

# UNRAVELLING THE INSULIN SIGNALLING PATHWAY USING MECHANISTIC MODELLING

MARTIN KIN LOK WONG

*Development and application of modelling tools to develop new pathway  
models*

A thesis submitted in fulfilment of the requirements for the degree of  
Doctor of Philosophy  
Faculty of Science  
The University of Sydney



March 2016



# ORIGINALITY STATEMENT

---

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked with, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

*Sydney, Australia, March 2016*

---

Martin Kin Lok Wong,  
May 30, 2017



# ABSTRACT

---

Type two diabetes affects 5% of the world's population and is increasing in prevalence. A key precursor to this disease is insulin resistance, which is characterised by a loss of responsiveness to insulin in liver, muscle and adipose tissue. This thesis focuses on understanding insulin signalling using the 3T3-L1 adipocyte cell model.

Computational modelling was used to generate quantitative predictions in the signalling pathways of the adipocyte, many of which are mediated by enzymatic reactions. This study began by comparing existing enzyme kinetic models and evaluating their applicability to insulin signalling in particular. From this understanding, we developed an improved enzyme kinetic model, the differential quasi-steady state model (dQSSA), that avoids the reactant stationary assumption used in the Michaelis Menten model. The dQSSA was found to more accurately model the behaviours of enzymes in large *in silico* systems, and in various coenzyme inhibited and non-inhibited reactions *in vitro*.

To apply the dQSSA, the SigMat software package was developed in the MATLAB environment to construct mathematical models from qualitative descriptions of networks. After the robustness of the package was verified, it was used to construct a basic model of the insulin signalling pathway. This model was trained against experimental temporal data at 1 nM and 100 nM doses of insulin. It revealed that the simple description of Akt activation, which displays an overshoot behaviour, was insufficient to describe the kinetics of substrate phosphorylation, which does not display the overshoot behaviour. The model was expanded to include Akt translocation and the individual phosphorylation at the 308 and 473 residues. This model resolved the discrepancy and predicts that Akt substrates are only accessible to Akt localised in the cytosol and that PIP3 sequestration of cytosolic Akt acts as a negative feedback.



# PUBLICATIONS AND PRESENTATIONS

---

Chapters 3 and 4 are published in:

- Wong MK, Krycer JR, Burchfield JG, James DE, Kuncic Z. A generalised enzyme kinetic model for predicting the behaviour of complex biochemical systems. *FEBS Open Bio* 2015; 9(5):226-239 doi: 10.1016/j.fob.2015.03.002.

Parts of chapters 3-6 have been presented at the following scientific meetings:

- Physics of Living Matter Symposium 2012, University College London, London, United Kingdom (Poster).
- MedPhys Symposium 2012, University of Sydney, Sydney (Oral).
- Australian Institute of Physics Congress 2012, University of New South Wales, Sydney (Poster).
- Systems Biology of Human Diseases 2013, German Cancer Research Centre, Heidelberg, Germany (Poster).
- St Vincent's Mater Health Sydney Research Symposium 2013, Garvan Institute of Medical Research, Sydney (Poster).
- Computational Models of Life Science Symposium 2013, CSIRO, North Ryde (Poster).
- Sydney Bioinformatics Symposium 2013, Garvan Institute of Medical Research, Sydney (Poster and Lightning Talk).
- International Biophysics Congress 2014, Brisbane Convention Centre, Brisbane (Poster).

- International Conference on Systems Biology 2014, Melbourne Convention Centre, Melbourne (Poster).
- The Hunter Systems Biology Meeting 2015, Hunter Valley, Hunter Valley (Poster).
- Ninth qBio Conference 2015, Virginia Tech, Blacksburg, USA (Poster).



## ACKNOWLEDGEMENTS

---

A PhD is certainly no simple thing to obtain. Originally I thought it would just be a long thesis. Instead it's turn into a long roller-coaster ride through many different flavours of sciences. I didn't realise the true scale of the challenge I would be facing until some-time in second year. So often I felt like I was going out of my depth, but I've managed to get through it and have a functional understanding of cellular biology and in particular insulin signalling. It really wouldn't have been possible without the help of many people.

Firstly my friends at the James Lab and Medical Physics. You're all a bunch of crazy but fun people. It was great to break up days of heavy science with some mindless banter, many initiated by my co-supervisor, David James', unique brand of humour. I want to thank him for giving me the opportunity to take up this project, which wouldn't have existed without his out of the box and adventurous approach to science. It's been an eye opening experience to experience his mind work in this space (even if I only become knowledgeable enough towards the end to be able to really appreciate it). I also need to thank my fellow PhD students in the lab: Dougall, Sheyda, Beverley, Annabel, Anastasios, Westa, Vinita and Tom. We've shared the many joys and pains that comes with the unpredictability of science. Good luck and there is a bright and blazing light at the end of the tunnel when all of your hard work comes together.

I need to acknowledge Dougall Norris specifically. Towards the end of my project, he became my closest collaborator. His confidence and competence in the lab has really reached great heights. Without his support, I would never have been able to achieve the experimental validation that was needed in my project. It has been a real pleasure collaborating with him. I hope my contribution to your project will be as significant as the contribution you have made to mine.

Next I would like to acknowledge my post-doc mentors. For some reason my post-doc mentors are all named James, which

makes it really easy to remember. James Burchfield was my cool and focused supporter, and man of few words. He helped me gain confidence in my project many times and often supported me in meetings when I was feeling unsure about my work. James Krycer was my supporter on the ground. He was essentially my guide when I decided to venture into the great unknown of lab work. He's taken so much of his time out to teach me the experimental techniques, explain the science behind them, and troubleshoot when things invariably went wrong. Just to highlight the degree of his dedication to my cause, he continues to offer help and advice even though (paraphrasing him) "I seem to have that X-factor. I run into every problem that can possibly happen". Unfortunately I haven't become that prodigy scientist that's a master of lab and computer work, but it's given me a lot of perspective about the data I'm dealing with and a new found respect for the people running around the lab all the time.

Thirdly, I'd like to acknowledge and thank my supervisor Zdenka Kuncic for her constant support throughout the four years of my PhD. She has given me a very fulfilling time by supporting my many trips overseas and interstate, as well as comforting me during the lows of my PhD. The connections and resources she has provided me have been instrumental in the completion of my thesis and I'm extremely grateful for her undying support over the years.

Finally, my friends, fellow PhDs (Jonathan, Ben and Annie) and non-PhDs (Alisa, Lucy, Leo-nnie, Vivian, Mark, James-Lisa, Anna-Karl, Tulara-John, Mia, Jess) and all of my other friends who I haven't been able to name. I've often used my PhD as an excuse to not meet up with you, but you never gave up on our friendship. The times you have been able to drag me away from my work I've been immensely grateful for. I needed the time to get away to recuperate and come back to my work with fresh eyes and new ideas. This would not have been possible without your support.

*Thank you to all of you.*

# CONTENTS

---

1	INTRODUCTION	1	
2	LITERATURE REVIEW	3	
2.1	The Disease Context	3	
2.1.1	Diabetes: A non-infectious epidemic	3	
2.1.2	Insulin Action and Insulin Resistance	4	
2.2	The Insulin Signalling Pathway	6	
2.2.1	Identifying the Pathway to Solve the Disease	6	
2.2.2	Mechanisms of Signalling Pathways	8	
2.2.3	Overview of the core pathway	10	
2.3	Mathematical Modelling of Pathways	16	
2.3.1	Signalling Pathways as Biological Circuits	16	
2.3.2	Overview of Modelling Approaches	18	
2.3.3	Kinetic Models and Mass Action	22	
2.3.4	Building and Verifying Models of Pathways	26	
2.4	Models of the Insulin Signalling Pathway	28	
2.4.1	Models of Insulin Signalling are highly varied	28	
2.4.2	Logical and Discrete Models of Insulin Signalling	30	
2.4.3	Kinetic Models of Insulin Signalling	33	
2.5	Motivation and Aims for this Thesis	40	
3	DEVELOPMENT OF A NEW ENZYME KINETIC MODEL	43	
3.1	Summary of notations in this chapter	44	
3.2	Introduction	44	
3.3	Simplifying Enzyme Kinetics	47	
3.3.1	Reversible enzyme kinetic model in a cyclic reaction system	48	
3.3.2	Uncoupled Reversible enzyme kinetic models in cyclic reactions	50	
3.3.3	Irreversible enzyme kinetic models in cyclic reactions	51	
3.3.4	The Michaelis-Menten Model in cyclic reactions	51	

3.4	Development of the Differential Quasi-Steady State Model	53
3.4.1	The Quasi-Steady State	55
3.4.2	Decoupling rate equation by vectorisation	60
3.4.3	Calculation of the Initial Condition	63
3.5	Ordered-Sequential Bi-Bi Reactions	64
3.6	Discussion	67
4	VALIDATION OF NEW ENZYME KINETIC MODEL	71
4.1	Introduction	72
4.2	Method	72
4.2.1	In Silico Validation in a hypothetical complex network	72
4.2.2	In Vitro Validation Using a Biological Enzyme	77
4.2.3	In Silico Validation of Two Substrate dQSSA	83
4.3	Result	84
4.3.1	In Silico Validation of dQSSA	84
4.3.2	In Vitro Validation of dQSSA	89
4.3.3	In Silico Validation of Two Substrate Model	93
4.4	Discussion	95
5	DEVELOPMENT OF MODELLING TOOLS	99
5.1	Introduction	100
5.2	Implementing Compartmentalisation	101
5.2.1	Discretised model of compartmentalisation	101
5.2.2	Elementary Reactions	103
5.2.3	Compartmentalising the dQSSA	104
5.3	Creation of modelling algorithm	105
5.3.1	Layout of SigMat Model Files	106
5.3.2	Dynamic Inference of Rate Equation and its Matrix Composition	112
5.3.3	Validation of the SigMat Algorithm	118
5.4	Network Motifs in Signalling	121
5.4.1	Merging of Signalling Inputs	121
5.4.2	Sustained Response from Transient Stimulation	125
5.5	Discussion	130
6	INFERRING MECHANISM OF AKT ACTIVATION WITH MODELLING	135
6.1	Introduction	137
6.2	Methods	138

6.2.1	Mathematical Modelling	138
6.2.2	Immunoprecipitation	140
6.2.3	In-vitro Kinase Assay of Akt activity	141
6.3	Modelling of the Insulin Signalling Pathway	141
6.3.1	Modelling the core pathway	141
6.3.2	Modelling of the Akt Signalling Pathway	149
6.4	Biological Validation	159
6.4.1	Verification of separate pools of phosphorylated Akt	160
6.4.2	Determination of AKT activity using in Vitro Kinase Assays	164
6.5	Discussion	171
7	CONCLUSIONS AND FUTURE WORK	177
7.1	Conclusions	177
7.2	Future Work	179
A	ADDITIONAL METHODS	183
A.1	Cell culturing	183
A.2	Time Course Experiment	183
A.3	Western Blot Analysis	185
A.4	Total Internal Reflection Microscopy (TIRF)	185
B	ADDITIONAL DQSSA DERIVATIONS	187
B.1	Two enzyme one substrate model	187
C	DQSSA VALIDATION MODEL DEFINITIONS	193
C.1	In Silico Validation	193
c.1.1	dQSSA Model	193
c.1.2	Mass Action Model	195
c.1.3	Michaelis Menten Model	197
C.2	In Vitro Validation One Substrate Model	198
c.2.1	dQSSA Model	198
c.2.2	Michaelis Menten Model	199
C.3	In Vitro Validation Two Substrate Model	200
c.3.1	dQSSA Model (One Substrate and Two Substrate)	200
c.3.2	Mass Action Model	202
D	BIOLOGICAL MODEL DEFINITIONS AND EXTRA DATA	203
D.1	Insulins Signalling Model	203
D.2	Akt Model	204
D.3	Additional Data	207
	BIBLIOGRAPHY	209

## LIST OF FIGURES

---

- Figure 1 Comparative prevalence of diabetes globally [2]. 3
- Figure 2 Effect of insulin under (clockwise from the top left): normal fed condition, normal fasting condition, insulin resistant fed condition and insulin resistant fasting condition [11] 5
- Figure 3 Identifying the makeup of signalling pathways is like solving a jigsaw puzzle. Current pathways are often incomplete with missing connections and nodes, and orphaned pieces where their location in the system is unclear. 7
- Figure 4 Evolution of our understanding of signalling. (a) The traditional view is linear with a one to one relationship between input and output [18], (b) The contemporary "hour glass" view is complex with a many to many relationship between input and output [16] 9
- Figure 5 Illustration of the core components of the Insulin Signalling Pathway and how information is transmitted upon insulin stimulation. 11
- Figure 6 Illustration of the parallels between electrical and biochemical circuit boards. Both systems involve 1) A list of parts 2) which have well defined behaviours 3) are assembled in very specific ways 4) whose parts always have certain specifications. 17
- Figure 7 Different modelling methods. (A) Boolean modelling [86]. (B) Petri Net [98]. (C) Kinetic/chemical kinetics [99]. (D) Stochastic/-chemical master equation [100]. (E) Spatio-temporal [101]. 21

- Figure 8 Levels of abstraction of a kinetic model. (A) Reactions occur when molecules within a confined space meet. (B) This can be described as a chemical equation. (C) This leads to a rate equation for the concentration of species involved based on the raw law of the reaction type. 23
- Figure 9 Illustration of the mechanism of enzyme action. The enzyme and substrate first associate, and then they go through successive steps of catalysis before reaching their final thermodynamic form before dissociating into product and substrate. Each step in this reaction is in principle reversible. 24
- Figure 10 Illustration of the iterative process between model generation, validation and expansion. This also shows the link between the interface between mathematical insights and biological insight and where they sit in the iterative process. 27
- Figure 11 Simplified comparison of input data passed through a pure and fuzzy logic model. 31
- Figure 12 Illustration of the way multiple cases are handled in the fuzzy logic framework. This allows the outcome of multiple satisfied cases to be aggregated to form the final output. 32
- Figure 13 Evolution of Insulin Receptor and Insulin Receptor Substrate models in the literature. Starting with (A) Quon's endocytosis model [151]. (B) Wanant's insulin binding model [152]. (C) Brannmark's feedback endocytosis model. (D) Nyman's enhanced feedback endocytosis model. 34
- Figure 14 Large abstracted models of Insulin Signalling (A) Kubota model [74], (B) Dalle Pezze model [127] (C) Brannmark T2D model [124] 35

- Figure 15 Akt phosphorylation data that was used to fit the various models. (A) Dalle Pezze model (adapted from Pezze et al 2012 [127]) (B) Kubota model [74]. (C) Brannmark T2D model [124] 39
- Figure 16 Different models of single substrate enzyme kinetics. (A) An illustration of the cyclic enzyme kinetic reaction that will be the focus of this chapter. An enzyme reaction converts substrate to product which is reversed by a spontaneous reaction. (B) Chemical equation of the fully reversible reaction mechanism. (C) Chemical equation of the coupled irreversible reaction mechanism. (D) Chemical equation of the irreversible reaction mechanism. 47
- Figure 17 Line graphs showing the complex concentrations simulated by the mass action model (solid lines) and tQSSA model (dashed lines) at five different values of the association constant  $k^a$ . The parameters  $k^d$  and  $k^c$  and concentrations  $[S]_0$  and  $[E]_0$  are set to unity for illustrative purposes. The tQSSA models do not agree with the mass action model at the beginning of the simulation, but quickly align. The complex concentration of both models later decrease together, which demonstrates that both the mass action model and tQSSA predict a change in the complex concentration, which would appear to be a violation of Eqn. 31 56



- Figure 18 Network topology of the hypothetical network used to test the accuracy of the dQSSA model compared to the mass action model of enzyme kinetics. Numbers inside species refers to their index in the model equations. Re refers to reactions and their numbers as they are discussed in the text. The network is designed to operate in two modes. In the first mode only reactions with black lines are active. This is because enzymes required for those reactions (more specifically species with grey backgrounds) are unavailable then. In the second mode, all reactions become active. This second mode is triggered by the present of the species I. 74
- Figure 19 Illustration of the reaction network used to simulate the LDH reversible conversion of pyruvate to lactate. The numbers within the species indicates the species index produced by the simulation used in the study. Rate parameters are as defined at the beginning of chapter 3.1 except for  $K^i$  which are dissociation constants for their corresponding reactions. Numbers inside species indicate their index used in the mathematical model in the models shown in Appendix C.2 78
- Figure 20 Accuracy of the (A) dQSSA and (B) Michaelis Menten model compared with the mass action model with 1000 different generated parameter sets. (C) Fraction increase in the computation time of each of the runs when compared with the mass action model. The errors are calculated averaged between the outcomes from the "I off" and "I on" simulations. 85

- Figure 21      Representative time course of the reaction with the parameters used shown in table 3 where the Michaelis-Menten model does not agree with the mass action or dQSSA models in both the "I off" and "I on" regime. Between  $t = [0, 10)$  I is not present and the system operates in the "I off" regime. Between  $t = [10, 12]$  I is added which initiates some changes in the system (shown in subpanel d). When  $t > 12$  the system is in the "I on" phase. The solid line indicates the mass action prediction which is used as the gold standard. Crosses indicates the dQSSA model while pluses indicates the Michaelis Menten model.      86
- Figure 22      Representative time course of the reaction with the parameters used shown in table 4. Between  $t = [0, 100)$  I is not present and the system operates in the "I off" regime. Between  $t = [100, 120]$  I is added which initiates some changes in the system (shown in subpanel d). When  $t > 120$  the system is in the "I on" phase. The solid line indicates the mass action prediction which is used as the gold standard. Crosses indicates the dQSSA model while pluses indicates the Michaelis Menten model.      87
- Figure 23      Inference of the parameter when Gaussian noise with a signal to noise ratio of 5 is added. The correct parameter (dotted vertical line) is contained within the blue curve (the likelihood distribution of the parameter)      90

- Figure 24 Illustration of the reaction network used to simulate the LDH reversible conversion of pyruvate to lactate. (a) and (b) shows the Lineweaver-Burke plots of the inverse initial reaction rate against the inverse concentrations of pyruvate and lactate respectively. Circles represent the experimental result while solid black and dashed grey lines represent the dQSSA and Michaelis Menten model predictions respectively. (c) - (e) shows a representative time course of the experiment with the two model predictions overlaid. In these panels, solid line represents the experimental data while crosses and pluses represent the dQSSA and Michaelis Menten model respectively. 91
- Figure 25 Validation result of the LDH model. The pyruvate to lactate reaction was tested with varying amount of  $\text{NAD}^+$ . The initial velocity of the reaction ( $V_i$ ) was then measured as a fraction of the initial velocity in the absence of  $\text{NAD}^+$  ( $V_{i0}$ ). This is shown as open circles. Error bars show the SEM. The dQSSA model (light grey line) and the Michaelis Menten Model (dark grey line) of this scenario was simulated and their results plotted. 92
- Figure 26 Distribution of errors in 1000 simulations with randomly generated parameters in the LDH model between the one substrate and two substrate dQSSA. 94

- Figure 27 Representative time course showing instability in the one substrate dQSSA model showing instability which does not occur in the two substrate model. In all panels, the mass action, one substrate dQSSA and two substrate dQSSA models are shown as a solid line, crosses and circles respectively. The upper panel shows the time course of total NADH in the three models, showing they produce similar predictions. The lower panel shows the concentrations of the various complexes of LDH, which reveals instability in the one substrate model. 95
- Figure 28 Figure showing the physical interpretation of the discretised model of transport between  $V_1$  to  $V_2$ .  $X_1$  within a unit volume (dash bordered cube) is transported to  $V_2$  at a rate of  $k^{ap}$  and disperses throughout  $V_2$ . All  $X_1$  within the red volume of  $V_1$  (with a depth of  $\hat{z}$ ) undergoes this transport process to form the total transported amount. The red volume is only a subset of the total volume of  $V_1$  because the clear parts of  $V_1$  are too deep within the compartment to be transported to  $V_2$ . This affects the accessibility of  $X_1$  available to be transported. 102
- Figure 29 Outline of the SigMat algorithm. The components within the dashed box are components within the algorithm. Components outside are passed as inputs to the algorithm and can be customised. 106
- Figure 30 Outline of the way models are constructed in SigMat. 109

- Figure 31 Model layout of (a) a single receptor that is the target of two ligands (b) two receptors with distinct ligands that target a common substrate. Panels (c) and (d) presents the reaction schemes used to model the network motif. In (c) the PI3K components (dashed line components) are not explicitly modelled because it is a direct substrate of pIR so is assumed to simply transfer the signal of pIR forward. All enzymatic reactions are modelled using the dQSSA raw law. All other reactions are modelled with first order mass action rate law. [122](#)
- Figure 32 Result of the pPI3K output at steady state for the (a) single receptor model, and (b) two receptors model, under different doses of the their respective ligands. [125](#)
- Figure 33 Illustration of the proposed signalling motif that will lead to a sustained signal upon transient stimulation. All enzymatic reactions are modelled using second order mass action rate laws. All other reactions are modelled with first order mass action rate law. Species with dashed outlines are inactives with species with solid outlines are active. [126](#)
- Figure 34 Time course and dose response of the transient stimulation responder network motif. [128](#)

- Figure 35 Dose response of the transient stimulation responder network motif with different rate at which the stimulus is removed (shown in this figure as  $t_{\frac{1}{2}}$ ). At small stimulus removal times ( $t_{\frac{1}{2}} \approx 1$  s), the dose response is a typical sigmoidal curve. At an intermediate stimulus removal half time ( $t_{\frac{1}{2}} \approx 20$  s), the output behaves like a log-linear converted (shown in the black line) with a broadened dose response. At large stimulus removal times ( $t_{\frac{1}{2}} \approx 10,000$  s), the dose response becomes binary at low dose. 129
- Figure 36 Setup of the model of the Insulin Signalling Pathway. 142
- Figure 37 Time course data of (a) phosphorylated Akt under 1 nM insulin stimulation (dashed lines) and 100 nM insulin stimulation (solid lines), and (b) and (c) some Akt substrates under 1 nM insulin stimulation (dashed lines) and 100 nM insulin stimulation (solid lines). In both figures, coloured lines correspond to the phosphorylation of individual sites as indicated in the figure. Black lines are the mean aggregate of all phosphorylation site shown within the subpanels. Error bars are standard deviations. [169] 144

