

UNIVERSITY OF TRENTO



International PhD Program in Biomolecular Sciences Centre
for Integrative Biology
XXX CYCLE



A Systems and Synthetic Biology Framework
for Regulatory Systems

Tutors:

Prof. Martin Hanczyc
Prof. Alberto Inga

Advisors:

Dr. Ozan Kahramanoğulları
Prof. Corrado Priami

Ph.D. Thesis of

Cansu Uluşeker

University of Trento Centre for Integrative Biology

The Microsoft Research – University of Trento Centre for Computational Systems Biology

October 2018

I would like to dedicate this thesis to my loving family; my mother, my father and my brother.

Declaration

I Cansu Uluşeker confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

A handwritten signature in black ink, appearing to read 'Cansu Uluşeker', with a stylized flourish at the end.

Cansu Uluşeker
October 2018

Acknowledgements

I had a remarkable research experience during my Ph.D. journey. I am deeply indebted to my tutor Martin Hanczyc and thesis advisor Ozan Kahramanoğulları for their invaluable guidance, and for giving me this wonderful opportunity to work on the LIAR project. This Ph.D. would not have been possible without their expertise, dedication and enthusiasm. Also, I am very thankful to Ozan Kahramanoğulları for his patience and support. This thesis has been thoroughly reviewed and proofread by him, for which I am very grateful.

I would like to express my sincere gratitude to my other project tutor Alberto Inga and advisor Prof. Corrado Priami for the continuous support of my Ph.D. study and related research, for their patience, motivation, and knowledge. Their guidance helped me in all the time of research.

Besides my advisor, I am grateful to my co-advisors Luca Marchetti and Alice Matone for all their help and guidance. They have invested a remarkable amount of time and effort in my scientific growth.

Research on this thesis was made possible by the fellowship provided by COSBI and LIAR project, for which I am truly grateful. I would like to thank all the researchers of COSBI, for their insightful comments and encouragement.

My sincere thanks also goes to Professor Juan Nogales and Jesus Torres, who provided me the opportunity to collaborate with their teams and gave access to their laboratory and research facilities during my period abroad. Without their precious support it would not be possible to conduct the experimental part of my research.

I thank my fellow Ph.D. students (and former ones) at COSBI and CIBIO for the stimulating discussions and for all the fun we have had in the last years.

I thank all my friends for all the supports and good times we have had together.

I would like to thank my family; my mother, for supporting me throughout the Ph.D. and my life in general. I thank my father, he has always believed in me and has supported me. I thank my brother, for being my role model always. Thank you for all.

Abstract

Biological regulatory systems are complex due to their role in living organisms in modulating precise responses to changes in internal and external conditions. In this respect, mathematical models have become essential tools to address their complexity for a better understanding of their mechanisms. The vision here, based on integrating experimental and theoretical techniques, provides a systematic means to quantitatively study the characteristics of the interactions that occur in living organisms. The outcome of such an endeavour should provide insights in terms of predictions and quantifications for further investigations in systems and synthetic biology.

In this thesis, we establish an integrated modelling framework that can ensure the interaction of experimental biology with the development of quantitative mathematical descriptions of biological systems. To this end, we develop a framework to simulate and analyse biological regulatory systems by integrating different layers of regulatory information. The work herein presents a biological model development workflow in terms of a step by step approach, highlighting challenges and “real life” problems associated with each stage of model development.

In the first part, we have focused on applying systems and synthetic biology modelling tools to the phosphate system at the cellular and genetic levels in *Escheria coli*. Then, we have analysed the interaction mechanisms and the dynamic behaviour of the phosphate starvation response deactivation and evaluated the role of phosphatase activity. We have investigated how the properties of these signalling systems depend on the network structure. Moreover, we have constructed detailed transcriptional regulatory network models and models for promoter design. In the second part, we have designed a multi-level dynamical set up by providing a novel closed loop whole body model of glucose homeostasis coupled with molecular signalling. We have then developed a system embracing the intracellular metabolic level, the cellular level involving the dynamics of the cells, the organ level, and the processes within the whole body. The output of each model directly has been fed with the variables and the parameters of the next aggregated model. This allowed us to observe the metabolic changes that occur at all levels and monitor inter-level communications for Type 2 Diabetes disease.

Table of contents

List of figures	xiii
List of tables	xxvii
Nomenclature	xxvii
1 Introduction	1
1.1 Systems and Synthetic Biology	2
1.2 Research Area	3
1.2.1 Phosphate management system in <i>Escherichia coli</i>	3
1.2.2 Multi-level model of glucose homeostasis	5
1.3 The Thesis Outline	6
2 Design Approaches in Systems and Synthetic Biology	9
2.1 Dynamic Mathematical Modelling in Biology	10
2.1.1 Modelling biomolecular reaction networks	11
2.1.2 Multi-level systems and models	14
2.2 Basic Features of Dynamic Mathematical Models	15
2.2.1 State variables and model parameters	15
2.2.2 Linearity and nonlinearity	15
2.2.3 Steady-state and transient behaviour	16
2.2.4 Deterministic and stochastic models	16
2.2.5 Model fitting algorithms	17
3 Mechanisms of Switching Response to External Phosphate Levels in <i>Escheria coli</i>	19
3.1 Introduction	20
3.2 Material and Methods	22
3.3 Results and Discussion	26

3.4	Conclusion	34
4	Quantifying Dynamic Mechanisms of Auto-regulation in <i>Escherichia coli</i> with Synthetic Promoters in Response to Varying External Phosphate Levels	37
4.1	Introduction	38
4.2	Material and Methods	41
4.3	Results and Discussion	45
4.4	Conclusion	56
5	A Novel Closed-loop Multi-level Model of Glucose Homeostasis for Normal Glucose Regulation and Type 2 Diabetes	59
5.1	Introduction	60
5.2	Materials and Methods	62
5.2.1	Glucose physiology at the whole-body level	62
5.2.2	Insulin signalling at the cellular (adipocyte) level	64
5.2.3	The hierarchical model	65
5.2.4	The mathematical model	67
5.3	Results	80
5.4	Discussion	84
5.5	Conclusion	89
6	Discussion and Future Directions	91
	References	99
	Appendix A Appendix A	111
	Appendix B Appendix B	113
	Appendix C Appendix C	131
C.1	Sensitivity analysis	146

List of figures

- 3.1 Schematic representation of the different stages of Pho regulon. Green coloured shape represents ABC transporter, dark blue one is PhoR protein, light blue shape stands for PhoU protein and pink stands for PhoB protein. The signalling processes of activation and deactivation correspond to different states of PhoR. **Activation:** When external inorganic phosphate (P_i) is depleted, the system is in growth condition. Under this condition, Pho regulon is active and produces its components. ABC transporter binds the external P_i , internalises and releases it to the cytosol. According to the current biological model in the literature, PhoR assesses P_i availability by monitoring the ABC transporter. This is done by relaying the signal via PhoU. When the system is active, PhoU does not stabilise PhoR and PhoR passes to the active state ($PhoR_a$). $PhoR_a$ phosphorylates PhoB. Phosphorylated PhoB then acts as a transcription factor for the operon. **Deactivation:** PhoR acts as a phosphatase on phosphorylated PhoB. When environmental P_i is in excess, Pho regulon is inhibited. PhoU stabilises PhoR. This prevents PhoR from autophosphorylating itself. The signal is thus propagated to PhoR, resulting in its inhibition state $PhoR_i$ 23

- 3.2 A schematic representation of Pho regulatory system interactions resembling a logic gate circuit. Figure demonstrates the conceptual presentation of TCS signalling processes from an alternative perspective. ABC transporter and PhoU proteins are expressed by the same promoter, that is, pABCU. Similarly, PhoR and PhoB proteins are expressed by the same promoter, that is, pPhoBR. These are depicted on the left side of the figure. Expressions of ABC transporter, PhoU, PhoR and PhoB proteins refer to system inputs together with external inorganic phosphate P_i . System outputs are repressive PhoR (R_r) and active PhoB (B_a). Colours are used as described in Figure 3.1. The shape resembling an AND gate requires both incoming signals to be present for output signal. The shape resembling a single input NOT gate requires the incoming signal to be absent to produce a signal. (1) P_i is taken in to cytoplasm by the ABC transporter. (2) When the P_i is abundant, ABC transporter relays the signal of external concentration change to PhoU and activates PhoU (U_a). (3) Active PhoU stabilises PhoR and prevents PhoR dimers from autophosphorylating, and PhoR becomes repressive (R_r). PhoR is now inhibited that prevents activation of PhoB. Otherwise, when the P_i is absent, shown by NOT gate, (4) PhoR is free to autophosphorylate itself. This allows PhoR to become active, (R_a) and (5) when PhoB is present, PhoRa phosphorylates transcription factor PhoB and activates it (B_a). 24
- 3.3 The response of the model to variations in the external P_i levels. Model 1 includes the reactions 1, 2, 3a, and 4a and Model 2 includes 1, 2, 3b, and 4b. The simulations are performed for the external P_i concentrations of 0, 50, and 100 μM and a time-course of 4.5 hours is considered. PhoR activity is plotted in blue when P_i concentrations is 0 μM , it is represented with red when external P_i concentrations is 50 μM , and orange colour is used when P_i concentrations is 100 μM . Response times are represented with vertical lines. Red line is the response time when the external P_i concentration is 50 μM and the orange one is for the external P_i concentration of 100 μM 27

- 3.4 Heatmap displaying the effects of the changes in reaction rates r_1 , r_2 , r_{3a} and r_{3b} . Each parameter is varied by orders of magnitude (OM) higher and lower. The one on the left is for response time and on the right is for PhoR level change. They are computed when the system reaches the steady-state. For the cases that the system does not have a steady state, response time and PhoR level change are displayed as N.A.. Response time, in seconds, is calculated as the time needed to reach halfway between the initial and final level. The outcome of PhoR level change is normalised with respect to the control model. The analysis is done for Model 1 (M1) and Model 2 (M2) when external P_i (P_{ext}) are $50 \mu\text{M}$ and $100 \mu\text{M}$. Red and green represent the decreasing and increasing effects, respectively. Reaction rate r_4 variation is excluded because variations on it imply a departure from the expected system dynamics. 29
- 3.5 Heatmap displaying the normalised PhoR level change as a result of the variations in r_1 and r_2 together with variations in r_4 rate values. The analysis is performed for Model 1 (M1) and Model 2 (M2) when the external P_i concentrations are $50 \mu\text{M}$ (A) and $100 \mu\text{M}$ (B). Rates r_1 and r_2 are varied by orders of magnitude higher and lower from -6 to 2. PhoR level change from the beginning to the end of the simulation is computed and displayed. The outcome of PhoR level change is normalised with respect to the control model. 30
- 3.6 Heatmap displaying the response time, in seconds, as a result of the variations in r_1 and r_2 together with variations in r_4 rate values. The analysis is performed for Model 1 (M1) and Model 2 (M2) when the external P_i concentrations are $50 \mu\text{M}$ (A) and $100 \mu\text{M}$ (B). Rates r_1 and r_2 are varied by orders of magnitude higher and lower from -6 to 2. 31
- 3.7 The response of the model with PhoU to variations in the external P_i levels. Model 1 includes the reactions 1, 2, 3', 4', 5a and 6a, and Model 2 includes 1, 2, 3', 4', 5b and 6b. In all the experiments, the external P_i concentrations are 0, 50, and $100 \mu\text{M}$ and a time-course of 4.5 hours is considered. PhoR activity is plotted in blue when P_i concentrations is $0 \mu\text{M}$, it is represented with red when external P_i concentrations is $50 \mu\text{M}$, and orange colour is used when P_i concentrations is $100 \mu\text{M}$. Response times are represented with vertical lines. The red line is the response time of dynamics when external P_i concentration is $50 \mu\text{M}$ and the orange line is for an external P_i concentration of $100 \mu\text{M}$ 33

- 4.1 The transmembrane signal transduction due to external P_i levels and the Pho regulon activity when extracellular inorganic phosphate (P_i) is in excess (left) and when it is depleted (right). The ABC transporter consists of the extracellular domain PstS, transmembrane domains PstA and PstC, and intracellular domains PstB (see the main text for the description). *Left:* when external P_i is in excess, the ABC transporter activity is limited as the cell does not consume ATP for the P_i transport. According to the current model in the literature, PhoR assesses P_i availability by monitoring the activity of Pst transporter, and relays the signal from PstB via PhoU to PhoR. When the P_i influx is increased, PhoU stabilises PhoR, which is depicted with the green bars around PhoU. This prevents PhoR dimers from autophosphorylating (red). Consequently, the transcription factor PhoB does not become phosphorylated by PhoR. *Right:* Due to the ABC transporter activity, the external P_i binds to the PstS component of the ABC transporter. It is then translocated to the inner membrane domain of the transporter through PstC and PstA. Following this, PstB changes its conformation by consuming ATP. P_i is internalised and released to the cytosol. When the P_i influx through the ABC transporter reduces, PhoU does not stabilise PhoR, which is depicted with the red bar next to PhoU. As a result of this, PhoR becomes free to perform its auto-kinase-phosphotransferase activity, whereby it phosphorylates PhoB. Phosphorylated PhoB then forms a dimer to act as a transcription factor for the operons, resulting in PhoA, PhoB and PhoR expression. 40

- 4.2 Schematic representation of the experimental data and the control model, and its dynamics in response to varying external P_i concentrations. **A.** The data for the PhoA and PhoB expression are obtained using PCR amplified DNA from *E. coli* MG1655 genome and transcriptionally fused to the translational coupler BCD2 and the fluorescent ms-fgfp gene. **B.** The control model has been obtained from the experimental data and the chemical reaction network (CRN) described in Methods by applying a fitting procedure with the physiological ranges obtained from the literature in Table 4.1, and verified by sensitivity analysis. The deterministic ODE and stochastic simulations are performed by applying the standard translation from CRNs based on stoichiometry. The blue colour denotes the proteins, orange denotes the promoters in their active and inactive forms, and purple denotes the mRNA molecules. Filled arrowheads denote the reversible reactions. The red arrows denote degradation reactions, the green arrows denote complexations, the black arrows denote phosphorylation and dephosphorylation, and the blue arrows denote the transcription and translation reactions. The model species that are plotted in C and D are distinguished with frames. **C.** The dynamics of the highlighted species of the control model in panel B as a result of the fitting procedure together with the experimental data, as described in Methods, are plotted. In the first of the two steps, the blue curves are obtained by using only the experimental values of PhoA levels. In the second step, that delivers the control model, the red curves are obtained by using the experimental values of both PhoA and PhoB levels. Inclusion of the PhoB data highlights the contribution of the feedback mechanism to the response dynamics, which is otherwise not represented. The stochastic dynamics, plotted in grey, display the fluctuations in the control model that are due to small molecule numbers and are not observable in the deterministic simulations. **D.** The response of the control model to variations in the external P_i levels, which are represented as fold change factors, applied to the autophosphorylation propensities of PhoR. A higher external P_i concentration corresponds to a smaller factor and vice versa. 49
- 4.3 The variations in the mRNA and active promoter levels due to the unbinding rate of the promoter and the active transcription factor. An unbinding rate of 100/sec (A) results in much less spread in the steady state distributions in comparison to unbinding rates of 1000/sec (B) and 5000/sec (C). The variations are quantified as the ratio of the standard deviation and the mean. 50

- 4.4 Heatmaps for the activity of various promoter designs for pPhoA, and pPhoB and the resulting PhoA expression (bottom-row) under different external P_i concentration conditions. The heatmaps are ordered decreasingly from left to right according to the external P_i concentration given by the fold changes applied to the PhoR autphosphorylation reactions r1 and r2. The left most column with 1 as the fold change value is the starvation condition with $0\mu M$ external P_i . Each heatmap scans 100 simulations by applying 10 different fold change values to the promoter binding rates r10 and r11 as well as 10 different fold change values to the promoter unbinding rates r10r and r11r. The upper row displays the resulting steady state levels of the active promoter pPhoAa, whereas the lower row displays the yield of PhoA gene expression measured as the area under the curve (AUC). The intersection of the dashed lines in the left column delivers the experimentally observed regime observed in Figure 4.2. The levels of this regime, that display the starvation response, are highlighted in all the heatmaps. 53
- 4.5 Comparison of the experimental data on PhoA expression with the synthetic promoter Pliar00117 together with the simulation results with varying external P_i concentrations and promoter parameters that model synthetic designs. The experimental data on the starvation response with the synthetic promoter Pliar00117 is represented as hollow circles. As in Figure 4.4, the plots are ordered decreasingly from left to right according to the external P_i concentration given by the fold changes applied to the PhoR autphosphorylation reactions r1 and r2. The left most column with 1 as the fold change value is the starvation condition with $0\mu M$ external P_i . Each plot displays four simulations with varying fold change values applied to promoter binding and unbinding rates that model various promoter designs. A modified promoter (blue curve) can reproduce the starvation response in low as well as high external P_i concentration, and reproduce the experimental data under Pliar00117 starvation conditions. 54
- 4.6 Heatmap displaying the results of the sensitivity analysis by considering the physiological interval in Table 4.1. For each parameter, the maximum and minimum values within its physiological range are considered for simulation, and the area under the curve (AUC) for each species is computed. The difference of the AUC for the maximum and minimum parameter values are then normalised with the AUC of the control model. Red represents the decreasing effect and green represents the increasing effect. 55

- 5.1 Graphical representation of the whole-body glucose metabolism as considered in our model, according to the notation introduced in [61]. Only the organs/tissues for which a variable has been explicitly included in the model are depicted in the figure (other key organs/tissues of glucose metabolism, like pancreas and brain, are not displayed in the figure even if their effect has been indirectly taken into account in model equations, see Results). Adipose tissue is coloured in yellow to highlight that it is the part for which a model at the cellular level is also provided (see Figure 5.2). Green ovals (hormones) and orange rectangles represent model variables; arrows represent mass transfer (white head), stimulation (black head) and inhibition (T head). . . . 63
- 5.2 Graphical representation of the model describing the insulin signalling in adipocytes at the cellular level, according to the notation introduced in [61]. Solid arrows represent state modification, while dashed arrows indicate reaction stimulation. Protein complexes are coloured in yellow, green ovals represent the active and inactive feedback protein, while the orange rectangles represent all the other components of the cellular model. The plasma membrane of the adipose cell is represented in yellow and it separates the cytosol (light yellow horizontal lines) from the interstitial fluid (blue and white vertical lines). The variables I and G indicate insulin and glucose concentration in plasma (compartment not represented), which regulate the amount of interstitial insulin (INS_A) and glucose (Gt_A), respectively. For the sake of simplicity, we highlighted the five variables linking the cellular level to the whole body description (namely plasma insulin, interstitial insulin, plasma glucose, interstitial glucose and intra-adipocitary glucose) by adding the corresponding names in parenthesis. 66
- 5.3 Model dynamics at the whole body level. Each plot represents one variable dynamics. The normal glucose regulation (NGR) and T2DM conditions are shown in green and black, respectively. The red and blue lines delimit the physiological lower and upper ranges of variables (see also Table 5.1) . . . 81
- 5.4 Model dynamics of the insulin signalling in adipocytes at the cellular level (only a subset of key variables is represented). The normal glucose regulation (NGR) and T2DM conditions are shown in green and black, respectively. . 83

5.5	Model simulation and data fitting, normal glucose regulation (NGR) condition. Each plot represents the corresponding time courses for the indicated insulin signalling intermediaries. The experimental data are taken from Nyman et al. [127] and are represented with circles and error bars (a.u. indicates arbitrary units). The time course represents the model simulation.	84
5.6	Model simulation and data fitting, T2DM condition. Each plot represents the corresponding time courses for the indicated insulin signalling intermediaries. The experimental data are taken from Nyman et al. [127] and are represented with circles and error bars (a.u. indicates arbitrary units). The time course represents the model simulation.	85
A.1	The ODEs for the Model 1 reactions in Chapter 3.	111
A.2	The ODEs for the Model 2 reactions in Chapter 3.	111
A.3	The ODEs for the PhoU Model 1 reactions in Chapter 3.	112
A.4	The ODEs for the PhoU Model 2 reactions in Chapter 3.	112
B.1	The ODEs for the model reactions in Chapter 4.	114
B.2	The deterministic ODE simulation results that are fitted with only the experimental data on PhoA expression levels are plotted in blue. The deterministic ODE simulation results that are fitted with the experimental data on both PhoA and PhoB expression levels are plotted in red. In all the experiments, the external P_i concentration is $0\mu M$ and a time-course of 4.5 hours is considered. For the simulations that consider both PhoA and PhoB data, stochastic simulation results, plotted in grey, are also provided. In the simulations that include both of the PhoA and PhoB data, the values for r14 and r15 in Table 1 in the main text are obtained by increasing these parameters from 0.05 to 0.13 and from 0.03 to 0.035, respectively. Similarly, the disassociation rates, r3r, r5r, r7r, and r8r are increased from 74.94 to 94.94 to obtain comparable DiPhoBpp levels.	115
B.3	The cumulative output of simulations under the effect of random perturbations on the reaction rate r14. The reaction rates in the model can vary within a range in different cell types and cellular environments. Therefore, we applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.01 to 2, which are larger than the physiological range. The 1 fold change is plotted in black.	116

-
- B.4 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{15} . The reaction rates in the model can vary within a range in different cell types and cellular environments. Therefore, we applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.01 to 2, which are larger than the physiological range. The 1 fold change is plotted in black. 117
- B.5 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{3r} . The reaction rates in the model can vary within a range in different cell types and cellular environments. We applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.1 to 10. The 1 fold change is plotted in black. 118
- B.6 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{5r} . The reaction rates in the model can vary within a range in different cell types and cellular environments. We applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.1 to 10. The 1 fold change is plotted in black. 119
- B.7 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{7r} . The reaction rates in the model can vary within a range in different cell types and cellular environments. We applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.1 to 10. The 1 fold change is plotted in black. 120
- B.8 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{8r} . The reaction rates in the model can vary within a range in different cell types and cellular environments. We applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.1 to 10. The 1 fold change is plotted in black. 121

- B.9 Stochastic simulation results when DNA unbinding rates r_{10} and r_{11} are set to $100s^{-1}$ together with deterministic simulation results. The unbinding kinetics can have a significant impact on the stochastic dynamics of gene expression network. When the promoter unbinding rate increases, the promoter activity, which works at a faster time-scale, leads to much more noisy fluctuations (see Supplementary Figures B.10 and B.11). A stochastic simulation with this modified rate is plotted in gray. The deterministic simulations with our model (red) and the deterministic simulations with the changes above (green) display similar dynamics. 122
- B.10 Stochastic simulation results when DNA unbinding rates r_{10} and r_{11} are set to $500s^{-1}$ together with deterministic simulation results. The unbinding kinetics can have a significant impact on the stochastic dynamics of gene expression network. When the promoter unbinding rate increases, the promoter activity, which works at a faster time-scale, leads to much more noisy fluctuations (see Supplementary Figures B.9 and B.11). A stochastic simulation with this modified rate is plotted in gray. The deterministic simulations with our model (red) and the deterministic simulations with the changes above (green) display similar dynamics. 123
- B.11 Stochastic simulation results when DNA unbinding rates r_{10} and r_{11} are set to $5000s^{-1}$ together with deterministic simulation results. The unbinding kinetics can have a significant impact on the stochastic dynamics of gene expression network. When the promoter unbinding rate increases, the promoter activity, which works at a faster time-scale, leads to much more noisy fluctuations (see Supplementary Figures B.9 and B.10). A stochastic simulation with this modified rate is plotted in gray. The deterministic simulations with the changes above with slower unbinding rates (green) lead to slightly slower dynamics in comparison to deterministic simulations with our model with this rate set to $1000s^{-1}$ (red). 124
- B.12 Simulation results with varying external P_i concentrations. A decrease in the external P_i concentration is modelled as a fold change, listed in the legend, which is applied to the autophosphorylation reactions of PhoR. A fold change of 1 gives the starvation condition with $0\mu M$ external P_i . A lower fold change value represents an increased PhoU activity. Active PhoU inhibits PhoR, and thereby prevents it from autophosphorylating, which otherwise relays the signal downstream to the promoters pPhoA and pPhoB. 125

- B.13 Deterministic simulations with the model with varying fold changes applied to r_8 , the association rate of PhoR and PhoB, whereby PhoR acts as a phosphatase. The plots display the effect of the perturbations on DiPhoBpp in comparison to the case, where the fold change is 1. 126
- B.14 Deterministic simulation with the model with varying fold changes applied to degradation and dilution terms r_{17} , r_{18} , r_{19} , r_{20} , and r_{21} . Together with the Supplementary Figures B.15 and B.16, the plots display the effect of degradation and dilution under the conditions of reduced phosphorylation rates for PhoR resulting from an increase in external P_i concentration. . . . 127
- B.15 Deterministic simulation with the model with varying fold changes applied to degradation and dilution terms r_{17} , r_{18} , r_{19} , r_{20} , and r_{21} with a concomitant fold change of 0.5 applied to r_1 and r_2 . Together with the Supplementary Figures B.14 and B.16, the plots display the effect of degradation and dilution under the conditions of reduced phosphorylation rates for PhoR resulting from an increase in external P_i concentration. 128
- B.16 Deterministic simulation with the model with varying fold changes applied to degradation and dilution terms r_{17} , r_{18} , r_{19} , r_{20} , and r_{21} with a concomitant fold change of 0.1 applied to r_1 and r_2 . Together with the Supplementary Figures B.14 and B.15, the plots display the effect of degradation and dilution under the conditions of reduced phosphorylation rates for PhoR resulting from an increase in external P_i concentration. 129
- B.17 Heatmap displaying the results of the sensitivity analysis by considering all the model parameters. Each parameter is varied by 3 orders of magnitude higher and lower and the area under the curve (AUC) for each species is computed. The outcome is normalised with the AUC of the original model. Red represents the decreasing effect and green represents the increasing effect. The model is sensitive to a parameter when a variation in the input parameter causes a change in the model output. The amount of change quantifies the sensitivity. The sensitivity analysis results predict the system behaviour under varying conditions. 130

C.1	Heatmap resuming the results of the sensitivity analysis by considering the parameter estimates computed for the normal glucose regulation (NGR) condition. For each parameter, the differences between the two AUC ratios computed with the maximum and the minimum considered value of k are provided with respect to each model variable of the whole body model. Values lower than 0.15 have been considered negligible and are not displayed in the heatmap. The parameter c_2 does not affect model dynamics in the NGR condition, however, it has been estimated in the T2DM condition, where an impact on the system dynamics can be appreciated.	148
C.2	Heatmap resuming the results of the sensitivity analysis by considering the parameter estimates computed for the T2DM condition. For each parameter, the differences between the two AUC ratios computed with the maximum and the minimum considered value of k are provided, with respect to each model variable of the whole body model. Values lower than 0.15 have been considered negligible and are not displayed in the heatmap.	149
C.3	Oral glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	150
C.4	Stomach glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	150
C.5	Intestine glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	151
C.6	Plasma glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	151
C.7	Plasma insulin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	152
C.8	Plasma incretin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	152
C.9	Plasma glucagon chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	153
C.10	Liver glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	153
C.11	Muscle glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	154
C.12	Adipose glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	154

C.13 Plasma leptin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	155
C.14 Plasma ghrelin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition	155
C.15 Oral glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	156
C.16 Stomach glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	156
C.17 Intestine glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	157
C.18 Plasma glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	157
C.19 Plasma insulin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	158
C.20 Plasma incretin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	158
C.21 Plasma glucagon chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	159
C.22 Liver glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	159
C.23 Muscle glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	160
C.24 Adipose glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	160
C.25 Plasma leptin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	161
C.26 Plasma ghrelin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition	161

List of tables

3.1	The reaction rates and units of the control model.	26
4.1	Reactions and deterministic rates obtained from the physiological ranges in [2, 12, 129].	46
5.1	Physiological ranges of the whole body variables in the normal glucose regulation (NGR) condition. The ranges for H, S and L have not been specified (N.A. not available), as there is no clear upper limit to the amount of food (or glucose) one could ingest	78
C.1	Initial values of model variables in agreement with a morning fasting state. For each variable two initial values are provided, one for the normal glucose regulation (NGR) condition and one for T2DM. Values have been computed according to the estimation procedure described in the text.	132
C.2	Parameter estimates of the model in the normal glucose regulation (NGR) and T2DM conditions. Values have been computed according to the estimation procedure described in the text.	137

Chapter 1

Introduction

The goal of biology is to understand the functioning principles of biological systems. In early 1950s, Watson and Crick identified the structure of DNA [171]. This identification revolutionised the outlook on biological systems, especially from the point of evolution as a force acting on these systems, and the complex phenomena such as diseases. Since then, our understanding of the mechanisms of biological systems has significantly increased [91]. Now, it is known that the functioning of living organisms is controlled by networks of genes, proteins, small molecules, and their mutual interactions [130]. Moreover, the interactions between different components are universal characteristics of biological systems. The presence of such interactions determine the complexity of biological networks.

The complexity of biological phenomena highlight the need for looking at each component and its interactions, and consequently using proper mathematical tools to analyse and predict their functioning. Inter-disciplinary research in biology have made biological systems more comprehensible thanks to the emerging fields of systems and synthetic biology [27, 108]. Systems and synthetic biology are relatively new fields of biology that aim to develop a quantitative understanding of biological systems [91]. Prediction, control, design and understanding arise mainly from modelling biological systems by using formal models [108].

In this dissertation, we have used various modelling techniques in combination with wet-lab experiments to address biological questions related to *Escheria coli* (*E. coli*) phosphate response system and human glucose homeostasis. Within this framework, this thesis contains articles I contributed to that focus on these topics. These works, which are published in Artificial life conference ECAL 2017 [164] and in the journal of PloS ONE [165], are included in Chapters 4 and 5. At the time of writing, the work presented in Chapter 3 is accepted for publication in the proceedings of ALIFE 2018.

Here, I outline the scope of this chapter, which focuses, in particular, on models that arise in systems and synthetic biology. Section 1.1 gives an overview of systems and synthetic biology. Section 1.2 reviews the biological background of the systems studied in the models. The chapter concludes in Section 1.3 with an outline of the contributions.

1.1 Systems and Synthetic Biology

Systems biology aims at developing formal understanding of complex biological systems in their entirety by integrating all levels of information into cohesive models [158]. Biological systems consist of large numbers of components from gene networks to complete organisms [155]. The behaviour of a system at any given level of biological organisation is also dependent on the outputs and properties of systems at various levels [155].

The working assumption of the systems approach to biology is that the function and behaviour of biological systems can be understood by studying the component parts and their interactions. A complementary point of view in systems biology is that the hierarchy of biological levels and the ways in which they interact are also important [40]. This requires methods for studying different levels of biological organisation [39].

Multi-level methods help to analyse each component separately and investigate the effect of their interactions. Application of multi-level methods to biological complexity provides new observations. Such a perspective can help to to diagnose, define disease predilection, and develop the treatment strategies. This way, it offers an opportunity to redefine our approach to disease [127].

The rapidly-developing discipline of synthetic biology can be defined as the engineering of biology by exploiting the knowledge gained through systems biology [137]. The relationship between the two fields is analogous to the difference between forward and reverse engineering. Synthetic biology aims to build new biological networks to perform specific tasks, and modify the networks in order to achieve desired functionalities [108]. Thus, it collaborates with biology, mathematical modelling, information technology and biotechnology. Such a combination of disciplines allows us to construct robust and predictable synthetic networks. Moreover, it provides new perspectives for the production of chemicals in more environmentally friendly manners, with objectives in bioremediation, pollutant detection, and less expensive and more efficient energy production [25, 140].

Synthetic biology approach has a significant role in the development of bioremediation and programmable aspects of biology for sustainable environment strategies. Microorganisms can be redesigned using synthetic biology methods to produce environmental friendly products and used in bioremediation. Examples to these include biosensors that are designed

to signal the presence of environmental contaminants [89]. Other examples are engineered microorganisms that are enabled to transform the hazardous environmental pollutants, and recycle waste by converting agricultural waste into useful products [43].

Modelling in synthetic biology has contributed to the development of numerous modules for the precise control of protein expression and new devices such as toggle switch [57], an oscillator [42] and gene regulatory networks [109]. There is an increasing number of new results in this field. Moreover, through the development of applications from the design of gene circuits to robotic control, synthetic biology is becoming a global research enterprise.

The use of systems and synthetic biology models provide unprecedented opportunities in health and fight against disease, agriculture, manufacturing, energy production, and environmental remediation. The models in both systems and synthetic biology are able to formalise the knowledge about the biological process. They are also able to identify inconsistencies between hypotheses and observations [108]. Therefore, models can be used to predict the biological behaviours in untested conditions. Larger-scale models, which capture the interactions between hierarchical levels in the system, can provide predictions on the broader aspects of the phenotype. From this point of view, dynamical systems, as in this thesis, can have a fundamental role in investigations on biological systems, as well as in the design of synthetic living devices.

1.2 Research Area

1.2.1 Phosphate management system in *Escherichia coli*

The world is facing the problem of various forms of environmental pollution. For example, nutrient imbalance in water can give rise to the water contamination. Inorganic phosphate (P_i) is one of the main nutrients contributing the water quality problems when it is high in amount [150]. Microorganisms can be used in bioremediation to degrade P_i and to overcome the water contamination due to P_i [94]. Microorganisms can survive in many different environmental conditions due to their rich and versatile regulatory capabilities. Moreover, the nutritional capacity of microorganisms is completely varied, which brings about the possibility to use microorganisms as bioremediation of environmental pollutants. In this respect, the development of eco-friendly bioreactors that input waste water, carbon dioxide, and sunlight to output clean water, electricity, biomass and other mineral resources is a frontier in biotechnology with important implications.

E. coli provide diverse mechanisms and pathways to transform and transfer the P_i [168, 169]. *E. coli* have been used in many scientific investigations due to their ease of genetic

manipulation that enables them to be used in bioremediation with bioreactors [22, 166]. This makes *E. coli* one of the most well studied microorganisms. Therefore, *E. coli* provide a well suited test-bed for studying P_i intake together with protein-protein interactions in the intake pathways as well as in relation to genetic components.

The aim of the work in this dissertation in the context of the European project Living Architecture (LIAR) is to provide an integrated modelling framework across different scales to quantify the metabolites and biochemical mechanisms that feedback to the larger scale output of the entire system. To this aim, we have constructed detailed phosphate regulation and transcriptional regulatory network models that describe how *E. coli* sense the environmental P_i and import it into the cell. This way, we explore how the coupling of modelling with wet-lab experiments at different scales in synthetic promoter design can drive the technologies targeted at the inorganic phosphate intake and their use in bioremediation and synthetic biology applications.

The physiological characteristics of P_i transport have been extensively studied, especially in *E. coli*. In *E. coli*, there are two major phosphate transport systems; the low affinity phosphate inorganic transport (Pit) system, and the high affinity phosphate specific transport (Pst) system [168, 169]. Pit system is constitutive, and is generated by electron transport. It is used by *E. coli* when the external inorganic phosphate is abundant [168, 169]. On the other hand, Pst system is an ATP-binding cassette transporter (ABC transporter) and it is active when the external P_i concentration is limited [81, 95, 168, 169]. *E. coli* sense the external P_i changes via the Pst system. In other words, the Pst system is responsible for sensing the environmental changes, also importing the P_i into the cell. Moreover, the Pst system relays the incoming signal to the two-component system (TCS) to adapt to current conditions. TCS allows *E. coli* to receive the signal on the environmental changes from Pst, and responds to the changes by regulating the expression of Phosphate (Pho) regulon genes [56, 95, 168, 169].

Pho regulon is a global regulatory circuit involved in bacterial phosphate management [95, 168, 169]. Previous studies have shown that the two-component system, the sensor kinase PhoR and the response regulator PhoB, participate in sensing the P_i level in the environment and regulate the expression of genes that are directly involved in P_i , forming the Pho regulon [95, 168, 169]. The expression of genes of the Pho regulon is inhibited when the environmental P_i is in excess. This inhibition requires an inhibitory form of PhoR and the protein PhoU. PhoU is essential for the repression of the Pho regulon under the high-phosphate conditions [168, 169]. While the precise mechanism of how PhoU acts is not yet understood, it is known that PhoU acts as a messenger protein and interacts with Pst and TCS, in particular with PhoR. Moreover, the members of the Pst and TCS system and

PhoU protein are induced by Pho regulon [56, 95, 168, 169]. This is a topic that I address in Chapter 3.

Although much is known about the molecular aspects of this signal transduction pathway, a comprehensive and structured mechanistic model of the Pho regulon is currently not available. A better understanding of the Pho regulon system can be supported and propagated by modelling in combination with synthetic biology methods. Results of modelling can highlight challenges and suggest modified strategies for developing and characterising artificial circuits. This allows us to ask whether the system as a whole may function more robustly and operate as an artificial biological circuit capable of programmed responsiveness towards the desired outcomes.

1.2.2 Multi-level model of glucose homeostasis

The human body requires continuous and stable glucose supply for maintaining its biological functions. Stable glucose supply comes from the homeostatic regulation of the blood glucose level. The maintenance of blood glucose homeostasis, which is controlled by various glucose consuming or producing organs, tissues and hormones, is complex and vital. Moreover, the impairment of glucose homeostasis can lead to the severe disease Type II Diabetes Mellitus (T2DM) [46]. The development of T2DM goes through different stages of impairment of the glucose-insulin system, and often through insulin resistance. Insulin resistance is the impairment of the regular crosstalk between insulin and the cells [36]. The latter does not respond adequately to normal levels of insulin [175], requiring always more insulin to stimulate the glucose uptake. The more the cells are insulin resistant, the more severe are the consequences for the organism. After decades of investigations, it is becoming clear that diabetes is a complex and highly heterogeneous disease [65]. The knowledge gap still hampers a comprehensive understanding of the etiologic processes at the level of individual organs or tissues, and subcellular derangements ultimately affecting the whole body metabolism.

The etiology and pathophysiology of T2DM are still partly unknown and have been investigated for decades. The crosstalk between glucose/insulin and the cells is a very complex mechanism, where many actors play different roles in the promotion or inhibition of the consequent glucose uptake. In this context, mathematical models of glucose and insulin dynamics are becoming fundamental tools for the diagnosis, description and understanding of T2DM. Such models describe the physiology of the glucose insulin system according to different levels of accuracy [33, 161]. A first such modelling milestone is the so called “minimal model”, which considers only glucose and insulin and is still widely used. Models including a broader range of physiological variables were introduced later, for instance,

Celeste et al. [20] introduced the glucagon hormone. Toghaw et al. [161] introduced ghrelin effects; Brubaker et al. [15] and later De Gaetano et al. [33] proposed models including the incretin hormones.

In this work, a step forward was made by proposing a novel multi-level and closed loop model of glucose homeostasis and insulin signaling pathway. This model simultaneously considers glucose, insulin, glucagon, incretins, ghrelin and leptin at the same time, and key tissues and organs such as the stomach, intestine, liver, pancreas, muscles and adipocytes.

The closed loop in the herein introduced model allows us to connect the input and the output of the system. This possibility dramatically increases the modeller capabilities of investigating the biological system, while providing a way of testing regulative phenomena that work at different time scales and that can have a delayed effect on the overall dynamics of the system. The idea here is to consider at the same time different (hierarchical) levels of abstraction of the same biological phenomenon. Whole body, organ, cellular and molecular levels are all strictly interconnected layers and physiological variations in one will inevitably affect the others. As most diseases, T2DM occurs and is diagnosed at the whole body level, but it arises and is maintained at the molecular and cellular levels. A multi-level approach helps us to understand the mechanism of the disease. For example there are multi-scale models that include the pancreas [51], the intestine [17] or the adipose tissue [126].

Since insulin resistance in adipocytes can influence other tissues, such as muscles and liver [72], and obesity is one of the main risk factors for the development of T2DM, the present work focuses on the molecular level of adipocytes and on the interactions with the whole body level through glucose, insulin and leptin. In this thesis, a novel and comprehensive multi level and closed loop model of whole body glucose dynamics is enriched with an additional layer that zooms on to the adipocyte molecular processes of glucose intake.

1.3 The Thesis Outline

Biological systems involve many types of interactions, including metabolic reactions, intra and extra-cellular interactions, protein interactions, and transcriptional regulations. Consequently, biological systems form complex networks and play important roles in living organisms to adapt to the environment, and develop different phenotypes. Many mechanisms and interactions of these networks are still not clear. Therefore, modelling plays a key role in the process to systematically understand, reconstruct, analyse and integrate these complex systems. The aim of this dissertation is to develop a such framework to analyse and reconstruct biological systems by integrating different layers of information.

This dissertation is organised as follows:

Chapter 2: This chapter gives an overview of the general concepts in using mathematical modelling to describe biological systems. First, the concepts of model, system and simulation are explained. Some definitions and theoretical background are presented here in order to better understand the theories employed. This chapter serves as a basis of modelling knowledge for later chapters of the dissertation.

Chapter 3: We study how PhoU protein interacts and PhoU regulon adapts to environmental changes. We present comprehensive analysis of the switching response of two component system (TCS) that reveal the impact of PhoU activity on the switch system. A full mathematical analysis is developed for two possible interactions of the PhoU and different external inorganic concentration levels. We present preliminary results about the modelling and construction of the intracellular organisation of Pho regulon proteins in *E. coli*. We carry out a systematic comparison of PhoU interaction predictions obtained by deterministic models. By comparing the results of the two predicted frameworks, we conclude that the type of biochemical mechanisms have a significant effect on the response time. The research performed in this chapter has been funded by COSBI and the European Union project LIAR. The work presented here appears in the proceedings of the ALIFE conference in July 2018.

Chapter 4: We present a mathematical and computational model of the *E. coli* TCS. We build the model for TCS signaling mechanism together with the regulatory promoter analysis. Our approach takes into account the key regulatory interactions of the Pho regulon and potential synthetic promoters. We demonstrate the feasibility of using the library of synthetic promoters in different external P_i conditions in *E. coli*. We detail the differential equations of the model based on standard translation of chemical reaction networks with respect to mass action kinetics. By using both deterministic and stochastic simulations, we analyse the system response and the emergent dynamics. All the steps are reported, which include model derivation, experimental design, parameter identification and model validation. The data presented in this Chapter have been derived in collaboration with Juan Nogales group (CNB, Madrid). Part of the work presented here has been published in the proceedings of the ECAL conference in September 2017 [164]. The research performed in this chapter has been done for the European Union project LIAR. At the time of writing, the work in this chapter has been submitted to the journal of Scientific Reports.

Chapter 5: A novel multi-level and close loop glucose homeostasis model is presented. The glucose dynamics zoomed on the adipocyte molecular processes of glucose intake is described in the model. We have detailed the physiology of the system for the whole body and molecular level for healthy and diabetics conditions. All the equations, the parameters and the initial values of the model are reported. Model results and data are expressed with the relative parameter fitting and validation methods. The research performed in this chapter has been funded by the COSBI. The work presented here has been published in the journal of PloS ONE in February 2018 [165].

Chapter 6: Conclusions are drawn and possible directions for future work are suggested.

Chapter 2

Design Approaches in Systems and Synthetic Biology

The complexity of biological phenomena inspires humans to obtain insights into the mechanism of biological systems in order to design new products or find solutions to various problems of biological nature. Systems and synthetic biology are interdisciplinary approaches that aim at improving our ability to understand and predict these complex biological mechanisms, and possibly control them.

Systems and synthetic biology both explore the biological paths by the application of other disciplines such as physics, computer science, mathematics, chemistry, and engineering. Systems biology sets out for a quantitative and mechanistic understanding of the functionality of the biological systems [158]. Moreover, synthetic biology focuses on designing artificial components to achieve particular functions of the biological systems [137]. In this endeavour, understanding and design are interdependent aspects, therefore, there is no simple distinction between basic and applied science [63]. Moreover, the new knowledge obtained via systems biology guides the design of better synthetic biology tools, which can in turn provide insights to systems biology.

This work addresses dynamic models of biochemical and multi-level molecular networks. These models are abstractions of reality. We take up the complementary use of modelling and computation in the study of intracellular networks and their interactions with other cellular components. We analyse such mathematical models that simulate the behaviour of the networks and design the analysis of evolutionarily engineered constructs.

2.1 Dynamic Mathematical Modelling in Biology

A biological system is a complex network of interacting components. Better understanding of complex biological systems allows us to identify and characterise the individual molecules in the system. Moreover, it also helps us to obtain a better understanding of the interactions between molecules and pathways. For example, the human body consists of approximately 10^{14} individual cells. Each of them is a complex system comprising thousands of different proteins and other biomolecules [11]. Although we might have the information on all the genes, we often lack an understanding of the interactions within components. The knowledge of individual components does not reveal functions that arise through the interactions of components. It is thus necessary to analyse the interactions and pathways for understanding complex biochemical mechanisms as well as diseases such as diabetes.

Mathematical and computational models are major tools for understanding how complex regulatory networks are connected. Mathematical models can assist in developing more comprehensive pictures of biological processes and investigate different experimental conditions for the biological systems. Moreover, they can be helpful for systematically determining the relevance of a specific molecule or pathway for the overall behaviour of the system. In other words, models can help to reveal how biological components interact and form networks, and how the networks generate whole cell functions [130]. Comparison of computer simulations and actual experimental data may help the researcher to identify the features of the biological system. Therefore, numerous mathematical methods for biological systems have been developed to address different categories of biological systems, such as metabolic processes or signalling and regulatory pathways [103, 119, 123].

Today, modelling approaches are essential for biologists, enabling them to analyse complex physiological processes, also for the industry, as a means for supporting drug discovery and development programs. Moreover, modelling in systems and synthetic biology lead to the simulation and analysis of biological systems with large numbers of components and interactions. Biological systems under different conditions can be relatively easily simulated *in silico* once a mathematical model is available. To describe the system behaviour, the dynamical model requires identification of the variables, and the values of the variables and how their interactions change over time. The dynamic model represents the system itself, whereas the simulation represents the operation of the system over time. Moreover, the simulations can serve to make qualitative predictions. Simulations can then be checked for consistency with existing data [78].

There is a large variety of modelling approaches due to the need for organising biological knowledge as models. This section describes two modelling approaches, which are used in this dissertation to represent biological systems in terms of computational models. The use

of mathematical models in the construction of intracellular signalling networks and cellular interactions will be explained in more detail in Chapters 3 and 4 for an *E. coli* system and in Chapter 5 for human glucose homeostasis. Computer simulations in many cases require relatively low investment and less time compared with the typically more time consuming and expensive biological experiments. Thus, modelling efforts can accelerate the wet-lab process by narrowing down the experimental search space. Model based design is also used in synthetic biology. It is applied for choosing of components or predicting the effective combination for the system performance. The use of model-based design in the construction of synthetic promoter will be explained in more detail in Chapter 4.

2.1.1 Modelling biomolecular reaction networks

Chemical reaction systems are tools for understanding biology on a molecular level using dynamic models of intra- and extracellular processes [48]. There is a large variety of reaction models [48] in the literature. Here, we review the most common ones that are used in modelling dynamic biochemical interactions.

Mass action kinetics: It is the proposition that the rate of a chemical reaction is directly proportional to the product of the concentrations of the reactants. It was first formulated by Cato Maximilian Guldberg and Peter Waage in 1864 [67]. It remains as one of the most common kinetic assumptions used in modelling. This is the method we use extensively in this thesis and describe below in further detail.

Michaelis-Menten equations: Its name comes after the collaboration of biochemist Leonor Michaelis and physician Maud Menten [112]. Michaelis-Menten equations describe the rate of enzymatic reactions, by relating reaction rate to the concentration of the substrates. They arise from the general reaction dynamics



for an enzymatic reaction, where E is the enzyme, S is the substrate, ES is the enzyme-substrate complex, and P is the product [14]. The enzyme directly interacts with the substrate in a stoichiometric manner. The interaction results in the enzyme-substrate complex, leads to equilibrium [14].

The general scheme for the Michaelis-Menten dynamics is given by the equation:

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

Here, v is the initial velocity of the reaction and V_{max} represents the maximum velocity achieved by the system, at maximum substrate concentrations. K_M , the Michaelis constant, which shows the substrate concentration when the reaction velocity is 50% of the V_{max} . $[S]$ is the concentration of the substrate S [14].

Michaelis-Menten equations require certain assumptions to be true. For example, Michaelis-Menten equations are used when the steady-state assumption is valid. They assume that the enzyme-substrate complex reaches equilibrium. The enzyme concentration is considered to be lower than the substrate concentration. Moreover, only initial velocity and initial rates of the reactions are measured.

Hill functions: It was introduced by A. V. Hill in 1910 [74] to describe the binding of oxygen to hemoglobin. It describes the fraction of a molecule saturated by ligand as a function of the ligand concentration [98, 172]. However, it has been widely used in biochemistry, physiology, and mathematical models [98, 172]. The Hill function parameter provides a measure of the ligand affinity for the receptor. The flow is sublinear for low substrate and saturates for large substrate concentrations.

The Hill equation is commonly expressed as [122]

$$\theta = \frac{[L]^n}{(K_A)^n + [L]^n}$$

where θ is the fraction of the receptor protein concentration that is bound to ligand. L is the unbound ligand concentration. (K_A) is the half-maximal concentration constant. n is the Hill coefficient. It provides a measure of the cooperativity in a binding process. It describes the cooperativity of ligand binding in the following way [122]

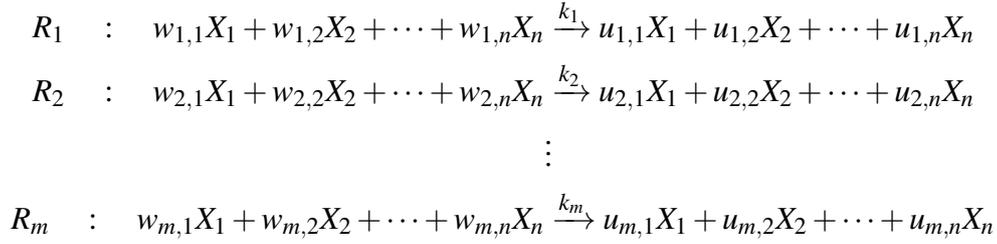
$n > 1$ Positive cooperativity: The reaction exhibits positive cooperativity with respect to substrate binding to the protein. Therefore, the binding of one molecule of ligand increases the binding of others.

$n < 1$ Negative cooperativity: The reaction exhibits negative cooperativity with respect to substrate binding, which means the binding of one molecule of ligand decreases the binding of others.

$n = 1$ No cooperativity: There is no cooperativity with respect to substrate binding to the protein. If $n = 1$, the Hill equation is reduced to its more familiar form known as the Michaelis-Menten equation.

In this thesis, we focus mainly on mass action kinetics, which can be described as systems of chemical reaction networks. A system of chemical reactions may be formulated using the

deterministic or the stochastic paradigms with the same formulation. Consider a system of chemical reactions with m reactions, denoted by R_1, R_2, \dots, R_m and n species, denoted by X_1, X_2, \dots, X_n . This $m \times n$ system may be represented as the following system of chemical reactions:



The corresponding, $m \times n$ sized, stoichiometric matrix, v , may be written as:

$$v = \begin{bmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,n} \\ v_{2,1} & v_{2,2} & \dots & v_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ v_{m,1} & v_{m,2} & \dots & v_{m,n} \end{bmatrix}$$

in which,

$$v_{i,j} = u_{i,j} - w_{i,j} \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n$$

The above notation of a system of chemical reactions and the corresponding stoichiometric matrix is common to deterministic and stochastic formulations. In spite of this similarity, both formulations have completely different governing equations which involve widely different physical and mathematical quantities and definitions. An important quantity is the parameter k_i associated with each reaction R_i . For the deterministic formulation, the parameter $k_i, i = 1, 2, \dots, m$ represent the reaction rate constants, whereas for the stochastic formulation, k_i is the stochastic rate constant or the reaction parameter [16]. Note that the deterministic reaction rate constant and the stochastic rate constants are not necessarily equal. However, one can be converted to the other. The stochastic reaction rate can be calculated from a deterministic rate. Deterministic simulations give a measure of concentrations of molecules whereas stochastic simulations give the numbers of molecules.

Here, we restrict the explanations to the deterministic simulations and not further expand on stochastic simulations. A detailed development of stochastic models can be found in Gillespie, 1976 [59] and many other works that build on it.

Given an $m \times n$ system as the one described before, the state of the system at time t , when modelled as a deterministic chemical reaction network is denoted by the vector-valued variable $\mathbf{x}(t) \in \mathbb{R}^n$. The i^{th} element of $x(t)$, $x_i, i = 1, 2, \dots, n$, represents the concentration of i^{th} species $X_i, i = 1, 2, \dots, n$. The variable $\mathbf{r} \in \mathbb{R}^m$ denotes the vector of rate of reactions. The rate of change of the state of the system may then be written as

$$\frac{d}{dt}\mathbf{x} = \mathbf{v}^T \mathbf{r}$$

with the initial conditions, $\mathbf{x}(0) = \mathbf{x}_0$. This system of equations obtained this way forms the deterministic formulation of the chemical reaction system $m \times n$.

2.1.2 Multi-level systems and models

Multi-level models have been recognised as relevant for the understanding of biological systems because of their emphasis of a strong connection between biological levels of functioning [130]. In particular, multi-level models are suitable for guiding biology from a qualitative to a quantitative point of view. This can be challenging, even if each model has similar processes and evolve on similar time- and length-scales [69].

