, \ldots, \mathbf{b}_*^m\right] = \underset{\mathbf{b}^1, \ldots, \mathbf{b}^m}{\text{argmax}}\left[\det\left(\mathbf{I}_{\boldsymbol{\theta}}\right)\right] \qquad (5.10)$$

where some constraint must be applied to the input vectors to keep the maximization from being unbounded. In paper E, optimal input fluxes are found by this method for the steroidogenesis pathway using the parameter values discussed in Sec. 5.1 that are estimated from real data in paper C.

Chapter 6

Data analysis in paper D

So far have we considered dynamical models of reaction networks, and all the papers except paper D are based on such models. In paper D, however, experimental data are analysed using traditional statistical methods without a dynamical model.

The aim of Paper D [9] was to assess exposure effects of chlorinated, brominated and perflourinated chemicals to the steroidogenesis pathway [6] discussed in Sec. 1.1.3 and shown in Fig. 6.1. Both single and combined mixtures were considered, giving a total of seven different mixtures. Each of these mixtures have four different concentrations denoted low, medium, high and very high corresponding to 1, 10, 100 and 1000 times the estimated concentration in human blood, respectively. For the brominated and perflourinated mixture, the low concentration is replaced by a concentration 10 000 times the blood value. In addition to the four different concentrations is the experiment performed both for unstimulated (treated with DMSO) and forskolin stimulated cells, which can be viewed as two different conditions. This gives a total of

$$(7 \text{ mixtures}) \cdot (4 \text{ concentrations}) \cdot (2 \text{ conditions}) = 56 \text{ cases}. \tag{6.1}$$

For each of these, the concentrations of 16 steroids were measured after 48 hours exposure. Each concentration was measured a total of nine times using three independent biological replicates with three technical replicates each. Undetected values and concentration values below the detection limit were replaced by half the detection limit.

Since the main objective of the study is the response to the various exposures, the actual concentrations values may not be too interesting. In addition, there could be differences between the biological replicates such that the concentration values from different replicates not necessarily are comparable. For this reason, we transform all the concentrations to fold changes before doing any analyses. The *fold change* (FC) for a measurement of a steroid concentration is given by the fraction

$$FC = \frac{\text{measurement}}{\text{control}} \tag{6.2}$$

where the control value for the measurement is the average concentration of the steroid for the same biological replicate without any exposure.

Since all concentrations are positive, the FC will also be a positive value. A value below one is referred to as a downregulation and a value above one an upregulation.

When working with FC values it is common to do a log-transform since the range for downregulation is given only by the interval $(0, 1)$, while an upregulation is given
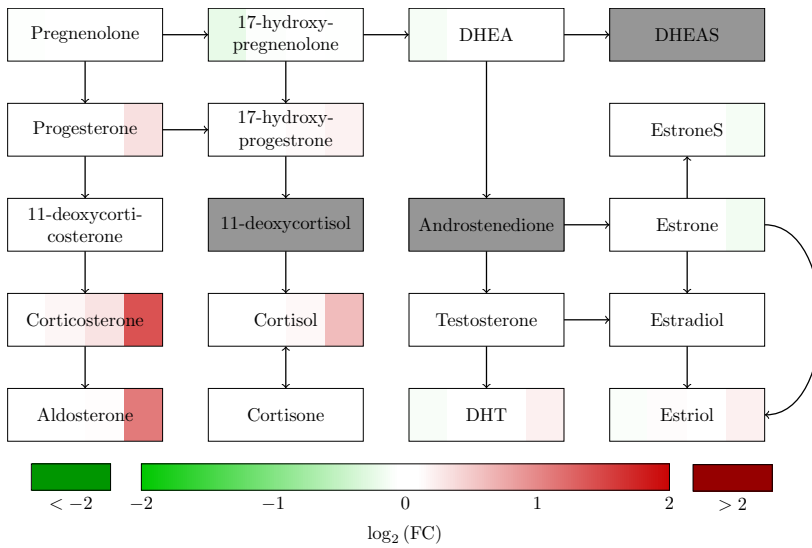
Figure 6.1: *Fig. 2A of paper D. Average fold change values of the steroids in the steroidogenesis pathway for forskolin stimulated cells under exposure of a chlorinated and perflourinated mixture. The box for each steroid is divided into four parts corresponding to (from left to right) low, medium, high and very high mixture concentration. Each value is colored according to the $\log_2$-value of the fold change of Eq. 6.3, where the colorbar is shown at the bottom and a value of $\log_2(\text{FC}) = 0$ corresponds to no mixture effect. Gray boxes are missing data.*

by $(1, \infty)$. In paper D, we use the logarithm with base 2 giving

$$\log_2(\text{FC}) = \log_2\left(\frac{\text{measured}}{\text{control}}\right) = \log_2(\text{measured}) - \log_2(\text{control}) \quad (6.3)$$

which can be any number in $\mathbb{R}$. Downregulation is then given by a negative value and upregulation by a positive value. Note that $\log_2(\text{FC})$ is symmetric in the sense that a value of -1 corresponds to half the control and a value of 1 to double the control, which is desirable since the two values correspond to what can be seen as inverse events. Also note that some texts refer directly to what we denote $\log_2(\text{FC})$ as the fold change.

The average fold change values for the four different concentrations of the chlorinated and perflourinated mixture and forskolin stimulated cells are shown in Fig. 6.1. The averages were computed after transforming to the fold change values of Eq. 6.1, but before performing the log-transform of Eq. 6.3.

Since there are 16 steroids measured for each of the 56 cases, the experiment results in 896 different average fold change values. To test for significant effects, we use a standard t-test [41] for each of the 896 different values. For each of these we use the nine replicates after they are transformed to the fold change values of Eq. 6.1, but without the log-transform of Eq. 6.3. The null hypothesis for the test of no effect then corresponds to a fold change value of one. All the 896 tests result in a p-value where a p-value below the significance value of $\alpha = 0.05$ is considered a significant effect. However, due to the large number of tests we need to perform a correction for multiple testing to reduce the number of false positives. This was done using the method of
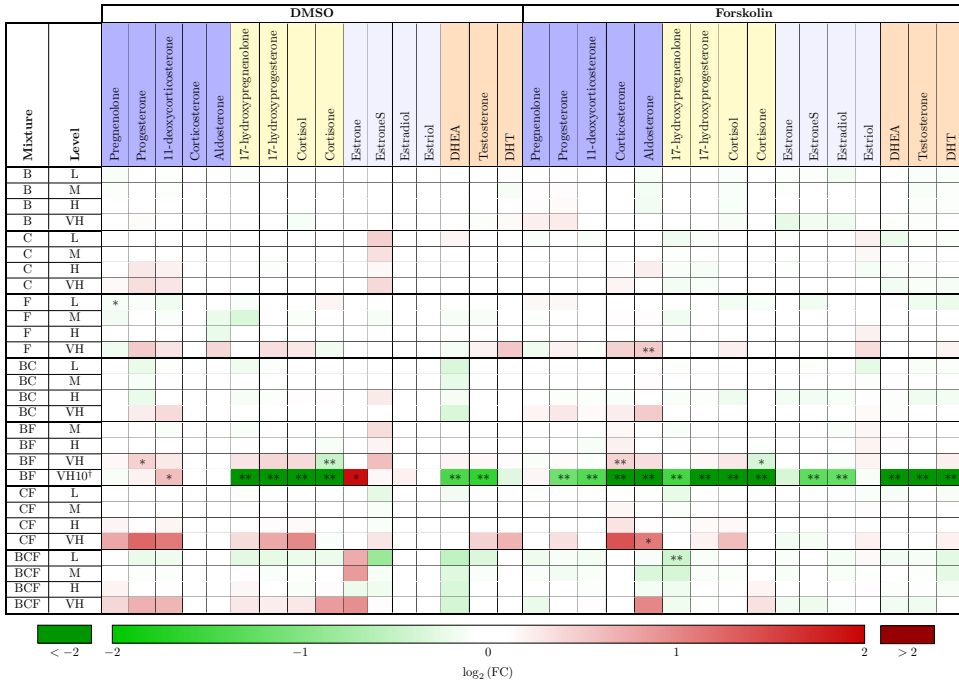
Figure 6.2: *Fig. 1 of paper D. Average fold change values for all the 16 measured steroids in unstimulated (DMSO) and forskolin stimulated cells. Each value is colored according to the* $\log_2$*-value of the fold change of Eq. 6.3 where the colorbar is shown at the bottom of the figure. The abbreviations for the mixtures are B for brominated, C for chlorinated and F for perflourinated. The abbreviations for the levels (concentrations) are L for 1, M for 10, H for 100, VH for 1000 and VH10 for 10 000 times the concentration in human blood. An asteriks (*) indicates a significant effect with $p \leq 0.05$ and a double asterisk (**) a strong significant effect with $p \leq 0.01$ where the p-values are adjusted for multiple testing. The coloring of the steroid names is according to a classification of the steroids.*

Holm [42] that calculates adjusted p-values and keep the significance level.

The average fold changes for all the 896 concentrations are plotted in Fig. 6.2 together with significance indicators from the adjusted p-values. We see that most of the significant effects are for the highest concentration of the brominated and perflourinated mixture, but this mixture has a 10 times higher concentration than the highest concentration of the other mixtures. This could also affect the p-values of the other tests since the p-values for this mixture are much lower than for the other mixtures. It should also be noted that there are only nine replicates for each test, with what seems to be a high natural variability. For this reason is it natural that we do not get many significant effects for the lower mixture concentrations.

In paper D, we also tested if the effect for a combined mixture of two compounds could be predicted by the effects of the corresponding single mixtures. The null hypothesis is that the effect is additive in the fold changes such that

$$FC = FC_1 + FC_2 - 1 \tag{6.4}$$

where FC is the fold change of Eq. 6.1 for the combined mixture and FC_1 and FC_2 are the fold changes for the two single mixtures. To test for significant non-additive effects, we performed ANOVA for linear models with and without interaction [22]. Also here, the resulting p-values were adjusted for multiple testing using Holm [42]. The result for the brominated and chlorinated mixture and the forskolin stimulated cells are shown in Fig. 6.3. We see that there is only one statistically significant non-additive effect.
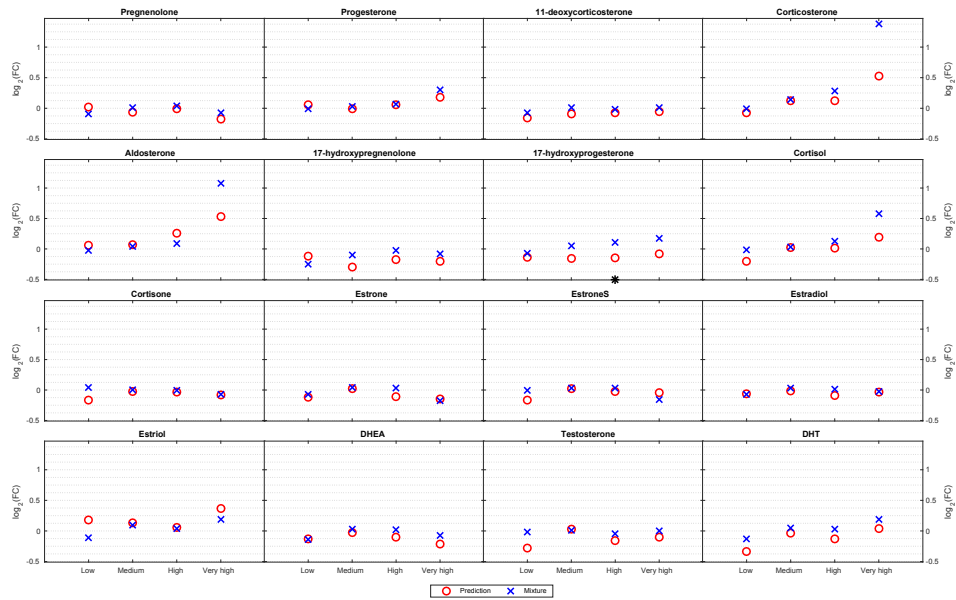
Figure 6.3: *Fig. 4 of paper D. Average fold change values for the steroids in forskolin stimulated cells for the brominated and chlorinated mixture. The four values for each steroid is for the low, medium, high and very high concentrations that correspond to 1, 10, 100 and 1000 times the concentration in human blood, respectively. The red circles are fold change values predicted from the single mixtures by Eq. 6.4, while the blue crosses are the observed values. The 17-hydroxyprogesterone for the high concentration has the only statistically significant non-additive effect with $p \leq 0.05$, and is marked with an asterisk (*).*

Chapter 7

Overview of the papers

We end part I of the thesis by a short overview of the papers and how they are connected. This is visualized in Fig. 7.1 where both existing and potential connections between the five papers are included in the figure.
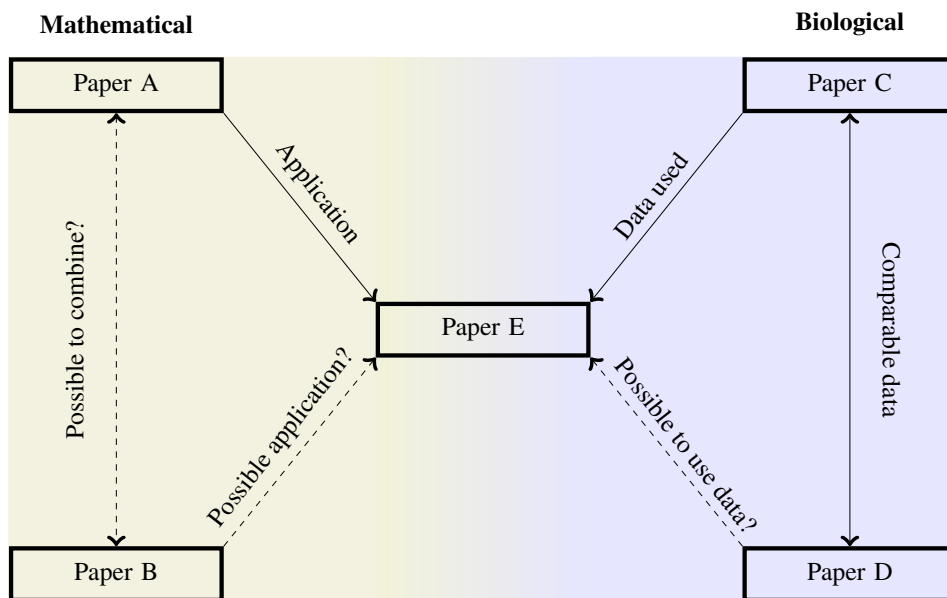
Figure 7.1: *Overview of the papers in the thesis and the connections between them. A solid line represents an existing connection, while a dashed line represents a possible connection. The papers on the left hand side have a mathematical focus, while the papers on the right hand side have a biological focus. Paper E is mathematical, but uses measurements from paper C.*

Paper A is the most theoretical work and develops a framework for experimental design. The framework assumes a dynamical model with zero order kinetics for the input fluxes and zero order kinetics for the remaining fluxes. Such a model is constructed in paper C for the steroidogenesis pathway where the parameter estimates are obtained from measurements. This is made use of in paper E where the framework developed in paper A is applied to the model presented in paper C.

Paper B is a mathematical paper and develops a method to evaluate model reduction under parameter uncertainty. As indicated in Fig. 7.1 could it potentially be combined to the method in paper A, e.g. by reducing the network to obtain identifiability.

Paper D could be viewed as an alternative approach to the remaining papers. In this paper, no dynamical model is assumed or created. Instead, measured concentrations are evaluated using statistical methods and there is no attempt to give a mechanistic description of the system. The experimental setup, however, is completely different to paper C where a dynamical model is constructed. Instead of measuring the concentrations at different times in the same experiment, the concentrations in paper D are measured at a given time for different experiments. Both papers, however, contain measurements of concentrations for steroidogenesis such that the data are comparable.

The dashed lines in Fig. 7.1 suggest some candidates for further work. Another option could be to apply some of the methods in the papers to other pathways than steroidogenesis. Obvious candidates for this are pathways that are relevant for salmon louse or cod based on the biological motivation of Sec. 1.1.

Bibliography

[1] Centre for Digital Life Norway. https://digitallifenorway.org/gb/. Accessed: 2019-10-29. 1.1.2

[2] Fatty acid metabolism. https://www.genome.jp/kegg-bin/show_pathway?map01212. Accessed: 2019-10-29. 1.1.2

[3] KEGG: Amino sugar and nucleotide sugar metabolism - Drosophila melanogaster (fruit fly). https://www.genome.jp/kegg-bin/show_pathway?org_name=dme&mapno=00520. Accessed: 2019-10-28. 1.1.1, 1.3, 1.2.2

[4] Overview of metabolic pathways. https://www.genome.jp/kegg-bin/show_pathway?map01100. Accessed: 2019-11-02. 1.2

[5] Sea Lice Research Centre. https://slrc.w.uib.no/. Accessed: 2019-10-28. 1.1.1, 1.1

[6] Steroid hormone biosynthesis. https://www.genome.jp/kegg-bin/show_pathway?org_name=map&mapno=00140. Accessed: 2019-10-30. 1.1.3, 1.2.2, 5.1, 6

[7] dCod 1.0: Decoding the systems toxicology of Atlantic cod. https://www.uib.no/en/dcod. Accessed: 2019-10-28. 1.1.2

[8] Kareem Eldin Mohammed Ahmed, Håvard G Frøysa, Odd André Karlsen, Jørn V Sagen, Gunnar Mellgren, Steven Verhaegen, Erik Ropstad, Anders Goksøyr, and Ralf Kellmann. LC-MS/MS based profiling and dynamic modelling of the steroidogenesis pathway in adrenocarcinoma H295R cells. *Toxicology in Vitro*, 52:332–341, 2018. 1.1.3, 4.1.3, 5.1

[9] Kareem Eldin Mohammed Ahmed, Håvard G Frøysa, Odd André Karlsen, Nello Blaser, Karin Elisabeth Zimmer, Hanne Friis Berntsen, Steven Verhaegen, Erik Ropstad, Ralf Kellmann, and Anders Goksøyr. Effects of defined mixtures of pops and endocrine disruptors on the steroid metabolome of the human h295r adrenocortical cell line. *Chemosphere*, 218:328–339, 2019. 1.1.3, 6

[10] D.H. Anderson. *Compartmental Modeling and Tracer Kinetics*. Lecture Notes in Biomathematics. Springer, 1983. ISBN 9783540123033. 4

[11] Syed Murtuza Baker, C Hart Poskar, Falk Schreiber, and Björn H Junker. A unified framework for estimating parameters of kinetic biological models. *BMC bioinformatics*, 16(104), 2015. 5.2

[12] Åke Bergman, Jerrold J Heindel, Susan Jobling, Karen Kidd, Thomas R Zoeller, World Health Organization, et al. *State of the science of endocrine disrupting chemicals 2012*. World Health Organization, 2013. 1.1.3

[13] Upinder S Bhalla and Ravi Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283(5400):381–387, 1999. 1.3

[14] Cristina Bizarro, Marta Eide, Daniel J Hitchcock, Anders Goksøyr, and Maren Ortiz-Zarragoitia. Single and mixture effects of aquatic micropollutants studied in precision-cut liver slices of atlantic cod (gadus morhua). *Aquatic Toxicology*, 177:395–404, 2016. 1.1.2

[15] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002. ISBN 9780534243128. 5.3

[16] Oana-Teodora Chis, Julio R Banga, and Eva Balsa-Canto. Structural identifiability of systems biology models: a critical comparison of methods. *PLoS one*, 6(11): e27755, 2011. 5.2

[17] William J. Conover. *Practical nonparametric statistics*. John Wiley & Sons, New York, 1971. 4.2

[18] Carsten Conradi, Elisenda Feliu, Maya Mincheva, and Carsten Wiuf. Identifying parameter regions for multistationarity. *PLoS computational biology*, 13(10): e1005751, 2017. 2.1, 4.2

[19] Mark J Costello. The global economic cost of sea lice to the salmonid farming industry. *Journal of fish diseases*, 32(1):115–118, 2009. 1.1.1

[20] Karina Dale, Mette Bjørge Müller, Zhanna Tairova, Essa Ahsan Khan, Kristin Hatlen, Merete Grung, Fekadu Yadetie, Roger Lille-Langøy, Nello Blaser, Hans J Skaug, et al. Contaminant accumulation and biological responses in atlantic cod (gadus morhua) caged at a capped waste disposal site in kollevåg, western norway. *Marine environmental research*, 145:39–51, 2019. 1.1.2

[21] Mousumi Debnath, Godavarthi BKS Prasad, and Prakash S Bisen. Omics technology. In *Molecular Diagnostics: Promises and Possibilities*, pages 11–31. Springer, 2010. 1.3

[22] Annette J Dobson and Adrian G Barnett. *An introduction to generalized linear models*. Chapman and Hall/CRC, 2008. 6

[23] JS Edwards and BO Palsson. The escherichia coli mg1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97(10):5528–5533, 2000. 1.2.1

[24] C Eichner, M Dondrup, and F Nilsen. Rna sequencing reveals distinct gene expression patterns during the development of parasitic larval stages of the salmon louse (lepeophtheirus salmonis). *Journal of fish diseases*, 41(6):1005–1029, 2018. 1.1.1

[25] Christiane Eichner, Lars Are Hamre, and Frank Nilsen. Instar growth and molt increments in lepeophtheirus salmonis (copepoda: Caligidae) chalimus larvae. *Parasitology international*, 64(1):86–96, 2015. 1.1.1

[26] M Eide, OA Karlsen, H Kryvi, PA Olsvik, and A Goksøyr. Precision-cut liver slices of atlantic cod (gadus morhua): An in vitro system for studying the effects of environmental contaminants. *Aquatic toxicology*, 153:110–115, 2014. 1.1.2

[27] Marta Eide, Halfdan Rydbeck, Ole K Tørresen, Roger Lille-Langøy, Pål Puntervoll, Jared V Goldstone, Kjetill S Jakobsen, John Stegeman, Anders Goksøyr, and Odd A Karlsen. Independent losses of a xenobiotic receptor across teleost evolution. *Scientific reports*, 8(1):10404, 2018. 1.1.2

[28] V.V. Fedorov. *Theory Of Optimal Experiments*. Probability and Mathematical Statistics. Elsevier Science, 1972. ISBN 9780323162463. 5.3

[29] Elisenda Feliu and Carsten Wiuf. Simplifying biochemical models with intermediate species. *Journal of the royal society interface*, 10(87):20130484, 2013. 1.2.3, 4.2

[30] David A Fell and J Rankin Small. Fat synthesis in adipose tissue. an examination of stoichiometric constraints. *Biochemical journal*, 238(3):781, 1986. 3.4

[31] Håvard G Frøysa, Shirin Fallahi, and Nello Blaser. Evaluating model reduction under parameter uncertainty. *BMC Systems Biology*, 12(1):79, 2018. 4.2

[32] Changdai Gu, Gi Bae Kim, Won Jun Kim, Hyun Uk Kim, and Sang Yup Lee. Current status and applications of genome-scale metabolic models. *Genome biology*, 20(1):121, 2019. 1.2.1

[33] Steinn Gudmundsson and Ines Thiele. Computationally efficient flux variability analysis. *BMC bioinformatics*, 11(1):489, 2010. 3.4

[34] Cato M Guldberg and Peter Waage. Studies concerning affinity. *CM Forhandlinger: Videnskabs-Selskabet i Christiana*, 35(1864):1864, 1864. 4

[35] Lars A Hamre, Christiane Eichner, Christopher Marlowe A Caipang, Sussie T Dalvin, James E Bron, Frank Nilsen, Geoff Boxshall, and Rasmus Skern-Mauritzen. The salmon louse lepeophtheirus salmonis (copepoda: Caligidae) life cycle has only two chalimus stages. *PloS one*, 8(9):e73539, 2013. 1.1.1

[36] Hulda María Harðardóttir, Rune Male, Frank Nilsen, and Sussie Dalvin. Effects of chitin synthesis inhibitor treatment on lepeophtheirus salmonis (copepoda, caligidae) larvae. *PloS one*, 14(9):e0222520, 2019. 1.1.1

[37] Hulda María Harðardóttir, Rune Male, Frank Nilsen, Christiane Eichner, Michael Dondrup, and Sussie Dalvin. Chitin synthesis and degradation in lepeophtheirus salmonis: Molecular characterization and gene expression profile during synthesis of a new exoskeleton. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 227:123–133, 2019. 1.1.1

[38] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004. 4.2

[39] Archibald Vivian Hill. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *j. physiol.*, 40:4–7, 1910. 4.1.2

[40] Kai Höffner, Stuart M Harwood, and Paul I Barton. A reliable simulator for dynamic flux balance analysis. *Biotechnology and bioengineering*, 110(3):792–802, 2013. 3.4

[41] Robert V Hogg, Elliot A Tanis, and Dale L Zimmerman. *Probability and statistical inference*, volume 993. Macmillan New York, 1977. 6

[42] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979. 6, 6

[43] Padraig Houlahan and John Scalo. Recognition and characterization of hierarchical interstellar structure. ii-structure tree statistics. *The Astrophysical Journal*, 393:172–187, 1992. 4.2

[44] Brian P Ingalls. *Mathematical modeling in systems biology: an introduction*. MIT press, 2013. 3

[45] Kenneth A Johnson and Roger S Goody. The original michaelis constant: translation of the 1913 michaelis–menten paper. *Biochemistry*, 50(39):8264–8269, 2011. 4.1.1

[46] Minoru Kanehisa and Susumu Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000. 1.2

[47] Kenneth J Kauffman, Purusharth Prakash, and Jeremy S Edwards. Advances in flux balance analysis. *Current opinion in biotechnology*, 14(5):491–496, 2003. 3.4

[48] Hiroaki Kitano. Systems biology: a brief overview. *science*, 295(5560):1662–1664, 2002. 1.3

[49] Costas Kravaris, Juergen Hahn, and Yunfei Chu. Advances and selected recent developments in state and parameter estimation. *Computers & chemical engineering*, 51:111–123, 2013. 5.2

[50] Clemens Kreutz. An easy and efficient approach for testing identifiability. *Bioinformatics*, 34:1913–1921, 2018. 5.2

[51] Kasper Kristensen, Anders Nielsen, Casper Berg, Hans Skaug, and Bradley Bell. TMB: Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software*, 70(1):1–21, 2016. ISSN 1548-7660. 5.1

[52] James C Lamb IV, Paolo Boffetta, Warren G Foster, Julie E Goodman, Karyn L Hentz, Lorenz R Rhomberg, Jane Staveley, Gerard Swaen, Glen Van Der Kraak, and Amy L Williams. Critical comments on the who-unep state of the science of endocrine disrupting chemicals–2012. *Regulatory Toxicology and Pharmacology*, 69(1):22–40, 2014. 1.1.3

[53] D.C. Lay. *Linear Algebra and Its Applications*. Pearson Education, Limited, 2012. ISBN 9781447911234. 3.2