- Figure 38 Result of calibrated model (solid line shows mean of best parameter sets, dashed lines shows the  $\pm$  one standard deviation spread from optimal parameters, defined by the parameter set with the largest likelihood) compared to the experimental data (open circle for mean, triangles for  $\pm$  on standard deviation from the mean). (a)-(c) Clusters 1-3 of Akt response to insulin. (d) shows the means of the Akt responses of the three clusters plotted together. (e)-(g) Clusters 1-3 of Akt substrate response to insulin. (h) shows the means of the Akt substrate responses of the three clusters plotted together. 145
- Figure 39 Plots of the (a) model fit to the pT308 Akt phosphorylation time profile and (b) resulting substrate phosphorylation fit. (c) model fit to the pS473 Akt phosphorylation time profile and (d) its resulting substrate phosphorylation fit. Solid line represents the mean of all accepted parameter sets while dotted lines show one standard deviation of the range of simulated kinetics across accepted parameter sets. 147
- Figure 40 The resulting substrate temporal profiles that can be achieved by varying the reaction rate of the substrate reactions based on the model fitted to the pT308 phosphorylation time profile. 148

- Figure 41      Setup of the model for Akt phosphorylation and activation. Nodes with dotted outlines lie within the plasma membrane and nodes with solid outlines lie in the cytosol. Reactions can only occur when reactants are in the same compartment (e.g. PDK1 can only phosphorylate Akt in the plasma membrane). PIP3 dimerizes with Akt when Akt is recruited to the plasma membrane. PIP3 is released back into the plasma membrane when Akt dissociates from the plasma membrane. The input for the model is PIP3, where its total concentration is controlled external to the model.      150
- Figure 42      Time course of Akt translocation at 1 nM Insulin. Open circles and their associated error bars are TIRF microscopy results ( $n=5$ ). The open diamonds are extrapolated from the post-overshoot decay using an exponential decay function. The grey line shows the average PM bound Akt response predicted by the model across the accepted parameter sets. [Experiment performed by Dougall Norris]      151
- Figure 43      Time course of Akt phosphorylation at 1 nM Insulin. Open circles and the associated error bars are the experimental mean and standard deviation of  $n = 5$  [38]. The black line shows the average of the Akt phosphorylation on the (a) 308 site and (b) 473 site across the accepted parameter sets.      152



- Figure 44 Time course of Akt phosphorylation at 1 nM Insulin at the (a) 308 site only at the plasma membrane, (b) 308 site only at the cytosol, (c) 473 site only at the plasma membrane, (d) 473 site only at the cytosol, (e) 308 and 473 site only at the plasma membrane, (f) 308 and 473 site only at the cytosol. The Deepness of the red at point shows the number of parameter sets that shows that level of response at that particular time point. This shows aggregated results from 1000 parameter sets. Distribution of the parameter sets in parameter space is shown in Fig. ?? in Appendix D.3 154
- Figure 45 Distribution of the predicted partition of phosphorylated Akt between its different phosphorylated forms as well as their locations at steady state. (a) Indicates the fraction in the cytosol (b) indicates the fractions in the plasma membrane. 156
- Figure 46 Temporal profiles of the different forms of Akt produced by the non-overshooting parameters set listed in Table. 11. Note there is no distribution shown because this is the result of a single parameter set. States without any suffixes are cytosolic. 156
- Figure 47 Substrate temporal profiles that can be achieved by varying  $k_{cat}$  of the  $\text{Substrate} \rightarrow \text{pSubstrate}$  and  $k$  of the  $\text{pSubstrate} \rightarrow \text{Substrate}$  reactions based on the model using parameters of the model that produces a non-overshooting cytosolic phosphorylated Akt (listed in Table. 11). This is compared to experimental data of Akt substrates aggregated in the same way as in the investigation of the insulin signalling model in section 6.3. 158
- Figure 48 Immunoprecipitation result of Akt after stimulation with 1 nM of Insulin. The analysis of the two phosphorylation sites were performed on the same membrane. 160

- Figure 49 Illustration of the various fractions of phosphorylated Akt that are in the various Western blotting samples shown in Fig. 48. The whole cell lysate and flow through samples are loaded to the same total protein amount, which means unaltered protein contents stay the same between those two samples. The IP sample only contains samples extracted by the IP process, but they become highly enriched. However, the concentration ratios between extracted species are the same as their ratio in their original whole cell lysate samples assuming the antibody binding efficiency is constant between species. 161
- Figure 50 Results of optimisation *In vitro* kinase assay experiments performed using purified active Akt. The y-axis normalised to the initial and maximal fluorescence with recombinant Akt. Dotted lines indicate  $\pm$  one standard deviation of the experimental mean with  $n = 3$ . (a) Optimising the amount of Akt required. (b) Optimising the buffer components. (c) Optimising the lysis buffer used and observing their effects (d) Validation of the inhibitors used and their effects on assay behaviour. 165

Figure 51 (a) *In vitro* kinase assay results with cell lysates under different conditions, with y-axis normalised to the initial and maximal fluorescence with recombinant Akt. Dotted lines indicate  $\pm$  one standard deviation from the experimental mean with  $n = 3$ . Cells are stimulated with insulin and other drugs. W is Wortmannin during stimulation, and C is Calyculin A during stimulation. G is GDC0068 and is added during the kinase assay only. N is the non-Akt inhibitor cocktail (see method for full list), also added during the kinase assay only. (b) Western blot analysis of the whole cell lysates used in the kinase assays to verify the change in phosphorylation state of Akt and TSC2 and AS160 as an indication of Akt activity. (c) Kinase assay performed with SOX-peptide already phosphorylated by recombinant Akt to demonstrate the phosphatase activity of whole cell lysate stimulated with 100 nM insulin. 168

Figure 52 Verification of the compatibility of immunoprecipitation conditions with *in vitro* kinase assay. The y-axis normalised to the initial and maximal fluorescence with recombinant Akt. Dotted lines indicate  $\pm$  one standard deviation from the mean with  $n = 3$ . (a) shows the change in activity of recombinant Akt with antibody addition. (b) shows the efficiency of immunoprecipitation of recombinant Akt where control indicates reaction run with just recombinant Akt, beads indicates addition of empty beads to the control reaction, Ab indicates addition of beads bound with recombinant Akt (without free recombinant Akt) and supernatant is from the immunoprecipitation process (again without further addition of recombinant Akt). 170

- Figure 53 Western blot analysis showing amount of Akt phosphorylated at the Thr308 site and total Akt from 15  $\mu$ g of cell lysate from 3T3-L1 adipocytes stimulated with 100 nM insulin for 10 minutes, and various amount of purified Akt. [171](#)
- Figure 54 Visualisation of the distribution of 1000 parameter sets achieved in the Akt-Akt substrate model using 3D heat maps. Red indicates high density of accepted parameters while blue shows a low density (deep blue is zero density). [207](#)

## LIST OF TABLES

---

Table 1	Summary of the different insulin signalling models found in literature. <a href="#">29</a>
Table 2	Comparison of the fitted initial concentration of IR and binding rate between Insulin and IR in various models of insulin signalling. <a href="#">38</a>
Table 3	Parameters used to generate the time course shown in Fig. 21. <a href="#">88</a>
Table 4	Parameters used to generate the time course shown in Fig. 22. <a href="#">89</a>
Table 5	Rate parameters of the LDH system that are achieved when fitted to the Michaelis-Menten and dQSSA models. The residuals of the best fits are also given. <a href="#">93</a>
Table 6	Rules relating to how parameters should be assigned and how they are interpreted by SigMat. <a href="#">111</a>
Table 7	Rules relating to reaction rates are inferred from the structure of the reaction that is parsed by SigMat. Dashed line separates groups of reactions that use rate laws of the same form. <a href="#">114</a>
Table 8	Index assignment for matrices within Eqn. <a href="#">138</a> for each reaction type <a href="#">115</a>
Table 9	Overview of the layout of the Matrix value pre-matrix. Cells marked as Target Rate, Reactant, Substrate and Enzyme indicate their indices. Crosses indicate unused columns. <a href="#">117</a>

Table 10	Spread of the log transformed, normalised error for each reaction type when simulated with sigMat compared to when modelled with a manually compiled mass action model. A value of -2 equates to an error of 1%, with lower values equating to lower errors. <a href="#">120</a>
Table 11	List of the parameters used to achieve the model with a cytosolic phosphorylated Akt temporal profile that has a suppressed overshoot. Non-conventional units are used for the concentration and volume. The model is non-dimensionalised with respect to the PIP3 concentration and volume of the cytoplasm, thus N indicates concentration as a fraction of PIP3 concentration, and V indicates volume as fraction of cytosol volume. <a href="#">157</a>

# ABBREVIATIONS

---

## GENERAL

**T1D** - Type 1 Diabetes  
**T2D** - Type 2 Diabetes  
**GLUT4** - Glucose transporter 4  
**GTP** - Guanine triphosphate  
**GDP** - Guanine diphosphate  
**PP2A** - Protein Phosphatase 2A  
**ADP** - Adenosine diphosphate  
**ATP** - Adenosine triphosphate  
**PTM** - Post translational modification  
**PI3K** - PI3-Kinase  
**IRS** - Insulin receptor substrate  
**IR** - Insulin receptor  
**SH2** - Src Homology 2  
**PI** - Phosphatidylinositol  
**PI(4,5)P2** - phosphatidylinositol (4,5) bisphosphate  
**PI(3,4,5)P3** - phosphatidylinositol (3,4,5) triphosphate  
**PH** - Pleckstrin Homology  
**mTORC1** - Mammalian target of rapamycin complex 1  
**mTORC2** - Mammalian target of rapamycin complex 2  
**PS** - Phosphatidylserine  
**CPN** - Coloured Petri Nets  
**CME** - Chemical Master Equation  
**ODE** - Ordinary Differential Equations  
**XML** - Extensible Markup Language  
**SBML** - Systems Biology Markup Language  
**PM** - Plasma membrane  
**LDH** - Lactate Dehydrogenase  
**NAD<sup>+</sup>** - Nicotinamide adenine dinucleotide  
**NADH** - Reduced nicotinamide adenine  
**NADP<sup>+</sup>** - Nicotinamide adenine dinucleotide phosphate

## MATHEMATICAL

**t** - Time $\hat{t}$  - Non-dimensional time**V** - Volume**S** - Substrate**E** - Enzyme**P** - Product**ES** - Enzyme-Substrate complex**EP** - Enzyme-Product complex**EO** - Enzyme-Coenzyme complex**EOS** - Enzyme-Coenzyme-Substrate complex $k_j^i$  - Reaction rate constant

- when  $i$  contains
  - $a$  - Association
  - $d$  - Dissociation
  - $c$  - Catalysis
- when  $j$  contains
  - $f$  - Forward reaction
  - $r$  - Reverse reaction
- when the above do not apply,  $k$  is a standard mass action reaction rate.

**K** - Equilibrium constants**tQSSA** - Total Quasi Steady State assumption**dQSSA** - Differential Quasi-Steady State assumption**MM** - Michaelis Menten**MV** - Matrix Value pre-matrix**PI** - Parameter Index pre-matrix**G** - dQSSA Prefactor Matrix**Y** - Bimolecular reaction matrix**Z** - Unimolecular reaction matrix $\sigma$  - Zeroth order reaction vector**H** - Hill function matrix**f** - Custom function



# INTRODUCTION

---

Diabetes, characterised by the dysregulation of glucose homeostasis, is a complex disease that is quickly becoming a leading cause of mortality in the modern world. One of the key risk factors of the Type 2 variant is insulin resistance. Significant research effort is shifting towards determining the intracellular causes of insulin resistance in key tissues that regulate glucose homeostasis: muscle, fat and liver.

The conventional method for interrogating the intracellular behaviour of cells is through a reductionist, node by node perturbation of the cell. In recent decades, the use of computational approaches to complement the reductionist experimental approach has gained prominence. This is the field of systems biology. One component of this field is the simulation of biological networks by mathematical modelling. This involves simulating the collection of biochemical reactions that occur within the cell to reproduce, predict and understand the emergent properties created by the network. This will enhance our understanding of the defects contributing to insulin, and their effects on the regulatory mechanisms of glucose homeostasis. However, to do this it is necessary for the model to accurately replicate the behaviour of each individual biochemical reaction.

Mathematical models need to be constructed in close association with experimental data derived from the cell model that is studied. As there are many cell models used to study insulin signalling and insulin resistance, there are similarly many mathematical models that have been customised for each cell line. However, there is currently no mathematical model designed specifically for the common cell line used for studying insulin signalling in fat cells, the 3T3-L1 adipocyte.

This has led to two key questions pursued in this thesis:

1. Are existing modelling methods sufficiently accurate, and if they are not, can they be improved?

2. What does a quantitative model of the common cellular model system for studying insulin signalling and insulin resistance, the 3T3-L1 adipocyte, look like?

These research questions are significant because together, they begin the process of quantitatively integrating the enormous volume of research data (and uncountable number of sacrificed adipocytes) that have been accrued using this cell model. From there, the mathematical model can then be used to quantitatively understand the system. These simulations can then be used to identify gaps and inconsistencies in the model, and refine hypotheses to explain gaps and resolve inconsistencies with novel mechanisms. Finally, the quantitative predictions can be used to experimentally validate the model with increased precision.

## LITERATURE REVIEW

---

### 2.1 THE DISEASE CONTEXT

#### 2.1.1 Diabetes: A non-infectious epidemic

Diabetes is a non-communicable disease which is characterised by a loss of glucose homeostatis which can lead to many serious and sometimes fatal complications [1]. Its severity is increased by its growing prevalence throughout the world as shown in Fig. 1 [2]. In 2008, there were an estimated 347 million sufferers of diabetes world wide, making up approximately 5% of the world population [3]. In 2012, an estimated 1.5 million deaths were caused by this disease [4]. Even within Australia, there are around 1.1 million sufferers of diabetes with 280 new cases appearing every day [5]. Diabetes also creates a significant financial burden on health care globally. In 2010, the global expenditure of diabetes was USD 376 billion, which represents 12 % of global health expenditure. In Australia, the average annual cost per person of

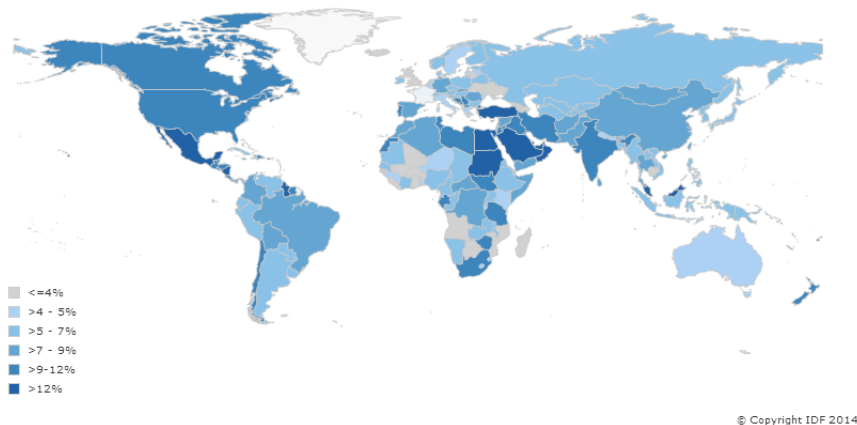


Figure 1: Comparative prevalence of diabetes globally [2].

diabetes treatment is AUD 4669, increasing to about AUD 17,000 when patients present with vascular complications [6]. Diabetes leads to an increased risk of complications such as cardiovascular disease, blindness, kidney failure, neuropathy, and reduced resistance against infections [7].

There are two broad ways this diseases can arise. Type I Diabetes (T1D), is genetic in origin and accounts for around 10% of cases. Type II Diabetes (T2D) on the other hand, has both genetic and environmental origins and accounts for around 90% of cases. T2D is characterised by insulin resistance in the patient and only partial loss of insulin production [7, 8]. While there is a strong genetic component to this form of diabetes, its development is strongly correlated with obesity and increased abdominal adiposity for those already predisposed [8, 9]. Given that the link between obesity and insulin resistance is already known, investigation of this link has become the focus of T2D research and the focus of this thesis.

### 2.1.2 Insulin Action and Insulin Resistance

Regulation of plasma glucose levels is essential for whole body homeostasis [10]. Insulin and glucagon are two components of the feedback system that maintains glucose homeostasis. Glucagon is secreted by the  $\alpha$ -cells of the pancreas and increases blood glucose levels during starvation. Conversely, insulin, which is secreted by the  $\beta$ -cells of the pancreas, primarily causes a system wide increase in glucose storage due to increased glucose uptake in muscle and fat cells via recruitment of the glucose transporter 4 (GLUT4) to the cell plasma membrane, and reduced glucose output from the liver [10].

Insulin, the focus of this thesis, acts primarily on three types of tissue in the body: liver, muscle and fat (Fig. 2) [11, 12]. In liver cells, or hepatocytes, insulin increases lipogenesis and promotes glycogen storage through inhibition of glycogen breakdown and increasing glycogen synthesis from glucose [10]. Glycogen synthesis is also promoted in muscle cells, or myocytes, in preparation for exercise. Fat storage is controlled by insulin in fat cells, or adipocytes, by decreasing lipolysis which reduces the release of free fatty acids [11]. In addition to these functions related to glucose homeostasis, insulin increases gene expression, protein

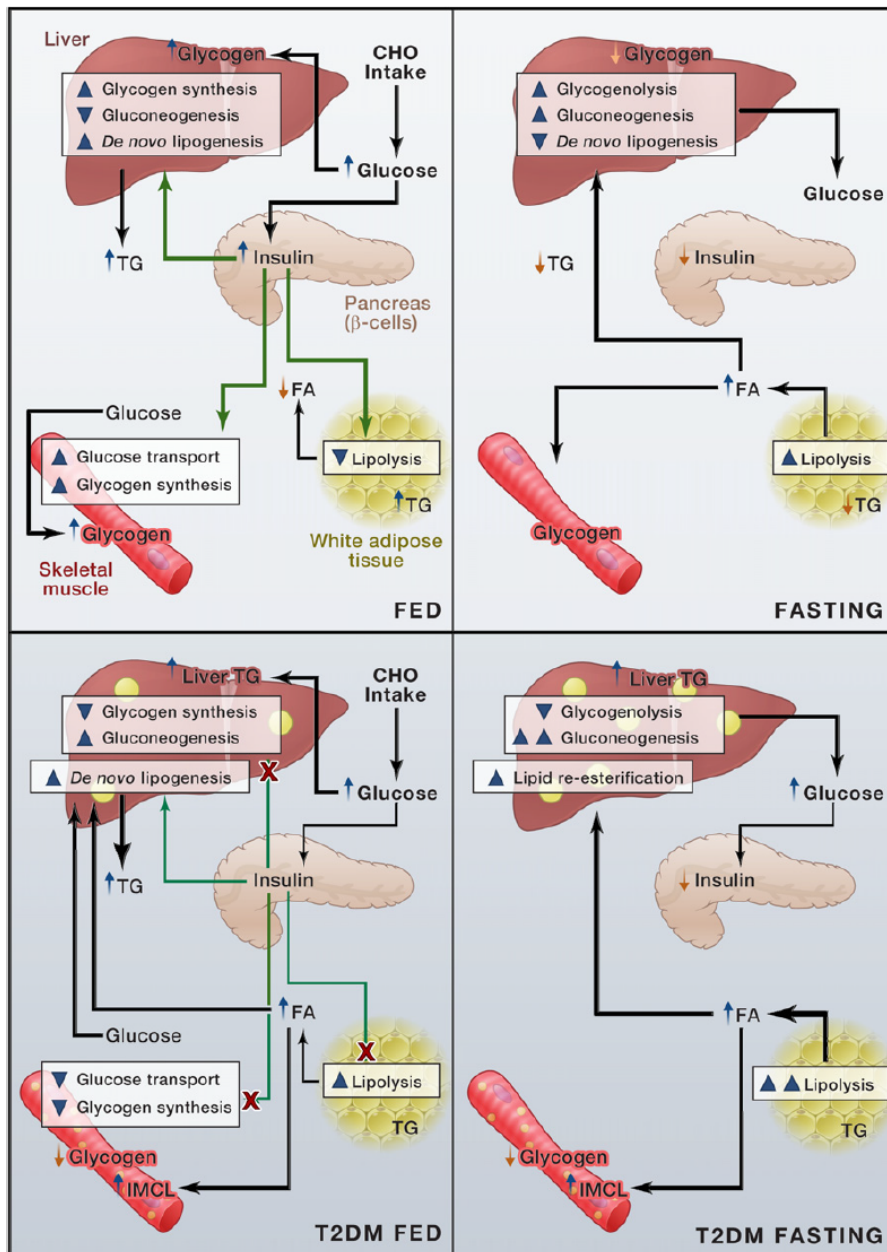


Figure 2: Effect of insulin under (clockwise from the top left): normal fed condition, normal fasting condition, insulin resistant fed condition and insulin resistant fasting condition [11]

synthesis and amino acid uptake, and decreases gluconeogenesis (the conversion of pyruvate to glucose), apoptosis (programmed cell death) and autophagy (degradation of cellular components)

[13]. These effects are triggered by the signalling cascade that is initiated when insulin binds to its cognate receptor.

Insulin resistance is characterised by an impairment in the ability of insulin sensitive tissue to respond to insulin. As there is a strong link between T2D and obesity, insulin resistance is also strongly correlated with obesity [9]. Beyond this, the biochemical defects responsible for insulin resistance are highly contentious. While defects in the insulin receptor itself is a possible cause of the disease, it is only correlated with some cases of insulin resistance [9]. Early insulin resistance research showed that other defects downstream of the insulin receptor such as the insulin receptor substrate and  $\text{PIP}_3$  production could also contribute to insulin resistance. However, the full complexity of the disease is only now appreciated as a growing range of possible causes are being identified [11]. This can be demonstrated by a Google Scholar search of "causes of insulin resistance" which returns 24,000 search results within just the last two years. It is clear that the causes of insulin resistance are both complex and variable. Thus a deeper understanding of the mechanisms involved with insulin signalling is necessary before treatment strategies can be designed to reverse it.