Majority of multi-level models follow different perspectives to build the system. In order to link different scales, two major strategies can be employed: top-down or bottom-up [151]. The top-down approach starts with a large model and breaks it down into smaller more detailed segments. On the other hand, the bottom-up approach begins with the study of the subsystem components and connect them to describe and predict the behaviour of the entire system [151]. Another emerging approach is the middle out approach, which starts with an appropriate scale and exploit expanding knowledge to include both smaller and larger scales [151]. Each of these approaches has two subtypes with respect to their coupling. These are input-output coupling and direct coupling [69]. In input-output coupling, each variable depends either on the state of the low level system or on the state of the high level system. Therefore, variables can be inputs and outputs for the individual scales. On the other hand, in direct coupling, variables share the state between lower and higher levels. Models with input-output coupling are generally easier to simulate and to analyse in comparison to those with direct couplings [69].

Walpole et al. [167] introduced the terms series simulation, parallel simulation and integrated simulation to categorise different approaches. Series simulation gets the information from one scale and delivers it to be used as input for the other level (without feedback). Parallel and integrated simulation require communication between components simulations. They can often be computationally challenging. To overcome this, an approximation method

based on spatial homogenisation and time-scale is employed. This method decreases the overall complexity of the simulation. For example, on the tissue level, it is considered that all cells behave similarly. Therefore, a representative cell is evaluated in the model. Moreover, differences in time-scales of each processes can be used to simplify simulations.

Multi-scale models nowadays allow for the efficient simulation of coupled ODE-PDE systems [173]. Recently, Agent Based Modelling (ABM), also known as Individual Based Modelling (IBM), has used as time-resolved simulation of diffusion is based on time-scale separation [80]. It has become a very popular and powerful tool because of fast increasing computational power, which makes even large systems accessible [29].

2.2 Basic Features of Dynamic Mathematical Models

This section introduces some of the concepts in dynamic mathematical modelling that are relevant to the material covered in the rest of this thesis.

2.2.1 State variables and model parameters

The primary components of a dynamic mathematical model symbolise the species involved in the system. The existence of each species is assigned to a state variable within the model. The collection of group of these state variables is called the state of the system [78]. Such a representation provides a complete description of the system's condition at any given time.

Besides state variables, models also include parameters, whose values are fixed. Model parameters represent the environmental effects and interactions among system components as well as changes in the variables. Consequently, the model parameters typically have constant values during simulation: these values can be varied to explore system behaviour under perturbations or in altered environments. For instance, in Chapter 3 we focus on models of a signalling network in *E. coli*. In these models, external concentrations are varying state variables.

If all of the input parameters of the model are known, the output of the model can be computed. Therefore, the model is said to be explicit. However, sometimes the output parameters are known, and the corresponding inputs must be solved for by an iterative procedure, which is commonly termed as model fitting (Section 2.2.5).

2.2.2 Linearity and nonlinearity

If all the operators in a mathematical model exhibit direct proportionality, the resulting mathematical model is defined as linear. A model is considered to be nonlinear, when the

relationship of variables are not linear [78]. In other words, nonlinear connections do not follow any specific order. In mathematical modelling, if the constraints are represented entirely by linear equations, then the model is regarded as a linear model [49, 78]. If one or more of the objective functions, which is the function that it is to be maximised or minimised [49], or constraints are represented with a nonlinear equation, then the model is known as a nonlinear model [78].

2.2.3 Steady-state and transient behaviour

Simulations of dynamic models are time dependent and specified as steady-state or transient. In most of the deterministic models, biological characteristics do not change with time. Models of biological processes reach to steady behaviours in the long run. This steady behaviour, which models exhibit, is called a steady-state [78]. In continuous time, the steady-state behaviour is reached when the rate of change of the variable with respect to time is zero. In many systems, a steady-state is not achieved until some time after the system is started or initiated. This time course that leads from the initial state to the asymptotic behaviour is referred to as a transient state [78].

2.2.4 Deterministic and stochastic models

A mathematical model is called deterministic if its behaviour is dependent on a specified states and conditions, without influence of any other factors [78]. Therefore, repeated simulations under the same conditions are always in perfect agreement and can be perfectly replicated. However, real biological systems can have implicit influences on the model. Moreover, these influences can affect the analysis of the systems. In this regard, there are various methods to extend deterministic models to express the possible variations in dynamics [175]. A way of modelling these elements is by including stochastic influences. Stochastic models allow for randomness in their behaviour. The behaviour of a stochastic model result is affected both by specified conditions and by randomness [78]. Each repetition of a stochastic simulation thus yields a distinct sample of system behaviour.

Another difference between deterministic and stochastic models is seen in the simulation methods. The deterministic models are commonly simulated by solving a system of differential equations, more commonly ordinary differential equations (ODEs) due to their capability of capturing complex dynamics such as mass action kinetics. An ODE system is a set of differential equations that contain one independent variable such as time and one or more ordinary differential equations with respect to that independent variable. These variables are typically concentrations of model species.

MATLAB software has several libraries of functions for solving ODEs. In this dissertation, we have used Matlab ode15s solver for deterministic simulations. ode15s is designed to solve both stiff and non-stiff problems. A stiff equation is a differential equation where implicit methods for solving the equation can function better than explicit ones. Implicit and explicit methods are approaches used in numerical analysis for obtaining numerical approximations to the solutions of ODEs. An implicit method finds the solution by solving an equation involving the current state of the system and the later one, while an explicit method calculate the state of a system at a later time from the current time. For linear systems, a system of differential equations is termed stiff if the ratio between the largest and the smallest eigenvalue is large. Therefore, the numerical method must take small steps to obtain satisfactory results [96].

The stochastic simulations require a probabilistic method, which involves repeated generation of random numbers. For the case of simulating chemical reaction networks as in this thesis, this typically involves a version of the Gillespie's algorithm [59].

2.2.5 Model fitting algorithms

Mathematical models require parameters to describe the phenomena they are representing. Some of these model parameters are unknown or they are given within a range. In the absence of effective methods to determine parameter estimates, a model can produce a distorted representation of the observed phenomena. This can cause the rejection of its mechanistic description and mislead the data analysis. To parametrise the models, we resort to a fitting procedure, which often uses the least square method via various optimisation algorithms. The least square method can be described as solving an overdetermined system. The method minimises the sum of squared differences between the data values and corresponding modelled values to define best approximation [124].

There are various methods for optimisation including multi-start approach and Nelder-Mead method [114] When the system is non-linear, non-linear least squares methods get more solutions and return the local solutions. In order to provide a global solution of the system, the procedure can be repeated starting from different set of parameter estimates, that are encoded as different starting points. The best solution received by repeated procedure is then selected to ensure that the result is global. This procedure is known as multi-start approach [73]. In this dissertation, we have used Matlab lsqnonlin function from optimisation Toolbox for multi-start approach for least squares methods. This toolbox is commonly used to solve nonlinear least square problems that are applied for fitting experimental data.

Chapter 3

Mechanisms of Switching Response to External Phosphate Levels in *Escheria coli*

The phosphate economy in cells is essential in many biochemical processes from signal transduction to energy metabolism to DNA and RNA synthesis. All living systems therefore acquire and regulate phosphate in order to survive and reproduce. *Escheria coli* (*E. coli*), for example, regulate the inorganic phosphate (P_i) uptake in order to survive under phosphate-limiting conditions. To achieve this, *E. coli* have developed an accurate control mechanism, Pho regulon, to adapt to environmental perturbations of P_i , controlled by the PhoR/PhoB two-component regulatory system (TCS). The signalling of the TCS is delivered by interactions with the ABC transporter via PhoU. However, the exact mechanisms of interaction are unknown. Here, we propose mechanistic explanations for these mechanisms via a quantitative computational analysis, whereby we model plausible ABC and TCS state transitions. We analyse the interaction mechanism and the dynamic behaviour of TCS system deactivation in relation to the external P_i levels. We show that the behaviour of this system depends on the network structure. In particular, we use alternative models to demonstrate that variation in interaction patterns affect the response time of the system. Overall we show how to model a system where some key interactions are as yet unknown and to provide testable predictions that can easily be verified in the lab. This way, modelling is being used to increase our mechanistic understanding of important biological systems by defining and driving wet-lab experiments and to increase our biological understanding of the often complex relationship between an organism and its environment.

3.1 Introduction

Living cells, like bacteria, face a wide range of challenging environmental conditions such as nutrient limitation or exposure to antibiotics. Therefore, they must sense and rapidly produce the appropriate response to their environment. Response networks of *E. coli* are selected for fast and reliable adaptation to environmental conditions [19]. A mechanism that governs these responses is given by the two-component systems (TCSs) [148]. TCSs transmit information between a histidine kinase (HK) sensor and a cognate response regulator (RR) by phosphorylation [148]. A particularly well-characterised example of a TCS histidine kinase is PhoR in *E. coli*. PhoR responds to the changes in the external inorganic phosphate (P_i) level and controls the phosphorylation of the response regulator PhoB [142, 168]. The phosphorous compounds are essential nutrients for many biomolecules and have important roles in cell function and life [168]. Therefore, through such metabolic architectures, proper P_i signaling produces robust growth of *E. coli*.

Many parts of this P_i signalling pathway are known. *E. coli* control P_i metabolism through global phosphate management Pho regulon [95, 142, 168]. It is central to assimilation of P_i and regulation of P_i metabolism. It includes a number of P_i starvation-inducible genes such as [168, 169]

- four components of the ABC transporter Pst (PstSCAB), which consist of an extracellular binding protein (PstS), two transmembrane proteins (PstC, PstA) that form the transmembrane domain (TMD), and a dimer of cytosolic peripheral proteins (PstB), i.e., the nucleotide-binding domain;
- two that are members of the large family of two component system (TCS) that perform both positive and negative regulation, a sensor histidine kinase PhoR and a transcriptional response regulator PhoB;
- the chaperone-like PhoR/PhoB inhibitory protein PhoU.

Figure 3.1 displays a schematic representation of the Pho regulon system. Limiting concentrations of extracellular P_i activates the system, resulting in the phosphorylation of transcription factor PhoB by PhoR [21, 142, 168]. Signal transmission occurs through autophosphorylation of PhoR, followed by transfer of the phosphoryl group to PhoB [21, 95, 142, 168, 169]. Phosphorylated PhoB activates Pho regulon by binding to a consensus DNA sequence within the promoters of Pho regulon genes [21, 95, 142, 168, 169].

A return to the repression state occurs by a transition from the growth state to a phosphate-rich environment. The activation signal is interrupted by PhoR acting as a phosphatase on phosphorylated PhoB. System deactivation upon stimulus removal has an important task of

coordinating and regulating the cell growth [21, 95, 142, 168, 169]. Although, there has been much research on activation dynamics of TCS [52, 149], it is not well understood how PhoR receives the signal from PhoU and regulates the PhoB. It is known that, under environmental P_i depleted conditions, the Pho regulon is not induced (off state) and PhoB is maintained in the non-phosphorylated form by PhoR [21, 95].

It has been suggested that the PhoR/PhoB proteins assess P_i availability by monitoring the activity of the ABC transporter. In addition to the ABC transporter, PhoU is also required for P_i signal transduction [56, 139, 168]. PhoU constitutes an intermediate step between the ABC and TCS systems, inhibiting PhoR when the ABC system actively transports P_i [56, 139, 168]. When PhoU is mutated or deleted, PhoR continuously phosphorylates PhoB [56]. Then, phosphorylated PhoB delivers the high expression of the Pho regulon genes. PhoU is involved not only in the control of autokinase activity, but also in control of ABC transporter in order to avoid an uncontrolled P_i uptake. Moreover, it is known that PhoU and ABC transporter are also required for dephosphorylation of phospho-PhoB [168, 169]. However, the exact mechanism of PhoU action is not fully elucidated. Here, we address these uncertainties by applying alternative models based on plausible interaction patterns. This allows us to predict the unknown mechanisms of the system, which can be verified by experiments.

In order to address these questions, we have analysed the signal transfer pattern from the ABC transporter to the histidine kinase PhoR under the conditions of various external P_i concentrations. For this, we have designed alternative models to quantify and evaluate the events of the deactivation mechanism in the system. The main contribution of our model is thus the mechanistic quantification of the continuous relationship between the external P_i concentration and the PhoR activity in the cell, delivered by the signal transduction mechanism through ABC transporter and PhoU. This, in return, provides predictions on the effects of the ABC transporter and PhoU to the system deactivation.

In the following, we distinguish different types of interactions, and show how these patterns can correspond to signalling mechanisms. For this, we have first implemented our models in terms of hypothetical direct interactions between ABC transporter and PhoR. Thus, our preliminary model includes an implicit representation of the PhoU effect on the system. We have then extended and refined this implicit representation with an explicit mechanistic characterisation of PhoU that supports the preliminary observations. We have performed a computational analysis by simulations to enhance the understanding of the effects of multiple pathways on response dynamics and protein homeostasis in different environmental conditions. Our simulations with various mechanisms provide predictions on the effect of the system structure and PhoU activity. In particular, we have analysed how

strong phosphatase activity suppresses TCS activation, and quantified how fast the system is turned off once the stimulus due to external P_i concentration is removed.

3.2 Material and Methods

In a system of *E. coli*, it is important to balance P_i accumulation [95, 168, 169]. When external P_i levels are low, it becomes a growth limiting factor [133]. Therefore, the cell invests resources to produce proteins, rather than P_i accumulation [76, 133]. However, when the external P_i is in excess, the *E. coli* begins to enhance P_i accumulation and switches the TCS off. To explain this system, we have previously designed and analysed a two component system model that quantifies the dynamic mechanisms of auto-regulation in *E. coli* (for more details see Chapter 4 and [164]). At the time of writing, the work in Chapter 4 has been submitted to the journal. We have presented a dynamical model of Pho regulon, coupled with a layered description of its regulation in the experimental conditions of phosphate starvation. The model describes the dynamics of TCS together with the key regulatory promoter PhoB and experimental data on promoters PhoA and PhoB. Moreover, we have explored and verified emerging phenotypes with synthetic promoters and the response of the model to variations in the external P_i levels [164]. The model is parameterized according to the feasible range given in the literature, and fitted to the dynamic response of our experimental data. Our model provides a dynamic description of the core determinants of Pho regulon and promoter activities and their response to the change of external phosphate level.

In the following, to better illustrate the response of Pho regulon and TCS signalling to the different external P_i levels, we present a model as a conceptual representation of switching of TCS. We have considered a set of reactions for the signal of external P_i with the TCS histidine kinase as the target for delivery. As a preliminary step, the conceptual model contains implicit representation of the interactions with the TCS. We describe the formation of repression complex, which involves PhoR and ABC transporter, together with an implicit effect of PhoU protein. The model is consistent with observations in the literature, and maintains accurate relationships between all the components in the system. This enables the model to demonstrate the effect of PhoU protein and its integration to the system.

From an alternative perspective, Figure 3.2 shows how the regulatory system can be represented as a logic gate. TCS can be viewed in terms of its inputs and outputs without details of internal workings. Thus, it allows us to implement the role of PhoU implicitly. It is more precise representation of the signal. Inputs for deactivation of Pho regulon are P_i , ABC, PhoU, PhoR and PhoB. With the help of schematic and logic gate representation of the system, we have built a preliminary model to analyse how the signal for external

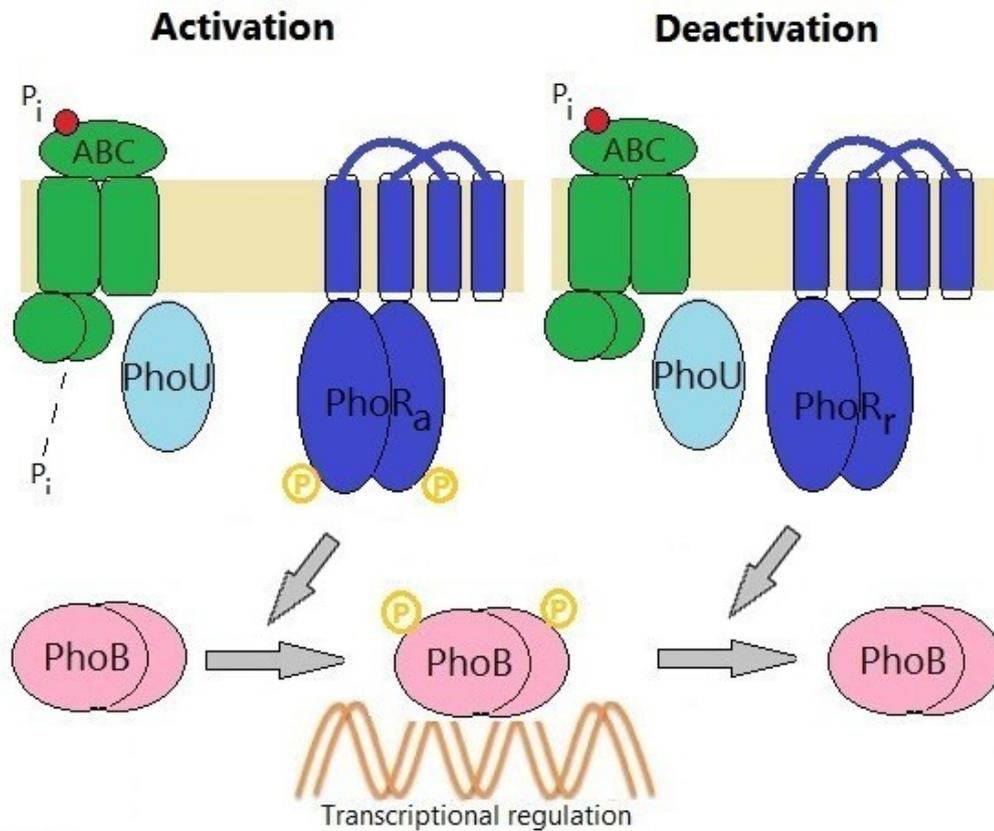


Fig. 3.1 Schematic representation of the different stages of Pho regulon. Green coloured shape represents ABC transporter, dark blue one is PhoR protein, light blue shape stands for PhoU protein and pink stands for PhoB protein. The signalling processes of activation and deactivation correspond to different states of PhoR.

Activation: When external inorganic phosphate (P_i) is depleted, the system is in growth condition. Under this condition, Pho regulon is active and produces its components. ABC transporter binds the external P_i , internalises and releases it to the cytosol. According to the current biological model in the literature, PhoR assesses P_i availability by monitoring the ABC transporter. This is done by relaying the signal via PhoU. When the system is active, PhoU does not stabilise PhoR and PhoR passes to the active state ($PhoR_a$). $PhoR_a$ phosphorylates PhoB. Phosphorylated PhoB then acts as a transcription factor for the operon.

Deactivation: PhoR acts as a phosphatase on phosphorylated PhoB. When environmental P_i is in excess, Pho regulon is inhibited. PhoU stabilises PhoR. This prevents PhoR from autophosphorylating itself. The signal is thus propagated to PhoR, resulting in its inhibition state $PhoR_r$.

changes is delivered to TCS, especially to PhoR. Moreover, logic gate representation allows

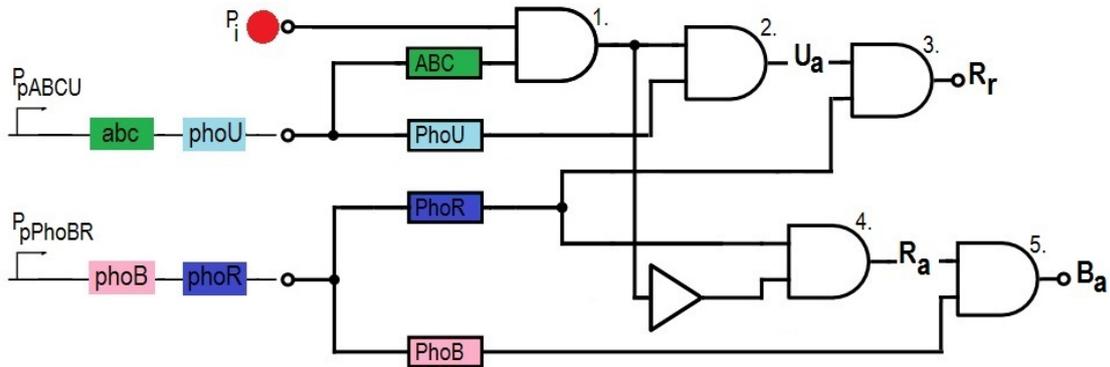


Fig. 3.2 A schematic representation of Pho regulatory system interactions resembling a logic gate circuit. Figure demonstrates the conceptual presentation of TCS signalling processes from an alternative perspective. ABC transporter and PhoU proteins are expressed by the same promoter, that is, p_{ABCU} . Similarly, PhoR and PhoB proteins are expressed by the same promoter, that is, p_{PhoBR} . These are depicted on the left side of the figure. Expressions of ABC transporter, PhoU, PhoR and PhoB proteins refer to system inputs together with external inorganic phosphate P_i . System outputs are repressive PhoR (R_r) and active PhoB (B_a). Colours are used as described in Figure 3.1. The shape resembling an AND gate requires both incoming signals to be present for output signal. The shape resembling a single input NOT gate requires the incoming signal to be absent to produce a signal. (1) P_i is taken in to cytoplasm by the ABC transporter. (2) When the P_i is abundant, ABC transporter relays the signal of external concentration change to PhoU and activates PhoU (U_a). (3) Active PhoU stabilises PhoR and prevents PhoR dimers from autophosphorylating, and PhoR becomes repressive (R_r). PhoR is now inhibited that prevents activation of PhoB. Otherwise, when the P_i is absent, shown by NOT gate, (4) PhoR is free to autophosphorylate itself. This allows PhoR to become active, (R_a) and (5) when PhoB is present, PhoRa phosphorylates transcription factor PhoB and activates it (B_a).

us to observe possible interaction mechanisms. This figure concerns how *E. coli* detects environmental P_i to regulate genes of the Pho regulon by the TCS. It includes the proteins are required for the P_i signaling pathway and their interactions. It is seen that P_i signaling involves different states of the PhoR: an inhibition state, an activation state. The signal transduction from PhoR to gene regulation via TCS has been quantified in our previous work [164]. In the current work, this framework is used as a basis to describe mechanistically the effect of external P_i concentration change and TCS inhibition. The model is based on a mechanistic description of the system dynamics within a chemical reaction system representation with respect to mass action dynamics.

ABC transporter regulates the translocation of P_i to the cytosol. ABC transporter autophosphorylates itself and switches between the closed (ABC) and open (ABC_o) dimer conformations as a result of ATP binding. External P_i (P_{ext}) intake takes place with the conformational changes in the ABC transporter. This causes an increase in the amount of internal P_i (P_{in}) [75, 168].



PhoU is an essential protein for the repression of the Pho regulon at high P_i conditions [56, 168]. Although the mechanism behind the Pho regulon deactivation is unknown, one hypothesis is that PhoU inhibits the activation of PhoR by interacting with PhoR and ABC, converts PhoR to the repressive state (PhoR_r) [56, 139, 152]. It is assumed that the ABC transporter senses the P_i levels and transfers the signal to the TCS via PhoU [56].

Therefore, as an initial step of the model component interactions, we have expressed the inhibition as a direct interaction between ABC and PhoR. We have then constructed two alternative models for assessing interactions between the ABC transporter and TCS with potentially distinct mechanisms for diminished PhoR activities. These two alternative models are abstractions that contain PhoU implicitly, which we refine below. Our motivation behind the use of alternative reaction designs is to examine how the type of the chemical reactions impacts the switch dynamics.

First possible interaction is:



and other one is expressed as:



For these two alternative models, we have implemented the deterministic ordinary differential equation (ODE) systems in Matlab by using the standard translation from the chemical reaction network above, based on stoichiometries. The models have been tested with 0, 50, and 100 μM external P_i . The concentrations of proteins PhoR and ABC are set to 0.22 μM based on literature and verified by our previous work [38, 164]. The parameters for the rates are listed in Table 3.1. The rate of reaction 1 is obtained in accordance with our previous work [164]. The unknown rates of reaction 2 and 3 have been estimated to reproduce decreasing dynamics of the PhoR according to the external P_i concentration change with respect to the

switching of dynamics. The rate of reverse reaction 4 has been set to the value that delivers the expected dynamics. We have used these values as a calibration for the control models and explored the possible variations as described below.

Reaction N.	Rate Symbol	Value	Unit
1	r1	25.3658	sec^{-1}
2	r2	0.001	$\mu M^{-1} sec^{-1}$
3a and 3b	r3a and r3b	1	$\mu M^{-1} sec^{-1}$
4a and 4b	r4a and r4b	0.000001	sec^{-1}

Table 3.1 The reaction rates and units of the control model.

3.3 Results and Discussion

Figure 3.3 shows a dynamic representation of the two different models used in our analysis. Model 1 includes reactions 1, 2, 3a and 4a and Model 2 includes the reactions 1, 2, 3b and 4b. The model describe two possible mechanisms for switching off the TCS in response to the P_i depletion. The deterministic ODE simulations display the average dynamic behaviour of the concentrations for the simulated 4.5 hours. In the two models, the parameter values of all the processes are the same. This guarantees that the differences in dynamics are only due to the types of the interactions.

The ABC transporter can act as an enzyme (Model 1) or binds the histidine kinase PhoR (Model 2) in response to an external signal. Both forms of ABC transporter mechanisms affect the TCS with similar affinities. Both models examine the functionality of the system without going into the details of its internal structures with respect to PhoU. When ABC acts as an enzyme as in Model 1, it indirectly affects PhoR and inhibits PhoR activation while being preserved. In the case of Model 2, ABC participates directly and binds the unphosphorylated PhoR to form a complex ABC_PhoR. This complex formation causes the stabilisation of PhoR. The stabilised PhoR does not modulate the phosphorylation and activity of relevant proteins. The designed interaction of the regulatory system functions as expected in response to various external P_i levels.

The plots clearly show that when the external P_i level increases, activity of PhoR decreases in both models (Figure 3.3). Moreover, the system is able to maintain PhoR activity, thus protein synthesis to maximal when the external P_i is 0 μM . Figure 3.3 provides a comparison of the functional effectiveness of the two mechanisms: Model 1 has faster response time than Model 2 for all the external P_i levels (Figure 3.3). Response time is the time to reach halfway

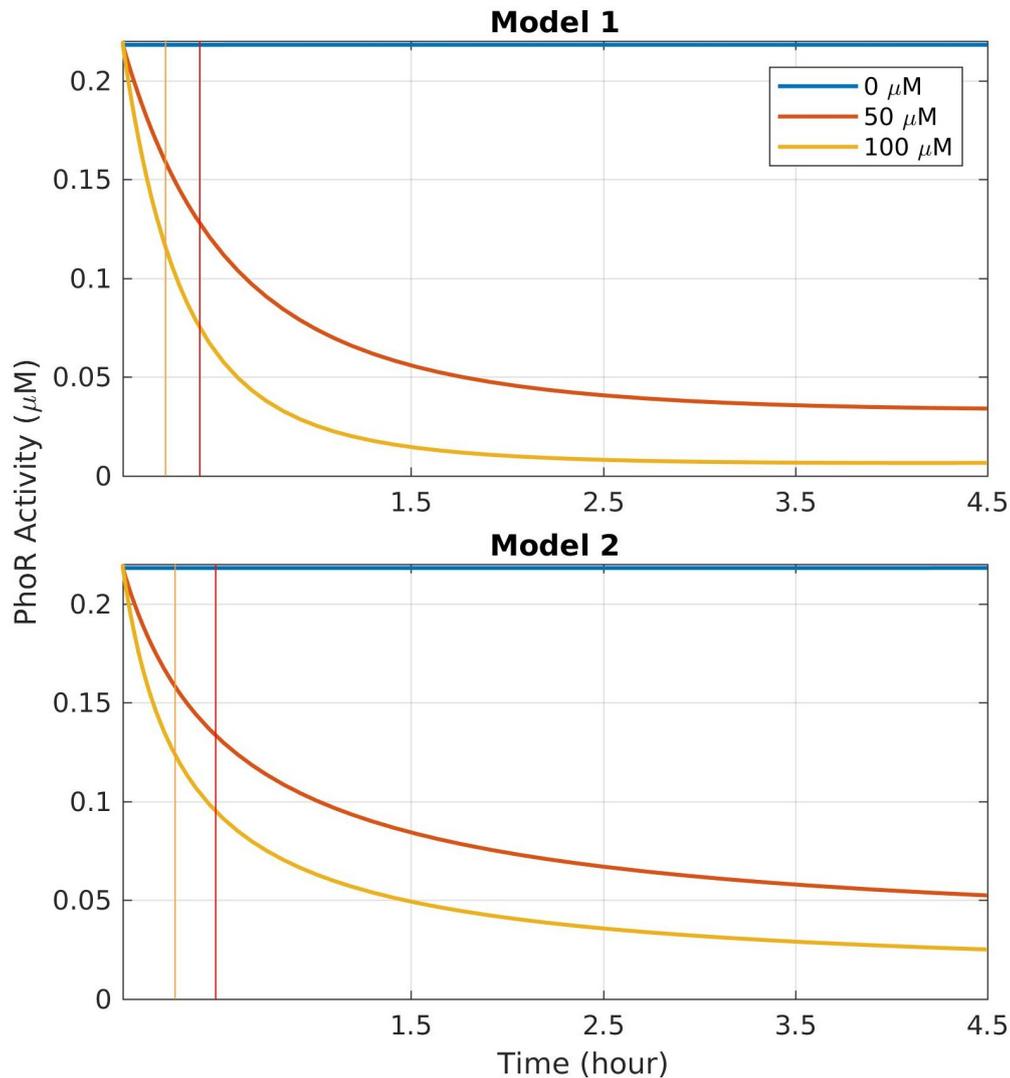


Fig. 3.3 The response of the model to variations in the external P_i levels. Model 1 includes the reactions 1, 2, 3a, and 4a and Model 2 includes 1, 2, 3b, and 4b. The simulations are performed for the external P_i concentrations of 0, 50, and 100 μM and a time-course of 4.5 hours is considered. PhoR activity is plotted in blue when P_i concentrations is 0 μM , it is represented with red when external P_i concentrations is 50 μM , and orange colour is used when P_i concentrations is 100 μM . Response times are represented with vertical lines. Red line is the response time when the external P_i concentration is 50 μM and the orange one is for the external P_i concentration of 100 μM .

between the initial and final levels in the dynamic process [2]. It is an important measure for the speed at which PhoR levels change as it quantifies the effectiveness, especially in fluctuating environmental conditions.

To better understand the TCS switching dynamics, we have performed a mathematically controlled comparison of the two models [3]. We have set the system as equivalent as possible by reaction rates, external conditions and initial values. This allows us to observe the differences in the physiological behaviour between designs. The comparison analysis was done using the specific set of parameter values reported in Table 3.1. By exploring the effect of the changes in the reaction rates r_1 , r_2 , and r_{3a} and r_{3b} in orders of magnitude, we have classified the similarity in dynamics with possible variations in the rates.

Next, we have examined the response time and PhoR degradation for alternative models with the selected parameters. We have extended our calculations and reproduced the dynamics in Figure 3.4. Steady-state levels of PhoR and response time have been cross-examined at reaction rates r_1 , r_2 , and r_3 , allowing evaluation of the stability of the system with different rates. We have omitted the reaction rate r_4 variation because variations on this rate at this regime imply a departure from the expected dynamics. However, variations on r_4 together with others preserve the dynamics, as we explore below.

Robust adaptation is a biological behaviour for *E. coli* that describes surviving in varying environmental conditions. We have performed an analysis to predict the robustness of network against variations of its parameters. We have observed that the model preserves the expected dynamics under perturbations of system parameters. That is, changes on rates result in robust adaptation of Pho regulon system. With variations on rates in Model 1, the switch from PhoR to repressed state occurs more quickly than in Model 2 (see Figure 3.4). As in Figure 3.4, a scaling of the system rates leads to a time scaling of the output response. The steady-state levels are lower in the Model 2 system. Moreover, Figure 3.4 demonstrates the role of the conformational change of ABC on the switch off dynamics. Lower autophosphorylation rate of ABC results in higher levels of PhoR repression.

The response of the Pho regulon has to be rapid enough for TCS to be switched off as the P_i level rises. In the models under comparison, the main difference is in the interaction of ABC and PhoR. In this situation, the response time of Model 1 is faster than Model 2 (see Figure 3.4). This comparison is thus relevant for understanding the differences in the dynamic behaviours that are intrinsic to the differences in design. The present analysis of the TCS switch off system thus provides a refined estimation of the phosphatase activity of PhoR for these models. The analysis here on Model 1 and 2 is used below to discuss the PhoU mechanism. Moreover, these models can be extended and refined to analyse the switch off mechanism of the other TCSs in *E. coli*.

Figure 3.5 and Figure 3.6 display the results of the analysis, whereby we have scanned the parameter values for r_{4a} and r_{4b} in orders of magnitude. We have increased the value for the reaction rate, r_{4a} and r_{4b} while keeping the reaction rates r_{3a} and r_{3b} the same. We

M1		Response Time (sec)					
		r1		r2		r3a	
OM	P _{ext}	50 μ M	100 μ M	50 μ M	100 μ M	50 μ M	100 μ M
2		0	0	18	12	0,05	0,05
1		2608	2608	151	85	136	67
0		1442	800	1442	800	1442	800
-1		164,5	80	18450	7912	2264	2264
-2		5,2	4,9	N.A.	N.A.	0	0
-3		0	0	N.A.	N.A.	0	0
-4		0	0	N.A.	N.A.	0	0
-5		0	0	N.A.	N.A.	0	0
-6		0	0	N.A.	N.A.	0	0

M1		PhoR Level Change					
		r1		r2		r3a	
OM	P _{ext}	50 μ M	100 μ M	50 μ M	100 μ M	50 μ M	100 μ M
2		0	0	1	1,01	1,183	1,035
1		0,204	0,329	1,011	0,99	1,183	1,035
0		1	1	1	1	1	1
-1		1,18	1,03	0,914	0,965	0,215	0,329
-2		1,183	1,035	N.A.	N.A.	0	0
-3		1,183	1,035	N.A.	N.A.	0	0
-4		1,183	1,035	N.A.	N.A.	0	0
-5		1,183	1,035	N.A.	N.A.	0	0
-6		1,183	1,035	N.A.	N.A.	0	0

M2		Response Time (sec)					
		r1		r2		r3b	
OM	P _{ext}	50 μ M	100 μ M	50 μ M	100 μ M	50 μ M	100 μ M
2		0	0	30	16	9,8	4,4
1		3100	2370	255	124	215	123
0		1740	980	1740	980	1740	980
-1		214	114	27500	13000	3030	2510
-2		16,7	11,5	N.A.	N.A.	0	0
-3		4,8	4,8	N.A.	N.A.	0	0
-4		4	4	N.A.	N.A.	0	0
-5		0	0	N.A.	N.A.	0	0
-6		0	0	N.A.	N.A.	0	0

M2		PhoR Level Change					
		r1		r2		r3b	
OM	P _{ext}	50 μ M	100 μ M	50 μ M	100 μ M	50 μ M	100 μ M
2		0	0	1	1,06	1,257	1,1
1		0,229	0,35	1,063	1,06	1,234	1,09
0		1	1	1	1	1	1
-1		1,25	1,096	0,857	0,9	0,229	0,35
-2		1,257	1,1	N.A.	N.A.	0	0
-3		1,257	1,1	N.A.	N.A.	0	0
-4		1,257	1,1	N.A.	N.A.	0	0
-5		1,257	1,1	N.A.	N.A.	0	0
-6		1,257	1,1	N.A.	N.A.	0	0

Fig. 3.4 Heatmap displaying the effects of the changes in reaction rates r_1 , r_2 , r_{3a} and r_{3b} . Each parameter is varied by orders of magnitude (OM) higher and lower. The one on the left is for response time and on the right is for PhoR level change. They are computed when the system reaches the steady-state. For the cases that the system does not have a steady state, response time and PhoR level change are displayed as N.A.. Response time, in seconds, is calculated as the time needed to reach halfway between the initial and final level. The outcome of PhoR level change is normalised with respect to the control model. The analysis is done for Model 1 (M1) and Model 2 (M2) when external P_i (P_{ext}) are 50 μ M and 100 μ M. Red and green represent the decreasing and increasing effects, respectively. Reaction rate r_4 variation is excluded because variations on it imply a departure from the expected system dynamics.

have scanned r_1 and r_2 values by orders of magnitude to observe the variations in system dynamics in Model 1 and 2. We have calculated for each set of reaction rate values the functional effectiveness with respect to PhoR level change and the response time. We have found that if the chosen reaction rate 2 is reduced, when the rate of reaction 4 is increased, then robust adaptation of dynamics is preserved. Moreover, when r_4 is increased, the system has slower response time and lower PhoR level change.

We have analysed the evolution of the *E. coli* in response to different external P_i concentrations. The analysis on two different models with variations in model parameters demonstrate the gain of the system under a broad spectrum of circumstances, represented by these parameters. We have observed that Model 1 has faster response time in all regimes. With the assumption that faster response is more favourable, these results suggest that the enzymatic interaction in Model 1 can be a prevalent regulatory pattern in biological signalling pathways. Two-component-signaling systems constitute the most common signaling pathways in bacteria [6]. The complete genome sequences of several bacteria have determined that the TCS exists in many bacteria and controls many processes such as nutrient intake, respiration, colony growth, drug resistance, etc. Moreover, extensive genome analysis has shown that the majority of the proteins identified in *E. coli* have homologues in other bacteria [6]. Considered together, these characteristics may explain that biochemical information transfer can be establish through enzymatic reactions in different TCS and bacteria.

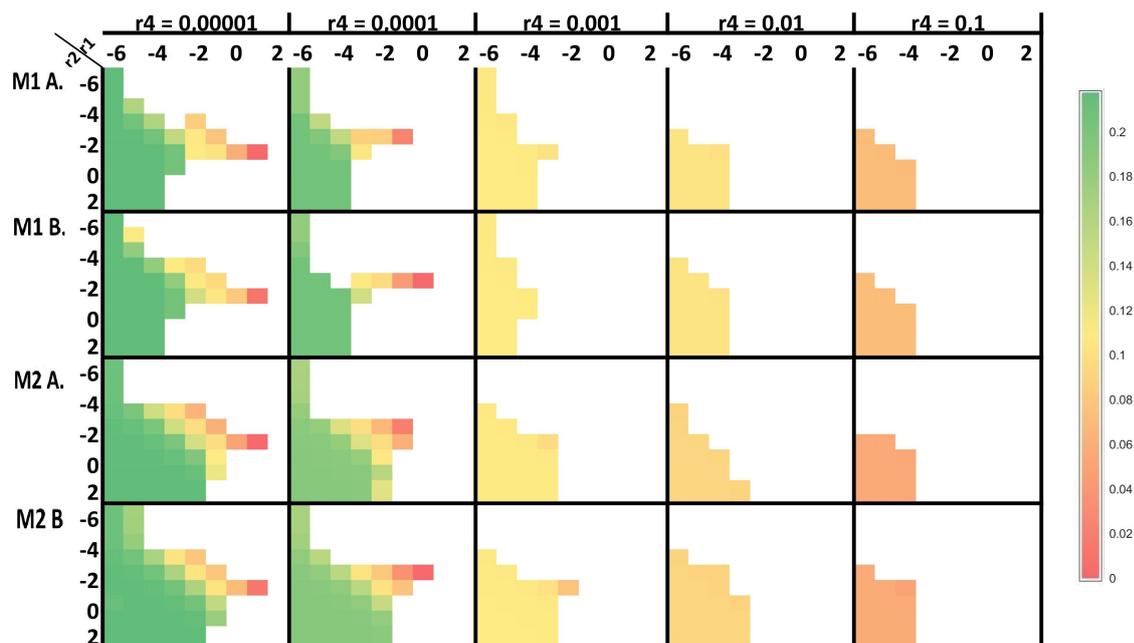


Fig. 3.5 Heatmap displaying the normalised PhoR level change as a result of the variations in r_1 and r_2 together with variations in r_4 rate values. The analysis is performed for Model 1 (M1) and Model 2 (M2) when the external P_i concentrations are $50 \mu\text{M}$ (A) and $100 \mu\text{M}$ (B). Rates r_1 and r_2 are varied by orders of magnitude higher and lower from -6 to 2. PhoR level change from the beginning to the end of the simulation is computed and displayed. The outcome of PhoR level change is normalised with respect to the control model.

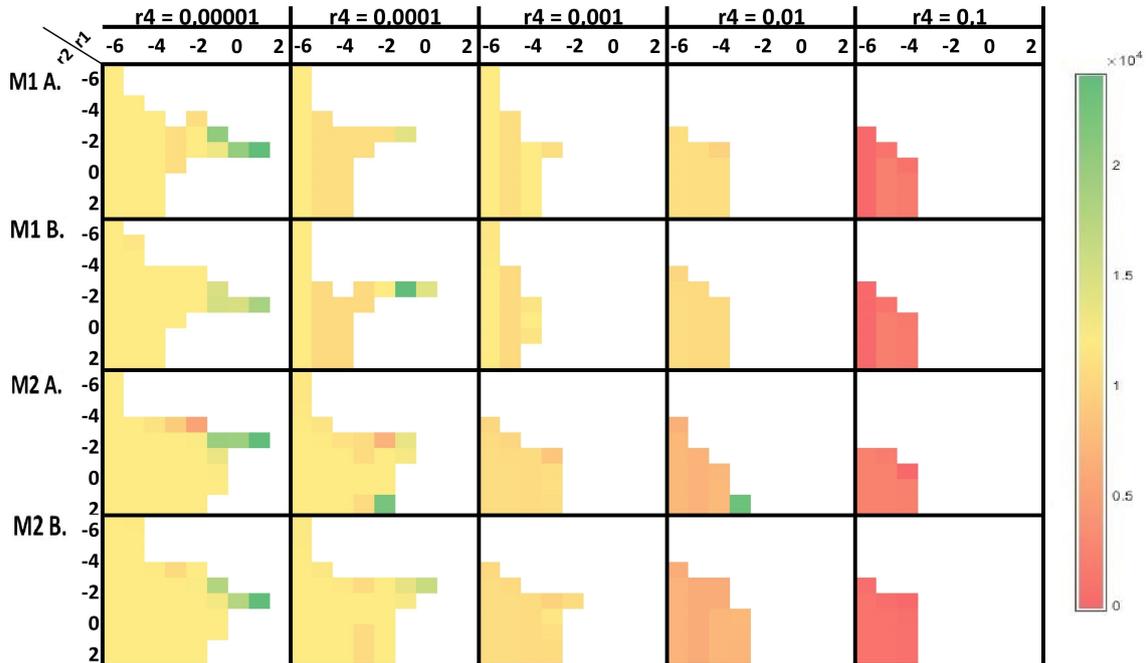


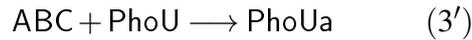
Fig. 3.6 Heatmap displaying the response time, in seconds, as a result of the variations in r_1 and r_2 together with variations in r_4 rate values. The analysis is performed for Model 1 (M1) and Model 2 (M2) when the external P_i concentrations are $50 \mu\text{M}$ (A) and $100 \mu\text{M}$ (B). Rates r_1 and r_2 are varied by orders of magnitude higher and lower from -6 to 2.

Pho regulon expresses many genes that are influenced by the environmental P_i level, and their expression is regulated by TCSs. TCS is thus a predominant form of signal transduction in *E. coli*.

Our results provide a quantitative description of how different proteins interact when the TCS is switched off as a result of the changes in external P_i concentration. The effect of the switch off mechanism is measured in terms of PhoR level change. Our results do not only provide predictions on the physiology of the Pho regulon, but also demonstrate how a strong signal due to phosphatase activity can cause a fast switch off response. The comparison of the effect of possible interaction types in our models becomes instrumental for understanding the differences in behaviour of biological circuits created by using synthetic biology techniques. Moreover, this work provides a framework to quantitatively analyse the interactions of PhoU with other system compartments, for example, as in [164].

To elucidate the mechanistic function of PhoU, we refine our models with explicit mechanism of PhoU action within the alternative models discussed above. In these refined

models, reactions 1 and 2 are kept as same as these reactions model the ABC transporter, which regulates the translocation of P_i to the cytosol. We have then considered the interactions of PhoU in the system in accordance with the two models above. As a result of this, we have replaced reactions 3 and 4 with their alternatives that integrate PhoU mechanistically. In this refined model, PhoU interacts with ABC, gets the external P_i concentration change signal, and then becomes active (PhoUa).



After passing to the active state, PhoUa transmits the signal to PhoR, inhibits PhoR activity, and causes the repression of PhoR.

Here, in accordance with the two models above, we have provided alternative reactions to express the affect of PhoU on the system. That is, PhoU can affect PhoR in similar ways as in Model 1 and Model 2.

We first consider the possibility that PhoUa can act like an enzyme, drawing parallels with Model 1.



Alternatively, we have considered the interaction as in Model 2. In this model, PhoU binds to the histidine kinase PhoR in a reversible manner.



Figure 3.7 displays a dynamic representation of these two models with PhoU, which are mathematically described by using the same procedures. PhoU Model 1 includes reactions 1, 2, 3', 4', 5a and 6a and PhoU Model 2 includes the reactions 1, 2, 3', 4', 5b, and 6b. In order to observe the differences that originate from the inclusion of an explicit PhoU mechanism, we have kept the parameter values as before. Therefore, the initial values are kept the same and PhoU initial value is set to 0.22 μM as in [38]. The rates of the reactions 1 and 2 are also set to the same values. Reactions 5a, 6a and 5b, 6b are assigned the values of the reactions 3a, 4a and 3b, 4b.

The plots in Figure 3.7 clearly display the robust adaptation of model with PhoU protein. When the external P_i level increases, PhoR concentration decreases in both models. Moreover, we have observed that PhoU Model 1 has a faster response time than PhoU Model 2 for different external P_i as shown in Figure 3.7.

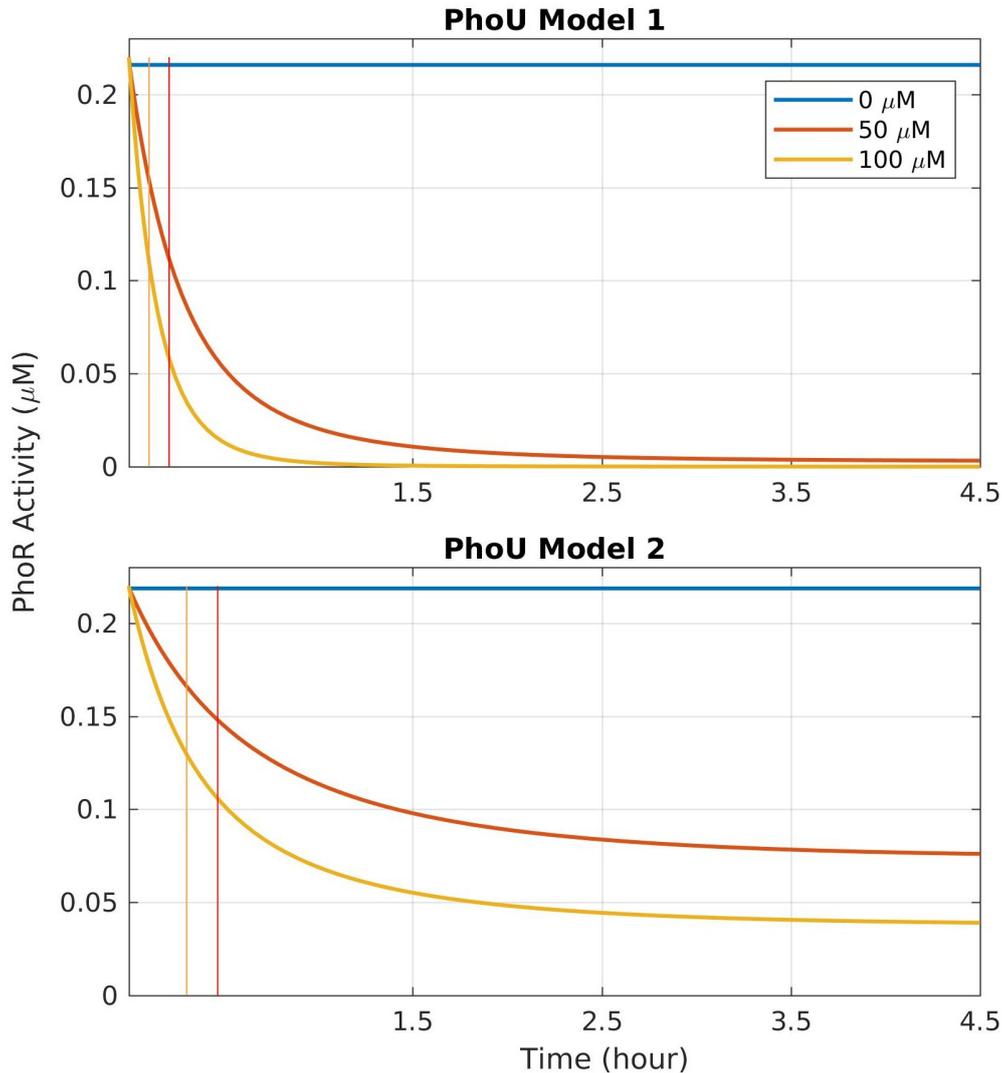


Fig. 3.7 The response of the model with PhoU to variations in the external P_i levels. Model 1 includes the reactions 1, 2, 3', 4', 5a and 6a, and Model 2 includes 1, 2, 3', 4', 5b and 6b. In all the experiments, the external P_i concentrations are 0, 50, and 100 μM and a time-course of 4.5 hours is considered. PhoR activity is plotted in blue when P_i concentrations is 0 μM , it is represented with red when external P_i concentrations is 50 μM , and orange colour is used when P_i concentrations is 100 μM . Response times are represented with vertical lines. The red line is the response time of dynamics when external P_i concentration is 50 μM and the orange line is for an external P_i concentration of 100 μM .

These results provide an analysis of mechanistic interactions of the PhoU, PhoR, and ABC transporter proteins. The analysis of the results above provides a framework for better understanding P_i signal transduction in *E. coli*. According to our signalling complex model, PhoR is able to sense through PhoU the conformational states of ABC transporter as a

consequence of P_i transport and then modulate its kinase/phosphatase in accordance with the appropriate response. This model can thus explain the mechanism that provides the adequate response time.

3.4 Conclusion

We have previously reported a quantitative description of the activation of a Pho regulon [164]. Here we have developed a model for understanding the role of ABC transporter and PhoU by an examination of the switch off dynamics. Our model provides descriptions of the possible interaction mechanisms between TCS and ABC transporter and predicts the mechanistic behaviour for different cases under the conditions of varying external Pi concentrations. Understanding the behaviours of TCS requires characterisation of not only the pathway organisation, but also the dynamic rates of individual activities in the cell. In addition, by providing testable predictions for the wet-lab, the understanding produced from this approach provides insights for future engineering of such biological processes.

Quantitative wet-lab analysis can be used to determine the system structure. Experimental data to estimate the amount of the protein in the cell helps us to quantify model results. The concentrations of system components such as PhoU and PhoR proteins can help to validate the results of model. Moreover, not only the concentrations of the proteins but also the structure, interactions and functions of all proteins within *E. coli* can be elucidated by experimental design. It is known that protein production and purification tools are designed for identifying which factors are important in the biological system. They can also help to find an optimum process in the mechanism. One strategy to determine protein function can be to identify the protein–protein interactions such as interaction of transcription factor PhoB with PhoR. Examining the phosphorylation of PhoR and PhoB proteins can be used as a proxy for system formalization. Moreover, artificial overexpression of proteins can be used to identify proteins involved in biological processes and examine their functions.

The model should also allow us to better understand the mechanisms for tuning the regulatory system to be sharper and more rapid. For example variations in the ABC phosphorylation rate can work as an inhibitor for the regulation of PhoR due to the changes in the external P_i concentration.

By applying such models to living systems and processes we begin to see mechanistically how an environmental condition produces a change of state in an organism both on the protein function level and gene expression level. It is through such an understanding that the meaning of adaptation and embeddedness in living cells becomes apparent and quantifiable.

Author contributions

This work is supported by COSBI and by the European Union's Horizon 2020 research and innovation programme under the grant agreement No 686585 - LIAR, Living Architecture. Ozan Kahramanođulları and I designed and performed simulations and sensitivity analysis. All figures have been produced under the supervision of Ozan Kahramanođulları. Ozan Kahramanođulları and I wrote and edited the manuscript. Martin M. Hanczyc supervised the work edited the manuscript.

Chapter 4

Quantifying Dynamic Mechanisms of Auto-regulation in *Escherichia coli* with Synthetic Promoters in Response to Varying External Phosphate Levels

Escherichia coli (*E. coli*) have developed one of the most efficient regulatory response mechanisms to phosphate starvation. The machinery involves a cascade with a two-component regulatory system (TCS) that relays the external signal to the genetic circuit, resulting in a response with feedback to the network. Achieving a quantitative mechanistic understanding of this system has implications in synthetic biology and biotechnology, for example, in applications for wastewater treatment. To this aim, we present a computational model with a detailed description of the TCS, consisting of PhoR and PhoB, together with the regulatory promoters and the mechanisms of gene expression. The model is parameterised according to the feasible range given in the literature, and fitted to the dynamic response of our experimental data on PhoB as well as PhoA, the product of this network that is used in alkaline phosphatase production. Deterministic and stochastic simulations with our model predict the regulation dynamics in higher external phosphate concentrations while reproducing the dynamics observed in the experimental conditions of phosphate starvation. In a cycle of simulations and experimental verification, our model predicts and explores phenotypes with various synthetic promoter designs that can optimise the inorganic phosphate intake in *E. coli*. Sensitivity analysis demonstrates that the expression of Pho-controlled genes has a significant influence over the phosphate response. Together with experimental findings, our model should thus provide insights for the investigations on engineering new sensors and regulators for living technologies.

