Computational Systems Biology
Exam

Dr. Jürgen Pahle
Aleksandr Andreychenko, M.Sc.

31 July, 2012

Name

Matriculation Number

Do not open this exam booklet before we ask you to. **Do** read this page carefully.

Make sure you write your name and matriculation number onto this exam booklet.

This is a closed-book exam. Please leave bags and jackets at the sides of the room. Turn your mobile phones off and leave them with your bag. You may only take writing utensils, drinks and food as well as your student ID and passport or identity card to your seat.

Leaving the room without handing in your exam booklet is regarded as an attempt of deception. If you need to use the bathroom, please hand your exam booklet to the supervisor. Only one person at a time may leave for the bathroom.

Write your solutions on the (printed) right pages of the exam booklet. Solutions written in a different language than English or on the (blank) left pages **will not be graded**. Should you run out of space, ask the supervisor for an additional sheet of paper. You may use a pencil.

The duration of the exam is **90 minutes**. The total number of points in the exam is 90. The number of points of an exercise is thus a rough indication of how much time you should spend on the solution of that exercise. 45 points will be sufficient to pass the exam.

Please put your identity card or passport and your student ID on the table next to you.

Have fun!

1	2	3	4	5	6
22.5	20	17	9	6.5	15

Sum
90

Result

Exercise 1. Comprehension questions (in all multiple choice questions only one answer is the correct one) ⟨1.5 points each⟩ or ⟨22.5 points total⟩

a) Systems biology can be defined as

- ☐ the systematic study of organisms using a set of established guidelines
- ☐ the detailed study of the components of biochemical pathways, such as receptor proteins, using simulation methods, such as molecular dynamics
- ☑ the combined study of biological systems integrating experimental and computational methods and focussing on the interactions between components
- ☐ the thorough study of the dynamics of the glycolysis pathway

b) A model in systems biology is usually **not**

- ☐ simplified
- ☐ a mathematical representation of biological processes
- ☐ the central element of systems biology research
- ☑ valid under all conceivable environmental conditions

c) Ordinary differential equation (ODE) models of biochemical networks

- ☑ are based on continuously-valued variables for the concentrations of chemical species
- ☐ describe biochemical systems in terms of probabilities of reactive events
- ☐ are usually linear
- ☐ can usually be solved analytically using sophisticated software, such as COPASI

d) Which of the following statements is **not** correct?

- ☐ One set of parameters determines unambiguously the variables in the model
- ☐ One set of variables can potentially be caused by many parameter sets
- ☐ The values of parameters do not change in the course of a simulation
- ☑ The fluxes of reactions are usually parameters of the model

e) In the iterative modelling cycle the so-called forward modelling, among other things, comprises

- ☐ optimisation
- ☐ parameter fitting/estimation
- ☑ stochastic simulation
- ☐ text mining in databases of scientific articles

f) Stoichiometric coefficients indicate

- ☐ the speed of a reaction
- ☑ the proportions in which substrates are consumed and products are produced
- ☐ the change in steady state concentrations upon a small increase of the rate of a reaction
- ☐ whether a steady state is stable or not

g) What is **not** needed when setting up and simulating an ordinary differential equation (ODE) model?

- ☑ the sensitivities (flux control coefficients)
- ☐ the structure of the reaction system (stoichiometry)
- ☐ the species concentrations at time point 0 (inital state)
- ☐ a description of the velocities of reactions (kinetic functions/rate laws)

h) Let us assume the reaction $S \rightleftharpoons P$ is in equilibrium. By how much does the ratio of product concentration to substrate concentration change if you double the substrate concentration and then let the system evolve to an equilibrium again?

- ☐ by 2
- ☐ by 1/2
- ☐ by 1/3
- ☑ it does not change at all

i) If the system is in a steady state

- ☐ the system always returns to the steady state after very small perturbations
- ☐ there can be no other stable steady state with the same parameter set
- ☐ the system shows periodic changes in species' concentrations
- ☑ the rate of change of species' concentrations is zero

j) A system's steady state is asymptotically stable if

- ☑ all eigenvalues of the corresponding Jacobian matrix have strictly negative real parts
- ☐ all eigenvalues of the corresponding Jacobian matrix have strictly positive real parts
- ☐ there is at least one eigenvalue of the corresponding Jacobian matrix with value greater than 1.0
- ☐ its stoichiometric matrix does not have full rank

k) Which of the following can **not** be used to find steady states of a system

- ☐ Newton method
- ☑ Gauss elimination algorithm
- ☐ forward integration
- ☐ backward integration (in reverse time direction)

l) Conservation relations are

- ☐ linear combinations of parameters that stay constant during a simulation
- ☐ nonlinear functions of reaction fluxes that stay constant during a simulation
- ☑ linear combinations of species' concentrations that stay constant during a simulation
- ☐ nonlinear functions of species' concentrations that stay constant during a simulation

m) Which of the following is **not** saved in a file that contains a model in the Systems Biology Markup Language (SBML) format?