## 2.2 THE INSULIN SIGNALLING PATHWAY

### 2.2.1 Identifying the Pathway to Solve the Disease

As research into T2D and insulin resistance progresses, there is increasing evidence that they are caused by intracellular defects. As a result, there is increasing interest in understanding the mechanisms behind the insulin signalling pathways. Cell signalling pathways can be described by a list of biomolecular reactions which occur between the pathway components. So to understand the pathway, it is necessary to understand the composition of the biochemical components.

Identifying the composition of the network is much like solving a jigsaw puzzle. The biomolecular makeup, analogous to the collection of puzzle pieces, is a list of unique species that differ in physical (e.g. location) and chemical (e.g. intrinsic protein or

post-translational modified) properties. This can be audited using -omics type studies [14, 15].

The list of interactions, analogous to the joints between puzzle pieces, are interactions between different biomolecular components. These are governed by physico-chemistry, which causes a change in one or more biomolecular species. This list is more difficult to audit and is the subject of many biological assays, such as yeast two hybrid studies, far western blotting and co-immunoprecipitation.

In order to solve and understand the signalling pathway, we must reconstruct the biomolecular makeup in an interaction network where each interaction is verified and the general behaviour is self consistent. This is much like the big picture of a jigsaw puzzle that can only be formed when all puzzle pieces fit in place without any clashes in the connecting points or the continuity of the picture (illustrated in Fig. 3).

The following sections will review the general mechanisms commonly found in signalling pathways. The focus will then turn specifically to insulin signalling, where the broad structure of the

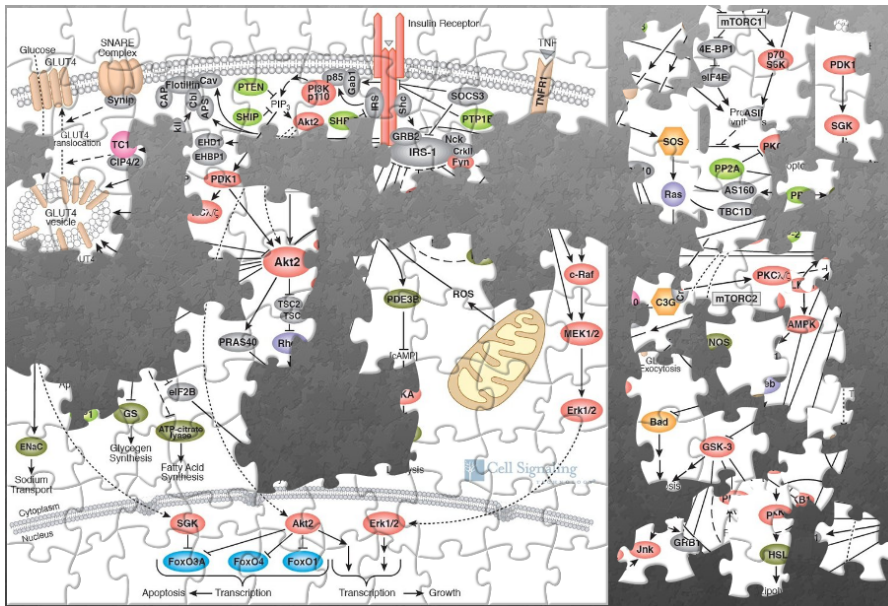


Figure 3: Identifying the makeup of signalling pathways is like solving a jigsaw puzzle. Current pathways are often incomplete with missing connections and nodes, and orphaned pieces where their location in the system is unclear.

pathway as it is currently understood is outlined. From there, the various feedbacks acting on the important nodes are discussed to demonstrate the immense complexity inherent in this system.

### 2.2.2 Mechanisms of Signalling Pathways

The traditional view of signalling pathways is a linear cascade, where the receptor activates a linear cascade, ending at one effector (Fig. 4A). However, the more contemporary view is a complex network of multiple inputs and multiple outputs which connect with each other in complex ways (Fig. 4B). The crucial feature of the contemporary view is that all inputs converge onto a selection of key regulatory nodes, which then disperse the signal through to the many effectors in the cell. This is known as the hourglass or bow-tie architecture of signalling [16, 17].

Information through a signalling cascades is often transmitted by enzymatic reactions [19]. Upon activation of the upper most receptor in the signalling pathway, it proceeds to activate its downstream molecular target, converting it from an "off" state to an "on" state. The activated molecule then proceeds to activate its target.

The most studied ways in which proteins become activated are phosphorylation and guanine triphosphate (GTP) binding. Phosphorylation can occur on many proteins that contain key amino acid residues that are phosphorylatable (serines, threonines and tyrosines), a process where a phosphate group is added to the residue. The addition of the phosphate is performed via enzymatically moving one phosphate from an adenosine triphosphate (ATP) to the protein, converting it to adenosine diphosphate (ADP). Enzymes that do this are known as kinases. Removal of the phosphate is also performed via enzymatic action by enzymes known as phosphatases. In general, phosphatases such as Protein Phosphatase 2A (PP2A) are assumed to be promiscuous and constitutively active while kinases change their activity depending on upstream signals [20]. Similarly, proteins can become activated by GTP binding by guanine nucleotide exchange factors, which then become inactivated through its conversion to guanine diphosphate by GTPase-activating proteins.

Phosphorylation represents one type of post translational modification (PTM). Other PTMs can also lead to the activation of pro-



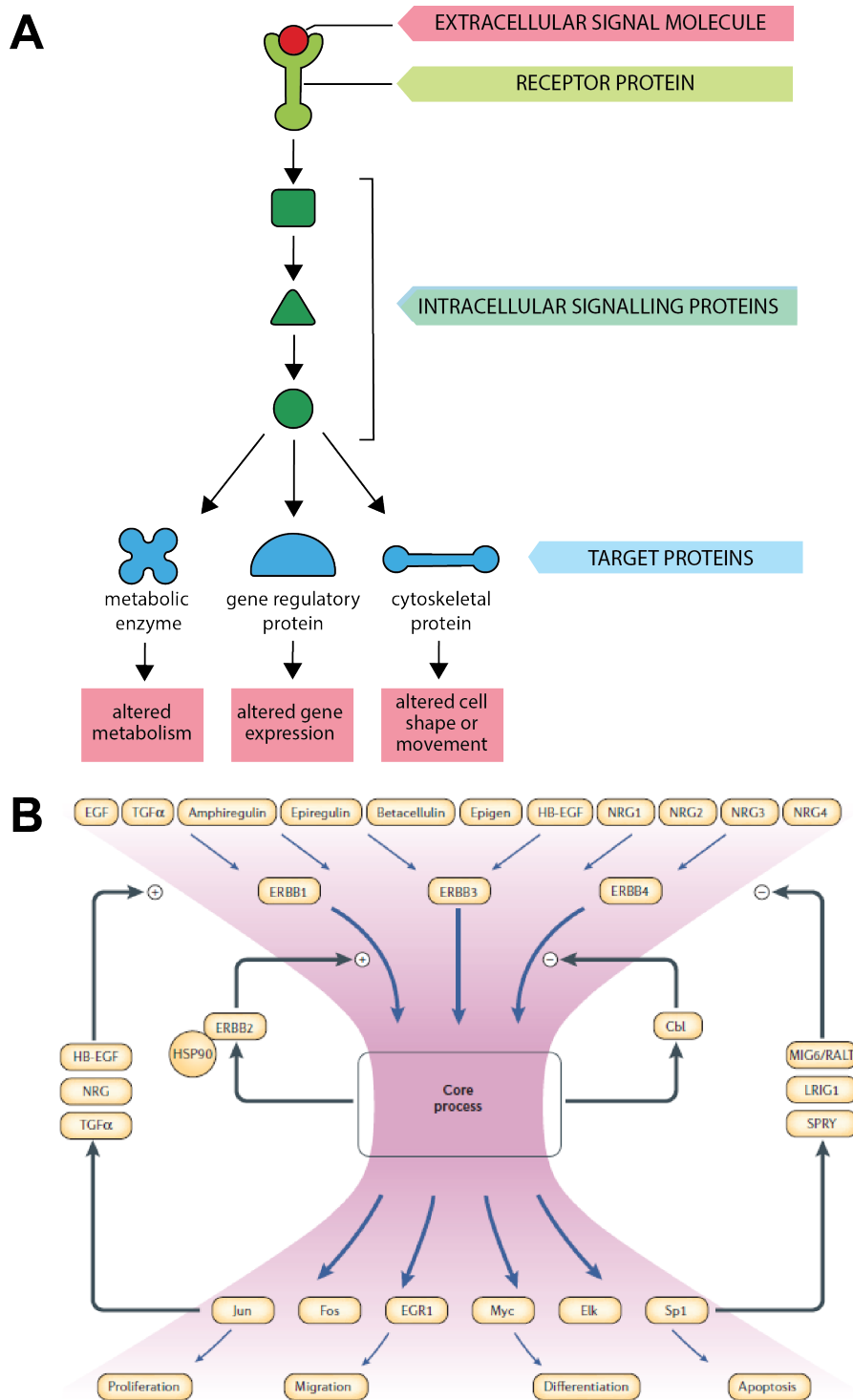


Figure 4: Evolution of our understanding of signalling. (a) The traditional view is linear with a one to one relationship between input and output [18], (b) The contemporary "hour glass" view is complex with a many to many relationship between input and output [16]

teins as well as other changes such as alterations in function, binding affinity or stability. Examples of these are methylation, acetylation, palmitoylation, ubiquitination, glycosylation and oxidation [18, 21, 22]. With continuing improvements in mass spectrometry techniques, it is becoming more practical to identify modifications in protein peptides. This has allowed the function of these other modifications to be studied [23].

Other molecular mechanisms may be required for the activation of signalling proteins such as translocation to specific sub-cellular compartments, or association with other proteins to form signalling scaffolds. More research is revealing that this is an important mechanism in signalling and is required to enable specificity from input signals [24].

### 2.2.3 Overview of the core pathway

The topology of the insulin signalling pathway is preserved between the different types of insulin responsive cells. The differences that lead to their distinct responses to insulin arises from variations in the expression levels of the nodes in the pathway [18, 25, 26]. Insulin signalling begins with the initial binding of insulin to the insulin receptor (IR) at its alpha subunit. This enables activation of the beta component which contains a tyrosine kinase which autophosphorylates the receptor [26]. This leads to the increased recruitment of the adaptor protein insulin receptor substrate (IRS) to the plasma membrane, leading to its subsequent phosphorylation [27].

Phosphorylation of IRS enables PI3-Kinase (PI3K) to bind to it via the Src Homology 2 (SH2) domain of the p85 regulatory subunit [28, 29]. This also recruits the p110 subunit of PI3K, which is already bound to the p85 subunit, to phosphorylate the membrane bound phospholipid phosphatidylinositol (4,5) bisphosphate (PI(4,5)P2) to generate phosphatidylinositol (3,4,5) triphosphate (PI(3,4,5)P3 or PIP3) [30, 31]. The PIP3 can be dephosphorylated into PI(4,5)P2, PI(3,4)P2 by the phosphatase PTEN or SHIP2, respectively [32, 33]. The Pleckstrin Homology (PH) domain of the serine/threonine kinase Akt (also known as Protein Kinase B) preferentially binds to PIP3 [34]. Thus, when PIP3 is generated after insulin stimulation, Akt is recruited to the plasma membrane [35]. At the plasma membrane, 3-phosphoinositide-dependent ki-

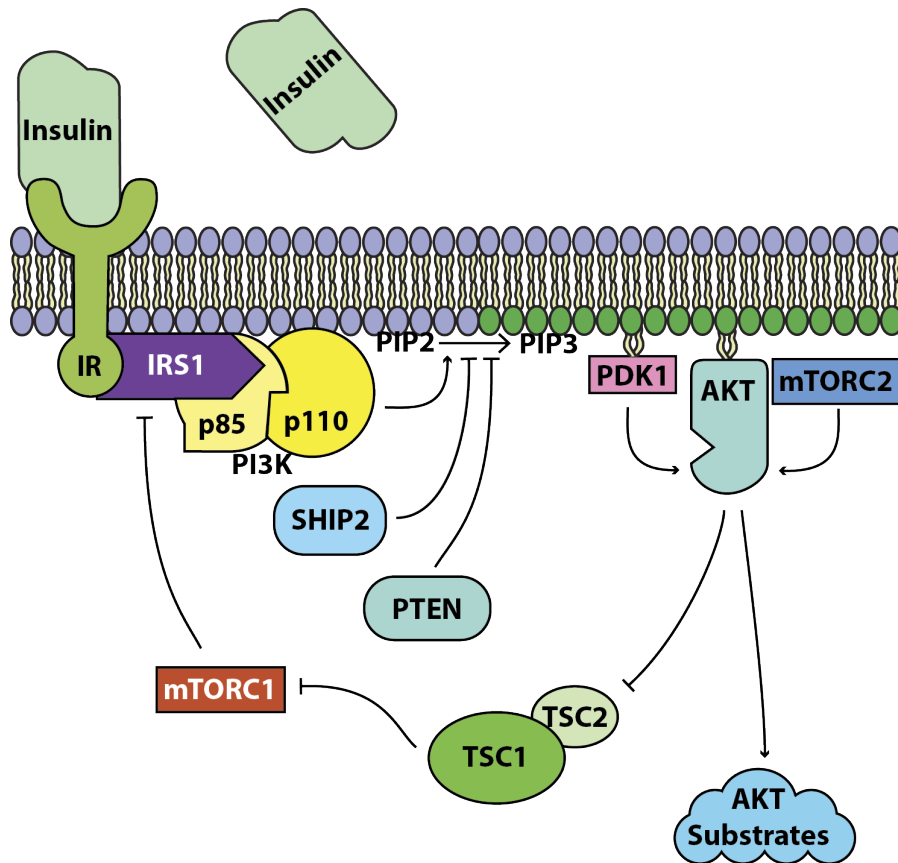


Figure 5: Illustration of the core components of the Insulin Signalling Pathway and how information is transmitted upon insulin stimulation.

nase (PDK1) and mammalian target of rapamycin complex 2 (mTORC2) phosphorylates Akt at the threonine 308 and serine 473 residues (in Akt1, 308 and 474 respectively in Akt2) [36, 37].

Phosphorylation of these two sites activates Akt, allowing it to phosphorylate and alter the activity of its many substrates, which will be discussed in more detail in a later section [38, 39]. One of the key substrates of Akt is TSC2, which indirectly activates the mammalian target of rapamycin complex 1 (mTORC1) [40]. mTORC1 then activates S6-Kinase (S6K), which subsequently phosphorylates IRS1 at an inhibitory serine site [40]. This initiates a negative feedback loop in the pathway [41].

### 2.2.3.1 *Insulin Receptor-Insulin Receptor Substrate*

IR is a transmembrane protein that has been shown to be localised on caveolae in the plasma membrane using gold labelled electron microscopy [42]. Caveolae are domains rich with cholesterol and lipids that are approximately 100 nm in diameter. They have a pit like morphology and are scattered throughout the surface of the plasma membrane [18, 42, 43]. Whilst the precise mechanisms of caveolae function are still not well understood, they are known to cause endocytosis. Based on caveolae's function and IR's proximity to it, it is inferred IR is subject to endocytosis through this mechanism [44].

There is some data to suggest that this mechanism of endocytosis is itself insulin regulated. This model proposes that under unstimulated conditions, caveolae endocytosis is inhibited by caveolins [18]. When caveolins are phosphorylated, its activity is inhibited which re-enables endocytosis [45]. Since it has been shown that caveolin phosphorylation itself is insulin sensitive, it would stand to reason that endocytosis is then insulin sensitive [45].

Caveolae mediated endocytosis represents only one method of receptor endocytosis. An alternative method is via clathrin coated pits. There is some data to suggest that IR activity is associated with a number of proteins associated with this process [46, 47]. endocytosis of the IR is mediated through this alternative pathway [47].

Whatever the mechanism may be though, there appears to be some species specific interaction, as IRS1 is permanently membrane bound in human adipocytes while in rat adipocytes IRS1 is only recruited upon insulin stimulation [27]. Based on these factors, it can be inferred that the endocytosis process has some effect on IRS1 activation and downstream signalling.

In addition to the feed forward mechanism of IRS1, it is also a potential target of feedback regulation in the pathway. Early studies showed that serine phosphorylation lead to its degradation [48]. Later, four serine residues have been identified: 302, 522, 632 and 635. Phosphorylation of only the S302, S632 and S635 appear to be associated with IRS1 degradation [49–51]. The S302 site appears to be targeted by S6K while the S632 and S635 sites are targeted by mTORC2. There is some controversy regarding the roles of these sites as some studies show a positive feedback effect from

phosphorylation of these residues [52]. S522 phosphorylation on the other hand, appears to reduce the tyrosine phosphorylation induced by active insulin receptor and is phosphorylated directly by Akt [53]. However, the inhibitory effect of serine phosphorylation of IRS1 has been questioned with mTOR dependent negative feedback shown to be effective even on upstream growth receptor signalling pathways that are IRS1 independent [38].

### 2.2.3.2 Mechanism of Akt Activation and Akt Substrates

Akt is the node that forms the neck of the "hour glass" of the insulin signalling pathway. As such, understanding its activation mechanism is very important in understanding how the insulin signal is carried forward to its substrates, driving phenotypic changes. The two main prerequisites for Akt activation are translocation to the plasma membrane upon PIP3 generation and T308 and S473 phosphorylation.

Phosphorylation of Akt at the T308 and S473 residues have been shown to be catalysed by PDK1 and mTORC2 using *in vitro* kinase assays and knock down in cell models. This has also been shown in embryonic stem cells, where knockdown of PDK1 does not abolish S473 phosphorylation [54]. mTORC2's ability to phosphorylate the S473 residue was verified using *in vitro* kinase assay of proteins purified from HEK293 cells [37, 55, 56]. Later addition of PDK1 in this assay showed that T308 phosphorylation can occur after S473 phosphorylation [37]. On the other hand, Sin1 knockout studies shows that T308 can be phosphorylated when mTORC2 is inactive in mouse embryonic fibroblasts [57]. These results show that the phosphorylation sites are neither intrinsically exclusive nor ordered. There is also some evidence emerging that PDK1 and mTORC2 are themselves regulated by insulin signalling. In the case of mTORC2, there is some evidence to suggest that one of its sub-components Sin1 is a substrate of Akt. Phosphorylation of Sin1 in turn further activates mTORC2, which formed a positive feedback loop [14]. PDK1 appears to be constitutively active and doesn't respond to additional growth factor stimulation [58, 59].

It has been shown that Akt translocation to the plasma membrane is sufficient to activate GLUT4 translocation, which suggests it is sufficient for Akt activity [60]. It has also been shown

that translocation is sufficient for Akt phosphorylation at T308 and S473, which leads to the idea that phosphorylation of these residues are the prerequisites of Akt activity [60, 61]. Translocation is believed to occur because of binding of the Akt Pleckstrin Homology (PH) domain to PIP3 as PH deletion mutants do not translocate [35]. In spite of this, it has been shown that some PH deletion mutants are still insulin sensitive [62]. Recently it has been argued that this is mediated by an alternative mechanism of Akt activation beginning with mTORC2 activation by PI3K. This leads to S473 phosphorylation that allows Akt to bind to the PIF pocket on PDK1. This allows T308 to become phosphorylated allowing Akt to become active independent of Akt translocation [63].

Additionally, it is unclear if Akt alone is translocated to the membrane, or if its kinases also undergo some translocation as part of the insulin signalling process. *In vitro* studies have suggested that other phospholipids, such as phosphatidylserine (PS), have a role to play in this translocation mechanism, with overexpression of a PS binding protein leading to the loss of PDK1 localisation in the PM [36, 64, 65]. It is also suggested that PDK1 translocation may be regulated by some membrane bound proteins that act as a scaffold between PDK1 and Akt [66]. In terms of mTORC2, the presence of a PH domain on one of its components, Sin1, suggests that it too is translocated to the PM upon insulin stimulation in addition to the phosphorylation mediated regulation discussed before [67].

Downstream functions of Akt involve the phosphorylation of its substrates, of which there are currently over a hundred and counting [41]. Most substrates appear to exhibit a pattern in their amino acid sequence around their Akt targeted phosphosite [68]. This consensus sequence is identified for Akt as R-X-R-X-X-S/T, where R is an arginine, X is any amino acid and S/T is the phosphorylation site (either serine or threonine) [14, 41]. It should be noted that additional regulatory mechanisms are likely to be required beyond the consensus motif because this sequence is often similar to the consensus motif of other kinases. Thus *in vitro*, these kinases are often promiscuous but this is not the case *in vivo* [41, 69]. In spite of this, the consensus sequence has been used as a powerful predictor of other Akt substrates [14, 41]. These substrates can be roughly grouped into a number of key

functions: glucose uptake, energy metabolism (glycogen synthesis, gluconeogenesis etc), survival, proliferation and growth (mitochondrial biogenesis, protein synthesis), and angiogenesis [41, 70, 71]. For more details of the proteins that are proposed to be responsible for each of these functions, the interested reader is referred to an excellent review by Manning [41].

Some of the key substrates that will be explored in this thesis are TSC2 and PRAS40 due to their relationship to the negative feedback loop in the insulin signalling pathway, AS160 due to its relationship to GLUT4 trafficking, GSK3 due to its relationship to glycogen synthesis, and FOXO1 due to its relationship to gluconeogenesis [25, 26, 72, 73]. A key question that emerges from this, is how Akt can activate its downstream substrates with different strengths while being the only trigger for their activity [38, 74]. This is still an area that is not well understood.