[54] E.L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer New York, 2006. ISBN 9780387227283. 5.3

[55] Arthur Lesk. *Introduction to bioinformatics*. Oxford University Press, 2019. 1.3

[56] Yajie Liu and Hans vanhauwaer Bjelland. Estimating costs of sea lice control strategy in norway. *Preventive veterinary medicine*, 117(3-4):469–477, 2014. 1.1.1

[57] EW Lund. Guldberg and waage and the law of mass action. *Journal of Chemical Education*, 42(10):548, 1965. 4, 4.1.1

[58] Michael MacGillivray, Amy Ko, Emily Gruber, Miranda Sawyer, Eivind Almaas, and Allen Holder. Robust analysis of fluxes in genome-scale metabolic pathways. *Scientific reports*, 7(1):268, 2017. 3.4

[59] Kevin AP McLean and Kim B McAuley. Mathematical modelling of chemical processes—obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *The Canadian Journal of Chemical Engineering*, 90:351–366, 2012. 5.2

[60] Miles S Okino and Michael L Mavrovouniotis. Simplification of mathematical models of chemical reaction systems. *Chemical reviews*, 98(2):391–408, 1998. 4.2

[61] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245, 2010. 3.4

[62] Jeffrey D Orth, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard Ø Palsson. A comprehensive genome-scale reconstruction of escherichia coli metabolism—2011. *Molecular systems biology*, 7(1), 2011. 1.2.1

[63] Bernhard Ø. Palsson. *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press, 2015. ISBN 9781316239940. 1.2.1, 1.2.2, 1.2.3, 1.3, 2, 2.1.2, 2.4, 3.1, 3.4, 3.4, 3.5

[64] Eleftherios Terry Papoutsakis. Equations and calculations for fermentations of butyric acid bacteria. *Biotechnology and Bioengineering*, 26(2):174–187, 1984. 3.4

[65] Eleftherios Terry Papoutsakis and Charles L Meyer. Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations. *Biotechnology and Bioengineering*, 27(1):50–66, 1985. 3.4

[66] Ovidiu Radulescu, Alexander N Gorban, Andrei Zinovyev, and Vincent Noel. Reduction of dynamical biochemical reactions networks in computational biology. *Frontiers in genetics*, 3:131, 2012. 4.2

[67] Shodhan Rao, Arjan Van der Schaft, Karen Van Eunen, Barbara M Bakker, and Bayu Jayawardhana. A model reduction method for biochemical reaction networks. *BMC systems biology*, 8(1):52, 2014. 2.2.3, 2.3, 2.3, 2.3, 2.3, 4.2, 4.2, 4.3, 4.2

[68] Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009. 5.2, 5.3

[69] Andreas Raue, Johan Karlsson, Maria Pia Saccomani, Mats Jirstrand, and Jens Timmer. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*, 30(10):1440 – 1448, 2014. 5.2

[70] Meritxell Sáez, Carsten Wiuf, and Elisenda Feliu. Graphical reduction of reaction networks by linear elimination of species. *Journal of mathematical biology*, 74 (1-2):195–237, 2017. 4.2

[71] Maurice Scheer, Andreas Grote, Antje Chang, Ida Schomburg, Cornelia Munaretto, Michael Rother, Carola Söhngen, Michael Stelzer, Juliane Thiele, and Dietmar Schomburg. Brenda, the enzyme information system in 2011. *Nucleic acids research*, 39(suppl_1):D670–D676, 2010. 4.1.1, 5

[72] Jan Schellenberger, Richard Que, Ronan MT Fleming, Ines Thiele, Jeffrey D Orth, Adam M Feist, Daniel C Zielinski, Aarash Bordbar, Nathan E Lewis, Sorena Rahmanian, et al. Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2. 0. *Nature protocols*, 6(9):1290, 2011. 1.3

[73] Wilhelmus HA Schilders, Henk A Van der Vorst, and Joost Rommes. *Model order reduction: theory, research aspects and applications*, volume 13. Springer, 2008. 4.2

[74] Christophe H. Schilling, David Letscher, and Bernhard Ø. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, 203(3):229 – 248, 2000. ISSN 0022-5193. 3.5, 3.5

[75] Thomas J Snowden, Piet H van der Graaf, and Marcus J Tindall. Methods of model reduction for large-scale biological systems: a survey of current methods and trends. *Bulletin of mathematical biology*, 79(7):1449–1486, 2017. 4.2

[76] Ines Thiele, Neil Swainston, Ronan MT Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K Aurich, Hulda Haraldsdottir, Monica L Mo, Ottar Rolfsson, Miranda D Stobbe, et al. A community-driven global reconstruction of human metabolism. *Nature biotechnology*, 31(5):419, 2013. 1.2.1

[77] Ole Torrissen, S Jones, Frank Asche, Atle Guttormsen, Ove Tommy Skilbrei, Frank Nilsen, Tor Einar Horsberg, and D Jackson. Salmon lice–impact on wild salmonids and salmon aquaculture. *Journal of fish diseases*, 36(3):171–194, 2013. 1.1.1

[78] L.N. Trefethen and D. Bau. *Numerical Linear Algebra*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, 1997. ISBN 9780898719574. 3.2.5

[79] Robert J Vanderbei et al. *Linear programming*. Springer, 2015. 3.4

[80] J Vanlier, CA Tiemann, PAJ Hilbers, and NAW Van Riel. Parameter uncertainty in biochemical models described by ordinary differential equations. *Mathematical biosciences*, 246:305–314, 2013. 5.2

[81] Daojing Wang and Steven Bodovitz. Single cell analysis: the new frontier in 'omics'. *Trends in biotechnology*, 28(6):281–290, 2010. 1.3

[82] Hao Wang, Simonas Marcišauskas, Benjamín J Sánchez, Iván Domenzain, Daniel Hermansson, Rasmus Agren, Jens Nielsen, and Eduard J Kerkhoven. Raven 2.0: A versatile toolbox for metabolic network reconstruction and a case study on streptomyces coelicolor. *PLoS computational biology*, 14(10):e1006541, 2018. 1.3

[83] M R Watson. Metabolic maps for the Apple II. *Biochemical Society Transactions*, 12(6):1093–1094, 12 1984. 3.4

[84] Ulrike Wittig, Renate Kania, Martin Golebiewski, Maja Rey, Lei Shi, Lenneke Jong, Enkhjargal Algaa, Andreas Weidemann, Heidrun Sauer-Danzwith, Saqib Mir, et al. Sabio-rk—database for biochemical reaction kinetics. *Nucleic acids research*, 40(D1):D790–D796, 2011. 4.1.1, 5

[85] Fekadu Yadetie, Xiaokang Zhang, Eileen Marie Hanna, Libe Aranguren-Abadía, Marta Eide, Nello Blaser, Morten Brun, Inge Jonassen, Anders Goksøyr, and Odd André Karlsen. Rna-seq analysis of transcriptome responses in atlantic cod (gadus morhua) precision-cut liver slices exposed to benzo [a] pyrene and 17α-ethynylestradiol. *Aquatic toxicology*, 201:174–186, 2018. 1.1.2

Part II

Included papers

Paper B:
Evaluating model reduction under parameter uncertainty

Håvard G. Frøysa, Shirin Fallahi and Nello Blaser

METHODOLOGY ARTICLE

Open Access

CrossMark

Evaluating model reduction under parameter uncertainty

Håvard G. Frøysa[*], Shirin Fallahi and Nello Blaser

Abstract

Background: The dynamics of biochemical networks can be modelled by systems of ordinary differential equations. However, these networks are typically large and contain many parameters. Therefore model reduction procedures, such as lumping, sensitivity analysis and time-scale separation, are used to simplify models. Although there are many different model reduction procedures, the evaluation of reduced models is difficult and depends on the parameter values of the full model. There is a lack of a criteria for evaluating reduced models when the model parameters are uncertain.

Results: We developed a method to compare reduced models and select the model that results in similar dynamics and uncertainty as the original model. We simulated different parameter sets from the assumed parameter distributions. Then, we compared all reduced models for all parameter sets using cluster analysis. The clusters revealed which of the reduced models that were similar to the original model in dynamics and variability. This allowed us to select the smallest reduced model that best approximated the full model. Through examples we showed that when parameter uncertainty was large, the model should be reduced further and when parameter uncertainty was small, models should not be reduced much.

Conclusions: A method to compare different models under parameter uncertainty is developed. It can be applied to any model reduction method. We also showed that the amount of parameter uncertainty influences the choice of reduced models.

Keywords: Model reduction, Parameter uncertainty, Clustering, Systems biology

Background

Modelling of biochemical networks

Biochemical networks consist of chemical reactions between compounds, such as enzymes and metabolites. Through these reactions, the various compounds are consumed and produced. Each of these reactions has a reaction rate (flux) that typically depends on the compound concentrations, giving a dynamical behaviour of the system. The compound concentrations can thus be modelled by systems of ordinary differential equations (ODEs) and such dynamical models of biochemical networks may give biological insight that could not be obtained by modelling the compounds individually. However, the network dynamics may be complex and difficult to model accurately. The chemical reactions could possess advanced

kinetics such as activation and inhibition. In addition, the dimensions of the network may be large, for example the central energy metabolism in E. *coli* consists of more than 50 metabolites and 100 reactions [1].

Model reduction

The potential high complexity of the ODEs in the model represents a major challenge in analysing the dynamics of the system. Model reduction is a method for studying biochemical networks as it aims to identify the main components governing the dynamics of the system. The reduced model should be simpler to analyse, but retain the dynamical behaviour of the original model. There are different approaches to reduce the complexity of biochemical reaction networks, with the most common ones being lumping, sensitivity analysis and time-scale analysis [2–4]. Lumping combines compounds with

*Correspondence: havard.froysa@uib.no
Department of Mathematics, University of Bergen, Mailbox 7803, 5020, Bergen, Norway

similar behaviour into pseudo-compounds and considers differential equations involving these lumped pseudo-compounds [5, 6]. By performing parameter sensitivity analysis, the parameters with the least effect on the system output are neglected [7, 8]. In time-scale separation, biological processes are split into fast and slow processes and then the focus is put on the relevant time scale [9–15].

Another challenge in the analysis of complex networks is the lack of information on the kinetic properties of the reactions and parameter values. Reduction approaches that are not influenced by parameter uncertainty or incompleteness are called parameter independent reduction methods. For example, some reduction techniques based on exact lumping methods [5, 6] or qualitative reduction methods [16, 17] are parameter independent. Such reduction methods have been used extensively for signalling networks. For most reduction techniques, including methods based on time-scale separation or sensitivity analysis, the full parametrization of the model is required. In parameter dependent reduction, model parameters can play a significant role in selecting the elements for reduction. For some biochemical networks, the accuracy and validity of the reduced model can be influenced by changing the range of parameters so that the reduced model is only valid locally [3]. For reaction networks with well separated parameter values, reduced models capture the dynamical behaviour of the original model with an acceptable level of accuracy for an extensive range of parameter values [11, 18]. This, however, is not the case for general networks.

While there is a large literature on model reduction techniques, there is a lack of methods for evaluating model reductions. Some ad-hoc methods are the difference or scaled difference between the full and reduced model [5, 9], an error integral [14] and a criterion based on the initial values [10]. We are not aware of any criteria for evaluation of model reductions that takes parameter uncertainty into account. We present a new way to evaluate model reductions that takes parameter uncertainty into account and show the benefit of this method on two example networks.

Methods
Mathematical framework
The state variables of the dynamical model are the concentrations of the compounds. These compounds occur in different combinations on the left and right hand side of the chemical reactions of the network, where such a combination is called a complex [14]. For example, the chemical reaction $X_1 + X_2 \rightarrow X_3$ consists of the compounds X_1, X_2 and X_3, and the complexes $X_1 + X_2$ and X_3. The complex on the left hand side of an equation being consumed is called the substrate complex of the reaction and the complex on the right hand side of the reaction being produced is called the product complex. All this

information can be represented mathematically by a stoichiometric matrix [1] which gives the structure of the network.

In the notation of Rao et al. [14] the complexes are given by a matrix Z where the columns are the non-negative integer stoichiometric coefficients of the different complexes. The internal reactions are given by the linkage matrix B where each column corresponds to a reaction. This column is zero except in the rows corresponding to the substrate and product complex where it is -1 and 1, respectively. Let $x_i(t)$ be the concentration of compound i at time t and $\mathbf{x}(t)$ the corresponding vector quantity. The dynamics of any biochemical network is given by the system

$$\dot{\mathbf{x}} = ZB\mathbf{v} + Z\mathbf{v}_b \qquad (1)$$

of ODEs where Z and B give the network structure as described above. The vector $\mathbf{v}$ provides the internal fluxes of the network and $\mathbf{v}_b$ the boundary fluxes, i.e. the fluxes entering or leaving the network. As the fluxes typically are functions of $\mathbf{x}$, we restrict the internal fluxes $\mathbf{v}$ to the form

$$v_j(\mathbf{x}) = k_j d_j(\mathbf{x}) \exp\left(Z_{\mathcal{S}_j}^T \mathrm{Ln}(\mathbf{x}) \right) \qquad (2)$$

considered in [14] where k_j is a kinetic proportionality constant of reaction j, $d_j(\mathbf{x})$ is any function of $\mathbf{x}$, $Z_{\mathcal{S}_j}$ is the column of Z corresponding to the substrate complex of reaction j and $\mathrm{Ln}(\mathbf{x})$ is the mapping defined by $(\mathrm{Ln}(\mathbf{x}))_i = \ln(x_i)$. Further, let $Z_{\mathcal{S}}$ be the matrix where column j is $Z_{\mathcal{S}_j}$, i.e. the substrate complex of the reaction.

The dynamical model (1) now has the parameters k_j in addition to potential parameters in $\mathbf{v}_b(\mathbf{x})$ and the functions $d_j(\mathbf{x})$. A given set of values for such a parametrization will be called a parameter set. The unreduced model described by (1) will be referred to as the full or original model.

Reduction
We use the reduction procedure of Rao et al. [14] to reduce the model for a given parameter set. The first step in this procedure is to specify a set $\mathcal{M}_I$ of compounds considered to be important in the view of experimental design, e.g. the ones that are possible to measure. Note that the choice of $\mathcal{M}_I$ is subjective, but plays a major role in the reduction as the dynamics of the compounds in $\mathcal{M}_I$ are the ones used to compare the different reduced models. Then, the complexes of the network are divided into two categories. The first category is the complexes containing at least one of the compounds in $\mathcal{M}_I$. These complexes will not be considered for reduction. The other category is the complexes not containing any of the compounds in $\mathcal{M}_I$, and these will be the complexes considered for reduction. The reduction is then based on the assumption that the model approaches some steady state that can be found by integrating the system for a long enough time and that

the model is asymptotically stable around the steady state. A complex is reduced by setting its concentration constant equal to the corresponding steady state value of the full model. This can be done simultaneously for any number of complexes.

Having the possibility to reduce any given set of complexes, an iterative method to choose the complexes to be reduced is presented in Rao et al. [14]. It is a greedy method that reduces one complex at the time, always choosing the one yielding the smallest error as defined below. Finally, it stops when an error threshold is reached.

However, since the reduced models are independent of the order of reduction, we consider all possible simultaneous reductions of complexes. Assume now that there are c complexes eligible for reduction. It is then possible to reduce anywhere from 0 to c complexes, where reducing 0 gives the full model. In total there are 2^c possible reduced models for a given original model and parameter set. For each of these models, the concentrations of the compounds in $\mathcal{M}_I$ are then used to compare the models. When having n different parameter sets for the same original model, we perform the described reduction procedure for all the parameter sets. This yields 2^c possible reduced models for each parameter set and a total of $n \cdot 2^c$ different reduced models.

Comparing models
We need to be able to compare the dynamics of the different reduced models. In Rao et al. [14] the difference between the original model and a given reduced model is measured by an error integral. Let the concentration at time t of compound number i be $x_{ir}(t)$ and $x_{if}(t)$ for the reduced and the full model, respectively. Further, let $\mathbf{x}_r$ and $\mathbf{x}_f$ be the corresponding vector quantities for all the compounds. Finally, let $n(\mathcal{M}_I)$ be the number of compounds in $\mathcal{M}_I$ and $[0, T]$ the time interval that we evaluate the dynamics over. The error integral is then given by

$$I_T\left(\mathbf{x}_r, \mathbf{x}_f\right) = \sum_{i \in \mathcal{M}_I} \frac{1}{Tn\left(\mathcal{M}_I\right)} \int_0^T \left|1 - \frac{x_{ir}(t)}{x_{if}(t)}\right| dt \qquad (3)$$

which gives the average relative difference between the full and reduced model for all the compounds in $\mathcal{M}_I$ over the given time interval. Note that the error integral is non-symmetric in its arguments. However, we need to compare any two (reduced) models without favouring one of them. For this reason we introduce the symmetric error measure

$$E_T\left(\mathbf{x}_1, \mathbf{x}_2\right) = \frac{1}{2}\left(I_T\left(\mathbf{x}_1, \mathbf{x}_2\right) + I_T\left(\mathbf{x}_2, \mathbf{x}_1\right)\right) \qquad (4)$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the compound concentrations of any two (reduced) models. Note that this error measure can be calculated also for two models having different parameters as long as they have the same set $\mathcal{M}_I$.

Clustering
We use single linkage clustering [19] with the symmetric error as dissimilarity measure to cluster all the $n \cdot 2^c$ models with different parameter sets and reductions. Single linkage clustering is an agglomerative clustering method, which means that initially every model is in its own cluster. The dissimilarity $d(C_1, C_2)$ between two clusters C_1 and C_2 is calculated as the minimal symmetric error $\min_{x \in C_1, y \in C_2} E_T(x, y)$. The two clusters with the lowest dissimilarity are combined into one cluster at a hight given by their dissimilarity. Clusters are iteratively combined until only one cluster remains. This stepwise process can be visualized in a dendrogram [20]. A dendrogram provides a complete description of the single linkage clustering. From such dendrograms it is apparent which models are most similar and which models are more different.

We then color the dendrogram according to the used reduction. Each reduction is mapped to a color and each leaf of the dendrogram receives the color associated to its reduction. Model reductions that cluster together with the original model do not change the model behaviour, while model reductions that are separated from the original model changed the model behaviour. So if the dendrogram separates colors, we consider the model reduction that causes the separation to change the model behaviour. The reduced models that are distributed in a similar way as the original model in the dendrogram are considered to be consistent for the given parameter uncertainty.

In order to analytically compare the distributions of different models in the dendrogram, we calculate the positions in the dendrogram for each model. We then use the test statistics of a Kolmogorov-Smirnov test [21] between a given model and the full model as score for the model. For a given threshold α, we say that models with a score lower than the threshold are consistent with the full model at threshold α. Finally, the best reduced model is then chosen to be the consistent model that uses the most reductions. In the case of several consistent models having the same number of reductions, the best model is the one with the lowest score. For the remainder of this article we use a threshold of $\alpha = 0.2$.

Simple example
To illustrate the method, we created a small example network consisting of four compounds as shown in Fig. 1. Each compound occurs only one place in the network and never in combination with other compounds, implying that the complexes are just the compounds. The set $\mathcal{M}_I$ of important compounds is chosen to be number 1 and 4 such that the intermediate compounds 2 and 3 are considered for reduction.

Fig. 1 Example network. Each node is a compound and each arrow a reaction. The k_j's are the kinetic parameters of the reactions. Important compounds $\mathcal{M}_I$ and candidate compounds for reduction are specified by pink and black rectangles, respectively. External fluxes are indicated by blue arrows
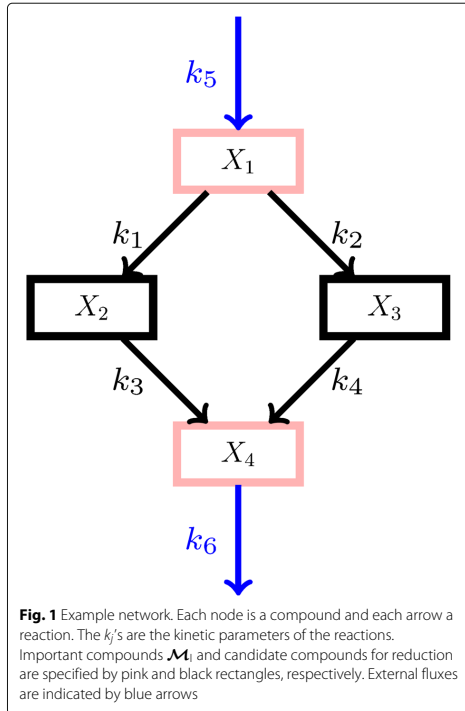
We apply mass action kinetics. Then k_j is the only kinetic parameter of reaction j. In the notation of [14] introduced earlier in the article, we have the matrices

$$Z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} B = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix} Z_{\mathcal{S}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$(5)$$

for the network. Using mass action we have $d_j(\mathbf{x}) = 1$ such that (2) becomes

$$v_j(\mathbf{x}) = k_j \exp\left(Z_{\mathcal{S}j}^T \mathrm{Ln}(\mathbf{x})\right) , \ j \in \{1, 2, 3, 4\} \quad (6)$$

for the internal fluxes of $\mathbf{v}$. The boundary fluxes are given by

$$\mathbf{v}_b = \begin{bmatrix} k_5 & 0 & 0 & -k_6 x_4 \end{bmatrix}^T \quad (7)$$

where the last entry is negative since the flux is leaving the network.

The dynamics are now given by (1) and we have six kinetic parameters k_j associated with one of the six

fluxes each. We sampled several parameter sets, which as expected lead to different reduction results. The parameter set that was chosen as reference because it gives particularly interesting reduction results is shown in Table 1. Then, 100 new parameter sets were sampled using this reference set by assuming the parameters to be independently log-normally distributed with the logarithm of the reference values as mean on the log scale and 0.1 as log standard deviation. We applied the reference initial values for all of the parameter sets, and the models were then reduced and clustered as described above.

Yeast glycolysis example

We also tested our method on a kinetic model of yeast glycolysis [22] shown in Fig. 2. This model was used in Rao et al. [14] to demonstrate the model reduction method which ignores parameter uncertainty. The model is asymptotically stable around the steady state and the governing equations of the system can be represented in the form of Eqs. 1 and (2) such that the reduction procedure can be applied. The important compounds to form $\mathcal{M}_I$ are Glci, TRIO, BPG, PYR, AcAld and NADH. Accordingly, the six candidates for reduction are F6P, G6P, P2G, P3G, PEP and F16BP, which leads to a total of $2^6 = 64$ possible reductions for a given parameter set including the full model.

The model has 89 parameters for the different reactions of the network. Each of these parameters should be nonnegative, and have a reference value used in [14]. To study the effect of parameter uncertainty on the reduction we sampled parameter sets using these reference values. We assumed the parameters to be independently log-normally distributed with mean equal to the reference value and standard deviation equal to the reference value divided by a scaling parameter. The parameters with reference value zero were set to zero in the sampling. We sampled 100 parameter sets for each of the values 3, 5, 10, 20, 50 and 100 of the scaling parameter. For each of the parameter sets we performed model reduction and clustered all the $100 \cdot 64 = 6{,}400$ resulting models for each scaling parameter as described above. We ended up with six dendrograms containing 6400 models each.

Table 1 Initial values and reference kinetic parameter values for the example network of Fig. 1

Parameter	Value	Initial value	Value
k_1	0.44	$x_1(0)$	0.4
k_2	0.03	$x_2(0)$	0.0
k_3	0.55	$x_3(0)$	0.5
k_4	0.44	$x_4(0)$	0.4
k_5	0.42		
k_6	0.33		

107

Frøysa *et al. BMC Systems Biology*   (2018) 12:79   Page 5 of 10

Fig. 2 Yeast glycolysis network. Each node is a compound and each arrow a reaction. Important compounds $\mathcal{M}_I$ and candidate compounds for reduction are specified by pink and black rectangles, respectively. NAD indicated by a cyan rectangle is not explicitly included in the model as the total amount of NAD and NADH is conserved. External fluxes are indicated by blue arrows. ∗, ⋆ and † show an irreversible reaction from NADH to NAD, a reversible reaction between NADH and NAD and an irreversible reaction from NAD to NADH, respectively. As indicated in the network, different types of these reactions bind to some of the fluxes

In order to check the sensitivity of the method to the number of parameter sets sampled, we also sampled 1000 parameter sets for the model with scaling parameter 50. For each parameter set we considered all model reductions with a Kolmogorov-Smirnov test score below a threshold of 0.5 for the 100 previous parameter sets. We performed model reduction and clustering as above.

All analyses were performed in MATLAB [23]. All code used to generate the results is available in the online supplementary material.

Results
Simple example

For the used parameter values, the model with both compounds number 2 and 3 reduced clustered together with the original model and had a Kolmogorov-Smirnov score of 0.17. Both the model with only compound 2 removed and the model with only compound 3 removed had a Kolmogorov-Smirnov score of 1.00. The models with only compound 3 reduced were the furthest from the cluster including the original model. Figure 3 shows the single linkage cluster dendrogram. The behaviour changes substantially for different parameter values and parameter uncertainties.

Yeast glycolysis example

The trajectories of the full model and all reduced models using the parameter set from [14] show no effect for Glci, two groups for TRIO, PYR and NADH, but no clear picture for BPG and ACALD (Fig. 4). For the reference parameter set, we found two big clusters. The first cluster contained the full models as well as all the models with compound F16BP not reduced, and the second cluster contained all models with F16BP reduced.

The clusterings for a distribution of parameters depended on the parameter distribution. When the standard deviation was high, there were no clear clusters and the full models were evenly distributed between the reduced models (Fig. 5, top left). This means that the uncertainty in the parameters had more effect than the model uncertainty due to reduction. The more certain the parameters were, the more we saw a clear picture emerge, with all models that had compound F16BP reduced clustering together and all other models forming a separate cluster (Fig. 5, top right, bottom left). When decreasing parameter uncertainty even further, the original models started forming a cluster of models where both compounds PEP and F16BP were not reduced (Fig. 5, bottom right).