What follows is the extended and improved work content of the article, published in the Conference ECAL in September 2017 [164].

4.1 Introduction

Mechanisms of inorganic phosphate intake within the context of cellular processes is a topic of extensive research effort, also due to potential applications in enhanced biological phosphorus removal (EBPR) from wastewater. Phosphorous, which is one of the major causes of water quality problems, occurs in wastewater almost solely in the form of phosphates such as inorganic phosphate (P_i). Microorganisms, which are key players in bioremediation, have potential to treat large amounts of the pollutants and hold promise for renewable sources [115]. In particular, the bacteria *Escherichia coli* can take and store excess P_i in the form of polyphosphates [168]. An in-depth understanding of the mechanisms controlling such processes should pave the way for metabolic engineering to lead to improvements in wastewater treatment. Phosphorous is one of the essential elements for all life forms, since it is required for many functional regulation mechanisms in the cell. P_i is the only form of phosphorus that can be directly utilized by cells [168]. Thus, microorganisms have developed mechanisms to assimilate inorganic phosphate-containing compounds. Bacterial two-component systems (TCS) are key signal transduction networks regulating global responses to environmental change.

The physiological characteristics of P_i transport in *E. coli* have been extensively studied [2, 12, 38, 54, 168]. Nonetheless, the regulatory interactions that control the P_i transport are complex, and they involve two major phosphate transport systems, depending on external P_i levels. The low affinity phosphate inorganic transport (Pit) system depends on the proton motive force; it is a coupled transporter of two different ions through the membrane [68, 168]. The phosphate specific transport (Pst) system, on the other hand, is P_i -repressible, and it is induced when the external P_i concentration is depleted [21, 81, 169].

Mechanistically, P_i signalling is associated with the Pst system proteins. These are PstS, PstC, PstA, PstB, which constitute an ABC transporter, and are abbreviated as PstSCAB. In *E. coli*, Pst is a negative process, that is, excessive P_i is required for turning the system off. Activation is the default state and occurs in conditions of P_i limitation [56, 168, 169]. Signal transduction by environmental P_i requires seven proteins (PhoR, PhoB, PstSCAB, PhoU), which are thought to interact in a membrane associated signalling complex.

Figure 4.1 displays a schematic representation of the system in the starvation condition. When P_i is limited in quantity outside the cell, PstS binds to the P_i following its diffusion to the cell surface [95, 168, 169]. The transmembrane domain of the ABC transporter, that

is, PstC and PstA are integral membrane proteins that span the entirety of the membrane. They regulate the translocation of P_i from PstS to the inner membrane [38, 95, 168, 169]. P_i intake happens with the conformational changes in the PstB as a result of ATP binding, also known as ATP-switch model. Principal conformations of the PstB are as follows: (i.) the formation of a closed dimer upon binding two ATP molecules; (ii.) dissociation to an open dimer facilitated by ATP hydrolysis. The switching between the open and closed dimer conformations induces conformational changes in the TMD resulting in substrate translocation of P_i transport from inner membrane to the cytosol. This way, the ABC transporter provides the required increase in the amount of phosphate in the cell [168, 169].

The low activity of the ABC transporter PstSCAB causes, in a number of steps, the autophosphorylation of the sensor kinase PhoR, which relays the signal to the transcription factor PhoB. The current evidence suggests that PhoR and eventually PhoB assess P_i availability by monitoring the activity of Pst transporter via PhoU [56, 95]. In mechanistic terms, when there is sufficient P_i flux, PhoU stabilises PhoR. The resulting stable conformation prevents PhoR from auto-phosphorylation. This, in return, prevents the down-stream signalling that would otherwise result in the expression of the genes that feedback to the Pst system for further P_i intake. In fact, when PhoU is deleted, PstB does not only continue to spend ATP and transport P_i , but PhoR acts as a constitutive PhoB kinase, leading to high expression of the Pho regulon genes [168, 169], and thereby, activating the expression of the Pst system.

All experimental evidence indicates that Pho regulon is controlled by external phosphate limitation rather than internal [38]. When the surrounding environment has abundant phosphate, *E. coli* uses as few resources as possible to facilitate the phosphate intake. However, when P_i becomes low outside the cell, it turns into a growth limiting factor and the cell spends energy to up-regulate the expression of target genes that are used to acquire phosphate. Previous studies have shown that the TCS plays a central role in sensing the P_i level in the environment and regulating the expression of genes that are directly involved in the intake [95, 168]. Here, we investigate the relationship between P_i -starvation, the TCS signalling, and promoter activity by using a combination of wet-lab experiments and dynamic modelling. Our model, describing the phosphate starvation response at the genetic level, is composed of a set of ordinary differential equations and the corresponding stochastic model that are derived from a chemical reaction network. Here we suggest a mechanism that can provide a quantitative description of how different processes interact to form a positively-regulated biological control system. The mechanism is based on a certain *E. coli* phosphate signaling system. It explains how specific proteins can work together to provide gene expression and increase the P_i intake. Moreover, it suggests an approach that can help to identify a wide family of promoter mechanisms that potentially have synthetic applications.

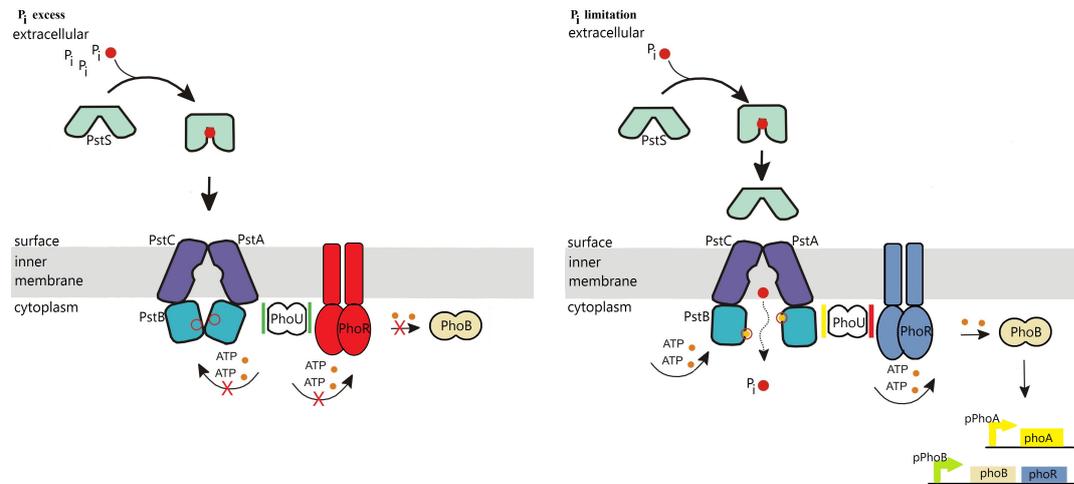


Fig. 4.1 The transmembrane signal transduction due to external P_i levels and the Pho regulon activity when extracellular inorganic phosphate (P_i) is in excess (left) and when it is depleted (right). The ABC transporter consists of the extracellular domain PstS, transmembrane domains PstA and PstC, and intracellular domains PstB (see the main text for the description). *Left*: when external P_i is in excess, the ABC transporter activity is limited as the cell does not consume ATP for the P_i transport. According to the current model in the literature, PhoR assesses P_i availability by monitoring the activity of Pst transporter, and relays the signal from PstB via PhoU to PhoR. When the P_i influx is increased, PhoU stabilises PhoR, which is depicted with the green bars around PhoU. This prevents PhoR dimers from autophosphorylating (red). Consequently, the transcription factor PhoB does not become phosphorylated by PhoR. *Right*: Due to the ABC transporter activity, the external P_i binds to the PstS component of the ABC transporter. It is then translocated to the inner membrane domain of the transporter through PstC and PstA. Following this, PstB changes its conformation by consuming ATP. P_i is internalised and released to the cytosol. When the P_i influx through the ABC transporter reduces, PhoU does not stabilise PhoR, which is depicted with the red bar next to PhoU. As a result of this, PhoR becomes free to perform its auto-kinase-phosphotransferase activity, whereby it phosphorylates PhoB. Phosphorylated PhoB then forms a dimer to act as a transcription factor for the operons, resulting in PhoA, PhoB and PhoR expression.

E. coli relies on accurate control of TCS signaling systems to adapt to environmental perturbations. TCS regulates physiological processes in response to environmental parameters and enable adaptation to changing conditions. To understand the signal transduction mechanism in bacteria, TCS has been studied widely. Therefore, the model includes TCS members and activation of the Pho regulon promoters pPhoB and pPhoA. In recent years significant advances have been made in the understanding of the role of TCS and signalling mechanisms. More than 180 different TCS and gene sequence of more than one thousand

sensors and response regulators have been identified in different bacteria [6]. In this work, we have described the first steps in the design of a synthetic biological system based on the use of TCS. This should constitute a contribution to the research on genetically modified bacteria that detect environmental changes and respond to higher inorganic phosphate levels.

One of our objectives is to construct a responsive model structure which can be used to predict the behaviour of engineered bacteria as biological sensors that detect changes in their environment and respond by increased synthesis and intake. An application of this would be to reduce the need for large-scale reaction networks and may be the basis for a new generation of bio-based materials. A quantitative model is presented, which integrates experimental data from a promoter activity within *E. coli*. We have presented the initial steps towards this application domain in the form of a design process, which integrates gene expression data from *E. coli* into a computer model. In order to understand the quantitative analyses of protein concentrations, modifications and activities in environment, we have focused on the activation response of the TCS components PhoB and PhoR. Moreover, this model can help us to understand the dynamic behavior of system deactivation and to quantitatively evaluate the role of phosphatase activity under abundant external P_i conditions as in Chapter 3.

In this chapter, we have developed an approach combining experimentation and modelling. Experimental data is used to fit the parameters to the feasible physiological range given in the literature, and to determine the relative sensitivity to the parameters. The simulations with our model provide a dynamic description of the mechanisms. With a combination of wet-lab experiments and computer simulations, we use our model to quantify dynamic mechanisms of auto-regulation in *E. coli* in response to varying external phosphate levels, and explore and verify emerging phenotypes with synthetic promoters. Simulations with the model do not only reproduce our experimental measurements, but also predict phenotypes with various synthetic promoter designs that can optimise the P_i intake in *E. coli*. Sensitivity analysis on the parameters demonstrate the influence on the expression of Pho-controlled genes and the gain of the system under variations in transcription efficiency in response to external phosphate concentration. The model can thus serve as a virtual lab, and can be used to test various promoter designs for enhanced P_i intake in biotechnology applications for phosphate sequestration.

4.2 Material and Methods

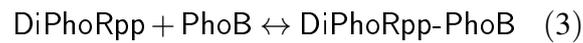
The model is based on a mechanistic description of the system dynamics within a chemical reaction system representation with respect to mass action kinetics. To build the simulation model, we have selected the part of the network from the TCS to gene regulation, and used

the promoter activity as well as the levels of active transcription factor as indicators for quantifying the response to external P_i levels. This is because the TCS proteins PhoR and PhoB are encoded by the same operon [168]. Moreover, the ABC transporter proteins and PhoU are encoded by the same operon [168]. The ABC transporter expression and the TCS expression are regulated by the same transcription factor PhoB [38]. The expression of these proteins depends on the external phosphate concentration. The Pst phosphate uptake system regulates gene expression in response to P_i availability by interacting with the two-component system.

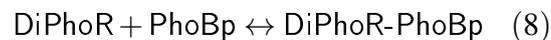
When the external P_i concentration is limited, resulting in a low activity of the ABC transporter, PhoU releases the otherwise inhibited PhoR. This allows PhoR to phosphorylate PhoB through an autokinase/phosphotransferase activity [168, 169]. PhoR is stable as a dimer, thus it is doubly phosphorylated.



PhoR is essential for the control of the activity of PhoB [52, 168, 169]. After autophosphorylating, PhoR relays the signal to the transcription factor PhoB. The bidirectional reactions 3 and 5 below model the association of phosphorylated PhoR dimer and PhoB, and the unidirectional reactions 4 and 6 model the phosphotransferase. PhoB has been reported to exist primarily as monomers and phosphorylation greatly enhances dimerisation, modelled by reaction 7 [81].



In *E. coli*, the sensor histidine kinase PhoR is a bifunctional enzyme that paradoxically performs two opposed tasks: in one direction, it catalyzes the phosphorylation of response regulator PhoB, and in the other, it also performs the dephosphorylation of PhoBp [53, 149]. The association of PhoR dimers with phosphorylated PhoB is modelled by the bidirectional reaction 8, whereas the phosphatase activity is given by reaction 9.

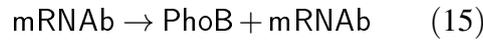
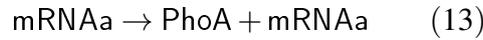


Phosphorylated dimer structure PhoB (DiPhoBpp) is enabled for activating Pho regulon by binding to a consensus promoter region. PhoB and PhoR in Pho regulon are encoded by the same operon, that is, the *phoBR* operon. Thus, the synthesis of the regulatory proteins PhoB and PhoR is under Pho regulon control [168–170].

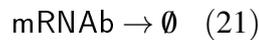
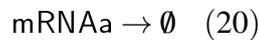
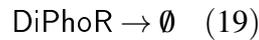
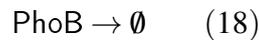
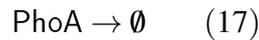
Based on experimental data, we consider the PhoA and PhoB promoters, whereby the PhoB promoter provides feedback to the system as this results in the expression of both PhoB and PhoR. PhoR expression during P_i limitation is dependent on the upstream PhoB promoter; the operon structure indicates that PhoR gene function requires expression from the PhoB promoter [170]. For this, phosphorylated PhoB dimers bind to the promoter as active transcription factors.



Active promoters pPhoAa and pPhoBa lead to the transcription of mRNA, which carry the information for the subsequent translation, resulting in protein synthesis [168, 169].



With the inclusion of the degradation/dilution terms, we obtain:



We have implemented the deterministic ordinary differential equations (ODE) model in Matlab by using the standard translation from the chemical reaction network above, based on stoichiometries. The ODEs are listed in Figure B.1. We have also implemented a version of the model for stochastic simulation, again based on mass action kinetics. The stochastic simulations capture fluctuations due to small molecule numbers, which are not captured by the deterministic simulations. By using the standard conversion factors for mass action kinetics, we could use the same rates for the deterministic and stochastic simulations.

The initial concentrations of the model variables have been derived from the literature or obtained from our experiments, described below. The control model has been calibrated for an initial culture containing $0 \mu\text{M}$ external P_i . Prior to P_i starvation, the concentrations of proteins PhoR and PhoB are approximately $0.22 \mu\text{M}$. The concentrations of active PhoR and active PhoB are $4 \cdot 10^{-8} \mu\text{M}$ and $6 \cdot 10^{-8} \mu\text{M}$, as determined by Keasling et al. [38]. With a single plasmid, average mRNA number is 2-3 in *E. coli* [129]. Therefore, the initial states of mRNA_a and mRNA_b are set to $0.00166 \mu\text{M}$ by taking *E. coli* volume as $1 \mu\text{m}^3$, and the initial promoters numbers are set to 10 for each. We assume that the ATP concentration stays constant throughout the considered time intervals.

The rates of chemical reactions are obtained in accordance with the variability of physiological ranges given in the literature [2, 12, 129]. The model includes 29 reactions, including the reverse reactions. The control model has been used to reproduce the data and the unknown parameters have been estimated by least square inference within the plausible physiological ranges. When possible, parameter values are fixed or estimated by using experimental measurements found in the literature. The parameter values taken from the literature and their physiological ranges for the rates, if applicable, are listed in Table 4.1. The parameters without a range are fitted to the experimental data by using the deterministic model to reproduce the response curves. The parameter estimation procedure has been carried out by using a multi-start approach. The rate values have been selected with respect to the best fit to the physiological ranges, also listed in Table 4.1, and the dynamics in accordance with the experimental findings in order to avoid discontinuities or states with unrealistic values.

The data for the PhoA and PhoB expression have been obtained using PCR amplified DNA from *E. coli* MG1655 genome and transcriptionally fused to the translational coupler BCD2 [118] and the fluorescent ms-fgfp gene. Subsequently the PphoA::BCD2-msfgfp fragment was cloned using the PacI/HindIII restrictions sites in pSEVA234 plasmid (<http://seva.cnb.csic.es/>), generating the pSEVA237PphoA vector.

The synthetic promoter Pliar00117 was obtained by PCR using a degenerative primer and the promoter pBG42 [178] as template. The Pliar00117 promoter was transcriptionally fused to BCD2 and msfgfp gene. The Pliar00117::BCD2-msfgfp fragment was cloned using the PacI/HindIII restrictions sites in pSEVA234 plasmid, generating the pSEVA237Pliar00117 vector. *E. coli* DH10B was used for cloning and for MsfGFP protein expression under PphoA or Pliar00117 regulation.

Activities of the PphoA and Pliar00117 promoters have been determined as follows. *E. coli* DH10B carrying pSEVA237PphoA or pSEVA237Pliar00117 have been grown over night at 37°C in MOPS medium [121] containing $100 \mu\text{M}$ KH_2PO_4 to OD600 of 2.0. The bacterial cells have been pelleted at $1500 \times g$, room temperature, 10 min and washed twice in

MOPS medium without KH_2PO_4 . The cells have been suspended in 250 μ l of MOPS buffer with increasing concentrations of KH_2PO_4 from 0 to 50 mM. The bacterial cell suspensions have been loaded in 96-well plates and incubated at 37°C, 200 rpm. The expression of MsfGFP was monitored at different times in a Varioskan Flash spectral scanning multimode reader (Thermo Scientific); excitation 488 nm, emission 509 nm.

4.3 Results and Discussion

Phosphate intake at starvation requires rapid activation of PhoB dimers. Figure 4.2 provides a schematic representation of the control model set up with the experimental data and the formal model. The control model is set up with respect to the initial phosphate starvation in accordance with the experimental data and the parametrisation and fitting procedure described in Methods. In first of the two steps, the blue curves in Figure 4.2 C and Supplementary Figure B.2 are obtained by using only the experimental values of PhoA levels. When external P_i is abundant, the Pst system inhibits the activation of TCS, and consequently the P_i intake; PhoA is then expressed at a basal level. However, when P_i is limiting, inhibition of TCS is relieved, resulting in the activation of PhoA transcription. The alterations of PhoA expression can thus be interpreted to an extent as an indicator for the changes in external phosphate level and P_i intake.

In these simulations, in response to external P_i level at 0 μ M, the system initiates the activation of PhoR, given by the autophosphorylation of both of the monomers in the stable dimer. This results in the subsequent transfer of phosphoryl groups. As a consequence, the response regulator PhoB rapidly becomes active and dimerises to form active transcription factors. The resulting rapid increase in the promoter activity delivers the mRNA transcription, and the consequent experimentally observed levels of PhoA.

At a second step in our analysis, to highlight and contrast the role of PhoB dynamics in the feedback mechanism, we have refined the model to include the experimental data on PhoB expression. The resulting red curves in Figure 4.2 C and Supplementary Figure B.2 are obtained by using the experimental levels of both PhoA and PhoB in the fitting procedure. The rates in Table 4.1 are obtained as a result of this fitting procedure that delivered the control model parameters, whereby we have enforced the displayed physiologically boundaries. The difference in the phenotype between red and blue curves in Figure 4.2 C should thus highlight the role of the feedback of the PhoB and PhoR expression to the system.

We have performed a large scale analysis of the system dynamics of the control model in terms of the fitted parameter values within a broader range. We have first analysed the broader effect of the the TCS disassociation rates (r_{3r} , r_{5r} , r_{7r} , r_{8r}) on the dynamics and

Reaction Number	Rate Symbol	Fit Value	Literature Value
1	r1	25.3658	$10 - 100 s^{-1}$
1 reverse	r1r	8.1165	$\ll 10 s^{-1}$
2	r2	25.3658	$10 - 100 s^{-1}$
2 reverse	r2r	8.1165	$\ll 10 s^{-1}$
3	r3	100	$100 \mu M^{-1} s^{-1}$
3 reverse	r3r	94.9411	N.A.
4	r4	21.3718	$17 - 23 s^{-1}$
5	r5	100	$100 \mu M^{-1} s^{-1}$
5 reverse	r5r	94.9411	N.A.
6	r6	21.3718	$17 - 23 s^{-1}$
7	r7	100	$100 \mu M^{-1} s^{-1}$
7 reverse	r7r	94.9411	N.A.
8	r8	100	$100 \mu M^{-1} s^{-1}$
8 reverse	r8r	94.9411	N.A.
9	r9	12.95	$\ll 17 s^{-1}$
10	r10	10000	$10000 \mu M^{-1} s^{-1}$
10 reverse	r10r	1000	N.A.
11	r11	10000	$10000 \mu M^{-1} s^{-1}$
11 reverse	r11r	1000	N.A.
12	r12	0.0540	$0.0025 - 0.2 s^{-1}$
13	r13	0.0302	$0.0006 - 0.05 s^{-1}$
14	r14	0.130	$0.0025 - 0.2 s^{-1}$
15	r15	0.035	$0.0006 - 0.05 s^{-1}$
16	r16	0.0302	$0.0006 - 0.05 s^{-1}$
17	r17	0.0001	$0.000096 - 0.00079 s^{-1}$
18	r18	0.0001	$0.000096 - 0.00079 s^{-1}$
19	r19	0.0001	$0.000096 - 0.00079 s^{-1}$
20	r20	0.0055	$0.0055 s^{-1}$
21	r21	0.0055	$0.0055 s^{-1}$

Table 4.1 Reactions and deterministic rates obtained from the physiological ranges in [2, 12, 129].

the PhoB transcription and translation parameters that deliver the expression of PhoB (r14, r15), which is a transcription factor. The cumulative output of different simulations with fold changes that cover the physiological intervals for r14 and r15 as well as a broad range for r3r, r5r, r7r, and r8r are depicted in Supplementary Figures B.3, B.4, B.5, B.6, B.7, B.8, respectively.

In these simulations, the dissociation rates of PhoR and PhoB, that is, r3r and r5r, have transient effects on the steady state concentrations of the complexes formed by these molecules. However, these rates do not modify the PhoB activity or the levels of PhoA and PhoB as shown in Supplementary Figures B.5, and B.6. Variations in the dissociation rate of PhoB, given by r7r, affect the activity of PhoB as a transcription factor, however this has a minor effect on the transcription and translation of PhoB and none on PhoA (Supplementary Figure B.7). On the other hand, the dissociation rate of the inactive PhoR from the active PhoB, given by r8r, affects the activation of PhoB as a transcription factor in a proportional way, however its effect on promoter activity is negligible (Supplementary Figure B.8). As it can be seen from Table 4.1, there are also other parameters that have an effect on both system dynamics and PhoB activity. However, these are not suitable candidates for modification at this first analysis due to their physiologic ranges, and they are discussed below.

As it can be observed in Figure 4.2, and supported by the analysis above, the simulations that include the PhoB data for parametrisation result in faster response dynamics, measured in terms of the time required to reach a peak state. This can be explained by the self-feeding role of the TCS, and the resulting increased requirement for the active transcription factors to sustain the experimentally observed protein levels due to their feedback to the network: because PhoR and PhoB are encoded by the same operon, not only PhoB levels, but also PhoR levels increase as a result of the changes in the reaction rates. This causes the cell to have more sensor histidine kinase, resulting in a more immediate response. The response time thus decreases with an increase in the amount of sensor histidine kinase as well as an increase in its activity.

The deterministic ODE simulations lead to observations that describe the average dynamic behaviour of the variable concentrations for the simulated 4.5 hours. To observe the possible fluctuations in the system, we have performed stochastic simulations. This way, we have been able to compare the mean behaviour with the regulatory dynamics that incorporates the noise due to smaller molecular numbers. Figure 4.2 as well as Supplementary Figure B.2 include a comparison of the deterministic and the stochastic simulations. The stochastic simulations are performed by applying the standard conversion to obtain molecule numbers from the concentrations. The grey fluctuating lines show the stochastic results at a single representative simulation; the simulation shows the expected fluctuations in the model species

with smaller numbers such as mRNA molecules as well as the qualitative agreement between the stochastic and the deterministic simulations.

Binding and unbinding of transcription regulators are a primary mechanism for gene regulation, whereby transcription factors operate at a much faster time-scale in comparison to the incoming signalling cascade. While the rate of binding of transcription regulators are known in many cells, little is known about how cells can modulate their unbinding for regulation [23]. The unbinding rate of an active transcription factor can thus vary over many orders of magnitude [2, 23]. Therefore, in the initial analysis, we have fixed the transcription factor unbinding rates (r_{10r} , r_{11r}) to $1000s^{-1}$ as shown in Table 4.1, and analysed the system behaviour with respect to variations.

To this end, we have experimented in stochastic simulations with different DNA unbinding rates, that is, r_{10r} and r_{11r} , from $100/s$ up to $5000/s$. In accordance with the common practice, we have used stochasticity to quantify the noise that arises from the binding of a regulatory protein to a promoter [87]. The resulting amplification in oscillations in stochastic simulations, shown in Figure 4.3 as well as Supplementary Figures B.9, B.10 and B.11 are due to the increase in the promoter unbinding rates. In Figure 4.3, we have quantified the decrease in noise in the steady state distribution of mRNA and active promoter levels in terms of the ratio of the standard deviation over the mean. In this respect, the deterministic simulations display how DNA binding rates affect the mean behaviour, while the stochastic simulations bring about the loss of coherence due to noise in gene expression. These results indicate that lower unbinding rates, as observed in saturation conditions, are required for stable gene regulation that is not affected by noise. This also reflects how both genetics and noise due to environmental factors can affect the development of targeted pathway interventions for faster P_i accumulation.

The results above show that our control model provides detailed predictions about the complex effects of production pathways of the P_i accumulation system. The simulations are in good agreement with the experimental data and the general concepts described in the literature for the functionality of Pho regulon [38, 168]. The control model proposed can thus serve as a virtual lab, which can be used to test and justify the theoretical approaches on the P_i intake system.

PhoR tunes the P_i intake both up and down. Besides the chemical properties of the proteins in the regulatory system [54], also the activity of the TCS proteins influence the P_i intake. Moreover, as displayed in Figure 4.2 C, the feedback mechanism due to the increased expression of sensor histidine kinase PhoR and the response regulator PhoB introduces a speed-up of an hour in comparison to the simulations, where this feedback mechanism is

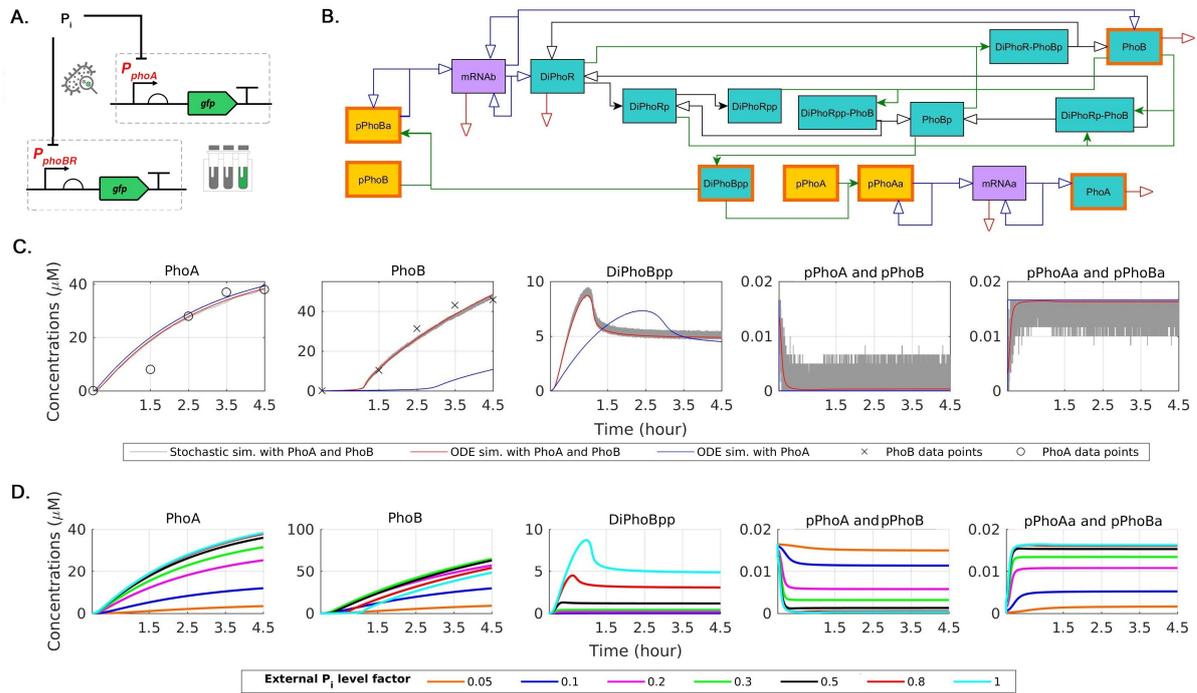


Fig. 4.2 Schematic representation of the experimental data and the control model, and its dynamics in response to varying external P_i concentrations. **A.** The data for the PhoA and PhoB expression are obtained using PCR amplified DNA from *E. coli* MG1655 genome and transcriptionally fused to the translational coupler BCD2 and the fluorescent ms-fgfp gene. **B.** The control model has been obtained from the experimental data and the chemical reaction network (CRN) described in Methods by applying a fitting procedure with the physiological ranges obtained from the literature in Table 4.1, and verified by sensitivity analysis. The deterministic ODE and stochastic simulations are performed by applying the standard translation from CRNs based on stoichiometry. The blue colour denotes the proteins, orange denotes the promoters in their active and inactive forms, and purple denotes the mRNA molecules. Filled arrowheads denote the reversible reactions. The red arrows denote degradation reactions, the green arrows denote complexations, the black arrows denote phosphorylation and dephosphorylation, and the blue arrows denote the transcription and translation reactions. The model species that are plotted in C and D are distinguished with frames. **C.** The dynamics of the highlighted species of the control model in panel B as a result of the fitting procedure together with the experimental data, as described in Methods, are plotted. In the first of the two steps, the blue curves are obtained by using only the experimental values of PhoA levels. In the second step, that delivers the control model, the red curves are obtained by using the experimental values of both PhoA and PhoB levels. Inclusion of the PhoB data highlights the contribution of the feedback mechanism to the response dynamics, which is otherwise not represented. The stochastic dynamics, plotted in grey, display the fluctuations in the control model that are due to small molecule numbers and are not observable in the deterministic simulations. **D.** The response of the control model to variations in the external P_i levels, which are represented as fold change factors, applied to the autophosphorylation propensities of PhoR. A higher external P_i concentration corresponds to a smaller factor and vice versa.

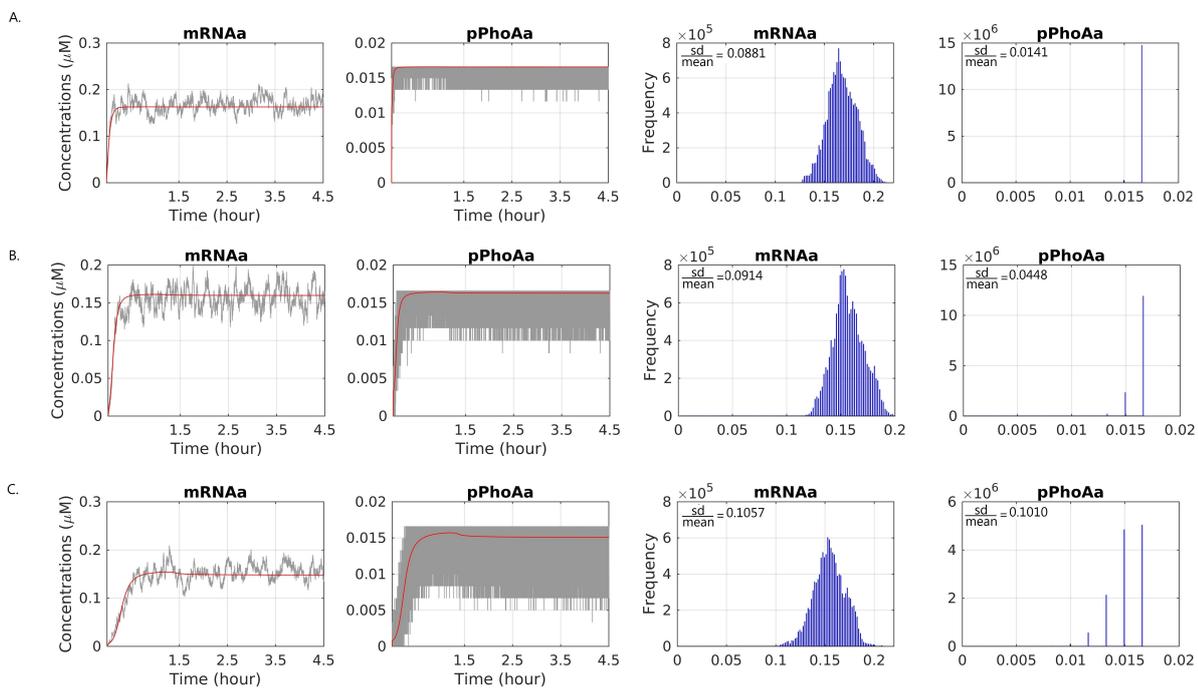


Fig. 4.3 The variations in the mRNA and active promoter levels due to the unbinding rate of the promoter and the active transcription factor. An unbinding rate of 100/sec (A) results in much less spread in the steady state distributions in comparison to unbinding rates of 1000/sec (B) and 5000/sec (C). The variations are quantified as the ratio of the standard deviation and the mean.

not taken into consideration. This indicates that a faster response in terms of P_i intake is delivered by an increase in the histidine kinase levels as well as the increase of its activity due to signalling.

A notable feature of the P_i response system is that the sensor histidine kinase is bifunctional: it participates in both phosphorylation and dephosphorylation of its cognate response regulator. In this respect, the TCS autoregulatory design is a distinct mechanism from the conventional positive feedback loops. The bifunctional PhoR component is an autokinase with concomitant opposing phosphatase activity [35, 113, 159].

The dual role of PhoR is a mechanism that enhances signal robustness [53, 149]. Moreover, it has also been shown that the phosphatase activity in TCS provides a rapid dephosphorylation mechanism that shuts off the system, and thereby restores it to the original state [55]. Such a dynamics can be triggered, for example, by an increase in the external P_i concentration and the consequent decrease in the autophosphorylation activity of PhoR.

We have analysed the effect of the changes in autophosphorylation rates to the system behaviour. Because autophosphorylation becomes possible when the starvation signal prevents PhoU from inhibiting PhoR, the propensity of autophosphorylation depends on the incoming signal, which is a function of the external P_i levels. By decreasing the autophosphorylation propensity by applying various fold changes, we can thus see the effect of an increase in external P_i concentration on the system as depicted in Figure 4.2 D and Supplementary Figure B.12. In these simulations, a decrease in PhoR activity due to increased external P_i concentration results in a proportional decrease in the active PhoB dimers, and a decrease in the promoter activity as well as the PhoR activity.

A complementary realisation of this mechanism is given by the association of PhoB to PhoR, that is, r_8 . Although the physiological range for this parameter is narrow, as a result of a hypothetical increase in the association rate of PhoB and PhoR, the PhoB concentration stays low for a longer time period and the levels of active PhoB dimers decrease proportionally as displayed in the fold change experiments in Supplementary B.13.

In our model, we assume that the growth in the cell culture within the considered time interval is negligible. However, due to cell cycle, which has a time scale in the order of an hour, the protein concentrations can be subject to dilution besides the active degradation of the molecules we have considered. To this end, Supplementary Figures B.14, B.15, and B.16 explore the effect of higher degradation rates due to dilution in growth conditions, (r_{17} , r_{18} , r_{19} , r_{20} , and r_{21}), together with higher external P_i levels, given with a decrease in the rates r_1 and r_2 . As it can be seen in Supplementary Figure B.12, a decrease in the autophosphorylation rates does not only lower the steady state levels, but also slows down the activation of the transcription factor by preventing the formation of an initial peak in DiPhoBpp levels.

A concomitant increase in the degradation and dilution rates delays reaching a steady state. However, this does not drastically alter the eventual active DiPhoBpp concentrations.

Starvation response can be obtained with synthetic promoters in the conditions of higher external P_i concentrations. The simulation results in Figure 4.2 D and Supplementary B.12 demonstrate the system's adaptation to the stimuli due to P_i concentration, whereby the autophosphorylation propensity of PhoR acts as a proxy for the external P_i levels. These simulations predict how changes in the external P_i concentration affect the Pho regulon, and in particular, how the promoter activity decreases with an increase in the external P_i concentration. These results thus confirm that the adaptation of gene expression is clearly dependent on the P_i response stimuli [95, 168]. Moreover, the model provides a mechanistic explanation for the interplay between the system components under the conditions of varying external P_i concentrations, which result in variations in the promoter activity.

It is well established that the protein production rate is greatly influenced by the specific nucleotide sequence of the promoter [70, 82, 83]. In this respect, synthetic biology and genetic engineering methods aim at synthesising promoters with the desired strength. To this end, in order to observe the possible variations in gene expression due to variations in promoter strength, we have performed a class of simulations. The results of these simulations in Figure 4.4 display measurements of the steady state levels of PhoA promoter activity (pPhoAa) as well as the PhoA yield of the system as the resulting product in terms of the area under the curve (PhoA AUC).

In these simulations, we have scanned regimes with varying external P_i concentrations simulated by applying fold changes to the rates r_1 and r_2 as in the simulations in Figure 4.2 D and B.12. We have applied a fold change of 1 for the control regime with $0\mu M$ external P_i , and 0.5, 0.2 and 0.1 for increasing external P_i levels. For each external P_i regime, we have scanned 100 different promoter designs by means of simulations that apply 10 fold change values from 0 to 2.5 with steps of 0.25 to the promoter binding rates r_{10} and r_{11} as well as 10 such fold change values to the promoter unbinding rates r_{10r} and r_{11r} . The heatmaps resulting from these 100 simulations are depicted in Figure 4.4, where the control values for 1 fold change for binding and unbinding rates are marked by dashed lines.

These results indicate that the steady state promoter activity and the PhoA yield are highly correlated in all the regimes and for all the promoter binding and unbinding rates. As expected, when the control system's output in P_i starvation condition is compared with the output in regimes with increased external P_i , we observe a decrease in PhoA yield. Moreover, these results predict that in order to obtain the starvation response in the conditions with higher external P_i concentration, promoter binding rates need to be increased and unbinding

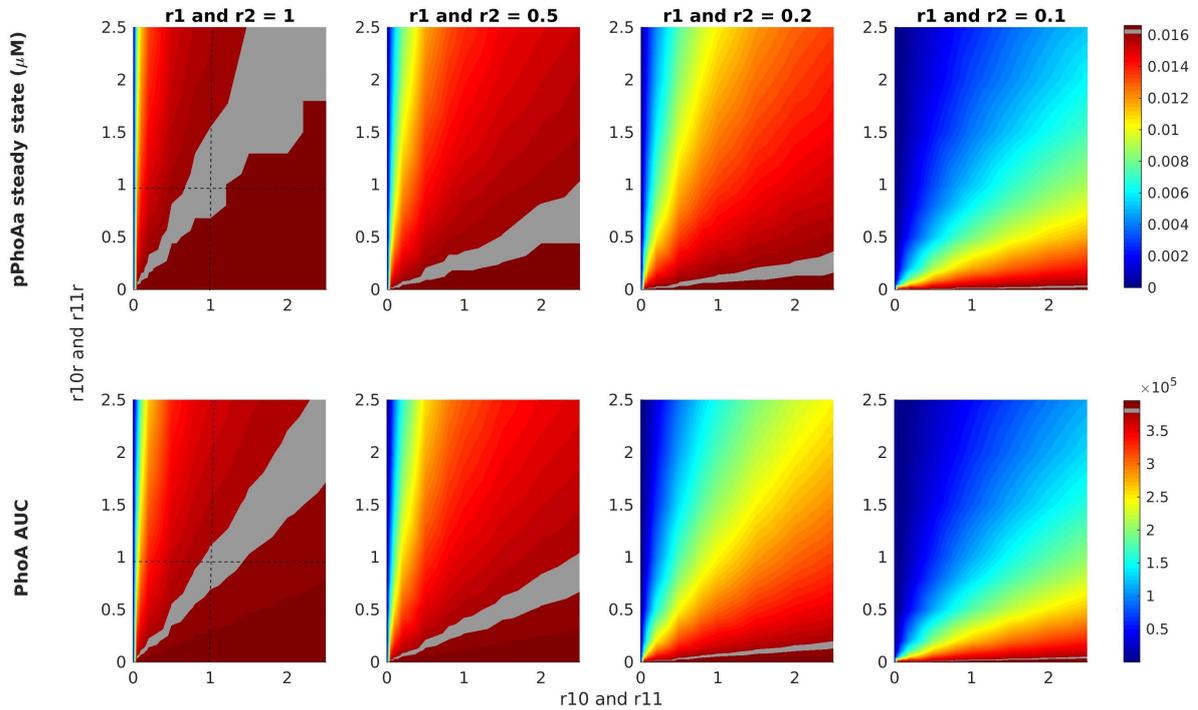


Fig. 4.4 Heatmaps for the activity of various promoter designs for pPhoA, and pPhoB and the resulting PhoA expression (bottom-row) under different external P_i concentration conditions. The heatmaps are ordered decreasingly from left to right according to the external P_i concentration given by the fold changes applied to the PhoR autphosphorylation reactions r_1 and r_2 . The left most column with 1 as the fold change value is the starvation condition with $0\mu M$ external P_i . Each heatmap scans 100 simulations by applying 10 different fold change values to the promoter binding rates r_{10} and r_{11} as well as 10 different fold change values to the promoter unbinding rates r_{10r} and r_{11r} . The upper row displays the resulting steady state levels of the active promoter pPhoAa, whereas the lower row displays the yield of PhoA gene expression measured as the area under the curve (AUC). The intersection of the dashed lines in the left column delivers the experimentally observed regime observed in Figure 4.2. The levels of this regime, that display the starvation response, are highlighted in all the heatmaps.

decreased. Promoters that provide the required strengths can be obtained by modifying the nucleotide sequences, for example, as in [34, 82, 99].

Through synthetic biology approaches, we have obtained a library of synthetic promoters while exhibiting a broad range of sensitivity in sensing P_i in *E. coli*. At the time of writing, our collaborators are working on the library and it is under development. We have tested the predictions of the model on experimental data obtained with one of the synthetic promoter Pliar00117 (pPliar00117) from the library under starvation conditions with $0\mu M$ external P_i

concentration. Moreover, we have used the model to explore the effect of possible synthetic promoter designs under various external P_i concentration conditions. The experimental data and the simulation results with our model are depicted in Figure 4.5. The selected Pliar00117 synthetic promoter verifies the model predictions as it has a similar behaviour as the control model for the pPhoA promoter in starvation conditions. Moreover, the model further predicts the synthetic promoters with increased strength, given with higher binding rates and lower unbinding rates, deliver responses similar to the starvation response, also in the presence of higher external P_i concentrations. The simulation results indicate that, within a modular framework, individual promoters can be easily replaced for various tasks. The simulations mechanistically quantify how changes in the genetic components affect the behaviour of the circuit.

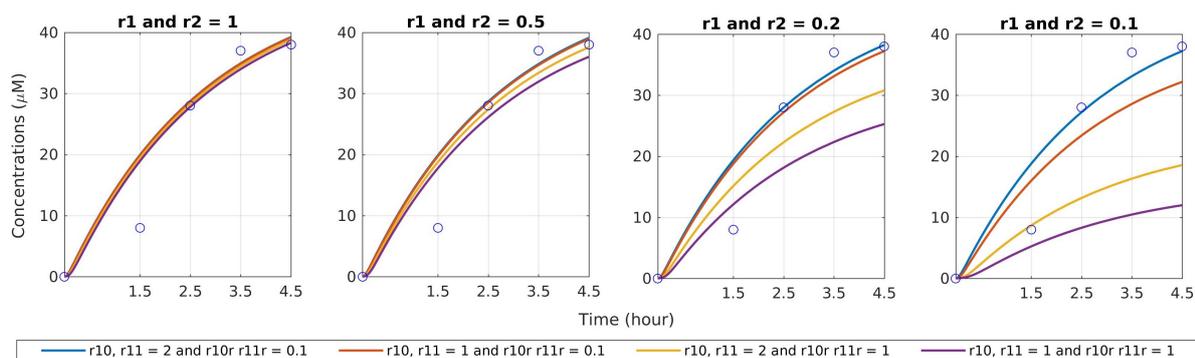


Fig. 4.5 Comparison of the experimental data on PhoA expression with the synthetic promoter Pliar00117 together with the simulation results with varying external P_i concentrations and promoter parameters that model synthetic designs. The experimental data on the starvation response with the synthetic promoter Pliar00117 is represented as hollow circles. As in Figure 4.4, the plots are ordered decreasingly from left to right according to the external P_i concentration given by the fold changes applied to the PhoR autphosphorylation reactions r_1 and r_2 . The left most column with 1 as the fold change value is the starvation condition with $0\mu M$ external P_i . Each plot displays four simulations with varying fold change values applied to promoter binding and unbinding rates that model various promoter designs. A modified promoter (blue curve) can reproduce the starvation response in low as well as high external P_i concentration, and reproduce the experimental data under Pliar00117 starvation conditions.

Sensitivity Analysis To assess the sensitivity of our model to the parameters, we have performed a two-step analysis. In the first step, we have considered the physiological interval of the parameters given in Table 4.1. For this analysis, we have only included the rate

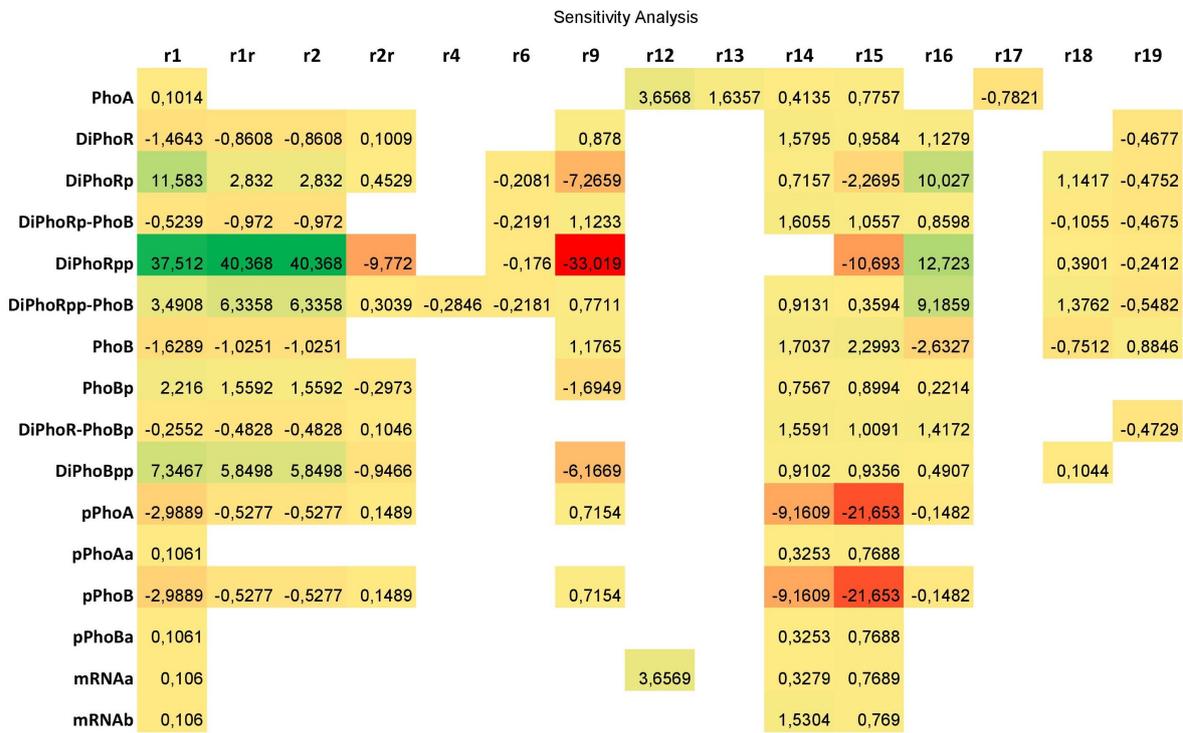


Fig. 4.6 Heatmap displaying the results of the sensitivity analysis by considering the physiological interval in Table 4.1. For each parameter, the maximum and minimum values within its physiological range are considered for simulation, and the area under the curve (AUC) for each species is computed. The difference of the AUC for the maximum and minimum parameter values are then normalised with the AUC of the control model. Red represents the decreasing effect and green represents the increasing effect.

parameters that have been taken from literature and have been estimated within a given range. For each parameter, we have run simulations by instantiating the model with the maximum and minimum values of its physiological range, and, for each species, we have computed the yield of the system in terms of the area under the curve (AUC). Figure 4.6 displays the results obtained by taking the difference of the AUC for the maximum and minimum parameter values, normalised with the AUC of the control model. The resulting heatmap quantifies the impact of each parameter on the system dynamics with respect to plausible variations within its physiological range. The results demonstrate that the changes in the translation parameters are more pronounced than in the others. Moreover, in accordance with the results above, the autophosphorylation rate of PhoR impacts the active transcription factor levels and the transient species that lead to it.

In the second step, we have performed a sweeping analysis by considering all the model reactions and species within a range of 3 orders of magnitude both up and down. That is, each reaction rate is multiplied with a fold-change factor within a spectrum of 6 orders of magnitude, that is, from 10^{-3} to 10^3 . We have then computed the AUC for each species and normalised the outcome with the AUC of the control model. The heatmap depicted in Supplementary Figure B.17 quantifies the impact of these changes and predicts the system behaviour under hypothetical conditions simulated by such variations in parameters.

4.4 Conclusion

We have presented a computational model and its experimental validation for quantifying dynamic mechanisms of auto-regulation in *Escherichia coli*, including the transcriptional regulatory network, in response to varying external phosphate levels. Our model provides a mechanistic explanation of the interplay between TCS, Pho regulon and promoter efficiency under the conditions of varying external P_i concentrations. Being parameterised with the physiological ranges of its components, the results provided by the model are in good agreement with the theory and the general concepts described in the literature for the functionality of Pho regulon [38, 168]. In particular, the output of the model in terms of gene regulation delivers the expected system dynamics without the need for external intervention. Moreover, the model provides predictions for the complex effects of TCS activity and consequent dynamics. A direct validation of these predictions is provided by the good fit of the experimental data.

Besides the deterministic simulations with the ODEs, we have employed stochastic simulations to highlight the effect of noise in the system [147]. As expected, the stochastic simulations are consistent with the deterministic simulations, which display the mean behaviour, thus are more adequate for sensitivity analysis. Stochastic simulations, on the other hand, capture the noise due to concomitant fast and slow reactions and the consequent fluctuations observed in experimental observations. In particular, our stochastic simulations highlight the fluctuations due to binding and unbinding of the transcription factors with the promoters, which operate at a much faster time scale in comparison to the preceding signalling cascade [44, 138, 177]. In these simulations, higher unbinding rates result in greater fluctuations that correlate with the decrease in binding saturation, a requirement for a robust signal. While confirming the notion that the complex networks of interacting molecules within cells should be robust [117, 176], these results expose the additional role of the transcription factor unbinding rate in tuning the protein synthesis.

The simulations with our model confirm that the dynamics of TCS and its responsiveness to both genetic and environmental perturbations play a key role in tuning the *E. coli* P_i response. The model can thus serve as a virtual lab for the P_i intake system, and can be used to explore and test various promoter designs, for example, in biotechnology applications such as sensors for wastewater treatment or detecting environmental pollutants to relocate towards them.

Author contributions

This work is supported by COSBI and by the European Union's Horizon 2020 research and innovation programme under the grant agreement No 686585 - LIAR, Living Architecture. It is result of a collaboration between CIBIO and Juan Nogales's group at CNB, Madrid, Spain. The conference article is the result of the joint effort of both the institutions. Ozan Kahramanođulları and I designed and performed simulations and sensitivity analysis. Juan Nogales, Jesus Torres, and José L. García provided the experimental data. All figures have been produced under the supervision of Ozan Kahramanođulları. Ozan Kahramanođulları and I wrote and edited the manuscript. Martin M. Hanczyc supervised the work. All the authors reviewed the manuscript.

Chapter 5

A Novel Closed-loop Multi-level Model of Glucose Homeostasis for Normal Glucose Regulation and Type 2 Diabetes

Type 2 diabetes (T2DM) patients suffer a pathology that involves failure at different levels and subsystems. It is diagnosed at the phenotypic level by inspecting fasting glucose, but it arises and it is maintained at molecular level by impaired intracellular insulin signalling. The pathophysiologic processes underlying the regulation of glucose homeostasis are considerably complex at both cellular and systemic level. Therefore, we need to understand the links between the several layers of abstraction of glucose metabolism to control the disease progression. As a first step in this direction, researchers are developing the hierarchical description provided by multi-level models that are able to reproduce the glucose homeostasis.