Variation within Akt itself also contributes to complexity in this pathway. In humans there are three isoforms of Akt which share approximately 80% amino acid homology [75]. Regulation of their different functions is mediated through differential expression of the isoforms throughout the body. Akt1 and 2 are expressed ubiquitously while Akt3 is primarily expressed in the brain [76, 77]. In the context of glucose homeostasis, Akt2 is believed to be the most important isoform as Akt2 knockout mouse become diabetic and Akt1 overexpression is unable to rescue this effect [78]. The multiple isoforms appear to have some redundant functions, as whole body Akt2 and 3 double knockout mice remain viable albeit with various growth and glucose homeostasis related defects [79]. The high degree of homology between the Akt isoforms has made them difficult to perturb in a specific manner [80]. As a result, knockout strategies are used to isolate phenotypic outcomes with specific isoforms. However, this approach is limited by the compensatory effects that are initiated when proteins are down-regulated for an extended period of time [81]. To overcome this limitation, Kajno et. al. developed a mutant of Akt that is resistant to the inhibitory effects of the small molecule inhibitor MK2206. Since this W80 mutation prevents the inhibitor from binding to the mutant Akt's PH domain, its ability to translocate to the plasma membrane no longer allosterically inhibited. Thus, this MK resistance mutant of Akt can remain active in the presence of MK2206, while endogenous protein become inhibited, which



isolates the effect and function of the mutant Akt isoform [81]. Using this strategy, Kajno found that Akt mediated GLUT4 translocation is not an isoform specific function while control of differentiation was specific for Akt1 [81].

## 2.3 MATHEMATICAL MODELLING OF PATHWAYS

### 2.3.1 Signalling Pathways as Biological Circuits

Signalling pathways, as previously mentioned, are the network in which biological signals are transmitted from a biological sensor, like a receptor, to a biological outcome, like the activation of some biological machinery. While these concepts were originally simple to understand, as the field advanced, the complexity of pathways also increased. This is analogous to increasing the size and complexity of jigsaw puzzles. When the puzzle is relatively simple, it can be solved by eye. However, as complexity and the number of pieces increases, the problem becomes intractable. There is now a general sense that these systems cannot be understood or accurately predicted using intuition. Given this complexity, how can we begin to comprehend the nature of real systems?

A possible answer is inspired by other disciplines that routinely deal with complex systems. One such system is electronics engineering, which in its essence is not too dissimilar to the study of cell signalling pathways [82]. Cell signalling pathways contain fundamental elements (ligands, receptors, enzymes, molecular motors, etc.) just like electronic circuits (batteries, resistors, capacitors, inductors, etc.) and complex behaviours can be created by combining them in novel ways. As such, signalling pathways can really be considered biological circuits. Given that engineers routinely investigate the behaviour of electronic circuits through mathematical modelling, it would stand to reason that mathematical modelling can be a helpful tool in understanding the behaviour of biological systems.

In engineering, knowledge of the relationship between structure and function in specific electronic circuits can be applied in two ways. Firstly, the emergent function of the circuit allows it to be applied in other contexts for synthetic purposes (analogous to synthetic biology). Secondly, understanding the behaviour of



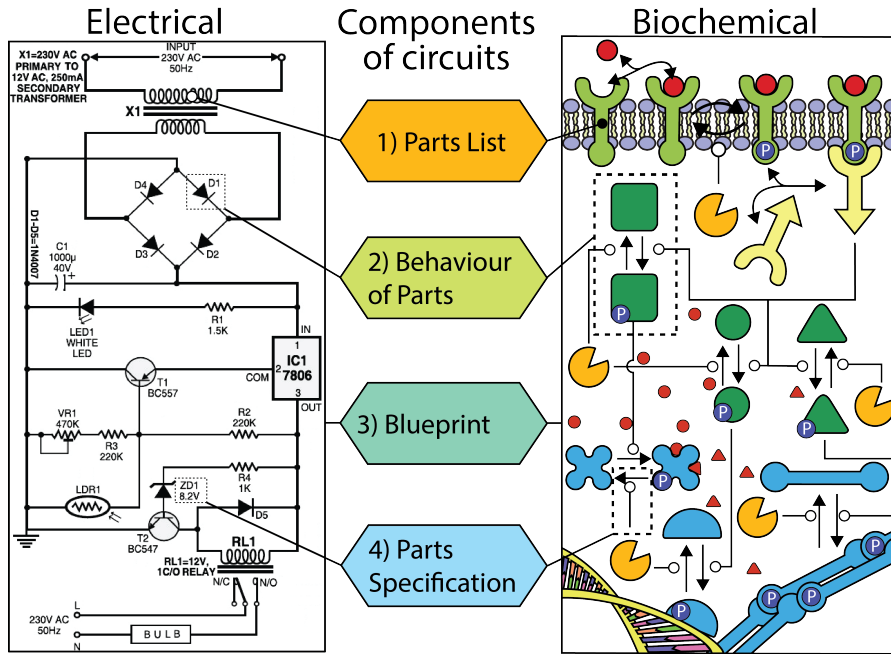


Figure 6: Illustration of the parallels between electrical and biochemical circuit boards. Both systems involve 1) A list of parts 2) which have well defined behaviours 3) are assembled in very specific ways 4) whose parts always have certain specifications.

the circuit helps us understand the origins of defects and how they can be reversed (analogous to disease research and the design of therapeutics). In order to achieve this, it is necessary to have knowledge of the structure of the circuit, which can either be learnt through the blueprint or by studying the wiring in the circuit board. This process is similarly followed when signalling pathways are studied. However, in this context, there is the additional difficulty that the underlying blueprint is not known a priori, and the network cannot be easily studied through their physical wiring. Thus, reconstruction of the system is a necessary first step before the structure-function relationship can be investigated.

In order to reconstruct the system, four pieces of knowledge are generally required (Fig. 6). First is the parts list of the system (e.g. the electrical components). In the case of cell signalling pathways, this is equivalent to knowledge of the proteome, transcriptome, genome and other -omics data. Sophisticated techniques in this area already exist and these -omics level data can be readily

collected or are available in the literature [14]. Second is an understanding of how parts interact and behave (e.g. circuit theory). There are well refined theoretical concepts from physics and biochemistry that are relevant to this system, which enables highly accurate quantitative models to be constructed. Third is the blue print of the system (e.g. the circuit diagram), which customises the model for the particular system investigated. Finally, the specifications of the components (e.g. the precise resistance and capacitances) must be known in order to tune the system to its correct operational parameters. In the case of biochemical systems, these are the rate parameters and dissociation/association constants. In this section, each of these factors (except for the -omics level data) will be discussed with regard to the relevant computational techniques.

### 2.3.2 Overview of Modelling Approaches

The fundamental element of a biochemical circuit are the fundamental reactions between biochemical species, analogous to the fundamental components of electrical circuits: the resistors, inductors and capacitors. These components have a theoretical description, which can be used to describe and predict their behaviour. In this section, we summarise the various approaches from the least complex logical models to most complex spatial-temporal and stochastic models. Models of low complexity typically require less data to validate but are less descriptive of the mechanisms in the system and conversely, high complexity models may produce behaviours that are due entirely to the large degree of freedom in the system which requires a much larger data set to validate.

Boolean models, such as the Saez-Rodriguez model of T-cell signalling, conceptualise proteins as binary nodes that are either on or off [83]. The interdependence of responses between nodes, such as activation and inhibition, are conceptualised as computational OR, AND or NOT gates (Fig. 8A). This simple and unambiguous method allows the model to focus on the direction and pattern of information flow in the system through emergent properties such as attractors and transient points [84]. A shortcoming of this method is that populations of proteins do not always act in an on/off manner, and different nodes can become

activated with only partial activation of a previous node. Thus, some Boolean models, such as the one proposed by Bornholdt, introduce a continuous signalling strength between nodes while retaining their binary on/off status. The strength of activation depends on the length of time the node has been on for and flips the node when a threshold is reached [85]. Alternatively, fuzzy logic is used to enable intermediate states of activation [86]. Additional drawbacks of this approach include the inability to infer biochemical mechanisms and that reaction speeds are not considered. A variety of programs are available for performing this type of analysis [87–91]. A standard has recently been created for designing and interpreting boolean models in an Extensible Markup Language (XML) format that was designed to improve consistency and reproducibility in modelling results [92].

Petri net models conceptualise the model as a system of nodes, which contain resources and reactions to transfer resources between nodes. This type of model focuses on the way resources are moved throughout the network using this concept [93, 94]. This simulates the level of activity of each node as well as the stoichiometry of reactants in interactions (Fig. 8B). In this framework, there are very clear rules for how resource nodes are consumed (as inputs to reactions) and produced (as outputs to reactions) [95]. This method, introduced by Reddy into the systems biology field, builds on Boolean networks by including not just on and off states, but other discrete intermediate levels of activation, abstracted as the population size in each node [96]. These can be consumed which adds more mechanistic detail to the model. The drawback of this is reactions are traditionally untimed, which like the Boolean model, does not reveal temporal information of the system. This can be overcome using the more advanced Coloured Petri Nets (CPN) initially adapted into biochemical modelling by Lee [97]. CPNs allows the algorithm to model timed interactions, and it allows individual resource unit to remember predefined attributes (e.g. phosphorylation state of a site). In algorithm, reactions don't just move resources, they can also modify the attributes.

It is important to note that as timing information and advanced states are included, the mathematics behind Petri nets begin to resemble mass action based kinetic models. For more informa-

tion, the interested reader is referred to the excellent review by Chaouiya [98].

Kinetic models remove the concept of discrete states, replacing it with a continuous function designed to represent the concentration of system states as a population. Like the Petri Net framework, species can be consumed in kinetic models but timing information is encoded explicitly and is the focus of this modelling technique [102]. Kinetic models are constructed using ordinary differential equations (ODEs) that expresses the rate of change of each state as a function of time and the value of other system states (Fig. 8C). This enables chemical kinetics to be utilised in modelling interactions in the system. The chief rate equations used are based on mass action kinetics from which rate equation for more complex reactions are derived. Spatial effects can be approximate by discretisation of spatial components and simulating transport between compartments as mass action reactions. This requires that intra-compartment diffusion is much faster than inter-compartment transport. The limitation of this modelling approach is that states are assumed to be at high enough concentrations that stochastic effects are negligible, and the system is not diffusion limited or contains advective transport and thus spatial aspects are also negligible. When these assumptions do not hold, these complexities must be accounted by using other modelling methods.

Stochastic effects can be accounted for by using models which simulate chemical reactions from the molecular basis (Fig. 8D). In these stochastic models, interactions are modelled probabilistically using the chemical master equation (CME), originally proposed by Gillespie [100]. The CME is the stochastic form of the mass action model and explores the probability of each molecule interacting at any given time. The result is a probability distribution function of all concentrations of a species at any point in time. This tends to diverge from the kinetic model when copy number is less than approximately 100, and so this approach is appropriate for low concentration species like transcription factors and rare events such as gene expression [103, 104]. Because the CME is a differential equation involving all possible states in the system, it suffers from the issue of combinatorial explosion and is thus impossible to solve in practice. As such, a number of methods have been developed which can approximate the solution that can be

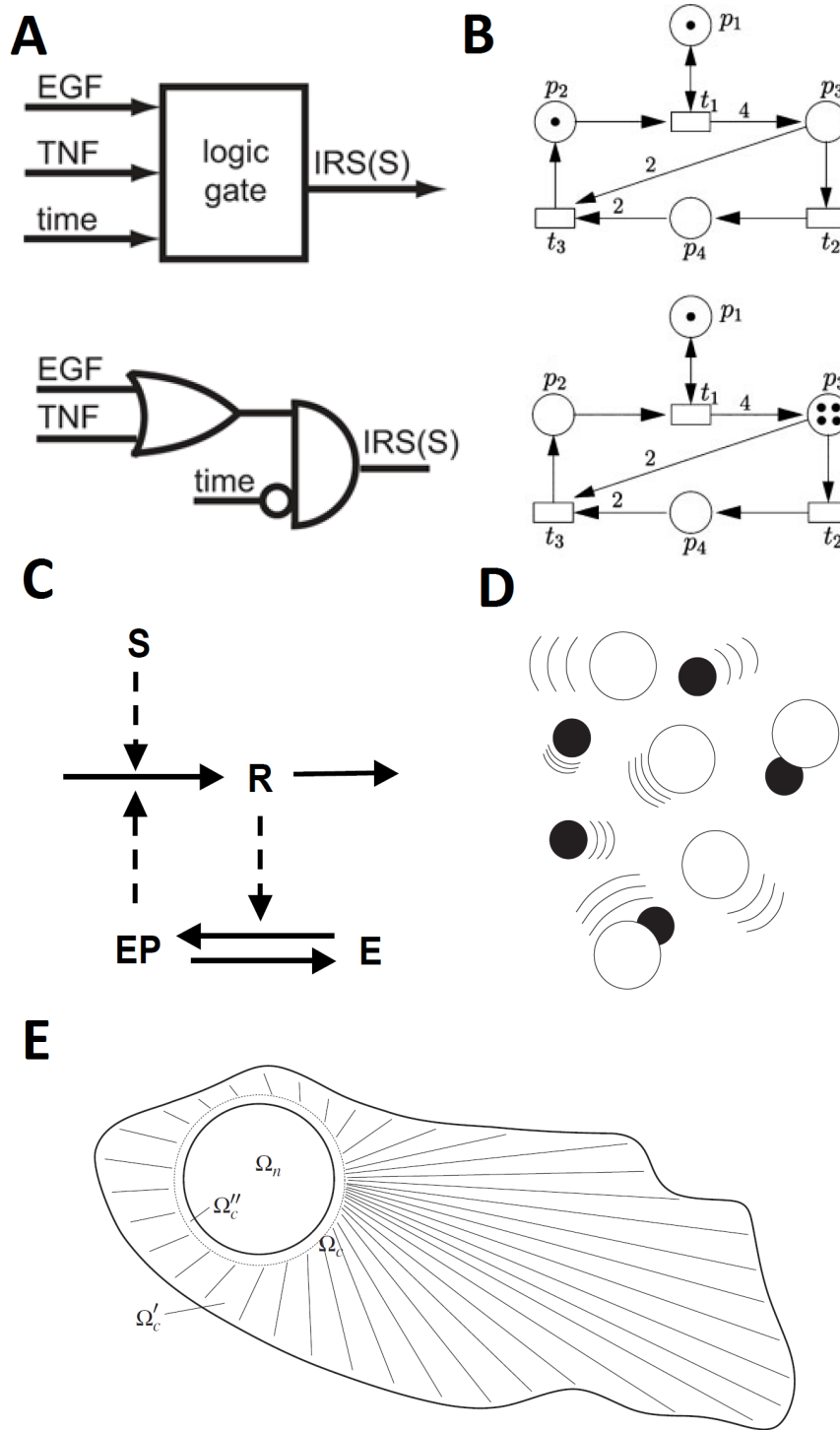


Figure 7: Different modelling methods. (A) Boolean modelling [86]. (B) Petri Net [98]. (C) Kinetic/chemical kinetics [99]. (D) Stochastic/chemical master equation [100]. (E) Spatio-temporal [101].

obtained numerically. We provide some references for interested readers to follow on this topic [105–107]. Apart from the computational cost of this approach, the main limitation is that the spatial aspect of signalling is neglected.

Spatio-temporal models are the same as kinetic models except the spatial component is not smoothed out (Fig. 8E). Typically a spatio-temporal model utilises reaction-diffusion partial differential equations (PDEs) to model all states in the system, although there is an increased interest in reaction-advection-diffusion PDEs [101, 108–110]. This approach is more powerful than the pure kinetic model and has been used to model phenomena such as crowding within the cellular environment which limits diffusion in general, and active transport mechanisms through the microtubule and actin filament networks [111–115]. Spatial models do have the disadvantage of increasing complexity. Firstly, the inclusion of extra spatial dimensions greatly increases the computational cost. Secondly, the use of PDEs requires more complex and less robust numerical methods such as finite element analysis [109, 116, 117]. Thus far this form of rigorous spatial modelling has only been successfully achieved for the simplest geometry and simplest cases [110]. Apart from the computational cost of this approach, the main limitation is that the stochastic effects are neglected.

Finally, it should be noted that the stochastic approach and the spatio-temporal approach essentially have complementary strengths and limitations. As such, a logical method of resolving this is simply to integrate the two methods. While there is no theoretical limitation to this, the two approaches are computationally expensive when applied alone. Thus, merging the two approaches will significantly increase this cost. It is likely that for this reason, a stochastic-spatio-temporal approach has not yet been implemented.

### 2.3.3 Kinetic Models and Mass Action

Kinetic models are deterministic models that are motivated by chemical kinetics. They consider signalling pathways as a collection of chemical reactions. By modelling the kinetics of each reaction, the kinetics of the signalling event can be accurately reproduced, (Fig. 8). Mass action kinetics is the mathematical model

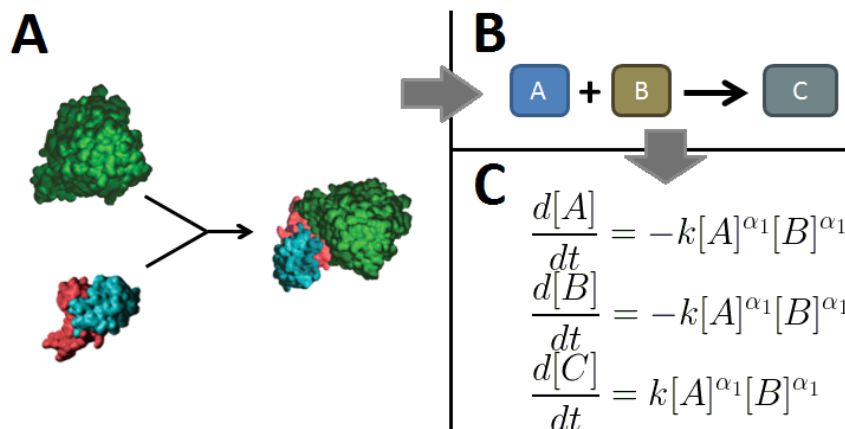


Figure 8: Levels of abstraction of a kinetic model. (A) Reactions occur when molecules within a confined space meet. (B) This can be described as a chemical equation. (C) This leads to a rate equation for the concentration of species involved based on the raw law of the reaction type.

describing these chemical reactions. It began as an empirically determined relationship between the concentrations of reactants and the velocity of the reaction. Mass action based rate laws typically look like:

$$\frac{d[A]}{dt} = -k[A]^{\alpha}[B]^{\beta} \quad (1)$$

$$\frac{d[P_i]}{dt} = -\frac{d[A]}{dt} \quad (2)$$

where  $[B_i]$  and  $[A]$  are the concentrations of the reactants,  $[P_i]$  is the concentration of the  $i$ th product,  $k$  is a rate constant and  $\alpha$  and  $\beta$  are empirically determined constants [118]. This describes the following reaction:



where the association reaction can lead to the formation of multiple decay products.

As these systems were studied in more detail, a physical interpretation of these empirically determined constants were given.  $k$ , was found to be a temperature and pressure sensitive variable

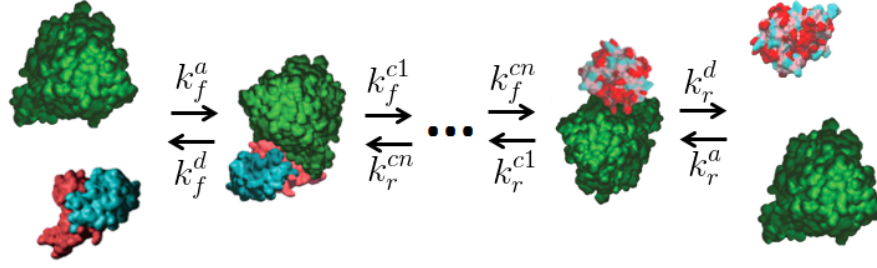
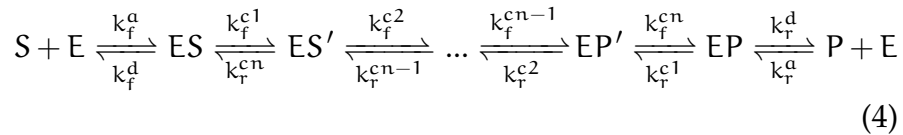


Figure 9: Illustration of the mechanism of enzyme action. The enzyme and substrate first associate, and then they go through successive steps of catalysis before reaching their final thermodynamic form before dissociating into product and substrate. Each step in this reaction is in principle reversible.

that describes the reaction rate at unit concentrations of the reactant [118]. The exponents  $\alpha$  and  $\beta$  are attributed to the stoichiometry of the reaction (i.e. number of molecules of the species involved with the reaction), and molecular crowding in the medium, which is an additional multiplicative factor that increases as the environment becomes more crowded [119].

Elementary reactions are the most fundamental units of kinetic models. These are defined as single, direct interactions between species that results in a chemical reaction [118]. Elementary reactions only involve two reactants because the probability of tri-molecular (and greater) interactions occurring is infinitesimal [100, 120]. Chemical reactions that follow higher order mass action kinetics are likely to be made up of multiple, fast elementary reactions [118, 121].

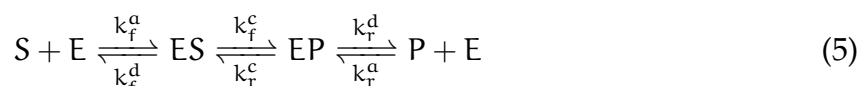
From elementary reactions, enzyme kinetic reactions that are relevant to signalling cascade systems can be constructed [19]. Basic enzyme kinetics, which ignores crowding effects, is modelled using the following series of reactions, which is also illustrated in Fig. 9:



where  $S$ ,  $E$ ,  $ES$ ,  $EP$  and  $P$  denote the substrate, enzyme, enzyme-substrate complex, enzyme-product complex and product, respec-



tively. Primes above the ES and EP complex indicates further conformations the complex may take before undergoing catalysis and dissociation. In general, this can be summarised into an enzyme-substrate complex and an enzyme-product complex where the chemical and thermodynamic change caused by the actual catalytic step defines the distinction between the two complexes. This leads to the general reaction scheme for enzyme action [122, 123].



A number of simplifications have been made to this reaction scheme in order to reduce the dimensionality and the mathematical complexity of the model. The first simplifying assumption is to assume enzymes are irreversible and dissociate quickly upon completion. By assuming  $k_r^c = 0$ ,  $k_r^a = 0$  and very large  $k_r^d$ , this reduces the model to three kinetic parameters [19, 74, 103, 124–130].