Fig. 3 Model clustering. Dendrogram from single linkage clustering of all model reductions with 100 parameters sets. Parameters were sampled from a log-normal distribution with log standard deviation 0.1
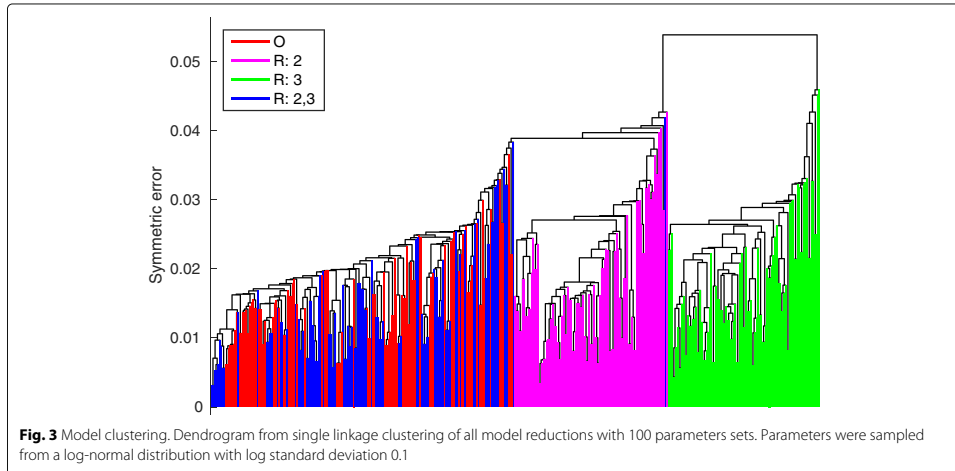
In addition to finding clusters that are inconsistent with the model uncertainty, we studied the distribution of the reduced models in the dendrogram. In the case of large parameter uncertainty (scaling parameters 3, 5, 10) the distribution of the fully reduced model in the dendrogram was similar to the distribution of the original model (Kolmogorov-Smirnov 0.11 or smaller). In the case of relatively large uncertainty (scaling parameter 20), all the models that did not reduce F16BP were distributed similarly to the original model (Kolmogorov-Smirnov 0.01).

Fig. 4 Yeast trajectories. The $\mathcal{M}_1$ states are shown for the reference parameters. Each color corresponds to a different model reduction
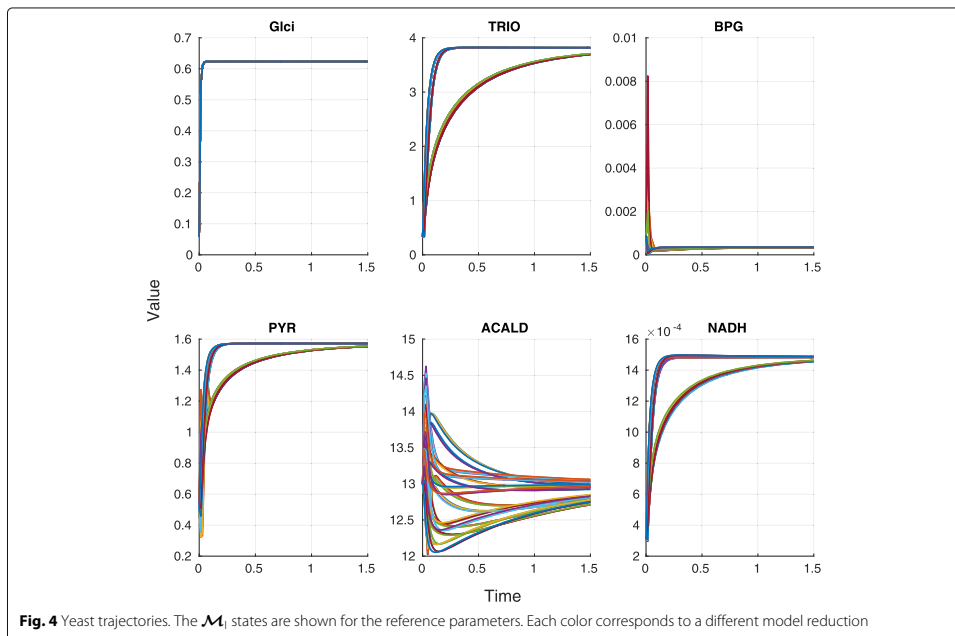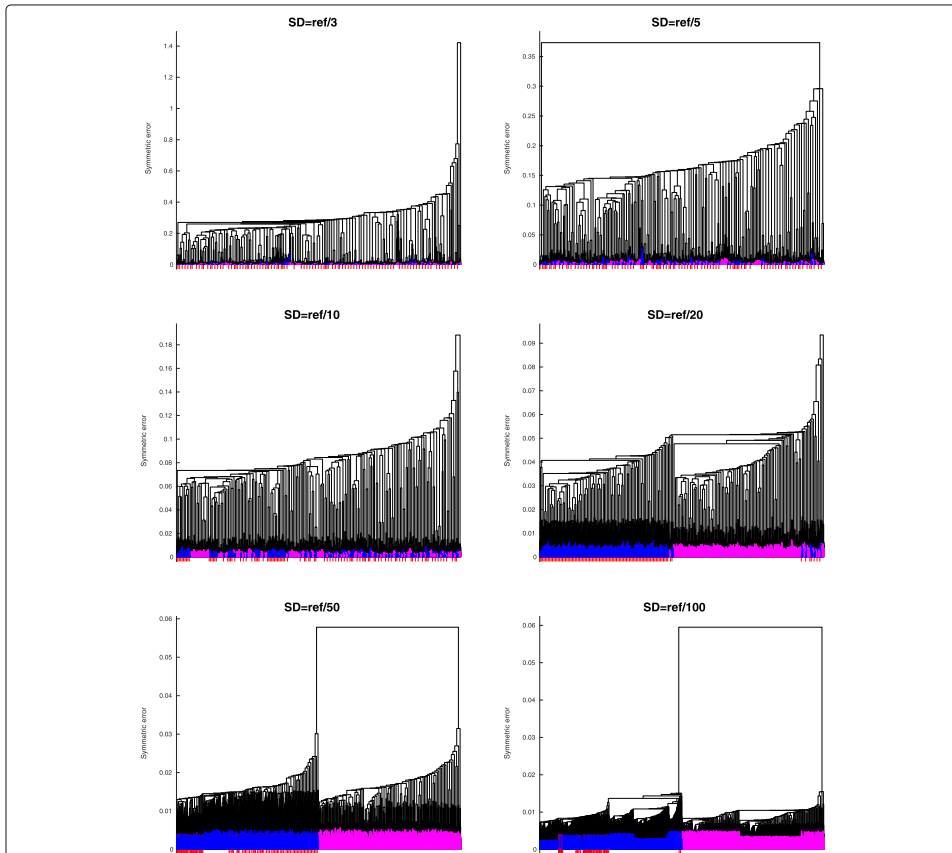
Fig. 5 Model clustering of F16BP reduced models. Dendrogram from single linkage clustering of all the model reductions using 100 parameters sets. Parameters were sampled from a log-normal distribution with standard deviation as reference value divided by 3 (top left), 5 (top right), 10 (center left), 20 (center right), 50 (bottom left) and 100 (bottom right). The original models are shown in red, models where F16BP was reduced are purple and all other models are blue

When the uncertainty was relatively low (scaling parameter 50), all models with F16BP and PEP not reduced clustered together with the full model (Kolmogorov-Smirnov 0.01 or 0.02). However, in the case of very low uncertainty (scaling parameter 100) the only model whose distribution in the dendrogram was similar to the distribution of the original model was the one where only F6P was reduced (Kolmogorov-Smirnov 0.01). The sensitivity analysis showed that whether or not a reduction was consistent for a given uncertainty did not dependent on the number of parameter sets (Fig. 6).

Discussion

We developed a new method to evaluate model reductions under parameter uncertainty based on the symmetric error measure in (4). In the yeast glycolysis example we showed that the amount of parameter uncertainty influences the model reduction. In particular, model uncertainty and parameter uncertainty are positively related. When the model parameters are uncertain, the model can be reduced further without increasing uncertainty in the model dynamics. We have also demonstrated empirically that if a model can be reduced to a certain degree for a

Reduction						Scaling						
G6P	F6P	F16P	P3G	P2G	PEP	3	5	10	20	50	50L	100
						0.01	0.01	0.01	0.01	0.02	0.06	0.50
						0.01	0.01	0.01	0.01	0.01	< 0.01	0.01
						0.10	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.02	0.03	0.46
						0.01	0.01	0.01	0.01	0.02	0.03	0.46
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.01	0.01	0.01	0.01	0.02	0.06	0.50
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.02	0.07	0.54
						0.01	0.01	0.01	0.01	0.02	0.07	0.54
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.02	0.03	0.46
						0.01	0.01	0.01	0.01	0.02	0.03	0.46
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.10	0.08	0.11	0.92	1.00	–	1.00
						0.10	0.08	0.11	0.92	1.00	–	1.00
						0.10	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.02	0.03	0.46
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.02	0.07	0.54
						0.01	0.01	0.01	0.01	0.02	0.07	0.54
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.02	0.07	0.54
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.02	0.03	0.46
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.10	0.08	0.11	0.92	1.00	–	1.00
						0.10	0.08	0.11	0.92	1.00	–	1.00
						0.10	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.02	0.07	0.54
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.10	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.01	0.01	0.01	0.01	0.38	0.86	0.98
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00
						0.09	0.08	0.11	0.92	1.00	–	1.00

Fig. 6 Kolmogorov-Smirnov test scores. Kolmogorov-Smirnov test scores for all the model reductions using 100 parameters sets as well as the sensitivity analysis with 1000 parameter sets (50L). The compounds in gray are reduced in the model of the corresponding row. Parameters were sampled from a log-normal distribution with standard deviation as reference value divided by the scaling factor. Models that are consistent with the original model are shown in light green and the best reduced model for each case is shown in dark green

given amount of uncertainty, then it can be reduced to at least the same degree if the uncertainty increases. If a model is used to analyse different scenarios, the parameters for all the scenarios should be considered when reducing a model. A full model should only be reduced to a model that is consistent for all considered scenarios. In addition to parameter values, uncertainty in initial values should also be considered. Our analysis shows that the

Frøysa *et al. BMC Systems Biology* (2018) 12:79

Page 9 of 10

reduction of Rao et al. [14] for the yeast model agrees with our best reduction for a relatively high amount of uncertainty, but becomes inappropriate for low or very large uncertainty.

In the simple example we demonstrated that it is sometimes better to reduce two complexes than just one. This also shows that even without parameter uncertainty the iterative approach used in [14] may not find the best reduction. Whether or not the best reduction is found depends on the symmetric error cut-off value. In the example, the reduced model would be found with symmetric error cut-off value at least 0.04, even though the symmetric error is only 0.02. The reference values in Table 1 for the parameters were chosen to illustrate this behaviour.

The novelty of our approach is a new way to evaluate model reduction. This model reduction evaluation criterion can be applied together with any model reduction method. Our criterion does not assume that the full model with a given parameter set is optimal. Instead it compares the full model with a wide range of parameter values to reduced models with the same range of parameter values to find a reduced model with the same properties, including model uncertainty. A reduced model with lower uncertainty in the trajectories could lead to overconfidence in the results.

A limitation of our method is that we need to choose a set $\mathcal{M}_I$ of important compounds. This choice is subjective and affects the resulting reduced model. However, there are some natural choices for the set $\mathcal{M}_I$, which depend on the model purpose. Of course $\mathcal{M}_I$ should contain all the compounds the study is investigating. It should also contain all the compounds whose concentrations are measured experimentally. Another limitation of our approach is that we have to choose the length T of the time series. It is important that at time T the trajectories are close to the steady state, because otherwise the error integral does not cover the entire model dynamics. On the other hand T should not be too large because otherwise the error integral reduces to the difference in steady states. If the model does not approach a steady state the dissimilarity measure we use may not be appropriate. There may also be some scaling issues with our proposed approach. Already in the case where we have to evaluate 64 models, we have to calculate a 6400×6400 matrix of dissimilarity measures using 100 parameter sets. For most practical examples, however, it is possible to reduce the sample space of reductions to a manageable size. In our sensitivity analysis with 1000 parameter sets, we have solved the issue by using the first 100 parameter sets to exclude some model reductions, which lead to a $32,000 \times 32,000$ dissimilarity matrix. The calculation of this matrix is the computational bottleneck of the method, but parallel computing can be applied. Moreover, it is possible to iteratively compare only a few

models at a time. We suggest that investigators adapt their strategies for model reduction based on model size, complexity and choice of the set $\mathcal{M}_I$. The Kolmogorov-Smirnov score leads to an automatic way of choosing the best reduced model. However, we believe that it is important to look at the dendrograms and not choose the model reduction only based on the Kolmogorov-Smirnov scores.

Conclusions

We presented a new method for evaluating models under parameter uncertainty and applied it for comparing full models to reduced models. We showed that multiple reductions can result in better models than individual reductions and that the amount of parameter uncertainty influences the choice of reduced models.

Additional file

Additional file 1: Scripts. Archive file containing all the scripts needed to produce and analyse the data of this paper. These scripts also produce Figs. 3, 4, and 5. (ZIP 23.3 kb)

Authors' contributions
All authors designed the method, wrote and revised the manuscript, and contributed to the interpretation of the results and the final version of the manuscript. HGF and SF performed the model simulations and model reductions. NB and HGF performed the cluster analysis. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Palsson BØ. Systems Biology: Constraint-based Reconstruction and Analysis. Cambridge, United Kingdom: Cambridge University Press; 2015.
2. Okino MS, Mavrovouniotis ML. Simplification of mathematical models of chemical reaction systems. Chem Rev. 1998;98(2):391–408.
3. Radulescu O, Gorban A, Zinovyev A, Noel V. Reduction of dynamical biochemical reactions networks in computational biology. Front Genet. 2012;3:131.
4. Snowden TJ, van der Graaf PH, Tindall MJ. Methods of model reduction for large-scale biological systems: A survey of current methods and trends. Bull Math Biol. 2017;79:1–38.
5. Borisov NM, Markevich NI, Hoek JB, Kholodenko BN. Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity. Biophys J. 2005;89(2):951–66.
6. Conzelmann H, Saez-Rodriguez J, Sauter T, Kholodenko BN, Gilles ED. A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. BMC Bioinformatics. 2006;7(1):34.
7. Liu G, Swihart MT, Neelamegham S. Sensitivity, principal component and flux analysis applied to signal transduction: the case of epidermal growth factor mediated signaling. Bioinformatics. 2004;21(7):1194–202.
8. Maurya M, Bornheimer S, Venkatasubramanian V, Subramaniam S. Reduced-order modelling of biochemical networks: application to the gtpase-cycle signalling module. IEE Proc Syst Biol. 2005;152(4):229–42.
9. Maas U, Pope SB. Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space. Combust Flame. 1992;88(3):239–64.
10. Surovtsova I, Sahle S, Pahle J, Kummer U. Approaches to complexity reduction in a systems biology research environment (sycamore). In: Proceedings of the 38th Conference on Winter Simulation. Monterey: Winter Simulation Conference; 2006. p. 1683–1689.
11. Radulescu O, Gorban A, Zinovyev A, Lilienbaum A. Robust simplifications of multiscale biochemical networks. BMC Syst Biol. 2008;2(1):86.
12. Hardin HM. Handling biological complexity: as simple as possible but not simpler. 2010.
13. Kourdis PD, Palasantza AG, Goussis DA. Algorithmic asymptotic analysis of the nf-κb signaling system. Comput Math Appl. 2013;65(10):1516–34.
14. Rao S, Van der Schaft A, Van Eunen K, Bakker BM, Jayawardhana B. A model reduction method for biochemical reaction networks. BMC Syst Biol. 2014;8(1):52.
15. West S, Bridge LJ, White MR, Paszek P, Biktashev VN. A method of 'speed coefficients' for biochemical model reduction applied to the NF-κB system. J Math Biol. 2015;70(3):591–620.
16. Naldi A, Remy E, Thieffry D, Chaouiya C. A reduction of logical regulatory graphs preserving essential dynamical properties. In: Computational Methods in Systems Biology. Bologna: Springer; 2009. p. 266–280.
17. Naldi A, Remy E, Thieffry D, Chaouiya C. Dynamically consistent reduction of logical regulatory graphs. Theor Comput Sci. 2011;412(21):2207–18.
18. Gorban A, Radulescu O. Dynamic and static limitation in multiscale reaction networks, revisited. Adv Chem Eng. 2008;34:103–73.
19. Florek K, Łukaszewicz J, Perkal J, Steinhaus H, Zubrzycki S. Sur la liaison et la division des points d'un ensemble fini. Colloq Math. 1951;2: 282–2851952.
20. Houlahan P, Scalo J. Recognition and characterization of hierarchical interstellar structure. ii - structure tree statistics. Astrophys J. 1992;393: 172–87.
21. Conover WJ. Practical Nonparametric Statistics. New York: Wiley; 1971.
22. van Eunen K, Kiewiet JA, Westerhoff HV, Bakker BM. Testing biochemistry revisited: how in vivo metabolism can be understood from in vitro enzyme kinetics. PLoS Comput Biol. 2012;8(4):1002483.
23. The Mathworks, Inc. MATLAB Version 8.6.0.267246 (R2015b). Natick, Massachusetts: The Mathworks, Inc; 2015.

Paper C:

LC-MS/MS based profiling and dynamic modelling of the steroidogenesis pathway in adrenocarcinoma H295R cells

Kareem Eldin Mohammed Ahmed, **Håvard G. Frøysa**, Odd André Karlsen, Jørn V. Sagen, Gunnar Mellgren, Steven Verhaegen, Erik Ropstad, Anders Goksøyr and Ralf Kellmann

LC-MS/MS based profiling and dynamic modelling of the steroidogenesis pathway in adrenocarcinoma H295R cells

Kareem Eldin Mohammed Ahmed[a,1], Håvard G. Frøysa[b,1], Odd André Karlsen[a], Jørn V. Sagen[c,d], Gunnar Mellgren[c,d], Steven Verhaegen[e], Erik Ropstad[e], Anders Goksøyr[a], Ralf Kellmann[c,*]

[a] Department of Biological Sciences, University of Bergen, P.O. Box 7803, N-5020 Bergen, Norway
[b] Department of Mathematics, University of Bergen, P.O. Box 7803, N-5020 Bergen, Norway
[c] Hormone Laboratory, Haukeland University Hospital, N-5021 Bergen, Norway
[d] Department of Clinical Science, University of Bergen, P. O. box 7804, 5020 Bergen, Norway
[e] Faculty of Veterinary Medicine and Biosciences, Department of Production Animal Clinical Sciences, Norwegian University of Life Sciences (NMBU), P.O. Box 8146, Dep. N-0033, Oslo, Norway

ABSTRACT

Endocrine disrupting chemicals have been reported to exert effects directly on enzymes involved in steroid biosynthesis. Here, we present a new liquid chromatography-tandem mass spectrometry (LC-MS/MS) method for profiling the steroid metabolome of H295R human adrenocarcinoma cells. Our method can simultaneously analyse 19 precursors, intermediates and end-products, representing the adrenal steroid biosynthesis pathway. In order to obtain better insights into the processes of steroidogenesis, we investigated the dose-response relationship of forskolin, an activator of adenylate cyclase, on steroid production in H295R cells. We observed that 1.5 μM forskolin stimulated steroid production at approximately 50% of the maximum rate for most steroids. Hence, we studied the time course for steroid synthesis over 72 h in H295R cells that were stimulated with forskolin. At 24 h, we observed a peak in steroid levels for the intermediate metabolites, such as progesterone and pregnenolone, while end-products such as testosterone and cortisol continued to increase until 72 h. Finally, we show how global data provide a unique basis to develop a comprehensive, dynamic model for steroidogenesis using first order kinetics. The timeline data made it possible to estimate all reaction rate constants of the network. We propose this method as a unique and sensitive screening tool to identify effects on adrenal steroidogenesis by endocrine disrupting compounds.

1. Introduction

Endocrine disrupting chemicals (EDCs) are chemicals which can interfere with hormonal systems of animals and humans (WHO, 2012). Production and use of such chemicals inevitably leads to their release as environmental contaminants (Fox, 2004; Toppari et al., 1996). Earlier studies have been focusing on the actions of EDCs on hormone receptors such as estrogen receptor (ER) and androgen receptor (AR). However, it has been established that EDCs can exert effects directly on enzymes involved in steroid biosynthesis and metabolism (Cai et al., 1995). In order to screen the large number of chemicals being produced for potential endocrine disrupting effects, programs such as REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) of the European Commission and the United States Environmental Protection Agency (EPA) 's EDSP (Endocrine Disruptor Screening Program) were implemented (Odermatt et al., 2016).

The EDSP program consists of a battery of in vitro and in vivo assays that assess the capability of xenobiotic compounds to act as agonists or antagonists to the ER, AR and steroidogenesis (O'Connor et al., 2002). Assays targeting aspects of female and male steroidogenesis exist, but are expensive and time consuming. The rat uterotropic assay utilizes the uterine weight as parameter for evaluating estrogenic activity of compounds (Kanno et al., 2001; Tomoya et al., 2003), and the minced testis assay measures the production of testosterone from testicular tissue which is harvested and incubated in cultured media (Charles, 2004). Hence, there is a need for more accurate, high throughput and

less expensive assays (Odermatt et al., 2016). As an alternative, many researchers use cell culture systems such as bovine adrenocortical primary cells (Cheng and Hornsby, 1992), the mouse Y1 cell line or the human H295R cell line (Cohen et al., 1957; Gazdar et al., 1990) to study steroid biosynthesis. The H295R cell model has been proposed as an alternative assay to be used by EPA for screening programs to investigate the effect of pesticides and other chemicals on the human population (Harvey and Everett, 2003; Hecker et al., 2006a; Hilscherova et al., 2004; Zhang et al., 2005), and the Organisation for Economic Co-operation and Development (OECD) has developed a test guideline for use of this assay (Guideline 456, (OECD, 2011)).

The H295R cell line originated from the parent NCI-H295 cell line, which was established from excised adrenocortical carcinoma (Gazdar et al., 1990). Previous analysis showed that H295R cells have all of the adrenocortical enzyme systems which give them the capacity to produce 30 different steroids (Bird et al., 1996; Gazdar et al., 1990; Rainey et al., 1994, 1993). In adrenal steroidogenesis, physiological stimulation of the steroidogenesis pathway occurs by binding of adrenocorticotropic hormone (ACTH) to the ACTH receptor, initiating a cyclic adenosine monophosphate (cAMP)-dependent response. However, H295R cells express low levels of the ACTH receptor. This leads to low response or even a complete resistance to ACTH stimulation (Mountjoy et al., 1994). Therefore, stimulation of the cAMP-pathway in these cells can be performed using a cAMP-elevating agent such as forskolin (Rainey et al., 1993). Forskolin stimulation of H295R cells results in a rapid up-regulation of cytochrome P450 (CYP) enzymes and subsequent steroid production (Denner et al., 1996; Weisser et al., 2016).

Steroids are commonly measured using immunological approaches such as radioimmunoassays (RIA), enzyme-linked immunosorbent assays (ELISA), and fluoroimmunoassays (FIA) (Gracia et al., 2006; Kjærstad et al., 2010; Szécsi et al., 2004). Different steroids have highly similar structures that differ only by their hydroxyl or carbonyl groups, which can cause significant cross-reactivity with specific antibodies (Heald et al., 2006; Hecker et al., 2006b; Middle, 2007; Penning et al., 2010). Moreover, immunoassays are prone to interference by the biological matrix, in particular when measuring steroids at low concentrations (Kushnir et al., 2011). Also, studies have revealed that immunoassays suffer from poor accuracy at low concentrations (Singh, 2008).

The use of liquid chromatography-tandem mass spectrometry (LC-MS/MS) can overcome many of the immunoassay deficiencies. By implementing LC-MS/MS methods, multiple analytes can be measured simultaneously during the same run (Rauh, 2010). Moreover, in the last decade newly developed triple quadrupole LC-MS/MS instruments provide a high precision and sensitivity enabling the quantitation of steroids at low concentrations with imprecisions of < 10% (Faupel-Badger et al., 2010; Hoofnagle and Wener, 2009; Stanczyk et al., 2007; Stenman, 2013). These methods are robust and can be used in a high-throughput environment (Alder et al., 2006; Guo et al., 2006).

In order to improve the interpretation of data from *in vitro* steroidogenesis assays, mechanistic mathematical models are useful tools (Breen et al., 2011). Such models may be used to assist with estimating the effects of EDCs in H295R cells and their concentration-response behaviour (Breen et al., 2010). In addition, by utilizing these mechanistic models the interpretation of data from H295R steroidogenesis assays could be improved by helping to define mechanisms of action for poorly characterized environmental toxicants (Breen et al., 2011, 2010; Mangelis et al., 2016). Moreover, guided by such mechanistic models more accurate extrapolations of toxic response of low dose exposures can be achieved (Conolly and Lutz, 2004).

Although the OECD guideline 456 only focuses on the production of estradiol and testosterone as endpoints, there is a need for a method that can detect and quantitate all steroids, many of which have physiological importance, in H295R cells (Hecker et al., 2006b; Winther et al., 2013). Here, we have developed an LC-MS/MS-based method to measure the biosynthesis of 19 steroids in H295R cells. Additionally, we

have investigated steroid production in relation to time and chemical stimulation with forskolin. Finally, we present a mathematical model for steroid biosynthesis, which take into account this comprehensive overview of the steroidogenesis pathway.

2. Materials and methods

2.1. H295R cell culturing

The H295R cell line was purchased form American Type Culture Collection (ATCC). Cells were cultured in 75 cm^2 flasks in Dulbecco's modified Eagle medium/HamF12 (DMEM/F12) containing HEPES buffer, L-glutamine and pyridoxine HCl (Gibco, Invitrogen, Paisley, UK). Additional supplements were added to the medium, including 1% insulin, human transferrin and selenous acid (ITS + premix) (BD Biosciences, Bedford, MA) and 5% charcoal stripped fetal bovine serum (F7524, Sigma Aldrich). H295R cells were incubated at 37 °C with 5% CO_2 in a humidified atmosphere. The medium was changed every 2–3 days and cells passaged at approximately 80% confluence by brief exposure to 0.25% trypsin/0.53 mM EDTA (Gibco, Invitrogen). The cells from passages 4–6 were used in experiments.

2.2. Forskolin exposure

After seeding H295R cells for 24 h in 6 well plates at a cell density of 1.2×10^6 cell per well, fresh medium containing different forskolin concentrations, (0.312 μM, 0.625 μM, 1.25 μM, 2.5 μM, 5 μM, 10 μM, 20 μM) was added to the cells for 48 h. Each concentration had 6 replicate wells.

2.3. Steroid production timeline

H295R cells ($1,2 \times 10^6$) were seeded in 6-well plates in 6 replicate wells per treatment condition and incubated for 24 h. Fresh medium containing 1.5 μM forskolin was added after 24 h of incubation and media was collected at 0, 2,4, 6, 12, 24, 36, 48, 72 h for analyses.

2.4. Cell viability

Cell viability was evaluated using Alamar Blue TM assay (Invitrogen) on the 96-well microplates (VWR, USA). Approximately 50,000 cells were seeded for 24 h before exposure for 48 h. DMSO control, 10 μM forskolin exposure was performed in triplicate. The medium was removed and replaced with 100 μl of fresh medium for 3 h at 5% CO_2 at 34 °C. A PerkinElmer (EnSpire 2300 Multilabel Reader) spectrophotometer was used to read the plates. The absorbance was read at 570 nm and 600 nm and viability was expressed as percentage of control (medium with 0.25% DMSO). Triton X-100 (10%) was used as a positive control of cell death.