In this framework, this chapter provides a novel multi-level closed-loop model of whole-body glucose homeostasis. This model is coupled with the molecular specifications of the insulin signalling cascade in the adipose tissue (abstracted as adipocyte), under the experimental conditions of normal glucose regulation and type 2 diabetes. Moreover, it is important to observe the system in a long period and how the levels of abstraction react to each other. The coupling of the low level model with a closed loop whole body model allows seeing the effect on adipocytes not only after one meal but in a perpetual fashion.

To this end, we introduced the closed-loop model allows for the input and the output to be connected in such a way that it is possible to run long self-sustained simulations. Model takes into account key hormones like glucagon, incretins, ghrelin and leptin and key compartments such as the liver, muscle and adipose tissue. The combination of a multi-level and closed-loop modelling approach provided a fair dynamic description of the core determinants of glucose homeostasis at both cellular and systemic scales.

The ordinary differential equations of the model were used to describe the dynamics of glucose and key regulatory hormones and their reciprocal interactions among gut, liver, muscle and adipose tissue. The model was designed for being embedded in a modular, hierarchical structure. The closed-loop model structure allowed self-sustained simulations to represent an ideal *in silico* subject that adjusts its own metabolism to the fasting and feeding states, depending on the hormonal context and invariant to circadian fluctuations. The cellular level of the model provided a seamless dynamic description of the molecular mechanisms downstream the insulin receptor in the adipocytes by accounting for variations in the surrounding metabolic context.

What follows is the content of the article, published in the journal of PLoS ONE in February 2018 [165].

5.1 Introduction

The glucose homeostasis is a physiological closed-loop, which is able to maintain the plasma glucose levels within a narrow physiological range, as a result of an interaction among many components [47, 120]. A disruption in the governance of the glucose-insulin system can lead to variable degrees of altered glucose regulation, which may ultimately result in overt diabetes mellitus. According to the most recent estimates, diabetes mellitus affects over 415 million individuals worldwide [46], is characterised by severe cardiovascular complications leading to early death [37], and its prospective incidence trends highlight it as a global burden of pandemic proportion. Most diabetes cases are classified as type 2 diabetes mellitus (T2DM), which is the result of genetic predisposition and life style factors and it is usually developed in adult age.

Over the past decades, a number of mathematical models of glucose and insulin dynamics have been developed to allow the description and interpretation of such processes, which are often not accessible to direct measurement *in vivo* [1]. In such models the physiology of the glucose-insulin system in different experimental conditions are described through a set of ordinary differential equations (ODEs) or delay differential equations (DDEs) [10, 15, 18, 20, 104, 105, 107, 131, 161]. In order to soften the burden of both experimental and modelling complexity, more parsimonious model was proposed and successfully employed by the so called minimal model [10], which was originally applied to estimate insulin sensitivity by inspecting the time courses of insulin and glucose after an intravenous glucose tolerance test (IVGTT).

In its most advanced specification, the concept of minimal models has been applied to more physiologic experimental conditions and in some instances has been enriched by

integrating a broader range of physiological variables, including several hormones and regulatory elements such as ghrelin [161], glucagon [20] and incretins [15, 33, 161]. However, at present, there is no evidence of a comprehensive and structured model, which summarises the dynamics governing the glucose absorption by the peripheral tissues after a meal, according to the regulation performed by insulin and other hormones at subcellular level. This is possibly due to the constraints inherent to the reductionist modelling approach applied so far, an issue that is currently amenable to solution with the hierarchical description provided by multi-level models. This approach, also referred to as hierarchical modelling, constitute an interesting trend in mathematical modelling of biological mechanisms, because they provide a broader and more detailed description of the system than classical single level models [29].

Recent applications of multi-level models have been proposed to describe the pathophysiology of beta-cells in the endocrine pancreas [51], and the whole body effects of the altered insulin signalling cascade in adipocytes [126], while other applications described the effects of inflammation on the onset of T2DM and its complications [17]. A holistic, rather than a reductionist approach, combining more levels of abstraction, is necessary to improve the understanding of the disease and shed light on the physiopathology.

A comprehensive, hierarchical description may be reasonably applied to T2DM, because the phenotypic hallmark of hyperglycaemia is the consequence of alterations involving complex hormonal and signalling networks, individual tissues and cell subtypes. Recently, Chew et al. [24] proposed a model of the glucose regulatory system combined with the insulin signalling model of Sedaghat et al. [145]. An interesting work of Nyman et al [126] combined the organ level model of Dalla Man et al. [106] with three different detailed versions of insulin signalling in the adipocytes. The most detailed version includes the model from Kiselyov et al. for the description of insulin binding to its receptor [90].

We moved one step forward, as we exploited a multi-level modelling approach to provide a novel closed-loop whole-body model of glucose homeostasis, coupled with a layer describing the insulin signalling cascade in adipocytes in the experimental conditions of normal glucose regulation and type 2 diabetes mellitus. The closed-loop structure was designed to allow self-sustained simulations, thus fostering the modeller's opportunities of investigating the biological system in its components, while providing a way to test regulative phenomena that work at different time scales and possibly have a delayed effect on the overall system dynamics. The model was tested *in silico* in both the conditions of normal glucose regulation (NGR) and T2DM.

Since insulin resistance constitutes one of the key pathophysiologic determinants of T2DM, and given the increasing relevance of the adipose tissue as an endocrine organ influencing systemic energy balance and glucose homeostasis [72, 88], we integrated the

herein presented whole-body model with an additional layer, drawn from the most recent insulin signalling model proposed by Nyman et al. [127], that provides a detailed specification of the intracellular signalling cascade in the adipocytes.

The presented hierarchical architecture of glucose homeostasis might be relevant to strategies aimed at improving the molecular descriptions of other organs and tissues that are here considered only at the whole-body level.

5.2 Materials and Methods

5.2.1 Glucose physiology at the whole-body level

A graphical representation of whole body glucose metabolism is shown in Figure 5.1 where all the variables and compartments considered in the high level model that is proposed are shown.

Under physiological circumstances, blood glucose concentration is around 5 mM , and many organs and hormones are involved in the maintenance of this delicate equilibrium [134]. Whole body glucose regulation mechanisms at the organ level are described by the following sequence of events.

The breakdown of the meal constituents after oral feeding occurs at time 0, glucose transits to the stomach and then to the intestine, and it is ultimately resulting in an inward flux of micro- and macro-nutrients, including glucose, to the bloodstream [33]. These events triggers the endogenous insulin secretion, which is amplified by a concomitant increase in the circulating levels of the incretin hormones, including the gastric inhibitory polypeptide (GIP) and the glucagon-like peptide-1 (GLP1) [33]. Among the other neuroendocrine and gastro-intestinal tract effects, the incretins potentiate the release and the *de novo* synthesis of insulin from pancreatic beta cells, thus contributing to the proper glucose disposal in peripheral tissues and to maintain plasma glucose levels within the physiological range [161]. The incretin effect is normally responsible for about 50 to 70% of insulin production [161]. From the intestine, glucose is absorbed to the plasma, also causing beta cell insulin production. Therefore, insulin release is modelled here as a consequence of direct effects exerted by the glucose and indirect effects mediated by the incretins.

The liver has a major role in the homeostasis of glucose, which is stored as glycogen in the postprandial state and released at fasting through endogenous hepatic gluconeogenesis. These mechanisms play a crucial role in maintaining constant plasmatic glucose concentrations, in case of prolonged fasting, for instance during the night. Endogenous glucose release from the liver decreases rapidly and is suppressed by nearly 80% during five hours postprandial

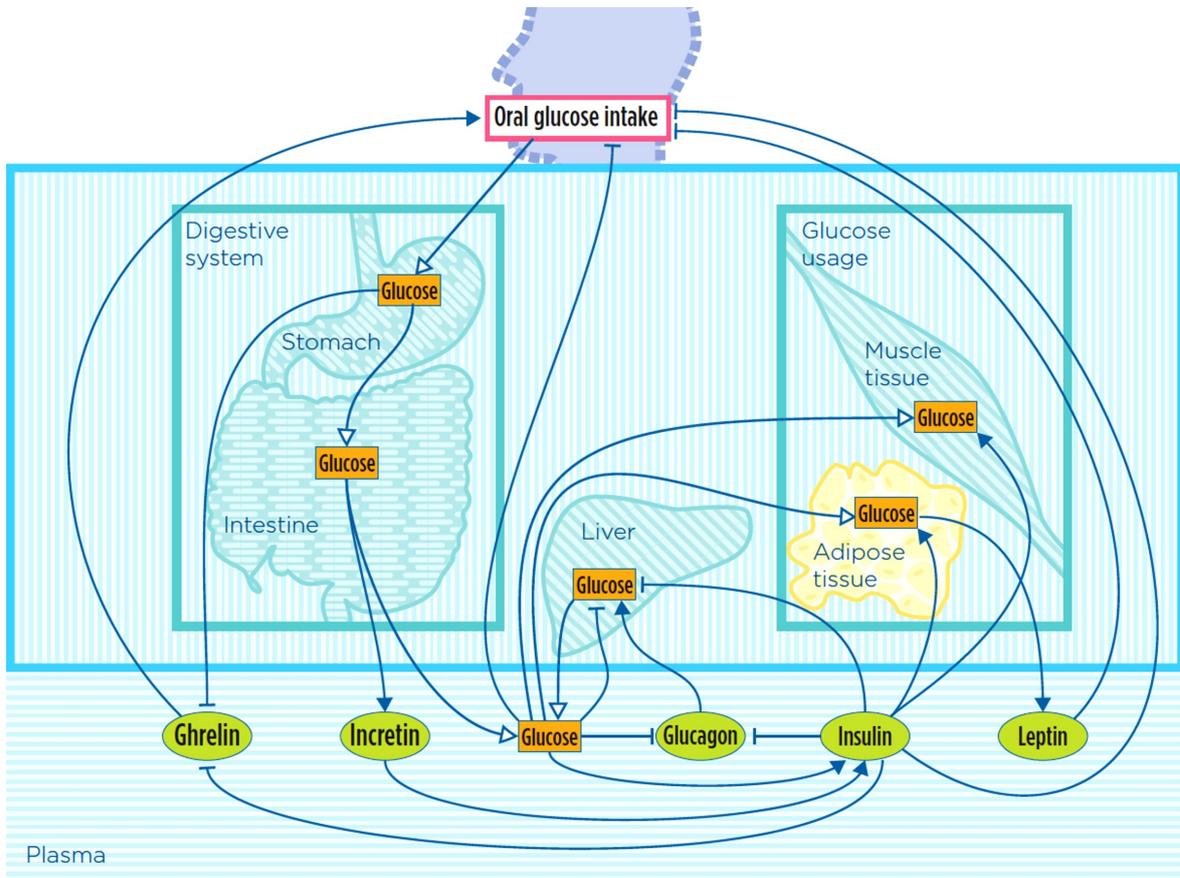


Fig. 5.1 Graphical representation of the whole-body glucose metabolism as considered in our model, according to the notation introduced in [61]. Only the organs/tissues for which a variable has been explicitly included in the model are depicted in the figure (other key organs/tissues of glucose metabolism, like pancreas and brain, are not displayed in the figure even if their effect has been indirectly taken into account in model equations, see Results). Adipose tissue is coloured in yellow to highlight that it is the part for which a model at the cellular level is also provided (see Figure 5.2). Green ovals (hormones) and orange rectangles represent model variables; arrows represent mass transfer (white head), stimulation (black head) and inhibition (T head).

period. Endogenous glucose production is suppressed by high levels of both glucose and insulin [58, 106, 154]. Insulin acts on the liver by inhibiting gluconeogenesis and glycogen breakdown [58]. Another important player in liver glucose metabolism is glucagon, a hormone secreted from pancreatic alpha cells. Glucagon has the opposite effect of insulin, stimulating glycogen breakdown as plasma glucose levels decrease [64, 134]. Glucagon secretion is inhibited by hyperglycemia and stimulated by hypoglycemia [64, 134, 154, 156].

Glucose tissue uptake happens when insulin binds to its cell receptors on target organs, including, among others, the liver, muscle and adipose tissues, where glucose is stored as glycogen [102]. The most substantial percentage of glucose (15% - 20%) is utilised from skeletal muscle, while the adipose tissue utilises 2% to 4%, while relatively smaller glucose proportions eventually reach the brain, kidney, blood cells, adipose tissue and the splanchnic organs the rest is utilised from the brain, kidney, blood cells and splanchnic organs [134].

Glucose metabolism is also regulated by hunger and satiety, through a number of finely regulated biological mechanisms. Among the several hormones potentially involved in the regulation of glucose homeostasis, we focused on the reciprocal interactions occurring among insulin, ghrelin and leptin [28], as their mathematical specifications were available from existing literature. Hunger is here defined as the amount of food needed by organism [79]. Leptin, is a hormone so called satiety hormone, is mainly secreted from peripheral tissues (mainly the white adipose tissue) after glucose uptake, and its production rate is directly related to the size of adipose tissue mass in the body [26, 84]. After its production leptin is transported from plasma to the brain where its signal is responsible for hunger inhibition [26]. Fall in serum leptin levels -as it happens during prolonged starvation- in fact leads to downstream neuroendocrine alterations that prevent reproduction, reduce thyroid function, activate the hypothalamic-pituitary-adrenal axis, and inhibit the growth hormone [26, 50, 143].

On the other side, ghrelin, this counteracts leptin, acts as the hunger hormone. It is secreted from the empty stomach, stimulating oral glucose intake [141], and it is inhibited by insulin [9,34] [160, 161]. It circulates in the blood and serves as a peripheral signal to the central nervous system to stimulate appetite [161]. Ghrelin, contrarily to leptin, acts on the short term. Plasma ghrelin concentration has been shown to increase before a meal and to decrease to its minimum concentration about one hour after food intake. Fasting plasma ghrelin and leptin levels are negatively correlated.

The latter, hunger is also inhibited by high levels of plasma glucose and insulin [66, 135, 153]. Also insulin regulates hunger: directly, signalling satiety to the brain, and indirectly, inhibiting ghrelin release [161]. In order to build a closed-loop model, oral glucose intake is represented as the amount of glucose needed from the organism, an abstracted hunger signal, and it is computed considering the current levels of leptin, plasma glucose, insulin and ghrelin.

5.2.2 Insulin signalling at the cellular (adipocyte) level

A schematic representation of the molecular level model for insulin signalling in the adipocyte is shown in Figure 5.2. The cellular part of the hierarchical model was taken from [134]

and then modified. Insulin surrounding the adipocyte binds to the insulin receptor (IR) on the cell surface. The insulin receptor is composed of two extracellular α subunits and two transmembrane β subunits linked together by disulphide bonds. Binding of insulin to α subunit induces a conformational change resulting in the autophosphorylation of a number of tyrosine residues present in the β subunit. These residues are recognised by phosphotyrosine-binding (PTB) domains of adaptor proteins such as members of the insulin receptor substrate family (IRS). Together with the autophosphorylation process the insulin binding with the receptors causes the endocytosis of the IR.

Receptor activation leads to the phosphorylation of key tyrosine residues on the insulin receptor substrate protein 1 (IRS1) inside the cell. These phosphorylations are used as docking sites by downstream effector molecules. The subsequent signalling cascades involve phosphoinositide 3-kinase (PI3K), protein kinase B (PKB) and mammalian target of rapamycin (mTOR).

Insulin signalling is involved in different intersected pathways, such as glycogen synthesis, protein synthesis and cell survival. Among these, the next important step in this path is the phosphorylation of PKB at Thr 308 from IRS1 and at Ser 473 from mTOR complex 2 (mTORC2). Both phosphorylations are needed in order for PKB to be fully active and specific to its substrate. Targets of PKB include glycogen synthase kinase-3 (GSK3), which is involved in control of glycogen synthesis, and AS160, which is involved in the control of glucose uptake. Indeed AS160 regulates the translocation of vesicles including glucose transporter 4 (GLUT4) from the cytosol to the plasma membrane. The vesicles merge into the membrane where GLUT4.

5.2.3 The hierarchical model

The construction of a hierarchical model for glucose homeostasis and insulin signalling is based on the connection between the two levels of abstraction: the whole body and the cellular levels. To connect the two levels of abstraction, the knowledge of physiology of the human body has been necessary.

Plasma insulin and glucose are the variables linking the whole-body level to the cellular one. The whole body model contains the adipocyte tissue glucose mass which uptakes glucose plasma via plasma insulin. Instead, the cellular level as it was built in [127] takes as input insulin and glucose as the constant value of 100 nM and 5 mM, respectively. The output of the cellular model is the glucose uptake from the adipose cell.

Concerning the link between the whole body model and the intracellular one, the presence of the interstitial fluid for both insulin and glucose is crucial to combine the two levels of abstraction and to get a more realistic result. This because the insulin signalling occurs in

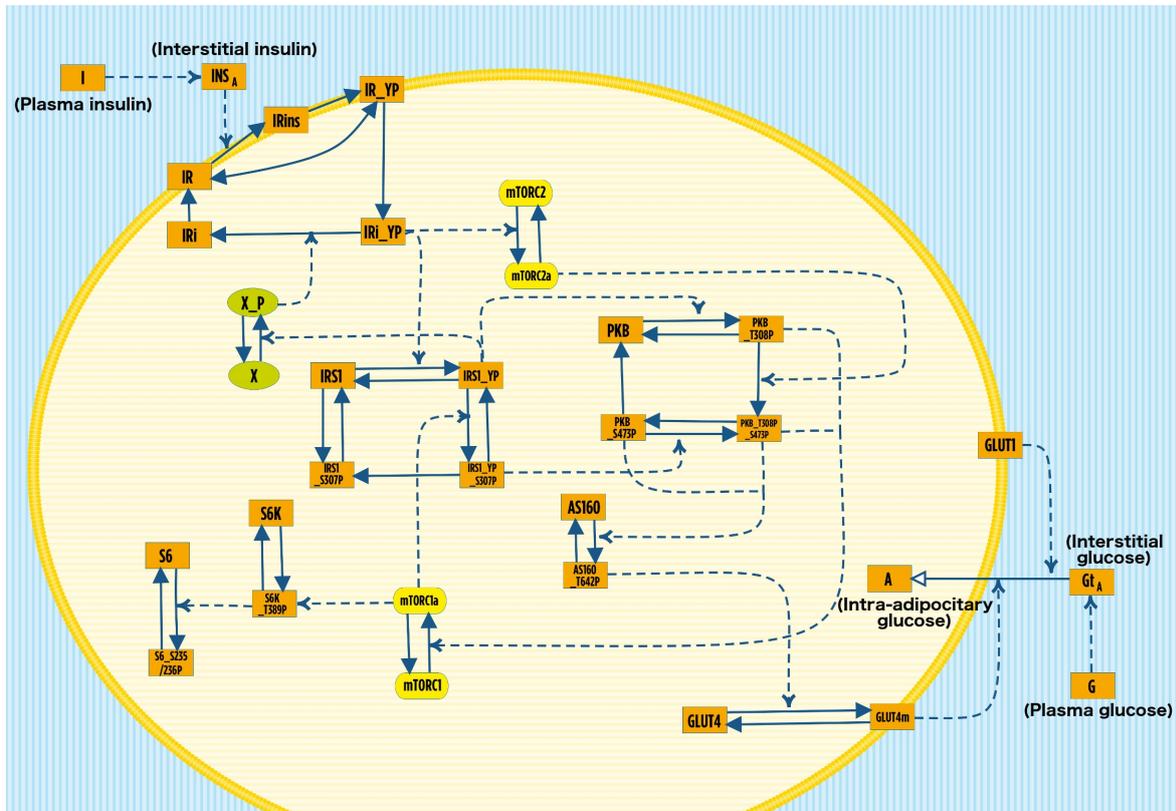


Fig. 5.2 Graphical representation of the model describing the insulin signalling in adipocytes at the cellular level, according to the notation introduced in [61]. Solid arrows represent state modification, while dashed arrows indicate reaction stimulation. Protein complexes are coloured in yellow, green ovals represent the active and inactive feedback protein, while the orange rectangles represent all the other components of the cellular model. The plasma membrane of the adipose cell is represented in yellow and it separates the cytosol (light yellow horizontal lines) from the interstitial fluid (blue and white vertical lines). The variables I and G indicate insulin and glucose concentration in plasma (compartment not represented), which regulate the amount of interstitial insulin (INS_A) and glucose (Gt_A), respectively. For the sake of simplicity, we highlighted the five variables linking the cellular level to the whole body description (namely plasma insulin, interstitial insulin, plasma glucose, interstitial glucose and intra-adipocitary glucose) by adding the corresponding names in parenthesis.

cells that are not surrounded by plasma but by the interstitial fluid. Interstitial fluid is a thin layer of fluid which surrounds the tissue cells of multicellular animals.

The interstitial fluid is found in the interstices, the space between cells, and acts as a kind of fuelling station [174]. Water, ions, and small solutes are continuously exchanged between plasma and interstitial fluids across the walls of capillaries. Plasma, the major component in

blood, communicates freely with interstitial fluid in other words, the nutrients in interstitial fluid come from blood capillaries [174].

Therefore, whole body model and the cellular one were combined together adding two new equations: one for the interstitial insulin and one for the interstitial glucose. The connection between the levels is designed to work through the interstitial fluid, which has not been modelled as a separate compartment, but rather assuming that the interstitial concentrations of insulin and glucose (*i.e.* the amount surrounding the cell) would proportionally correspond to those of insulin and glucose in the plasma. The equation for the interstitial insulin was taken from the literature [106]. The input of the cellular model is interstitial insulin, which depends directly on the value of plasma insulin, so on the whole body model. Interstitial insulin binds its receptor on the cell membrane and prompts the auto-phosphorylation of the receptor and its endocytosis. The internalised phosphorylated receptor starts a cascade of phosphorylation and activation events, according to the model introduced by Nyman et al. [127], here simplified in some parts according to [13]. The model takes into account key actors, such as insulin receptor substrate 1 (IRS1), feedback protein X_P, PKB, mTORC1 and mTORC2 complexes, P70 ribosomal S6 kinase, ribosomal protein S6 and Akt substrate (AS160), which regulate the translocation of GLUT4 from the cytosol to the plasma membrane. The output of the cellular model is connected with the whole-body model: the amount of glucose uptake by the adipocyte (intra-adipocitary glucose) is regulated by the amount of GLUT1 and GLUT4 on the plasma membrane [126, 127].

5.2.4 The mathematical model

We introduce a closed-loop multi-level mathematical model describing glucose homeostasis in NGR and T2DM conditions. In order to be consistent with the previous results the molecular level of the model describing the insulin signalling in adipocytes has been fitted by considering the same experimental data used in the paper of Nyman et al. [127].

The proposed model includes 40 ODEs and 80 kinetic parameters. They were implemented in Matlab 2015b and numerically simulated by means of the state of the art ODE solver ode15s. Initial values and parameters have been obtained from the literature and estimated by the steady state analysis or nonlinear optimisation (for details see the Table C.1, and Table C.2)

All the ODEs of the hierarchical model are listed below starting from the ones of the whole body level, then the two of the interstitial compartment and at the end all the ones of the cellular level.

The whole-body model is described by Eqs. (1) to (14), as follows.

Eq. (1) describes stomach glucose dynamics (S):

$$\frac{dS(t)}{dt} = b_9H(t) - b_8S(t) \quad S(0) = S_0 \quad (1)$$

The first term represents ingested glucose, which depends on glucose intake (H, Eq. (12)) and on rate (b_9). The second term models stomach emptying and depends on the amount of stomach glucose and transfer rate (b_8) according to [33, 161].

Eq. (2) models intestine glucose transit (L), as described by Toghaw et al [161]:

$$\frac{dL(t)}{dt} = b_8S(t) - b_{10}L(t) \quad L(0) = L_0 \quad (2)$$

The first term represents the glucose entry from stomach, which coincides with the amount of glucose exiting the stomach in Eq. (1). The second term accounts for the glucose absorption into the plasmatic compartment, which depends on the amount of glucose in the intestine (L) and on the rate (b_{10}).

The dynamic of plasma glucose concentration (G) dynamics is described in Eq. (3).

$$\frac{dG(t)}{dt} = f \frac{b_{10}L(t)}{v} + f \frac{b_5C(t)}{v} - b_1G(t) - b_3I(t)G(t) \quad G(0) = G_0 \quad (3)$$

The first term, as from Toghaw et al. [161], represents the glucose appearance in plasma from the intestine, where b_{10} is the intestine to plasma transfer rate, v is the glucose distribution volume and f is a fraction of absorption, accounting for the part of glucose is lost in the transfer. The second term describes the liver glucose production, where the glucose coming from the liver (C, see Eq. (7)) is multiplied by the transfer rate b_5 and by f/v , similarly to the first term. The third and fourth terms represent blood glucose elimination through insulin-independent and insulin-dependent mechanisms, respectively. The third term models glucose uptake from brain and other tissues such as, blood cells, renal medulla and splanchnic tissue, which is insulin independent and depends only on G and on the rate b_1 . Instead the last term models the glucose uptake from adipose and muscle tissues, which depends both on plasma insulin concentration (I, see Eq. (4)), and G and the rate (b_3) [161] [33].

Eq. (4) represents the dynamics of plasma insulin concentration (I):

$$\frac{dI(t)}{dt} = b_4G(t) + cW(t)G(t) - b_2I(t) \quad I(0) = I_0 \quad (4)$$

The first and second terms are simplifications of the insulin dynamics described by Toghaw et al. [161]. The first term describes the glucose-dependent insulin secretion where b_4 is glucose dependent insulin secretion rate. The second term represents the incretin-dependent insulin production, which is proportional to plasma glucose and incretins concentration (W , see Eq. (5)), c being the incretin-dependent insulin secretion rate. The last term represents insulin elimination, which depends on I and on insulin disappearance rate constant (b_2).

Eq. (5) describes the variation in concentration of plasma incretins concentration (W) [161] [160]:

$$\frac{dW(t)}{dt} = b_6L(t) + b_7W(t) + s \quad W(0) = W_0 \quad (5)$$

The first term accounts for incretins appearance due to the glucose transit through the intestine (L): it depends on L and on the incretin production rate b_6 . The second term represents incretin elimination, depending on W and their disappearance rate constant b_7 . The last term is the constant incretin appearance rate (s).

In Eq. (6) the dynamics of plasma glucagon (E) is described, according to Sulston et al. [154]:

$$\frac{dE(t)}{dt} = c_0 + \frac{c_1}{c_2 + I(t)e} (G_e - G(t))u(G_e - G(t)) - c_3E(t) \quad E(0) = E_0 \quad (6)$$

The first term models the basal level of glucagon secretion c_0 , which happens at normal fasting glucose levels. The second term represents the dependency of glucagon secretion from plasma glucose concentration where $u(G_e - G(t))$ indicates the Heaviside step function:

$$u(G_e - G(t)) = \begin{cases} 1 & \text{when } G_e - G(t) \geq 0 \\ 0 & \text{when } G_e - G(t) < 0 \end{cases}$$

When G is above the threshold level G_e this part of glucagon secretion is suppressed, resulting in an equilibrium value of glucagon achieved when $\frac{dE(t)}{dt} = 0$, that is, when $E = \frac{c_0}{c_3}$. Otherwise, the term represents glucagon secretion in the α cells of the pancreas, with the secretion increasing at low glucose levels but being suppressed by high insulin levels, according to parameters c_1 and c_2 . The parameter e models insulin effectiveness to represent the cell sensitivity too insulin action, which is compromised in insulin resistance and in T2DM: this rate will be at its maximum in the NGR condition while it is lower in T2DM according

to [154]. The last term describes plasma glucagon elimination, which depends on E and its degradation rate c_3 [154].

The variable C, described in Eq. (7), represents the glucose mass in the liver ready to be secreted, which has been produced from the glycogen breakdown:

$$\frac{dC(t)}{dt} = b_{23} - b_{25}I(t)e - b_{22}G(t) + b_{21}E(t) - b_5C(t) \quad C(0) = C_0 \quad (7)$$

The equation is obtained by combining the works of Dalla Man et al. [106] and Sulston et al. [154], where variable and parameter units have been converted in accordance with the model. The first term models the basal rate of liver glucose production, b_{23} . The second and third terms represent the inhibiting effect of I and G on liver glucose production, according to the rate constant b_{25} and b_{22} and the parameter modelling insulin effectiveness (e). The fourth term accounts for the plasma glucagon (E) effect in stimulating glycogen breakdown, where b_{21} is the rate of liver glucose production, which is glucagon dependent [154]. The last term represents the glucose transfer from liver to plasma according to the transmission rate b_5 . [161]

Eq. (8) describes the dynamics of glucose mass in the muscle tissue (M):

$$\frac{dM(t)}{dt} = 0.1 \frac{v}{f} b_3 G(t) I(t) e - b_{27} M(t) \quad M(0) = M_0 \quad (8)$$

The first term represents glucose entry in the tissue, which depends on glucose and insulin concentrations, on the insulin effectiveness (e) [154], and on the utilisation rate b_3 . The scaling factors $0.1 \frac{v}{f}$ have been introduced to convert plasma glucose concentration to a mass and to set the muscle glucose uptake to the 10% of the whole body glucose uptake [134]. The last term in the equation represents muscle glucose elimination, which depends on muscle glucose mass (M) and on the elimination rate b_{27} .

Eq. (9) represents the adipose tissue glucose mass (A):

$$\frac{dA(t)}{dt} = k_8 GLUT4 m(t) \frac{G_A(t)}{K_m G_4 + G_A(t)} + GLUT1 \frac{G_A(t)}{K_m G_1 + G_A(t)} - k_{gluc} A(t) \quad A(0) = A_0 \quad (9)$$

Eq. (9) is one of the links between the whole body model and the cellular one and includes variables from both. The first two terms of equation come from the work of Nyman et al. [126]. These terms represent glucose entry in adipocytes mediated by glucose transporter 1

(GLUT1) and by glucose transporter 4 at the adipocyte membrane (GLUT4m). GLUT1 and GLUT4m are variables of the cellular model. GLUT1 does not depend on time in Eq. (9) because its amount is assumed to be constant according to [126, 127]. Both terms depend on interstitial glucose concentration (Gt_A , see Eq. (14)), where $KmG4$ and $KmG1$ are two parameters modelling the saturation of glucose internalisation. The last term of the equation represents glucose elimination from the adipose tissue, which depends on the amount of internalised glucose on the elimination rate k_{gluc} .

Eq. (10) describes the dynamics of plasma leptin (Y):

$$\frac{dY(t)}{dt} = b_{13}A(t)Fat - b_{14}Y(t) \quad Y(0) = Y_0 \quad (10)$$

The first term represents leptin secretion, which depends on the amount of glucose in adipose tissue (A) [97, 116], on leptin secretion rate b_{13} and on the *Fat* parameter indicating the averaged total fat mass in humans [62]. The second term models plasma leptin degradation, which depends on Y and on the constant elimination rate b_{14} according to [62, 157].

The dynamics of the ghrelin concentration in plasma (Q) is described in Eq. (11):

$$\frac{dQ(t)}{dt} = (b_{12}exp^{-lS(t)})exp^{-mI(t)} - b_{11}Q(t) \quad Q(0) = Q_0 \quad (11)$$

The first term represents ghrelin secretion, which is modelled as being exponentially inhibited both by the presence of S and I. The equation term depends also on three parameters: the S-dependent decay rate l , the ghrelin secretion rate b_{12} and the I-dependent decay rate m . The last term of the equation accounts for the linear elimination of ghrelin, which depends on the elimination rate b_{11} [160, 161].

Eq. (12) describes the glucose intake (H):

$$\frac{dH(t)}{dt} = \frac{b_{17}Q(t)}{b_{18}Y(t) + 1}exp^{-rI(t)} - b_{19}G(t)H(t) - b_9H(t) \quad H(0) = H_0 \quad (12)$$

H indicates the amount of glucose needed from the body in the current state. In order to build a close loop model, glucose intake has been modelled equal to this signal, which can be thought of as the hunger signal. The latter has been initially introduced in the rat model of Jacquier et al. [79] and here corresponding equation has been adapted to model human physiology. The first term represents the effect of plasma insulin (I), leptin (Y) and ghrelin (Q) on hunger. I and Y inhibit food intake while Q increases it. The effect of leptin and ghrelin is mediated by the parameters b_{18} and b_{17} , respectively, according to [60, 79].

Insulin exponentially inhibits H through the r parameter by modelling the negative effect on appetite which arises when the I is high. The second term describes the glucose intake reduction that depends on plasma glucose and on H itself [66]. The last term accounts for the glucose absorption to the stomach, which depends on the amount of ingested glucose and on the glucose transfer rate b_9 .

Eq. (13) and Eq. (14) provide the link between the whole body model and the molecular one, by modelling the interstitial insulin (INS_A) and the interstitial glucose (Gt_A) surrounding adipocytes:

$$\frac{dINS_A(t)}{dt} = -p_{2U}INS_A(t) + p_{2U}(I(t) - I_b) \quad INS_A(0) = INS_{A0} \quad (13)$$

$$\frac{dGt_A(t)}{dt} = -q_1Gt_A(t) + q_2(G(t) - G_b) \quad Gt_A(0) = Gt_{A0} \quad (14)$$

The first equation is derived from Dalla Man et al. [106], the second one has been written by following the same modelling approach. The first term in both the equations models degradation, according to the parameters p_{2U} and q_1 , respectively. The second term describes the amount of plasma insulin (I) and plasma glucose (G) that moves to the interstitial compartment. In both the cases, we assume that only the part exceeding the basal level can be considered, where I_b and G_b are the basal levels of insulin and glucose in plasma, respectively, and p_{2U} and q_2 are the rates regulating the transfer.

Equations from 15 to 41 describes the insulin signalling cascade in adipocytes, starting from the binding of interstitial insulin INS_A with the free insulin receptor on the adipocyte membrane (IR) and ending with the translocation of GLUT4 from the cytosol to the plasma membrane. Eqs from (15) to (39) are derived from Nyman et al. [127], while Eqs (40) and (41) are from Brannmark et al. [13]. The cellular model reconnects with the whole body model through Eq. (9), where the increase of glucose mass in adipose tissue is modelled according to the amount of GLUT4 and GLUT1 [126, 127].

The last five equations of the Nyman's model introduced in [127] were not included here, as they describe regulative phenomena related to S6 and S6K which are not relevant for the scope of the present work. Therefore, Eqs (40) and (41), which model the dynamics of S6 and S6K, were taken from Brannmark et al. [13].

All the cellular equations are modelled through mass action kinetics as from [127] and [13]. The description and the value of all the parameters are provided in Table C.2 Here we reported the list of model equations with a short description of the variables. We refer to the

next section Model simulations for the description of the cellular dynamics and to [127] and [13] for any additional insight.

Eq. (15) describes the dynamics of the free insulin receptor (IR) on the adipocyte membrane:

$$\frac{dIR(t)}{dt} = -k_{1a}IR(t)INS_A(t) - k_{1basal}IR(t) + k_{1r}IRi(t) + k_{1g}IR_YP(t) \quad IR(0) = IR_0 \quad (15)$$

Eq. (16) describes the dynamics of the phosphorylated insulin receptor (IR_YP). Phosphorylation can be insulin independent (parameter k_{1basal}) and dependent (parameter k_{1c}):

$$\begin{aligned} \frac{dIR_YP(t)}{dt} &= k_{1basal}IR(t) + k_{1c}IRins(t) - k_{1d}IR_YP(t) + k_{1g}IR_YP(t) \\ IR_YP(0) &= IR_YP_0 \end{aligned} \quad (16)$$

Eq (17) represents the dynamics of the insulin receptor that is bound to insulin but not already phosphorylated (IRins):

$$\frac{dIRins(t)}{dt} = k_{1a}IR(t)INS_A(t) - k_{1c}IRins(t) \quad IRins(0) = IRins_0 \quad (17)$$

Eq (18) represents the phosphorylated insulin receptor that has been endocytosed from the adipocyte (IRi_YP):

$$\frac{dIRi_YP(t)}{dt} = k_{1d}IR_YP(t) - k_{1f}IRi_YP(t)X_P(t) \quad IRi_YP(0) = IRi_YP_0 \quad (18)$$

Eq. (19) represents the dynamics of the free internalised insulin receptor (IRi):

$$\frac{dIRi(t)}{dt} = k_{1f}IRi_YP(t)X_P(t) - k_{1r}IRi(t) \quad IRi(0) = IRi_0 \quad (19)$$

Eqs from (20) to (23) describe the insulin receptor substrate 1 (IRS1) in its four phosphorylation forms. IRS1 is not phosphorylated, IRS1_YP is phosphorylated at the tyrosine

site, $IRS1_YP_S307P$ is phosphorylated at both the tyrosine and serine sites, $IRS1_S307P$ is phosphorylated only at the serine site:

$$\begin{aligned} \frac{dIRS1(t)}{dt} &= k_{2b}IRS1_YP(t) + k_{2g}IRS1_S307P(t) - k_{2a}IRi_YP(t)IRS1(t) - k_{2basal}IRS1(t) \\ IRS1(0) &= IRS1_0 \end{aligned} \quad (20)$$

$$\begin{aligned} \frac{dIRS1_YP(t)}{dt} &= k_{2a}IRS1(t)IRi_YP(t) + k_{2d}IRS1_YP_S307P(t) - k_{2b}IRS1_YP(t) \\ &+ k_{2c}IRS1_YP(t)mTORC1a(t)k_{fb} \quad IRS1_YP(0) = IRS1_YP_0 \end{aligned} \quad (21)$$

$$\begin{aligned} \frac{dIRS1_YP_S307P(t)}{dt} &= k_{2c}IRS1_YP(t)mTORC1a(t)k_{fb} - k_{2d}IRS1_YP_S307P(t) \\ &- k_{2f}IRS1_YP_S307P(t) \quad IRS1_YP_S307P(0) = IRS1_YP_S307P_0 \end{aligned} \quad (22)$$

$$\begin{aligned} \frac{dIRS1_S307P(t)}{dt} &= k_{2basal}IRS1(t) + k_{2f}IRS1_YP_S307P(t) - k_{2g}IRS1_S307P(t) \\ IRS1_S307P(0) &= IRS1_S307P_0 \end{aligned} \quad (23)$$

Eqs (24) and (25) represent the dynamics of the feedback protein X that, in its active form X_P , enhances the dephosphorylation of the internalised insulin receptor:

$$\frac{dX(t)}{dt} = k_{3b}X_P(t) - k_{3a}X(t)IRS1_YP(t) \quad X(0) = X_0 \quad (24)$$

$$\frac{dX_P(t)}{dt} = -k_{3b}X_P(t) + k_{3a}X(t)IRS1_YP(t) \quad X_P(0) = X_P_0 \quad (25)$$

Eqs from (26) to (29) describe the four different forms of the protein kinase b: not phosphorylated (PKB), phosphorylated only at the threonine site (PKB_T308P), only at the serine site (PKB_S473P) and at both sites (PKB_T308P_S473P):

$$\frac{dPKB(t)}{dt} = -k_{4a}PKB(t)IRS1_YP(t) + k_{4b}PKB_T308P(t) + k_{4h}PKB_S473P(t) \quad (26)$$

$$PKB(0) = PKB_0$$

$$\begin{aligned} \frac{dPKB_T308P(t)}{dt} &= k_{4a}PKB(t)IRS1_YP(t) - k_{4b}PKB_T308P(t) + \\ &-k_{4c}PKB_T308P(t)mTORC2a(t) \quad PKB_T308P(0) = PKB_T308P_0 \end{aligned} \quad (27)$$

$$\begin{aligned} \frac{dPKB_S473P(t)}{dt} &= -k_{4e}PKB_S473P(t)IRS1_YP_S307P(t) - k_{4h}PKB_S473P(t) \\ &+k_{4f}PKB_T308P_S473P(t) \quad PKB_S473P(0) = PKB_S473P_0 \end{aligned} \quad (28)$$

$$\begin{aligned} \frac{dPKB_T308P_S473P(t)}{dt} &= k_{4c}PKB_T308P(t)mTORC2a(t) - k_{4f}PKB_T308P_S473P(t) \\ &+k_{4e}PKB_S473P(t)IRS1_YP_S307P(t) \\ PKB_T308P_S473P(0) &= PKB_T308P_S473P_0 \end{aligned} \quad (29)$$

Eqs (30) and (31) describe the protein complex mTORC1 (mammalian target of rapamycin mTOR in complex with raptor) in its inactive (mTORC1) and active (mTORC1a) forms:

$$\begin{aligned} \frac{dmTORC1(t)}{dt} &= k_{5b}mTORC1a(t) - mTORC1(k_{5a1}PKB_T308P_S473P(t) + \\ &+k_{5a2}PKB_T308P(t)) \quad mTORC1(0) = mTORC1_0 \end{aligned} \quad (30)$$

$$\begin{aligned} \frac{dmTORC1a(t)}{dt} &= -k_{5b}mTORC1a(t) + mTORC1(k_{5a1}PKB_T308P_S473P(t) + \\ &+ k_{5a2}PKB_T308P(t)) \quad mTORC1a(0) = mTORC1a_0 \end{aligned} \quad (31)$$

Eqs (32) and (33) represent the protein complex mTORC2 (mammalian target of rapamycin mTOR in complex with rictor) in its inactive (mTORC2) and active (mTORC2a) forms:

$$\frac{dmTORC2(t)}{dt} = -k_{5c}mTORC2(t)IRi_YP(t) + k_{5d}mTORC2a(t) \quad (32)$$

$$mTORC2(0) = mTORC2_0$$

$$\frac{dmTORC2a(t)}{dt} = k_{5c}mTORC2(t)IRi_YP(t) - k_{5d}mTORC2a(t) \quad (33)$$

$$mTORC2a(0) = mTORC2a_0$$

Eqs (34) and (35) describe AS160, the substrate of PKB, and its phosphorylated form AS160_T642P:

$$\begin{aligned} \frac{dAS160(t)}{dt} &= k_{6b}AS160_T642P(t) - AS160(t)(k_{6a1}PKB_T308P_S473P(t) + \\ &+ k_{6a2}PKB_S473P(t)) \quad AS160(0) = AS160_0 \end{aligned} \quad (34)$$

$$\begin{aligned} \frac{dAS160_T642P(t)}{dt} &= -k_{6b}AS160_T642P(t) + AS160(t)(k_{6a1}PKB_T308P_S473P(t) + \\ &+ k_{6a2}PKB_S473P(t)) \quad AS160_T642(0) = AS160_T642_0 \end{aligned} \quad (35)$$

Eqs (36) and (37) represent glucose transporter 4 inside the adipocyte cytosol (GLUT4) and on the cell membrane (GLUT4m):

$$\frac{dGLUT4m(t)}{dt} = k_{7a}AS160_T642P(t)GLUT4(t) - k_{7b}GLUT4m(t) \quad (36)$$

$$GLUT4m(0) = GLUT4m_0$$

$$\frac{dGLUT4(t)}{dt} = -k_{7a}AS160_T642P(t)GLUT4(t) + k_{7b}GLUT4m(t) \quad (37)$$

$$GLUT4(0) = GLUT4_0$$

Eqs (38) and (39) describe the dynamics of the S6 kinase (S6K) and its phosphorylated form S6K_T389P:

$$\frac{dS6K(t)}{dt} = k_{9b}S6K_T389P(t) - k_{9a}mTORC1a(t)S6K(t) \quad S6K(0) = S6K_0 \quad (38)$$

$$\frac{dS6K_T389P(t)}{dt} = -k_{9b}S6K_T389P(t) + k_{9a}mTORC1a(t)S6K(t) \quad (39)$$

$$S6K_T389P(0) = S6K_T389P_0$$

Eqs (40) and (41) represent the ribosomal protein S6 and its phosphorylated form S6_S235_S236P:

$$\frac{dS6(t)}{dt} = k_{9b2}S6_S235_S236P(t) - k_{9f2}S6(t)S6K_T389P(t) \quad S6(0) = S6_0 \quad (40)$$

$$\frac{dS6_S235_S236P(t)}{dt} = -k_{9b2}S6_S235_S236P(t) + k_{9f2}S6(t)S6K_T389P(t) \quad (41)$$

$$S6_S235_S236P(0) = S6_S235_S236P_0$$

Initial values of all the model variables are listed in Table C.1, for both NGR and T2DM conditions. They have been derived from the literature or obtained by nonlinear optimisation constrained to the variability of physiological ranges. Initial values have been selected to find the best balance between human physiology and reliability of the model dynamics in order to avoid discontinuities or states with unrealistic variable values, such as negative values or values outside the physiological ranges listed in Table 5.1. For what concerns the cellular model, the variability range of model variables used during the optimisation has been inferred *in silico*: Nyman's model has been simulated according to [127] and the minimum and maximum values reached by each variable have been considered.

Simulations start at time 0, at which is considered as morning fasting state. In order to represent equally the physiological ranges for each glucose intake period, we set $S_0 = 4mmol$

Variable Description	Symbol	Range	Reference
Plasma glucose concentration	G	4.5 - 11 <i>mM</i>	[161] [58]
Plasma insulin concentration	I	38 - 400 <i>pM</i>	[71]
Plasma incretin concentration	W	5 - 50 <i>pM</i>	[146]
Plasma glucagon concentration	E	28.68 - 47.04 <i>pM</i>	[71]
Liver glucose mass	C	0 - 8 <i>mmol</i>	[136]
Muscle tissue glucose mass	M	2 - 13 <i>mmol</i>	[136]
Adipose tissue glucose mass	A	30 - 120 <i>mmol</i>	[136]
Plasma leptin concentration	Y	0 - 0.6 <i>nM</i>	[4]
Plasma ghrelin concentration	Q	8 - 146 <i>pM</i>	[45]
Amount of glucose intake	H	N.A.	
Stomach glucose mass	S	N.A.	
Intestine glucose mass	L	N.A.	

Table 5.1 Physiological ranges of the whole body variables in the normal glucose regulation (NGR) condition. The ranges for H, S and L have not been specified (N.A. not available), as there is no clear upper limit to the amount of food (or glucose) one could ingest

and $L_0 = 10\text{mmol}$ in the model. The amount of glucose needed from body is set to $H_0 = 200\text{mmol}$, derived from the amount of needed glucose reported in [41].

Fasting plasma glucose is set to $G_0 = 5\text{mM}$ and fasting plasma insulin to $I_0 = 60\text{pM}$ as reported in literature by [134]. Moreover, plasma incretin fasting level is assigned to $W_0 = 10\text{pM}$ [77], fasting plasma glucagon level is $E_0 = 37.7\text{pM}$ [71], plasma ghrelin fasting level $Q_0 = 120\text{pM}$ and basal plasma leptin concentration $Y_0 = 0.4\text{nM}$ [92] [45]. Initial value for fasting glucose mass in muscle tissue M_0 was reported by Pratt et al. [136] to be 0.1mmol/l . The unit of measurement of this value was changed to be consistent with adipose tissue glucose mass and the other system variables. For this reason, we calculated the glucose mass in muscle tissue, M_0 as 25mmol . It is known that volume of skeletal muscle is 25l , and skeletal muscle tissue density is 1.06kg/l [136]. In the NGR simulation, skeletal muscle was estimated to be 40% of body weight. Same method is applied for the initial value of glucose mass in liver. After converting units from Pratt et al. [136], the glucose mass in liver ready to be secreted has been set to $C_0 = 3\text{mmol}$. Adipose tissue glucose mass initial value was computed taking into account the presence of the molecular level. Therefore, the initial value A_0 was set to 53.19mmol in order to remain physiological range for each oscillation as reported in [136].

There are not many reported initial values in literature for the interstitial compartment because of difficulty of calculating with experiments. Therefore, Nyman et al. [126] reported an initial value for interstitial insulin (INS_A) of 0pM . However, such a small value can

create some negative values for INS in the first time of the simulation. For this reason, in this model the initial value of INS_A was increased to 20 pM , the minimum value allow to avoid negative values in the first part of the simulation. However, interstitial glucose initial value Gt_{A0} was set to 135 mg/kg as reported by Dalla Man et al. [106]. For the cellular part of the hierarchical model all initial values were taken from the model of Nyman et al. [127] and then re-optimised in the physiological range according to their value when interstitial insulin goes back to INS_{A0} .

In order to simulate T2DM condition, initial values of model variables were modified again according to the literature and the model behaviour. By following the similar strategy as in NGR case, fasting stomach and intestine glucose are set to $S_0 = 14 \text{ mmol}$ and $L_0 = 25 \text{ mmol}$, respectively. The amount of ingested glucose is set to $H_0 = 240 \text{ mmol}$ derived from the amount of needed glucose per day in T2DM case [41]. Fasting plasma glucose was reported to be $G_0 = 7.5 \text{ mM}$ and fasting insulin $I_0 = 180 \text{ pM}$ for diabetic individuals [161]. Since plasma incretin levels in diabetes are controversial, its fasting value is set equal to as in the healthy case ($W_0 = 10 \text{ pM}$). Fasting plasma glucagon is known to be higher in T2DM and it is reported to be $E_0 = 42 \text{ pM}$ in [71]. Instead, plasma ghrelin in T2DM is lower and is set to $Q_0 = 60 \text{ mM}$ as reported in [45]. Fasting plasma leptin was set to $Y_0 = 1.93 \text{ nM}$ due to the model behaviour oscillations. Fasting value for glucose mass in muscle tissue and liver remain unchanged in T2DM case. Adipose tissue glucose mass is affected directly by the presence of the molecular level in the model. Initial value A_0 is set to 32 mmol taking into account the molecular model dynamics in T2DM and physiological behaviour.

For the interstitial compartment there is a lack of data in the literature for diabetics condition. For this reason, interstitial insulin initial value INS_0 is set to 70 pM and interstitial glucose Gt_0 to 216 mg/kg making a proportion between initial values for plasma and interstitial levels. The proportion was made to maintain the same ratio, between plasma value and interstitial value, in healthy also for diabetic. For the cellular level, as reported from Nyman et al. [127], the sum of insulin receptor (in all its forms) decreases from 100% to 55% in T2DM compared to the healthy case. Similarly, the total number of glucose transporter 4 (GLUT4 and GLUT4m) is half in T2DM compared to the healthy case. All initial value for molecular state variables were computed as for the healthy case according to their value when interstitial insulin goes back to INS_{A0} .

Parameter estimates have been determined following different methodologies. When a reference from the literature was available, parameter estimates were directly taken from the literature or derived by following the same procedure indicated in the reference paper. We refer to Table C.2 for any further insight on the employed estimation procedures and for a complete list of all parameter estimates computed for the NGR and T2DM conditions.

Parameters b_4 , b_5 , b_{12} , b_{13} , b_{17} , b_{27} , c and c_3 have been estimated through steady state analysis, that is, by imposing a steady state condition at time 0 on the corresponding equations as indicated in the literature. In the other cases, parameters have been estimated by nonlinear optimisation constrained to obtain values within the physiological ranges discussed in the literature. The remaining parameters b_9 , r , k_{gluc} , q_1 and q_2 were derived by unconstrained optimisation to obtain a model dynamics consistent with physiology, that is, a dynamics without discontinuities or unrealistic variable values (*i.e.*, negative values or values outside physiological ranges). For what concerns the cellular model, parameters are all taken from [127] and [13], except for k_{1a} , which has been re-estimated within the same optimisation range of [127] in order to have the minimum value of the IRins dynamics of Eq. (17) consistent with that reported by Lodish et al. [101] and to preserve the model fits introduced by Nyman et al. [127] and provided in Figures 5.5 and 5.6.

5.3 Results

The model has been simulated for 1000 minutes (three consecutive meals) starting from an initial condition representing the fasting state ($t=0$). The model initial values and parameter estimates have been computed as introduced in Methods and they are reported in Tables C.1 and C.2, respectively. Figure 5.3 shows the dynamics of each model variable at the whole body level. The green and black lines represent the NGR and T2DM conditions. The physiological upper and lower ranges for each variable are shown in blue (higher line, HL) and red (lower line, LL) straight lines, according to the available estimates from the literature (see also Table 5.1). The model exhibited an oscillatory behaviour in both the NGR and T2DM conditions through alternate parameter sets that accounted for the reciprocal interaction among the constituting variables in the two conditions.

In order to define a closed loop model, the output signal representing the amount of glucose needed from the body was connected with the input signal representing oral glucose uptake. This allowed to simulate the body's behaviour by means of a system that auto feeds itself without the need of external intervention. System variables exhibit an oscillatory behaviour that is consistent with physiology by considering both each variable in respect of the others, and the concentration ranges computed during the simulation. It can indeed be observed that the model dynamics in the healthy case have all system variable fluctuations within the indicated physiological range.