The next assumption is the use of the quasi-steady state and the reactant stationary assumptions. The quasi-steady state assumption assumes that the formation of the enzyme-substrate complex occurs very quickly, such that the catalytic rate becomes dependent only on the substrate and enzyme concentrations. The reactant stationary assumption on the other hand assumes the complex concentration is extremely low, which allows its concentration to be ignored in the model. The resulting kinetic model, known as the Michaelis-Menten model, was the highlight of the work of Michaelis, Menten, Briggs and Haldane and has the benefit of further reducing the model dimensionality to two and reducing the number of state variables to three [125, 131, 132].

Unfortunately, the *in vivo* context violates the reactant stationary assumption of the Michaelis-Menten model. This has been recognised by various researchers who have attempted to explore other approaches by performing alternative transformations of the basic reaction [133–136]. An alternative model of note is the total quasi-steady state model, originally proposed by Lim but recently revived by Tzafriri [136, 137]. By deriving the quasi-steady state assumption using the total substrate concentration, Tzafriri is able to avoid invoking the reactant stationary assumption, which

enabled investigation of systems at high enzyme concentrations while retaining the dimensionality reduction of the Michaelis-Menten model [138]. However, this model is mathematically complex and requires reanalysis when applied to promiscuous enzymes or substrates [139].

#### 2.3.4 Building and Verifying Models of Pathways

The next component of the biochemical circuit that's needed is the blueprint. There are general problems with forming the blueprint. First is the discovery of the blueprint, which as we previously stated, is necessary because the blueprint is unknown in this context. Second is the verification of the discovered blueprint, which we want to be confident is correct before it is used to determine the next direction of discovery. Thus, reconstruction of the system's blueprint requires an iterative approach that is described in Fig. 10 [102, 140]. The stages of this process begins with a discovery stage, where the biological mechanisms to be tested are determined and then implemented into a kinetic model. The process then moves into the verification stage, where the model is fitted to experimental data in order to test whether it conforms to known information. If it does, further predictions are made based on the behaviour of accepted models. These behaviours are then carried into the biological domain and tested experimentally in order to verify the predictions. From there, the process returns to the discovery stage. If the predictions were not validated, then the focus will be on the discovery of new mechanisms for explaining the original data. If the prediction was validated, then the model is expanded to include new states and mechanism that may not have been previously included.

The model setup stage has been discussed in the kinetic modelling section. The model analysis stage consists of a number of substages. The first step in the process is parameter fitting. This is usually performed by minimising the least squares residual between the simulated data and the experimental data [52, 141, 142]. This can be done either with local or global optimisation approaches (the books by William Press for local approaches and Eric Weise of global approaches is recommended) [143, 144]. In general, global approaches are favoured [145, 146]. However, many modellers will apply a local optimisation routine seeded

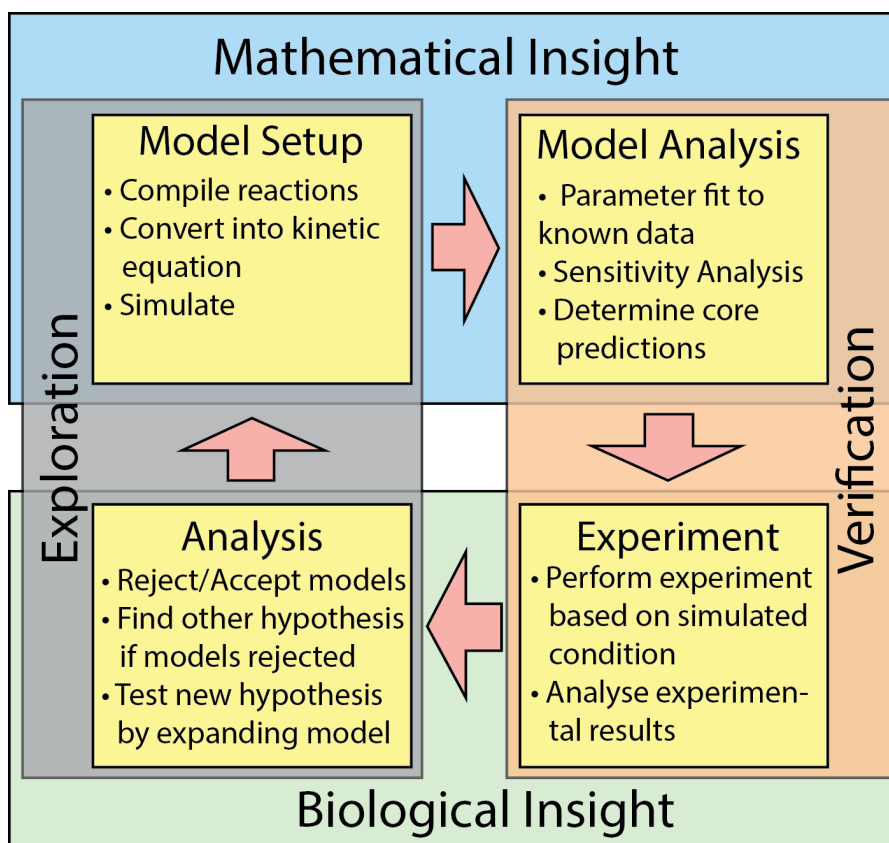


Figure 10: Illustration of the iterative process between model generation, validation and expansion. This also shows the link between the interface between mathematical insights and biological insight and where they sit in the iterative process.

from the solution of the global results to obtain the best fitted solution [52, 141, 142].

In theory, for physicochemical parameters (more specifically rate parameters and equilibrium coefficients) there is only one biologically true value that truly describes the system. This is because these parameters are controlled by the thermodynamics of the reaction which is defined by the physical and chemical properties of the medium (such as temperature, pH and viscosity), or the biochemical makeup of the reactants (such as protein sequence, conformation and post-translational modification state). The former are properties that are tightly controlled through homeostatic processes which if altered can caused system wide changes to physicochemical parameters. The latter are properties that are re-

actant dependent and thus species with different chemical make-ups are typically distinguished in the model to allow them to react with different reaction parameters. However in principle for a distinct species, the reaction parameter should be well defined. In spite of this theoretical argument, in practice the true value cannot be well constrained due to experimental noise or incomplete data [103]. This variability in the goodness of fit is known as parameter non-identifiability or parameter sloppiness [147]. Due to this effect, a single model fitted to one set of data, may not be constrained by a unique parameter set. This leads to an ensemble of predicted behaviours. This effect necessitates the next step, which is parameter sensitivity analysis. In this analysis, the parameters are varied in order to determine these ensembles of predicted behaviour. A number of strategies exist to address this. The most common approach is simply to individually (local) or as a group (global) vary the parameters to find the region around a central well fitting points to determine its degree of sloppiness. Erguler favours the Bayesian statistics approach, where a posterior distribution is empirically reconstructed and then pockets of well fitting parameters are clustered as well fitting sets [148]. Brannmark on the other hand, suggested a more sophisticated strategy that involves aggregating all well fitting parameters by describing them based on their common behaviours. This behaviour is then used as a candidate for verifying the model and is known as a core prediction [149]. With a core or ensemble of predictions determined, the next stage in the process is verification of the prediction by experiment. The model is then adjusted based on the result of the verification step as we have previously discussed, which initiates the next iteration of model development.

## 2.4 MODELS OF THE INSULIN SIGNALLING PATHWAY

### 2.4.1 Models of Insulin Signalling are highly varied

Given that the biology of the insulin signalling pathway is reasonably well studied, there has been increasing interest in creating mathematical models of the insulin signalling pathway. Most of these have focused on the kinetic modelling approach. Although

Table 1: Summary of the different insulin signalling models found in literature.

Model	Model Type	Pathway Scope	Fitted Data	Biological Insight
Wu 2009 [150]	Asynchronous Boolean	Insulin and PI3K pathway, parallel Akt, MAPK and mTORC1/S6K negative feedback	S312, T451 and Y941 IRS1 and T308 Akt in HepG2 cells	PP2A does not dephosphorylate Akt in the HepG2 cell line.
Aldrige 2009 [86]	Fuzzy Logic	TNF, EGF and Insulin as input signals. The downstream components then combine the signals and with negative feedback systems via ERK and Akt	Phosphorylation of 8 proteins such as Akt, IRS1, and MK2 in HT-29 cells	Negative cross talk between ERK and the Akt pathway.
Quon 1991 [151]	Mass Action	Insulin receptor only in the plasma membrane and cytosol.	Insulin receptor localisation in BC3H-1 myocytes.	Mechanism of insulin binding to and internalisation of IR
Wanant 2000 [152]	Mass Action	Dimerisation of IR and its divalent bonding to Insulin.	IR imaging and double probe labelling of IR.	Order of binding of Insulin to the IR.
Sedaghat 2002 [153]	Mass Action	Extension of the insulin receptor model by including downstream components such as Akt and GLUT4 translocation.	Concentrations and rate parameters from previous experiments in 3T3-L1 adipocytes.	Response of the wider insulin signalling pathway.
Brannmark 2010 [149]	Mass Action and Hill Function	Insulin receptor internalisation model	Insulin receptor location and phosphorylation in primary human adipocytes	Explains IR internalisation mechanism and new pathway connection to explain overshoot.
Nyman 2012 [154]	Mass Action and Hill Function	Insulin receptor internalisation model in murine cells	Same as Brannmark 2010 but in primary murine adipocytes	Explains pathway differences between human and murine cells.
Kubota 2012 [74]	Mass Action	Feed forward/back by Akt substrates	Time course of Akt and various substrates in Fao cells	Network motifs lead to low or high pass filtering of signals
Dalle Pezze 2012 [127]	Mass Action	Regulation of mTORC2 and Akt activation.	Temporal and inhibitor treated responses of Akt and TSC2	A negative feedback independent of PI3K is hypothesised.
Brannmark 2013 [124]	Mass Action and Hill Function	Brannmark 2010's IR to Akt and GLUT4 pathways to explore insulin resistance.	Temporal study of Akt + substrates in human adipocytes	A number of mechanisms of insulin resistance identified

others have adopted alternative approaches such as discrete or fuzzy logic boolean networks [86, 150].

In spite of the vast biological network that has been mapped using biological experiments, mathematical models of insulin signalling are often significantly simplified. In fact, different models will incorporate different nodes and connections from the known network, and consolidate other nodes into a single node in order to reduce complexity. This is because different models are focused on different aspects of the network's behaviour. Thus, modellers typically only choose to include the components that are relevant to the area of focus. This enables model simplification, which as we discussed in section 2.3, is a major requirement for making kinetic models tractable. Boolean models do not face this challenge and thus address more complex network topologies.

In this section we will begin by discussing the various discrete and logical models of insulin signalling, then move to the various kinetic models of the system in the literature. As an overview, these models have been listed and summarised in table 1.

#### 2.4.2 Logical and Discrete Models of Insulin Signalling

The two insulin signalling models of this type are the Wu et. al. discrete model of liver insulin signalling, and the Aldridge et. al. fuzzy logic model of TNF/EGF and insulin signalling in carcinoma cells.

The Wu et. al. discrete model aims to model the behaviour of HepG2 liver cells with the goal of testing and verifying a novel negative feedback mechanism mediated by the protein PKR acting on IRS1. In this model, each node can be in one of three states: 0 (off), 1 (intermediate) and 2 (on). When an activator/inhibitor acts on a target at a level higher than the control, it increases/decreases the state of the target node by one respectively. This model uses asynchronous updates for its nodes to model, which in principle assumes phosphorylation events all occur within the same time scale. Cell to cell variations are modelled by randomising the initial conditions. Using this modelling setup, the simulation was repeated 10,000 times to form a simulated population from each simulated individual cellular responses [150].

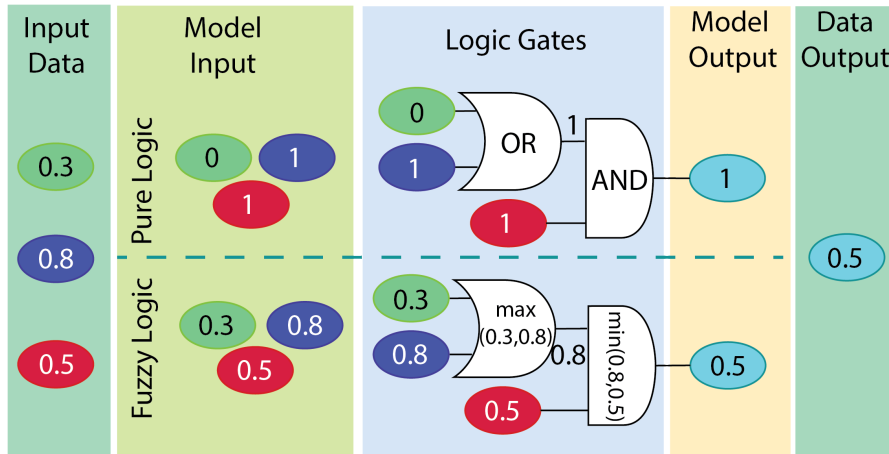


Figure 11: Simplified comparison of input data passed through a pure and fuzzy logic model.

When all components of the known pathway were included, the predicted temporal behaviour in an arbitrary time scale did not match experimental data. After analysis of the results it was concluded that the dephosphorylation of Akt by PP2A cannot be true as it is the source of the mismatch [150]. There is some experimental evidence in the literature that this is the case in HepG2 cells, unlike in cells like 3T3-L1 adipocytes [155]. Further analysis of the model led to the conclusion that the ERK negative feedback did not agree with experimental data and was excluded. However it should be noted that this was not rigorously corroborated by existing literature [150].

The second model is the Aldridge model, which aimed to predict the combined behaviour of the TNF, insulin and EGF pathway in HT-29 cells. Due to the complex web of crosstalks that characterises this model, a fuzzy logic model was applied. In short, fuzzy logic models allow logical operators (like AND and OR) to be applied to inputs and produce outputs with values between 0 and 1. Logical operators are applied using mathematical operators instead. I.e. OR is implemented either as the maximum, or the probabilistic OR ( $a + b - ab$ ) of the two inputs. AND is implemented either as the minimum or the product of the inputs. A simple example is illustrated in Fig. 11 [86].

For a given input, a pure logical model satisfies only one case and generates a single exclusive output. For fuzzy logic models, these inputs can satisfy multiple cases, due to its non-discrete



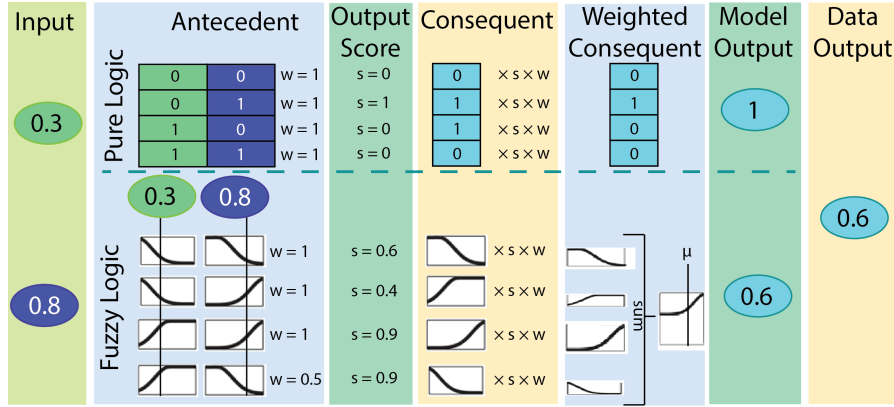


Figure 12: Illustration of the way multiple cases are handled in the fuzzy logic framework. This allows the outcome of multiple satisfied cases to be aggregated to form the final output.

nature. This collection of cases is called the Antecedent. Sometimes, a case is modelled as having less of a contribution to the outcome. These cases can be given a weighting (which are in practice, optimised to the experimental data) to account for this. From there, cases in the antecedent are assigned a corresponding output probability density function (PDF). This collection of output PDFs is called the Consequent. To obtain the final output value, an output "score" is first obtained for each antecedent based on the logical gate rule. The distribution of the corresponding Consequent is then multiplied by the score and the weighting for that antecedent. All weighted Consequents are then summed together and normalised to the area under the summed consequent curve. Finally the mean of the normalised summed consequent curve is calculated which forms the output of the fuzzy logic gate. The application of the Antecedent and Consequent is illustrated in Fig. 12 [86].

The Aldridge fuzzy logic model was fitted to experimental data of cells stimulated under various combinations and doses of the three inputs [86]. The outcome was compared to temporal phosphorylation data of nodes within the model. It was found the fuzzy logic model was significantly better at explaining the data compared to a discrete logical model (analogous to the Wu model) [150]. In spite of this, there were still nodes which were not well explained by the model, such as the IRS1 serine phosphorylation response to EGF stimulation.



### 2.4.3 Kinetic Models of Insulin Signalling

Kinetic models of the insulin signalling pathway can be roughly broken down into two types.

- Models focused on the upstream insulin-insulin receptor components.
- Models more generally focused on the whole pathway, some of which are more comprehensive than others.

We will go through each of these types in turn and describe the models that fit within these descriptions.

#### 2.4.3.1 *Focused models of Insulin Signalling*

As the insulin receptor (IR) is the sensor for insulin signalling, the mechanistic details of its activation is crucial for the response of all downstream components. In fact, the dynamic and kinetic mechanisms of these two nodes are extremely complex as we described in detail in section 2.2.3.1. Work on simulating the dynamics of this system began in the 1990's with Quon who explored various models of this system [151]. He looked firstly at an endocytosis model of the IR and then at their singly and doubly insulin bound monomer and dimer forms (shown in Fig. 13 A and B respectively) [151, 152]. These models were compared to temporal data and steady state dose response data respectively, and the parameters of these early models were fitted by inspection rather than by computation. These two models became the basis of the Sedaghat model of the insulin signalling pathway, which is a more general model which we will discuss in more depth later.

With an improved kinetic modelling strategy and model comparison techniques, this system was revisited by the Stralfors and Cedersund group a decade later [149]. Using novel global optimisation techniques and improved computational capabilities, the group constructed both Quon models and compared both models against new phosphorylation data of the IR and Insulin Receptor Substrate 1 (IRS1) nodes in human adipocytes. Instead of determining the parameter values by inspection, they were determined by parameter fitting and the models were evaluated using a concept called "core predictions". Roughly speaking, this

involves looking at the range of predicted quantitative behaviour within all accepted parameter models and then identifying those behaviours that encompasses all accepted models. This prediction is a core prediction because if it fails, it has the ability to not just invalidate the parameter models, but the entire topological model [149].

In using this approach, a number of models, particularly those that do not include an endocytosis mechanism (such as ones similar to the Wanant model) were excluded. These models failed for a number of reasons including an inability to replicate temporal features like overshoot of the phosphorylation kinetics, incorrect timing of the overshoot peak, unrealistic degradation rate of insulin and dependence of overshoot behaviour on the internalisation mechanism. This ultimately led to the identification of an additional mechanism promoting the recycling of endocytosed

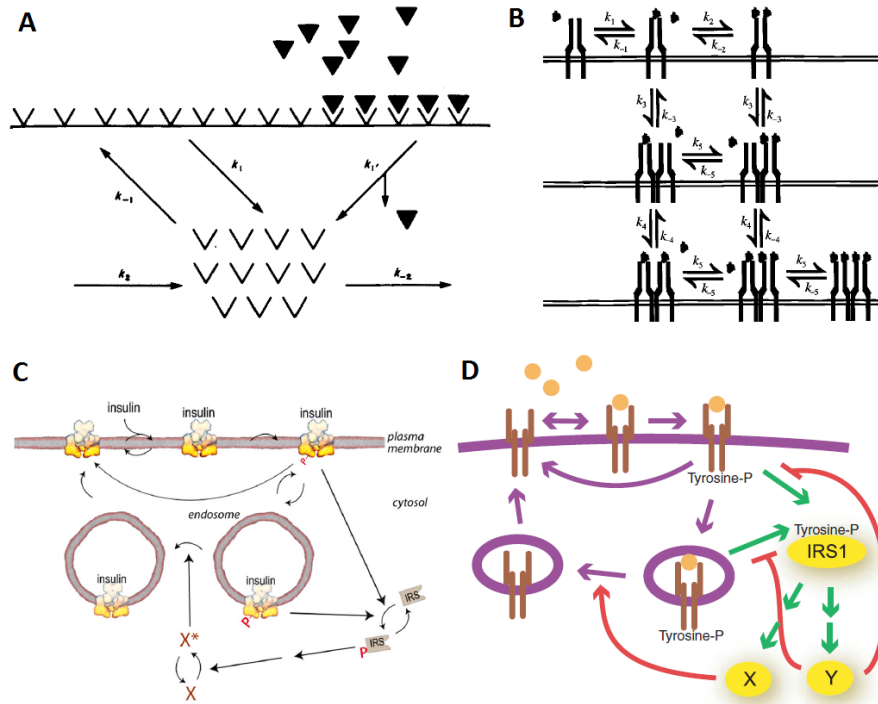


Figure 13: Evolution of Insulin Receptor and Insulin Receptor Substrate models in the literature. Starting with (A) Quon's endocytosis model [151]. (B) Wanant's insulin binding model [152]. (C) Brannmark's feedback endocytosis model. (D) Nyman's enhanced feedback endocytosis model.