2.5. Steroid-profiling by LC-MS/MS

2.5.1. Sample extraction

Samples of H295R cell medium were extracted using liquid-liquid extraction on a Hamilton Star pipetting robot, and 85 μl of sample was used for analysis in addition to 10 μl of internal standard that was added to all samples. Samples were equilibrated for one hour, and then extracted with 850 μl ethylacetate:hexane (80:20). 650 μl organic phase was evaporated under a stream of nitrogen at 45 °C, and samples reconstituted with 50 μl 25% methanol.

2.5.2. LC-MS/MS analysis

LC-MS/MS analysis was carried out on a Waters Xevo TQ-S triple quadrupole mass spectrometer that was coupled to a Waters i-class Acquity UPLC. Ionization was achieved by electrospray ionization (ESI) in positive and negative mode. The following LC conditions were used:

Table 1

Multiple reaction monitoring (MRM) based analytical parameters for 19 steroid analytes measured in H295R cells and deuterium-labelled internal standards, MRM transitions (Quantifier and Qualifier), analyte retention times, ionization for each analyte.

Analyte	Precursor > Quantifier/Qualifier (*m/z*)	Internal standard	IS Precursor > Quantifier/Qualifier (m/z)	Retention time (min)	Ionization
Aldosterone	361.1 > 189/331	D8-aldosterone	367.2 > 304/194	3.32	ESI-
Androstenedione	287.1 > 97/109	D7-androstenedione	294.1 > 100/113	5.60	ESI +
Corticosterone	347.0 > 121/97	D2–11 deoxycortisol	349.2 > 97/109	4.85	ESI +
Cortisol	363.1 > 297/282	D4-cortisol	335.3 > 301/286	3.97	ESI −
Cortisone	361.1 > 137/123	D4-cortisol	335.3 > 301/286	3.69	ESI −
11-Deoxycortisol	347.0 > 97/109	D2–11 deoxycortisol	349.2 > 97/109	5.00	ESI +
Dehydroepiandrosterone	271.2 > 253/213	D6-DHEA	277.2 > 219/258	6.18	ESI +
Dehydroepiandrosterone Sulfate	271.2 > 96/79	D6-DHEA sulphate	373.2 > 98/80	1.68	ESI −
Estrone	269.1 > 145/183	D4-estrone	273.2 > 187/147	5.36	ESI −
Estrone Sulfate	349.1 > 269/145	D4-E1sulphate	353.1 > 273/147	1.07	ESI −
Estriol	522 > 145/171	13C3-estriol	290.1 > 174/148	2.70	ESI −
Estradiol	255.2 > 145/183	D4-estradiol	275.2 > 187/147	5.31	ESI −
17-hydroxyprogesterone	331.1 > 97/109	13C3–17-hydroxy-progesterone	334.2 > 112/100	6.17	ESI +
17-hydroxypregnenolone	297.1 > 303/287	13C3–17-hydroxy-progesterone	334.2 > 112/100	6.13	ESI −
Progesterone	315.2 > 97/109	D9-progesterone	324.2 > 100/113	7.26	ESI +
Pregnenolone	299.1 > 159/281	D9-progesterone	324.2 > 100/113	7.76	ESI +
Testosterone	289.1 > 97/109	D3-testosterone	292.1 > 97/109	5.91	ESI +
Dihydrotestosterone	291.2 > 159/255	D3-testosterone	292.1 > 97/109	6.71	ESI +
21-hydroxyprogesterone	331.1 > 97/109	D3-testosterone	292.1 > 97/109	5.82	ESI +

chromatographic separation was achieved on a Waters Acquity BEH-C18 column (2.1 × 100 mm, 1.7 μm particle size, pore-size 130 Å). The column temperature was set at 60 °C. Mobile phase A consisted of Milli-Q water with 0,05% (vol/vol) ammonium hydroxide solution (25%), and mobile phase B consisted of methanolwith 0,5% (vol/vol) ammonium hydroxide solution (25%). The sample injection volume was 4 μl. Steroid hormones were detected and quantitated by isotope-dilution mass spectrometry by multi-reaction monitoring (MRM). Quantifier and qualifier MRM-transitions are listed in Table 1.

Methanol was used to dissolve all steroid hormones separately before they were added together. The mixture used for the standard curve had concentration of 100 times that of the highest working solution. The standard curve was prepared by serial diluting the mixture 1:4 in methanol and then adding 2 ml of each dilution to 198 ml of H295R growth medium with FBS serum. The final standard curve range is shown in Table 2. The standard curve has six levels for each steroid and a blank control that consist of H295R growth medium with FBS serum.

The internal standard was made of several labelled hormones that were dissolved separately in methanol. Quality control (QC) consist of H295R growth media of stimulated cells. Steroids in the media were measured and each steroid was added to a final concentration corresponding to 0.5, 1.56 and 25% of the highest standard level. The standard curve, internal standard and QCs were stored in −80 °C.

Our standard curve was run in parallel with a second serum based standard curve, which is fully validated with external quality control program (Methlie et al., 2013).

2.6. Dynamic model

Obtaining timeline data for the extracellular concentrations of steroids enabled us to construct a dynamic model for all the metabolites of steroidogenesis. The model has one compartment where the concentration of each steroid is assumed to be its measured extracellular concentration. A schematic overview of the network considered for

Table 2

Illustrates R^2 coefficient of determination, lower limit of detection (LLoD) and calibration ranges.

Metabolites	Range standard curve(ng/ml)	Limit of Detection (nM)	Regression Coefficients (R^2)	Average of Slope (m)	Standard deviation of Intercept (b)
Aldosterone	0.02–20	0.11	0.999	2.077	0.07
Pregnenolone	0.39–400	23.80	1.000	0.0002	0.002
Progesterone	0.19–198	0.03	1.000	0.065	0.001
Dihydrotestosterone	0.10–100	2.39	1.000	0.001	0.001
21-Hydroxyprogesterone	0.20–200	0.05	1.000	0.203	0.003
17-Hydroxypregnenolone	0.49–500	9.07	0.999	0.0005	0.001
17-Hydroxyprogesterone	0.43–435	7.46	0.999	0.034	0.076
Testosterone	0.09–88	0.03	1.000	0.269	0.002
Estrone	0.49–500	2.77	1.000	0.042	0.035
Estrone Sulphate	0.88–901	12.02	1.000	0.013	0.048
Cortisone	0.36–372	0.12	1.000	0.049	0.001
Cortisol	0.73–742	0.15	1.000	0.002	0.043
Dehydroepiandrosterone	1.95–2000	23.01	0.999	0.208	1.448
Dehydroepiandrosterone Sulphate	0.49–500	1.21	1.000	0.035	0.013
Estradiol	1.47–1500	0.53	1.000	0.009	0.001
Corticosterone	0.20–200	2.77	0,999	0.095	0.08
Estriol	0.49–500	0.46	1.000	0.007	0.001
11-Deoxycortisol	0.10–100	1.03	0,998	0.106	0.033
Androstenedione	0.32–219	0.32	1.000	0.07	0.006

Fig. 1. Schematic overview of steroidogenesis used for modelling. The boxes are the steroids measured experimentally and the arrows represent reactions with reaction rates v_i. The reaction rates are modelled by first order kinetics such that each rate v_i is proportional to its precursor with kinetic parameter k_i, e.g. $v_{13} = k_{13}C_{DHEA}$. An adjusted model is shown in Fig. 5.

Table 3

Quality control parameters for LC-MS/MS H295R cells steroid hormone measurements assay. Three different concentrations of quality controls (QC) were run in replicates of 6 in each experiment. Average concentrations (CONC) reported in Nano-molar (nM). (N) number of actual readings for each QC. Precision was calculated as the relative standard deviation (%RSD) of the standards.

Quality control levels	Low			Medium			High		
Metabolites	Average CONC	Precision (% RSD)	N	Average CONC	Precision (% RSD)	N	Average CONC	Precision (% RSD)	N
Aldosterone	0.95	14.7	21	2.6	13	21	20.2	12.1	21
Pregnenolone	13.5	8.8	18	23.3	11.7	21	135	8.3	20
Progesterone	6	9.3	19	8.8	8.5	18	42.7	7.3	18
Dihydrotestosterone	3.2	15.8	18	5.7	11.7	13	32.2	10.4	15
21-Hydroxyprogesterone	7	14	18	12.9	13.5	17	74.1	10.8	17
17-Hydroxypregnenolone	17	14.8	16	31.2	12.2	17	207.3	10.2	16
17-Hydroxyprogesterone	14	9	20	22	11.4	21	115.7	9.2	21
Testosterone	3.2	10	21	6	9.7	20	37.6	14.9	19
Estrone	19.1	14.3	19	41	13.2	19	271	8.8	16
Estrone Sulphate	30.2	13	22	44.4	8.2	18	219.3	8.6	20
Cortisone	11	7	21	17.2	6.7	21	91	8	22
Cortisol	24.2	5	21	40.3	5	21	222.7	6	22
Dehydroepiandrosterone	67.6	11.9	20	98.2	7.2	18	430.7	9.9	20
Dehydroepiandrosterone Sulphate	23.6	11.5	20	61	9.4	20	494.6	9	18
Estradiol	66	8.2	19	166.8	6.6	19	1312.7	9.7	18
Corticosterone	6	12	18	8.8	25.8	17	41.5	13.3	12
Estriol	20.3	6	21	48.3	6.1	20	370.4	6.3	22
11-Deoxycortisol	31.9	7.5	21	123.9	12.1	21	1775.3	20.4	6
Androstenedione	16.4	8.1	18	55.4	7.5	21	216.7	15.4	13

modelling is shown in Fig. 1, where each box is a measured steroid and each arrow a reaction (flux) with a reaction rate v_i. Note that the starting point of the model is pregnenolone.

The model equation for the concentration of *e.g.* DHEA will then be

$$\frac{dC_{DHEA}}{dt} = v_{11} - v_{12} - v_{13}$$

where the different reaction rates v_i of the network must be specified. Here, all reaction rates v_i of the model are taken to be first-order. Each reaction rate is then given by a rate constant k_i and the concentration of its precursor, e.g. $v_{13} = k_{13}C_{DHEA}$.

The replicated measurements were averaged to get a single concentration value for each steroid at the various time points. For each steroid, the first time-point was taken as the initial value for the concentration. The initial concentrations of steroids that were not detected at the start point were set to zero. The pregnenolone concentrations between the measurements were interpolated to give a time continuous input to the rest of the network. Having the pregnenolone concentration and initial concentrations for the remaining steroids, the system can be integrated for given reaction rate constants k_i. To find the k_i that best

fitted the data, a weighted least square estimate was calculated using the R package Template Model Builder (TMB)(Kristensen et al., 2016).

3. Results

3.1. Cell viability

Cell viability was evaluated using the Alamar Blue assay. No deviation from control was observed with 10 μM forskolin exposure, whereas the positive control triton X-100 showed decrease cell viability (data not shown).

3.2. LC-MS/MS analysis

Successive chromatographic separation of 19 adrenal steroids synthesised by H295R cell was achieved within 11 min for sample elution, column washing and re-equilibration (Table 1). With the exception of 17OH-pregnenolone and 17OH-progesterone, baseline separation of all other metabolites was achieved with a C18 column over the entire concentration range (Supplementary Fig. 1).

K.E.M. Ahmed et al.
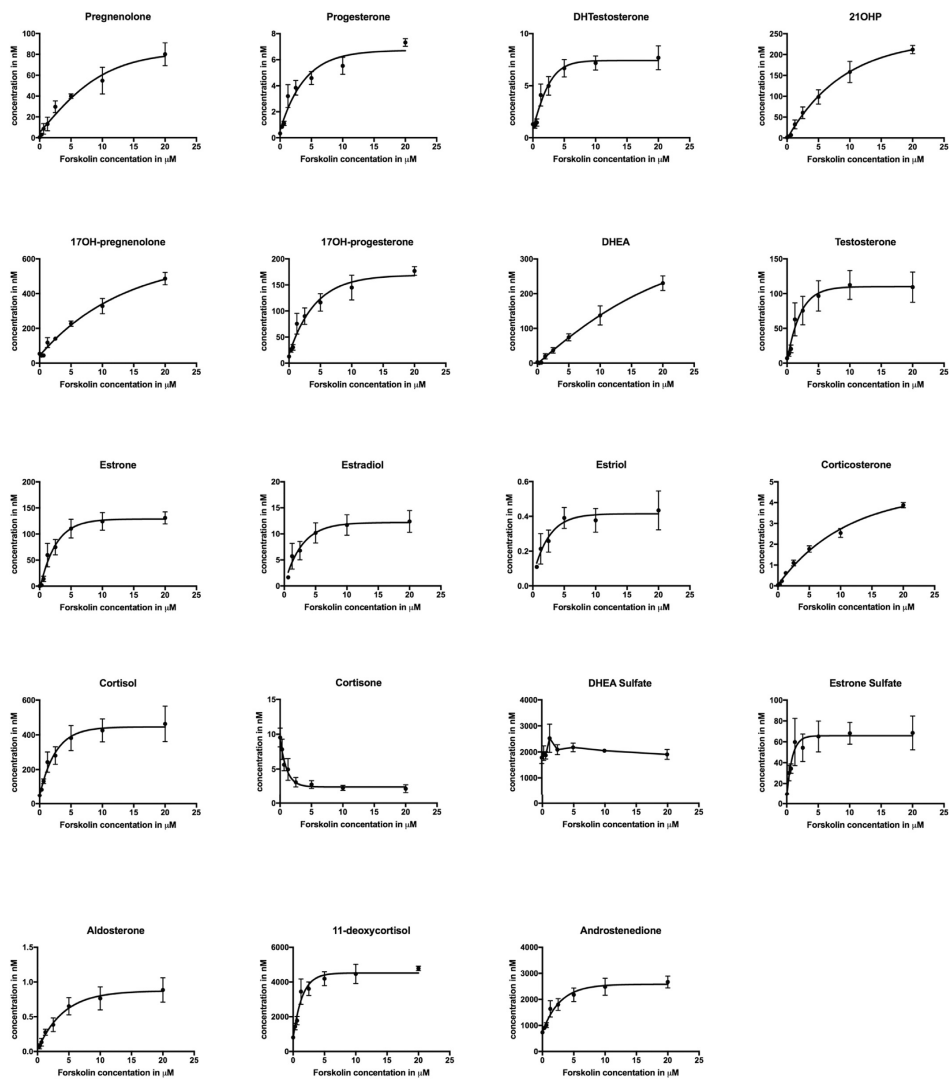
Fig. 2. Steroid production in the H295R cell line in 48 h of forskolin stimulation. Following the OECD guideline, steroids were extracted and analysed using the developed method. Each data point represents 6 samples from three independent experiments (*n* = 6). Error bars are standard deviations. Note the difference in y-axis scale.

The chromatographic separation starts with estrone sulfate in negative mode at 1.26 min followed by dehydroepiandrosterone sulfate (DHEAS), estriol, aldosterone, cortisone, cortisol and estradiol respectively all in negative mode (Supplementary Fig. 1 A). Furthermore, corticosterone (CCST), 11-deoxycortisol, androstenedione, testosterone, dihydrotestosterone (DHT), dehydroepiandrosterone (DHEA), 21-hydroxyprogesterone (21OHP), 17-hydroxypregnenolone (17OH-PREG), 17-hydroxy-progesterone (17OHP), progesterone and pregnenolone follow in positive mode (Supplementary Fig. 1).

Assay coefficients of variation (CV) were determined from repeated
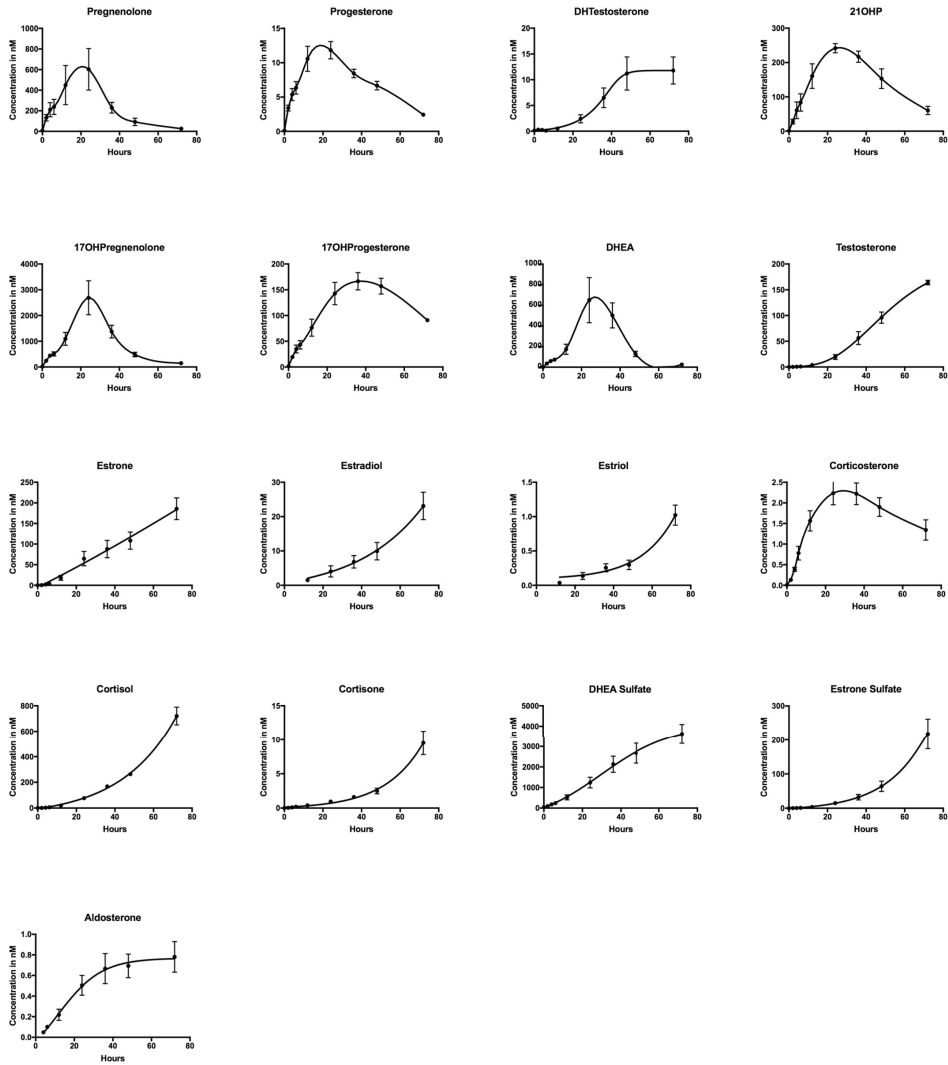
Fig. 3. Profile of steroid metabolites in H295R cells over a time course study after forskolin stimulation. Cells were treated with 1.5 μM forskolin, and measurements were taken at several time points from 0 to 72 h. Each data point represents 6 samples from three independent experiments (n = 6). *Error bars* are standard deviations. Note the difference in y-axis scale.

measurements of in-house prepared quality control (QC) samples at three different levels (Table 3). Lower limit of detection (LLOD) were determined as recommended by Armbruster and Pry, (2008). Estimated LLOD values can be found in Table 2. The accuracy of cortisol, cortisone, 11-deoxycortisol, progesterone, testosterone, 17-OH-progesterone, androstenedione and aldosterone have been determined (Supplementary Table 1).

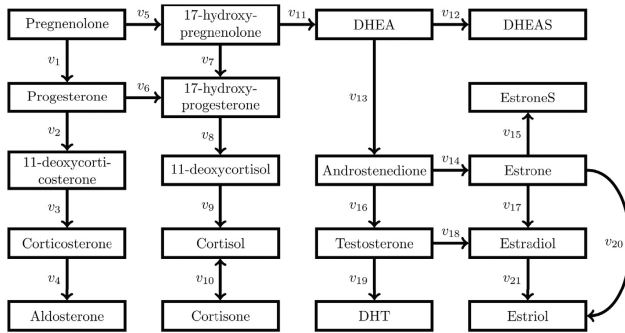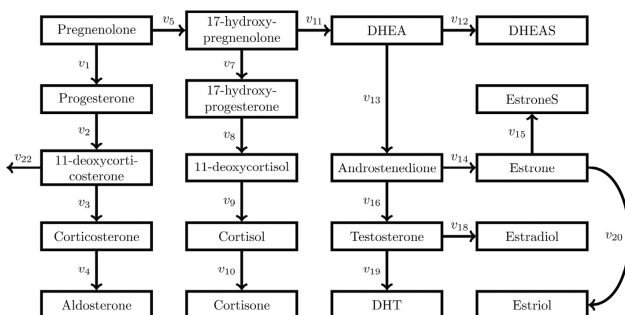**Fig. 4.** Adjusted schematic overview of steroidogenesis used for modelling. The boxes are the steroids measured experimentally and the arrows represent reactions with reaction rates vi. The reaction rates are modelled by first order kinetics such that each rate vi is proportional to its precursor with kinetic parameter ki, *e.g.* v13 = k13CDHEA. The original model is shown in Fig. 1.

Table 4

Estimated reaction rate constants with standard deviation for the adjusted model shown in Fig. 4. The values are estimated from the measurements (circles) in Fig. 5 and used to make the model predictions (lines) in Fig. 5. For technical details of the estimation see the subsections "Dynamical model" in the sections "Methods" and "Results".

Parameter	Point estimate	Standard deviation
k_1	$2.94 \cdot 10^{-1}\,\mathrm{h}^{-1}$	$1.02 \cdot 10^{0}\,\mathrm{h}^{-1}$
k_2	$1.24 \cdot 10^{1}\,\mathrm{h}^{-1}$	$4.36 \cdot 10^{1}\,\mathrm{h}^{-1}$
k_3	$3.43 \cdot 10^{-4}\,\mathrm{h}^{-1}$	$2.15 \cdot 10^{-4}\,\mathrm{h}^{-1}$
k_4	$1.02 \cdot 10^{-2}\,\mathrm{h}^{-1}$	$6.12 \cdot 10^{-3}\,\mathrm{h}^{-1}$
k_5	$5.99 \cdot 10^{-1}\,\mathrm{h}^{-1}$	$1.43 \cdot 10^{-1}\,\mathrm{h}^{-1}$
k_7	$6.49 \cdot 10^{-2}\,\mathrm{h}^{-1}$	$4.24 \cdot 10^{-2}\,\mathrm{h}^{-1}$
k_8	$6.78 \cdot 10^{-1}\,\mathrm{h}^{-1}$	$4.10 \cdot 10^{-1}\,\mathrm{h}^{-1}$
k_9	$4.29 \cdot 10^{-3}\,\mathrm{h}^{-1}$	$3.19 \cdot 10^{-3}\,\mathrm{h}^{-1}$
k_{10}	$5.67 \cdot 10^{-4}\,\mathrm{h}^{-1}$	$5.12 \cdot 10^{-4}\,\mathrm{h}^{-1}$
k_{11}	$7.94 \cdot 10^{-2}\,\mathrm{h}^{-1}$	$4.74 \cdot 10^{-2}\,\mathrm{h}^{-1}$
k_{12}	$1.89 \cdot 10^{-1}\,\mathrm{h}^{-1}$	$1.32 \cdot 10^{-1}\,\mathrm{h}^{-1}$
k_{13}	$1.03 \cdot 10^{-1}\,\mathrm{h}^{-1}$	$6.91 \cdot 10^{-2}\,\mathrm{h}^{-1}$
k_{14}	$6.54 \cdot 10^{-3}\,\mathrm{h}^{-1}$	$3.94 \cdot 10^{-3}\,\mathrm{h}^{-1}$
k_{15}	$3.98 \cdot 10^{-2}\,\mathrm{h}^{-1}$	$3.00 \cdot 10^{-2}\,\mathrm{h}^{-1}$
k_{16}	$3.31 \cdot 10^{-3}\,\mathrm{h}^{-1}$	$2.19 \cdot 10^{-3}\,\mathrm{h}^{-1}$
k_{18}	$5.79 \cdot 10^{-3}\,\mathrm{h}^{-1}$	$4.89 \cdot 10^{-3}\,\mathrm{h}^{-1}$
k_{19}	$3.33 \cdot 10^{-3}\,\mathrm{h}^{-1}$	$2.66 \cdot 10^{-3}\,\mathrm{h}^{-1}$
k_{20}	$1.84 \cdot 10^{-4}\,\mathrm{h}^{-1}$	$1.49 \cdot 10^{-4}\,\mathrm{h}^{-1}$
k_{22}	$6.12 \cdot 10^{-1}\,\mathrm{h}^{-1}$	$2.32 \cdot 10^{0}\,\mathrm{h}^{-1}$

3.3. Forskolin exposure

Measured concentrations of steroid levels in media of H295R cells after 48 h of exposure to different concentrations of forskolin showed that levels of the majority of analytes increased with elevated forskolin concentration as shown in Fig. 2. As expected the increase in steroid secretion reached a saturation level at around 10 µM of forskolin exposure (Fig. 2). The half-maximum production for most steroids occurred at approximately 1.5 µM forskolin, which was chosen to stimulate H295R cells for the timeline study.

3.4. Steroid production timeline

In the timeline experiment, precursors such as pregnenolone and progesterone reached peak production at 24 h (Fig. 3). On the other hand, levels of end products like testosterone, cortisol and aldosterone showed a continuous increase for up to 72 h. In addition, we observed that androstenedione and 11-deoxycortisol and DHEAS constitute around 86% of total steroids secreted by H295R cells, while estriol and aldosterone contributed only around 0.01% of the total steroid hormone production (Fig. 3).

3.5. Dynamic model

Based on the timeline data, optimization showed that the best fit for the dynamic model was attained by $k_6 = k_{17} = k_{21} = 0$ (Fig. 1), suggesting that the corresponding fluxes can be deleted from the model. In addition, the concentrations of corticosterone and aldosterone were too small to match the degradation of 11-deoxycorticosterone that takes place after 24 h. To compensate for this, an extra exchange flux from 11-deoxycorticosterone leaving the system is introduced. This could *e.g.* indicate leakage to another pathway. Finally, v_{10} can under the current conditions be modelled as an irreversible reaction from cortisol to cortisone due to the propagation of cortisone. Altogether this suggests the adjusted model of Fig. 4, for which we again performed parameter estimation using TMB.

The optimized reaction rate constants are listed in Table 4. TMB calculates standard deviations for each parameter in addition to the point estimates. These calculations show that k_1, k_2 and k_{22} are highly uncertain while the rest of the parameters have smaller standard deviations.

The concentrations predicted by the dynamic model are plotted in Fig. 5 together with the measured values.