In the NGR condition, the system started at fasting by simulating an oral glucose intake (Figure 5.3.1) and, after a time lag accounting for the transit time among compartments, was followed by subsequent transitions through the stomach (Figure 5.3.2) to the intestine (Figure

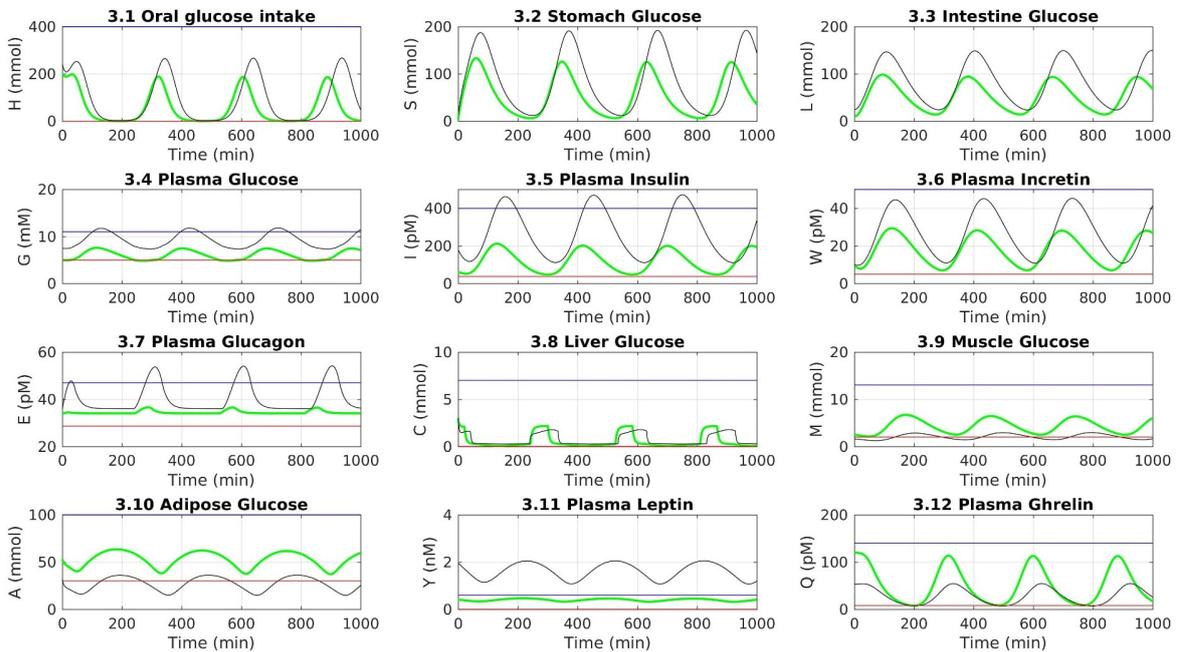


Fig. 5.3 Model dynamics at the whole body level. Each plot represents one variable dynamics. The normal glucose regulation (NGR) and T2DM conditions are shown in green and black, respectively. The red and blue lines delimit the physiological lower and upper ranges of variables (see also Table 5.1)

5.3.3). The glucose absorption from the intestine to the bloodstream was characterised by a further time lag (Figure 5.3.4) and it triggered the increase in circulating insulin levels (Figure 5.3.5). We modelled the glucose transit through the intestine as a stimulus for the secretion of incretins (Figure 5.3.6), which ultimately resulted in an amplification of the endogenous insulin secretion. The secretion of glucagon (Figure 5.3.7) was modelled as being inhibited by high glucose and insulin concentrations, and increased in case of markedly low plasma glucose levels, thus stimulating the endogenous glucose output from the liver (Figure 5.3.8). The glucose uptake by the muscle and adipose tissues was favoured in case of high insulin concentrations, thus resulting in a net increase of the glucose mass in these tissues (Figures 5.3.9 and 5.3.10).

The whole-body model has been linked to the adipocyte cellular level through the interstitial fluid surrounding the individual cells. Here we assumed the interstitial fluid being in direct communication with the plasma. Therefore, the two layers of abstraction (*i.e.* the whole-body and the adipocyte levels) were bound through the interstitial insulin and glucose (INSA and GtA, Figure 5.2) in a conceptual framework closely mirroring the physiology

of insulin signalling. The interstitial fluid, rather than the plasma, is surrounding the cells targeted by insulin, it flows in our model from the plasma to the interstitial space with a time shift (Figure 5.4), and then binds its membrane receptors, triggering the downstream cascade of signalling events. The inactive and un-phosphorylated components of the cascade, such as the unbound insulin receptor (IR) or the un-phosphorylated protein kinase B (PKB), proportionally decreased at incremental concentrations of interstitial insulin (Figures 5.4.3 and 5.4.7). The opposite occurred to the active and phosphorylated components, such as the bound insulin receptor (IRins) or the phosphorylated insulin receptor substrate 1 (IRS1_YP), which increased at higher levels of interstitial insulin (Figures 5.4.4 and 5.4.6), thus allowing the activation of the insulin signalling cascade and eventually leading to the translocation of glucose transporter type 4 (GLUT4) to the cell membrane. The output of the cellular layer was linked to the whole-body output through the amount of GLUT4 on the adipocyte membrane, which directly affected the glucose uptake by the adipose tissue (Figure 5.4.11). This latter variable was in fact shared between the two layers together with interstitial insulin and glucose.

In the T2DM condition, the initial value of several model variables, such as plasma glucose and insulin, and some of the parameters, were modified, as described in Methods and reported in Tables C.1, in order to simulate a T2DM condition of a drug-naïve individual patient. It can be observed that, as compared to NGR, glucose dynamics (Figures 5.3.1 to 5.3.4) showed a broader range in the T2DM condition, as well as that of insulin, incretin and glucagon (Figures 5.3.5 to 5.3.7). The total mass and the output rate of hepatic glucose production were reduced (Figure 5.3.8), as well as glucose uptake by muscle and adipose tissues (Figures 5.3.9 and 5.3.10). According to T2DM pathophysiology, insulin effectiveness was reduced in the T2DM condition. Despite higher absolute insulin levels, the glucose uptake mechanism was impaired in the muscles, adipocytes and liver. The impaired insulin signalling cascade affects glucose uptake, resulting in decreased glucose uptake by the adipose tissue and leptin secretion. An increased leptin concentration was observed in the T2DM experimental condition (Figure 5.3.11), as the individual fat mass was set to an increased level, according to Grasman et al. [62].

The estimates of some model parameters, including the number of IRs and GLUT4, and the positive feedback from mTORC1, have been modified, according to Nyman et al. [127], to simulate the T2DM condition at the cellular level. The diminished total number of IRs led to reduced IR binding and phosphorylation (Figures 5.4.3, 5.4.4 and 5.4.6). Similarly, the reduced total concentration of GLUT4 affected the amount of GLUT4 eventually docking to the adipocyte membrane (Figure 5.4.9). The reduced positive feedback from mTORC1 had a more general regulatory effect on all the components of the insulin signalling cascade starting

from a lower level of IRS1_YP, where the protein complex mTORC1 acts directly (Figure 5.4.6). All changes applied in the previously described model parameters harmonically worked together to reduce the amount of GLUT4, eventually leading to reduced glucose uptake by the adipocytes (Figure 5.4.11).

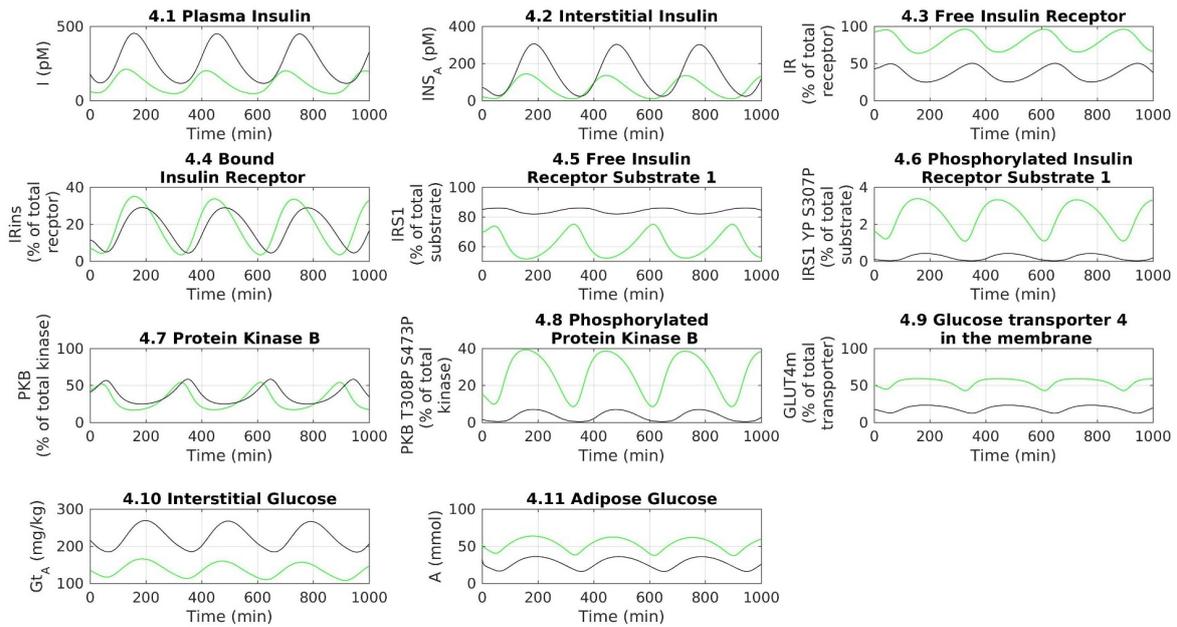


Fig. 5.4 Model dynamics of the insulin signalling in adipocytes at the cellular level (only a subset of key variables is represented). The normal glucose regulation (NGR) and T2DM conditions are shown in green and black, respectively.

In order to test the consistency of the model with physiology, we fitted the same experimental values employed by Nyman et al. [127] at the cellular level in both the NGR and T2DM conditions, as shown in Figures 5.5 and 5.6, respectively. According to the physiology governing the reciprocal interactions among leptin, insulin and ghrelin, we were also able to roughly reproduce the fluctuations of circulating ghrelin levels, which are typically characterised in humans by a marked reduction after meal ingestion and by a rebound to baseline before the next meal [46,47]. Of note, although we were unable to capture, by design, the circadian fluctuations of ghrelin (which usually increases after an overnight fast) and other hormones, our model successfully reproduced the physiologic dynamics of ghrelin by inversely paralleling those of insulin (Figures 5.3.5 and 5.3.12). Conversely, in accordance with the role of leptin as prototypical regulator of energy homeostasis and its dependence from adipose tissue mass, the dynamics of leptin returned by the model in both T2DM and

NGR conditions showed much dampened fluctuations, as compared to those of ghrelin and insulin (Figure 5.3.11).

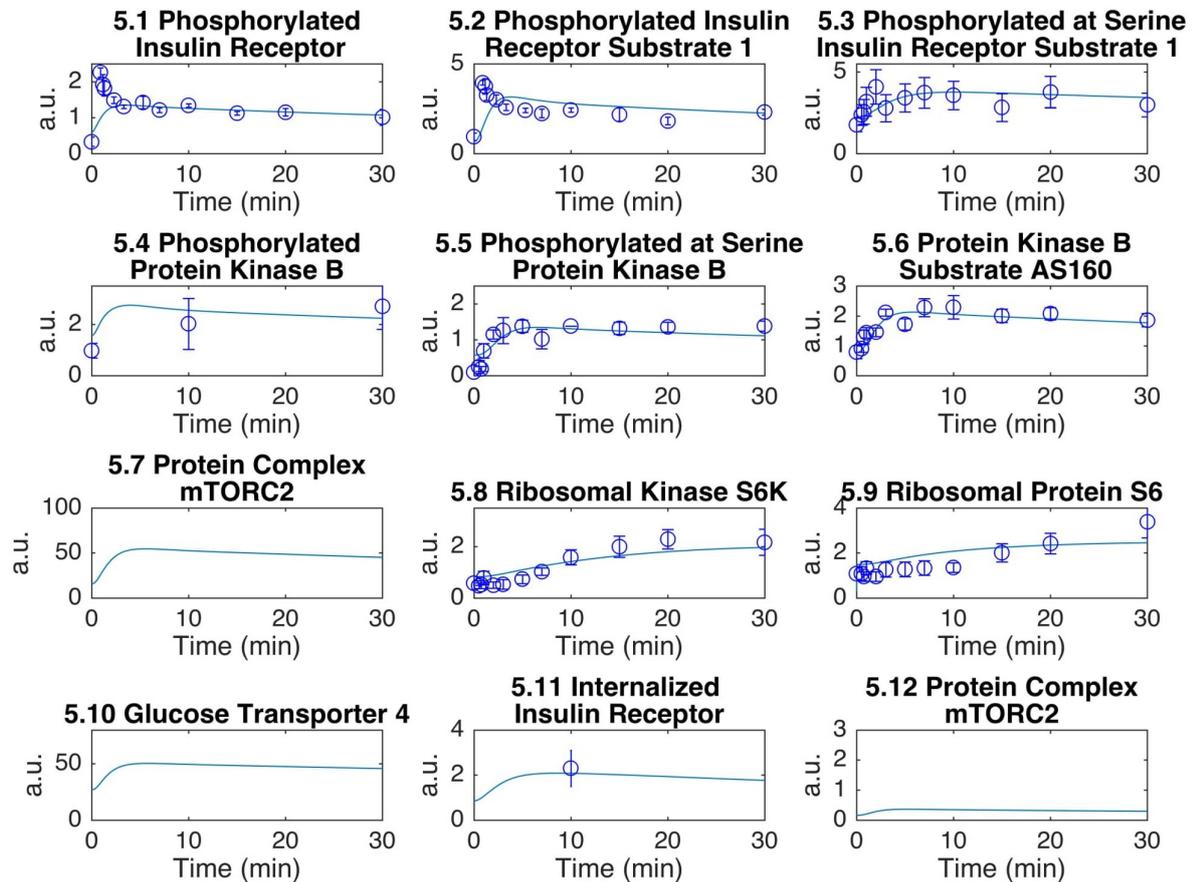


Fig. 5.5 Model simulation and data fitting, normal glucose regulation (NGR) condition. Each plot represents the corresponding time courses for the indicated insulin signalling intermediaries. The experimental data are taken from Nyman et al. [127] and are represented with circles and error bars (a.u. indicates arbitrary units). The time course represents the model simulation.

5.4 Discussion

The aim of this work was to introduce a closed-loop multi-level model of human glucose homeostasis, describing, in a hierarchical multi-scale system architecture, the contribution of its main determinants to the NGR and T2DM conditions, at both the whole-body and cellular (adipocyte) levels. The modelling strategy merged two different physiological levels, the

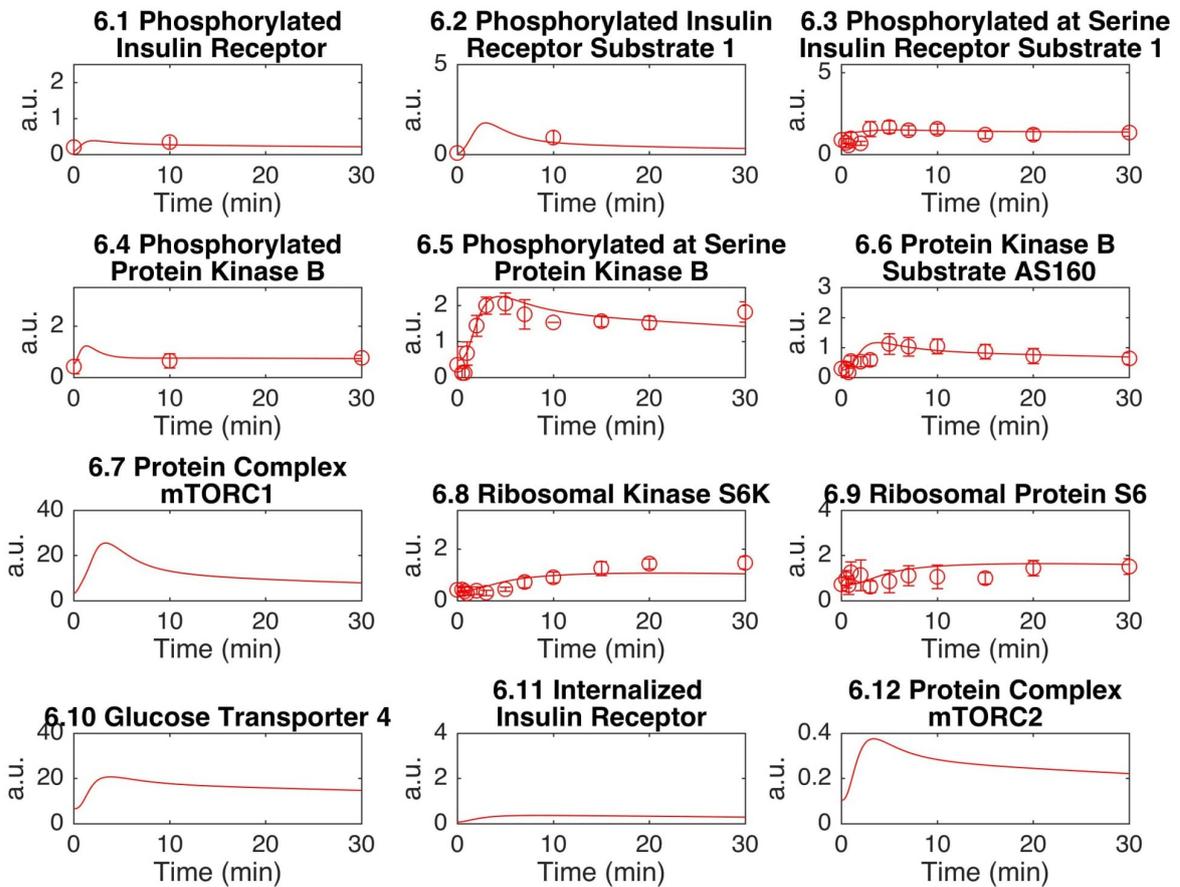


Fig. 5.6 Model simulation and data fitting, T2DM condition. Each plot represents the corresponding time courses for the indicated insulin signalling intermediaries. The experimental data are taken from Nyman et al. [127] and are represented with circles and error bars (a.u. indicates arbitrary units). The time course represents the model simulation.

organ and the cellular one, to ground the basis for the inclusion of the other main players in glucose homeostasis (such as muscle, pancreatic and liver cells) as additional compartments of the cellular level.

The model was successfully tested *in silico* for NGR and T2DM conditions, which have been described through alternate initial conditions and parameter estimates. The output of the model (whole-body glucose needs) coincided with the input (oral glucose intake) in a closed-loop fashion, which allowed to perpetually simulate whole body dynamics, according to a self-feeding system. Whenever possible, model equations have been directly derived from the literature and then adapted in order to work together through the identification of suitable initial states and parameter estimates. The shape of model equations has been left

unchanged in most instances for consistency with those introduced in the original works. This approach provides the following advantages: (i) it allows to rely on very established model equations, that have been extensively analysed in the literature to describe the physiology of interest; (ii) it does not require to estimate *de novo* several parameters (as it would be required in case of equation reshaping), which can rather be derived directly from the literature; (iii) it allows a fairer comparison of the results herein presented with those already discussed in the literature. However, this modelling strategy has the disadvantage that some equations could look different, even if they model similar processes. For example, saturation has been modelled either by considering Michaelis-Menten terms as in Eq. (9) or by relying on more abstracted exponential terms as in Eq. (11). We remark that this discrepancy does not affect the reliability of model simulations because the set of ODEs has been parameterised to have all model variables within their physiological ranges during the simulations. Therefore, model equations are computed in the same conditions considered in the papers where they have been originally introduced.

The closed-loop system, which has the unique advantage to simulate experimental conditions for long time windows without external intervention, was achieved through the inclusion of the hunger signal, here intended as the amount of glucose needed from the body. Hunger description was possible through leptin and ghrelin, which work as complementary molecules to regulate food intake and energy balance in close concert to insulin [28, 163]. Ghrelin is a fast-acting hormone secreted when the stomach is empty [9] and it stimulates food intake. Leptin concentration depends on fat mass [84] and acts on the long term, without major changes within hours or days (Figure 5.3.11), as a “satiety” signal to the brain [8, 50]. Of note, although T2DM individuals are often characterised by increased leptin concentrations due to increased fat mass, a mechanism of leptin resistance occurs, thus making them relatively insensitive to leptin [143].

The dual role of the liver as both glucose storage and production site is a novel feature of our model, since the inclusion of both the regulative effects of insulin and glucagon has never been considered in previous models [86, 106]. Glucagon is a hormone secreted from pancreatic alpha cells at low glycemic conditions, which signals the liver to release glucose from glycogen storages, thus maintaining the euglycaemic state at fasting [125]. Sub-diabetic hyperglycemic states and overt T2DM are often characterised by high fasting plasma glucose levels, due to an excessive glucose output from the liver, as a consequence of liver insensitivity to insulin or abnormally high glucagon [[30].

Although a number of models have been previously developed by including the dynamics of glucagon [20, 86, 154], incretins [15, 160, 161], leptin [132] and ghrelin [161], we have considered all these factors together for the first time. The inclusion of these components, as

well as the distinction of adipose and muscle glucose uptake, previously reported together [154], provided a better description of the reciprocal connections existing between the whole-body and the cellular levels and allowed us to physically “close the loop” among different layers of abstraction.

Of note, we have observed that, as compared to the NGR state, the T2DM condition in the model displayed slower dynamics. Although the reciprocal changes in ghrelin (decrease) and insulin (increase) would drive the need for increased glucose intake, the resulting emptying rate of the stomach is slower, as well as the glucose uptake from adipose tissue and muscles, thus leading to a delayed dynamics in the whole system. These dynamics may be interpreted in light of the constituting principle regulating the model, which takes into account the supposed individual energy requirements. Therefore, our model correctly reflects the physiological response of the organism to maintain the glucose homeostasis within a physiological range in case of an imbalance between energy requirements and energy intake (*i.e.* overfeeding) often observed in patients with T2DM.

Our study has however a number of limitations that should be addressed. We do acknowledge the peculiar use of the term “closed-loop” herein employed, which is usually linked, in the dictionary of the diabetes community, to the so-called “artificial pancreas”, *i.e.* a system that is able to automatically predict *in silico* the adjustments of external insulin delivery needed to keep the circulating plasma glucose within a narrow range of physiologic fluctuations. In the context of the present study, we were not limiting the term “closed-loop” to a sort of “artificial beta cell”. We rather sought to describe the governance of glucose homeostasis by applying a broader, holistic approach. It should however be pointed out that our modelling effort lacks of a comprehensive description of the hormonal networks and molecular cascades occurring at each organ and tissue involved in the regulation of glucose homeostasis. The rationale surrounding our choice of focusing on the intracellular molecular cascade occurring within the adipocytes was not solely dictated by the relevance of the adipose tissue as an “endocrine organ” influencing systemic energy balance and glucose homeostasis [24,25], but also by the availability of detailed mathematical specifications of the insulin signalling cascade recently provided by the group of Nyman et al. [127].

Furthermore, it could be argued that a whole-body description similar to the one herein proposed can be defined without the addition of a cellular level. However, the integration of the two levels of abstraction within a single model allows the detailed observation of the reciprocal effects of changes occurring between the constituents of the cellular and whole-body levels. Thus, the hierarchical modelling strategy allows to simply zoom in on specific areas of interest (in our case, the adipocytes) in order to investigate regulatory effects that may occur between the two levels of abstraction. For instance, the action of a molecule

on a receptor could be easily included in the specifications of the cellular level, and its effects at the whole-body level could be observed, thus allowing the identification of the changes in the organ variables caused by variations of the cellular ones. Therefore, the addition of a level in a hierarchical modelling structure does not imply that the rest of the model would not stand by itself, but it is rather there to allow the consideration of other (e.g. molecular) effects within a wider framework.

The functioning of the biological systems involve various structural levels. It includes proteins, cells, tissues, organs and whole body process and interactions between them. The multi-level nature represents a challenge for researchers and clinicians. Multi-level modelling can provide a conceptual framework for the integration of information and improve the capability of generating experiments and testing hypotheses about systems.

Traditional approaches to modelling focus on one level and it is assumed that processes at other levels are negligible from the point of view of model variables. However, for the cases where strong interconnections exist between levels, a function at a given level can depend on the sublevels but also on higher levels. In such a context, the need to model at multi-level is preferable due to physiological interactions of the system. Multi-level modelling approach offers an alternative view to understand the integration of information from cellular, tissue and organ levels up to the whole body level. With the aim of gaining a deeper insight over biological complexity, modelling the interaction of the smaller scales with larger scales can support the understanding of the biological system as well as the experimental data. It enhances the clinical decision-making, allow predictions and improve the diagnosis for individual patients. This way, multi-level approach provides a detailed understanding of the relation between the whole body and cellular levels. Without properly addressing this fundamental relation, controlling the disease progression can be slower and insufficient.

Nevertheless, despite its intrinsic limitations, our hierarchical modelling effort demonstrated sufficient robustness to provide a fair description of the core determinants of glucose homeostasis at both cellular and systemic scales. As such, given its unique modular architecture, the multi-level model herein tested constitutes a promising backbone to annex further layers of detail.

The model describes the normal glucose regulation and the diabetic states through alternate parameter sets, where the constraints for those parameters have been mostly derived from the available literature as detailed in Materials and Methods. This approach is blind to distinguish between primary changes that drive disease progression and secondary changes that are consequences or adaptations to the primary ones. Other independent research groups, such as Topp et al. [162] and De Gaetano et al. [32], have addressed, at variance with our approach, the argument of glucose homeostasis regulation from a pathogenesis modelling

standpoint. Topp et al. [162] for instance, have specifically investigated the link between beta-cell mass and beta-cell function by a set of nonlinear ODEs, where the glucose and insulin dynamics are designed to be fast relative to beta-cell mass dynamics. On the contrary, when we compared with our model the estimates obtained in the T2DM condition to those from the NGR state, it was impracticable to distinguish whether the reduction in the secretory capacity (parameter b_4) reflected reduced beta-cell mass or function, or both. Of note, even in the absence of a pre-specified and detailed mathematical description of the beta-cell function machinery, the value of this parameter was not imposed in advance, but it rather represents a consequence of the steady state analysis applied to the equation modelling insulin concentration in plasma (Eq. (4)), according to what introduced by Toghaw et al. [161]. As previously anticipated, and similarly to the case of other key determinants of glucose homeostasis, this encouraging result may be considered as a rough indicator of the goodness of the model, which leaves the beta-cell component open to further hierarchical refinements.

5.5 Conclusion

Coupling the cellular level model with a closed-loop whole body model allowed us to evaluate the behaviour of adipocytes not only during one meal but in a perpetual fashion. The simulation of the system over such a long time frame highlighted the reciprocal reactions occurring between the two levels of abstraction, *i.e.* the organ and the cellular levels. The model provided a seamless dynamic description of the molecular mechanisms downstream the insulin receptor in adipocytes, thus demonstrating the usefulness of a multi-level approach to the modelling of glucose homeostasis at both cellular and systemic scales. As for the potential applications, the herein proposed model architecture is intrinsically open to integrate supplementary layers of specifications for individual components. As such, more detailed and advanced versions of the present model could potentially be applied to investigate *in silico* the effect of specific drugs pointing to one or more of the model constituents or to identify currently unmet molecular targets amenable to pharmacological intervention.

Author contributions

I, Luca Marchetti, Marco Dauriz, Alice Matone and Corrado Priami designed and tested the whole-body model implemented by me. I, Giulia Simoni, Luca Marchetti, Marco Dauriz, Alice Matone and Corrado Priami designed and tested the cellular model implemented by Giulia Simoni. Figures 5.4, 5.5 and 5.6 are produced by Giulia Simoni under the supervision of all authors. All authors contributed to the integration of the whole-body model and the cellular model as well as to the writing of the PloS ONE paper manuscript.

Chapter 6

Discussion and Future Directions

Biological systems consist of large numbers of interacting components and involve processes operating across a variety of spatial, temporal and biological scales. Systems biology aims at facilitating the global understanding of natural living systems by focusing on complex interactions within biological systems. Synthetic biology takes advantage of this better understanding and focuses on the design, construction and optimisation of new-to-nature biological functions. Mathematical and computational modelling constitute a part of the systems and synthetic biology visions, these two interdependent fields provide a method for formally defining and analysing the structure and function of a biological entities as engineered systems. Along these lines, the results in this dissertation have addressed the analysis and design of time dependent signalling mechanism and the study of multi-level models by the application of systems and synthetic biology methods. The models described here, as models in general, rest on assumptions and abstractions. They, thus, may be refined or extended according to the context they are used in. We discuss in this chapter the contributions of this research and some of the possible developments and further extensions.

Pho regulon in *E. coli*. Water pollution is one of the universal problems with great impact on environment. Industrial and daily life waste contains inorganic and organic matters, such as inorganic phosphate, which cause the contamination of groundwater [100]. In this respect, bioreactors can play significant role to produce polished water from waste water. In this part of my Ph.D. research, in a collaboration with our partners in the European project Living Architecture (LIAR) we provide a modelling framework, which combines protein-protein interaction, signal transduction and transcriptional regulatory network. This model provides a detailed description of inorganic phosphate regulation of *E. coli*. *E. coli* can take inorganic phosphate and store excess phosphate in the form of polyphosphates [168, 169]. When the environmental inorganic phosphate levels are scarce, *E. coli* will spend energy to upregulate

expression of genes used to acquire inorganic phosphate from the environment. These genes comprise the Pho regulon, which is controlled by a two component system, PhoR and PhoB, Pst system (ABC transporter PstSCAB) and PhoU protein. Some of the protein-protein interactions of these 7 proteins in phosphate signal transduction has remained indeterminate in the literature.

We proposed models that describe the interactions between Pst system, PhoU and PhoR, whereby the external inorganic phosphate level changes from abundant to limited. The aim of this study is to answer questions on the type of interaction between Pst system and PhoU and between PhoR and PhoU. When the environmental inorganic phosphate levels are high, the phosphate signal is inhibited and *E. coli* stops to spend energy for the intake of inorganic phosphate. This causes the down-regulation of the gene expression [168, 169]. It is known that PhoU mediates the formation of a signalling complex between the Pst system (ABC transporter) and PhoR [56]. However, it is not known how PhoU functions or interacts within the signalling pathway. PhoU takes active role in inhibition of the system.

The primary technique we have used to characterise the function of PhoU is a control model, which has the PhoU effect implicitly. The control model describes a direct interaction mechanism between Pst and PhoR for different external inorganic phosphate concentrations. We describe the interactions within a chemical reaction network representation and use the standard translation from chemical reaction networks to ODEs based on stoichiometry and reaction rates. To better study how these interactions might affect the signal transduction, we have proposed alternative reactions to analyse the protein-protein interactions. The implicit control model with alternative reactions provides a simple setting to study the deactivation of PhoR and the consequent response in terms of inorganic phosphate intake. As a further step, we have simulated the system with an explicit model variable for PhoU, which acts as a messenger for PhoR sensor histidine kinase. Alternative interactions as described in control model have been kept the same. We have been able to identify possible protein-protein interactions between PhoU and Pst, as well as PhoR. This method of testing the interaction of PhoU within the Pho regulon has provided a quantitative description of how proteins interact. Moreover, it allowed us to construct the TCS switch off mechanism as a function of the change of the external inorganic phosphate concentration. For our purposes, this system is an effective confirmatory tool to verify the activation of PhoU protein and deactivation mechanism of Pho regulon.

Besides determining, Pst, PhoR and PhoU interactions and the consequent Pho regulon inhibition, we have established a model to analyse the active Pho regulon. It is known that when Pho regulon is active PhoU does not interact with PhoR and Pst. This causes *E. coli* intake inorganic phosphate and up-regulate the gene expression [168, 169]. We have used a

combination of deterministic and stochastic simulations to provide a dynamic account of the mechanistic interplay between the phosphate starvation response system variables. We have validated the model experimental data on PhoA and PhoB promoters. Moreover, we have used our model to explore the effect of synthetic promoters to develop a multi-task consortia based on signalling and regulatory mechanisms for different levels of inorganic phosphate concentration. We have quantified the effect of each single protein in the pathway through sensitivity analysis, identifying the main regulatory reactions of the two component system. We have used the available experimental data in literature as well as our own experimental data to identify the parameters of the model for corresponding variables.

With these insights, this work offers a mechanistic understanding of inorganic phosphate intake, realised by Pho regulon in a way that connects signalling with the genetic level. It provides a quantitative description of how different proteins interact to form a biological control system. It also describes the control of the phosphate-starvation response at the genetic level. Our work provides measurements of protein, phosphorylation, and promoter activity levels that are fundamental to define features of TCS circuits. One of the major outcomes is that our results do not only explain the observed experimental data, but also provide predictions on the physiology of the Pho regulon and insights for the synthetic promoter design. Moreover, our work has implications for applications in artificial life and for others in biotechnology that exploit such mechanisms. As the model architecture is intrinsically open to integrate supplementary layers, together with experimental findings, it should provide insights in investigations on engineering new dynamic sensors and regulators for living technologies.

For future applications, we suggest that the approach followed in Chapter 3 is promising for identifying the type of the PhoU interaction with the Pst system. Our model provides predictions for quantifying the response to increased levels of inorganic phosphate. The model can thus help to quantify the response of PhoU protein in Pho regulon system. Moreover, it allows one to design Pho regulon circuits and analyse its expected behaviour. Having learned more about the nature of communication between the proteins in Pho regulon, we can propose different mutations for tuning the active site. This quantitative model should help to optimise the system design, and guide the system analysis.

Additionally, synthetic signalling promoter systems discussed in Chapter 4 can be used to implement an artificial promoter logic in *E. coli*. This can result in programmable artificial gene networks that can control cell behaviour such as P_i intake. Through synthetic biology approaches, a library of synthetic promoters sensing P_i can be created and expressed while exhibiting a broad range of sensitivity. This library can be used to trigger a bioremediation process upon detection of high amount of inorganic phosphate in water. Moreover, the

reactions required for the accumulation of Polyphosphate (PolyP) in *E. coli* can be added to its corresponding metabolic model in Chapter 4. Therefore, the impact of PolyP accumulation on the bioremediation of phosphate contamination in wastewater can be increased.

In this thesis, we use modelling as an investigative tool to describe physiological responses of Pho regulon triggered by external P_i levels. Approaches explained in Chapter 3 and 4 are particularly suitable to characterise the genetic mechanisms of expression involved in the regulation of biological processes of inorganic phosphate. To further analyse the nature of Pho regulon, we are currently integrating the Pho regulon deactivation network in Chapter 3 with active Pho regulon network in Chapter 4. This integration should allow us to analyse the interactions of the system components and retrieve and compare data.

Multi-level modelling of glucose homeostasis. Type 2 Diabetes Mellitus (T2DM) is a major problem of modern society. Between 2010 and 2030, there is an expected 70% increase in numbers of adults with diabetes in developing countries and a 20% increase in developed countries [85]. T2DM is characterised from the inability of the body to maintain healthy blood glucose concentrations. Many organs and tissues in the human body have roles in preserving a correct glucose homeostasis, through absorption, uptake and release of different molecules and hormone production. However, the etiology of diabetes has not been clearly understood yet. Therefore, description and interpretation of the underlying physiology through various methods, such as mathematical models, is essential.

We have introduced a closed-loop, multi-level model of glucose and insulin metabolism at the whole body level, connecting the low level cellular model for adipocyte metabolism. The ability to map all interactions of system variables is a general strategy of our model. First, a knowledge base of the biological structure of states is acquired through the literature. In a second step, mathematical relations, which demonstrate the changes of states of systems are shown by equations. The model is based on differential equations. Parameters have been taken from available literature or optimised. To aid the model validation and design, sensitivity analysis is performed on the whole body glucose model for healthy and T2DM individuals. This analysis was successfully used to test the robustness of the results, and to understand the relationships between the variables and the parameters.

Normally, our body is equipped with physiological and biochemical responses to counter-act day-to-day fluctuations in food intake, which is intrinsically linked to hunger. The closed loop system has been possible through the description of hunger dependent on leptin and ghrelin. Ghrelin is a fast-acting hormone, of which the circulatory levels show food intake related changes. It is known that the daytime secretion of ghrelin is increased, thus regulating food intake [26]. Our model suggests that, ghrelin is the one of the main responsible factor

for the food intake. Moreover, leptin is known as satiety hormone and effects the food intake. However, it is observed that its concentration varies in the long term. Although diabetic subjects may have higher leptin concentration, as they usually have a higher fat mass, they develop a resistance mechanism that make them insensitive to leptin, known as leptin resistance [143]. Ghrelin and leptin are complementary molecules, inversely correlated, within the same regulating mechanism, informing the central nervous system in normal conditions about food intake necessity and energy balance [163].

The explicit role of the liver in glucose storage and production is also demonstrated in the model. Hepatic glucose metabolism has a fundamental role in T2DM development, especially regarding impaired fasting glucose (IFG), considered a pre-diabetic stage. IFG is due to the excessive release of glucose from the liver and can depend on liver insensitivity to insulin or on abnormally high glucagon, [30] which is observed in diabetic subjects. Glucagon has a key role in the regulation of glucose homeostasis. It is produced by alpha cell and keeps blood glucose levels high enough for the body to function. The incretin hormones, which stimulate insulin secretion in the absence of plasma glucose, also have a main role on glycemic control and their secretion and action is impaired in diabetes [111].

A main novelty of our whole body model is the inclusion of all the main physiological actors described as fundamental in glucose metabolism. These variables are important because they allow for a better description of the system for the implementation of the closed loop system and for the connection with molecular signaling. Here, we have shown that the closed-loop whole body glucose regulation model can also be used functionally to understand the biology, estimate parameters, and improve intuitions. Moreover, we have demonstrated that our model is able to correctly reproduce the dynamics of whole body glucose cycle and food intake.

Another main contribution of this work is the closed-loop model, which is able to perform routine operations independently and simulate for longer time and would provide complete explanations. Moreover, the closed-loop model is more stable and able to update itself. It allow us match the output of the model (glucose needed from the body) with the input (glucose intake).

We have created a hierarchical system adding the adipocyte glucose metabolism using a simplified version of the model from Nyman et al. [127]. One of the aims of the present work is to merge two different physiological levels to lay the basis for the inclusion of the other main players in the system (such as muscle, pancreatic and liver cells). The connection between the two levels happens through the interstitial fluid. We considered that only a portion of plasmatic values would surround the cell in accordance with human physiology.

Adding the low level model to a closed loop system allowed us to see the effect during long time period. The opportunity to observe the system in a long term period of time can also allow us to see how the two biological levels, the organ and the cellular ones, react to each other. An impaired insulin signalling cascade affects glucose uptake level resulting in decreased adipose glucose and leptin secretion.

Insulin secretion from the organ level impacts the cellular level: the two levels of the model are interdependent as it is in the body. A holistic representation of the system, with the inclusion of the main components, should allow the description and interpretation of a more realistic system, allowing to simulate, for instance, different hypothesised molecular mechanisms responsible for disease development, or the action of available or experimental therapeutic molecules.

In the current work, the dynamics of the T2DM is also illustrated. With the modified model parameter for T2DM, it is possible to test how glucose homeostasis would occur and how body would react to glucose intake for T2DM. The results of the present work have shown that the reductions in glucose uptake from tissues may help to explain some of the differences of the metabolic responses between NGR and T2DM. The result indicate that T2DM is slower to return to baseline levels than NGR and long-acting forms have developed. It is known that steady state process takes more length of time in type 2 diabetic individuals [136, 161]. Moreover, in type 2 diabetes, individuals have more glucose intake. That can be the result of insulin resistance. Body muscles do not get the energy, which is taken from the food intake. Insulin resistance prevents glucose from entering the muscle and providing energy. Therefore, especially the muscles and other tissues send a “hunger” message to get enough energy into the body.

Some of the aspects of the model can be improved with the collection of data, among which glucose and insulin concentrations after a glucose test as well as incretin levels. Considering that a meal is usually composed of proteins and fats, not only of glucose. This will be considered in future model developments. This will also be important for extending the model to type 1 diabetes. Including such phenomena can result in a better description of the system and help enhancing our understanding of the mechanisms behind these adaptations. An immediate extension of the research presented in this dissertation is to use the proposed methodology to explore the other component functions in greater detail. This may provide information on the mechanisms, which are more important in producing the system behaviours and are observed in patients with diabetes.

Models of biological systems will continue to have a key role in future system and synthetic biology research. This is because models help us to understand the biological mechanisms and develop the tools that are needed for constructing the biological systems. They can concisely communicate a large amount of information about systems via uniform representations. They will speed up the identification of disease mechanisms, biomarkers and drug discovery process. Moreover, modelling will be in great demand to design or modify the systems for new functions. It will allow us to create new artificial biological systems that do not exist in nature but have beneficial purposes. Modelling will keep pushing the boundaries of biology and it will be possible to model a greater variety of biological systems. The more structures are determined by modelling, the identification of missing links and information will become easier.

Most of the knowledge in science comes from performing experiments. However, experimental data can be constructed with evidence obtained from other sources such as modelling. Increasing amount of research in modelling has produced enormous amounts of data and is accelerating the knowledge discovery of biological systems. Modelling of biological systems helps us to gain a deeper conceptual insight into the organism of interest and correlate the experimental and theoretical studies. Combination of experimental data and simulation results generate hypothesis that are consistent with the data. Mathematical and computational models are used to interpret data and make predictions about different experimental conditions. They can cover abstractions that provide insights into the systems, which are responsible for a particular behaviour. Moreover, mathematical models overcome the challenges for experimental designs. They enable the simulation of complex biological systems and generate hypotheses and suggest experimental design. Mathematical models can explore the wealth of experimental data (Big Data) and find fundamental correlations and contingencies, changing the way experimental research, development and applications are conducted.

References

- [1] Ajmera, I., Swat, M., Laibe, C., Novère, N. L., and Chelliah, V. (2013). The impact of mathematical modeling on the understanding of diabetes and related complications. *CPT Pharmacometrics Syst Pharmacol*, 2.
- [2] Alon, U. (2007). *An introduction to system biology - design principles of biological circuits*. Chapman and Hall/CRC.
- [3] Alves, R. and Savageau, M. A. (2000). Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics*, 16:786–798.
- [4] Askari, H., Tykodi, G., Liu, J., and Dagogo-Jack, S. (2010). Fasting plasma leptin level is a surrogate measure of insulin sensitivity. *The Journal of Clinical Endocrinology & Metabolism*, 95(8):3836–3843.
- [5] Azeloglu, E. and Iyengar, R. (2014). Signaling networks: Information flow, computation, and decision making. *Cold Spring Harb Perspect Med*.
- [6] Bandyopadhyayi, A. (2014). *Theoretical and computational study of biochemical signaling pathways*. PhD thesis, University of Calcutta, Department of Chemistry.
- [7] Banks, W. A. (2012). Role of the blood-brain barrier in the evolution of feeding and cognition. *Annals of the New York Academy of Sciences*, 1264(1):13–19.
- [8] Banks, W. A., Kastin, A. J., Huang, W., Jaspan, J. B., and Maness, L. M. (1996). Leptin enters the brain by a saturable system independent of insulin. *Peptides*, 17:305–311.
- [9] Barazzoni, R. (2014). Ghrelin and insulin secretion in humans: not a tale of two hormones? *Diabetes*, 63:2213–2215.
- [10] Bergman, R. N., Ider, Y. Z., and Bowden, C. R. (1979). Quantitative estimation of insulin sensitivity. *Am J Physiol*, 236:E667–77.
- [11] Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M. C., Tassani, S., Piva, F., Perez-Amodio, S., Strippoli, P., and Canaider, S. (2013). An estimation of the number of cells in the human body. *Annals of Human Biology*, 40:463–471.
- [12] Bloch, W. and Schlesinger, M. J. (1973). The phosphate content of *Escherichia coli* alkaline phosphatase and its effect on stopped flow kinetic studies. *JBC*, 248:5794–5805.

- [13] Brännmark, C., Nyman, E., Fagerholm, S., Bergenholm, L., Ekstrand, E. M., Cedersund, G., and Stralfors, P. (2013). Insulin signaling in type 2 diabetes: experimental and modelling analysis reveal mechanisms of insulin resistance in human adipocyte. *Journal of Biological Chemistry*, 288(14):9867–9880.
- [14] Briggs, G. E. and Haldane, J. B. (1925). A note on the kinetics of enzyme action. *Biochem J*, 19:338–339.
- [15] Brubaker, P. L., Ohayon, E. L., D’Alessandro, L. M., and Norwich, K. H. (2007). A mathematical model of the oral glucose tolerance test illustrating the effects of the incretins. *Annals of Biomedical Engineering*, 35(7):1286–1300.
- [16] Cardelli, L. (2008). On process rate semantics. *Theoretical Computer Science*, 391:190–215.
- [17] Castiglione, F., Tieri, P., De Graaf, A., Franceschi, C., Liò, P., Ommen, B. V., Mazzà, C., Tuchel, A., Bernaschi, M., Samson, C., Colombo, T., Castellani, G. C., Capri, M., Garagnani, P., Salvioli, S., Nguyen, V. A., Bobeldijk-Pastorova, I., Krishnan, S., Cappozzo, A., Sacchetti, M., Morettini, M., and Ernst, M. (2013). The onset of type 2 diabetes: Proposal for a multi-scale model. *JMIR Research Protocols*, 2(2):e44.
- [18] Caumo, A., Bergman, R. N., Cobelli, C., Raffaele, S., and California, S. (2000). Insulin sensitivity from meal tolerance tests in normal subjects: a minimal model index. *Clin Endocrinol Meta*, 85:4396–4402.
- [19] Celani, A. and Vergassola, M. (2009). Bacterial strategies for chemotaxis response. *Proc Natl Acad Sci USA*, 107:1391–1396.
- [20] Celeste, R., Ackerman, E., Gatewood, L. C., Reynolds, C., and Molnar, G. D. (1978). The role of glucagon in the regulation of blood glucose: Model studies. *Bulletin of Mathematical Biology*, 40(1):59–77.
- [21] Chekabab, S. M., Harel, J., and Dozois, C. M. (1993). Interplay between genetic regulation of phosphate homeostasis and bacterial virulence. *Current Opinion in Microbiology*, 175:6797–6809.
- [22] Chen, S. and Wilson, D. B. (1997). Construction and characterization of Escherichia coli genetically engineered for bioremediation of Hg(2+)-contaminated environments. *Appl Environ Microbiol*, 63:2442–2445.
- [23] Chen, T., Santiago, A. G., Jung, W., Krzeminski, L., Yang, F., Martell, D. J., Helmann, J. D., and Chen, P. (2015). Concentration- and chromosome-organization-dependent regulator unbinding from DNA for transcription regulation in living cells. *Nature Communications*, 6(7445).
- [24] Chew, Y. H., Shia, Y. L., Lee, C. T., Majid, F. A. A., Chua, L. S., Sarmidi, M. R., and Aziz, R. A. (2009). Modeling of glucose regulation and insulin-signaling pathways. *Molecular and Cellular Endocrinology*, 303(1-2):13–24.
- [25] Chopra, P. and Kamma, A. (2006). Engineering life through synthetic biology. *In Silico Biology*, 6:401–410.

- [26] Crespo, C. S., Cachero, A. P., Jiménez, L. P., Barrios, V., and Ferreiro, E. A. (2014). Peptides and food intake. *Frontiers in Endocrinology*, 5.
- [27] Cuccato, G., Gatta, G. D., and di Bernardo, D. (2009). Systems and synthetic biology: tackling genetic networks and complex diseases. *Heredity*, 102:527–532.
- [28] Cummings, D. E. (2001). A preprandial rise in plasma ghrelin levels suggests a role in meal initiation in humans. *Diabetes*, 50:1714–1719.
- [29] Dada, J. O. and Mendes, P. (2011). Multi-scale modelling and simulation in systems biology. *Integr Biol*, 3:86–96.
- [30] D’Alessio, D. (2011). The role of dysregulated glucagon secretion in type 2 diabetes. *Diabetes, Obes Metab*, 13:126–132.
- [31] De Gaetano, A. and Arino, O. (2000). Mathematical modelling of the intravenous glucose tolerance test. *Journal of Mathematical Biology*, 40(2):136–168.
- [32] De Gaetano, A., Hardy, T., Beck, B., Abu-raddad, E., and Palumbo, P. (2008). Mathematical models of diabetes progression. *Am J Physiol Endocrinol Metab*, 295:E1462–E1479.
- [33] De Gaetano, A., Panunzi, S., Matone, A., Samson, A., Vrbikova, J., Bendlova, B., and Pacini, G. (2013). Routine OGTT: A robust model including incretin effect for precise identification of insulin sensitivity and secretion in a single individual. *PLoS ONE*, 8(8):e70875.
- [34] De Mey, M., Maertens, J., Lequeux, G. J., Soetaert, W. K., and Vandamme, E. J. (2007). Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnology*, 7(34).
- [35] de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M., and Vogel, C. (2010). Two-component signaling circuit structure and properties. *Curr Opin Microbiol*, 13:184–189.
- [36] Diabetes, A. A. (2016). 2. classification and diagnosis of diabetes.
- [37] Diabetes-Atlas-Group, I. (2015). Update of mortality attributable to diabetes for the idf diabetes atlas: estimates for the year 2013. *Diabetes Res Clin Pract*, 109:461–465.
- [38] Dien, S. J. V. and Keasling, J. D. (1997). A dynamic model of the *Escherichia coli* phosphate-starvation response. *J. Theor. Biol*, 190:37–49.
- [39] Dubitzky, W. (2006). Understanding the computational methodologies of systems biology. *Briefings in Bioinformatics*, 7:315–317.
- [40] Editorial (2004). End of the interlude? *Nature Biotechnology*, 22:1191.
- [41] EFSA Panel on Dietetic Products, N. and (NDA), A. (2010). Scientific opinion on dietary reference values for carbohydrates and dietary fibre. volume 8, page 1462. Wiley-Blackwell.
- [42] Elowitz, M. B. and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338.

- [43] Elowitz, M. B. and Leibler, S. (2010). Synthetic biology: challenges, opportunities. *Biotechnology*, 325.
- [44] Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186.
- [45] Engström, B. E., Öhrvall, M., Sundbom, M., Lind, L., and Karlsson, F. A. (2006). Meal suppression of circulating ghrelin is normalized in obese individuals following gastric bypass surgery. *Int J Obes Relat Metab Disord*, 31(3):476–480.
- [46] Federation, I. D. (2006). Third congress, international diabetes federation. *Diabetes*, 6:1689–1699.
- [47] Ferrannini, E. and DeFronzo, R. A. (2015). *Insulin actions in vivo: glucose metabolism*, page 211–233. John Wiley and Sons.
- [48] Fey, D. (2012). *Parameter identification for biological models*. PhD thesis, University of Liège, Department of Electrical Engineering and Computer Science.
- [49] Fourer, R., Gay, D. M., and Kernighan, B. W. (1997). *AMPL: A Modeling Language for Mathematical Programming*. Scientific Press.
- [50] Friedman, J. M. and Halaas, J. L. (1998). Leptin and regulation of body weight in mammals. *Nature*, 395:763–770.
- [51] Gallenberger, M., zu Castell, W., Hense, B. A., and Kuttler, C. (2012). Dynamics of glucose and insulin concentration connected to the β -cell cycle: model development and analysis. *Theor Biol Med Model*, 9(1):46.
- [52] Gao, R. and Stock, A. M. (2009). Biological insights from structures of two-component proteins. *Annu Rev Microbiol.*, 63:133–154.
- [53] Gao, R. and Stock, A. M. (2012). Probing kinase and phosphatase activities of two-component systems in vivo with concentration-dependent phosphorylation profiling. *PNAS*, 110:672–677.
- [54] Gao, R. and Stock, A. M. (2013). Evolutionary tuning of protein expression levels of a positively autoregulated two-component system. *PLoS Genet*, 9(10):e1003927. doi:10.1371/journal.pgen.1003927.
- [55] Gao, R. and Stock, A. M. (2017). Quantitative kinetic analyses of shutting off a two-component system. *mBio*, 8:e00412–17.
- [56] Gardner, S. G., Johns, K. D., Tanner, R., and McCleary, W. R. (2014). The PhoU protein from *Escherichia coli* interacts with PhoR, PstB, and metals to form a phosphate-signaling complex at the membrane. *J Bacteriol*, 196:1741–1752.
- [57] Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339–342.
- [58] Gerich, J. E. (2000). Physiology of glucose homeostasis. *Diabetes, Obes Metab*, 2:345–350.

- [59] Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434.
- [60] Golden, P. L., Maccagnan, T. J., and Pardridge, W. M. (1997). Human blood-brain barrier leptin receptor: binding and endocytosis in isolated human brain microvessels. *Journal of Clinical Investigation*, 99(1):14–18.
- [61] Gostner, R., Baldacci, B., Morine, M. J., and Priami, C. (2014). Graphical modeling tools for systems biology. *CSUR*, 47(2):1–21.
- [62] Grasman, J. (2013). Reconstruction of the drive underlying food intake and its control by leptin and dieting. *PLoS ONE*, 8(9):e74997.
- [63] Green, S. (2017). *Philosophy of Systems and Synthetic Biology*. Metaphysics Research Lab, Stanford University, winter 2017 edition.
- [64] Gromada, J., Franklin, I., and Wollheim, C. B. (2007). α -cells of the endocrine pancreas: 35 years of research but the enigma remains. *Endocr Rev*, 28:84–116.
- [65] Groop, L. and Pociot, F. (2014). Genetics of diabetes are we missing the genes or the disease? *Mol Cell Endocrinol*, 382:726–739.
- [66] Gruetter, R., Ugurbil, K., and Seaquist, E. R. (1998). Steady-state cerebral glucose concentrations and transport in the human brain. *Journal of Neurochemistry*, 70(1):397–408.
- [67] Guldberg, C. and Waage, P. (1864). Studies concerning affinity. *C. M. Forhandlinger: Videnskabs-Selskabet i Christiana*, 35.
- [68] Harris, R. M., Webb, D. C., Howitt, S. M., and Cox, G. B. (2001). Characterization of PitA and PitB from *Escherichia coli*. *J Bacteriol.*, 183:5008–5014.
- [69] Hasenauer, J. J., Jagiella, N., Hross, S., and Theis, F. J. (2015). Data-driven modelling of biological multi-scale processes. *Journal of Coupled Systems and Multi-scale Dynamics*, 3:101–121.
- [70] Hawley, D. K. and McClure, W. R. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Research*, 11(8):2237–2255.
- [71] Henkel, E., Menschikowski, M., Koehler, C., Leonhardt, W., and Hanefeld, M. (2005). Impact of glucagon response on postprandial hyperglycemia in men with impaired glucose tolerance and type 2 diabetes mellitus. *Metabolism*, 54(9):1168–1173.
- [72] Herman, M. A. and Kahn, B. B. (2006). Glucose transport and sensing in the maintenance of glucose homeostasis and metabolic harmony. *Journal of Clinical Investigation*, 116(7):1767–1775.
- [73] Hickernell, F. J. and Yuan, Y. (1997). A simple multistart algorithm for global optimization.
- [74] Hill, A. V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol*, 40.