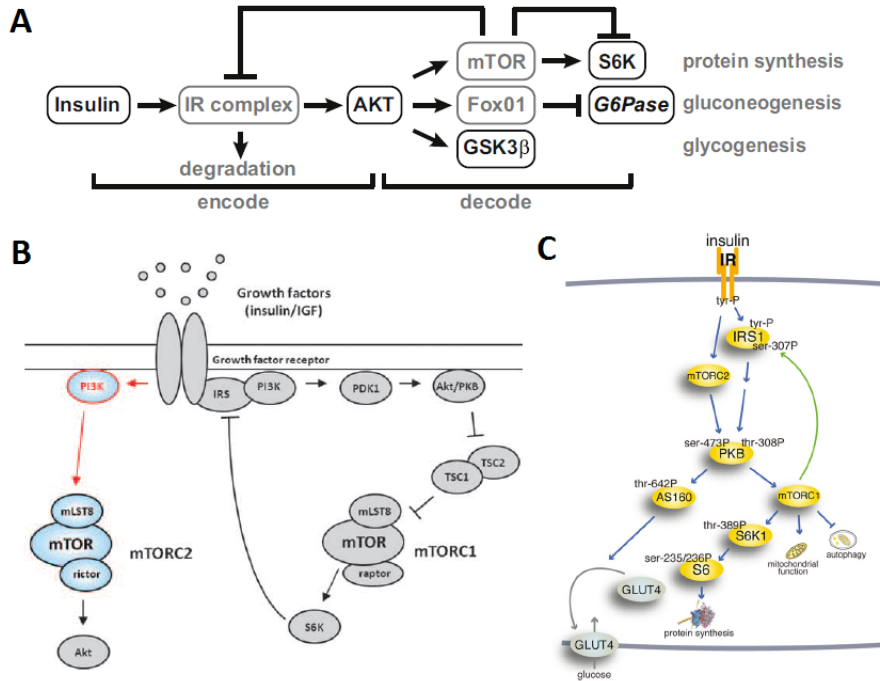


Figure 14: Large abstracted models of Insulin Signalling (A) Kubota model [74], (B) Dalle Pezze model [127] (C) Brannmark T2D model [124]

receptors while degrading their attached insulin, which is illustrated in Fig. 13C [149].

The resulting Brannmark IR-IRS model of receptor endocytosis was further improved by comparing the behaviour of this system in other species. When performing the same experiment and analysis on murine adipocytes, it was found that while the overshoot behaviour was consistent, the order of the overshoot between IR and IRS1 was reversed. This indicated that further development of the model was required, at least in the case of the murine cell model. The development of this model resulted in the inclusion of a new mechanism in their model of the insulin signaling pathway. This new mechanism involves a novel substrates that is activated by IRS1, which subsequently inhibits the activation of IRS1 by IR. Fig. 13D illustrates the structure of the updated model with the new mechanism included [52].

#### 2.4.3.2 *General Models of Insulin Signalling*

General models of insulin signalling attempt to build a comprehensive network of the pathway from the IR to nodes that are general proxies for phenotypic outputs. There are generally two subapproaches to this. First is a simplified model that aims to capture the essence of the pathway. These models do not incorporate all known mechanisms and greatly abstract the pathway. The second approach attempts to capture as much mechanistic detail as possible, where validated and only for components that are of interest to the study. These models usually build on another model, such as those described in the previous section, section 2.4.3.1. In our discussion of these general models, we will explore the abstracted models first, then discuss the more complex models later.

Abstracted models generally come in two forms. One form is simplified models that reduce the dimensionality of the system by removing intermediate nodes or leaving out mechanisms which appear to be unimportant. The first of these examples is the Kubota model of insulin signalling (shown in Fig 14A), which had the primary aim of examining the temporal behaviour of the signalling pathway, which makes use of one of the key advantages of kinetic modelling. This only requires the various feedback mechanisms of the pathway to be captured, thus intermediates in the pathway were omitted. This model concluded that some temporal responses that were experimentally observed from Fao cells could be explained with certain network motifs. For example, the inability of S6K to generate any sustained response was resolved by incorporating a Incoherent Feed Forward Loop (IFFL) onto the S6K node, where the upstream kinase is first activating, but later inactivating. This required a new unknown protein which transmitted the feedback component of the network motif to be incorporated into the model. Furthermore, the model shows that varying the reaction rate of a Feed Forward mechanism can change the sensitivity of the downstream components to acute stimulations [74].

Other simplified models take advantage of the mechanistic nature of kinetic models instead. In the case of the Dalle Pezze model (shown in Fig 14B), the aim was to evaluate and validate the mTORC2 to Akt negative feedback mechanism and explored the impact of alterations to the PIP3-AKT components of the path-

way. Again, the simulation prediction was fitted to temporal data from HeLa cells. In this model, because mTORC2 regulation was the key focus, the model was designed to include the detailed mechanism of mTOR associated with its various binding partners which form the mTORC1 and mTORC2 complexes, both of which are integral parts of the signalling network [127].

The base model was validated using S6K phosphorylation levels under different efficiencies of Raptor knockdown. Upon validation, the fitted model was used to test various hypothesis of mTORC2 activation (assumed to be tied to phosphorylation of mTOR on the 2481 site). The PI3K dependent (tested with Wortmannin), TSC2 dependent (TSC2 knockdown) and mTORC1 negative feedback (Raptor knockdown) hypotheses curated from the literature were tested and rejected as they did not agree with model predictions. From there, a PI3K-dependent step that was immune to the mTOR/S6K negative feedback loop (i.e. different PI3K than the one that generates the PIP3 which Akt binds to) was proposed. In effect, this implies that the new hypothesis was Akt independent, which led to a series of validating experiments. Transfection of constitutively active Akt did not result in any additional phosphorylation of S2481 phosphorylation [127]. This conclusion appears to contradict Humphrey et al's finding that mTORC2 activity is tied to Sin1 phosphorylation, which is regulated by Akt [14].

Other models are constructed as incremental improvements of previous models. The commonly known Sedaghat model is one such example, as it was built on the Quon and Wanant models [153]. This model is built from three separate models which focused on: IR binding, IR recycling and GLUT4 recycling [151, 152, 156]. The IR components form the input module while the GLUT4 components the output module. At this point there was no model of the signalling machinery linking the input to the output. As such, this model bridged the gap using a linear cascade of IRS1, PIP3 and Akt, modelled using empirical linear kinetics. Furthermore, the rate parameters used in this case was determined using dimensional reduction and theoretical arguments rather than by experimentation [153].

Another model that followed the same concept is the Brannmark T2D model of insulin signalling (shown in Fig 14C) [124]. In this study, the authors sought to determine the potential defects

Table 2: Comparison of the fitted initial concentration of IR and binding rate between Insulin and IR in various models of insulin signalling.

Model	$[\text{IR}]_{\text{total}}(\text{nM})$	$k(\text{nM}^{-1}\text{s}^{-1})$
Wanant 2000 [152]	0.1	$1 \times 10^{-3}$
Sedaghat 2002 [153]	$9 \times 10^{-3}$	$1 \times 10^{-3}$
Brannmark 2010 [149]	10	0.00649
Nyman 2012 [154]	10	0.00649
Kubota 2012 [74]	46.2	0.130
Dalle Pezze 2012[127]	12.1	0.00248
Brannmark 2013[124]	100	0.0106

that are created when an adipocytes changes to the T2D state. The phenotypic changes associated with the T2D state were IR and GLUT4 concentrations had previously been shown to decrease to 55% and 30-70% , respectively. Because GLUT4 was determined to be a marker of the T2D state, it needed to be introduced into the model. Thus, the original Brannmark model of IR-IRS1 signalling pathway was expanded into a general model of insulin signalling. In order to identify the potential defects, the model was fitted to time course data from human adipocytes from T2D subjects. No parameter was found that could fit the new dataset and thus the working model was rejected and a new model mechanism was invoked. Using various theoretical arguments, a positive feedback loop from mTORC1 to IRS1 was incorporated. From there, the new model was refitted to the data to generate the new working model [124].

As we can see, there are a wide range of quantitative insulin signalling models and on face value they appear to be identical. However, a closer look at the models reveals that there are subtle differences between the models. The least subtle is of course the wiring of the networks. The Kubota model and the Brannmark T2D model both include a feedback mechanism from mTORC1/S6K back to the IR-IRS1, but it was implemented as negative and positive feedbacks respectively [74, 124]. More subtly, the parameter values and initial concentrations of the components used are different. To illustrate this, Table 2 compares the initial concentration

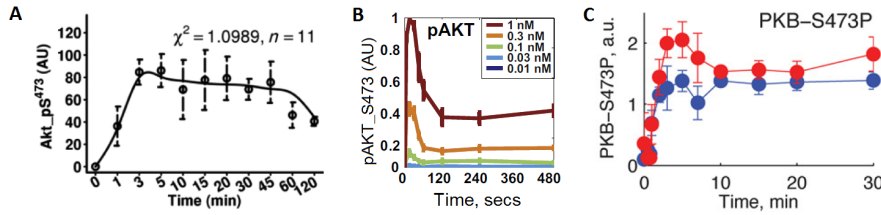


Figure 15: Akt phosphorylation data that was used to fit the various models. (A) Dalle Pezze model (adapted from Pezze et al 2012 [127]) (B) Kubota model [74]. (C) Brannmark T2D model [124]

of the insulin receptor in the various models as well as the binding rate between it and extracellular insulin.

As this comparison demonstrates, the predicted concentrations and kinetic rate parameters can vary across 3-4 orders of magnitude between the models. This means they are not consistent with each other and should in principle generate quantitative differences in predictions. Does this mean they are wrong? Before jumping to the conclusion that they are, we must consider that these models are fitted to their respective data. As such, when each model is considered in isolation with their datasets, they actually fit quite well, and often the model has undergone some validation in their model system. So if the models are data driven, why are they inconsistent with each other? The answer is as simple as the data being inherently different from each other.

By placing the experimentally obtained time course of the same state (for example in Fig. 15) it can be seen that they have different features and time scales. The Dalle Pezze dataset does not show any overshoot in the phosphorylation dynamics of Akt, while the overshoots seen in the Kubota and Brannmark datasets appear to occur in different time scales, with the Kubota overshoot occurring much earlier [74, 124]. In fact, if we take a closer look at the origins of the datasets, although they are performed with the same experimental conditions, that is stimulation with 100 nM insulin (except for the Kubota case, where it was 1nM), they were in fact carried out on different cell types, ranging from human primary adipocytes to HeLa cells, C2C12 cells and Fao cells [74, 124, 127]. Thus, what this demonstrates is while each model has performed well in the context of their model systems, it is still

questionable as to whether models can be translated between different cell models while retaining their accuracy.

## 2.5 MOTIVATION AND AIMS FOR THIS THESIS

The main aim of this thesis is to develop a mechanistic model of the insulin signalling pathway for the 3T3-L1 adipocyte. The purpose of the model will be to:

- Evaluate new connections discovered in the network by making experimentally testable implications of the network.
- Make predictions about the behaviour within the network that can be further constrained by experimentation to improve the fit of the model.

In order to do this, three objectives must first be achieved:

1. To improve the mechanistic accuracy of the modelling approach and balance it with the number of free parameters required, enzyme kinetic models need to be re-explored in order to develop and validate a method that minimises the number of necessary assumptions.
2. A model creation algorithm needs to be developed that can utilise the improved kinetic model if it cannot be integrated with existing tools.
3. A parameter fitting technique that accounts for inherent non-identifiability in the system by thoroughly exploring the parameter space in an efficient manner.

With the tools in place, a model will be constructed anew using the 3T3-L1 adipocyte as the model template. The reason for reconstructing this model of insulin signalling afresh is existing models were developed based on data from different cell types. Hence, these models may incorporate signalling mechanisms which do not exist in 3T3-L1 adipocytes and thus lead to unrealistic predictions.

The initial goal will be to construct with the insulin receptor because it forms the input of the system. Subsequently, the Akt pathway will also be explored. This is because there is significant



uncertainty in the mechanisms involved in the pathway, and little work has been performed in comprehensively investigating this pathway using mathematical modelling. Thus, this pathway is ripe for the application of quantitative modelling approaches.



## DEVELOPMENT OF A NEW ENZYME KINETIC MODEL

---

### Abstract

Enzymatic reactions underpin the regulatory mechanisms in biochemical networks. In order to accurately model the biochemical behaviour of an entire signalling network, the model must predict each individual reaction accurately to avoid errors propagating and amplifying throughout the simulated system. In this chapter, we theoretically analyse common enzyme kinetic models in order to understand the origins of the assumptions that underpin them. Then we perform some reanalysis of the single and two substrate models to remove the low enzyme concentration (or reactant stationary) assumption in order to broaden the conditions under which the model remains accurate. We will also demonstrate that this model can be applied universally without mathematical reanalysis of the model irrespective of the network topology of the system under investigation.

Some of this work has been published in:

- Wong MK, Krycer JR, Burchfield JG, James DE, Kuncic Z. A generalised enzyme kinetic model for predicting the behaviour of complex biochemical systems. FEBS Open Bio 2015; 9(5):226-239 doi: 10.1016/j.fob.2015.03.002.

### 3.1 SUMMARY OF NOTATIONS IN THIS CHAPTER

The following notations are used throughout:

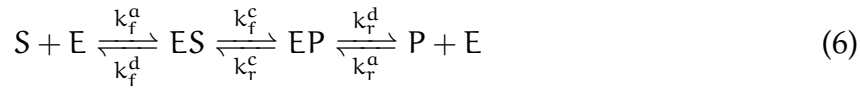
- Species names enclosed in square brackets implies their concentration, e.g.  $[S]$  is the concentration of  $S$ .
- Overhead dots of state variables indicate the total time derivative. e.g.  $\dot{[S]} = d[S]/dt$ .
- Species can contain a subscript which are used to describe the following properties:
  - $T$  indicate total (free and bound) concentrations e.g.  $[S_T] = [S] + [ES]$ .
  - $0$  indicates initial concentrations e.g.  $[S(t = 0)] = [S]_0$ .
  - $\infty$  indicates steady state concentrations e.g.  $[S(t \rightarrow \infty)] = [S]_\infty$ .
- Rate parameters  $k$  contains subscript that defines the reaction direction (f or r for forward or reverse) and a number. E.g. subscript  $k_{f_1}$  indicates rate parameter of the forward direction of the first enzymatic reaction. When the reaction only occurs in one direction, the letter is dropped and the subscript thus only contains one number (e.g.  $k_2$  indicates rate parameter of reaction 2).
- Rate parameters contain superscript a, d, or c which indicates the association, dissociation and catalytic rates of the enzyme reaction (e.g.  $k_{f_1}^c$  is a catalytic rate).
- Variables in bold font denote tensors and their numerical subscripts indicate their indices.

### 3.2 INTRODUCTION

Systems modelling of intracellular biochemical processes can provide quantitative insight into a cell's response to stimuli and perturbations [157]. If the model is mechanistic, it has the power to infer molecular mechanisms and predict biological responses [158].

This requires the simulation of biochemical reaction kinetics typically described using ordinary differential equations (ODEs). Modelling enzymatic cascade networks, however, requires the simulation of multiple reactions. This inevitably increases the complexity of the ODE model, which increases the number of free kinetic parameters. It then becomes more difficult to constrain all parameters simultaneously using the limited experimental data available [147]. This can result in the derivation of multiple well fitting models with limited predictive power because of their non-uniqueness. Thus, an optimum parameter dimensionality should be selected to reduce non-uniqueness without reducing the topological complexity required to capture key kinetic features in the system [102].

Of the biochemical processes that need to be modelled, many are enzyme reactions [19]. Enzyme reactions involve the chemical modification of a reactant (the substrate) that is aided by another reactant (the enzyme) by reducing the activation energy. In these reactions the enzyme maintains its chemical state throughout the reaction. Enzymatic cascades are based on enzyme kinetics where the product of one reaction forms the enzyme of the next. Within which additional interactions such as inhibition and allosteric effects can be included using mass action kinetics [159]. Basic enzyme kinetics is modelled using the following series of reactions:



where  $S$ ,  $E$ ,  $ES$ ,  $EP$  and  $P$  denote the substrate, enzyme, enzyme-substrate complex, enzyme-product complex and product, respectively. The full mass action description of this reaction requires six kinetic parameters:  $k_f^a$ ,  $k_f^d$  and  $k_f^c$  are the forward association, dissociation and catalytic rate parameters, respectively, and  $k_r^a$ ,  $k_r^d$  and  $k_r^c$  are the corresponding reaction rate parameters in the reverse direction.

Many models of biochemical systems use the simplified irreversible form of the reaction, which only requires three kinetic parameters [19, 74, 103, 124–130]. While this is an approximation of real enzyme action, in vitro spectroscopic studies of single molecule enzyme kinetics have shown that this approximation is sufficient in experiments where there is no product in-

hibition [123, 160]. Further simplifications have led to other enzyme kinetic models such as the Michaelis-Menten model and the Tzafriri Total Quasi-Steady State Assumption (tQSSA) model [103, 125, 133–136]. Whilst the Michaelis-Menten model is more widely used, it is strictly accurate at low enzyme concentrations. Since this may not be true under *in vivo* conditions, unrealistic conclusions may be drawn from models using the Michaelis-Menten equation [136, 139, 161, 162]. The tQSSA is not subject to the same limitation, but it has a more complex mathematical form that requires reanalysis for each distinct network to which it is applied [139]. Currently, systems modellers must choose between complex enzyme models with high parameter dimensionality, or simpler models at the cost of accuracy.

A further compounding factor is that *in vitro* investigations of enzyme action are generally performed in closed thermodynamic systems which achieve thermodynamic equilibrium, as reflected in the model described by Eqn. 6. Cellular systems, however, are not thermodynamically closed, and so achieve only homeostatic equilibrium. This is achieved by constant energy inflow through coenzymes such as ATP. This allows the network to form cyclic reactions made of counteracting enzymatic reaction pairs which maintain and regulate this equilibrium. Examples of cyclic reactions are the cyclic interconversion of nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ) and nicotinamide adenine dinucleotide phosphate ( $\text{NADP}^+$ ), mediated by NAD kinase and  $\text{NADP}^+$  phosphatase in metabolism, and the cyclic interconversion of phosphatidylinositol (4,5) - biphosphate to (3,4,5) - triphosphate, mediated by PI3K kinase and PTEN phosphatase in insulin and cancer signalling [32, 58, 163]. Thus, models of cellular systems need to account for the continual energy consumption in these cyclic reactions. Conventionally, the global coenzyme concentration is not the focus of study, hence systems models implicitly account for the effects of coenzyme concentration  $[\text{CoE}]$  by asserting that  $\bar{k}^c = k^c[\text{CoE}]$  and then directly varying the catalytic rate  $\bar{k}^c$  to vary energy input rate. This allows the thermodynamically closed enzyme kinetic model to be used in a thermodynamically open context [102].

To address these issues, we have developed a generalised enzyme kinetic model that retains its mathematical form for systems with multiple enzymes, whilst minimising the number of sim-

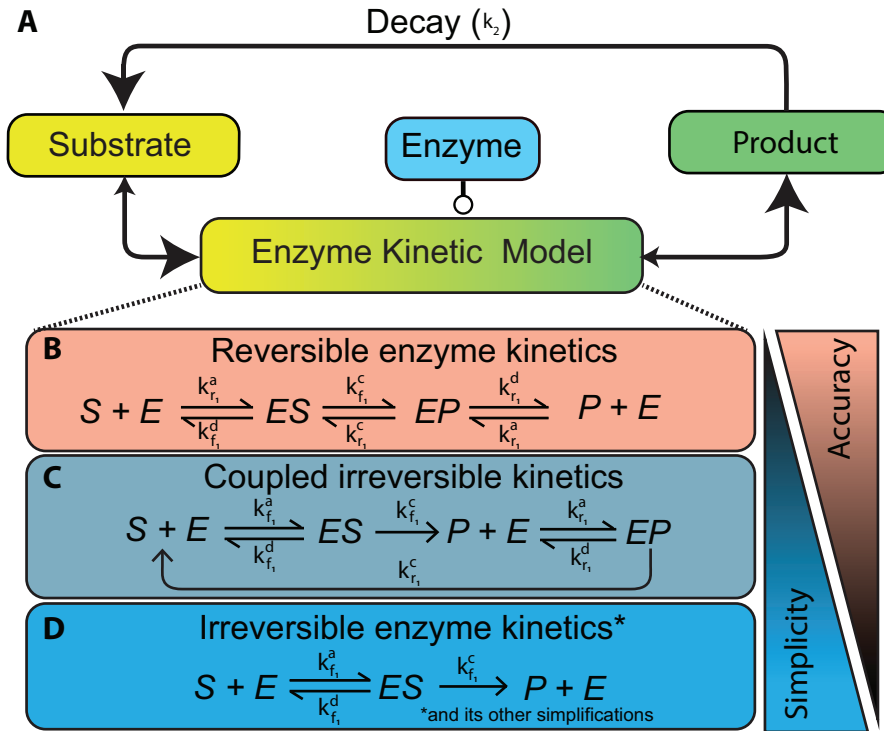


Figure 16: Different models of single substrate enzyme kinetics. (A) An illustration of the cyclic enzyme kinetic reaction that will be the focus of this chapter. An enzyme reaction converts substrate to product which is reversed by a spontaneous reaction. (B) Chemical equation of the fully reversible reaction mechanism. (C) Chemical equation of the coupled irreversible reaction mechanism. (D) Chemical equation of the irreversible reaction mechanism.

plifying assumptions and parameters needed to characterise the system. This enables more accurate simulation of the biochemical mechanisms involved. In this chapter, we perform the theoretical analysis for developing this improved model. Later in chapter 4, the model is numerically tested for accuracy both as an isolated reaction and within a more complex network.

### 3.3 SIMPLIFYING ENZYME KINETICS

Since enzymes form the basis of many biochemical processes, enzyme kinetic models similarly form the basis of mathematical representations of biochemical networks. The difficulty in imple-

menting enzyme kinetic models lies in its variety in the literature. In this section, we demonstrate and compare the limits of various simplifications of enzyme kinetic models, beginning with the more complex mass action based reversible enzyme kinetic model in a reaction cycle. The reaction cycle will involve an enzymatic reaction that favours substrate to product formation, coupled with an irreversible decay reaction of the product back to substrate as illustrated in Fig. 16A. Note that in enzyme kinetic models, it is generally assumed that the stoichiometric constants are unity as it is expected that one molecular of enzyme will react with one molecular of substrate.