4. Discussion

The H295R cell line is considered a unique model for the study of steroidogenic pathways, but also for the evaluation of endocrine disruption caused by xenobiotics (OECD, 2011; Rijk et al., 2012; Wang et al., 2014). Here we report a new steroidogenesis assay to profile the adrenal steroid metabolome of H295R cells. We have developed a robust and high throughput method to simultaneously analyse 19 precursors, intermediates and end-products of the steroid biosynthesis pathway. Our assay can be used to study metabolite fluxes in steroid biosynthesis, and to identify the targets of substances that interfere with this pathway. The low sample volume, combined with an automated sample extraction, and a short a chromatographic run-time of 11 min (Table 1) provide a throughput capacity of 130 samples per 24 h with minimal hands-on time by a single operator.

Baseline separation was achieved for all steroids in the liquid chromatography step, with the exception of 17OH-PREG and 17OHP. We did not observe any spectral interference between 17OH-PREG and 17OHP, but the continuous polarity switching required for their concurrent measurement may contribute to an increased imprecision (Table 1). The majority of calibration curves were linear, as shown by the coefficient of determination (r^2) being ≥0.99 (Table 2). However, similarly to previous reports (Abdel-Khalik et al., 2013), our data show poorer precision and non-linearity at low concentrations for DHT,
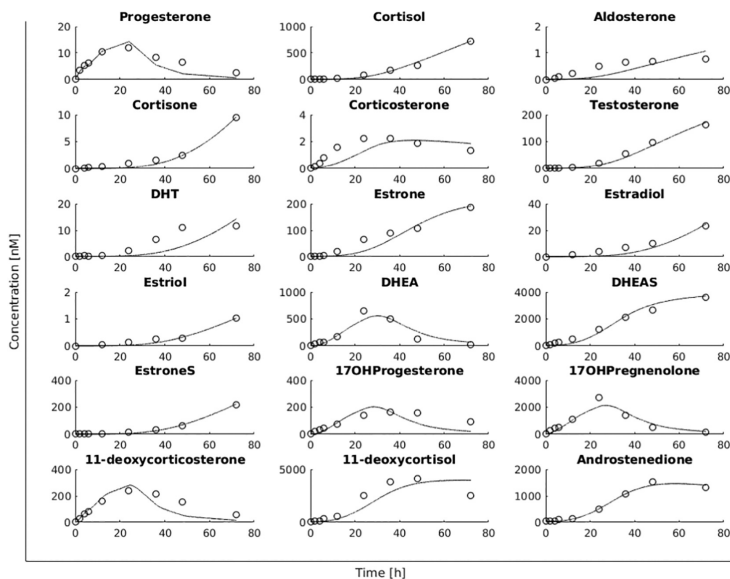
Fig. 5. Steroid production of H295R cells after 1.5 μM forskolin stimulation, based on measurements (o) and dynamic prediction (line). The x-axes are time [h] and the y-axes are concentration [nM]. The circles represent measurements performed experimentally and the lines are predictions made using the dynamical model presented in the paper. A schematic overview of the model is shown in Fig. 4. The measured data points are used to estimate the kinetic parameters of the model and the resulting values are shown in Table 4. For technical details of the model see the subsections "Dynamical model" in the sections "Methods" and "Results".

estrone and corticosterone. According to the FDA guidelines from 2001, the precision of measurement relative standard deviation (RSD) should not exceed 15%. Our data indicated that the LC-MS/MS assay had an RSD lower than 15% for most measured steroids, although corticosterone in the medium concentration and 11-deoxycortisol at the high concentration showed an RSD of 25.8% and 20.4% respectively (Table 3).

Several studies that utilize the H295R cell line as a model use 10 μM forskolin to stimulate steroid production (Seamon et al., 1981; Winther et al., 2013). However, several publications showed a dose related increase of steroid production with forskolin exposure (Hecker et al., 2006b; Weisser et al., 2016). Our results showed a similar increase in steroid production, but we observed a maximum stimulation point at 10 μM forskolin exposure for the majority of steroids. We chose a forskolin concentration of 1.5 μM to induce steroid production at approximately the half-maximum rate, to obtain more insights into the dynamics of steroid production in H295R cells.

By measuring the steroid metabolome at several different time points, we gain a more accurate assessment of the impact of pharmaceutical compounds and environmental toxicants on steroidogenesis (Mangelis et al., 2016). In the second part of this study, we measured changes in the 19 steroids from the metabolome from H295R cells cultured at multiple time points from 0 to 72 h. We found that precursor steroids such as progesterone, pregnenolone, dehydroepiandrosterone and 21-hydroxyprogesterone reached their highest production point at 24 h after treatment with 1.5 μM forskolin. Moreover, hormone production of end-products such as estradiol, testosterone and cortisol continue to increase until 72 h after treatment similar to previous reports (Rainey et al., 1993).

Although this method provides a comprehensive overview of the

steroidogenesis pathway in H295R cells, it lacks the measurement of cholesterol, which is the primary precursor for this process (Cherradi et al., 2001). In steroidogenic cells, sources for cholesterol are *de novo* synthesis, intracellular cholesteryl ester and lipoprotein cholesterol from the blood, with the latter being the primary source of cholesterol used in steroidogenic cells (Preslock, 1980; van Leusden and Villee, 1965). Extracellular cholesterol has been reported to account for around 80% of adrenal steroid production (Borkowski et al., 1970; Gwynne and Strauss, 1982), which suggests that any reduction in blood cholesterol or in cell culture medium may affect steroid production (Azhar et al., 1981; Christie et al., 1979). However, this key aspect of hormone biosynthesis has not been addressed in previous studies of steroidogenesis (Boggs et al., 2016; Weisser et al., 2016).

In our modelling approach, we have assumed first order kinetics. This was also used by Mangelis et al. (2016), and is valid for reactions under the Michaelis-Menten assumption that are not saturated (Johnson and Goody, 2011). The forskolin concentration of our experiment was 1.5 μM compared to 10 μM in Mangelis et al. (2016). This justifies that the assumption could be applied also here, which our results support (Fig. 2). Furthermore, we only consider the extracellular concentrations and treat the system as one compartment. This assumption can be justified since the intracellular concentrations of steroids are small compared to the extracellular concentrations and the reaction rates then are likely to be governed by the extracellular concentrations. In addition, the main point of this work is not to study the exact reaction rates of steroidogenesis, but rather motivate how measurements of all the steroids could help build more complete models for the steroidogenesis.

The obtained parameter values show that some of the reactions are not needed to fit the data and were therefore deleted. These reactions,

however, may be active and of importance *in vivo*. The large standard deviations of k_1, k_2 and k_{22} are due to the introduction of the flux v_{22}. There is no restriction on how much that should flow out of the system, making several parameter values possible (Uemura et al., 2010). To compensate for this, one could for instance introduce an extra term in the objective function to penalize large outflows. However, this flux is necessary to be able to fit the concentrations on the left branch of Fig. 4.

We have shown that the data produced by this approach can be used to build a dynamic model for all the steroid concentrations of steroidogenesis. Our model is able to predict the main trends of the measurement data using first order kinetics. Other mathematical models of steroidogenesis have previously been presented elsewhere (Breen et al., 2011; Mangelis et al., 2016). Both of these studies model the intracellular and extracellular concentrations of the various steroids and apply more advanced kinetics. However, these models do not include all the steroid hormones as considered in our model.

5. Conclusion

Using the adrenocortical H295R cell model we have developed a sensitive LC-MS/MS based method that enables us to measure all the components of the total steroidogenesis pathway except for its precursor cholesterol. Based on timeline studies of H295R cells treated with forskolin at 50% of saturation, we showed that such data can be used to develop a dynamic model. This dynamic model can enhance our understanding of the steroidogenesis process and our ability to predict the effects of drugs and environmental toxicants on this pathway. This method can be used in mechanistic studies of H295R adrenal steroid production as well as for a more in-depth view of the intermediate metabolome which could be of environmental and toxicological importance.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tiv.2018.07.002.

References

Abdel-Khalik, J., Björklund, E., Hansen, M., 2013. Development of a solid phase extraction method for the simultaneous determination of steroid hormones in H295R cell line using liquid chromatography-tandem mass spectrometry. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 935, 61–69. https://doi.org/10.1016/j.jchromb.2013.07.013.

Alder, L., Greulich, K., Kempe, G., Vieth, B., 2006. Residue analysis of 500 high priority pesticides: better by GC-MS or LC-MS/MS? Mass Spectrom. Rev. 25, 838–865. https://doi.org/10.1002/mas.20091.

Armbruster, D.A., Pry, T., 2008. Limit of blank, limit of detection and limit of quantitation. Clin. Biochem. Rev. 29 (1), S49–S52 Suppl. doi:citeulike-article-id:3416410.

Azhar, S., Menon, M., Menon, K.M.J., 1981. Receptor-mediated gonadotropin action in the ovary demonstration of acute dependence of rat luteal cells on exogenously supplied steroid precursor (sterols) for gonadotropin-induced steroidogenesis. Biochim. Biophys. Acta (BBA)/Lipids Lipid Metab. 665, 362–375. https://doi.org/10.1016/0005-2760(81)90248-4.

Bird, I.M., Pasquarette, M.M., Rainey, W.E., Mason, J.I., 1996. Differential control of 17 alpha-hydroxylase and 3 beta-hydroxysteroid dehydrogenase expression in human adrenocortical H295R cells. J. Clin. Endocrinol. Metab. 81, 2171–2178. https://doi.org/10.1210/jcem.81.6.8964847.

Boggs, A.S.P., Bowden, J.A., Galligan, T.M., Guillette, L.J., Kucklick, J.R., 2016.

Development of a multi-class steroid hormone screening method using liquid chromatography/tandem mass spectrometry (LC-MS/MS). Anal. Bioanal. Chem. 1, 12. https://doi.org/10.1007/s00216-016-9512-1.

Borkowski, A., Levin, S., Delcroix, C., Klastersky, J., 1970. Equilibration of plasma and adrenal cholesterol in man. J. Appl. Physiol. 28, 42–49.

Breen, M.S., Breen, M., Terasaki, N., Yamazaki, M., Conolly, R.B., 2010. Computational model of steroidogenesis in human H295R cells to predict biochemical response to endocrine-active chemicals: model development for metyrapone. Environ. Health Perspect. 118, 265–272. https://doi.org/10.1289/ehp.0901107.

Breen, M., Breen, M.S., Terasaki, N., Yamazaki, M., Lloyd, A.L., Conolly, R.B., 2011. Mechanistic computational model of steroidogenesis in H295R cells: role of oxysterols and cell proliferation to improve predictability of biochemical response to endocrine active chemical-metyrapone. Toxicol. Sci. 123, 80–93. https://doi.org/10.1093/toxsci/kfr167.

Cai, W., Benitez, R., Counsell, R.E., Djanegara, T., Schteingart, D.E., Sinsheimer, J.E., Wotring, L.L., 1995. Bovine adrenal cortex transformations of mitotane [1-(2-chlorophenyl)-1-(4-chlorophenyl)-2,2-dichloroethane; o,p'-DDD] and its p,p'- and m,p'-isomers. Biochem. Pharmacol. 49, 1483–1489. https://doi.org/10.1016/0006-2952(95)00028-X.

Charles, G.D., 2004. In vitro models in endocrine disruptor screening. ILAR J. https://doi.org/10.1093/ilar.45.4.494.

Cheng, C.Y., Hornsby, P.J., 1992. Expression of 11 beta-hydroxylase and 21-hydroxylase in long-term cultures of bovine adrenocortical cells requires extracellular matrix factors. Endocrinology 130, 2883–2889. https://doi.org/10.1210/endo.130.5.1572301.

Cherradi, N., Bideau, M., Arnaudeau, S., Demaurex, N., James, R.W., Azhar, S., Capponi, A.M., 2001. Angiotensin II promotes selective uptake of high density lipoprotein cholesterol esters in bovine adrenal glomerulosa and human adrenocortical carcinoma cells through induction of scavenger receptor class B type I. Endocrinology 142, 4540–4549. https://doi.org/10.1210/en.142.10.4540.

Christie, M., Strauss III, J., Flickinger, G., 1979. Effect of reduced blood cholesterol on sterol and steroid metabolism by rat luteal tissue. Endocrinology 105, 92.

Cohen, A.I., Bloch, E., Celozzi, E., 1957. In vitro response of functional experimental adrenal tumors to Corticotropin (ACTH). Exp. Biol. Med. 95, 304–309. https://doi.org/10.3181/00379727-95-23202.

Conolly, R.B., Lutz, W.K., 2004. Nonmonotonic dose-response relationships: mechanistic basis, kinetic modeling, and implications for risk assessment. Toxicol. Sci. 77, 151–157. https://doi.org/10.1093/toxsci/kfh007.

Denner, K., Rainey, W.E., Pezzi, V., Bird, I.M., Bernhardt, R., Mathis, J.M., 1996. Differential regulation of 11 beta-hydroxylase and aldosterone synthase in human adrenocortical H295R cells. Mol. Cell. Endocrinol. 121, 87–91 doi:0303720796038531. (pii).

Faupel-Badger, J.M., Fuhrman, B.J., Xu, X., Falk, R.T., Keefer, L.K., Veenstra, T.D., Hoover, R.N., Ziegler, R.G., 2010. Comparison of liquid chromatography-tandem mass spectrometry, RIA, and ELISA methods for measurement of urinary estrogens. Cancer Epidemiol. Biomark. Prev. 19, 292–300. https://doi.org/10.1158/1055-9965.EPI-09-0643.

Fox, J.E., 2004. Chemical communication threatened by endocrine-disrupting chemicals. Environ. Health Perspect. 112, 648–653.

Gazdar, A.F., Oie, H.K., Shackleton, C.H., Chen, T.R., Triche, T.J., Myers, C.E., Chrousos, G.P., Brennan, M.F., Stein, C.A., La Rocca, R.V., 1990. Establishment and characterization of a human adrenocortical carcinoma cell line that expresses multiple pathways of steroid biosynthesis. Cancer Res. 50, 5488–5496.

Gracia, T., Hilscherova, K., Jones, P.D., Newsted, J.L., Zhang, X., Hecker, M., Higley, E.B., Sanderson, J.T., Yu, R.M.K., Wu, R.S.S., Giesy, J.P., 2006. The H295R system for evaluation of endocrine-disrupting effects. Ecotoxicol. Environ. Saf. 65, 293–305. https://doi.org/10.1016/j.ecoenv.2006.06.012.

Guo, T., Taylor, R.L., Singh, R.J., Soldin, S.J., 2006. Simultaneous determination of 12 steroids by isotope dilution liquid chromatography-photospray ionization tandem mass spectrometry. Clin. Chim. Acta 372, 76–82. https://doi.org/10.1016/j.cca.2006.03.034.

Gwynne, J.T., Strauss, J.F., 1982. The role of lipoproteins in steroidogenesis and cholesterol metabolism in steroidogenic glands. Endocr. Rev. 3, 299–329. https://doi.org/10.1210/edrv-3-3-299.

Harvey, P.W., Everett, D.J., 2003. The adrenal cortex and steroidogenesis as cellular and molecular targets for toxicity: critical omissions from regulatory endocrine disruptor screening strategies for human health? J. Appl. Toxicol. https://doi.org/10.1002/jat.896.

Heald, A.H., Butterworth, A., Kane, J.W., Borzomato, J., Taylor, N.F., Layton, T., Kilpatrick, E.S., Rudenski, A., 2006. Investigation into possible causes of interference in serum testosterone measurement in women. Ann. Clin. Biochem. 43, 189–195. https://doi.org/10.1258/000456306776865106.

Hecker, M., Newsted, J.L., Murphy, M.B., Higley, E.B., Jones, P.D., Wu, R., Giesy, J.P., 2006a. Human adrenocarcinoma (H295R) cells for rapid in vitro determination of effects on steroidogenesis: hormone production. Toxicol. Appl. Pharmacol. 217, 114–124. https://doi.org/10.1016/j.taap.2006.07.007.

Hecker, M., Newsted, J.L., Murphy, M.B., Higley, E.B., Jones, P.D., Wu, R., Giesy, J.P., 2006b. Human adrenocarcinoma (H295R) cells for rapid in vitro determination of effects on steroidogenesis: hormone production. Toxicol. Appl. Pharmacol. 217, 114–124. https://doi.org/10.1016/j.taap.2006.07.007.

Hilscherova, K., Jones, P.D., Gracia, T., Newsted, J.L., Zhang, X., Sanderson, J.T., Yu, R.M.K., Wu, R.S.S., Giesy, J.P., 2004. Assessment of the effects of chemicals on the expression of ten steroidogenic genes in the H295R cell line using real-time PCR. Toxicol. Sci. 81, 78–89. https://doi.org/10.1093/toxsci/kfh191.

Hoofnagle, A.N., Wener, M.H., 2009. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. J. Immunol. Methods. https://doi.

org/10.1016/j.jim.2009.06.003.

Johnson, K.A., Goody, R.S., 2011. The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. Biochemistry 50, 8264–8269. https://doi.org/10.1021/bi201284u.

Kanno, J., Onyon, L., Haseman, J., Fenner-Crisp, P., Ashby, J., Owens, W., 2001. The OECD program to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses: phase 1. Environ. Health Perspect. 109, 785–794. https://doi.org/10.2307/3454820.

Kjærstad, M.B., Taxvig, C., Nellemann, C., Vinggaard, A.M., Andersen, H.R., 2010. Endocrine disrupting effects in vitro of conazole antifungals used as pesticides and pharmaceuticals. Reprod. Toxicol. 30, 573–582. https://doi.org/10.1016/j.reprotox.2010.07.009.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B., 2016. TMB: automatic differentiation and Laplace approximation. J. Stat. Softw. 70, 1–21. https://doi.org/10.18637/jss.v070.i05.

Kushnir, M.M., Rockwood, A.L., Roberts, W.L., Yue, B., Bergquist, J., Meikle, A.W., 2011. Liquid chromatography tandem mass spectrometry for analysis of steroids in clinical laboratories. Clin. Biochem. 44, 77–88. https://doi.org/10.1016/j.clinbiochem.2010.07.008.

Mangelis, A., Dieterich, P., Peitzsch, M., Richter, S., Jühlen, R., Hübner, A., Willenberg, H.S., Deussen, A., Lenders, J.W.M., Eisenhofer, G., 2016. Computational analysis of liquid chromatography-tandem mass spectrometric steroid profiling in NCI H295R cells following angiotensin II, forskolin and abiraterone treatment. J. Steroid Biochem. Mol. Biol. 155, 67–75. https://doi.org/10.1016/j.jsbmb.2015.09.038.

Methlie, P., Hustad, S.S, Kellmann, R., Almås, B., Erichsen, M.M., Husebye, E., Løvås, K., 2013. Multisteroid LC-MS/MS assay for glucocorticoids and androgens, and its application in Addison's disease. Endocr. Connect. 2, 125–136. https://doi.org/10.1530/EC-13-0023.

Middle, J.G., 2007. Dehydroepiandrosterone sulphate interferes in many direct immunoassays for testosterone. Ann. Clin. Biochem. 44, 173–177. https://doi.org/10.1258/000456307780118082.

Mountjoy, K.G., Bird, I.M., Rainey, W.E., Cone, R.D., 1994. ACTH induces up-regulation of ACTH receptor mRNA in mouse and human adrenocortical cell lines. Mol. Cell. Endocrinol. 99, R17–R20. https://doi.org/10.1016/0303-7207(94)90160-0.

O'Connor, J.C., Frame, S.R., Ladics, G.S., 2002. Evaluation of a 15-day screening assay using intact male rats for identifying antiandrogens. Toxicol. Sci. 69, 92–108. https://doi.org/10.1093/toxsci/69.1.92.

Odermatt, A., Strajhar, P., Engeli, R.T., 2016. Disruption of steroidogenesis: cell models for mechanistic investigations and as screening tools. J. Steroid Biochem. Mol. Biol. 158, 9–21. https://doi.org/10.1016/j.jsbmb.2016.01.009.

OECD, 2011. Test No. 456: H295R steroidogenesis assay. In: OECD Guidelines for the Testing of Chemicals, Section 4. https://doi.org/10.1787/9789264122642-en.

Penning, T.M., Lee, S.H., Jin, Y., Gutierrez, A., Blair, I.A., 2010. Liquid chromatographymass spectrometry (LC-MS) of steroid hormone metabolites and its applications. J. Steroid Biochem. Mol. Biol. 121, 546–555. https://doi.org/10.1016/j.jsbmb.2010.01.005.

Preslock, J.P., 1980. A review of in vitro testicular steroidogenesis in rodents, monkeys and humans. J. Steroid Biochem. 13, 965–975. https://doi.org/10.1016/0022-4731(80)90172-7.

Rainey, W.E., Bird, I.M., Sawetawan, C., Hanley, N.A., Mc Carthy, J.L., Mc Gee, E.A., Wester, R., Mason, J.I., 1993. Regulation of human adrenal carcinoma cell (NCI-H295) production of C19 steroids. J. Clin. Endocrinol. Metab. 77, 731–737. https://doi.org/10.1210/jcem.77.3.8396576.

Rainey, W.E., Bird, I.M., Mason, J.I., 1994. The NCI-H295 cell line: a pluripotent model for human adrenocortical studies. Mol. Cell. Endocrinol. 100, 45–50. https://doi.org/10.1016/0303-7207(94)90277-1.

Rauh, M., 2010. Steroid measurement with LC-MS/MS. application examples in pediatrics. J. Steroid Biochem. Mol. Biol. https://doi.org/10.1016/j.jsbmb.2009.12.007.

Rijk, J.C.W., Peijnenburg, A.A.C.M., Blokland, M.H., Lommen, A., Hoogenboom, R.L.A.P., Bovee, T.F.H., 2012. Screening for ModulatoryEffects on SteroidogenesisUsing the human H295R adrenocortical cell line: a metabolomics approach. Chem. Res. Toxicol. 25, 1720–1731.

Seamon, K.B.K.B.B., Padgett, W., Daly, J.W.J.W.W., 1981. Forskolin: unique diterpene activator of adenylate cyclase in membranes and in intact cells. Proc. Natl. Acad. Sci. U. S. A. 78, 3363–3367. https://doi.org/10.1073/pnas.78.6.3363.

Singh, R.J., 2008. Validation of a high throughput method for serum/plasma testosterone using liquid chromatography tandem mass spectrometry (LC-MS/MS). Steroids 73, 1339–1344. https://doi.org/10.1016/j.steroids.2008.07.006.

Stanczyk, F.Z., Lee, J.S., Santen, R.J., 2007. Standardization of steroid hormone assays: why, how, and when? Cancer Epidemiol. Biomarkers & Prev. 16, 1713–1719. https://doi.org/10.1158/1055-9965.EPI-06-0765.

Stenman, U.-H., 2013. Standardization of hormone determinations. Best Pract. Res. Clin. Endocrinol. Metab. 27, 823–830. https://doi.org/10.1016/j.beem.2013.10.007.

Szécsi, M., Tóth, I., Gardi, J., Nyári, T., Julesz, J., 2004. HPLC-RIA analysis of steroid hormone profile in a virilizing stromal tumor of the ovary. J. Biochem. Biophys. Methods 61, 47–56. https://doi.org/10.1016/j.jbbm.2004.04.008.

Tomoya, Y., Ueda, S., Yoshioka, K., Kawamura, S., Seki, T., Okuno, Y., Mikami, N., 2003. Lack of estrogenic or (anti-)androgenic effects of d-phenothrin in the uterotrophic and Hershberger assays. Toxicology 186, 227–239. https://doi.org/10.1016/S0300-483X(02)00750-3.

Toppari, J., Larsen, J.C., Christiansen, P., Giwercman, A., Grandjean, P., Guillette, L.J., Jégou, B., Jensen, T.K., Jouannet, P., Keiding, N., Leffers, H., McLachlan, J.A., Meyer, O., Müller, J., Rajpert-De Meyts, E., Scheike, T., Sharpe, R., Sumpter, J., Skakkebaek, N.E., Skakkebaek, N.E., 1996. Male reproductive health and environmental xenoestrogens. Environ. Health Perspect. 104 (Suppl 4), 741–803.

Uemura, M., Honma, S., Chung, S., Takata, R., Furihata, M., Nishimura, K., Nonomura, N., Nasu, Y., Miki, T., Shuin, T., Fujioka, T., Okuyama, A., Nakamura, Y., Nakagawa, H., 2010. 5aDH-DOC (5α-dihydro-deoxycorticosterone) activates androgen receptor in castration-resistant prostate cancer. Cancer Sci. 101, 1897–1904. https://doi.org/10.1111/j.1349-7006.2010.01620.x.

Wang, S., Rijk, J.C.W., Besselink, H.T., Houtman, R., Peijnenburg, A.A.C.M., Brouwer, A., Rietjens, I.M.C.M., Bovee, T.F.H., 2014. Extending an in vitro panel for estrogenicity testing: the added value of bioassays for measuring antiandrogenic activities and effects on steroidogenesis. Toxicol. Sci. 141, 78–89. https://doi.org/10.1093/toxsci/kfu103.

Weisser, J.J., Hansen, C.H., Poulsen, R., Larsen, L.W., Cornett, C., Styrishave, B., 2016. Two simple cleanup methods combined with LC-MS/MS for quantification of steroid hormones in in vivo and in vitro assays. Anal. Bioanal. Chem. 408, 4883–4895. https://doi.org/10.1007/s00216-016-9575-z.

WHO, 2012. State of the science of endocrine disrupting chemicals - 2012. World Heal. Organ. 1–289. https://doi.org/10.1016/j.toxlet.2012.03.020.

Winther, C.S., Nielsen, F.K., Hansen, M., Styrishave, B., 2013. Corticosteroid production in H295R cells during exposure to 3 endocrine disrupters analyzed with LC-MS/MS. Int. J. Toxicol. 32, 219–227. https://doi.org/10.1177/1091581813484366.

Zhang, X., Yu, R.M.K., Jones, P.D., Lam, G.K.W., Newsted, J.L., Gracia, T., Hecker, M., Hilscherova, K., Sanderson, J.T., Wu, R.S.S., Giesy, J.P., 2005. Quantitative RT-PCR methods for evaluating toxicant-induced effects on steroidogenesis using the H295R cell line. Environ. Sci. Technol. 39, 2777–2785. https://doi.org/10.1021/es048679k.