- [75] Hsieh, Y. and Wanner, B. L. (2010). Global regulation by the seven-component P_i signalig system. *Current Opinion in Microbiology*, 13:198–203.
- [76] Hudek, L., Premachandra, D., Webster, W. A. J., and Bräu, L. (2016). Role of phosphate transport system component PstB1 in phosphate internalization by *Nostoc punctiforme*. *Appl Environ Microbiol*, 82:6344–6356.
- [77] Idorn, T., Knop, F. K., Jorgensen, M., Holst, J. J., Hornum, M., and Feldt-Rasmussen, B. (2013). Postprandial responses of incretin and pancreatic hormones in non-diabetic patients with end-stage renal disease. *Nephrology Dialysis Transplantation*, 29(1):119–127.
- [78] Ingalls, B. P. (2013). *Mathematical modelling in systems biology: an Introduction*. MIT Press.
- [79] Jacquier, M., Crauste, F., Soulage, C. O., and Soula, H. A. (2014). A predictive model of the dynamics of body weight and food intake in rats submitted to caloric restrictions. *PLoS ONE*, 9(6):e100073.
- [80] Jagiella, N. (2012). *Parameterization of lattice-based tumor models from data*. PhD thesis, Universite Pierre et Marie Curie.
- [81] Jansson, M. (1988). Phosphate uptake and utilization by bacteria and algae. *Kluwer Academic Publishers*, 170:177–189.
- [82] Jensen, P. R. and Hammer, K. (1998a). Artificial promoters for metabolic optimization. *Biotechnology and Bioengineering*, 58(2-3):191–195.
- [83] Jensen, P. R. and Hammer, K. (1998b). The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Applied and Environmental Microbiology*, 64(1):82–87.
- [84] Jéquier, E. and Tappy, L. (1999). Regulation of body weight in humans. *Physiol Rev*, 79:451–480.
- [85] Joseph, J. (2010). Incidence and risk factors of type 2 diabetes in general population:the Tromsø Study. *Scand J Public Health*, 38(7):768–775.
- [86] Kang, H., Han, K., and Choi, M. (2012). Mathematical model for glucose regulation in the whole-body system. *Islets*, 4(2):84–93.
- [87] Karapetyan, S. and Buchler, N. E. (2015). Role of DNA binding sites and slow unbinding kinetics in titration-based oscillators. *Physical review E Statistical, nonlinear, and soft matter physics*, 92(6):062712.
- [88] Kershaw, E. E. and Flier, J. S. (2004). Adipose tissue as an endocrine organ. *Clin Endoc Meta*, 89:2548–2556.
- [89] Khalil, A. S. and Collins, J. (2010). Synthetic biology applications come of age. *Nature Revs Genetics*, 5:367–379.

- [90] Kiselyov, V. V., Versteheyhe, S., Gauguin, L., and Meyts, P. D. (2009). Harmonic oscillator model of the insulin and IGF1 receptors' allosteric binding and activation. *Mol Syst Biol*, 5.
- [91] Kitano, H. (2001). *Foundations of Systems Biology*. The MIT Press.
- [92] Klein, S., Coppack, S. W., Mohamed-Ali, V., and Landt, M. (1996). Adipose tissue leptin production and plasma leptin kinetics in humans. *Diabetes*, 45(7):984–987.
- [93] König, M. and Holzhutter, H. G. (2012). Kinetic modeling of human hepatic glucose metabolism in type 2 diabetes mellitus predicts higher risk of hypoglycemic events in rigorous insulin therapy. *Journal of Biological Chemistry*, 287(44):36978–36989.
- [94] Krishnaswamy, U., Muthuswamy, M., and Perumalsamy, M. (2009). Studies on the efficiency of the removal of phosphate using bacterial consortium for the biotreatment of phosphate wastewater. *Eur. J. Appl. Sci*, 1:6–15.
- [95] Lamarche, M. G., Wanner, B. L., Crépin, S., and Harel, J. (2008). The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. *Federation of European Microbiological Societies*, 32:461–473.
- [96] Lambert, J. D. (1992). *Numerical Methods for Ordinary Differential Systems*. New York Wiley.
- [97] Levy, J. R., Gyarmati, J., Lesko, J. M., Adler, R. A., and Stevens, W. (2000). Dual regulation of leptin secretion: intracellular energy and calcium dependence of regulated pathway. *Am J Physiol Endocrinol Metab*, 278:E892–901.
- [98] Lewis, J. (2003). Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr. Biol.*, 13:1398–1408.
- [99] Li, J. and Zhang, Y. (2014). Relationship between promoter sequence and its strength in gene expression. *The European Physical Journal E*, 37(86).
- [100] Liandong, Z. (2014). *Sustainable biodiesel production from microalgae cultivated with piggery wastewater*. PhD thesis, University of Vaasa, Faculty of Technology Department of Production.
- [101] Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., P. Scott, M., Bretscher, A., Ploegh, H., and Matsudaira, P. (6th edition). *Molecular Cell Biology*. W. H. Freeman and Company.
- [102] Lozada, S. (2010). Glucose regulation in diabetes. *New York University, Courant Institute for Mathematical Sciences*.
- [103] Lukas, T. (2004). A signal transduction pathway model prototype i: from agonist to cellular endpoint. *Biophysical Journal*, 87:1406–1416.
- [104] Makroglou, A., Li, J., and Kuang, Y. (2006). Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. *Appl Numer Math*, 56:559–573.

- [105] Man, C. D., Caumo, A., Basu, R., Rizza, R., and Toffolo, G. (2005). Measurement of selective effect of insulin on glucose disposal from labeled glucose oral test minimal model. *Am J Physiol Endocrinol Metab*, 289:904–914.
- [106] Man, C. D., Rizza, R. A., and Cobelli, C. (2007). Meal simulation model of the glucose-insulin system. *IEEE Transactions on Biomedical Engineering*, 54(10):1740–1749.
- [107] Marchetti, L., Reali, F., Dauriz, M., Brangani, C., and Boselli, L. (2016). A novel insulin glucose model after a mixed meal test in patients with type 1 diabetes on insulin pump therapy. *Scientific Report*, 6.
- [108] Marucci, L. (2010). *Mathematical modelling and non-linear analysis of Synthetic Gene Regulatory Networks*. PhD thesis, University of Napoli Federico II, Department of Computer and Systems Engineering.
- [109] McAdams, H. H. and Shapiro, L. (1995). Circuit simulation of genetic networks. *Science*, 269.
- [110] Meier, J. J. and Nauck, M. A. (2010). Is the diminished incretin effect in type 2 diabetes just an epi-phenomenon of impaired cell function? *Diabetes*, 59(5):1117–1125.
- [111] Meloni, A. R., DeYoung, M. B., Lowe, C., and Parkes, D. G. (2013). GLP-1 receptor activated insulin secretion from pancreatic β -cells: mechanism and glucose dependence. *Diabetes Obes Metab*, 15:15–27.
- [112] Michaelis, L. and Menten, M. (1913). Die kinetik der invertinwirkung. *Biochemistry Zeitung*, 49:333–369.
- [113] Miyashiro, T. and Goulian, M. (2008). High stimulus unmasks positive feedback in an autoregulated bacterial signaling circuit. *PNAS*, 105:17457–17462.
- [114] Mordecai, A. (2003). *Nonlinear Programming: Analysis and Methods*. Dover Publishing.
- [115] Mosa, K. A., Saadoun, I., Kumar, K., Helmy, M., and Dhankher, O. P. (2016). Potential biotechnological strategies for the cleanup of heavy metals and metalloids. *Frontiers in Plant Science*, 7:303.
- [116] Mueller, W. M., Gregoire, F. M., Stanhope, K. L., Mobbs, C. V., and Mizuno, T. M. (1998). Evidence that glucose metabolism regulates leptin secretion from cultured rat adipocytes. *Endocrinology*, 139:551–558.
- [117] Mukhopadhyay, A., Gao, R., and D. G. Lynn, D. (2004). Integrating input from multiple signals: the VirA/VirG two-component system of *Agrobacterium tumefaciens*. *Chembiochem*, 5:1535–1542.
- [118] Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, 10:354–360.

- [119] Nakayama, Y., Kinoshita, A., and Tomita, M. (2005). Dynamic simulation of red blood cell metabolism and its application to the analysis of a pathological condition. *Theoretical Biology and Medical Modelling*, 2.
- [120] Natali, A., Prato, S. D., and Mari, A. (2015). *Normal beta-cell function, section 8*, pages 108–124. John Wiley and Sons.
- [121] Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974). Culture medium for enterobacteria. *J Bacteriol.*, 119:736–747.
- [122] Nelson, D. L. and Cox, M. M. (2013). *Lehninger principles of biochemistry (6th ed.)*. W.H. Freeman, New York.
- [123] Noble, D. (2007). From the hodgkin-huxley axon to the virtual heart. *Journal of Physiology*, 580:15–22.
- [124] Nocedal, J. and Wright, S. (2006). *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering.
- [125] Nussey, S. S. and Whitehead, S. A. (2001). *Endocrinology: an integrated approach*. Oxford: BIOS Scientific Publishers.
- [126] Nyman, E., Brannmark, C., Palmer, R., Brugard, J., Nystrom, F. H., Stralfors, P., and Cedersund, G. (2011). A hierarchical whole-body modeling approach elucidates the link between in vitro insulin signaling and in vivo glucose homeostasis. *Journal of Biological Chemistry*, 286(29):26028–26041.
- [127] Nyman, E., Rajan, M. R., Fagerholm, S., Brännmark, C., Cedersund, G., and Stralfors, P. (2014). A single mechanism can explain network-wide insulin resistance in adipocytes from obese patients with type 2 diabetes. *Journal of Biological Chemistry*, 289(48):33215–33230.
- [128] O’Connor, C. M. and Adams, J. U. (2010). *Essentials of Cell Biology*. Cambridge, MA: NPG Education.
- [129] OpenWetWare (2007). Parameter estimation in e.coli. Available at http://www.openwetware.org/wiki/Computational_Biology/Gene_Expression_modeling.
- [130] Palsson, B. O. (2006). *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press.
- [131] Palumbo, P., Ditlevsen, S., Bertuzzi, A., and De Gaetano, A. (2013). Mathematical modeling of the glucose-insulin system: a review paper. *Math Biosci*, 244:69–81.
- [132] Pattaranit, R. and van den Berg, H. A. (2008). Mathematical models of energy homeostasis. *J R Soc Interface*, 5:1119–1135.
- [133] Peterson, C., Mandel, M. J., and Silhavy, T. J. (2005). Escherichia coli starvation diets: essential nutrients weigh in distinctly. *Journal of Bacteriology*, 187:22:7549–7553.
- [134] Poretzky, L. (2010). *Principles of Diabetes Mellitus*. Springer Science.

- [135] Porte, D. and Woods, S. C. (1981). Regulation of food intake and body weight by insulin. *Diabetologia*, 20:274–280.
- [136] Pratt, A. C., Wattis, J. A., and Salter, A. M. (2015). Mathematical modelling of hepatic lipid metabolism. *Mathematical Biosciences*, 262:167–181.
- [137] Prescott, T. P. (2014). *Large-scale layered systems and synthetic biology: model reduction and decomposition*. PhD thesis, University of Oxford.
- [138] Raj, A. and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226.
- [139] Rice, C. D., Pollard, J. E., Lewis, Z. T., and McCleary, W. R. (2009). Employment of a promoter-swapping technique shows that PhoU modulates the activity of the PstCAB2 ABC transporter in *Escherichia coli*. *Applied and Environmental Microbiology*, 75:573–582.
- [140] Ruder, W. C. and Collins, J. (2011). Synthetic biology moving into the clinic. *Science*, 333:1248–1252.
- [141] Sakata, I. and Sakai, T. (2010). Ghrelin cells in the gastrointestinal tract. *Int J Pept*, pages 1–7.
- [142] Santos-Beneit, F. (2015). The Pho regulon: a huge regulatory network in bacteria. *Frontiers in Microbiology*, 6:402:1–13.
- [143] Scarpace, P. J. and Zhang, Y. (2009). Leptin resistance: a predisposing factor for diet-included obesity. *Am J Physiol Regul Integr Comp Physiol*, 296:R493–R500.
- [144] Schirra, J., Katschinski, M., Weidmann, C., Schäfer, T., Wank, U., Arnold, R., and Göke, B. (1996). Gastric emptying and release of incretin hormones after glucose ingestion in humans. *Journal of Clinical Investigation*, 97(1):92–103.
- [145] Sedaghat, A. R., Sherman, A., and Quon, M. J. (2002). A mathematical model of metabolic insulin signaling pathways. *American Journal of Physiology - Endocrinology And Metabolism*, 283(5):E1084–E1101.
- [146] Seino, Y., Fukushima, M., and Yabe, D. (2010). GIP and GLP-1, the two incretin hormones: similarities and differences. *J Diabetes Investig*, 1:8–23.
- [147] Shahrezaei, V. and Swain, P. S. (2008). The stochastic nature of biochemical networks. *Current Opinion in Biotechnology*, 19:369–374.
- [148] Shimizu, K. (2014). Regulation systems of bacteria such as *Escherichia coli* in response to nutrient limitation and environmental stresses. *Metabolites*, 4:1–35.
- [149] Shinar, G., Milo, R., Matínez, M. R., and Alon, U. (2007). Input output robustness in simple bacterial signaling systems. *Proc. Natl. Acad. Sci.*, 104:19931–19935.
- [150] Sivakumar, G., Xu, J., Thompson, R. W., Yang, Y., Randol-Smith, P., and Weathers, P. J. (2012). Integrated green algal technology for bioremediation and biofuel. *Bioresour Technol*, 107:1–9.

- [151] Southern, J., Pitt-Francis, J., Whiteley, J., Stokeley, D., Kobashi, H., Nobes, R., Kadooka, Y., and Gavaghan, D. (2008). Multi-scale computational modelling in biology and physiology. *Progr Biophys Molecul Biol*, 96.
- [152] Steed, P. M. and Wanner, B. L. (2014). Use of the rep technique for allele replacement to construct mutants with deletions of the pstSCAB-phoU operon: evidence of a new role for the PhoU protein in phosphate regulon. *Virulence*, 5:8:786–793.
- [153] Stockhorst, U., Fries, D. D., Steingrueber, H. J., and Scherbaum, W. A. (2004). Insulin and the cns: effects on food intake, memory, and endocrine parameters and the role of intranasal insulin administration in humans. *Physiol Behav*, 83:47–54.
- [154] Sulston, K. W., Ireland, W. P., and Praught, J. C. (2006). Hormonal effects on glucose regulation. *Atlantic Electronic*, 1(1).
- [155] Sumner, T. (2010). *Sensitivity analysis in systems biology modelling and its application to a multi-scale model of blood glucose homeostasis*. PhD thesis, University College London, Centre for Mathematics and Physics in the Life Sciences and Experimental Biology.
- [156] Taborsky, G. J. (2010). The physiology of glucagon. *J Diabetes Sci Technol*, 4:1338–1344.
- [157] Tam, J., Fukumura, D., and Jain, R. K. (2009). A mathematical model of murine metabolic regulation by leptin: energy balance and defense of a stable body weight. *Cell Metab*, 9:52–63.
- [158] Thiel, K. (2006). Systems biology, incorporated? *Nature Biotechnology*, 24:1055–1057.
- [159] Tiwari, A., Ray, J. C., Narula, J., and Igoshin, O. A. (2011). Bistable responses in bacterial genetic networks: designs and dynamical consequences. *Math Biosci*, 231:76–89.
- [160] Toghaw, P. and Lenbury, Y. (2013). The glucose-insulin-incretin model for bariatric surgery and T2DM improvement mechanisms with two delays. *Int. J. Math. Comput. Simul.*, 7(439-447).
- [161] Toghaw, P., Matone, A., Lenbury, Y., and De Gaetano, A. (2012). Bariatric surgery and T2DM improvement mechanisms: a mathematical model. *Theor Biol Med Model*, 9(1):16.
- [162] Topp, B., Promislow, K., Devries, G., and Finegood, D. T. F. (2000). A model of β -cell mass, insulin, and glucose kinetics: pathways to diabetes. *J Theor Biol*, 206:605–619.
- [163] Tshöp, M., Smiley, D. L., and Heiman, M. L. (2000). Ghrelin induces adiposity in rodents. *Nature*, 407:908–913.
- [164] Uluşeker, C., Torres, J., García, J. L., Hanczyc, M. M., Nogales, J., and Kahramanoğulları, O. (2017). A dynamic model of the phosphate response system with synthetic promoters in *Escherichia coli*. *Proc. of the International Conference on Artificial Life, Lyon*.

- [165] Uluseker, C., Simoni, G., Marchetti, L., Dauriz, M., Matone, A., and Priami, C. (2018). A closed-loop multi-level model of glucose homeostasis. *PLoS ONE*, 13(2):e0190627.
- [166] Vijayadeep, C. and Sastry, P. S. (2014). Effect of heavy metal uptake by *E. coli* and *Bacillus* spp. *J Bioremed Biodeg*, 5.
- [167] Walpole, J., Papin, J. A., and Peirce, S. M. (2013). Multiscale computational models of complex biological systems. *Annu. Rev. Biomed. Eng.*, 15:137–154.
- [168] Wanner, B. L. (1996). *Phosphorus assimilation and control of the phosphate regulon*, pages 1357–1381. ASM Press, Washington, D.C.
- [169] Wanner, B. L., Jiang, W., Kim, S., Yamagata, S., Haldimann, A., and Daniel, L. L. (1995). *Are the multiple signal transduction pathways of the Pho regulon due to cross talk or cross regulation?*, pages 297–315. R.G. Landes Company, Texas USA.
- [170] Warner, B. L. and Chang, B. D. (1987). The *phoBR* operon in *Escherichia coli* K-12. *Bacteriol*, 169:5569–5574.
- [171] Watson, J. D. and Crick, F. H. (1953). Molecular structure of nucleic acids: A structure for deoxyribose. *Nucleic Acid. Nature*, 171:737–738.
- [172] Weiss, J. N. (1997). The hill equation revisited: uses and misuses. *FASEB J.*, 11.
- [173] Whiteley, J. P. (2008). An efficient technique for the numerical solution of the bidomain equations. *Ann. Biomed. Eng.*, 36:1398–1408.
- [174] Wiig, H. and Swartz, M. A. (2012). Interstitial fluid and lymph formation and transport: physiological regulation and roles in inflammation and cancer. *Physiol Rev*, 93:1005–1060.
- [175] Wilkinson, D. J. (2011). *Stochastic Modelling for Systems Biology*. CRC Press.
- [176] Williams, C. L. and Cotter, P. A. (2007). Autoregulation is essential for precise temporal and steady-state regulation by the *Bordetella BvgAS* phosphorelay. *J Bacteriol*, 189:1974–1982.
- [177] Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X. S. (2006). Probing gene expression in live cells, one protein molecule at a time. *Science*, 311(5767):1600–1603.
- [178] Zobel, S., Benedetti, I., Eisenbach, L., de Lorenzo, V., Wierckx, N., and Blank, L. M. (2015). Tn7-based device for calibrated heterologous gene expression in *Pseudomonas putida*. *ACS Synth Biol.*, 4:1341–1351.

Appendix A

Appendix A

Here, we represent the Supplementary Materials for Chapter 3.

$$\begin{aligned}dABC(t)/dt &= -r1.ABC(t) + r2.ABCo(t).Pext(t) \\dABCo(t)/dt &= r1.ABC(t) - r2.ABCo(t).Pext(t) \\dPext(t)/dt &= -r2.ABCo(t).Pext(t) \\dPin(t)/dt &= r2.ABCo(t).Pext(t) \\dPhoR(t)/dt &= -r3a.ABC(t).PhoR(t) + r4a.PhoRr(t) \\dPhoRr(t)/dt &= r3a.ABC(t).PhoR(t) - r4a.PhoRr(t)\end{aligned}$$

Fig. A.1 The ODEs for the Model 1 reactions in Chapter 3.

$$\begin{aligned}dABC(t)/dt &= -r1.ABC(t) + r2.ABCo(t).Pext(t) \\&\quad - r3b.ABC(t).PhoR(t) + r4b.ABC_PhoR(t) \\dABCo(t)/dt &= r1.ABC(t) - r2.ABCo(t).Pext(t) \\dPext(t)/dt &= -r2.ABCo(t).Pext(t) \\dPin(t)/dt &= r2.ABCo(t).Pext(t) \\dPhoR(t)/dt &= -r3b.ABC(t).PhoR(t) + r4b.ABC_PhoR(t) \\dABC_PhoR(t)/dt &= r3b.ABC(t).PhoR(t) - r4b.ABC_PhoR(t)\end{aligned}$$

Fig. A.2 The ODEs for the Model 2 reactions in Chapter 3.

$$\begin{aligned}
dABC(t)/dt &= -r1.ABC(t) + r2.ABCo(t).Pext(t) \\
dABCo(t)/dt &= r1.ABC(t) - r2.ABCo(t).Pext(t) \\
dPext(t)/dt &= -r2.ABCo(t).Pext(t) \\
dPin(t)/dt &= r2.ABCo(t).Pext(t) \\
dPhoU(t)/dt &= -r3'.ABC(t).PhoU(t) + r4'.PhoUa(t) \\
dPhoUa(t)/dt &= r3'.ABC(t).PhoU(t) - r4'.PhoUa(t) \\
dPhoR(t)/dt &= -r5a.PhoUa(t).PhoR(t) + r6a.PhoRr(t) \\
dPhoRr(t)/dt &= r5a.PhoUa(t).PhoR(t) - r6a.PhoRr(t)
\end{aligned}$$

Fig. A.3 The ODEs for the PhoU Model 1 reactions in Chapter 3.

$$\begin{aligned}
dABC(t)/dt &= -r1.ABC(t) + r2.ABCo(t).Pext(t) \\
dABCo(t)/dt &= r1.ABC(t) - r2.ABCo(t).Pext(t) \\
dPext(t)/dt &= -r2.ABCo(t).Pext(t) \\
dPin(t)/dt &= r2.ABCo(t).Pext(t) \\
dPhoU(t)/dt &= -r3'.ABC(t).PhoU(t) + r4'.PhoUa(t) \\
dPhoUa(t)/dt &= r3'.ABC(t).PhoU(t) - r4'.PhoUa(t) \\
&\quad -r5b.PhoUa(t).PhoR(t) + r6b.PhoUa_PhoR(t) \\
dPhoR(t)/dt &= -r5b.PhoUa(t).PhoR(t) + r6b.PhoUa_PhoR(t) \\
dPhoUa_PhoR(t)/d &= r5b.PhoUa(t).PhoR(t) - r6b.PhoUa_PhoR(t)
\end{aligned}$$

Fig. A.4 The ODEs for the PhoU Model 2 reactions in Chapter 3.

Appendix B

Appendix B

Here, we represent the Supplementary Materials for Chapter 4.

$$\begin{aligned}
dGfp(t)/dt &= r_{18}.mRNAa(t) - r_{22}.Gfp(t) \\
dDiPhoR(t)/dt &= r_{6r}.DiPhoRp(t) - r_6.DiPhoR(t) + r_{11}.DiPhoRp-PhoB(t) + r_{21}.mRNAb(t) \\
&\quad - r_{24}.DiPhoR(t) - r_{13}.DiPhoR(t).PhoBp(t) + r_{13r}.DiPhoR-PhoBp(t) + r_{14}.DiPhoR-PhoBp(t) \\
dDiPhoRp(t)/dt &= r_6.DiPhoR(t) - r_{6r}.DiPhoRp(t) - r_7.DiPhoRp(t) + r_{7r}.DiPhoRpp(t) \\
&\quad + r_9.DiPhoRpp-PhoB(t) + r_{10r}.DiPhoRp-PhoB(t) - r_{10}.DiPhoRp(t).PhoB(t) \\
dDiPhoRp-PhoB(t)/dt &= r_{10}.DiPhoRp(t).PhoB(t) - r_{10r}.DiPhoRp-PhoB(t) \\
&\quad - r_{11}.DiPhoRp-PhoB(t) \\
dDiPhoRpp(t)/dt &= r_7.DiPhoRp(t) - r_{7r}.DiPhoRpp(t) - r_8.DiPhoRpp(t).PhoB(t) \\
&\quad + r_{8r}.DiPhoRpp-PhoB(t) \\
dDiPhoRpp-PhoB(t)/dt &= r_8.DiPhoRpp(t).PhoB(t) - r_{8r}.DiPhoRpp-PhoB(t) \\
&\quad - r_9.DiPhoRpp-PhoB(t) \\
dPhoB(t)/dt &= -r_8.DiPhoRpp(t).PhoB(t) + r_{8r}.DiPhoRpp-PhoB(t) - r_{10}.DiPhoRp(t).PhoB(t) \\
&\quad + r_{10r}.DiPhoRp-PhoB(t) + r_{20}.mRNAb(t) + r_{23}.PhoB(t) + r_{14}.DiPhoR-PhoBp(t) \\
dPhoBp(t)/dt &= r_9.DiPhoRpp-PhoB(t) - 2.r_{12}.PhoBp(t)^2 + r_{11}.DiPhoRp-PhoB(t) \\
&\quad + 2.r_{12r}.DiPhoBpp(t) + r_{13r}.DiPhoR-PhoBp(t) - r_{13}.DiPhoR(t).PhoBp(t) \\
dDiPhoR-PhoBp(t)/dt &= r_{13}.DiPhoR(t).PhoBp(t) - r_{13r}.DiPhoR-PhoBp(t) \\
&\quad - r_{14}.DiPhoR-PhoBp(t) \\
dDiPhoBpp(t)/dt &= r_{12}.PhoBp(t)^2 - r_{12r}.DiPhoBpp(t) - r_{15}.DiPhoBpp(t).pPhoA(t) \\
&\quad + r_{16r}.pPhoBa(t) + r_{15r}.pPhoAa(t) - r_{16}.DiPhoBpp(t).pPhoB(t) \\
dpPhoA(t)/dt &= -r_{15}.DiPhoBpp(t).pPhoA(t) + r_{15r}.pPhoAa(t) \\
dpPhoAa(t)/dt &= r_{15}.DiPhoBpp(t).pPhoA(t) - r_{15r}.pPhoAa(t) \\
dpPhoB(t)/dt &= -r_{16}.DiPhoBpp(t).pPhoB(t) + r_{16r}.pPhoBa(t) \\
dpPhoBa(t)/dt &= r_{16}.DiPhoBpp(t).pPhoB(t) - r_{16r}.pPhoBa(t) \\
dmRNAa(t)dt &= r_{17}.pPhoAa(t) - r_{25}.mRNAa(t) \\
dmRNAb(t)/dt &= r_{19}.pPhoBa(t) - r_{26}.mRNAb(t)
\end{aligned}$$

Fig. B.1 The ODEs for the model reactions in Chapter 4.

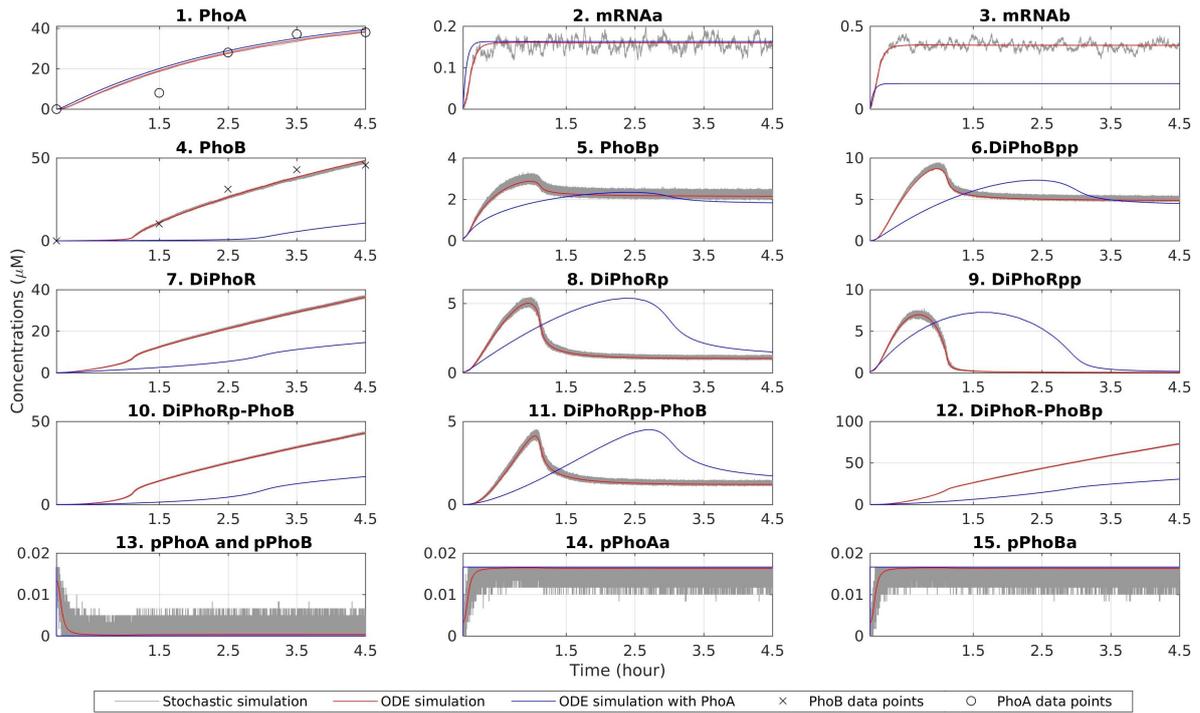


Fig. B.2 The deterministic ODE simulation results that are fitted with only the experimental data on PhoA expression levels are plotted in blue. The deterministic ODE simulation results that are fitted with the experimental data on both PhoA and PhoB expression levels are plotted in red. In all the experiments, the external P_i concentration is $0\mu\text{M}$ and a time-course of 4.5 hours is considered. For the simulations that consider both PhoA and PhoB data, stochastic simulation results, plotted in grey, are also provided. In the simulations that include both of the PhoA and PhoB data, the values for r_{14} and r_{15} in Table 1 in the main text are obtained by increasing these parameters from 0.05 to 0.13 and from 0.03 to 0.035, respectively. Similarly, the disassociation rates, r_{3r} , r_{5r} , r_{7r} , and r_{8r} are increased from 74.94 to 94.94 to obtain comparable DiPhoBpp levels.

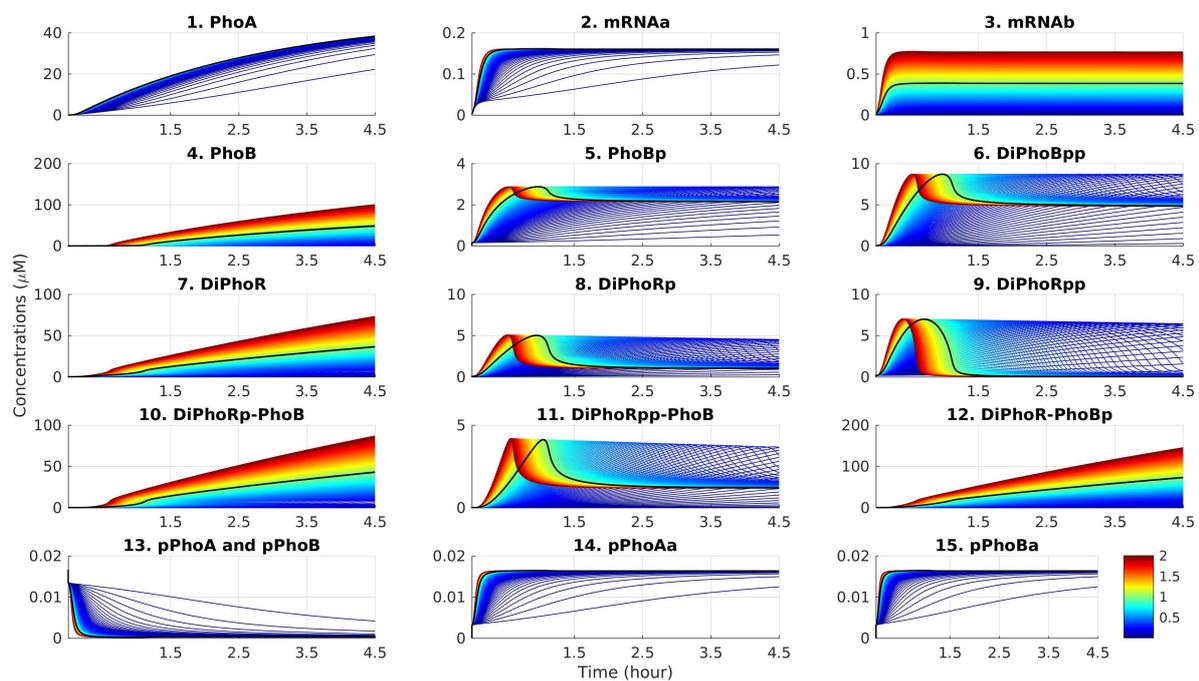


Fig. B.3 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{14} . The reaction rates in the model can vary within a range in different cell types and cellular environments. Therefore, we applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.01 to 2, which are larger than the physiological range. The 1 fold change is plotted in black.

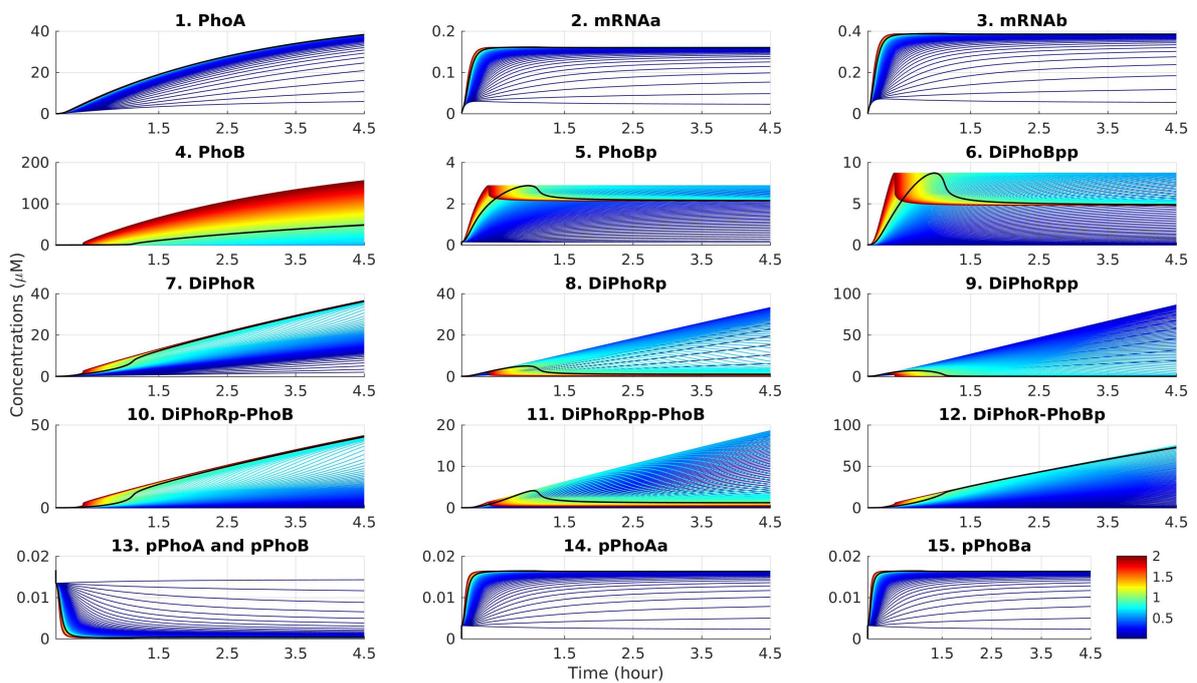


Fig. B.4 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{15} . The reaction rates in the model can vary within a range in different cell types and cellular environments. Therefore, we applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.01 to 2, which are larger than the physiological range. The 1 fold change is plotted in black.

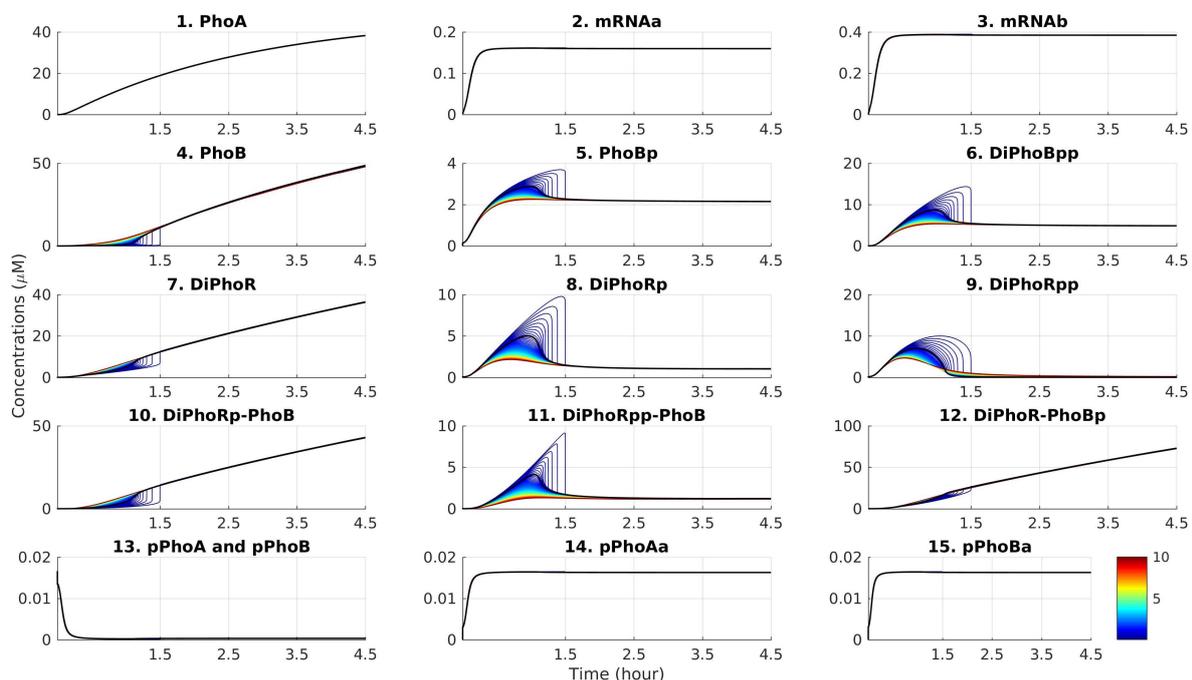


Fig. B.5 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{3r} . The reaction rates in the model can vary within a range in different cell types and cellular environments. We applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.1 to 10. The 1 fold change is plotted in black.

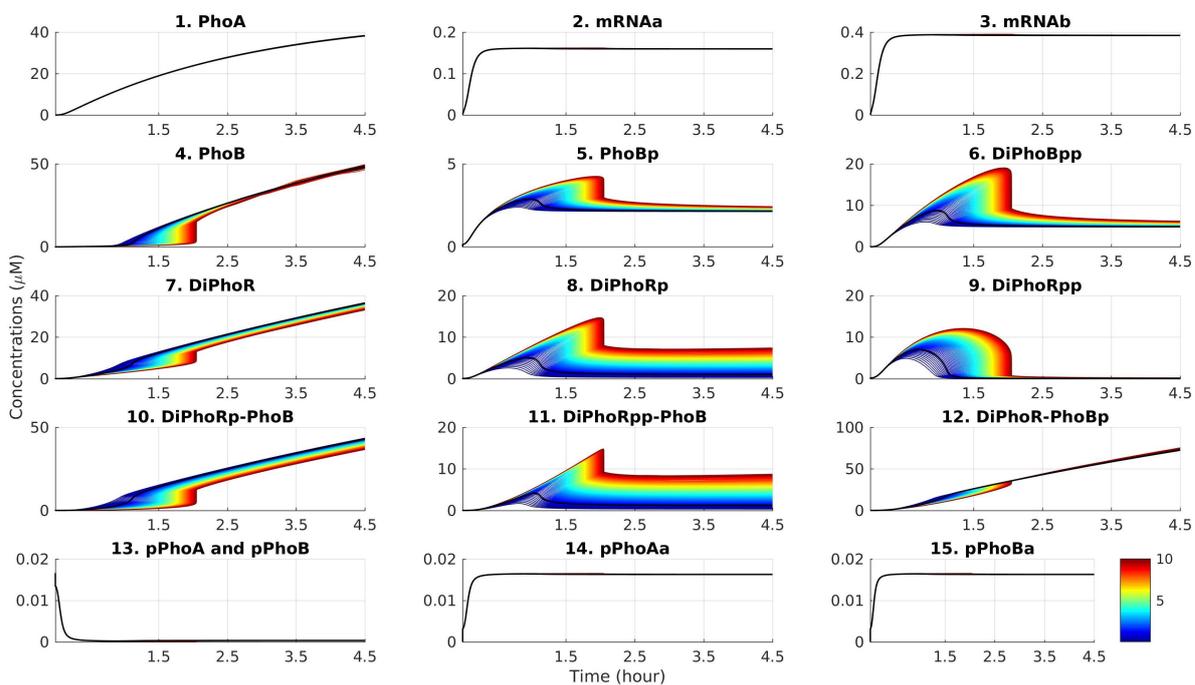


Fig. B.6 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{5r} . The reaction rates in the model can vary within a range in different cell types and cellular environments. We applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.1 to 10. The 1 fold change is plotted in black.

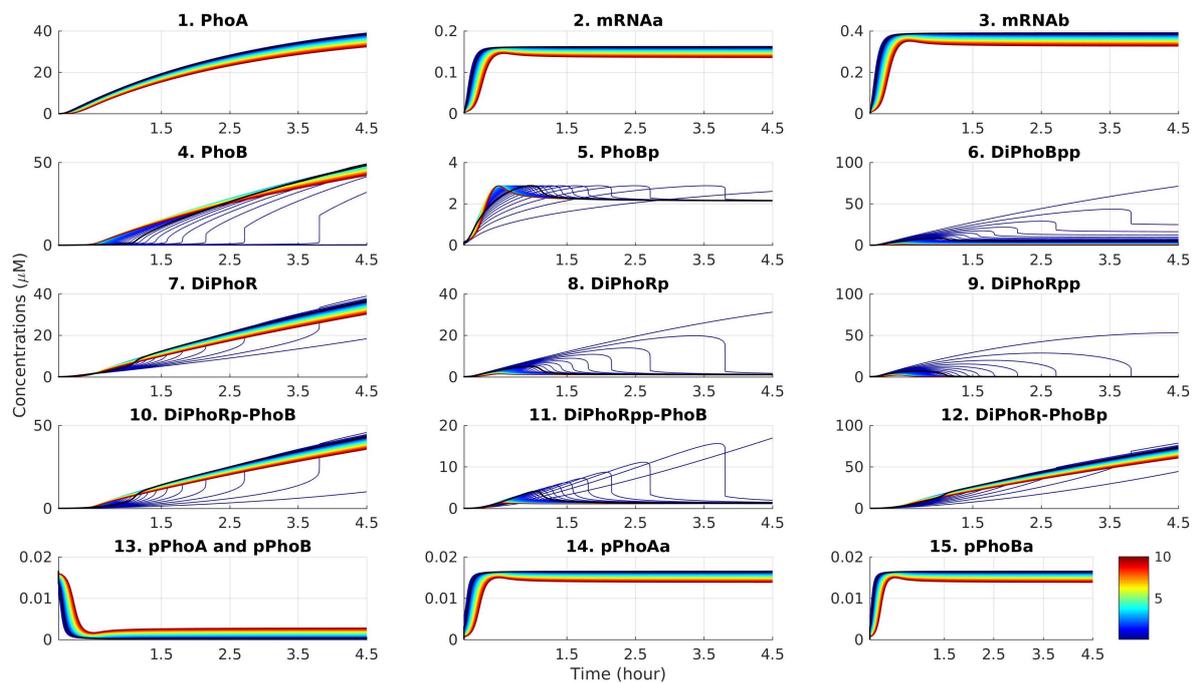


Fig. B.7 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{7r} . The reaction rates in the model can vary within a range in different cell types and cellular environments. We applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.1 to 10. The 1 fold change is plotted in black.

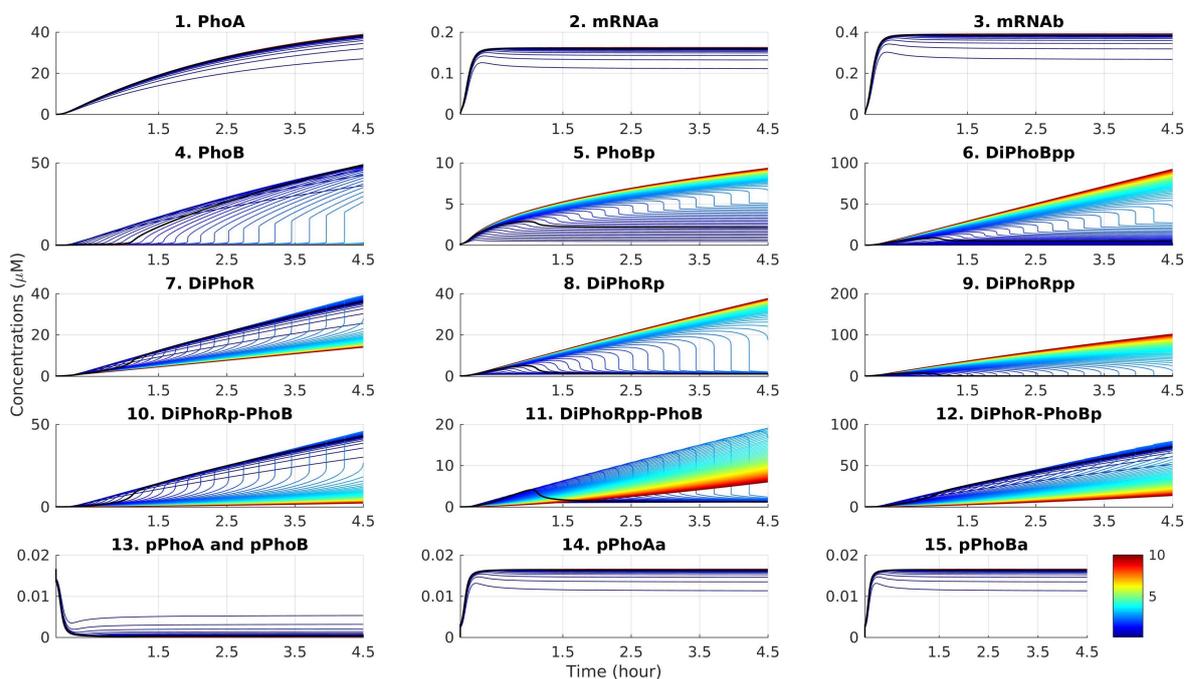


Fig. B.8 The cumulative output of simulations under the effect of random perturbations on the reaction rate r_{8r} . The reaction rates in the model can vary within a range in different cell types and cellular environments. We applied sensitivity analysis to quantify the model outputs with respect to such random perturbations. The rate is modified by applying a varied range of fold changes between 0.1 to 10. The 1 fold change is plotted in black.

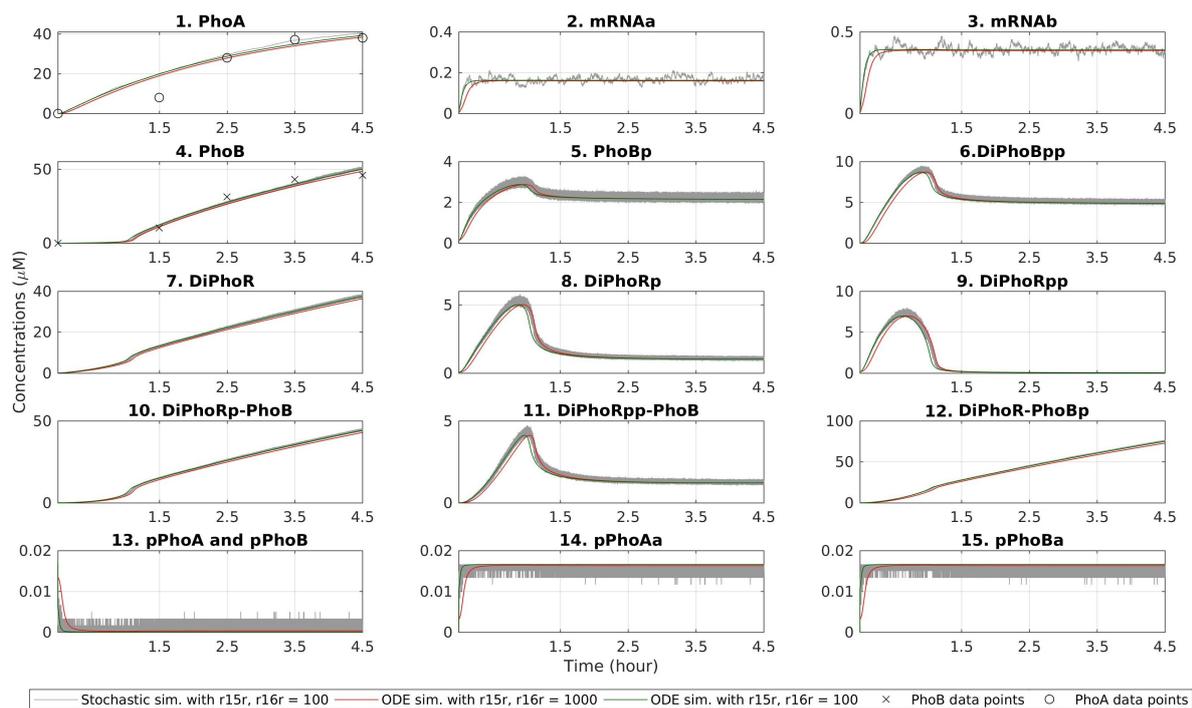


Fig. B.9 Stochastic simulation results when DNA unbinding rates r_{10} and r_{11} are set to 100s^{-1} together with deterministic simulation results. The unbinding kinetics can have a significant impact on the stochastic dynamics of gene expression network. When the promoter unbinding rate increases, the promoter activity, which works at a faster time-scale, leads to much more noisy fluctuations (see Supplementary Figures B.10 and B.11). A stochastic simulation with this modified rate is plotted in gray. The deterministic simulations with our model (red) and the deterministic simulations with the changes above (green) display similar dynamics.

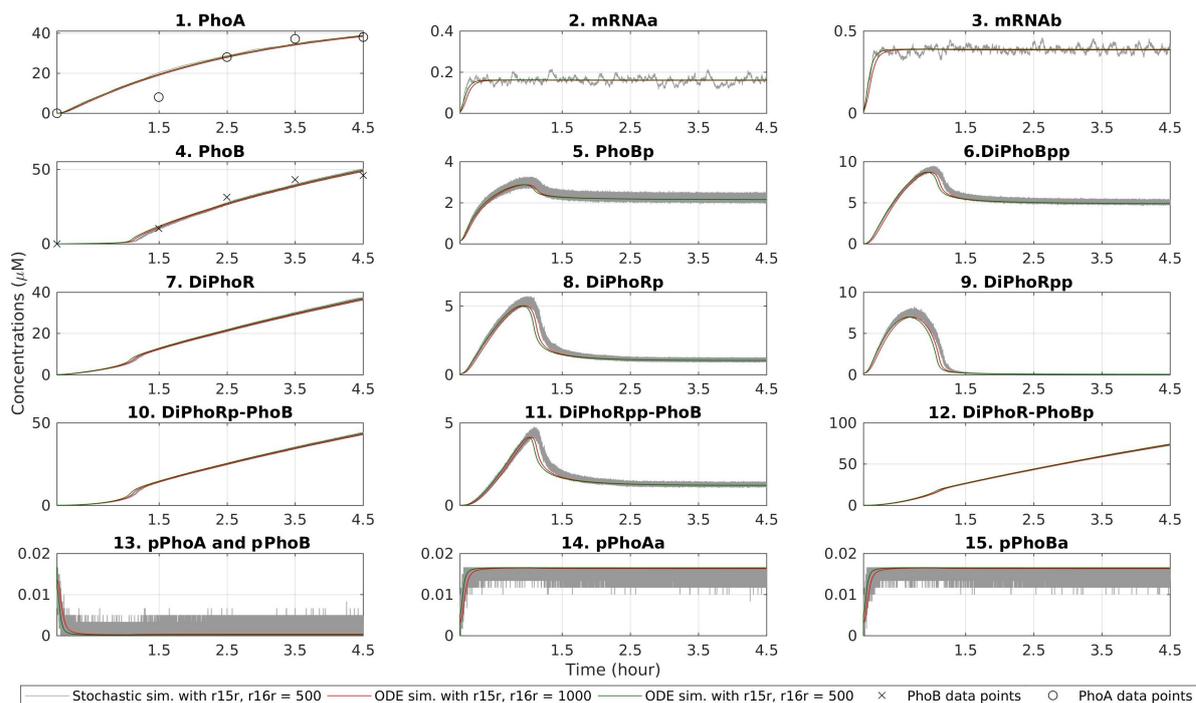


Fig. B.10 Stochastic simulation results when DNA unbinding rates r_{10} and r_{11} are set to $500s^{-1}$ together with deterministic simulation results. The unbinding kinetics can have a significant impact on the stochastic dynamics of gene expression network. When the promoter unbinding rate increases, the promoter activity, which works at a faster time-scale, leads to much more noisy fluctuations (see Supplementary Figures B.9 and B.11). A stochastic simulation with this modified rate is plotted in gray. The deterministic simulations with our model (red) and the deterministic simulations with the changes above (green) display similar dynamics.