While the overall goal of this work is to achieve an optimally accurate dynamic model, we will instead focus on the steady state product enzyme catalysis speed  $v_\infty$  to simplify the comparison. Where:

$$v_\infty = k_{f_1}^c [ES] - k_{r_1}^c [EP] \quad (7)$$

This is because the homeostatic enzymatic reaction speed is easier to compare analytically than the full temporal behaviour, yet an incorrect homeostatic equilibrium state implies the kinetic behaviour of the model is incorrect. Also, most enzyme kinetic models have the form [159]:

$$v_\infty = [E_T] \frac{k_{f_1}^c [S]/K_1 - k_{r_1}^c [P]/K_2}{1 + [S]/K_3 + [P]/K_4} \quad (8)$$

Expressing each model in this form allows us to compare the form of the coefficients  $K_1$ ,  $K_2$ ,  $K_3$  and  $K_4$  which allows us to understand the simplifying assumptions between successive models and thus the scenarios where the various models are valid.

### 3.3.1 Reversible enzyme kinetic model in a cyclic reaction system

For the reversible cyclic reaction, described by the chemical equation shown in Fig. 16B, the mass action model can be expressed minimally as [159, 164]:



$$\dot{[E]} = k_{f_1}^a [S][E] + k_{r_1}^c [EP] - (k_{f_1}^d + k_{f_1}^c)[ES] \quad (9)$$

$$\dot{[EP]} = k_{r_1}^a [P][E] + k_{f_1}^c [ES] - (k_{r_1}^d + k_{r_1}^c)[EP] \quad (10)$$

$$\dot{[P]} = -k_{r_1}^a [P][E] + k_{r_1}^d [EP] - k_2[P] \quad (11)$$

$$[S_T]_0 = [S] + [P] + [ES] + [EP] \quad (12)$$

$$[E_T]_0 = [E] + [ES] + [EP] \quad (13)$$

Setting all derivatives to zero, solving Eqn. 9 and 10 simultaneously, and then substituting the result into Eqn. 7, leads to the following expression for the catalysis speed at complete steady state:

$$\begin{aligned} v_\infty &= k_{f_1}^c [ES] - k_{r_1}^c [EP] \\ &= [E_T]_\infty \frac{k_f^c [S]_\infty / A_{f_1 r_1} - k_r^c [P]_\infty / A_{r_1 f_1}}{1 + [S]_\infty / M_{f_1 r_1} + [P]_\infty / M_{r_1 f_1}} \end{aligned} \quad (14)$$

where

$$\begin{aligned} A_{ij} &= K_i^m \frac{1 - \alpha_i \alpha_j}{1 - \alpha_j} \\ M_{ij} &= K_i^m \frac{1 - \alpha_i \alpha_j}{1 + \alpha_j \frac{k_i^c}{k_j^c}} \\ \alpha_i &= \frac{k_i^c}{k_i^c + k_i^d} \\ K_i^m &= \frac{k_i^c + k_i^d}{k_i^a} \end{aligned} \quad (15)$$

The subscript  $i$  in  $A$ ,  $\alpha$  and  $K^m$  determines whether the forward and reverse reactions are referred to when  $i$  is  $f$  or  $r$  respectively. Note that this model creates six unique coefficients within the enzyme kinetic model ( $k_{f_1}^c$ ,  $k_{r_1}^c$ ,  $A_{f_1 r_1}$ ,  $A_{r_1 f_1}$ ,  $M_{f_1 r_1}$  and,  $M_{r_1 f_1}$ ); therefore all six kinetic parameters of the enzymatic reaction are needed to fully define the coefficients and the reversible enzyme kinetic model. Thus for the reversible enzyme kinetic model, a quasi-steady state approximation does not simplify that model when compared to the full mass action model of the full set of reactions.

### 3.3.2 Uncoupled Reversible enzyme kinetic models in cyclic reactions

The coupled irreversible cyclic reaction described by the chemical equation shown in Fig. 16C, is similar to the full reversible system with a subtle difference. The catalysis step of both the forward and reverse reactions immediately dissociate into their respective product species and free enzyme. The purpose of making this subtle change is to make the reversible reaction appear like two coupled irreversible enzyme reactions, which can be expressed minimally as:

$$[E\dot{S}] = k_{f_1}^a [S][E] - (k_{f_1}^d + k_{f_1}^c)[ES] \quad (16)$$

$$[E\dot{P}] = k_{r_1}^a [P][E] - (k_{r_1}^d + k_{r_1}^c)[EP] \quad (17)$$

$$[\dot{P}] = -k_{r_1}^a [P][E] + k_{f_1}^c [ES] + k_{r_1}^d [EP] - k_2[P] \quad (18)$$

$$[S_T]_0 = [S] + [P] + [ES] + [EP] \quad (19)$$

$$[E_T]_0 = [E] + [ES] + [EP] \quad (20)$$

Again, setting all derivatives to zero and substituting Eqn. 16 and 17 into Eqn. 7, leads to the following expression for the catalysis rate at steady state:

$$v_\infty = \frac{k_{f_1}^c \frac{[E_T]_\infty [S]_\infty}{K_{f_1}^m} - k_{r_1}^c \frac{[E_T]_\infty [P]_\infty}{K_{r_1}^m}}{1 + \frac{[S]_\infty}{K_{f_1}^m} + \frac{[P]_\infty}{K_{r_1}^m}} \quad (21)$$

This (Eqn. 21) is similar to the analogous expression for the fully reversible system (Eqn. 14) under the conditions  $k^d \gg k^c$  which causes  $\alpha_f, \alpha_r \approx 0$ . This then leads to  $M_{f_1 r_1} \approx A_{f_1 r_1} \approx K_f^m$  and  $M_{r_1 f_1} \approx A_{r_1 f_1} \approx K_r^m$  which is required for Eqn. 14 to become identical to Eqn. 21. This implies that under the "rapid equilibrium" condition (where free enzymes and substrates rapidly equilibrate with their complexed forms) the reversible enzyme kinetic system can be modelled as two opposing irreversible enzyme kinetic systems without losing mechanistic detail. This assumption is identical to the original assumption made by Michaelis and Menten where  $k^d \gg k^c$  and thus  $K^m \approx K^i$  where  $K^i$  is the dissociation constant between enzyme, substrate and the enzyme-substrate complex [131].

### 3.3.3 Irreversible enzyme kinetic models in cyclic reactions

For the reaction to be modelled as completely irreversible, which can be described by the chemical equation shown in Fig. 16D, we simply set  $k_{r1}^c = 0$ . Dropping the f and r subscripts from parameters associated with reaction 1 (since all reverse terms are now dropped, removing any ambiguity that script 1 parameters refer to the forward direction of the reaction) this leads to the following result for the catalysis speed at complete steady state:

$$v_{\infty} = \frac{k_1^c [E_T]_{\infty} [S]_{\infty}}{K_1^m + [S]_{\infty}} \quad (22)$$

This model is now less realistic as enzymes do not behave irreversibly under physiological conditions. This is because enzyme simply speed up reactions and cannot change the equilibrium state as this is controlled by the free energy of the reaction [18]. However, by definition an irreversible enzyme would necessary push the system towards complete consumption of substrate, regardless of the actual equilibrium point. Thus this approximation is only suitable when the equilibrium point of the reaction is known to be complete consumption of the product. A significant consequence is that this model no longer includes product inhibition since no enzyme-product complex exists. This is an important result that will be revisited in a later section.

### 3.3.4 The Michaelis-Menten Model in cyclic reactions

Whilst the Michaelis-Menten (and Briggs-Haldane) model is a subset of the irreversible enzyme kinetic model, it is worthy of a standalone analysis as it is widely used in models of biochemical systems. Since the topology of reaction networks can itself be complicated, this model is often used to simplify the mathematics of the network model [19, 103, 124, 126, 129, 130]. However, the the Michaelis-Menten model may not be applicable under physiological conditions, such as for protein phosphorylation networks as various underlying assumptions of the model may be violated.

The Michaelis-Menten model adopts the quasi-steady state assumption, where the enzyme-substrate complex [ES] rapidly equi-

librates over small time scales ( $\dot{[ES]} = 0$  in Eqn. 16). This leads to the following expression for the enzyme-substrate complex:

$$[ES] = \frac{[S][E]}{K_1^m} \quad (23)$$

The state variable used is total enzyme concentration rather than free enzyme using the following substitution [125]:

$$[E_T] = [E] + [ES] \quad (24)$$

Combining Eqns. 7, 23 and 24 (noting that  $[EP] = 0$  in this case) leads to the following expression for the enzyme reaction rate  $v$ :

$$v = k_1^c \frac{[E_T][S]}{K_1^m + [S]} \quad (25)$$

This is very similar to the completely irreversible steady state enzyme turnover rate (Eqn. 22). So in principle the Michaelis-Menten model is consistent with the mass action based irreversible enzyme kinetics model. In practice, this is not the case because it is commonly assumed that:

$$[S] \approx [S_T] \quad (26)$$

This implies that the ES complex must have a negligible concentration:

$$[ES] \approx 0 \quad (27)$$

This is known as the reactant stationary assumption and is the origin of the low enzyme assumption implicit in the Michaelis-Menten model [103, 132, 161, 165, 166]. The condition required for this assumption to hold true is given by [132, 136]:

$$[E_T] \ll K_1^m + [S_T]_0 \quad (28)$$

This does not strictly require that  $[E_T] \ll [S_T]_0$  but in the situation where  $[E_T] \geq [S_T]_0$ , a low enzyme-substrate affinity associated with a large  $K_1^m$  is instead required for condition Eqn. 28 to hold [132, 166, 167].

The tQSSA proposed by Tzafriri avoids this limitation by not applying the assumption expressed in Eqn. 26 implicitly accounting for the enzyme-complex concentration by re-expressing the

series of equations in terms of the total substrate concentration  $[S_T]$  [136, 167, 168]. The resulting steady state concentration is:

$$[P]_{\infty} = \frac{k_1^c}{2k_2} \left( B - \sqrt{B^2 - 4[E_T][S_T]} \right) \quad (29)$$

$$B = K_1^m + [E_T] + [S_T] \quad (30)$$

There is some controversy surrounding the tQSSA model as it is argued by Lim that it is merely a restatement of the Michaelis-Menten model but with different state variables [137]. While it is true that the tQSSA is derived simply by making a change in state variable, it has a profound impact on the accuracy of the rate equation. This is due to a difference in their treatment of the reactant stationary assumption. The practical difference between the two models can be seen by the differences in the predicted concentrations as a function of time [136].

In spite of the increased accuracy of the tQSSA compared to the full mass action model of enzyme action, there is an additional yet common biological scenario in which it too can become inaccurate. In a system where an enzyme has multiple targets (e.g. Akt), the conservation law Eqn. 24 is no longer valid as there are additional enzyme-substrate complexes that contribute to the total pool of enzymes [169]. This causes the Michaelis-Menten and tQSSA models to break down as they both rely on this conservation law in their derivation. Hence, the rate equations for these quasi-steady state models would need to be rederived for each distinct biochemical network. As shown in other studies involving analysis using the tQSSA, this can become impractical as the networks become complex enough that the ensuing simultaneous equations become analytically intractable [139].

### 3.4 DEVELOPMENT OF THE DIFFERENTIAL QUASI-STEADY STATE MODEL

We have established that reversible enzyme kinetics can be approximated as two opposing irreversible enzyme reactions under rapid equilibrium conditions. We also showed that the Michaelis-Menten model is an approximation of the mass action irreversible enzyme reaction under quasi-steady state and reactant stationary assumptions. However, this model is often not appropriate in an

in vivo setting because the reactant stationary assumption is often violated in a physiological setting. As such to develop a physiologically relevant quasi-steady state enzyme kinetic model, it is necessary to overcome the reactant stationary assumption.

As we have outlined in the previous section, the Michaelis-Menten reactant stationary assumption assumes that the complex concentration is negligible, and thus does not include it as a state variable. The tQSSA model does overcome the reactant stationary model, but still does not explicitly model the complex concentration because it uses the total substrate concentration instead of the free substrate concentration as the state variable. Both of these methods of implicit modelling of the complex concentration leads to limitations when applied to more complex biological systems. Thus, if the quasi-steady state complex concentration can be explicitly modelled, this would overcome the reactant stationary assumption of the Michaelis-Menten model, and remove the need for rederivation in different networks as is the case for the tQSSA model. Thus the rapid equilibrium assumption would be the only remaining assumption in the model, thereby increasing its accuracy. We will show in this section that this can be achieved through a more in depth analysis of the quasi-steady state assumption.

In addition to this, we will show that our new model satisfies the following two challenges that need to be overcome for an enzyme kinetic model to be useful for general classes of biochemical networks:

1. Retain the accuracy and parameter reduction on par with the tQSSA irreversible enzyme kinetic model
2. General applicability so the model doesn't have to be reevaluated for novel networks.

To simplify the mathematics of this system, only the irreversible half reaction of the enzymatic reaction will be discussed in this section. However the reader should note that the derivation is designed to be general. That is, the reaction does not act in an isolate system and should be considered part of a wider network. More explicitly, this means that  $[E_T]$  is not constant because this enzyme may participate in another reaction and the definition  $[E_T]$  only includes the free substrate and the complex of the en-

zyme reaction in question. This will have other consequences that will be discussed when relevant.

### 3.4.1 The Quasi-Steady State

The conventional derivation of quasi-steady state enzyme kinetics is achieved by setting the rate of change of the complex concentration to zero and solving for the complex concentration. For the irreversible half reaction without any spontaneous decay from product back to substrate (i.e.  $k_2 = 0$  from the irreversible analysis in section 3.3.3, dropping the subscripts from the enzymatic reaction as all kinetic parameters now refer to the enzymatic reaction), this is:

$$[ES] = k^a[E][S] - (k^c + k^d)[ES] = 0 \quad (31)$$

$$[ES] = \frac{[E][S]}{K^m} \quad (32)$$

Where  $k^a$ ,  $k^d$  and  $k^c$  are as defined at the beginning of this chapter. Note that  $[ES]$  is assumed to only interact with this enzyme reaction in this analysis.

Setting  $[ES]$  to zero then implies the complex concentration is constant at all times. This has been shown to be false independently by Chen and Tzafriri: the complex concentration does change whilst the system is in quasi-steady state [103, 136]. This is shown in Fig. 17, where the complex concentration predicted by the mass action model, using the full enzyme kinetic system, is overlayed with the same prediction by the tQSSA model (this system does not contain other external reactions). The two models are in disagreement during the transient phase, where the complex concentration of the tQSSA model is constant while the mass action model shows an increasing complex concentration over time. Once the mass action model concentration reaches its peak, the two models begin to align as the model reaches the quasi-steady state. However, after this point of convergence, the two models then exhibits a reduction in the complex concentration. This demonstrates two things. Firstly, the tQSSA and mass action models are in agreement after the initial transient phase. Secondly, the complex concentration can change after during the quasi-static phase. This is seemingly contradictory with the quasi-steady state assumption expressed in Eqn. 31.

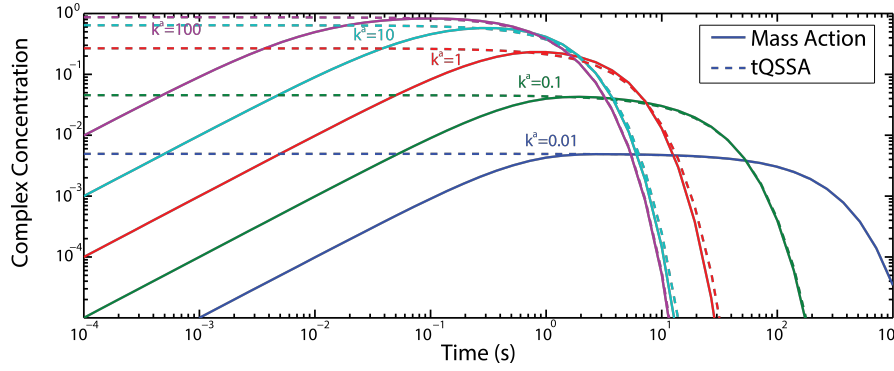


Figure 17: Line graphs showing the complex concentrations simulated by the mass action model (solid lines) and tQSSA model (dashed lines) at five different values of the association constant  $k^a$ . The parameters  $k^d$  and  $k^c$  and concentrations  $[S]_0$  and  $[E]_0$  are set to unity for illustrative purposes. The tQSSA models do not agree with the mass action model at the beginning of the simulation, but quickly align. The complex concentration of both models later decrease together, which demonstrates that both the mass action model and tQSSA predict a change in the complex concentration, which would appear to be a violation of Eqn. 31

To resolve this contradiction, the definition of the quasi-steady state must be carefully considered. To start, any quasi-steady state analysis assumes that the system can be divided into two time scales, a high speed transient time scale, and a slower quasi-static time scale. It is useful to recall that this is the conditions  $k^c \ll k^d$ , which is required for the full mass action enzyme kinetic model to be approximated as two irreversible enzyme kinetic half reactions. This condition implies that any reactions that operate in the transient time scale (specifically the complex formation steps) occurs fast enough that it can always reequilibrates before any reaction in the slow quasi-static time scale (more specifically the catalytic step) has proceeded to any appreciable degree. This distinction is important as it leads to a number of differences in setting up the derivation of the quasi-steady state assumption.



To begin let us examine the equations governing the system at the transient state. The minimal description of the system in this regime is:

$$\dot{[ES]} = k^a([E_T] - [ES])([S_T] - [ES]) - (k^d + k^c)[ES] \quad (33)$$

$$[E_T]_0 + \Delta = [E_T] = [ES] + [E] \quad (34)$$

$$[S_T]_0 + \Delta = [S_T] + [P] = [ES] + [S] + [P] \quad (35)$$

Where  $\Delta$  are other sources/sinks for the substrate and enzyme external to the enzymatic reaction (such as other complexes) that is not important to this analysis. In the quasi-steady state approximation of an enzyme kinetic reaction, the complex formation reactions are assumed to operate in the transient time scale while the catalytic conversion of substrate to product operates in the quasi-static time scale. If we consider the total concentration of substrate and enzyme (as opposed to distinguishing between free and complex bound forms), then  $[E_T]$  and  $[S_T]$  are approximately constant in the transient regime, and  $[P] = 0$ . With this setup, the dynamic system can be described only with one differential equation as the other conservation laws are trivial.

Traditionally, the complex concentration under the quasi-steady state approximation can be found by setting  $\dot{[ES]} = 0$ . This was found by Tzafiriri to be [136]:

$$[ES] = \frac{[E_T] + [S_T] + K^m - \sqrt{([E_T] + [S_T] + K^m)^2 - 4[E_T][S_T]}}{2} \quad (36)$$

Now let us consider what  $\dot{[ES]} = 0$  actually means.  $\dot{[ES]}$  is a function of three dependent variables,  $[ES]$ ,  $[E_T]$  and  $[S_T]$ , all of which themselves are individually functions of time. Thus:

$$\begin{aligned} \dot{[ES]} &= \dot{f}(t, [ES], [E_T], [S_T]) \\ &= k^a([E_T] - [ES])([S_T] - [ES]) - (k^d + k^c)[ES] \end{aligned} \quad (37)$$

The integral of this expression should yield:

$$[ES] = f(t, [E_T], [S_T]) \quad (38)$$

The derivative of this generic functional form using the chain rule yields:

$$\dot{[ES]} = \frac{\partial [ES]}{\partial t} + \frac{\partial [ES]}{\partial [E_T]} \dot{[E_T]} + \frac{\partial [ES]}{\partial [S_T]} \dot{[S_T]} \quad (39)$$

Now let us consider only the transient phase. We have previously asserted that in this regime,  $[E_T]$  and  $[S_T]$  are constant since they only operate in the quasi-static time scale. Thus,  $\dot{[E_T]} = 0$  and  $\dot{[S_T]} = 0$ . So in the transient regime,  $\dot{[ES]}$  is just the partial derivative of  $[ES]$  with time. Therefore in the transient regime, the following expressions are equivalent:

$$\frac{d[ES]}{dt} = \frac{\partial[ES]}{\partial t} = 0 \quad (40)$$

Thus solving this will reproduce the Tzafriri result [136]. However, with our reanalysis of the quasi-steady state assumption, the partial derivative and the full derivative are not equivalent. This is because the full derivative contains the  $\dot{[E_T]}$  and  $\dot{[S_T]}$  terms which are no longer zero. Now this is implicit in Tzafriri's result, because his expression for  $[ES]$  is a function of  $[S_T]$ , and  $[S_T]$  is a function of time.

It is useful to obtain an explicit solution for  $\dot{[ES]}$  because it allows us to explicitly model the kinetics of the complex state, rather than needing to infer it from  $[E_T]$  and  $[S_T]$ . Furthermore, it is more useful to use the free substrate and enzyme concentrations as the state variables as this simplifies the kinetic equation when multiple substrates are involved. Thus Tzafriri's analysis can be extended along these lines. The relevant transformations are:

$$[E] = [E_T] - [ES] \quad (41)$$

$$[S] = [S_T] - [ES] \quad (42)$$

Using this, the state variable of  $[ES]$  can be changed to:

$$[ES] = \frac{[E_T] + [S_T] + K^m - \sqrt{([E_T] + [S_T] + K^m)^2 - 4[E_T][S_T]}}{2} \quad (43)$$

$$= \frac{[E][S]}{K^m} \quad (44)$$

$\dot{[ES]}$  can then be found for the new variables by again using the chain rule, which gives:

$$\dot{[ES]} = \frac{\partial[ES]}{\partial t} + \frac{\partial[ES]}{\partial[E]} \dot{[E]} + \frac{\partial[ES]}{\partial[S]} \dot{[S]} \quad (45)$$

From the quasi-steady state assumption, we assert that the first term on the right hand side is zero. This produces:

$$[\dot{E}S] = \frac{\partial[ES]}{\partial[E]}[\dot{E}] + \frac{\partial[ES]}{\partial[S]}[\dot{S}] \quad (46)$$

The partial derivatives can be determined from Eqn. 44. Which gives:

$$\frac{\partial[ES]}{\partial E} = \frac{[S]}{K^m} \quad (47)$$

$$\frac{\partial[ES]}{\partial S} = \frac{[E]}{K^m} \quad (48)$$

Substituting this into the total time derivative for the complex Eqn. 46, and noting that the partial time derivative of  $[ES]$  is now zero, leads to:

$$[\dot{E}S] = \frac{[S]}{K^m}[\dot{E}] + \frac{[E]}{K^m}[\dot{S}] \quad (49)$$

With an explicit expression for  $[\dot{E}S]$ , it is now possible to write a new set of equations that describe the quasi-steady state reaction:

$$[\dot{E}S] = \frac{[S]}{K^m}[\dot{E}] + \frac{[E]}{K^m}[\dot{S}] \quad (50)$$

$$[\dot{P}] = k^c[ES] = \frac{k^c[S][E]}{K^m} \quad (51)$$

$$\begin{aligned} [\dot{S}] &= \frac{d([S_T]_0 - [ES] - [P])}{dt} = -[\dot{E}S] - \frac{k^c[S][E]}{K^m} \\ &= -\frac{[S]}{K^m}[\dot{E}] - \frac{[E]}{K^m}[\dot{S}] - \frac{k^c[S][E]}{K^m} \end{aligned} \quad (52)$$

$$\begin{aligned} [\dot{E}] &= \frac{d([E_T]_0 - [ES])}{dt} = -[\dot{E}S] \\ &= -\frac{[S]}{K^m}[\dot{E}] - \frac{[E]}{K^m}[\dot{S}] \end{aligned} \quad (53)$$

By accounting for the change in the complex concentration during the quasi-steady state, we no longer require the reactant stationary assumption and satisfy the first challenge. Additionally we have collapsed the model from three parameters ( $k^a$ ,  $k^d$  and  $k^c$ ) to two parameters ( $K^m$  and  $k^c$ ) However, the new dynamic equation cannot be solve directly as there are mixed time derivatives.