- ☐ initial conditions
- ☐ kinetic functions
- ☑ parameters for the simulation method, such as step size
- ☐ the chemical equations including all stoichiometric coefficients

n) Annotations in models, as defined in the Minimal Information Requested In the Annotation of Models (MIRIAM) standard, are important because they

- ☑ link (arbitrarily named) model entities to entities in the real world, such as specific proteins
- ☐ unify the naming of model entities across different models
- ☐ specify in a unique way how model entities should be displayed in a qualitative model diagram
- ☐ give ratings for the biological correctness of parts of the model

o) Name three different databases that are used in systems biology research.

E.g. Biomodels database, JWS, Sabio-RK, BRENDA, KEGG, Reactome, Bionumbers.
Not correct are, e.g., MIRIAM (it is a standard) or BioML and BioPax (they are languages).

Exercise 2. Models and simulation ⟨20 points total⟩

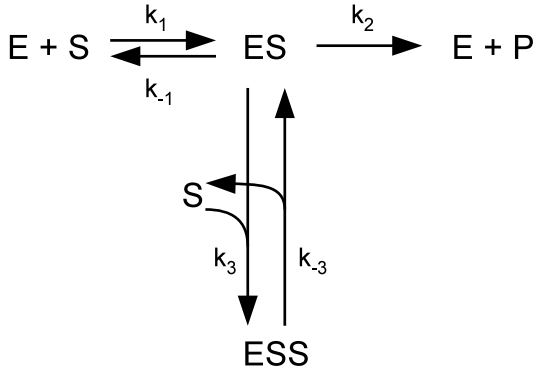
a) Write down the kinetic function for the following reversible reaction, assuming that it follows a mass action rate law: $2A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C$
(please use rectangular brackets, like in $[A]$, to denote concentrations of chemical species) ⟨3 points⟩

$v = k_1 \cdot [A]^2 \cdot [B] - k_2 \cdot [C]$

b) For the (mass action) reaction $S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} P$ write down the definition of the equilibrium constant K_{eq} one time using the rate constants k_1 and k_2, and one time using the species' concentrations $[S]$ and $[P]$. ⟨2 points⟩

$K_{eq} = \frac{k_1}{k_2}, \quad K_{eq} = \frac{[P]_{eq}}{[S]_{eq}}$

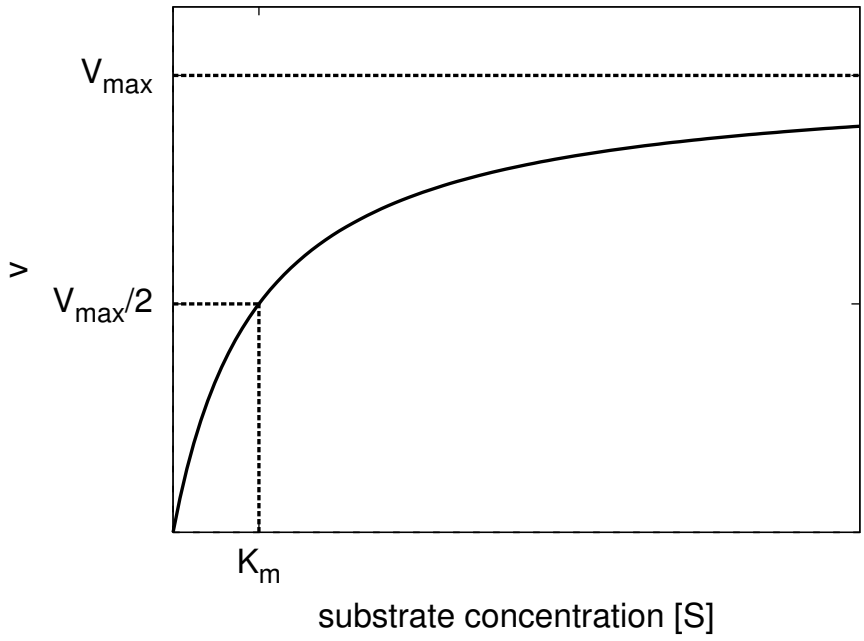
c) Set up the ordinary differential equation (ODE) system for the following mechanism of substrate inhibition (assume that all reactions have mass action rate laws/kinetic functions with the given rate constants. Please use rectangular brackets, like in $[ES]$, to denote concentrations of chemical species). ⟨10 points ⟩