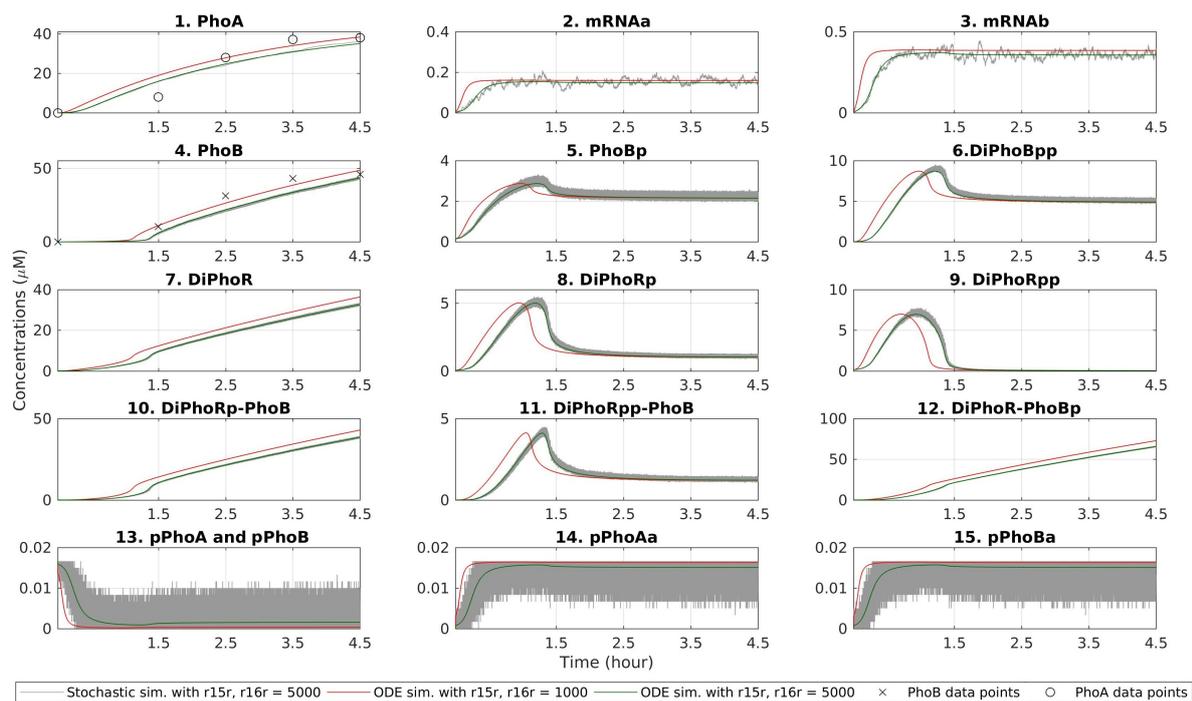


Fig. B.11 Stochastic simulation results when DNA unbinding rates r_{10} and r_{11} are set to 5000 s^{-1} together with deterministic simulation results. The unbinding kinetics can have a significant impact on the stochastic dynamics of gene expression network. When the promoter unbinding rate increases, the promoter activity, which works at a faster time-scale, leads to much more noisy fluctuations (see Supplementary Figures B.9 and B.10). A stochastic simulation with this modified rate is plotted in gray. The deterministic simulations with the changes above with slower unbinding rates (green) lead to slightly slower dynamics in comparison to deterministic simulations with our model with this rate set to 1000 s^{-1} (red).

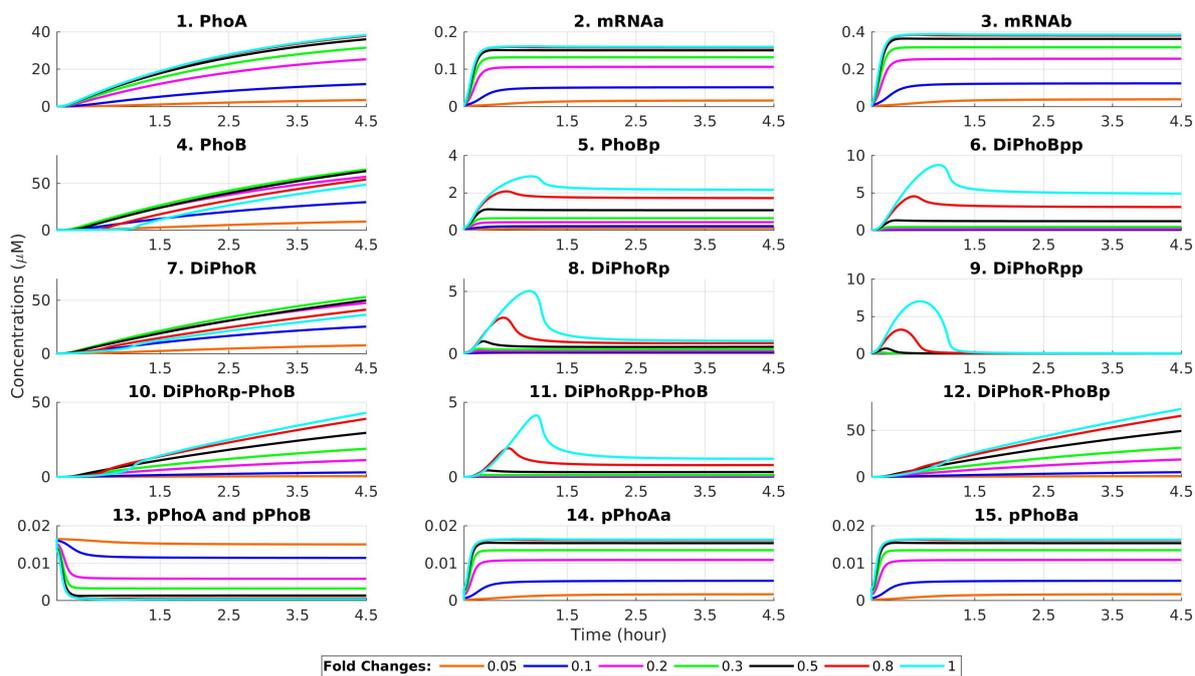


Fig. B.12 Simulation results with varying external P_i concentrations. A decrease in the external P_i concentration is modelled as a fold change, listed in the legend, which is applied to the autophosphorylation reactions of PhoR. A fold change of 1 gives the starvation condition with $0\mu\text{M}$ external P_i . A lower fold change value represents an increased PhoU activity. Active PhoU inhibits PhoR, and thereby prevents it from autophosphorylating, which otherwise relays the signal downstream to the promoters pPhoA and pPhoB.

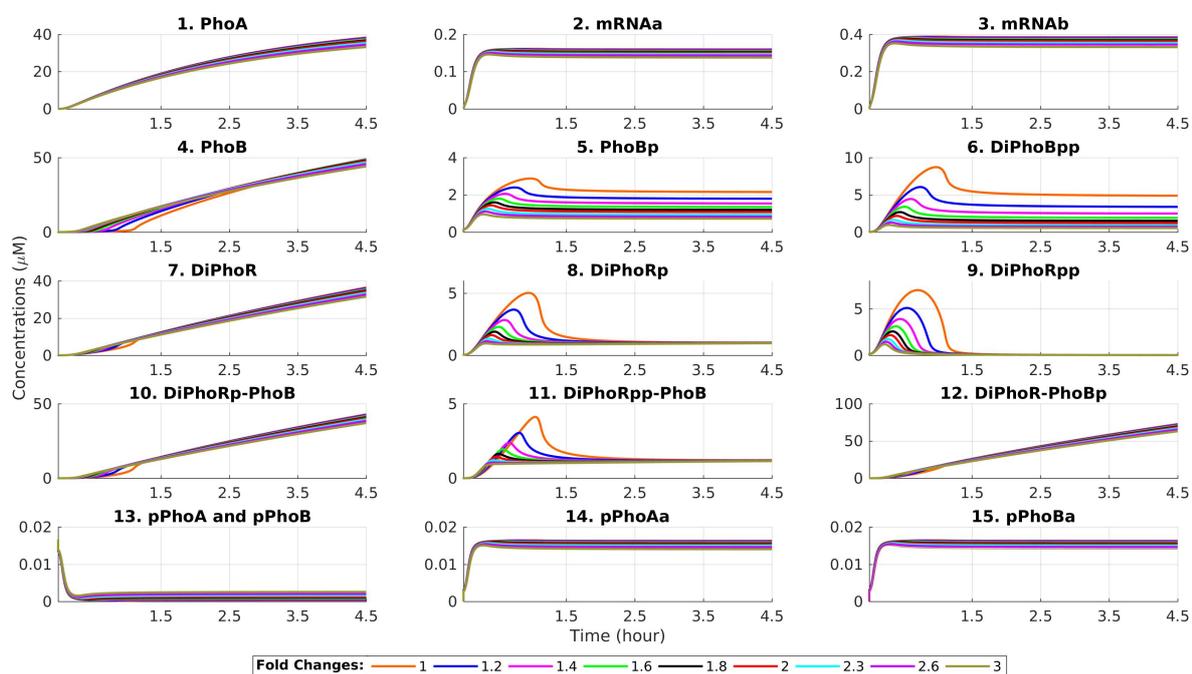


Fig. B.13 Deterministic simulations with the model with varying fold changes applied to r_8 , the association rate of PhoR and PhoB, whereby PhoR acts as a phosphatase. The plots display the effect of the perturbations on DiPhoBpp in comparison to the case, where the fold change is 1.

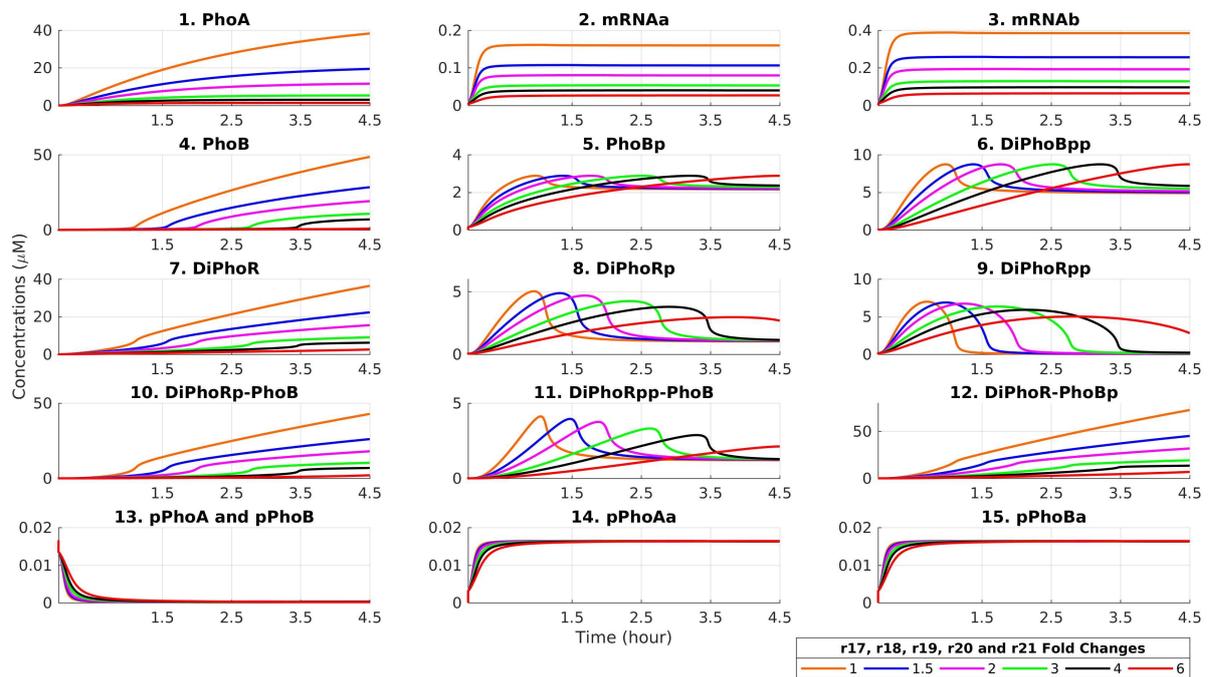


Fig. B.14 Deterministic simulation with the model with varying fold changes applied to degradation and dilution terms r_{17} , r_{18} , r_{19} , r_{20} , and r_{21} . Together with the Supplementary Figures B.15 and B.16, the plots display the effect of degradation and dilution under the conditions of reduced phosphorylation rates for PhoR resulting from an increase in external P_i concentration.

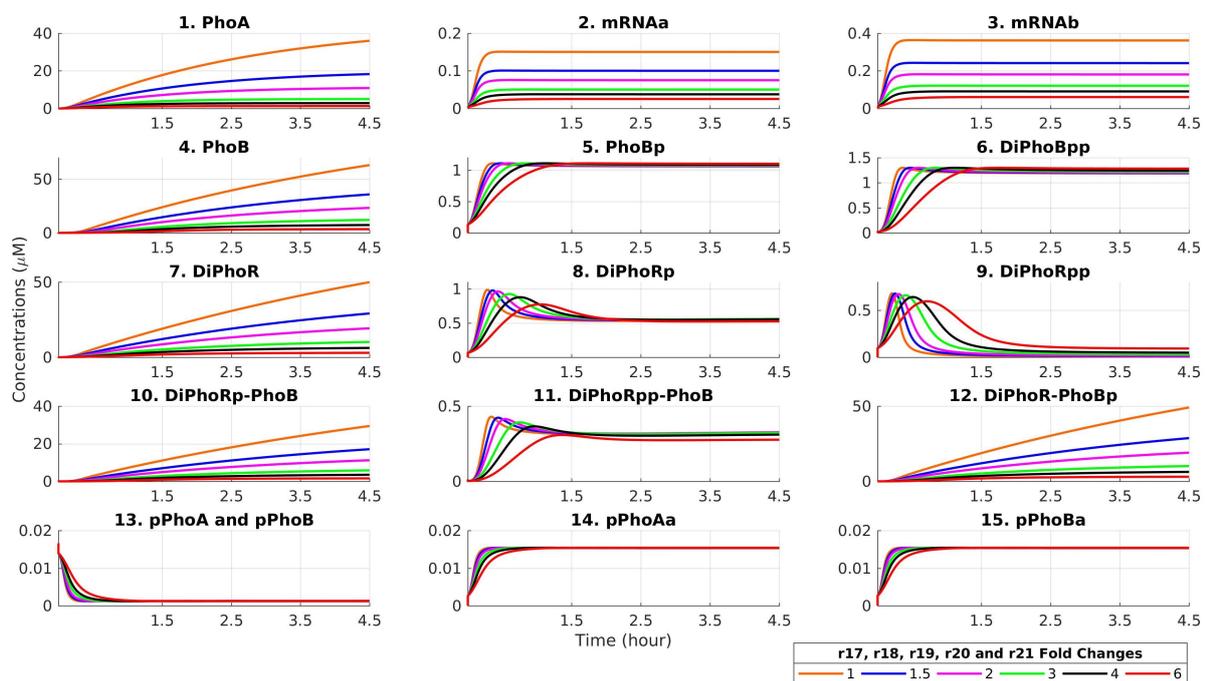


Fig. B.15 Deterministic simulation with the model with varying fold changes applied to degradation and dilution terms r_{17} , r_{18} , r_{19} , r_{20} , and r_{21} with a concomitant fold change of 0.5 applied to r_1 and r_2 . Together with the Supplementary Figures B.14 and B.16, the plots display the effect of degradation and dilution under the conditions of reduced phosphorylation rates for PhoR resulting from an increase in external P_i concentration.

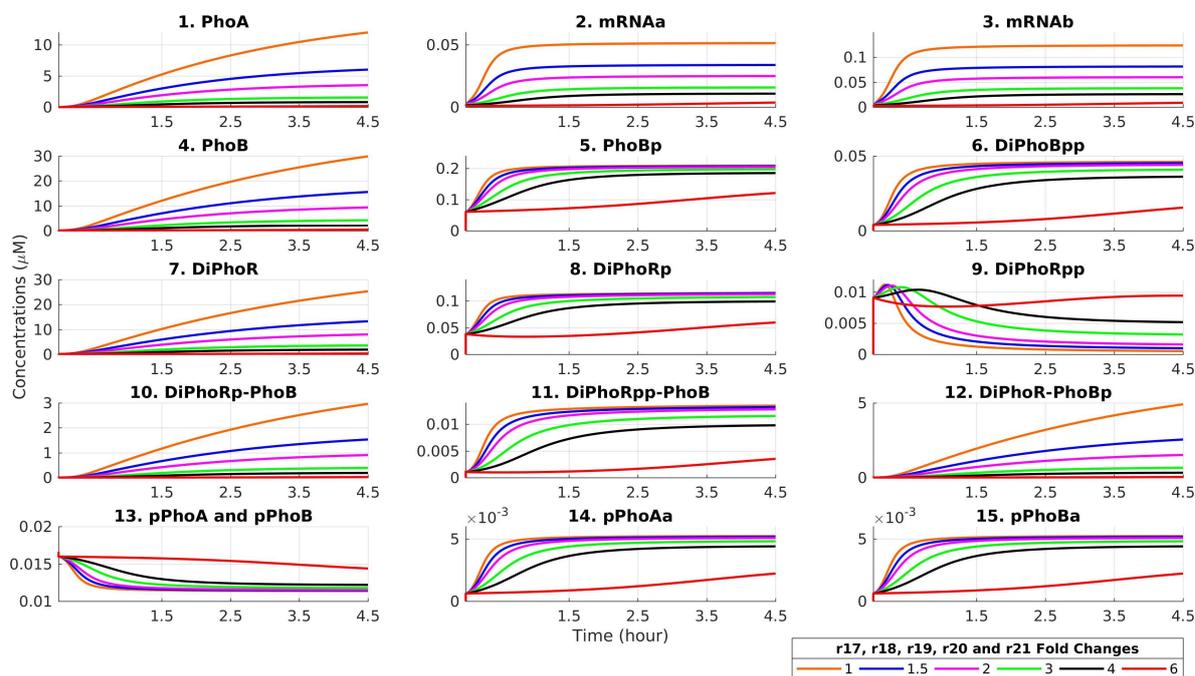


Fig. B.16 Deterministic simulation with the model with varying fold changes applied to degradation and dilution terms r_{17} , r_{18} , r_{19} , r_{20} , and r_{21} with a concomitant fold change of 0.1 applied to r_1 and r_2 . Together with the Supplementary Figures B.14 and B.15, the plots display the effect of degradation and dilution under the conditions of reduced phosphorylation rates for PhoR resulting from an increase in external P_i concentration.

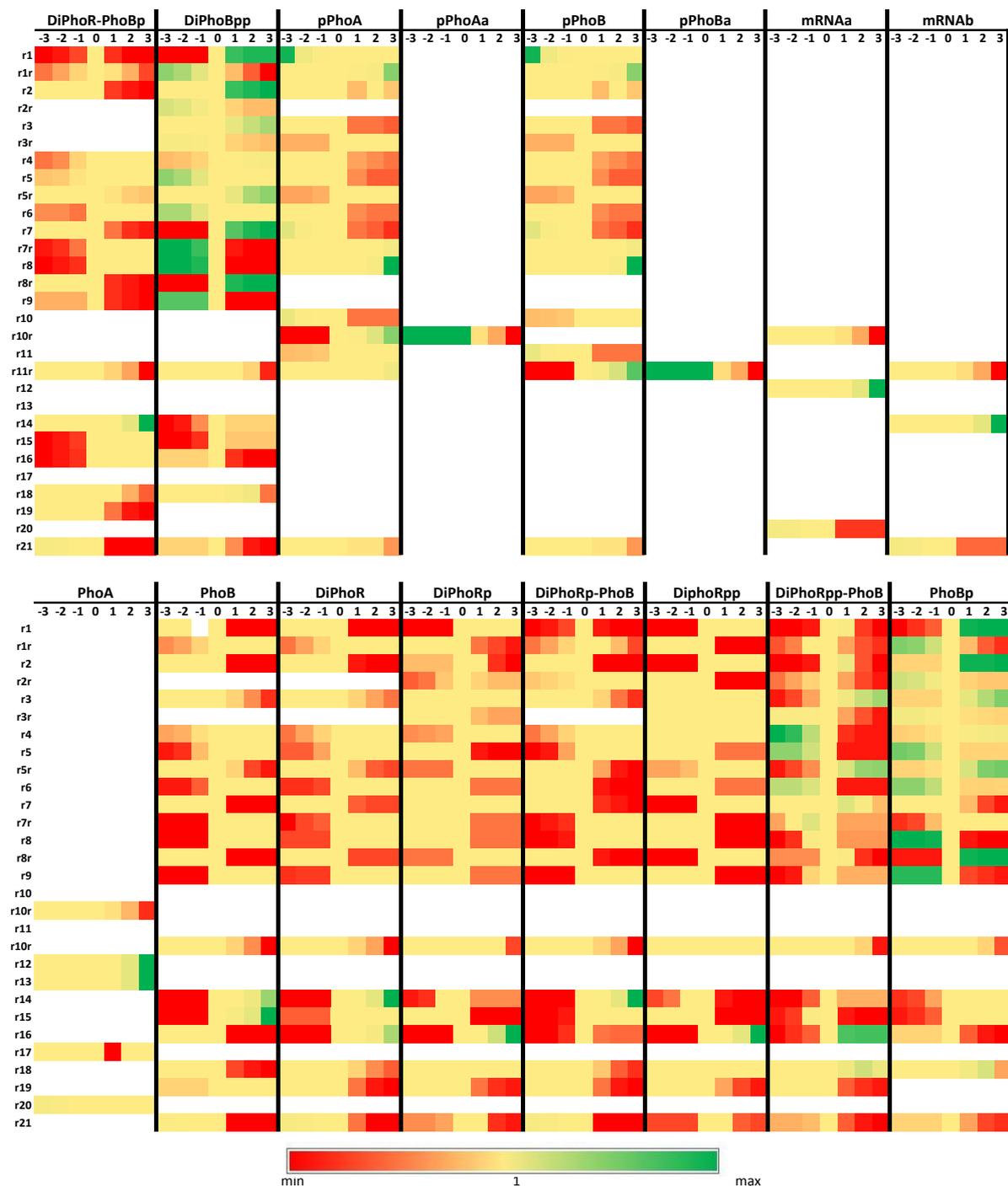


Fig. B.17 Heatmap displaying the results of the sensitivity analysis by considering all the model parameters. Each parameter is varied by 3 orders of magnitude higher and lower and the area under the curve (AUC) for each species is computed. The outcome is normalised with the AUC of the original model. Red represents the decreasing effect and green represents the increasing effect. The model is sensitive to a parameter when a variation in the input parameter causes a change in the model output. The amount of change quantifies the sensitivity. The sensitivity analysis results predict the system behaviour under varying conditions.

Appendix C

Appendix C

Here, we represent the Supplementary Materials for Chapter 5.

Unlike type 1 diabetes, type 2 diabetes patients' pancreas can usually produce insulin at normal levels; the problem is instead the ineffective use of the insulin produced. The adipocytes and muscle cells become less and less sensitive to insulin, which leads to increased levels of glucose in the bloodstream. Over time, diabetes can damage the heart, blood vessels, eyes, kidneys and nerves. Both genetics and environmental factors, such as excess body weight and physical inactivity, influence the outbreak of type 2 diabetes. Thus, the first step in treatment of the disease is a change of diet and more physical activity.

Also on a cellular level in type 2 diabetes there are many aspects to consider. One of these aspects is the insulin signalling network. Many research teams attempt to understand the mechanisms in the network and more details are discovered all the time. Mutations in the gene expressing IR do exist but are rare. Phosphorylation of IRS1 can be both increased and decreased. There can be defects at numerous points in the regulation system of the glucose transport. In this project the type 2 diabetes conditions at the cellular level was recreated according to Nyman et al. [127] through three parameters: the total number of IRs, the total number of GLUT4 and the entity of the positive feedback from the protein complex mTORC1, which have been all decreased according to literature. All the changes described work together to reduce the amount of GLUT4 that reaches the adipocyte membrane, which finally leads to a reduced glucose uptake from the adipose cell consistent with the type 2 diabetes conditions.

Table C.1 Initial values of model variables in agreement with a morning fasting state. For each variable two initial values are provided, one for the normal glucose regulation (NGR) condition and one for T2DM. Values have been computed according to the estimation procedure described in the text.

Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
S_0	4 mmol	optimised around 0 to avoid discontinuities in the dynamics	14 mmol	optimised around 0 to avoid discontinuities in the dynamics
L_0	14 mmol	optimised around 0 to avoid discontinuities in the dynamics	25 mmol	optimised around 0 to avoid discontinuities in the dynamics
G_0	5 mM	taken from [134]	7.5 mM	optimised in the range 8.05 ± 1.82 mM from [161]
I_0	60 pM	taken from [134]	180 pM	optimised in the range 171 ± 74 pM from [161]
W_0	10 pM	optimised in the range 10-19.4 pM from [77]	10 pM	as in NGR
E_0	34 pM	optimised in the range 37.86 ± 9.18 pM from [71]	36 pM	optimised in the range 42.19 ± 10.67 pM from [71]
C_0	3 mmol	optimised in the range 3-5 mmol from [136]	3 mmol	as in NGR
M_0	2.5 mmol	optimised around the value from [136] $\pm 5\%$	2.5 mmol	as in NGR

Table C.1: Continued on next page

Table C.1: continued from previous page

Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
A_0	53.19 <i>mmol</i>	optimised in the range 30-75 <i>mmol</i> from [136]	32 <i>mmol</i>	unconstrained optimisation
Y_0	0.4 <i>nM</i>	optimised in the range 0.11-1.75 <i>nM</i> from [92]	1.93 <i>nM</i>	unconstrained optimisation
Q_0	120 <i>pM</i>	optimised in the range 72.8-146.1 <i>pM</i> from [45]	53 <i>pM</i>	optimised in the range 50.6-78.9 <i>pM</i> from [45]
H_0	200 <i>mmol</i>	minimum daily glucose need taken from [41] divided by the number of meals	240 <i>mmol</i>	minimum daily glucose need taken from [41] divided by the number of meals
INS_{A0}	20 <i>pM</i>	unconstrained optimisation	70 <i>pM</i>	derived to preserve the same ratio I_0/INS_{A0} of the NGR condition
Gt_{A0}	135 <i>mg/kg</i>	taken from [106]	216 <i>mg/kg</i>	derived to preserve same the ratio G_0/Gt_{A0} of the NGR condition
IR_0	92.9271	optimised in the range 1.3298-99.8724 from [127]	43.4601	optimised in the range 0.7319-54.9280 from [127]

Table C.1: Continued on next page

Table C.1: continued from previous page

Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
IR_{YP_0}	0.0047	optimised in the range 0.0019-0.0438 from [127]	0.0058	optimised in the range 0.0010-0.0241 from [127]
$IRins_0$	6.7692	optimised in the range 0-98.3174 from [127]	11.1869	optimised in the range 0-54.0728 from [127]
IRi_{YP_0}	0.0433	optimised in the range 0.0200-0.3684 from [127]	0.0340	optimised in the range 0.0128-0.2007 from [127]
IRi_0	0.2557	optimised in the range 0.1057-2.4239 from [127]	0.3131	optimised in the range 0.0581-1.3338 from [127]
$IRS1_0$	70.8957	optimised in the range 32.8904-82.2711 from [127]	85.3981	optimised in the range 74.9450-86.2120 from [127]
$IRS1_{YP_0}$	0.0015	optimised in the range 0.0012-0.0079 from [127]	0.0027	optimised in the range 0.0010-0.0128 from [127]
$IRS1_{YP_{S307P_0}}$	1.6060	optimised in the range 0.3945-12.9125 from [127]	0.1011	optimised in the range 0.0108-2.6372 from [127]
$IRS1_{S307P_0}$	27.4967	optimised in the range 17.3332-61.8749 from [127]	14.4981	optimised in the range 13.7761-23.9259 from [127]

Table C.1: Continued on next page

Table C.1: continued from previous page

Symbol	NGR	Estimation Pro- cedure	T2DM	Estimation Pro- cedure
X_0	90.8250	optimised in the range 82.7446- 92.1517 from [127]	85.6979	optimised in the range 74.2898- 93.2612 from [127]
X_{P_0}	9.1750	optimised in the range 7.8483- 17.2554 from [127]	14.3021	optimised in the range 6.7388- 25.7102 from [127]
PKB_0	45.0895	optimised in the range 2.3591- 66.7261 from [127]	41.7141	optimised in the range 5.4955- 75.9049 from [127]
PKB_{T308P_0}	10.9438	optimised in the range 1.4914- 33.8840 from [127]	17.8142	optimised in the range 6.3922- 50.9079 from [127]
PKB_{S473P_0}	29.6843	optimised in the range 9.4834- 37.0490 from [127]	39.1462	optimised in the range 11.1110- 61.5802 from [127]
$PKB_{T308P_{S473P_0}}$	14.2824	optimised in the range 2.1594- 71.6300 from [127]	1.3253	optimised in the range 0.0771- 33.7498 from [127]
$mTORC1_0$	48.0015	optimised in the range 15.8212- 84.0562 from [127]	87.8940	optimised in the range 28.3447- 96.6769 from [127]

Table C.1: Continued on next page

Table C.1: continued from previous page

Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
$mTORC1a_0$	51.9985	optimised in the range 15.9438-84.1788 from [127]	12.1060	optimised in the range 3.3231-71.6553 from [127]
$mTORC2_0$	99.6572	optimised in the range 97.7928-99.8383 from [127]	99.7289	optimised in the range 98.9110-99.8963 from [127]
$mTORC2a_0$	0.3428	optimised in the range 0.1617-2.2072 from [127]	0.2711	optimised in the range 0.1037-1.0890 from [127]
$AS160_0$	56.5941	optimised in the range 24.6694-83.3321 from [127]	76.9252	optimised in the range 37.2542-93.1897 from [127]
$AS160_T642P_0$	43.4059	optimised in the range 16.6679-75.3306 from [127]	23.0747	optimised in the range 6.8103-62.7458 from [127]
$GLUT4m_0$	49.1528	optimised in the range 27.0715-62.6541 from [127]	16.9725	optimised in the range 6.5848-29.1440 from [127]
$GLUT4_0$	50.8472	optimised in the range 37.3459-72.9285 from [127]	33.0275	optimised in the range 20.8560-43.4152 from [127]

Table C.1: Continued on next page

Table C.1: continued from previous page

Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
$S6K_0$	86.3744	optimised in the range 83.8342-95.5400 from [127]	95.5986	optimised in the range 89.0410-99.0364 from [127]
$S6K_{T389P_0}$	13.6256	optimised in the range 4.4600-16.1658 from [127]	4.4014	optimised in the range 0.9636-10.9590 from [127]
$S6_0$	40.4757	optimised in the range 36.5506-67.6139 from [13]	69.5055	optimised in the range 45.9399-90.6217 from [13]
$S6_{S235_{S236P_0}}$	59.5243	optimised in the range 32.3861-63.4494 from [13]	30.4945	optimised in the range 9.3783-54.0601 from [13]

Table C.1: It ends from the previous page.

Table C.2 Parameter estimates of the model in the normal glucose regulation (NGR) and T2DM conditions. Values have been computed according to the estimation procedure described in the text.

Parameter description	Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
Insulin independent glucose utilisation	b_1	0.0059 min^{-1}	calculated as in [161]	0.0049 min^{-1}	taken from [161]
Insulin disappearance rate	b_2	0.1262 min^{-1}	taken from [31]	0.04 min^{-1}	taken from [161]

Table C.2: Continued on next page

Table C.2: continued from previous page

Parameter description	Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
Insulin dependent glucose utilisation	b_3	$0.00005 (pM.min)^{-1}$	taken from [31]	$0.00002 (pM.min)^{-1}$	taken from [161]
Glucose dependent insulin secretion	b_4	$0.4543 pM/mM.min$	calculated by steady state as in [161]	$0.15 pM/mM.min$	calculated by steady state as in [161]
Glucose transfer from liver to plasma	b_5	$0.185 min^{-1}$	calculated by steady state as in [161]	$0.3320 min^{-1}$	calculated by steady state as in [161]
Intestine glucose dependent incretin secretion	b_6	$0.0102 pM/mmol.min$	set equal to the estimate computed for T2DM according to [110]	$0.0102 pM/mmol.min$	optimised in the range 0.0102-0.0104 $pM/mmol.min$ from [161]
Incretin disappearance rate	b_7	$0.03 min^{-1}$	set equal to the estimate computed for T2DM according to [110]	$0.03 min^{-1}$	optimised in the range 0.03-0.06 min^{-1} from [161]
Stomach glucose emptying rate	b_8	$0.022 min^{-1}$	optimised in the range $0.020 \pm 0.002 min^{-1}$ from [144]	$0.022 min^{-1}$	optimised in the range $0.020 \pm 0.002 min^{-1}$ from [144]
Glucose transfer rate to the stomach	b_9	$0.022 min^{-1}$	unconstrained optimisation	$0.022 min^{-1}$	unconstrained optimisation

Table C.2: Continued on next page

Table C.2: continued from previous page

Parameter description	Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
Intestine glucose emptying rate	b_{10}	0.022 min^{-1}	optimised in the range $0.020 \pm 0.002 \text{ min}^{-1}$ from [144]	0.022 min^{-1}	optimised in the range $0.020 \pm 0.002 \text{ min}^{-1}$ from [144]
Ghrelin disappearance	b_{11}	0.02 min^{-1}	set equal to the estimate computed for T2DM	0.02 min^{-1}	taken from [161]
The appearance constant for ghrelin	b_{12}	$28.66 \text{ pM}/\text{min}$	calculated by steady state as in [161]	$7 \text{ pM}/\text{min}$	calculated by steady state as in [161]
Leptin secretion rate	b_{13}	$0.0000095 \text{ nM}/\text{mmol}.\text{min}.\text{kg}$	calculated by steady state	$0.000061 \text{ nM}/\text{mmol}.\text{min}.\text{kg}$	calculated by steady state
Leptin disappearance	b_{14}	0.0278 min^{-1}	taken from [62]	0.0278 min^{-1}	as in NGR
Ghrelin dependent glucose intake appearance	b_{17}	$0.7 \text{ mmol}/\text{pM}.\text{min}$	calculated by steady state	$1.2 \text{ mmol}/\text{pM}.\text{min}$	calculated by steady state
Leptin inhibition on glucose intake	b_{18}	0.35 nM^{-1}	taken from [60]	0.23 nM^{-1}	calculated by as in [7]
Glucose effect rate on glucose intake	b_{19}	$0.004 \text{ (mM}.\text{min)}^{-1}$	taken from [66]	$0.004 \text{ (mM}.\text{min)}^{-1}$	taken from [66]
Glucogan action on the liver	b_{21}	$0.008764 \text{ mmol}/\text{pM}.\text{min}$	taken from [154]	$0.008764 \text{ mmol}/\text{pM}.\text{min}$	as in NGR
Glucose action on the liver	b_{22}	0.0021 min^{-1}	taken from [106]	$0.008764 \text{ mmol}/\text{pM}.\text{min}$	taken from [106]

Table C.2: Continued on next page

Table C.2: continued from previous page

Parameter description	Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
Liver glucose constant production	b_{23}	$0.08 \text{ mmol}/\text{min}$	taken from [106]	$0.12 \text{ mmol}/\text{min}$	taken from [106]
Insulin action on the liver	b_{25}	$0.00026 \text{ mmol}/\text{pM}\cdot\text{min}$	taken from [106]	$0.00014 \text{ mmol}/\text{pM}\cdot\text{min}$	taken from [106]
Muscle glucose disappearance	b_{27}	0.014 min^{-1}	calculated by steady state	0.0099 min^{-1}	calculated by steady state
Incretin dependent insulin secretion	c	$0.1060 \text{ (mM}\cdot\text{min)}^{-1}$	calculated by steady state as in [161]	$0.035 \text{ (mM}\cdot\text{min)}^{-1}$	calculated by steady state as in [161]
Glucagon basal secretion	c_0	$1.8854 \text{ (pM}/\text{min)}$	taken from [154]	$1.8854 \text{ (pM}/\text{min)}$	taken from [154]
Glucose action on glucagon	c_1	$198 \text{ (pM}/\text{mM}\cdot\text{min)}$	taken from [154]	$198 \text{ (pM}/\text{mM}\cdot\text{min)}$	taken from [154]
Insulin action on glucagon	c_2	$94 \text{ (pM}/\text{mM}\cdot\text{min)}$	set equal to the estimate computed for T2DM	$94 \text{ (pM}/\text{mM}\cdot\text{min)}$	optimised to have the plasma glucagon dynamics inside the range discussed in [71]
Glucagon disappearance	c_3	0.0554 min^{-1}	calculated by steady state as in [154]	0.0524 min^{-1}	calculated by steady state as in [154]
Insulin effectiveness	e	1	taken from [154]	0.2	taken from [154]

Table C.2: Continued on next page

Table C.2: continued from previous page

Parameter description	Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
The fraction of absorbed glucose	f	0.9	taken from [161]	0.9	taken from [161]
Averaged total fat mass in humans	Fat	22 kg	calculated as in [62] assuming a standard body weight of 75 kg	27 kg	calculated as in [62] a standard body weight of 90 kg
Glucose threshold value	G_e	5 mM	taken from [93]	8 mM	taken from [93]
Glucose in stomach dependent decay rate of ghrelin	l	0.006 mmol ⁻¹	calculated as in [161]	0.006 mmol ⁻¹	calculated as in [161]
Insulin dependent decay rate of ghrelin	m	0.04 pM ⁻¹	calculated as in [161]	0.01 pM ⁻¹	calculated as in [161]
Insulin dependent decay rate of ghrelin glucose intake	r	0.04 pM ⁻¹	unconstrained optimisation	0.01 pM ⁻¹	unconstrained optimisation
Incretin constant secretion	s	0.03 pM/min	set equal to the estimate computed for T2DM [110]	0.01 pM/min	optimised around value $\pm 5\%$ as calculated in [15]
Glucose distribution volume	v	15 l	calculated as in [161]	18 l	calculated as in [161]

Table C.2: Continued on next page

Table C.2: continued from previous page

Parameter description	Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
Insulin binding with IR	k_{1a}	0.00333 ($pM.min$) ⁻¹	optimised in the range $1e^{-6} - 1e^6$ from [127]	0.00333 ($pM.min$) ⁻¹	optimised in the range $1e^{-6} - 1e^6$ from [127]
Basal phosphorylation of IR	k_{1basal}	0.0368 min^{-1}	taken from [127]	0.0368 min^{-1}	taken from [127]
Phosphorylation of IRins	k_{1c}	0.877 min^{-1}	taken from [127]	0.877 min^{-1}	taken from [127]
Endocytosis process of the receptor	k_{1d}	31.0 min^{-1}	taken from [127]	31.0 min^{-1}	taken from [127]
Feedback from protein X	k_{1f}	0.368 min^{-1}	taken from [127]	0.368 min^{-1}	taken from [127]
Basal dephosphorylation of IR	k_{1g}	1940 min^{-1}	taken from [127]	1940 min^{-1}	taken from [127]
Recycling rate of IRi	k_{1r}	0.547 min^{-1}	taken from [127]	0.547 min^{-1}	taken from [127]
Phosphorylation of IRS1 from the receptor	k_{2a}	3.23 min^{-1}	taken from [127]	3.23 min^{-1}	taken from [127]
Phosphorylation of IRS1 in the serine site from mTORC1a	k_{2c}	5760 min^{-1}	taken from [127]	5760 min^{-1}	taken from [127]
Basal phosphorylation of IRS1 in the serine site	k_{2basal}	0.0423 min^{-1}	taken from [127]	0.0423 min^{-1}	taken from [127]
Dephosphorylation of IRS1 in the tyrosine site	k_{2b}	3420 min^{-1}	taken from [127]	3420 min^{-1}	taken from [127]

Table C.2: Continued on next page

Table C.2: continued from previous page

Parameter description	Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
Dephosphorylation of IRS1 in the serine site	k_{2d}	281 min^{-1}	taken from [127]	281 min^{-1}	taken from [127]
Dephosphorylation of IRS1 in the tyrosine site	k_{2f}	2.91 min^{-1}	taken from [127]	2.91 min^{-1}	taken from [127]
Dephosphorylation of IRS1 in the serine site	k_{2g}	0.267 min^{-1}	taken from [127]	0.267 min^{-1}	taken from [127]
Activation of X protein	k_{3a}	6.90 min^{-1}	taken from [127]	6.90 min^{-1}	taken from [127]
Deactivation of X protein	k_{3b}	0.0988 min^{-1}	taken from [127]	0.0988 min^{-1}	taken from [127]
Phosphorylation of PKB in the threonine site from IRS1	k_{4a}	5790 min^{-1}	taken from [127]	5790 min^{-1}	taken from [127]
Dephosphorylation of PKB in the threonine site	k_{4b}	34.8 min^{-1}	taken from [127]	34.8 min^{-1}	taken from [127]
Phosphorylation of PKB in the serine site from mTORC2	k_{4c}	4.46 min^{-1}	taken from [127]	4.46 min^{-1}	taken from [127]
Phosphorylation of PKB in the serine site from IRS1	k_{4e}	42.8 min^{-1}	taken from [127]	42.8 min^{-1}	taken from [127]
Dephosphorylation of PKB in the threonine site	k_{4f}	144 min^{-1}	taken from [127]	144 min^{-1}	taken from [127]

Table C.2: Continued on next page

Table C.2: continued from previous page

Parameter description	Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
Dephosphorylation of PKB in the serine site	k_{4h}	0.536 min^{-1}	taken from [127]	0.536 min^{-1}	taken from [127]
Activation of mTORC1 complex from PKB	k_{5a1}	1.84 min^{-1}	taken from [127]	1.84 min^{-1}	taken from [127]
Activation of mTORC1 complex from PKB	k_{5a2}	0.0551 min^{-1}	taken from [127]	0.0551 min^{-1}	taken from [127]
Deactivation of mTORC1	k_{5b}	24.8 min^{-1}	taken from [127]	24.8 min^{-1}	taken from [127]
Activation of mTORC2 complex	k_{5c}	0.0858 min^{-1}	taken from [127]	0.0858 min^{-1}	taken from [127]
Deactivation of mTORC2 complex	k_{5d}	1.06 min^{-1}	taken from [127]	1.06 min^{-1}	taken from [127]
Phosphorylation of AS160 from PKB	k_{6a1}	2.65 min^{-1}	taken from [127]	2.65 min^{-1}	taken from [127]
Phosphorylation of AS160 from PKB	k_{6a2}	0.410 min^{-1}	taken from [127]	0.410 min^{-1}	taken from [127]
Dephosphorylation of AS160	k_{6b}	65.2 min^{-1}	taken from [127]	65.2 min^{-1}	taken from [127]
Translocation of GLUT4 to the membrane	k_{7a}	51.0 min^{-1}	taken from [127]	51.0 min^{-1}	taken from [127]
Translocation of GLUT4 to the cytosol	k_{7b}	2290 min^{-1}	taken from [127]	2290 min^{-1}	taken from [127]

Table C.2: Continued on next page

Table C.2: continued from previous page

Parameter description	Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
GLUT4 glucose uptake	k_8	0.5275 <i>mmol/min</i>	taken from [126] changing units	0.5644 <i>mmol/min</i>	taken from [126] changing units
GLUT1 glucose uptake	GLUT1	0.0283 <i>mmol/min</i>	taken from [126] changing units	0.032 <i>mmol/min</i>	taken from [126] changing units
Dependence on interstitial glucose saturated	$kmG4$	146.851 <i>mg/kg</i>	taken from [126]	146.851 <i>mg/kg</i>	kept same
Dependence on interstitial glucose saturated	$kmG1$	1.082 <i>mg/kg</i>	taken from [126]	1.082 <i>mg/kg</i>	kept same
Intra-adipocitary glucose elimination rate	k_{gluc}	0.25 min^{-1}	unconstrained optimisation	0.25 min^{-1}	unconstrained optimisation
Interstitial insulin rate of change	p_{2U}	0.033 min^{-1}	taken from [106]	0.033 min^{-1}	kept same
Interstitial glucose elimination rate	q_1	0.0031 min^{-1}	unconstrained optimisation	0.0025 min^{-1}	unconstrained optimisation
Glucose transfer rate from plasma to interstitium	q_2	0.4054 <i>(mg/kg)/mM.min</i>	unconstrained optimisation	0.4365 <i>(mg/kg)/mM.min</i>	unconstrained optimisation
Phosphorylation of S6K	k_{9a}	0.0013 min^{-1}	taken from ([127])	0.0013 min^{-1}	taken from [127]
Dephosphorylation of S6K	k_{9b}	0.0444 min^{-1}	taken from [127]	0.0444 min^{-1}	taken from [127]
Dephosphorylation of S6	k_{9b2}	30.9966 min^{-1}	taken from [13]	30.9966 min^{-1}	taken from [13]

Table C.2: Continued on next page

Table C.2: continued from previous page

Parameter description	Symbol	NGR	Estimation Procedure	T2DM	Estimation Procedure
Phosphorylation of S6	k_{ϕ_2}	3.3289 min^{-1}	taken from [13]	3.3289 min^{-1}	taken from [13]
Positive feedback from mTORC1 reduced in T2DM condition	k_{fb}	1	taken from [127]	0.15	taken from [127]

Table C.2: It ends from the previous page.

C.1 Sensitivity analysis

A state of the art sensitivity analysis has been computed for each variable of the whole body model to assess the effect on the model dynamics of each estimated parameter. During the analysis, we did not consider the cellular model because its equations and parameters have been all taken from [127], [13] and [101] and therefore we refer to such references for any additional insight related to this layer of the model. Following the same approach, we did not consider the parameters of the whole body model that have been taken from literature without the need of being estimated. Conversely, the effect of each estimated parameter has been analysed by considering their estimate both in the NGR condition and in T2DM, according to the values reported in C.2. For computing the analysis, each parameter estimate has been perturbed by multiplying its value by a parameter k varying in the range 0.5 - 2 (from halved to doubled parameter estimate). For each value of the k parameter, we ran a model simulation and then we measured the impact on the system dynamics of each model variable in terms of AUC ratio:

$$AUC_{ratio}^k = AUC_{perturbed}^k / AUC_{original}$$

where AUC indicates the area under the curve provided by the corresponding simulated behaviour in time. This allowed us to compute the set of charts included in Supplementary figures, which show the variation of the AUC ratio of each model variable for each value of k in the two modelled conditions (NGR and T2DM). To improve the readability of results,

only the parameters with

$$\max AUC_{ratio}^k - \min AUC_{ratio}^k \geq 15$$

have been included in the plots.

Interestingly, the computed AUC ratios exhibit close-to-linear differentiable patterns with respect to the value of k , when the parameter is close to 1. This could indicate that model equations have the right structure to allow parameter identifiability. Only few parameters exhibit not-differentiable points in the behaviour of their AUC ratio, but this is due to the thresholds employed in the model, which are all justified by literature. Moreover, all the estimated parameters affect the dynamics of at least one model variable, which is another key requirement for parameter identifiability. The only parameter that showed to have a negligible impact on the dynamics of all model variables in the NGR condition is c_2 , but this does not constitute a problem because it has been estimated in the T2DM condition, where an impact on the system dynamics can be appreciated. This can be easily checked by looking at the two heatmaps provided in Supplementary Figure C.1 and Figure C.2, which further resume the results of the sensitivity analysis. The two figures, one for the NGR condition and one for T2DM, depict in one single image the differences between the two AUC ratios computed with the maximum and the minimum considered value of k

$$AUC_{ratio}^{k=2} - AUC_{ratio}^{k=0.5}$$

for each model variable and for each estimated parameter (differences lower than 0.15 have been not displayed). This provides a general view of the impact of each estimated parameter in both the NGR and T2DM conditions.

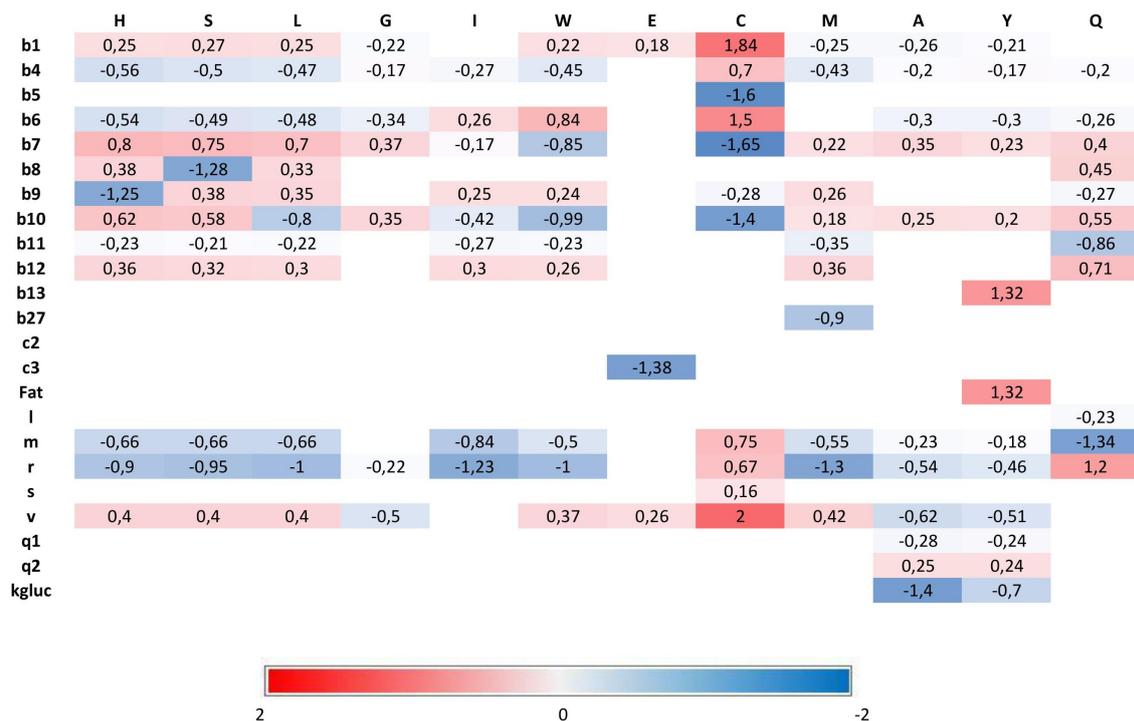


Fig. C.1 Heatmap resuming the results of the sensitivity analysis by considering the parameter estimates computed for the normal glucose regulation (NGR) condition. For each parameter, the differences between the two AUC ratios computed with the maximum and the minimum considered value of k are provided with respect to each model variable of the whole body model. Values lower than 0.15 have been considered negligible and are not displayed in the heatmap. The parameter c_2 does not affect model dynamics in the NGR condition, however, it has been estimated in the T2DM condition, where an impact on the system dynamics can be appreciated.

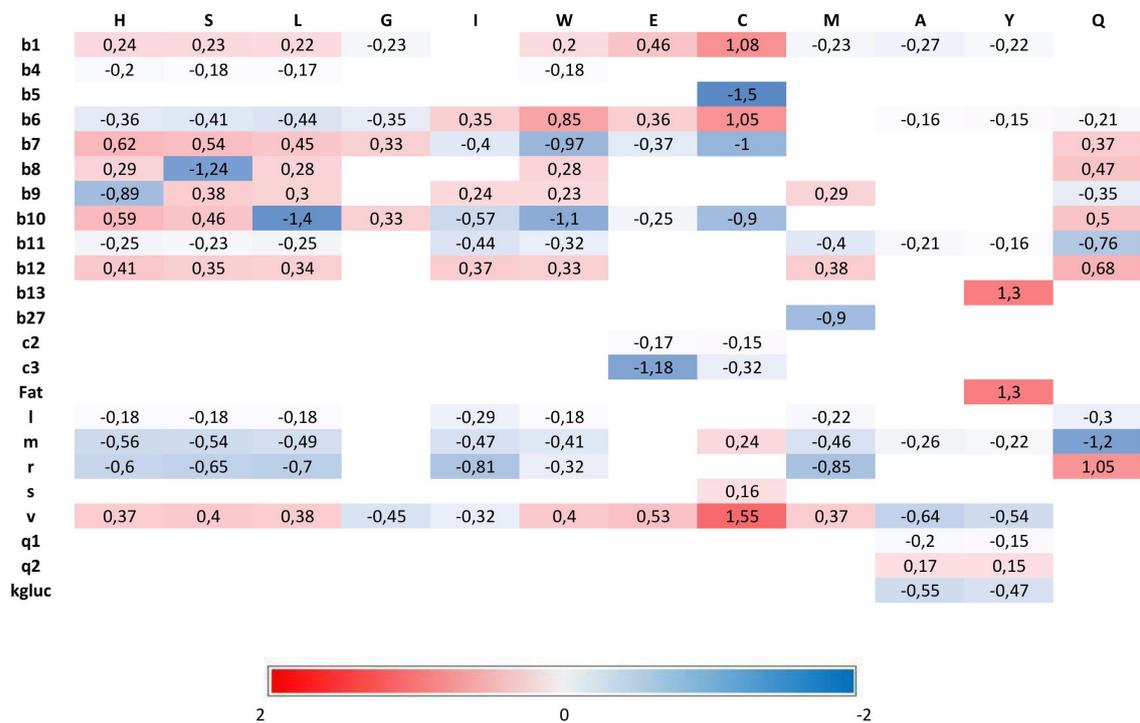


Fig. C.2 Heatmap resuming the results of the sensitivity analysis by considering the parameter estimates computed for the T2DM condition. For each parameter, the differences between the two AUC ratios computed with the maximum and the minimum considered value of k are provided, with respect to each model variable of the whole body model. Values lower than 0.15 have been considered negligible and are not displayed in the heatmap.