Paper D:

Effects of defined mixtures of POPs and endocrine disruptors on the steroid metabolome of the human H295R adrenocortical cell line

Kareem Eldin Mohammed Ahmed, **Håvard G. Frøysa**, Odd André Karlsen, Nello Blaser, Karin Elisabeth Zimmer, Hanne Friis Berntsen, Steven Verhaegen, Erik Ropstad, Ralf Kellmann and Anders Goksøyr

Effects of defined mixtures of POPs and endocrine disruptors on the steroid metabolome of the human H295R adrenocortical cell line

Kareem Eldin Mohammed Ahmed [a], Håvard G. Frøysa [b], Odd André Karlsen [a], Nello Blaser [b], Karin Elisabeth Zimmer [c], Hanne Friis Berntsen [d, e], Steven Verhaegen [d], Erik Ropstad [d], Ralf Kellmann [f], Anders Goksøyr [a, *]

[a] Department of Biological Sciences, University of Bergen, P.O. Box 7803, N-5020 Bergen, Norway
[b] Department of Mathematics, University of Bergen, P.O. Box 7803, N-5020 Bergen, Norway
[c] Department of Basic Sciences and Aquatic Medicine, Faculty of Veterinary Medicine, Norwegian University of Life Sciences (NMBU), P.O. Box 8146 Dep. N-0033, Oslo, Norway
[d] Department of Production Animal Clinical Sciences, Faculty of Veterinary Medicine, Norwegian University of Life Sciences (NMBU), P.O. Box 8146 Dep. N-0033, Oslo, Norway
[e] Department of Administration, Lab Animal Unit, National Institute of Occupational Health, P.O. Box 5330 Majorstuen, N-0304, Oslo, Norway
[f] Hormone Laboratory, Haukeland University Hospital, N-5021 Bergen, Norway

HIGHLIGHTS

- We assessed effects of environmentally relevant mixtures of contaminants on steroidogenesis.
- We utilized the H295R cell line model and LC-MS/MS to analyze all steroids in the pathway.
- The highest concentration of the combined brominated and perfluorinated mixture produced the strongest effect.
- Indication of synergistic effects were observed when a statistical model was used to analyze the data.

ARTICLE INFO

ABSTRACT

The presence of environmental pollutants in our ecosystem may impose harmful health effects to wildlife and humans. Several of these toxic chemicals have a potential to interfere with the endocrine system. The adrenal cortex has been identified as the main target organ affected by endocrine disrupting chemicals. The aim of this work was to assess exposure effects of defined and environmentally relevant mixtures of chlorinated, brominated and perfluorinated chemicals on steroidogenesis, using the H295R adrenocortical cell line model in combination with a newly developed liquid chromatography tandem mass spectrometry (LC-MS/MS) method. By using this approach, we could simultaneously analyze 19 of the steroids in the steroid biosynthesis pathway, revealing a deeper insight into possible disruption of steroidogenesis. Our results showed a noticeable down-regulation in steroid production when cells were exposed to the highest concentration of a mixture of brominated and fluorinated compounds (10,000-times human blood values). In contrast, up-regulation was observed with estrone under the same experimental condition, as well as with some other steroids when cells were exposed to a perfluorinated mixture (1000-times human blood values), and the mixture of chlorinated and fluorinated compounds. Interestingly, the low concentration of the perfluorinated mixture alone produced a significant, albeit small, down-regulation of pregnenolone, and the total mixture a similar effect on 17-hydroxypregnenolone. Other mixtures resulted in only slight deviations from the control. Indication of synergistic effects were noted when we used a statistical model to improve data interpretation. A

* Corresponding author.
E-mail addresses: Kareem.Ahmed@uib.no (K.E.M. Ahmed), Havard.Froysa@uib.no (H.G. Frøysa), odd.karlsen@uib.no (O.A. Karlsen), Nello.Blaser@uib.no (N. Blaser), karin.zimmer@nmbu.no (K.E. Zimmer), Hanne.Berntsen@stami.no (H.F. Berntsen), steven.verhaegen@nmbu.no (S. Verhaegen), erik.ropstad@nmbu.no (E. Ropstad), ralf.kellmann@uib.noand (R. Kellmann), Anders.Goksoyr@uib.no (A. Goksøyr).

potential for adverse outcomes of human exposures is indicated, pointing to the need for further investigation into these mixtures.

1. Introduction

Environmental pollution has detrimental consequences to our ecosystem and imposes harmful health effects to wildlife and humans (Bergman et al., 2013b). Approximately 16 million human deaths per year are reported to be attributed to chemical pollutant exposure (Landrigan et al., 2017). These pollutants include legacy and emerging persistent organic pollutants (POPs), which are persistent in the environment and resist degradation for decades (Barón et al., 2015), and can be found in various environmental compartments in low concentrations. Some compounds have the potential to interfere with the endocrine system (Rhind, 2008) and are designated as *endocrine disrupting chemicals* (EDCs) (Bergman et al., 2013a; Schug et al., 2016). Previous work have identified the adrenal cortex as the main target organ affected by endocrine disrupting chemicals (Bergman et al., 2013b).

The presence of POPs in the food chain is a consequence of their lipophilic characteristics, which accelerate their uptake by plants or animals (Guillette et al., 1996; Nimrod and Benson, 1996). Some POPs have the ability to disrupt endocrine functions, even when they are present in very low concentrations (Li et al., 2006). The Stockholm Convention on Persistent Organic Pollutants (SCPOP) is an international treaty aimed to eliminate or reduce POPs in the environment. The majority of POPs listed in SCPOP are halogenated molecules and comprise chlorinated, brominated, and fluorinated compounds (UNEP, 2008). Also, it is reported that exposure to POPs have been linked to adverse health effects including cancer, metabolic disorders, and reproductive defects (Berg et al., 2016; Hotchkiss et al., 2008; Zimmer et al., 2011).

Perfluorinated chemicals such as perfluoroalkyl acids, are commonly used in consumer products as a stain, water repellent material in clothing or as non-stick coatings on cookware (Stahl et al., 2011). Perfluorooctanoic acid (PFOA) was reported to increase estradiol levels in exposed humans and decrease testosterone levels in isolated rat Leydig cells (Olsen et al., 1998; Zhao et al., 2010). Also, decreased gene expression of key enzymes associated with steroidogenesis was observed in mice and rats exposed to perfluorododecanoic acid (PFDoDA) and perfluorooctanesulfonic acid (PFOS) (Gorrochategui et al., 2014; Wan et al., 2011). Due to the global distribution and persistence of the perfluorinated compounds in the environment concerns were raised over their potential human toxicity, especially their ability to act as endocrine disrupters (Du et al., 2013; Hines et al., 2009).

Chlorinated compounds encompass industrial chemicals such as polychlorinated biphenyls (PCBs) and pesticides such as dichlorodiphenyltrichloroethane (DDT). Although they are now banned in most of the world, they still persist in the environment due to their resistance to degradation (Craig et al., 2011; Tiemann, 2008), and continued uses in certain countries. DDT and its main metabolite, 1,1-dichloro-2,2-bis(4-chlorophenyl)ethylene, DDE, was reported to increase progesterone production in low concentration exposures and block progesterone synthesis at high concentration exposures *in vitro* (Craig et al., 2011; Crellin et al., 2001, 1999). Moreover, PCBs are reported to cause reproductive abnormalities, carcinogenicity and endocrine disruption (Folmar et al., 1996; Guillette et al., 1996; Ropstad et al., 2006; Safe, 1994).

Brominated flame retardants (BFRs) are utilized widely for their fire-resistance. They are applied to plastics used in electronic equipment and many household items such as curtains, furniture covers and carpets. Due to their chemical properties they may leak into the environment (Alaee et al., 2003; de Wit et al., 2010). The toxic effects of BFRs have been investigated in several *in vitro* and *in vivo* studies, most of which is caused by targeting androgen and oestrogen receptors (Hamers et al., 2006; Kitamura et al., 2005; Lilienthal et al., 2006). In humans, exposure to BFRs has been linked to congenital cryptorchidism (Darnerud, 2008; Main et al., 2007).

In a realistic exposure scenario, humans are exposed to multiple EDCs simultaneously, albeit in low concentrations (Diamanti-Kandarakis et al., 2009). Based on their mode of action, mixture toxicity can be sorted in two groups: 1) mixtures that show no direct interaction between compounds when they exert their effects and 2) mixtures that show interaction between compounds. Effects caused by the first category are normally explained by *addition*, which means that chemicals act independently of each other targeting the same tissue with similar modes of action. In the second category, compounds can act in a synergistic manner indicating higher than additive effect predictions, or an antagonistic manner, which indicate lower than predicted additive effects (McCarty and Borgert, 2006; Monosson, 2005; Rizzati et al., 2016). Current policies concerning EDC risk assessment regulation is based on single compound exposure (Bars et al., 2012; Fuhrman et al., 2015; Kortenkamp et al., 2007).

The United State Environmental Protection Agency (EPA) and the Organization for Economic Co-operation and Development (OECD) have developed guidelines for screening and examining EDCs (Gelbke et al., 2004). Among recommended assays are the Hershberger assay (Owens et al., 2007) and Uterotrophic assay (OECD, 2007), which are both *in vivo* assays utilizing rats. In addition, the *in vitro* minced testis assay, which is a rat primary cell model is also used. These models' need for animal sacrifice or acquisition of human tissue present a limitation for a high throughput screening. Therefore, there is a demand for alternative *in vitro* cell models for the growing number of chemicals that require testing for potential endocrine disrupting effects. The H295R cell line has been suggested by both the EPA and the OECD as a screening tool for investigating possible EDCs (Hecker and Hollert, 2011; Hilscherova et al., 2004; Karmaus et al., 2016; Zhang et al., 2005).

The H295R adrenocortical cell line originated from an excised adrenocortical carcinoma (Gazdar et al., 1990). H295R cells have all of the adrenocortical enzymes, which give them the capacity to produce 30 different steroids (Bird et al., 1996; Gazdar et al., 1990; Rainey et al., 1994, 1993). The H295R cells express low levels of the adrenocorticotropic hormone (ACTH) receptor, and thus unresponsive to ACTH, the physiological stimulus of adrenal steroidogenesis. However, stimulation of steroidogenesis can be initiated via forskolin, a natural diterpene compound and a cyclic adenosine monophosphate (cAMP)-dependent activator (Mountjoy et al., 1994; Rainey et al., 1993). Stimulation of H295R cells with forskolin leads to a rapid increase in steroid production (Denner et al., 1996; Weisser et al., 2016; Ahmed et al., 2018).

The triple quadrupole liquid chromatography tandem mass spectrometer (LC-MS/MS) is rapidly taking place as an instrument of choice to measure steroid hormones (Rauh, 2010). In this paper,

we utilized a newly developed steroidogenesis assay, which comprises both LC-MS/MS and the H295R cell line model. This assay has ultra-high performance and is capable of simultaneously assessing and quantifying 19 analytes in the steroidogenesis pathway, providing a comprehensive insight into steroid biosynthesis (Ahmed et al., 2018).

In the present paper, we investigated the endocrine disrupting potential of carefully defined environmentally relevant mixtures and sub-mixtures of POPs with known compositions. The selection of POPs was largely based on the compounds listed as such under the SCPOP, and further on a literature review of recent publications on POPs occurring at the highest levels in Scandinavian blood (Berntsen et al., 2017). We used the H295R cell line and studied effects of several POP mixtures on the steroidogenesis by analysing the steroid metabolome using our newly developed LC-MS/MS system, which take into account active and dormant condition of H295R cells (Ahmed et al., 2018).

2. Materials and methods

2.1. Chemicals

All PBDEs, PCBs and other organochlorines were originally purchased from Chiron As (Trondheim, Norway). All perfluorinated compounds were obtained from Sigma-Aldrich (St. Louis, MO, USA) except perfluorohexanesulfonic acid (PFHxS) that was obtained from Santa Cruz (Dallas, US). Hexabromocyclododecane (HBCD), phosphate buffered saline (PBS) and dimethyl sulfoxide (DMSO) were obtained from Sigma–Aldrich (Dorset, UK). More details about the chemicals can be found in Berntsen et al. (2017). Cell culture reagents were supplied by Life Technologies (Paisley, UK) and Sigma-Aldrich.

2.2. POP mixtures

Mixtures of the test POPs were designed in concentration ratios relevant to human exposure and premade at the Norwegian University of Life Sciences, Oslo as described (Berntsen et al., 2017). Seven mixtures were used in the cell assays (Table 1): (1) total mixture, containing all the 29 test compounds (B + C + F), (2) perfluorinated mixture (F), (3) brominated mixture (B), (4) chlorinated mixture (C), (5) brominated and perfluorinated mixture (B + F), (6) chlorinated and perfluorinated mixture (C + F) and (7) brominated and chlorinated mixture (B + C). The chemicals included in the mixtures and their respective concentrations in the stock solution are shown in Table 1.

2.3. Cell culture

The H295R cell line was purchased from American Type Culture Collection (ATCC). Cells were cultured in 75 cm^2 flasks in Dulbecco's modified Eagle medium/HamF12 (DMEM/F12) containing HEPES buffer, L-glutamine and pyridoxine HCl (Gibco, Invitrogen, Paisley, UK). Additional supplements were added to the medium which included 1% ITS + premix (BD Biosciences, Bedford, MA) and charcoal stripped foetal bovine serum 5% (F7524, Sigma Aldrich). H295R cells were incubated at 37 °C with 5% CO$_2$ in a humidified atmosphere. The medium was changed every 2–3 days and cells passaged at approximately 80% confluence by a brief exposure to 0.25% trypsin/0.53 mM EDTA (Gibco, Invitrogen). The cells were used between passages 4–6.

2.4. Mixture exposures

After seeding H295R cells for 24 h in 6 well plates at a cell

Table 1
The composition and measured concentrations of the various compounds in the total mixture stock in mg/ml. The estimated concentration of POPs in the C (Chlorinated compounds), F (Perfluorinated compounds) B (Brominated compounds) stock solutions, prepared as described in Berntsen et al. (2017).

Compound	Mixture stock concentration (mg/ml)					
	B + C + F	F	B	C	B + F	C + F	B + C
Perfluorinated compounds (F)						
PFOA	1.743	1.743			17.43	1.743	
PFOS	22.348	22.348			223.48	22.348	
PFDA	0.193	0.193			1.93	0.193	
PFNA	0.507	0.507			5.07	0.507	
PFHxS	3.422	3.422			34.22	3.422	
PFUnDA	0.190	0.190			1.90	0.190	
Brominated compounds (B)						
BDE-209	0.009			0.009	0.09		0.009
BDE-47	0.009			0.009	0.09		0.009
BDE-99	0.004			0.004	0.04		0.004
BDE-100	0.002			0.002	0.02		0.002
BDE-153	0.001			0.001	0.01		0.001
BDE-154	0.002			0.002	0.02		0.002
HBCD	0.035			0.035	0.35		0.035
Chlorinated compounds (C)						
PCBs						
PCB 138	0.155			0.155		0.155	0.155
PCB 153	0.252			0.252		0.252	0.252
PCB 101	0.008			0.008		0.008	0.008
PCB 180	0.134			0.134		0.134	0.134
PCB 52	0.006			0.006		0.006	0.006
PCB 28	0.008			0.008		0.008	0.008
PCB 118	0.045			0.045		0.045	0.045
OCPs						
p, p'-DDE	0.339			0.339		0.339	0.339
HCB	0.065			0.065		0.065	0.065
α-chlordane	0.010			0.010		0.010	0.010
oxychlordane	0.014			0.014		0.014	0.014
trans-nonachlor	0.044			0.044		0.044	0.044
α-HCH	0.005			0.005		0.005	0.005
β-HCH	0.022			0.022		0.022	0.022
γ-HCH (Lindane)	0.005			0.005		0.005	0.005
Dieldrin	0.021			0.021		0.021	0.021

density of 1.2×10^6 cells per well, fresh medium was added containing different mixture concentrations based on tenfold dilution. Concentrations were designated (Low, Medium, High, Very high) corresponding to 1, 10, 100 and 1000 times the estimated concentrations in human blood. The combined B + F sub-mixture was ten times more concentrated than the other mixtures (Berntsen et al., 2017), and tested at 10000x human blood levels as the highest concentration (denoted VH$_{10}$). H295R cells were exposed to the mixtures for 48 h. Each concentration had a total of 9 replicate wells acquired from three independent experiments. Parallel exposures were performed with unstimulated (DMSO-treated) and 1.5 μM forskolin stimulated cells.

2.5. Cell viability

Cell viability was evaluated using Alamar Blue TM assay (Invitrogen) on the 96-well microplates (VWR, USA). Approximately 50,000 cells were seeded for 24 h before exposure for 48 h. Each exposure was performed in triplicate. The medium was removed and replaced with 100 μl of fresh medium +10 μl Alamar Blue assay solution for 3 h at 5% CO$_2$ at 34 °C. A PerkinElmer (EnSpire 2300 Multilabel Reader) spectrophotometer was used to read the plates. The absorbance was read at 570 nm and 600 nm and viability was expressed as percentage of control (medium with 0.25% DMSO). Triton X-100 (10%) was used as a positive control of cell death.

2.6. LC-MS/MS

LC-MS/MS analysis was carried out by isotope-dilution mass spectrometry as described previously (Ahmed et al., 2018). Briefly, samples of H295R cell medium were extracted using liquid-liquid extraction on a Hamilton Star pipetting robot, and 85 μl of sample was used for analysis in addition to 10 μl of internal standard that was added to all samples. LC-MS/MS analysis was performed on a Waters Xevo TQ-S triple quadrupole mass spectrometer that was coupled to a Waters i-class Acquity UPLC. Ionisation was achieved by electrospray ionisation (ESI) in positive and negative mode. Steroid hormones were detected and quantitated by isotope-dilution mass spectrometry by multi-reaction monitoring (MRM).

2.7. Data analysis and statistical testing, and mixture effect modeling

All values were transformed to fold change values prior to data analysis. First, undetected values and concentration values below the low limit of detection were replaced with half the lowest limit of detection. The concentrations were then divided by the average concentration in the corresponding control group to obtain fold change values. Finally, the average fold change values across all the biological and technical replicates were calculated. To test if a treatment had an effect on a steroid a standard t-test was performed with a null hypothesis of no change (fold change of 1) and significance level $\alpha = 0.05$. The t-tests were adjusted for multiple testing to reduce the number of false significances (Holm, 1979).

Using the single mixtures, predictions for their combined exposure were made by assuming additive effects (Groten et al., 2001). We assume that FC_1 and FC_2 are the fold changes for exposure of single mixture 1 and 2, respectively. The additive model then predicts the fold change for their combined exposure to be $FC = FC_1 + FC_2 - 1$. Observed values lower or greater than these predictions are regarded as antagonistic or synergistic, respectively. To test if the observed effects were significantly non-additive an ANOVA was performed for linear models with and without interaction. Also, these tests were adjusted for multiple testing (Holm, 1979).

3. Results

3.1. Cell viability

Cell viability was evaluated using the Alamar Blue assay. All mixtures in all concentrations were tested. Fluorescence from the formed resorufin was measured and no deviation from solvent control was observed with any of the mixtures in any concentrations, whereas the positive control Triton X-100 showed an expected decrease in cell viability (Supplementary Figs. 1 and 2).

3.2. Separate exposures with brominated (B), chlorinated (C) and perfluorinated (F) mixtures

Unstimulated H295R cells exposed to the low concentration of perfluorinated mixture (F) showed a down-regulation in pregnenolone production ($p = 0.05$) (Fig. 1, Supplementary Table 1). In the high concentration exposure, an up-regulation of pregnenolone was indicated, although not statistically significant ($p = 0.39$). Forskolin-stimulated H295R cells showed up-regulation of aldosterone levels ($p = 0.002$) when exposed to the highest concentration of the F mixture (Fig. 1, Supplementary Table 1). Steroid production in unstimulated and stimulated H295R cells did not show any statistically significant deviation from the control when exposed to either brominated mixture (B) or chlorinated mixture (C) (Fig. 1).

3.3. Combined exposure with the brominated and chlorinated mixture (B + C)

Forskolin-stimulated H295R cells exposed to the B + C mixture showed no statistically significant deviation from the control, although a non-significant up-regulation of aldosterone production ($p = 0.13$) was observed in the highest concentration exposure. Steroid production in unstimulated H295R cells showed no difference from control (Figs. 1 and 3).

3.4. Combined exposure with the brominated and perfluorinated mixture (B + F)

Unstimulated H295R cells exposed to the highest concentration of the B + F mixture showed statistically significant down-regulation of all glucocorticoid steroids, as well as dehydroepiandrosterone (DHEA) and testosterone (Fig. 1, Supplementary Table 1). Significant up-regulation of hormone production was noticed for 11-deoxycorticosterone with the highest concentration of B + F mixture and progesterone in the high concentration (Figs. 1 and 2-A). Forskolin-stimulated H295R cells exposed to the highest concentration of this mixture showed a significant down-regulation for most steroids (Fig. 3). In the high concentration, up-regulation of hormone production was observed in corticosterone showing the strongest response, in addition to down-regulation of cortisone (Figs. 1 and 2-B, Supplementary Table 1).

3.5. Combined exposure with the chlorinated and perfluorinated mixture (C + F)

In unstimulated H295R cells no deviation from control was observed in steroid production (Fig. 1). In the forskolin-stimulated H295R cells a significant up-regulation of aldosterone was observed after exposure to the very high concentration of C + F mixture. Other exposure concentrations did not cause any deviations from the solvent control in steroid production (Figs. 1, 2C, 3).

3.6. Exposure with the total mixture (B + C + F)

Unstimulated H295R cells exposed to the highest concentration of the total mixture showed up-regulation trends in mineral and glucocorticoids, but not statistically significant (Fig. 1). Similarly, trends of down-regulation of androgens in the low concentration exposure were found. In the forskolin-stimulated H295R cells production of 17-hydroxy-pregnenolone was significantly down-regulated in the low concentration exposure. In the highest concentration exposure aldosterone indicated an up-regulation ($p = 0.19$) (Fig. 1, Supplementary Table 1).

3.7. Additive effect predictions

Additive effect predictions were made and compared to observed responses using the model described in Materials & Methods. Fig. 4 shows the additive effect predictions from single mixtures and the observed values for the chlorinated and perfluorinated (C + F) mixture in the forskolin stimulated condition. We observed a more than additive response in the 17-hydroxyprogesterone levels in the high exposure of this mixture when compared to single mixture levels of F and C compounds ($p < 0.05$). A similar deviation could be observed in aldosterone levels upon exposure to the very high concentration of the C + F mixture (Figs. 3 and 4). This non-additive effect had a p-value of 0.07 in the statistical test for interaction after correcting for multiple testing. In non-stimulated C + F cells, a less than additive response ($p < 0.05$) was observed with cortisone in the low
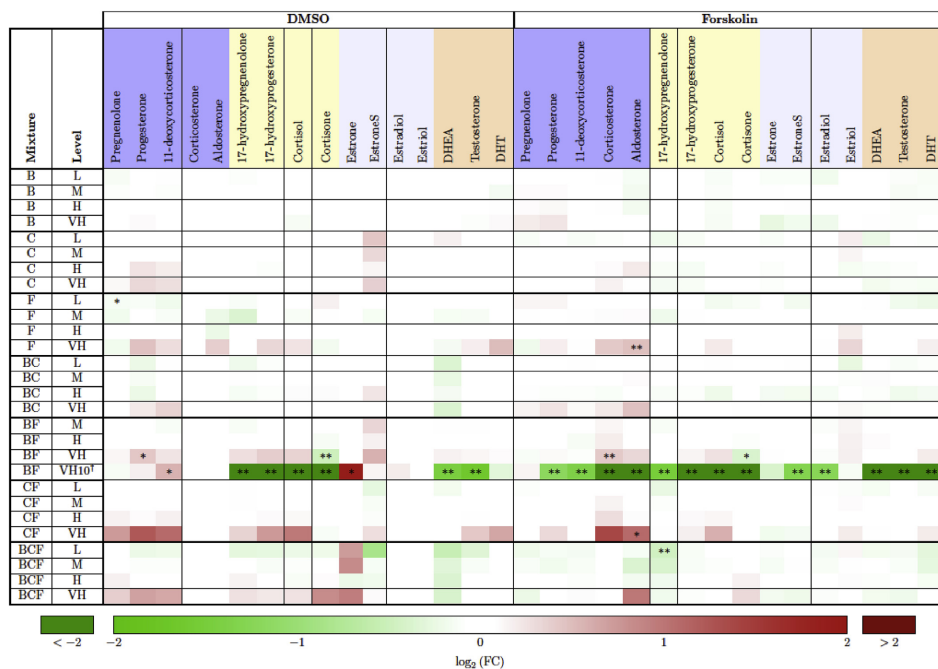
Fig. 1. Heat map of steroid production after mixture exposure of H295R cells, with and without forskolin stimulation. The fold change (FC) values for all the steroids and experiments. The mixture abbreviations are B for brominated, C for chlorinated and F for perfluorinated such that e.g. BC is the combined mixture of B and C. The level abbreviations are L for low, M for medium, H for high and VH for very high mixture concentrations. The colouring of the steroid names are according to their classification in steroidogenesis. Each value is coloured according to its log2-value where log2(FC) = 0 corresponds to no difference from control. An asterisk (*) indicates a significant effect with p ≤ 0.05 and a double asterisk (**) indicates a strong significant effect with p ≤ 0.01. All the p-values can be found in Table 1 in supplementary material. †VH$_{10}$: concentration in this combined mixture is 10 times higher than the VH concentration of B and F separate mixtures.

exposure group. This and all other model predictions are shown in Supplementary Fig. 4.

4. Discussion

Data from laboratory studies and exposed wildlife populations indicate that both the HPA axis and the adrenal glands are highly susceptible to endocrine disruption caused by POPs at environmentally relevant concentrations (Bergman et al., 2013b). In this study, we have used the H295R adrenocortical cell model and a recently established LC-MS/MS based method as an approach to identify changes in steroidogenesis after exposure to mixtures of POPs in various combinations. The most noteworthy finding was a significant down-regulation in glucocorticoids and mineralocorticoids in forskolin stimulated H295R cells when they were exposed to the highest concentration of a mixture of brominated and fluorinated compounds (10,000-times human blood values).

When exposed to brominated compounds alone, no changes in steroid production were observed in the H295R cells, although previous studies on BFRs using numerous *in vitro* and *in vivo* systems are indicative of endocrine disrupting effects (Hamers et al., 2006; Harju et al., 2007; Kitamura et al., 2005). For example, in rats exposed to BDE-99 for eight days during pregnancy a

significant decrease in estradiol levels in blood of the offspring was reported (Lilienthal et al., 2006). Also, similar effects have been reported in studies of BDE-47 exposed rat offspring (Talsness et al., 2006). Both of these compounds are present in our brominated mixture. However, the previous reports investigating BFRs using *in vitro* models indicated harmful effects only when used in much higher concentrations than in our mixture. For example, HepG2 cells exposed to 10 μM of BDE-209 showed decreased cell viability (Hu et al., 2007). This is an order of magnitude higher than the highest BDE-209 concentrations used in our mixture. Also reactive oxygen species formation increased in SH-SY5Y neuroblast cells when exposed to 2 μM BDE-47 which is 100 times more concentrated than in our exposure (He et al., 2008; Shao et al., 2008). Thus, the lack of endocrine disrupting effects in our results may be due to the lower concentrations used, but may also be due to the shorter duration of exposures as well as differences caused by using a mixture.