$$\frac{d[S]}{dt} = -k_1 \cdot [E] \cdot [S] + k_{-1} \cdot [ES] - k_3 \cdot [ES] \cdot [S] + k_{-3} \cdot [ESS]$$
$$\frac{d[E]}{dt} = -k_1 \cdot [E] \cdot [S] + (k_{-1} + k_2) \cdot [ES]$$
$$\frac{d[ES]}{dt} = k_1 \cdot [E] \cdot [S] - (k_{-1} + k_2) \cdot [ES] - k_3 \cdot [ES] \cdot [S] + k_{-3} \cdot [ESS]$$
$$\frac{d[ESS]}{dt} = k_3 \cdot [ES] \cdot [S] - k_{-3} \cdot [ESS]$$
$$\frac{d[P]}{dt} = k_2 \cdot [ES]$$

d) Write down the formula for the Michaelis-Menten rate law. Sketch the relation between substrate concentration $[S]$ and rate of the reaction v in a diagram where v on the y-axis is plotted against $[S]$ on the x-axis. Mark the two parameters of this rate law, V_{max} and K_m in the diagram. Give brief textual interpretations (one sentence each) of these two parameters. ⟨5 points⟩

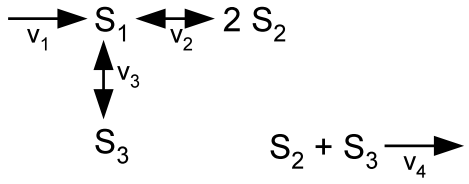
$v = \frac{V_{max} \cdot [S]}{K_m + [S]}$



V_{max} is the maximal velocity of the reaction, i.e. the velocity of the reaction converges to V_{max} when the concentration of the substrate, $[S]$, goes to infinity.
K_m is the substrate concentration at which the reaction reaches half of its maximal velocity V_{max}.

Exercise 3. Structural analysis ⟨17 points total⟩

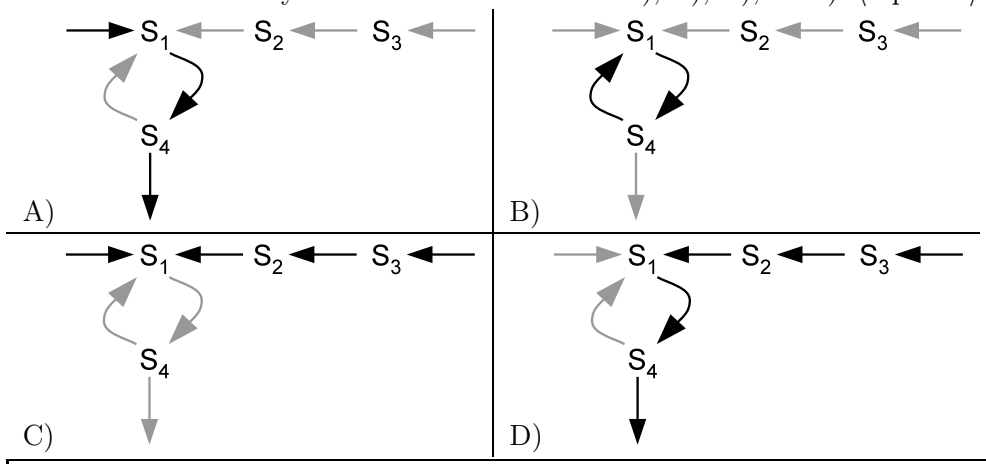
a) Given the following reaction system:



Write down the corresponding stoichiometric matrix including appropriate labels for the rows and columns. ⟨5 points⟩

	$R1$	$R2$	$R3$	$R4$
S_1	1	−1	−1	0
S_2	0	2	0	−1
S_3	0	0	1	−1

b) Which of the following sets of reactions (marked with black arrows) does **not** describe an elementary flux mode? Please circle A), B), C), or D). ⟨4 points⟩



Correct answer: C)

c) Given the following stoichiometric matrix:

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{pmatrix}$$

(i) What is the rank of this matrix? ⟨5 points⟩

rank = 3

(ii) How many dimensions does the space of all possible steady states have? (or, equivalently, how many basis vectors are there in the so-called nullspace of the stoichiometric matrix)? Explain in one sentence how this can be calculated (*e.g.* "The number of dimensions of the space of steady states is given by the _____ between the number of _____ and the _____"). ⟨1.5 points⟩

2
The number of dimensions of the space of steady states is given by the difference between the number of reactions and the rank of the stoichiometric matrix.