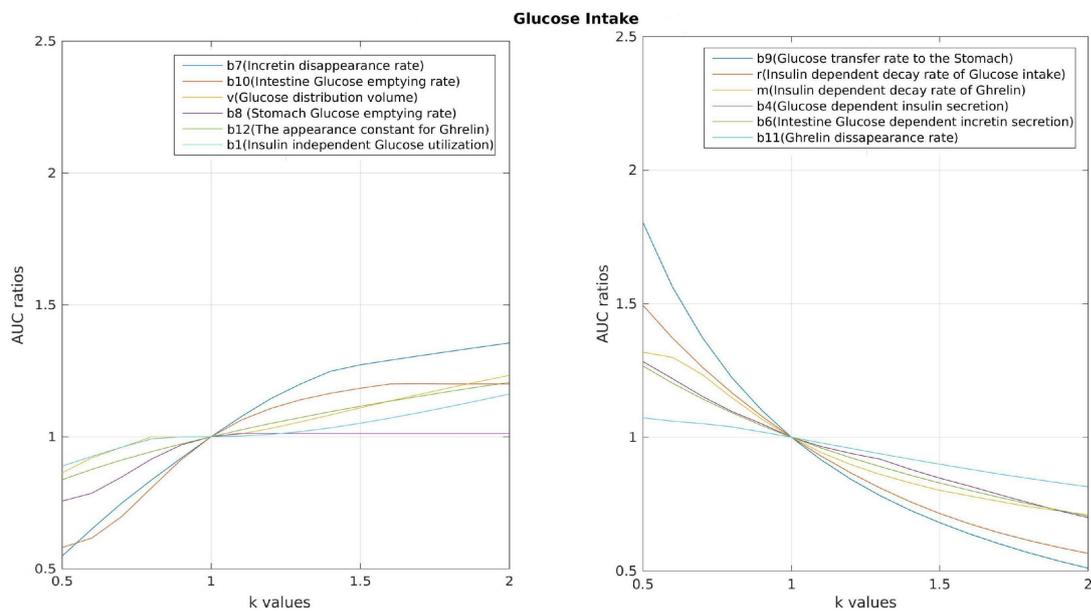


Fig. C.3 Oral glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

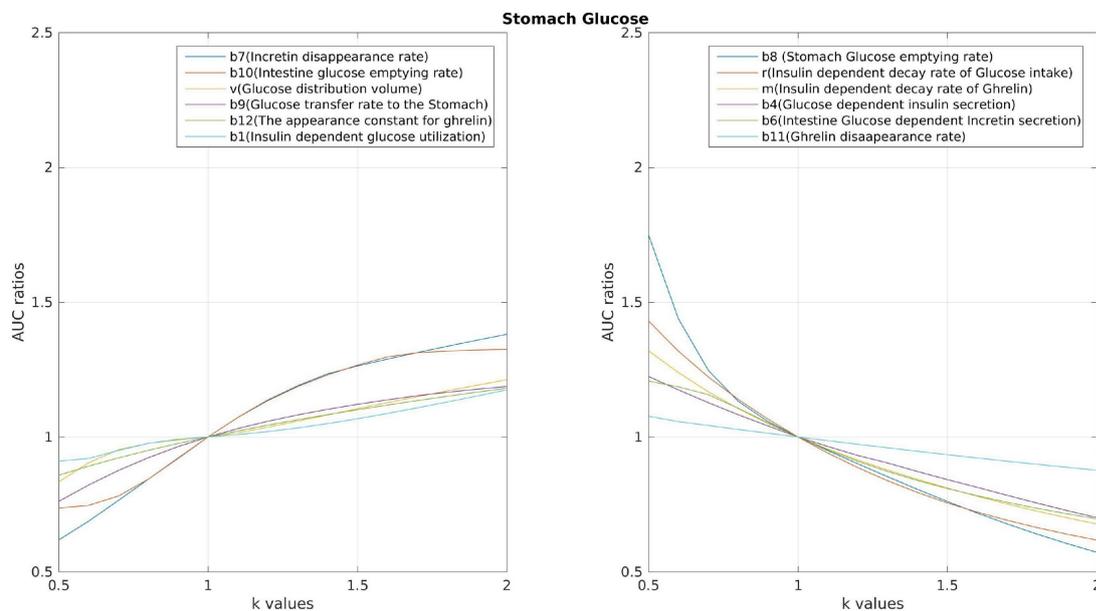


Fig. C.4 Stomach glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

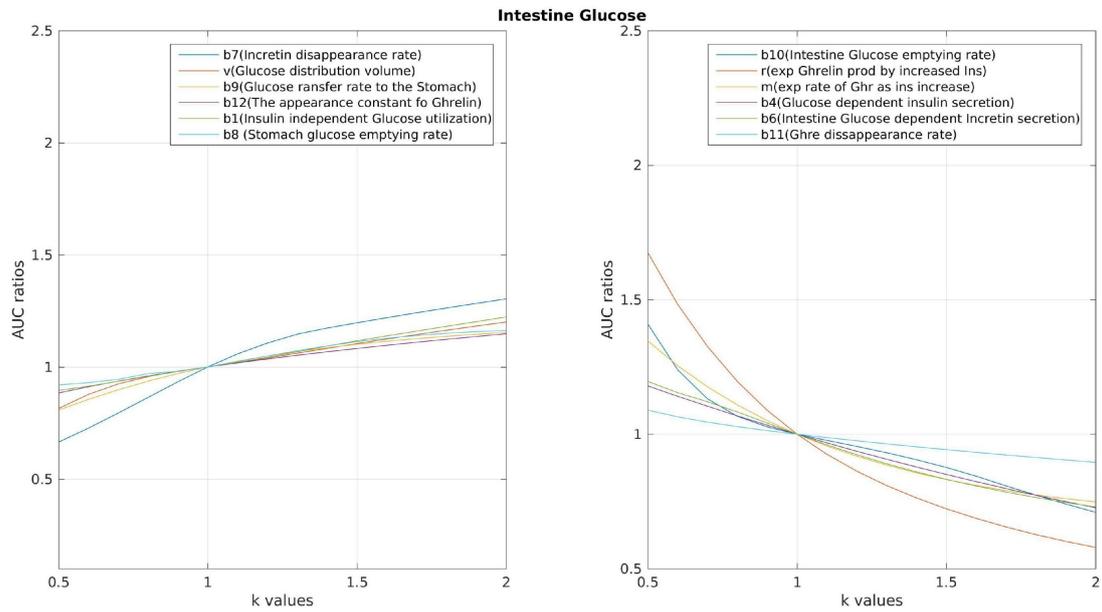


Fig. C.5 Intestine glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

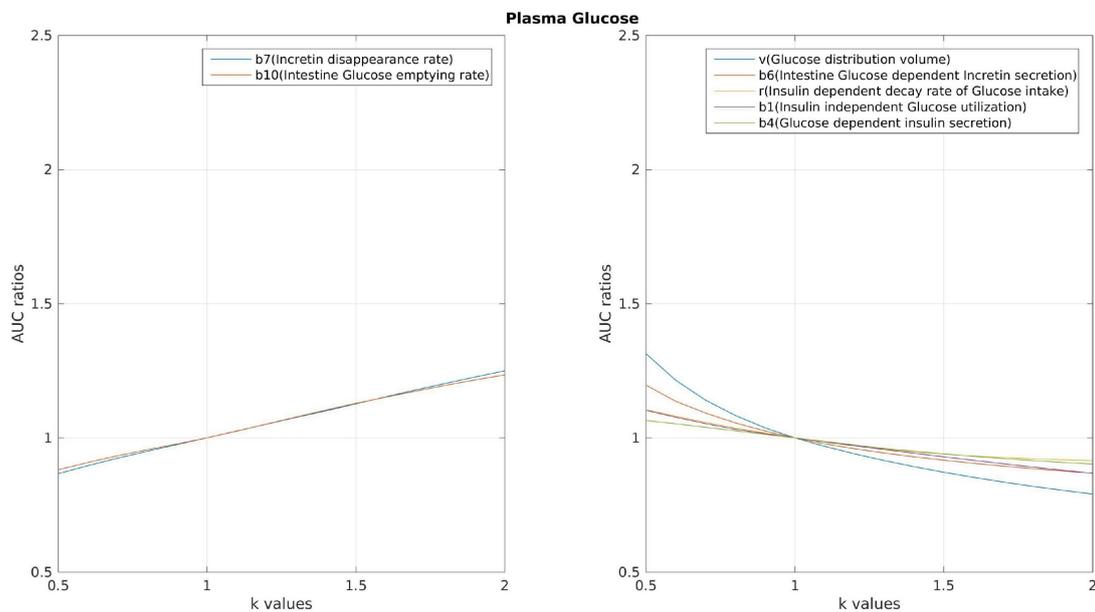


Fig. C.6 Plasma glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

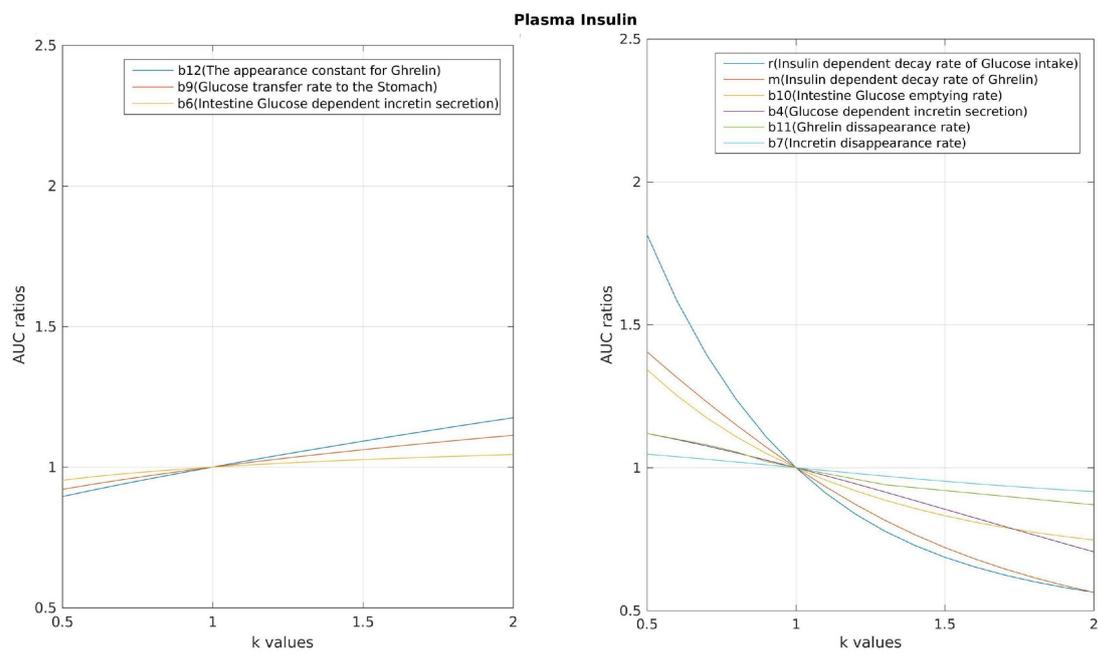


Fig. C.7 Plasma insulin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

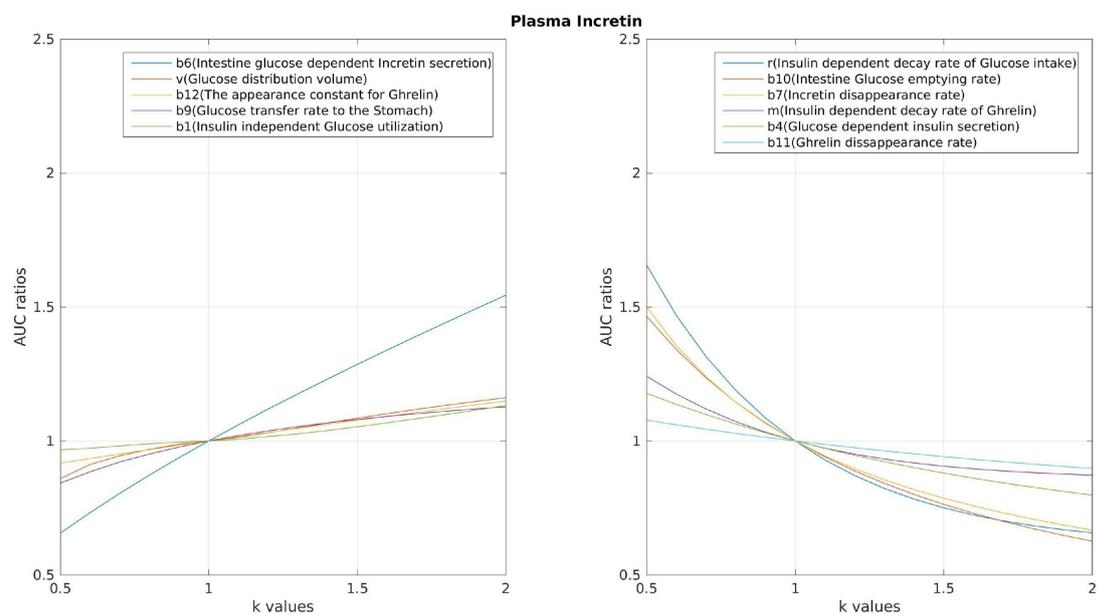


Fig. C.8 Plasma incretin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

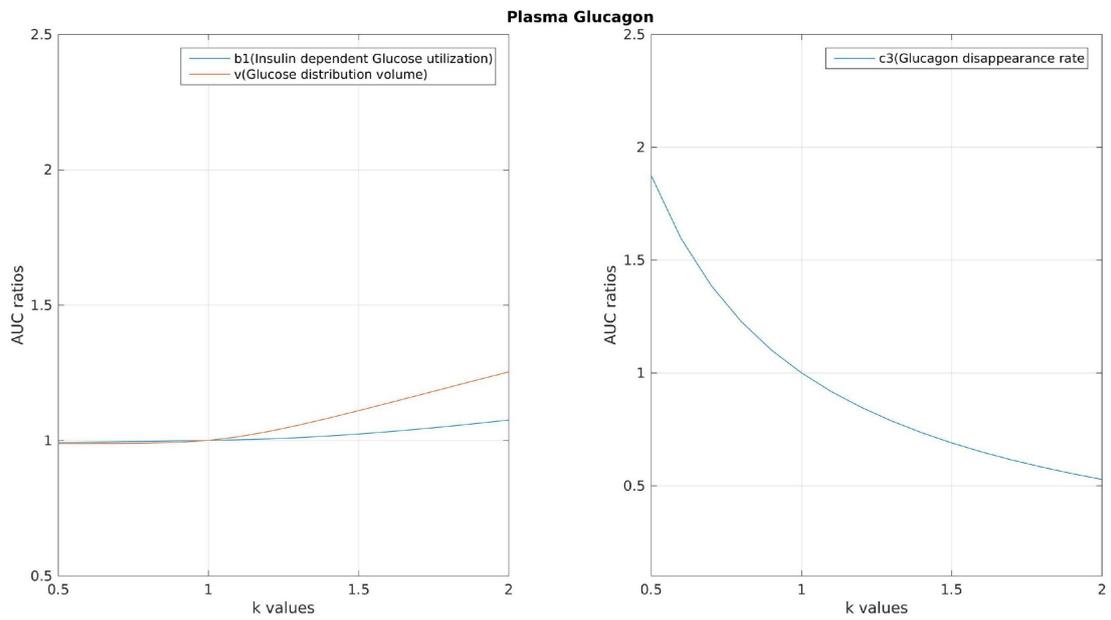


Fig. C.9 Plasma glucagon chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

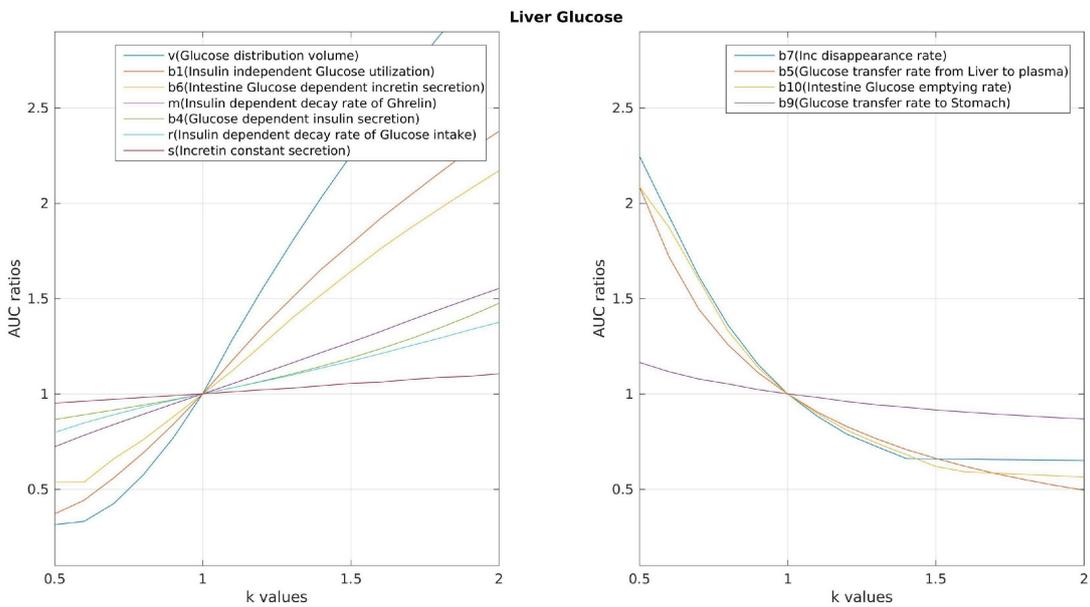


Fig. C.10 Liver glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

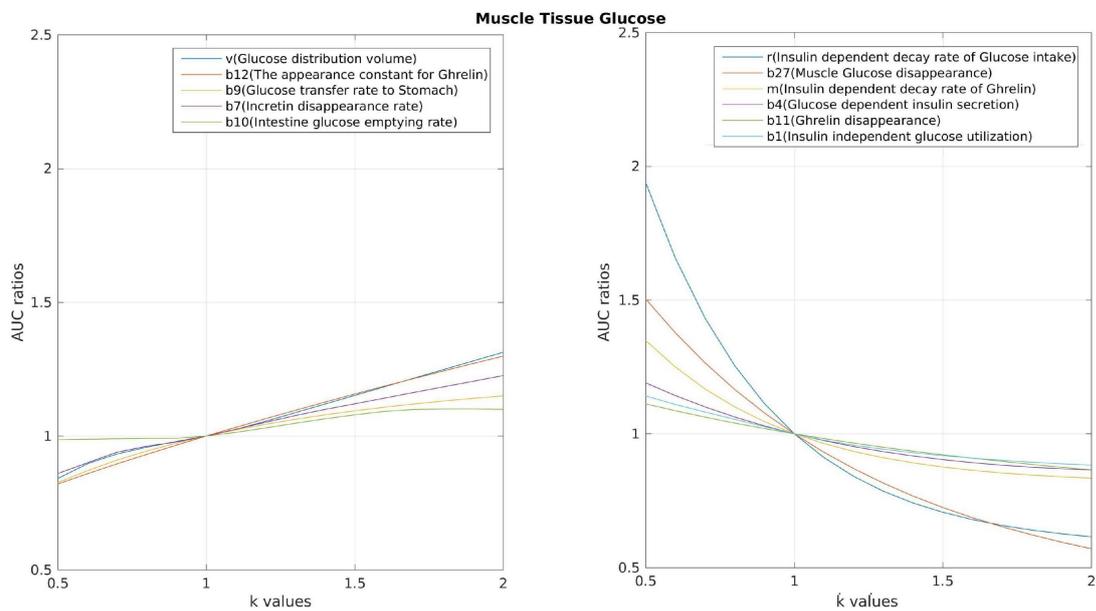


Fig. C.11 Muscle glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

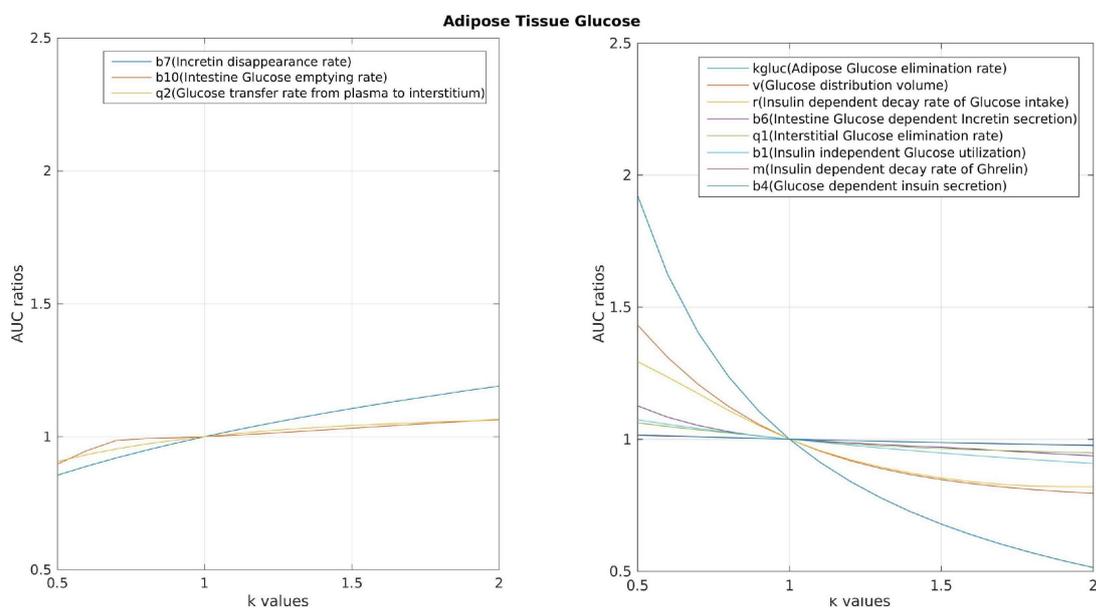


Fig. C.12 Adipose glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

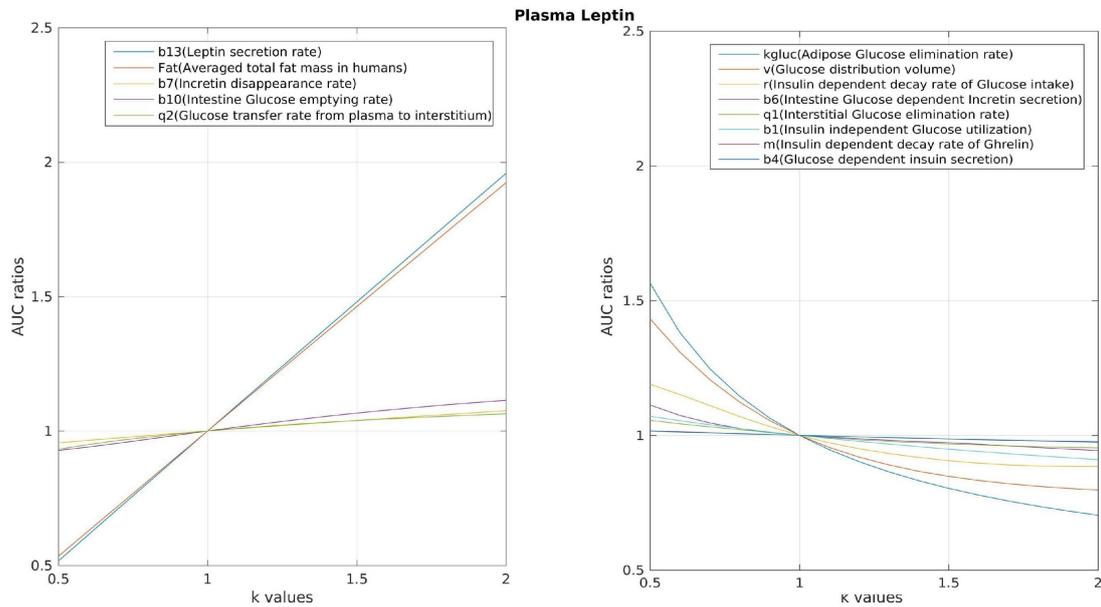


Fig. C.13 Plasma leptin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

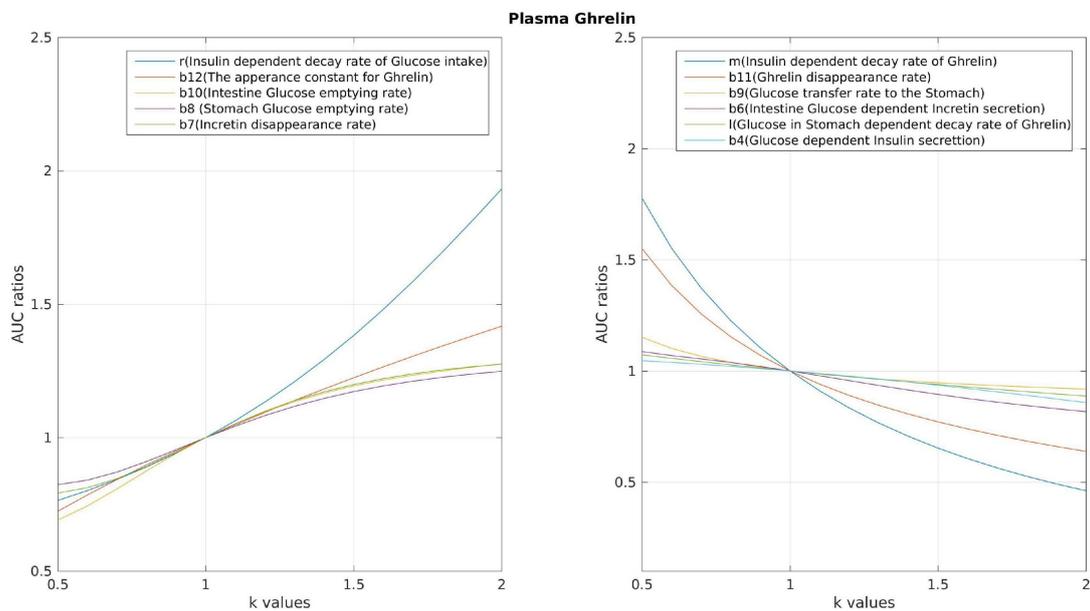


Fig. C.14 Plasma ghrelin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the NGR condition

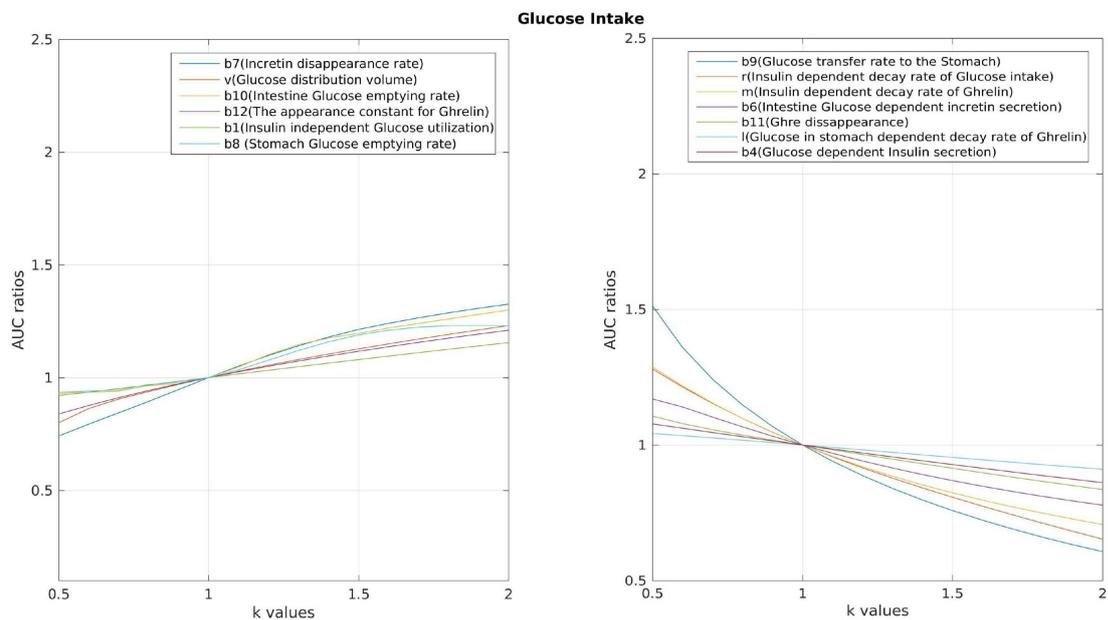


Fig. C.15 Oral glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

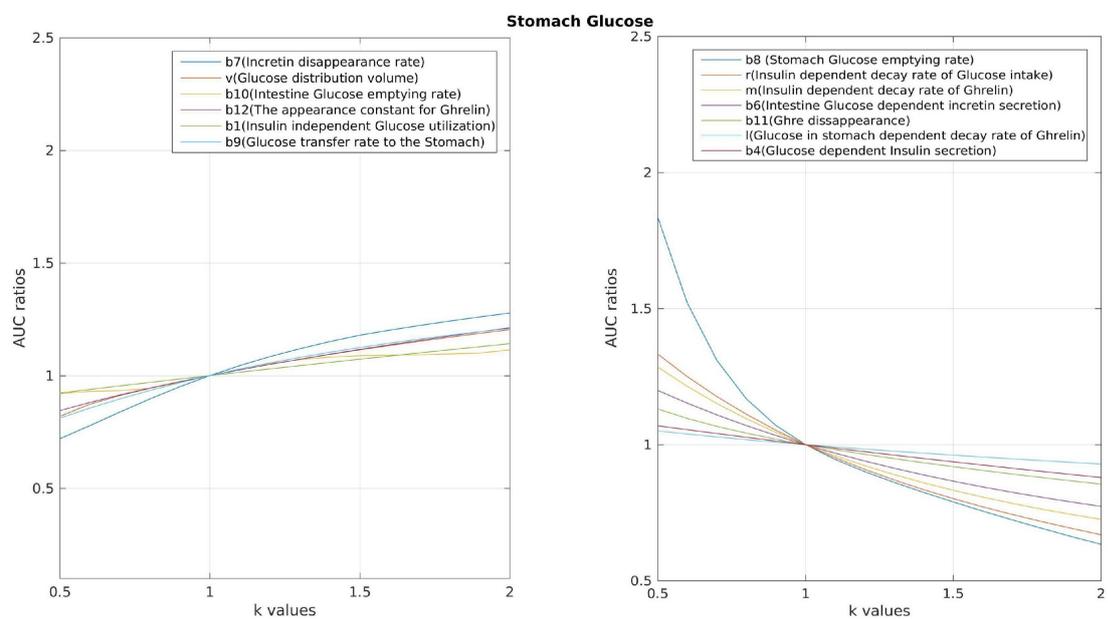


Fig. C.16 Stomach glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

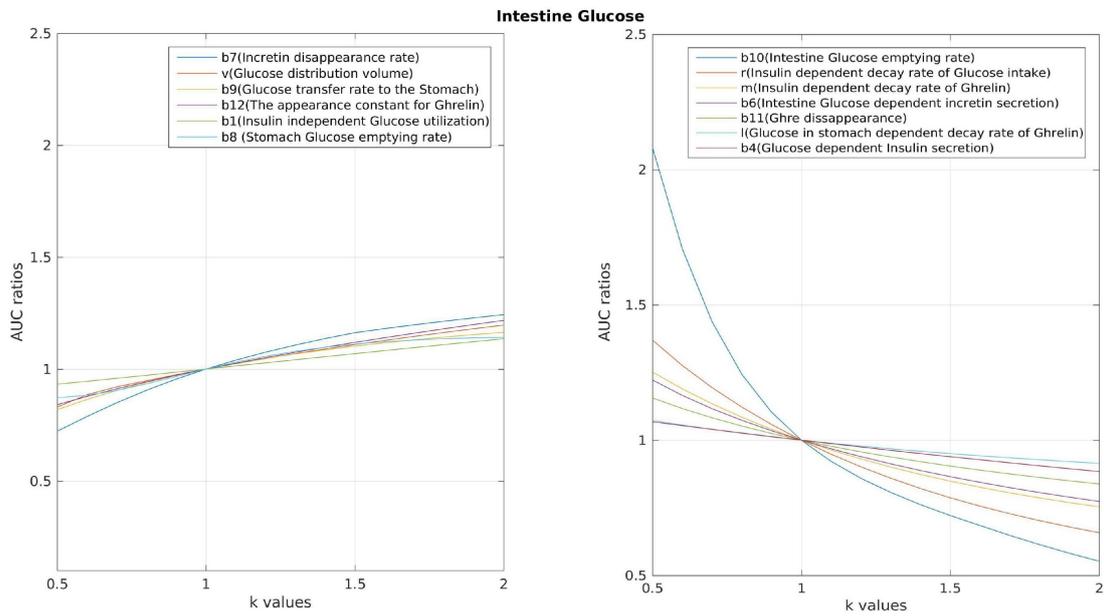


Fig. C.17 Intestine glucose intake chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

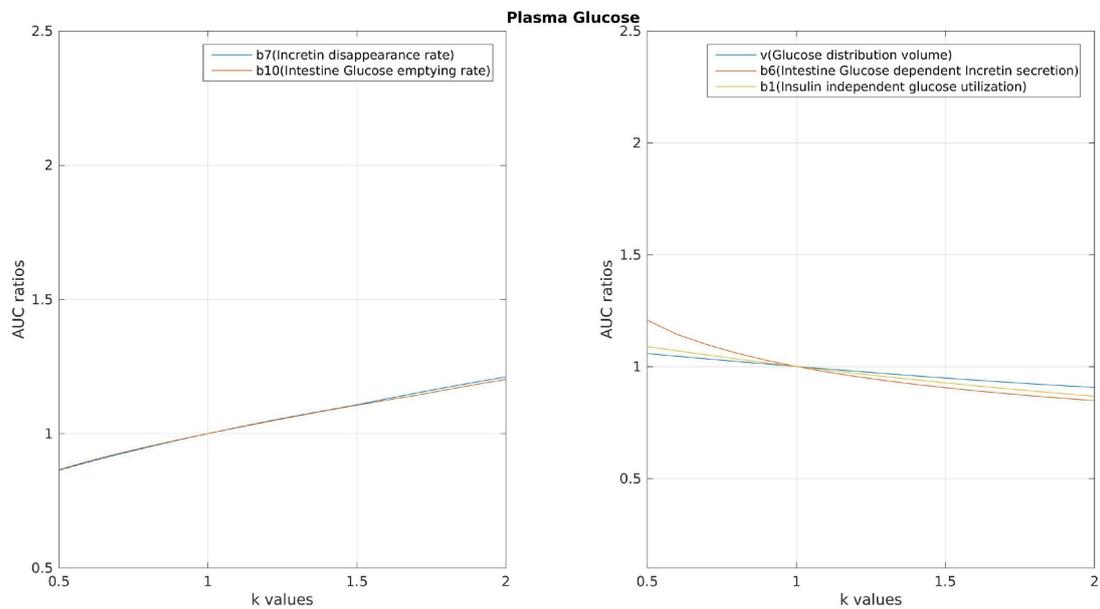


Fig. C.18 Plasma glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

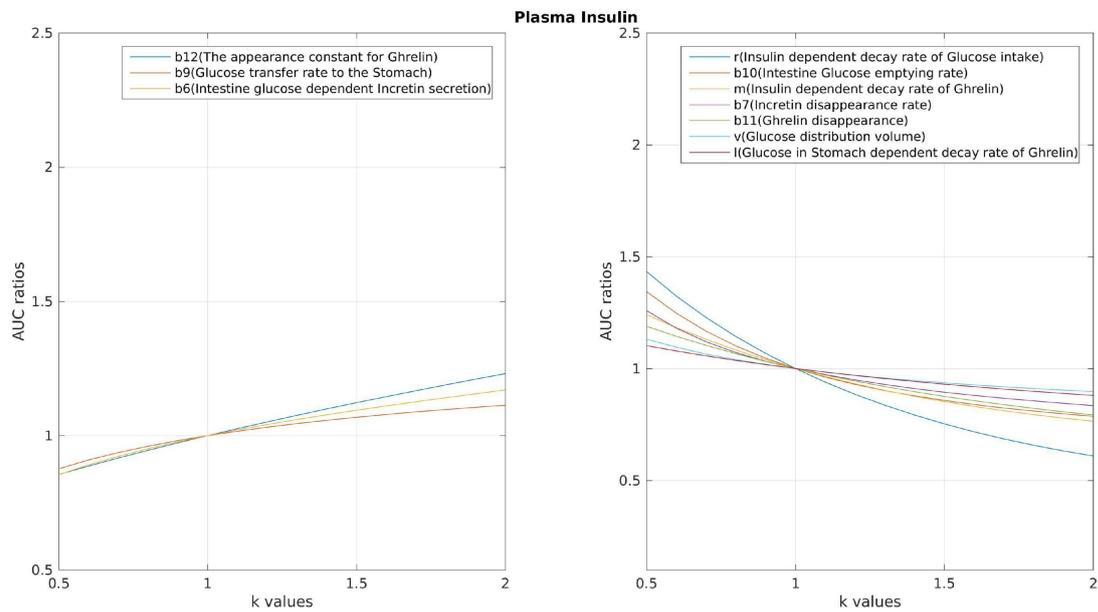


Fig. C.19 Plasma insulin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

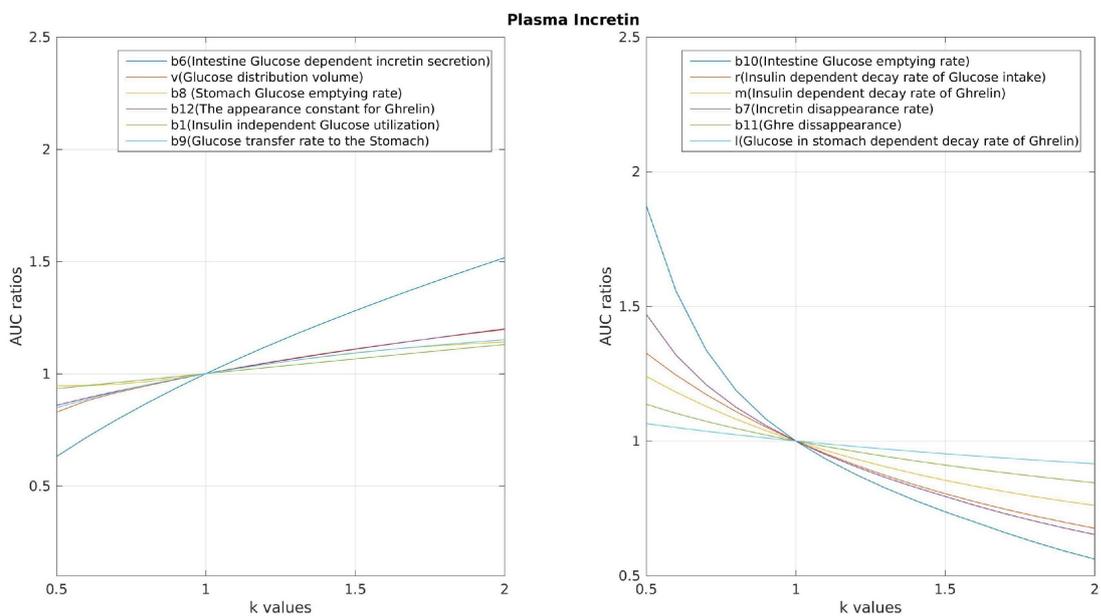


Fig. C.20 Plasma incretin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

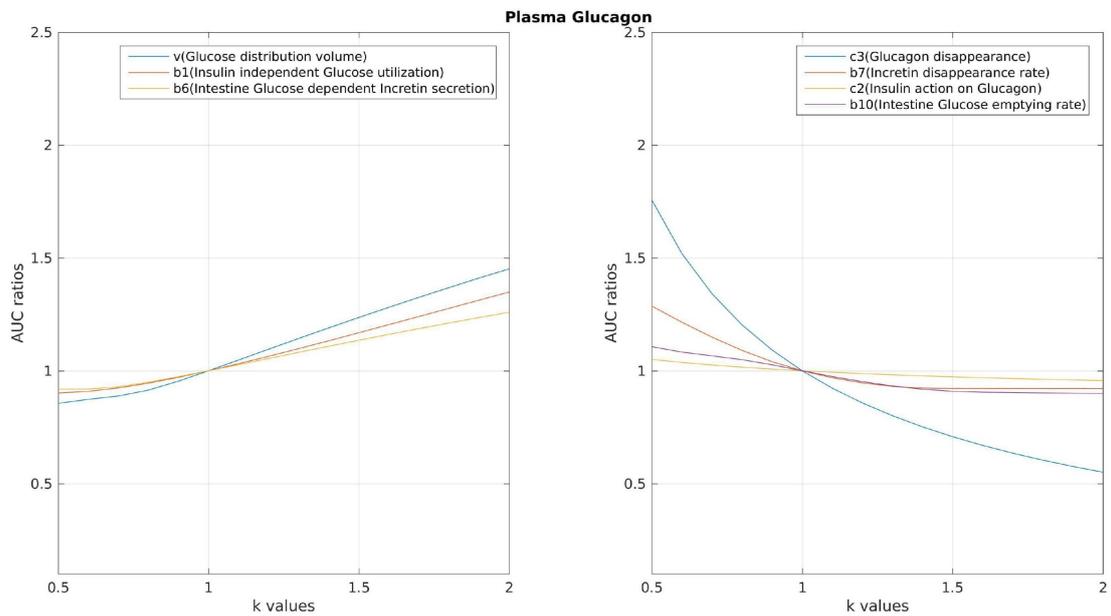


Fig. C.21 Plasma glucagon chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

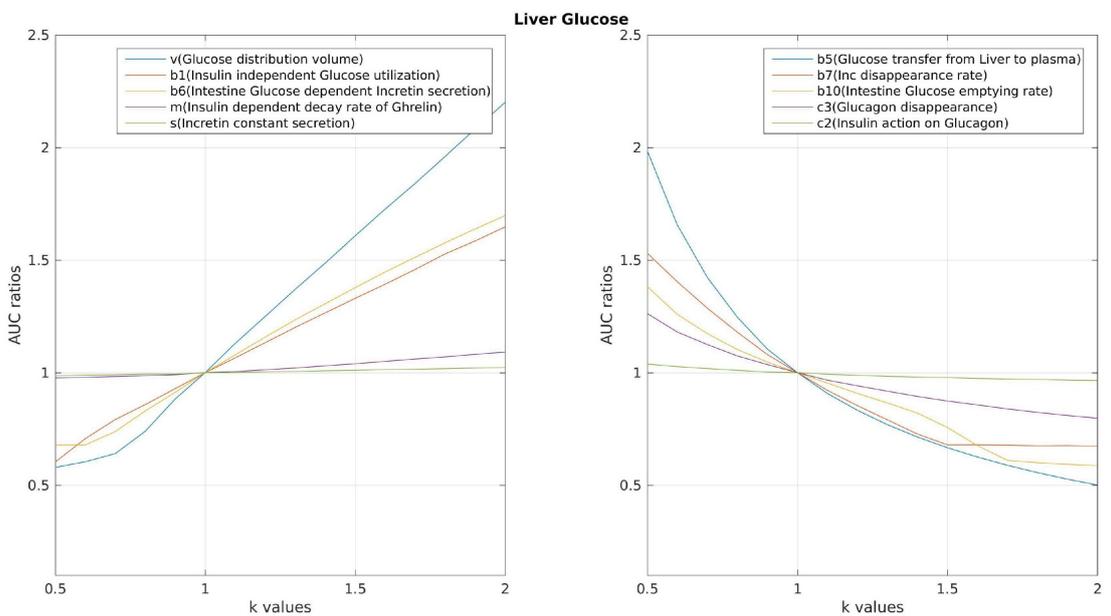


Fig. C.22 Liver glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

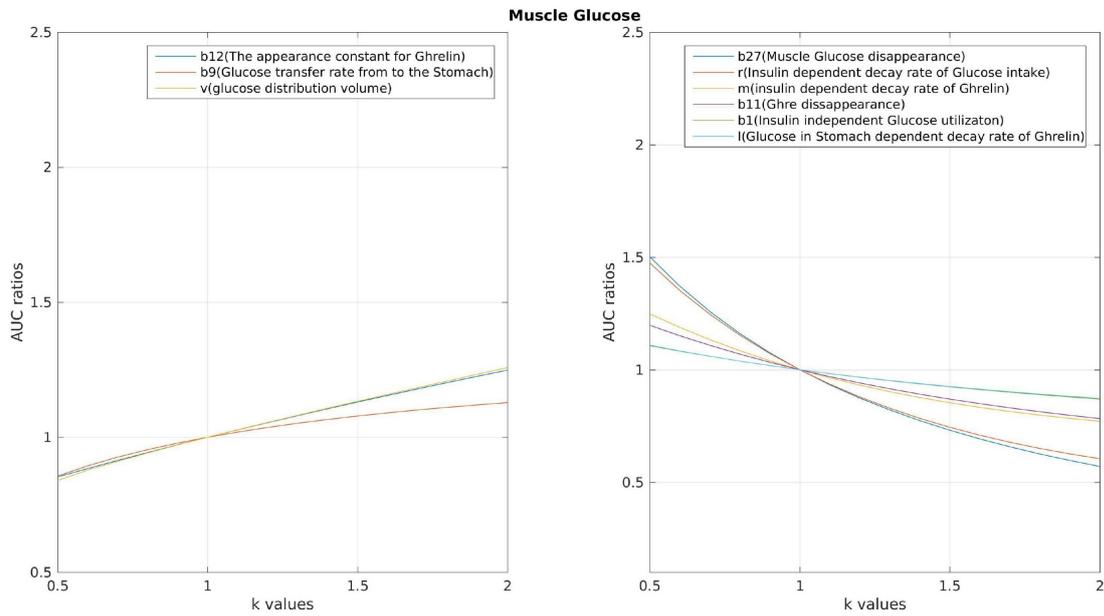


Fig. C.23 Muscle glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

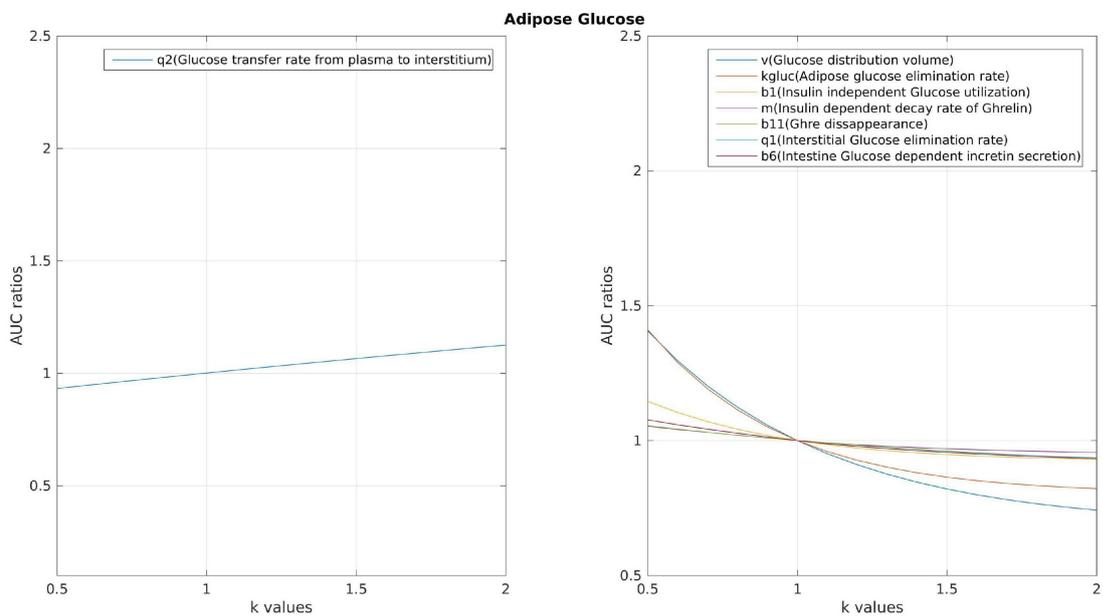


Fig. C.24 Adipose glucose chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

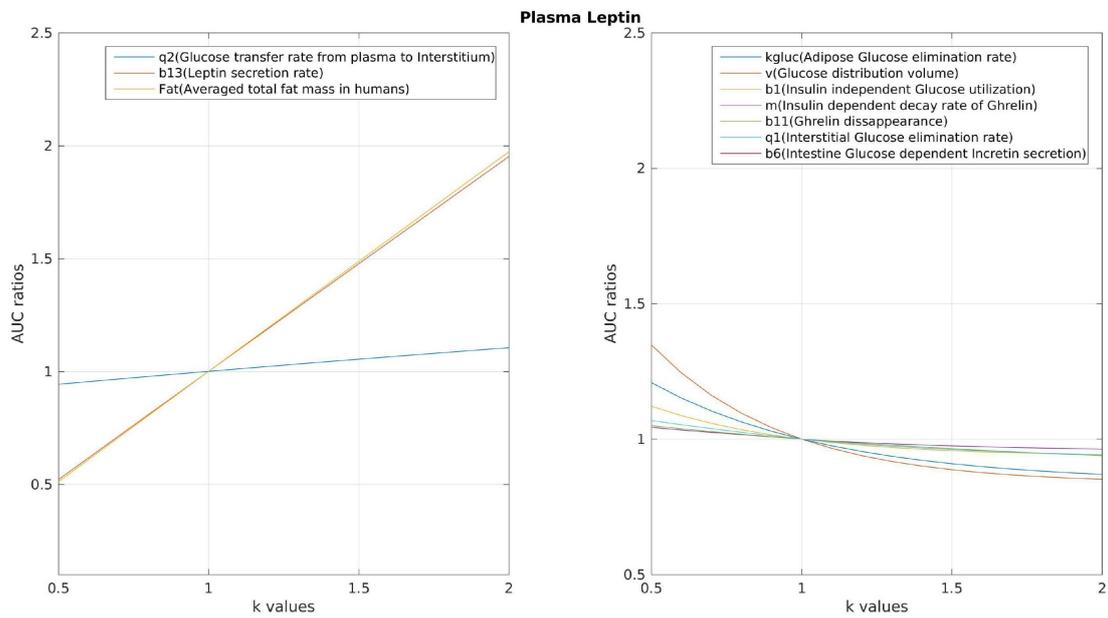


Fig. C.25 Plasma leptin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

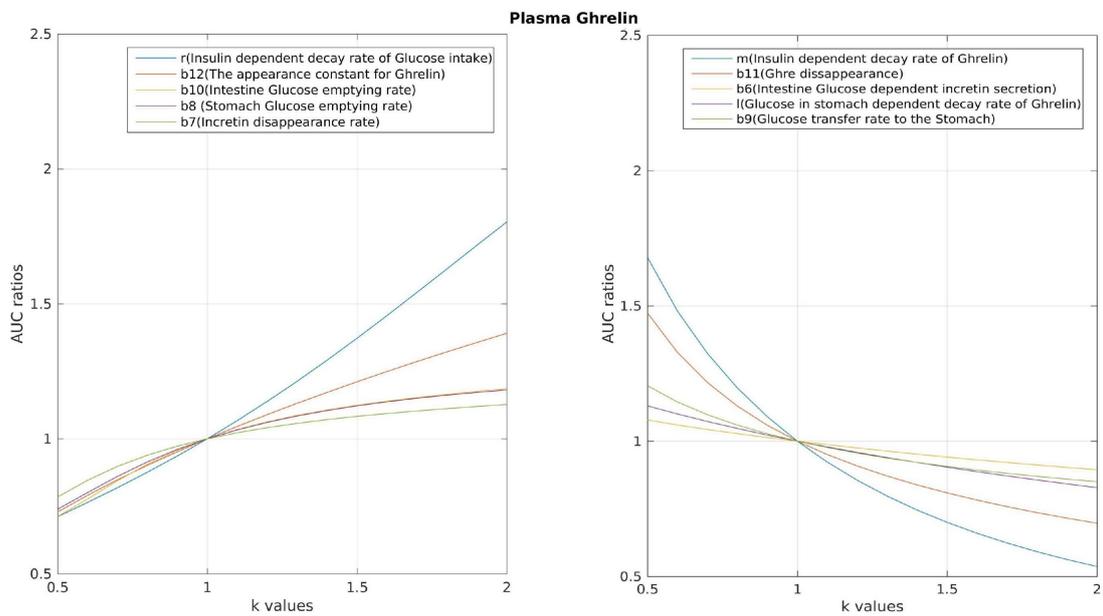


Fig. C.26 Plasma ghrelin chart providing the sensitivity analysis of estimated parameters for each variable of the whole body model in the T2DM condition