With regard to perfluorinated (F) compounds several *in vivo* experiments have reported endocrine disrupting potential (Benninghoff et al., 2011; Wang et al., 2011; Yang et al., 2009). PFOA and PFOS have caused an increase in estradiol levels and decrease in testosterone levels in a steroidogenic assay (Du et al., 2013). Kang et al. (2016) suggested that they target CYP19 (aromatase), which is
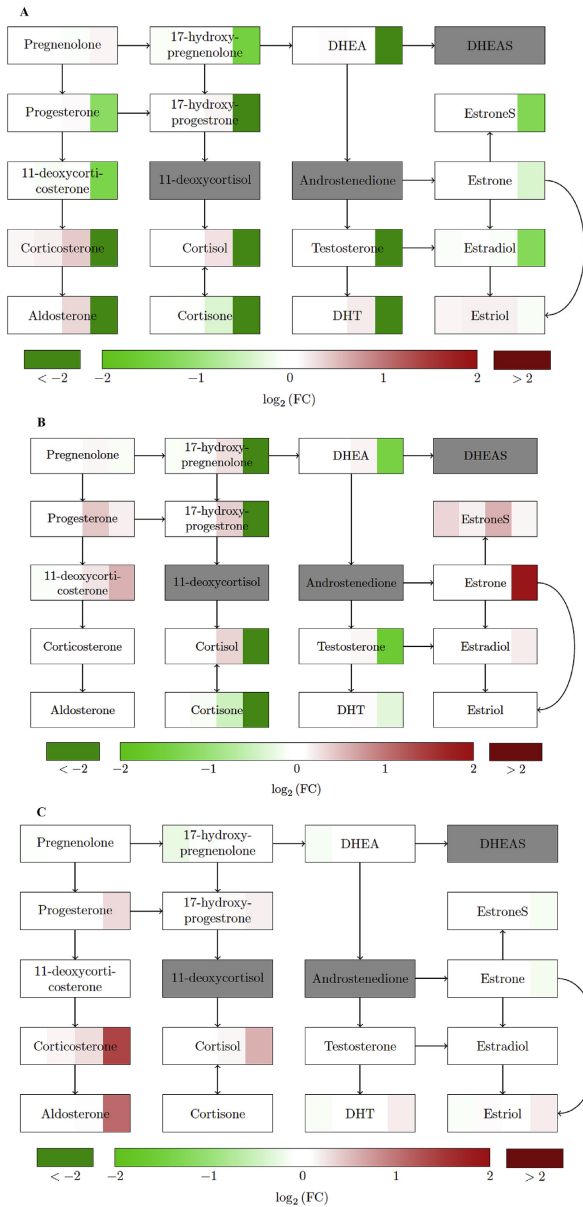
Fig. 2. Changes in steroid production in H295R cells after mixture exposures, plotted as fold change (FC) in a network representation. The box for each steroid is divided in four parts, corresponding to low, medium, high and very high mixture concentration from left to right. Each value is coloured according to its $\log_2$-value where $\log_2(FC) = 0$ corresponds to no effect. The grey boxes are missing data. A) brominated and perfluorinated (B + F) mixture in forskolin-stimulated H295R cells. B) brominated and perfluorinated (B + F) mixture in unstimulated cells. C) chlorinated and perfluorinated (C + F) mixture in stimulated cells.

the enzyme responsible for the biosynthesis of estrogens from androgens, such as testosterone. Also, H295R cells were previously exposed to PFOA, PFOS and perfluorononanoic acid (PFNA) alone, and only PFOS had effects on hormone levels, again at higher (20 times) concentration than used in our mixture (Kraugerud et al., 2011). In the current study, we observed aldosterone being elevated at the highest F exposure concentration in the forskolin stimulated condition.

Regarding the chlorinated C mixture, our data showed no effects in either stimulated or unstimulated conditions. These results contradict previous reports of PCBs affecting H295R cell hormone production by Tremoen et al. (2014), who found that exposure of PCB 118, 153 severely down-regulate testosterone production in H295R cells, while PCB 126 up-regulate cortisol production by 40%. In addition Kraugerud et al. (2010) reported significant up-regulation in estradiol and aldosterone production in single exposure of H295R cells to PCB 118, PCB 153 and PCB 126. In a previous study, a co-culture of theca and granulosa cells, an ovarian cell model, was exposed to a defined PCB mixture in addition to DDE, which caused a 15-fold increase in estradiol secretion and a strong inhibitory effect on testosterone levels (Gregoraszczuk et al., 2008). Also, Xu et al. (2006) reported a significant up-regulation of CYP11B1 and CYP11B2 in H295R cells, which could be correlated to production of cortisol and aldosterone respectively. However, the exposure concentrations used in those studies were more than 30 times higher than the highest concentrations used in the current study, again suggesting that the lack of responses observed here are due to the use of lower and more realistic concentration ranges.

A key issue, when dealing with compound mixture exposures is chemical *interaction*. Interaction could take place in several ways such as affecting the toxicokinetic phase, which include means of uptake, distribution, metabolism and excretion. In addition, interaction could modify responses of a receptor or a target organ to one chemical in the presence of another. Moreover, in reality we would be dealing with all previously mentioned scenarios of interaction at the same time, especially if the mixture contain more than two compounds (Groten, 2000; Spurgeon et al., 2010; Tipping, 1994).

The combination of the brominated and perfluorinated (B + F) mixtures resulted in the most severe changes in the steroid metabolome, especially at the highest concentration (VH$_{10}$,10,000 times blood levels), but also in the VH concentration, which is comparable to the very high concentration of the single mixtures where only a few effects were observed. Clear differences could also be observed between unstimulated and forskolin-stimulated cells (Fig. 1). Most notable is the shift from upregulation to down-regulation of estrone and down-regulation of mineralocorticoids in the stimulated cells. Previous studies on HepG2 cells that were exposed to the same mixture showed diminished cell viability, higher ROS production and decreased mitochondrial membrane potential (MMP) using the same concentrations (Wilson et al., 2016).

We observed a more than additive response in aldosterone levels upon exposure to the very high concentration of chlorinated + perfluorinated (C + F) mixture when compared to single mixture levels of C and F compounds (Fig. 3). This effect is also indicated in Fig. 4, where predicted vs observed responses are presented (p = 0.07 after correcting for multiple testing). In addition, in the same exposure concentration we observed antagonism in aldosterone levels, when we compared perfluorinated and brominated mixture to the single exposures (Fig. 3). A deviation was also observed for 17-hydroxyprogesterone in the high C + F group, which showed a significant more than additive effect (Fig. 4). These results correlate with previous work on the same mixtures as significant synergistic effects were observed for the increase of ROS production and MMP decrease in HepG2 cells (Wilson et al., 2016).

These results could lead us to speculate that there is some interaction between chemicals in the different mixtures causing more than additive or synergistic effects.

It is worth commenting, that in our results all p-values are corrected for multiple testing due to the high number of acquired measurements according to the Holm method (Holm, 1979). This procedure increases the p-values obtained from the t-tests for significant effects. Thus, effects that would be considered significant in individual mixture experiments could become non-significant with this approach. This can be observed in the bar graphs for aldosterone and cortisol where the p-values are corrected for multiple testing, but not the confidence intervals (Fig. 3). However, the correction strengthens the confidence in the effects classified as significant since stronger evidence is needed for a p-value to be significant after the correction.

In primary neonatal porcine Leydig cells, testosterone hormone production increases by more than 140-fold when stimulated with luteinizing hormone (LH) (Lervik et al., 2011). Exposure to 3-MeSO$_2$-DDE caused a dose dependent increase in testosterone levels in unstimulated Leydig cells and a down regulation in LH stimulated cells (Castellanos et al., 2013; Kalayou et al., 2016). Forskolin plays a similar role to LH in stimulating H295R cells to significantly increase the basal hormone production, e.g. an increase up to 17-fold in testosterone levels was observed in Ahmed et al. (2018). In the present study, we noticed a difference in response to the same mixture exposure depending on the physiological status of the cells. For example, progesterone and aldosterone showed a complete different secretion profile in the stimulated condition compared to the unstimulated one after exposure to the same mixture (Figs. 1, 2A and 2B). Moreover, in some instances apparently contradictory result appeared, such as the case with estrone and corticosterone (Fig. 1). These results indicate the need to use both stimulated and unstimulated H295R cells when testing for endocrine disrupting potential.

The H295R cell model is commonly used to assess *in vitro* impact of drugs and toxicants, such as environmental endocrine disruptors, on the steroidogenesis (Hecker et al., 2006; Rijk et al., 2012). Moreover, toxicological studies utilizing this model generally expose the cells unstimulated and use 10 μL forskolin as a positive control (van den Dungen et al., 2015; Wang et al., 2014). From our point of view, the use of forskolin to induce hormone production in H295R cells is necessary to utilize the full potential of the H295R assay. In the unstimulated condition several key metabolites such as aldosterone and estradiol are produced at undetectable levels (Fig. 1). In our previous work (Ahmed et al., 2018), we found that hormone secretion in H295R cells reaches 50% saturation when cells are stimulated with 1.5 μM forskolin. This is suggested to represent an optimal condition to investigate the effect of EDC mixtures on steroidogenesis, when cells are already stimulated, but steroid production is not saturated. Extrapolating to the *in vivo* situation, the unstimulated H295R cell would simulate a resting adrenal cortex, whereas the stimulated cell would reflect a stressed situation. Different toxicological responses under these conditions could be highly relevant for real life exposure situations.

There is a possibility that forskolin may interact with any of the compounds present in the mixtures and add to the complexity of the interpretation of the results. In order to address this point, we have for each mixture exposure a corresponding exposure in the DMSO condition, which we are taking into account. Several previous studies used forskolin induced H295R cells in their exposures as a base condition, e.g. Gracia et al. (2006) used 10 μM forskolin stimulation, while Krogh et al. (2010) used 5 μM forskolin. Other, more complex interactions, can of course not be excluded, but are not considered in this set-up. As one example, forskolin has been shown to interact with the transcription of several gene sets,
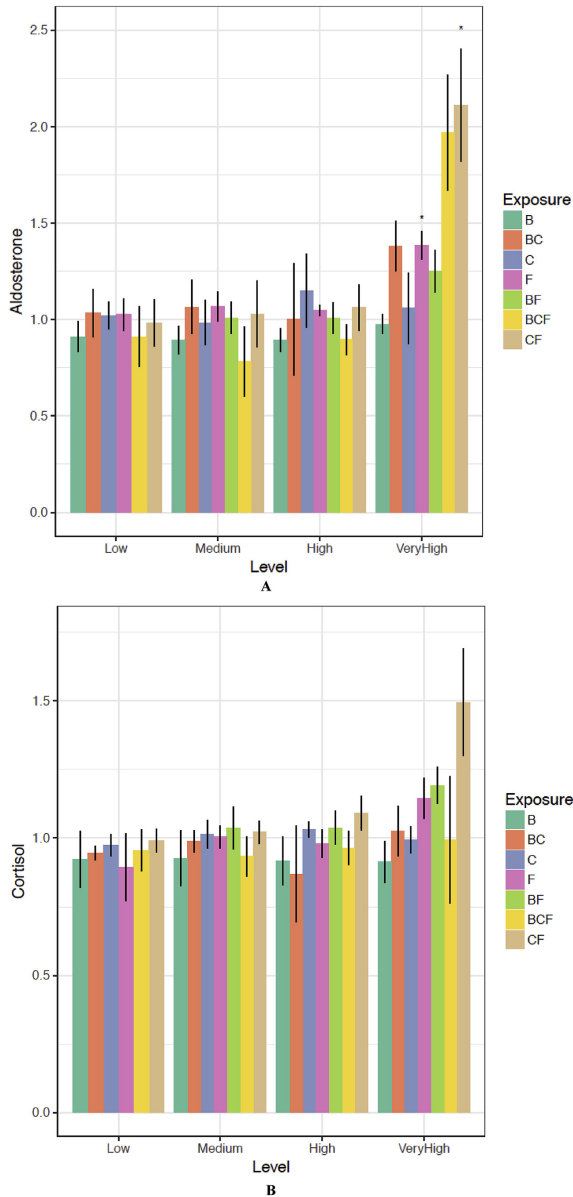
Fig. 3. Alterations in A) aldosterone and B) cortisol production after mixture exposure of H295R cells. Fold change is plotted for all mixtures used with forskolin-stimulated H295R cells. The mixture abbreviations are B for brominated, C for chlorinated and F for perfluorinated such that e.g. BC is the combined mixture of B and C. A star (*) indicates a significant effect with $p \leq 0.05$. In addition, a 95% confidence interval is plotted for each value. Note that the confidence intervals are not corrected for multiple testing such that an effect may not be significant even though the confidence interval does not contain 1.
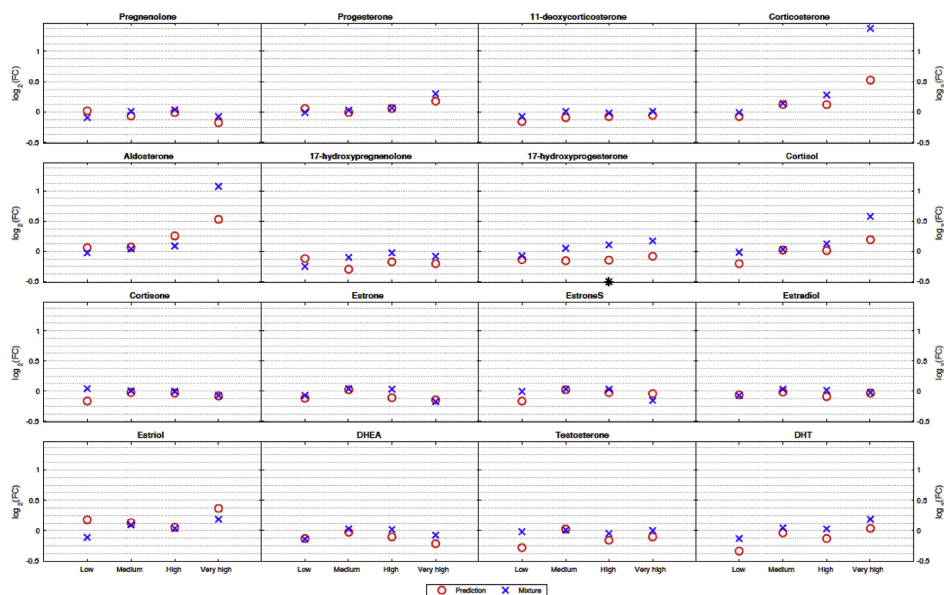
Fig. 4. Predicted (o) and observed (x) fold change (FC) values for all the steroids after exposure to the chlorinated and perfluorinated (C + F) mixture in H295R forskolin-stimulated cells. The predicted values are calculated using the single mixtures and assuming additive effects. The values are plotted on log2-scale such that log2(FC) = 0 corresponds to no effect. The difference between the predicted and observed value of 17-hydroxyprogesterone in the high concentration was the only statistically significant difference (p ≤ 0.05) and is marked with an asterisk (*).

including estrogen receptor in breast cancer cells. Upon activation by forskolin, cAMP binds to protein kinase A which in turn activates the estrogen receptor (Al-Dhaheri and Rowan, 2007). This could, potentially, change the toxicodynamic properties of EDCs in some cell types.

Statistical models provide an ideal platform to strengthen our understanding of data acquired from mixture experiments. Rajapakse et al. (2002) reported corroborating findings as they used the human breast cancer cell proliferation assay to assess a mixture of four low potency estrogenic chemicals at low effect concentrations. The excellent agreement between prediction response model and experimental observations, implied that every component in the mixture contributed to the total mixture effect, even when present at concentrations below zero effect levels (Rajapakse et al., 2002, 2001; Silva et al., 2002). In addition, a mixture of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), PCB-126 and 2,3,4,7,8-pentachlorodibenzofuran (PeCDF), which are aryl hydrocarbon receptor (AhR) agonists, were used in a 2-year rodent cancer bioassay, acted in agreement with the additive effect model (Khetan, 2014; Walker et al., 2005). In the current work, using an additive prediction model, deviations between measured and predicted results were noted (Fig. 4).

For a better insight and more comprehensive overview of mixture effects, models such as *dose additivity* and the *isobole method* could be considered (Bosgra et al., 2009; Groten et al., 2001). Both these models require exposures with more diverse dose combinations of the different mixtures, for instance the high concentration of C mixture combined with low concentration of B mixture. However, the amount of laboratory experiments needed to be carried out in order to provide the comprehensive data necessary for such models would be expensive and not feasible within normal project budgets (Bosgra et al., 2009; Groten, 1996).

Some EDCs are suggested to have effects on gene expression in the steroidogenesis pathway, however these effects are not always translated into effects on enzyme activity (Gracia et al., 2006; Maglich et al., 2014). This means that the H295R *in vitro* assay used here would not detect compounds that modulate only gene expression of steroidogenesis enzymes, if they do not affect hormone production levels. There is a high probability that a compound can target several points in the steroidogenesis pathway simultaneously. In order to get a clearer and deeper overview of mechanisms by which compounds exert their endocrine disrupting effects, measuring all or most of the metabolites in the steroidogenesis pathway would be the best approach (Nielsen et al., 2012). In addition, if we take into account both gene expression of key enzymes and hormone production levels of H295R cells as endpoints to predict endocrine disrupting potential of xenobiotic compounds, a more accurate prediction could be established (Maglich et al., 2014).

Translating these *in vitro* findings into hazard predictions for human exposure should be performed with caution, but some risk potentials emerge. A small, but significant down-regulation of pregnenolone production was observed in the low and environmentally relevant exposure of the F (perfluorinated) mixture.

Pregnenolone is a known neuro-steroid that is involved in cognitive and memory abilities in humans and animals, hence decreased pregnenolone levels may be linked to impaired memory performance and Alzheimer's disease (Cheney et al., 1995; Vallée et al., 2001). Aldosterone is an important hormone associated with increasing blood pressure and elevated aldosterone levels could lead to hypertension (Basavanagowdappa et al., 2016). An up-regulation in aldosterone production was observed with exposures to the very high concentration of both the F mixture and C + F mixture. The exposure to the B + F mixture at VH₁₀ concentrations caused an overall down-regulation of steroidogenesis. Implications of such changes could involve induction of cancer, male infertility, reproductive disorders in females, metabolic and cardiovascular complications (Fernandez and Olea, 2012; Thayer et al., 2012; Vuorinen et al., 2015). Hence, cognitive defects as well as metabolic and endocrine disruption may be possible adverse outcomes of human exposures to these environmentally relevant mixtures.

5. Conclusion

The highest concentration of the combined brominated and perfluorinated (B + F) mixture produced a strong steroidogenesis disrupting effect with H295R cells. Also, the very high concentration of the C + F combination produced apparent non-additive effects, which were not predicted from the single mixtures. Otherwise, smaller and mostly non-significant changes were observed, with the perfluorinated mixture and combinations of this with other mixtures giving the most significant responses. Interestingly, the low concentration of the perfluorinated mixture alone produced a significant, albeit small, down-regulation of pregnenolone, and the total mixture a similar effect on 17-hydroxypregnenolone. Stimulating the cells with forskolin produced a marked shift in effects towards mineralocorticoids being down-regulated compared to unstimulated cells in the B + F response. Also, an upregulation of estrone in unstimulated cells disappeared after forskolin stimulation. The difference in results observed in stimulated and unstimulated conditions indicates the need to use both conditions in future toxicological screening experiments. The use of statistical models is of great importance to better interpret data from mixed exposure regimes. The potential for adverse outcomes of human exposures is indicated, pointing to a need for further investigation into these mixtures, as several of the single compounds present in the mixtures have been reported previously to be endocrine disrupting chemicals, but at higher concentrations than used in the present study.

Declarations of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemosphere.2018.11.057.

References

Ahmed, K.E.M., Frøysa, H.G., Karlsen, O.A., Sagen, J.V., Mellgren, G., Verhaegen, S., Ropstad, E., Goksøyr, A., Kellmann, R., 2018. LC-MS/MS based profiling and dynamic modelling of the steroidogenesis pathway in adrenocarcinoma H295R cells. Toxicol. Vitro 52, 332–341. https://doi.org/10.1016/j.tiv.2018.07.002.

Al-Dhaheri, M.H., Rowan, B.G., 2007. Protein kinase a exhibits selective modulation of estradiol-dependent transcription in breast cancer cells that is associated with decreased ligand binding, altered estrogen receptor α promoter interaction, and changes in receptor phosphorylation. Mol. Endocrinol. 21, 439–456. https://doi.org/10.1210/me.2006-0059.

Alaee, M., Arias, P., Sjödin, A., Bergman, Å., 2003. An overview of commercially used brominated flame retardants, their applications, their use patterns in different countries/regions and possible modes of release. Environ. Int. 29, 683–689. https://doi.org/10.1016/S0160-4120(03)00121-1.

Barón, E., Giménez, J., Verborgh, P., Gauffier, P., De Stephanis, R., Eljarrat, E., Barceló, D., 2015. Bioaccumulation and biomagnification of classical flame retardants, related halogenated natural compounds and alternative flame retardants in three delphinids from Southern European waters. Environ. Pollut. 203, 107–115. https://doi.org/10.1016/j.envpol.2015.03.041.

Bars, R., Fegert, I., Gross, M., Lewis, D., Weltje, L., Weyers, A., Wheeler, J.R., Galay-Burgos, M., 2012. Risk assessment of endocrine active chemicals: identifying chemicals of regulatory concern. Regul. Toxicol. Pharmacol. 64, 143–154. https://doi.org/10.1016/j.yrtph.2012.06.013.

Basavanagowdappa, H., Basavanagowdappa, N., Ys, R., Kulkarni, P., Devegowda, D., 2016. Magnitude of Resistant Hypertension and Impact of Aldosterone to Renin Ratio in Resistant Hypertension, vol. 7, pp. 133–138.

Benninghoff, A.D., Bisson, W.H., Koch, D.C., Ehresman, D.J., Kolluri, S.K., Williams, D.E., 2011. Estrogen-like activity of perfluoroalkyl acids in vivo and interaction with human and rainbow trout estrogen receptors in vitro. Toxicol. Sci. 120, 42–58. https://doi.org/10.1093/toxsci/kfq379.

Berg, V., Kraugerud, M., Nourizadeh-Lillabadi, R., Olsvik, P.A., Skåre, J.U., Alestrøm, P., Ropstad, E., Zimmer, K.E., Lyche, J.L., 2016. Endocrine effects of real-life mixtures of persistent organic pollutants (POP) in experimental models and wild fish. J. Toxicol. Environ. Health Part A Curr. Issues 79, 538–548. https://doi.org/10.1080/15287394.2016.1171980.

Bergman, Å., Heindel, J.J., Jobling, S., Kidd, K.A., Zoeller, R.T., 2013a. State of the science of endocrine disrupting chemicals - 2012. World Health Organ. 1–289. https://doi.org/10.1016/j.toxlet.2012.03.020.

Bergman, Å., Heindel, J.J., Kasten, T., Kidd, K.A., Jobling, S., Neira, M., Zoeller, R.T., Becher, G., Bjerregaard, P., Bornman, R., Brandt, I., Kortenkamp, A., Muir, D., Drisse, M.N.B., Ochieng, R., Skakkebaek, N.E., Byléhn, A.S., Iguchi, T., Toppari, J., Woodruff, T.J., 2013b. The impact of endocrine disruption: a consensus statement on the state of the science. Environ. Health Perspect. 121 https://doi.org/10.1289/ehp.1205448.

Berntsen, H.F., Berg, V., Thomsen, C., Ropstad, E., Zimmer, K.E., 2017. The design of an environmentally relevant mixture of persistent organic pollutants for use in in vivo and in vitro studies. J. Toxicol. Environ. Health Part A Curr. Issues 80, 1002–1016. https://doi.org/10.1080/15287394.2017.1354439.

Bird, I.M., Pasquarette, M.M., Rainey, W.E., Mason, J.I., 1996. Differential control of 17 alpha-hydroxylase and 3 beta-hydroxysteroid dehydrogenase expression in human adrenocortical H295R cells. J. Clin. Endocrinol. Metab. 81, 2171–2178. https://doi.org/10.1210/jcem.81.6.8964847.

Bosgra, S., Van Eijkeren, J.C.H., Slob, W., 2009. Dose addition and the isobole method as approaches for predicting the cumulative effect of non-interacting chemicals: a critical evaluation. Crit. Rev. Toxicol. 39, 418–426. https://doi.org/10.1080/10408440902787592.

Castellanos, C.G., Sørvik, I.B., Tanum, M.B., Verhaegen, S., Brandt, I., Ropstad, E., 2013. Differential effects of the persistent DDT metabolite methylsulfonyl-DDE in nonstimulated and LH-stimulated neonatal porcine Leydig cells. Toxicol. Appl. Pharmacol. 267, 247–255. https://doi.org/10.1016/j.taap.2012.12.022.

Cheney, D.L., Uzunov, D., Guidotti, A., 1995. Pregnenolone sulfate antagonizes dizocilpine amnesia: role for allopregnanolone. Neuroreport Int. J. Rapid Commun. Res. Neurosci. https://doi.org/10.1097/00001756-199508000-00025. Lippincott Williams & Wilkins, US.

Craig, Z.R., Wang, W., Flaws, J.A., 2011. Endocrine-disrupting chemicals in ovarian function: effects on steroidogenesis, metabolism and nuclear receptor signaling. Reproduction 142, 633–646. https://doi.org/10.1530/REP-11-0136.

Crellin, N.K., Kang, H.G., Swan, C.L., Chedrese, P.J., 2001. Inhibition of basal and stimulated progesterone synthesis by dichlorodiphenyldichloroethylene and methoxychlor in a stable pig granulosa cell line. Reproduction 121, 485–492. https://doi.org/10.1530/reprod/121.3.485.

Crellin, N.K., Rodway, M.R., Swan, C.L., Gillio-Meina, C., Chedrese, P.J., 1999. Dichlorodiphenyldichloroethylene potentiates the effect of protein kinase A pathway activators on progesterone synthesis in cultured porcine granulosa cells. Biol. Reprod. 61, 1099–1103. https://doi.org/10.1095/biolreprod61.4.1099.

Darnerud, P.O., 2008. Brominated flame retardants as possible endocrine disrupters. Int. J. Androl. 31, 152–160. https://doi.org/10.1111/j.1365-2605.2008.00869.x.

de Wit, C.A., Herzke, D., Vorkamp, K., 2010. Brominated flame retardants in the Arctic environment - trends and new candidates. Sci. Total Environ. 408, 2885–2918. https://doi.org/10.1016/j.scitotenv.2009.08.037.