(iii) Are there any conservation relations present in this system? If yes, how many? Explain in one sentence how this can be calculated (*e.g.* "The number of conservation relations is given by the _____ between the number of _____ and the _____"). ⟨1.5 points⟩

Yes, 1
The number of conservation relations is given by the difference between the number of species and the rank of the stoichiometric matrix.

Exercise 4. Sensitivities and metabolic control analysis ⟨9 points total⟩

a) Complete the following sentence: "The concentration control coefficient $C^{S_i}_{v_k}$ of species S_i with respect to reaction rate v_k in a metabolic network describes ..." ⟨3 points⟩

> ... the change in the steady state concentration of species S_i upon an infinitesimal change in the reaction rate v_k.

b) One of the summation theorems of metabolic control analysis states that all the flux control coefficients in a metabolic network for one steady state flux (J_i) sum up to one, $\sum_{k=1}^{r} C^{J_i}_{v_k} = 1$, with r the number of reactions.
Briefly discuss the consequences of this, particularly in a situation where one tries to increase the flux over a pathway by making the reaction with the highest control ("rate limiting step" or "bottleneck") faster. ⟨3 points⟩

> If the flux control coefficients of a steady state flux J_i are very small with respect to all reactions except for one specific reaction R_j, then reaction R_j can be called a rate limiting step. It has the highest control over J_i in this situation and, therefore, is a promising target for manipulation if one wants to increase the steady state flux J_i. However, as reaction R_j is made faster, usually, the control coefficient $C^{J_i}_{v_j}$ decreases. This means that reaction R_j looses control over J_i or, in other words, bigger and bigger changes in the rate of R_j are needed for a certain change in J_i. According to the summation theorem all flux control coefficients of a certain flux have to add up to one. Therefore, as reaction R_j looses control over flux J_i (its control coefficient decreases) one or many other reactions have to gain control (their control coefficients increase). This inevitable shift of control means that now other reactions might be rate limiting and should be targeted if the flux J_i is to be increased.

c) Calculate the ϵ-elasticity, $\epsilon^v_S = \frac{[S]}{v}\frac{\partial v}{\partial [S]}$ with respect to the substrate of the following reversible mass action reaction $S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} P$. ⟨3 points⟩

> $\epsilon^v_S = \frac{[S]}{v} \cdot \frac{\partial v}{\partial [S]} = \frac{[S]}{k_1 \cdot [S] - k_2 \cdot [P]} \cdot k_1 = \frac{k_1 \cdot [S]}{k_1 \cdot [S] - k2 \cdot [P]} = \frac{1}{1 - \frac{v_r}{v_f}}$

Exercise 5. Stochastic modelling ⟨6.5 points total⟩

a) Name two advantages and two disadvantages of using stochastic modelling and simulation for the study of biochemical systems. ⟨3.5 points⟩

Advantages:

- Takes into account the stochastic timing of discrete reactive events in the system, i.e. stochastic effects such as noise-sustained oscillations or stochastic resonance can be studied
- Usually works on discrete particle numbers instead of concentrations, e.g. it is suitable even for very small particle numbers
- Appropriate for bi- or multistable systems. Here, the probability of reaching the different stable states can be calculated, or spontaneous switching between them can be studied
- Extinction of chemical species can be modelled correctly
- Suitable for modelling rare events

Disadvantages:

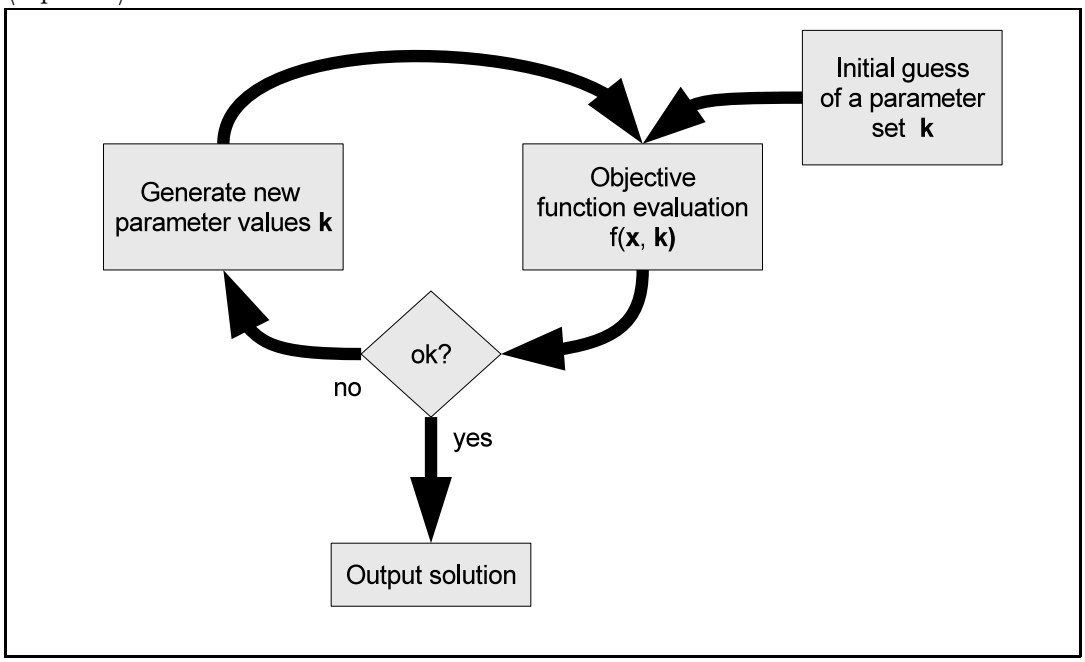
- Computationally demanding (very long computation time) if the system contains fast reactions
- It lacks behind in terms of analysis methods compared with modelling using ordinary differential equations
- Models must not contain reversible reactions, e.g. many models have to be adapted before they can be simulated stochastically
- Under what conditions exactly more complicated kinetic functions can be used still has to be studied

b) Most models that you can download from model databases have to be adapted before they can be simulated stochastically. Briefly describe the two most important points that have to be considered in this context and why. ⟨3 points⟩

- The volume of the model has to be adjusted to reflect (small) cellular volumes. Often, downloaded models are set to have an unrealistically high (default) volume, such as 1 litre. Because concentrations are particle numbers divided by volume, this means that the particle numbers in the system are much too big. This leads to very long computation times since every reactive event in the system has to be calculated, or can even make stochastic simulation impossible if particle numbers are too big to be held in integer variables.
- All reversible reactions (and the corresponding kinetic functions) have to be split into two reactions, one for the forward and one for the backward direction. Even if the average net flux of a reaction is zero (e.g. the forward flux equals the backward flux in a deterministic model) this reaction can still cause stochastic effects due to random fluctuations in the forward and backward direction.

Exercise 6. Optimisation & parameter estimation/fitting ⟨15 points total⟩

a) Sketch (draw a flowchart of) how most optimisation algorithms function in an iterative way to find parameter sets that lead to a desired behaviour of the system ⟨5 points⟩



b) Briefly explain what an objective function in optimisation is generally. What does the objective function describe when it is used in parameter fitting/estimation? In this special case, do we have to minimise or maximise the objective function? ⟨5 points⟩

An objective function is a function of the model behaviour. Its scalar value expresses the goodness of the model with respect to a certain desired behaviour. The model behaviour is determined by the specific parameter set used. Therefore, the objective function can be optimised by adjusting the parameters. The desired behaviour can correspond to a maximum or a minimum of the objective function and, consequently so, the task can be to minimise or to maximise the objective function to achieve the desired behaviour.
In the case of parameter fitting the objective function describes the difference between the simulated behaviour and an experimentally measured system behaviour. In this case the objective function has to be minimised. This minimum corresponds to the parameter set that leads to the smallest difference between simulated and experimentally observed behaviour.

c) Describe the problem of non-identifiability in parameter estimation/fitting. ⟨5 points⟩

Dependencies between parameters or insufficient data can make parameter fitting difficult or even impossible. This is called non-identifiability. Non-identifiability comes in two different types, structural and practical non-identifiability.

- Each parameter set uniquely determines the system's behaviour (steady states, time course, etc.). However, different parameter sets might lead to the same system behaviour, e.g. there exists a many-to-one relation between parameter sets and behaviours. In this case only combinations of parameters can be determined and equally good fits can be had with different single parameter values. If the model contains such dependencies then it is structurally non-identifiable no matter what data is available.
- Practical non-identifiability is when the available data is not sufficient, e.g. in terms of temporal resolution or measurement noise, to adequately fit parameters.