Denner, K., Rainey, W.E., Pezzi, V., Bird, I.M., Bernhardt, R., Mathis, J.M., 1996. Differential regulation of 11 beta-hydroxylase and aldosterone synthase in human adrenocortical H295R cells. Mol. Cell. Endocrinol. 121, 87–91 doi:

0303720796038531 [pii].

Diamanti-Kandarakis, E., Bourguignon, J.P., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., Zoeller, R.T., Gore, A.C., 2009. Endocrine-disrupting chemicals: an Endocrine Society scientific statement. Endocr. Rev. 30, 293—342. https://doi.org/10.1210/er.2009-0002.

Du, G., Hu, J., Huang, H., Qin, Y., Han, X., Wu, D., Song, L., Xia, Y., Wang, X., 2013. Perfluorooctane sulfonate (PFOS) affects hormone receptor activity, steroidogenesis, and expression of endocrine-related genes in vitro and in vivo. Environ. Toxicol. Chem. 32, 353—360. https://doi.org/10.1002/etc.2034.

Fernandez, M.F., Olea, N., 2012. Developmental exposure to endocrine disruptors and male urogenital tract malformations. In: Endocrine Disruptors and Puberty. Humana Press, Totowa, NJ, pp. 225—239. https://doi.org/10.1007/978-1-60761-561-3_8.

Folmar, L.C., Denslow, N.D., Rao, V., Chow, M., Crain, D.A., Enblom, J., Marcino, J., Guillette Jr., L.J., 1996. Vitellogenin induction and reduced serum testosterone concentrations in feral male carp (Cyprinus carpio) captured near a major metropolitan sewage treatment plant. Environ. Health Perspect. 104, 1096—1101.

Fuhrman, V.F., Tal, A, Arnon, S., 2015. Why endocrine disrupting chemicals (EDCs) challenge traditional risk assessment and how to respond. J. Hazard Mater. 286, 589—611. https://doi.org/10.1016/j.jhazmat.2014.12.012.

Gazdar, A.F., Oie, H.K., Shackleton, C.H., Chen, T.R., Triche, T.J., Myers, C.E., Chrousos, G.P., Brennan, M.F., Stein, C.A., La Rocca, R.V., 1990. Establishment and characterization of a human adrenocortical carcinoma cell line that expresses multiple pathways of steroid biosynthesis. Cancer Res. 50, 5488—5496.

Gelbke, H.P., Kayser, M., Poole, A., 2004. OECD test strategies and methods for endocrine disruptors. Toxicology 205, 17—25. https://doi.org/10.1016/j.tox.2004.06.034.

Gorrochategui, E., Pérez-Albaladejo, E., Casas, J., Lacorte, S., Porte, C., 2014. Perfluorinated chemicals: differential toxicity, inhibition of aromatase activity and alteration of cellular lipids in human placental cells. Toxicol. Appl. Pharmacol. 277, 124—130. https://doi.org/10.1016/j.taap.2014.03.012.

Gracia, T., Hilscherova, K., Jones, P.D., Newsted, J.L., Zhang, X., Hecker, M., Higley, E.B., Sanderson, J.T., Yu, R.M.K., Wu, R.S.S., Giesy, J.P., 2006. The H295R system for evaluation of endocrine-disrupting effects. Ecotoxicol. Environ. Saf. 65, 293—305. https://doi.org/10.1016/j.ecoenv.2006.06.012.

Gregoraszczuk, E.L., Ptak, A., Karniewska, M., Ropstad, E., 2008. Action of defined mixtures of PCBs, p,p???-DDT and its metabolite p,p???-DDE, on co-culture of porcine theca and granulosa cells: steroid secretion, cell proliferation and apoptosis. Reprod. Toxicol. 26, 170—174. https://doi.org/10.1016/j.reprotox.2008.07.003.

Groten, J.P., 2000. Mixtures and interactions. Food Chem. Toxicol. 38, 65—71. https://doi.org/10.1016/S0278-6915(99)00135-0.

Groten, J.P., 1996. Use of factorial designs in combination toxicity studies. Food Chem. Toxicol. 34, 1083—1089.

Groten, J.P., Feron, V.J., Suhnel, J., 2001. Toxicology of simple and complex mixtures. Trends Pharmacol. Sci. 22, 316—322. https://doi.org/10.1016/s0165-6147(00)01720-x.

Guillette, L.J., Pickford, D.B., Crain, D.A., Rooney, A.A., Percival, H.F., 1996. Reduction in penis size and plasma testosterone concentrations in juvenile alligators living in a contaminated environment. Gen. Comp. Endocrinol. 101, 32—42. https://doi.org/10.1006/gcen.1996.0005.

Hamers, T., Kamstra, J.H., Sonneveld, E., Murk, A.J., Kester, M.H.A., Andersson, P.L., Legler, J., Brouwer, A., 2006. In vitro profiling of the endocrine-disrupting potency of brominated flame retardants. Toxicol. Sci. 92, 157—173. https://doi.org/10.1093/toxsci/kfj187.

Harju, M., Hamers, T., Kamstra, J., Edwin, S., Boon, J., Tysklind, M., Andersson, P., 2007. Endocrine effects and metabolic stability involving 26 selected brominated flame retardants. Environ. Toxicol. Chem. 26, 816—826.

He, W., He, P., Wang, A., Xia, T., Xu, B., Chen, X., 2008. Effects of PBDE-47 on cytotoxicity and genotoxicity in human neuroblastoma cells in vitro. Mutat. Res. Genet. Toxicol. Environ. Mutagen 649, 62—70. https://doi.org/10.1016/j.mrgentox.2007.08.001.

Hecker, M., Hollert, H., 2011. Endocrine disruptor screening: regulatory perspectives and needs. Environ. Sci. Eur. 23, 15. https://doi.org/10.1186/2190-4715-23-15.

Hecker, M., Newsted, J.L., Murphy, M.B., Higley, E.B., Jones, P.D., Wu, R., Giesy, J.P., 2006. Human adrenocarcinoma (H295R) cells for rapid in vitro determination of effects on steroidogenesis: hormone production. Toxicol. Appl. Pharmacol. 217, 114—124. https://doi.org/10.1016/j.taap.2006.07.007.

Hilscherova, K., Jones, P.D., Gracia, T., Newsted, J.L., Zhang, X., Sanderson, J.T., Yu, R.M.K., Wu, R.S.S., Giesy, J.P., 2004. Assessment of the effects of chemicals on the expression of ten steroidogenic genes in the H295R cell line using real-time PCR. Toxicol. Sci. 81, 78—89. https://doi.org/10.1093/toxsci/kfh191.

Hines, E.P., White, S.S., Stanko, J.P., Gibbs-Flournoy, E.A., Lau, C., Fenton, S.E., 2009. Phenotypic dichotomy following developmental exposure to perfluorooctanoic acid (PFOA) in female CD-1 mice: low doses induce elevated serum leptin and insulin, and overweight in mid-life. Mol. Cell. Endocrinol. 304, 97—105. https://doi.org/10.1016/j.mce.2009.02.021.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65—70. https://doi.org/10.2307/4615733.

Hotchkiss, A.K., Rider, C.V., Blystone, C.R., Wilson, V.S., Hartig, P.C., Ankley, G.T., Foster, P.M., Gray, C.L., Gray, L.E., 2008. Fifteen years after "wingspread" - environmental endocrine disrupters and human and wildlife health: where we are today and where we need to go. Toxicol. Sci. 105, 235—259. https://doi.org/10.1093/toxsci/kfn030.

Hu, X.Z., Xu, Y., Hu, D.C., Hui, Y., Yang, F.X., 2007. Apoptosis induction on human hepatoma cells Hep G2 of decabrominated diphenyl ether (PBDE-209). Toxicol. Lett. 171, 19—28. https://doi.org/10.1016/j.toxlet.2007.04.002.

Kalayou, S., Granum, C., Berntsen, H.F., Groseth, P.K., Verhaegen, S., Connolly, L., Brandt, I., de Souza, G.A., Ropstad, E., 2016. Label-free based quantitative proteomics analysis of primary neonatal porcine Leydig cells exposed to the persistent contaminant 3-methylsulfonyl-DDE. J. Proteom. 137, 68—82. https://doi.org/10.1016/j.jprot.2015.12.007.

Kang, J.S., Choi, J.S., Park, J.W., 2016. Transcriptional changes in steroidogenesis by perfluoroalkyl acids (PFOA and PFOS) regulate the synthesis of sex hormones in H295R cells. Chemosphere 155, 436—443. https://doi.org/10.1016/j.chemosphere.2016.04.070.

Karmaus, A.L., Toole, C.M., Filer, D.L, Lewis, K.C., Martin, M.T., 2016. High-Throughput screening of chemical effects on steroidogenesis using H295R human adrenocortical carcinoma cells. Toxicol. Sci. 150, 323—332. https://doi.org/10.1093/toxsci/kfw002.

Khetan, S.K., 2014. Endocrine Disruptors in the Environment, Endocrine Disruptors in the Environment. https://doi.org/10.1002/9781118891094.

Kitamura, S., Suzuki, T., Sanoh, S., Kohta, R., Jinno, N., Sugihara, K., Yoshihara, S., Fujimoto, N., Watanabe, H., Ohta, S., 2005. Comparative study of the endocrine-disrupting activity of bisphenol A and 19 related compounds. Toxicol. Sci. 84, 249—259. https://doi.org/10.1093/toxsci/kfi074.

Kortenkamp, A., Faust, M., Scholze, M., Backhaus, T., 2007. Low-level exposure to multiple chemicals: reason for human health concerns? Environ. Health Perspect. 115, 106—114. https://doi.org/10.1289/ehp.9358.

Kraugerud, M., Zimmer, K.E., Dahl, E., Berg, V., Olsaker, I., Farstad, W., Ropstad, E., Verhaegen, S., 2010. Three structurally different polychlorinated biphenyl congeners (Pcb 118, 153, and 126) affect hormone production and gene expression in the human H295R in vitro model. J. Toxicol. Environ. Health Part A Curr. Issues 73, 1122—1132. https://doi.org/10.1080/15287394.2010.484338.

Kraugerud, M., Zimmer, K.E., Ropstad, E., Verhaegen, S., 2011. Perfluorinated compounds differentially affect steroidogenesis and viability in the human adrenocortical carcinoma (H295R) in vitro cell assay. Toxicol. Lett. 205, 62—68. https://doi.org/10.1016/j.toxlet.2011.05.230.

Krogh, K. Von, Harjen, H., Almås, C., Zimmer, K.E., Dahl, E., Olsaker, I., Taubøll, E., Ropstad, E., Verhaegen, S., 2010. The effect of valproate and levetiracetam on steroidogenesis in forskolin-stimulated H295R cells. Epilepsia 51, 2280—2288. https://doi.org/10.1111/j.1528-1167.2010.02702.x.

Landrigan, P.J., Fuller, R., Acosta, N.J.R., Adeyi, O., Arnold, R., Basu, N., Baldé, A.B., Bertollini, R., Bose-O'Reilly, S., Boufford, J.I., Breysse, P.N., Chiles, T., Mahidol, C., Coll-Seck, A.M., Cropper, M.L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., Hanrahan, D., Hunter, D., Khare, M., Krupnick, A., Lanphear, B., Lohani, B., Martin, K., Mathiasen, K.V., McTeer, M.A., Murray, C.J.L., Ndahimananjara, J.D., Perera, F., Potočnik, J., Preker, A.S., Ramesh, J., Rockström, J., Salinas, C., Samson, L.D., Sandilya, K., Sly, P.D., Smith, K.R., Steiner, A., Stewart, R.B., Suk, W.A., van Schayck, O.C.P., Yadama, G.N., Yumkella, K., Zhong, M., 2017. The Lancet Commission on pollution and health. Lancet https://doi.org/10.1016/S0140-6736(17)32345-0.

Lervik, S., von Krogh, K., Karlsson, C., Olsaker, I., Andresen, T., Dahl, E., Verhaegen, S., Ropstad, E., 2011. Steroidogenesis in primary cultures of neonatal porcine Leydig cells from Duroc and Norwegian Landrace breeds. Theriogenology 76, 1058—1069. https://doi.org/10.1016/j.theriogenology.2011.05.011.

Li, Q.Q., Loganath, A., Seng Chong, Y., Tan, J., Philip Obbard, J., 2006. Persistent organic pollutants and adverse health effects in humans. J. Toxicol. Environ. Health Part A 69, 1987—2005. https://doi.org/10.1080/15287390600751447.

Lilienthal, H., Hack, A., Roth-Härer, A., Grande, S.W., Talsness, C.E., 2006. Effects of developmental exposure to 2,2′,4,4′, 5-pentabromodiphenyl ether (PBDE-99) on sex steroids, sexual development, and sexually dimorphic behavior in rats. Environ. Health Perspect. 114, 194—201. https://doi.org/10.1289/ehp.8391.

Maglich, J.M., Kuhn, M., Chapin, R.E., Pletcher, M.T., 2014. More than just hormones: H295R cells as predictors of reproductive toxicity. Reprod. Toxicol. 45, 77—86. https://doi.org/10.1016/j.reprotox.2013.12.009.

Main, K.M., Kiviranta, H., Virtanen, H.E., Sundqvist, E., Tuomisto, J.T., Tuomisto, J., Vartiainen, T., Skakkebæk, N.E., Topparl, J., 2007. Flame retardants in placenta and breast milk and cryptorchildism in newborn boys. Environ. Health Perspect. 115, 1519—1526. https://doi.org/10.1289/ehp.9924.

McCarty, L.S., Borgert, C.J., 2006. Review of the toxicity of chemical mixtures: theory, policy, and regulatory practice. Regul. Toxicol. Pharmacol. 45, 119—143. https://doi.org/10.1016/j.yrtph.2006.03.004.

Monosson, E., 2005. Chemical mixtures: considering the evolution of toxicology and chemical assessment. Environ. Health Perspect. 113, 383—390. https://doi.org/10.1289/ehp.6987.

Mountjoy, K.G., Bird, I.M., Rainey, W.E., Cone, R.D., 1994. ACTH induces up-regulation of ACTH receptor mRNA in mouse and human adrenocortical cell lines. Mol. Cell. Endocrinol. 99, R17—R20. https://doi.org/10.1016/0303-7207(94)90160-0.

Nielsen, F.K., Hansen, C.H., Fey, J.A., Hansen, M., Jacobsen, N.W., Halling-Sørensen, B., Björklund, E., Styrishave, B., 2012. H295R cells as a model for steroidogenic disruption: a broader perspective using simultaneous chemical analysis of 7 key steroid hormones. Toxicol. Vitro 26, 343—350. https://doi.org/10.1016/j.tiv.2011.12.008.

Nimrod, A.C., Benson, W.H., 1996. Environmental estrogenic effects of alkylphenol ethoxylates. Crit. Rev. Toxicol. 26, 335—364. https://doi.org/10.3109/10408449609012527.

OECD, 2007. Uterotrophic Bioassay in Rodents : October.

Olsen, G.W., Gilliland, F.D., Burlew, M.M., Burris, J.M., Mandel, J.S., Mandel, J.H., 1998. An epidemiologic investigation of reproductive hormones in men with occupational exposure to perfluorooctanoic acid. J. Occup. Environ. Med. 40, 614–622. https://doi.org/10.1097/00043764-199807000-00006.

Owens, J.W., Gray, L.E., Zeiger, E., Walker, M., Yamasaki, K., Ashby, J., Jacob, E., 2007. The OECD program to validate the rat Hershberger bioassay to screen compounds for in vivo androgen and antiandrogen responses: phase 2 dose-response studies. Environ. Health Perspect. 115, 671–678. https://doi.org/10.1289/ehp.9666.

Rainey, W.E., Bird, I.M., Mason, J.I., 1994. The NCI-H295 cell line: a pluripotent model for human adrenocortical studies. Mol. Cell. Endocrinol. 100, 45–50. https://doi.org/10.1016/0303-7207(94)90277-1.

Rainey, W.E., Bird, I.M., Sawetawan, C., Hanley, N.A., Mc carthy, J.L., Mc gee, E.A., Wester, R., Mason, J.I., 1993. Regulation of human adrenal carcinoma cell (NCI-H295) production of C19 steroids. J. Clin. Endocrinol. Metab. 77, 731–737. https://doi.org/10.1210/jcem.77.3.8396576.

Rajapakse, N., Ong, D., Kortenkamp, A., 2001. Defining the impact of weakly estrogenic chemicals on the action of steroidal estrogens. Toxicol. Sci. 60, 296–304. https://doi.org/10.1093/toxsci/60.2.296.

Rajapakse, N., Silva, E., Kortenkamp, A., 2002. Combining xenoestrogens at levels below individual no-observed-effect concentrations dramatically enhances steroid hormone action. Environ. Health Perspect. 110, 917–921. https://doi.org/10.1289/ehp.02110917.

Rauh, M., 2010. Steroid measurement with LC-MS/MS. Application examples in pediatrics. J. Steroid Biochem. Mol. Biol. https://doi.org/10.1016/j.jsbmb.2009.12.007.

Rhind, S., 2008. Endocrine disruptors and other food-contaminating environmental pollutants as risk factors in animal reproduction. Reprod. Domest. Anim. 43, 15–22. https://doi.org/10.1111/j.1439-0531.2008.01138.x.

Rijk, J.C.W., Peijnenburg, A.A.C.M., Blokland, M.H., Lommen, A., Hoogenboom, R.L.A.P., Bovee, T.F.H., 2012. Screening for Modulatory Effects on Steroidogenesis Using the human H295R adrenocortical cell line: a metabolomics approach. Chem. Res. Toxicol. 25, 1720–1731.

Rizzati, V., Briand, O., Guillou, H., Gamet-Payrastre, L., 2016. Effects of pesticide mixtures in human and animal models: an update of the recent literature. Chem. Biol. Interact. 254, 231–246. https://doi.org/10.1016/j.cbi.2016.06.003.

Ropstad, E., Oskam, I., Lyche, J., Larsen, H., Lie, E., Haave, M., Dahl, E., Wiger, R., Utne Skaare, J., 2006. Endocrine disruption induced by organochlorines (OCs): field studies and experimental models. J. Toxicol. Environ. Health Part A 69, 53–76. https://doi.org/10.1080/15287390500259145.

Safe, S.H., 1994. Polychlorinated biphenyls (PCBs): environmental impact, biochemical and toxic responses, and implications for risk assessment. Crit. Rev. Toxicol. 24, 87–149. https://doi.org/10.3109/10408449409049308.

Schug, T.T., Johnson, A.F., Birnbaum, L.S., Colborn, T., Guillette, L.J., Crews, D.P., Collins, T., Soto, A.M., vom Saal, F.S., McLachlan, J.A., Sonnenschein, C., Heindel, J.J., 2016. Minireview: endocrine disruptors: past lessons and future directions. Mol. Endocrinol. 30, 833–847. https://doi.org/10.1210/me.2016-1096.

Shao, J., White, C.C., Dabrowski, M.J., Kavanagh, T.J., Eckert, M.L., Gallagher, E.P., 2008. The role of mitochondrial and oxidative injury in BDE 47 toxicity to human fetal liver hematopoietic stem cells. Toxicol. Sci. 101, 81–90. https://doi.org/10.1093/toxsci/kfm256.

Silva, E., Rajapakse, N., Kortenkamp, A., 2002. Something from "nothing" - eight weak estrogenic chemicals combined at concentrations below NOECs produce significant mixture effects. Environ. Sci. Technol. 36, 1751–1756. https://doi.org/10.1021/es0101227.

Spurgeon, D.J., Jones, O.A.H., Dorne, J.-L.C.M., Svendsen, C., Swain, S., Stürzenbaum, S.R., 2010. Systems toxicology approaches for understanding the joint effects of environmental chemical mixtures. Sci. Total Environ. 408, 3725–3734. https://doi.org/10.1016/j.scitotenv.2010.02.038.

Stahl, T., Mattern, D., Brunn, H., 2011. Toxicology of perfluorinated compounds. Environ. Sci. Eur. 23, 1–52. https://doi.org/10.1186/2190-4715-23-38.

Talsness, C.E., Kuriyama, S.N., Wichert Grande, S., Andrade, A., Sterner-Kock, A., Schnitker, P., Grote, K., Chahoud, I., 2006. Low dose effects on the rat female reproductive system following exposure to a single administration of PBDE-47. Organohalogen Compd. 68, 407–409.

Thayer, K.A., Heindel, J.J., Bucher, J.R., Gallo, M.A., 2012. Role of environmental chemicals in diabetes and obesity: a national toxicology program workshop review. Environ. Health Perspect. 120, 779–789. https://doi.org/10.1289/ehp.1104597.

Tiemann, U., 2008. In vivo and in vitro effects of the organochlorine pesticides DDT,

TCPM, methoxychlor, and lindane on the female reproductive tract of mammals: a review. Reprod. Toxicol. 25, 316–326. https://doi.org/10.1016/j.reprotox.2008.03.002.

Tipping, E., 1994. WHAM - a chemical equilibrium model and computer code for waters sediments and soils incorporating a discrete site electrostatic model of ion binding by humic substances. Comput. Geosci. 20, 973–1023.

Tremoen, N.H., Fowler, P.A., Ropstad, E., Verhaegen, S., Krogenæs, A., 2014. Exposure to the three structurally different PCB congeners (PCB 118, 153, and 126) results in decreased protein expression and altered steroidogenesis in the human adrenocortical carcinoma cell line H295R. J. Toxicol. Environ. Health Part A Curr. Issues 77, 516–534. https://doi.org/10.1080/15287394.2014.886985.

UNEP, 2008. Stockholm convention on persistent organic pollutants. Unit. Nation. Environ. Progr. https://doi.org/10.1111/1467-9388.00331.

Vallée, M., Mayo, W., Le Moal, M., 2001. Role of pregnenolone, dehydroepiandrosterone and their sulfate esters on learning and memory in cognitive aging. Brain Res. Rev. 37, 301–312. https://doi.org/10.1016/S0165-0173(01)00135-7.

van den Dungen, M.W., Rijk, J.C.W., Kampman, E., Steegenga, W.T., Murk, A.J., 2015. Steroid hormone related effects of marine persistent organic pollutants in human H295R adrenocortical cells. Toxicol. Vitro 29, 769–778. https://doi.org/10.1016/j.tiv.2015.03.002.

Vuorinen, A., Odermatt, A., Schuster, D., 2015. In silico methods in the discovery of endocrine disrupting chemicals. J. Steroid Biochem. Mol. Biol. 153, 93–101. https://doi.org/10.1016/j.jsbmb.2015.08.015.

Walker, N.J., Crockett, P.W., Nyska, A., Brix, A.E., Jokinen, M.P., Sells, D.M., Hailey, J.R., Easterling, M., Haseman, J.K., Yin, M., Wyde, M.E., Bucher, J.R., Portier, C.J., 2005. Dose-additive carcinogenicity of a defined mixture of dioxin-like compounds. Environ. Health Perspect. 113, 43–48. https://doi.org/10.1289/ehp.7351.

Wan, H.T., Zhao, Y.G., Wong, M.H., Lee, K.F., Yeung, W.S.B., Giesy, J.P., Wong, C.K.C., 2011. Testicular signaling is the potential target of perfluorooctanesulfonate-mediated subfertility in male Mice 1. Biol. Reprod. 84, 1016–1023. https://doi.org/10.1095/biolreprod.110.089219.

Wang, M., Chen, J., Lin, K., Chen, Y., Hu, W., Tanguay, R.L., Huang, C., Dong, Q., 2011. Chronic zebrafish PFOS exposure alters sex ratio and maternal related effects in F1 offspring. Environ. Toxicol. Chem. 30, 2073–2080. https://doi.org/10.1002/etc.594.

Wang, S., Rijk, J.C.W., Besselink, H.T., Houtman, R., Peijnenburg, A.A.C.M., Brouwer, A., Rietjens, I.M.C.M., Bovee, T.F.H., 2014. Extending an in vitro panel for estrogenicity testing: the added value of bioassays for measuring antiandrogenic activities and effects on steroidogenesis. Toxicol. Sci. 141, 78–89. https://doi.org/10.1093/toxsci/kfu103.

Weisser, J.J., Hansen, C.H., Poulsen, R., Larsen, L.W., Cornett, C., Styrishave, B., 2016. Two simple cleanup methods combined with LC-MS/MS for quantification of steroid hormones in in vivo and in vitro assays. Anal. Bioanal. Chem. 408, 4883–4895. https://doi.org/10.1007/s00216-016-9575-z.

Wilson, J., Berntsen, H.F., Zimmer, K.E., Frizzell, C., Verhaegen, S., Ropstad, E., Connolly, L., 2016. Effects of defined mixtures of persistent organic pollutants (POPs) on multiple cellular responses in the human hepatocarcinoma cell line, HepG2, using high content analysis screening. Toxicol. Appl. Pharmacol. 294, 21–31. https://doi.org/10.1016/j.taap.2016.01.001.

Xu, Y., Yu, R.M.K., Zhang, X., Murphy, M.B., Giesy, J.P., Lam, M.H.W., Lam, P.K.S., Wu, R.S.S., Yu, H., 2006. Effects of PCBs and MeSO2−PCBs on adrenocortical steroidogenesis in H295R human adrenocortical carcinoma cells. Chemosphere 63, 772–784. https://doi.org/10.1016/j.chemosphere.2005.08.013.

Yang, C., Tan, Y.S., Harkema, J.R., Haslam, S.Z., 2009. Differential effects of peripubertal exposure to perfluorooctanoic acid on mammary gland development in C57Bl/6 and Balb/c mouse strains. Reprod. Toxicol. 27, 299–306. https://doi.org/10.1016/j.reprotox.2008.10.003.

Zhang, X., Yu, R.M.K., Jones, P.D., Lam, G.K.W., Newsted, J.L., Gracia, T., Hecker, M., Hilscherova, K., Sanderson, J.T., Wu, R.S.S., Giesy, J.P., 2005. Quantitative RT-PCR methods for evaluating toxicant-induced effects on steroidogenesis using the H295R cell line. Environ. Sci. Technol. 39, 2777–2785. https://doi.org/10.1021/es048679k.

Zhao, B., Chu, Y., Hardy, D.O., Li, X. kun, Ge, R.S., 2010. Inhibition of 3β- and 17β-hydroxysteroid dehydrogenase activities in rat Leydig cells by perfluorooctane acid. J. Steroid Biochem. Mol. Biol. 118, 13–17. https://doi.org/10.1016/j.jsbmb.2009.09.010.

Zimmer, K.E., Montaño, M., Olsaker, I., Dahl, E., Berg, V., Karlsson, C., Murk, A.J., Skaare, J.U., Ropstad, E., Verhaegen, S., 2011. In vitro steroidogenic effects of mixtures of persistent organic pollutants (POPs) extracted from burbot (Lota lota) caught in two Norwegian lakes. Sci. Total Environ. 409, 2040–2048. https://doi.org/10.1016/j.scitotenv.2011.01.055.

uib.no