### 3.4.2 Decoupling rate equation by vectorisation

Although our new system of equations (Eqns. 50-53) appear to be able to resolve some of the challenges surrounding quasi-steady state expression of enzyme kinetic system. However, the expressions as it is currently presented is in implicit form and thus cannot easily be solved simply. More specifically solved directly using explicit techniques, which limits the numerical tools that can be used. In order to achieve an explicit form, we need to re-express the system of differential equations as a matrix equation with the state variables represented with the vector  $\mathbf{x}$  with the indices:

1. Substrate
2. Enzyme
3. Product
4. ES Complex

This enables the differential equation to be rearranged into an explicit form. The first step in achieving this is by isolating all derivatives from Eqns 50 - 53 to the left hand side. This leads to:

$$\left(1 + \frac{x_2}{K_m}\right) \dot{x}_1 + \frac{x_1}{K_m} \dot{x}_2 = -k^c x_4 \quad (54)$$

$$\frac{x_2}{K_m} \dot{x}_1 + \left(1 + \frac{x_1}{K_m}\right) \dot{x}_2 = 0 \quad (55)$$

$$\dot{x}_3 = k^c x_4 \quad (56)$$

$$-\frac{x_2}{K_m} \dot{x}_1 - \frac{x_1}{K_m} \dot{x}_2 + \dot{x}_4 = 0 \quad (57)$$

This can be written as the following matrix equation:

$$\begin{bmatrix} 1 + \frac{x_2}{K_m} & \frac{x_1}{K_m} & 0 & 0 \\ \frac{x_2}{K_m} & 1 + \frac{x_1}{K_m} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -\frac{x_1}{K_m} & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -k^c \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k^c \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad (58)$$

which can then be written in tensor form as:

$$(\delta_{jk} + \mathbf{G}_{ijk} x_k) \dot{x}_j = \mathbf{Y}_{il} x_l \quad (59)$$

where  $\delta$  is the Kronecker delta, and the tensors:

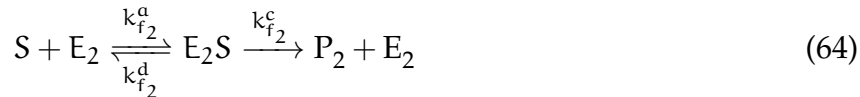
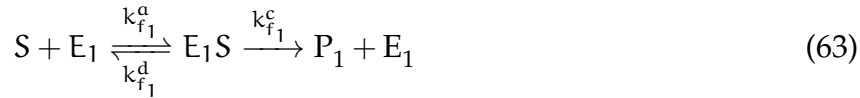
$$\mathbf{G}_{ijk} = \begin{cases} \frac{1}{K^m}, & \text{if } [i, j, k] = [1, 1, 2], [1, 2, 1], [2, 1, 2], [2, 2, 1] \\ -\frac{1}{K^m}, & \text{if } [i, j, k] = [4, 1, 2], [4, 2, 1] \\ 0, & \text{else} \end{cases} \quad (60)$$

$$\mathbf{Z}_{il} = \begin{cases} -k^c, & \text{if } [j, l] = [1, 4] \\ k^c, & \text{if } [j, l] = [3, 4] \\ 0, & \text{else} \end{cases} \quad (61)$$

The derivative terms can then be isolated from the non-derivative terms, leading to:

$$\dot{x}_j = (\delta_{jk} + \mathbf{G}_{ijk}x_k)^{-1} \mathbf{Z}_{il}x_l \quad (62)$$

This now leaves the equation as an explicit differential equation, with a matrix inversion term that can be determined using a variety of methods for solving a linear system of equations such as gaussian elimination. Additionally, the final form of this equation is also applicable to more complex networks of enzymatic reactions. Let us consider a system where two enzymes ( $E_1$  and  $E_2$ ) both act catalytically on one substrate with reaction equation shown below:



The system (without any reverse cyclic reactions) can be described as:

$$[S_T]_0 = [S] + [E_1S] + [E_2S] + [P_1] + [P_2] \quad (65)$$

$$[E_{1T}]_0 = [E_1] + [E_1S] \quad (66)$$

$$[\dot{P}_1] = k_1^c[E_1S] \quad (67)$$

$$[E_1\dot{S}] = k_1^a[S][E_1] - (k_1^d + k_1^c)[E_1S] \quad (68)$$

$$[E_{2T}]_0 = [E_2] + [E_2S] \quad (69)$$

$$[\dot{P}_2] = k_2^c[E_2S] \quad (70)$$

$$[E_2\dot{S}] = k_2^a[S][E_2] - (k_2^d + k_2^c)[E_2S] \quad (71)$$

Reaction rates are as defined at the beginning of this chapter. Using the indices  $x_1 = [S]$ ,  $x_2 = [E_1]$ ,  $x_3 = [P_1]$ ,  $x_4 = [E_1S]$ ,  $x_5 = [E_2]$ ,  $x_6 = [P_2]$  and  $x_7 = [E_2S]$ , and using the steps outlined in section 3.4 (shown in Appendix B.1), we generate the vectorised quasi-steady state equation with tensors:

$$\mathbf{G}_{ijk} = \begin{cases} \frac{1}{K_1^m}, & \text{if } [i, j, k] = [1, 1, 2], [1, 2, 1], [2, 1, 2], [2, 2, 1] \\ -\frac{1}{K_1^m}, & \text{if } [i, j, k] = [4, 1, 2], [4, 2, 1] \\ \frac{1}{K_2^m}, & \text{if } [i, j, k] = [1, 1, 5], [1, 5, 1], [5, 1, 5], [5, 5, 1] \\ -\frac{1}{K_2^m}, & \text{if } [i, j, k] = [7, 1, 5], [7, 5, 1] \\ 0, & \text{else} \end{cases} \quad (72)$$

$$\mathbf{Z}_{jl} = \begin{cases} -k_1^c, & \text{if } [i, l] = [1, 4] \\ k_1^c, & \text{if } [i, l] = [3, 4] \\ -k_2^c, & \text{if } [i, l] = [1, 7] \\ k_2^c, & \text{if } [i, l] = [6, 7] \\ 0, & \text{else} \end{cases} \quad (73)$$

It can be seen that the elements of the tensor in which  $K_1^m$  and  $k_1^c$  are inserted into are exactly the same as in the single enzyme case and do not conflict with the elements where the terms  $K_2^m$  and  $k_2^c$  are inserted into. We can surmise from this that the tensors  $\mathbf{G}$  and  $\mathbf{Z}$  for different reactions can be combined into their corresponding global version by summation. I.e.

$$\mathbf{G}_{ijk}^{(\text{Global})} = \sum_r^n \mathbf{G}_{ijk}^{(r)} \quad (74)$$

$$\mathbf{Z}_{il}^{(\text{Global})} = \sum_r^n \mathbf{Z}_{il}^{(r)} \quad (75)$$

Where  $n$  is the number of reactions within a network and  $r$  is the index of the reaction. Furthermore, the dynamic equation can be generalised to include other fundamental reactions and external inputs as arbitrary functions of time. These would appear as an additional sum on the right hand side of Eqn. 62. To show this,

consider Eqn. 65. If  $[S_T]_0$  is constant, but the total pool of substrate and its products contains external sources and sinks (as in Eqns 34 and 35), then the expression becomes:

$$\begin{aligned} [S_T]_0 + \int_0^t \sum_m x_m \sum_n k_{mn} x_n + \sum_p k_p x_p + k_s + f_s(t, x) dt \\ = [S] + [E_1 S] + [E_2 S] + [P_1] + [P_2] \end{aligned} \quad (76)$$

Where  $k_{lm}$  are second order rate parameters,  $k_n$  are first order,  $k_s$  is a zeroth order term and  $f(t, x)$  other arbitrary reaction velocities. If we follow through with the analysis we would find that:

$$\begin{aligned} [\dot{S}] + [E_1 \dot{S}] + [E_2 \dot{S}] = \left( \sum_m x_m \sum_n k_{mn} x_n + \sum_p k_p x_n + \right. \\ \left. k_s + f_s(t, x) \right) - [\dot{P}_1] - [\dot{P}_2] \end{aligned} \quad (77)$$

And similarly for the other state variables. What this shows is other elementary reactions when applied to the dQSSA simply appear as additional additive terms alongside  $Z_{il}x_l$ . Thus, the fully generalised equation, which includes these reactions becomes:

$$\dot{x}_j = (\delta_{jk} + \mathbf{G}_{ijk}x_k)^{-1} [\mathbf{Y}_{ilm}x_lx_m + \mathbf{Z}_{il}x_l + \sigma_i(t) + f_i(t, \vec{x})] \quad (78)$$

where  $\mathbf{Y}$  is a three dimensional tensor containing bimolecular rate constants,  $\mathbf{Z}$  is a two dimensional tensor containing unimolecular rate constants (which contains the catalytic constant for enzyme reactions),  $\sigma$  is a one dimensional tensor containing synthesis type rate parameters and  $f_i$  is a vector of arbitrary functions of external inputs or other kinetic models. We refer to Eqn. 78 as the dQSSA. By expressing the kinetic equation in this way, the model topology is now captured by the location and value of non-zero tensor elements.

### 3.4.3 Calculation of the Initial Condition

The dQSSA model in its current state simulates the beginning of the quasi-steady state to the full steady state. However, we have previously seen that the initial complex concentration during this state is not zero. Thus, we need to evaluate this before the simulation can begin. While there are many ways to achieve this, the

simplest way is to simulate the transient phase. Note that this transient phase is not mechanistically correct, but rather it is designed to obtain the initial condition for the quasi-steady state.

Noting that the rate parameters in the tensors  $Y$ ,  $Z$ ,  $\sigma$  and  $f$  are the non-quasi-static or "slow" reactions, these do not proceed during the transient phase. Thus, all of these components are set to zero during the transient phase. Next except with the reactants and enzymes injected into the system and allowed to equilibrate. This is done by providing a velocity profile for these states via the  $f_i(t, \vec{x})$  vector whose area under curve is equal to the initial concentrations prior to the transient phase. This is the only term within the square bracket that will be non-zero. Solving this differential equation will allow the correct initial quasi-steady state condition to be established for simulation because the differential equation should rapidly equilibrate the inject species into their quasi-steady state configuration. This approach will be verified in the next chapter.

### 3.5 ORDERED-SEQUENTIAL BI-BI REACTIONS

While most enzyme kinetic models are single substrate and single enzyme, enzymatic reactions in physiological scenarios are two substrate and single enzyme. This is because most reactions, such as phosphorylation, requires some cofactor, such as ATP to proceed [102]. Thus it may be necessary to model the effect of coenzyme depletion. In this case, modelling of the coenzyme concentration as a modification of the catalytic rate is not sufficient [102]. One mechanism by which this can happen is the ordered-sequential bi-bi shown in the following reaction equations [170]:



Again the rate parameters are as described in the beginning of the chapter.  $O$  is the substrate coenzyme (e.g. ATP) while  $\bar{O}$  is the



product coenzyme (e.g. ADP). The ODEs describing the unidirectional enzyme kinetic reaction is:

$$[\dot{E}] = -k^{a1} [E][O] + k_1^d [EO] + k^c [EOS] \quad (82)$$

$$[\dot{O}] = -k^{a1} [E][O] + k_1^d [EO] \quad (83)$$

$$[\dot{EO}] = k^{a1} [E][O] - k_1^d [EO] - k^{a2} [EO][S] + k^{d2} [EOS] \quad (84)$$

$$[\dot{S}] = -k^{a2} [EO][S] + k^{d2} [EOS] \quad (85)$$

$$[\dot{EOS}] = k^{a2} [EO][S] - k^{d2} [EOS] - k^c [EOS] \quad (86)$$

The intermediate complexes in this reaction are  $[EO]$ , the enzyme-coenzyme complex, and  $[EOS]$  the full enzyme-coenzyme-substrate complex (which is the complex that continues to the catalytic step). Using the same quasi-steady state simplification as the single substrate case in section 3.4, the quasi-steady state rate equations of the complexes are:

$$[\dot{EO}]_Q = K^2 \frac{[\dot{O}][E] + [\dot{E}][O]}{K^1 K^2 + \gamma[S]} + \gamma K^2 \frac{[\dot{S}][O][E]}{(K^1 K^2 + \gamma[S])^2} \quad (87)$$

$$[\dot{EOS}]_Q = \frac{[\dot{O}][E][S] + [\dot{E}][O][S] + [\dot{S}][E][O]}{K^1 K^2 + \gamma[S]} + \frac{\gamma[E][O][S][\dot{S}]}{(K^1 K^2 + \gamma[S])^2} \quad (88)$$

where

$$K^1 = k^{d1}/k^{a1} \quad (89)$$

$$K^2 = k^{d2}/k^{a2} \quad (90)$$

$$\gamma = k^c/k^{a1} \quad (91)$$

This is a far more complex expression compared with Eqn. 49, especially with the substrate concentration appearing in the denominator. However, it is still able to achieve a parameter reduction of two (as two states are set to quasi-steady state). It is possible to

further simplify the solution to appears in a similar form as Eqn 49 when  $\gamma \approx 0$ . Under this condition, Eqns. 87 and 88 simplify to:

$$[E\dot{O}] = [\dot{O}] \frac{[E]}{K^1} + [\dot{E}] \frac{[O]}{K^1} \quad (92)$$

$$\begin{aligned} [\dot{C}] &= \frac{[S]}{K^2} \frac{[\dot{O}][E] + [\dot{E}][O]}{K^1} + \frac{[\dot{S}]}{K^2} \frac{[E][O]}{K^1} \\ &= \frac{[S]}{K^2} [E\dot{O}] + \frac{[EO]}{K^2} [\dot{S}] \end{aligned} \quad (93)$$

This expression looks exactly the same as Eqn. 49, applied sequentially for each substrate. This means under these assumptions, even a two substrate enzyme kinetic reaction can be modelled using the dQSSA. By using this approximation, the model is simplified down from five parameters ( $k^{a1}$ ,  $k^{d1}$ ,  $k^{a2}$ ,  $k^{d2}$  and  $k^c$ ) to three parameters ( $K^1$ ,  $K^2$  and  $k^c$ ). Thus the tensor implementation for a two substrate model becomes:

$$\mathbf{G}_{ijk} = \begin{cases} \frac{1}{K^1}, & \text{if } [i, j, k] = [1, 1, 3], [1, 3, 1], [3, 1, 3], [3, 3, 1] \\ -\frac{1}{K^1}, & \text{if } [i, j, k] = [6, 1, 3], [6, 3, 1] \\ \frac{1}{K^2}, & \text{if } [i, j, k] = [2, 6, 2], [2, 2, 6], [6, 6, 2], [6, 2, 6] \\ -\frac{1}{K^2}, & \text{if } [i, j, k] = [7, 6, 2], [7, 2, 6] \\ 0, & \text{else} \end{cases} \quad (94)$$

$$\mathbf{W}_{il} = \begin{cases} -k^c, & \text{if } [j, l] = [1, 7] \\ -k^c, & \text{if } [j, l] = [2, 7] \\ k^c, & \text{if } [j, l] = [4, 7] \\ k^c, & \text{if } [j, l] = [5, 7] \\ 0, & \text{else} \end{cases} \quad (95)$$

Where the indices are: With the indices:

1. Coenzyme-1
2. Substrate
3. Enzyme
4. Coenzyme-2

5. Product
6. Enzyme-Coenzyme-1 complex
7. Enzyme-Coenzyme-1-Substrate complex

Biochemically, the  $\gamma \approx 0$  assumption states that  $k^c \ll k^{a_1}$ , that is the association of the coenzyme with the enzyme must occur very quickly compared to the catalytic rate. If this condition does not apply, then the quasi-steady state assumption breaks down because the rate limiting step becomes the coenzyme-enzyme binding step, rather than the catalytic step. It is expected that this condition should hold in *in vivo* because the coenzymes in this context are usually small molecules such as ATP or NADH, which are in large abundance, which means they should not be rate limiting. In spite of this, there are potentially scenarios where coenzymes are large molecules, such as enzyme complexes made of multiple proteins (such as the mTOR complexes). Under these conditions, the two substrate dQSSA must be applied with great care to ensure the complex formation steps are not rate limiting.

### 3.6 DISCUSSION

The motivation behind this study was to resolve the conflicting goals of model simplicity (both mathematical and dimensional) and accuracy when choosing enzyme kinetic models. We demonstrated that the fully reversible enzyme kinetic model, with the quasi-steady state assumption, requires all six rate parameters to fully define. Thus this model is not simpler than the full mass action description of enzyme kinetics. The simplest model, the Michaelis-Menten model, only requires two parameters. Each step of simplification requires increasingly restrictive assumptions, some of which are not necessarily true under physiological conditions. The reactant stationary assumption in the Michaelis-Menten model is an example of this.

The result of this study has shown that it is potentially possible to resolve this conflict between model accuracy and complexity. The dQSSA model is theoretically able to achieve the accuracy of the reversible model with the parameter reduction of the Michaelis-Menten and tQSSA models. This is because the the dQSSA

model only requires two assumptions. These are that the enzymatic reaction in question is approximately unidirectional (thus not requiring a fully reversible enzyme kinetic model), and that the dissociation rate constant is much larger than the catalytic rate constant. The MichaelisMenten model requires the additional reactant stationary assumption which assumes that the enzyme-substrate complex concentration is negligible. In addition to this, the dQSSA can be expressed with a universal form that does not require reanalysis, allowing it to be applied for a varied of biochemical system.

Furthermore, the reduction in parameter should improve the uniqueness of fitted models that utilise the dQSSA, while its predicted mechanistic accuracy improves predictive reliability of the model. This is especially important in large networks which have large parameter dimensionalities and limited kinetic data [145, 147, 148, 171]. Additionally, the standardised form of the dQSSA enables this model to be scaled up to be scaled up to larger biochemical systems which incorporate more pathways and molecular species. This ability will be important in ensuring modelling capabilities can keep pace with the increasing complexity of signalling pathways they are designed to study.

These two features of the dQSSA mean that it could have useful applications in the field of systems biology. It offers a more predictive and accurate modelling tool that can be applied with minimal reanalysis, and minimises the degrees of freedom of its models [102]. This enables model analysis to be focused on the structure rather than the parameter and the choice of kinetic model.

One significant weakness of the basic dQSSA model is its focus on single substrate reactions. While single substrate Michaelis-Menten models (and mass action type models) are the more commonly used enzyme kinetic models in biochemical systems, kinase reactions typically involved two substrates: that is a target substrate and ATP [102, 103]. To be able to comprehensively simulate signalling pathways, it is essential to be able to model at least two substrate reactions. This will cover a large group of post translational regulations that are relevant to biochemical networks [18]. The extended dQSSA model covering the ordered bi-bi reaction is a step towards this. While we have shown a universal quasi-steady state form for this reaction, it is only one of the three (and in fact the simplest) mechanisms that two-substrate reactions oc-

cur with. It is not unreasonable to expect that the dQSSA forms of the random-ordered bi-bi reaction and bi-bi ping-pong mechanisms will be more complicated even than the one presented in section 3.5. However, the ordered bi-bi mechanism we have considered will cover the kinase reactions that are of interest in the signalling networks we will study in later sections.

In the next chapter, we will take these theoretical models and verify their accuracy *in silico* and *in vitro*.



## VALIDATION OF NEW ENZYME KINETIC MODEL

---

### Abstract

While enzyme kinetic models can sometimes be considered empirical models, the most powerful kinetic models are those that are mechanistic and enable prediction of the underlying chemistry of the system. In order to validate the predictive nature of the model, they should ideally be verified experimentally. We have made two claims about our dQSSA model. Firstly that it is applicable for large networks using the universal form, and secondly that it accurately models real enzymatic reactions. In this chapter, we test these claims *in silico* and *in vitro*. We test the first claim by applying the dQSSA to a complex hypothetical network and comparing its predicted time course to that achieved by the mass action and Michaelis-Menten models. Next we tested the second claim by testing the ability of the dQSSA to predict the kinetics of the interconversion of pyruvate and lactate by lactate dehydrogenase (LDH) in the presence of different concentrations of coenzyme and enzyme. We found the dQSSA can accurately model large networks in its native mathematical form, and both the single substrate and two substrate dQSSA can predict the kinetics of LDH.

Some of this work has been published in:

- Wong MK, Krycer JR, Burchfield JG, James DE, Kuncic Z. A generalised enzyme kinetic model for predicting the behaviour of complex biochemical systems. FEBS Open Bio 2015; 9(5):226-239 doi: 10.1016/j.fob.2015.03.002.

## 4.1 INTRODUCTION

Kinetic models of biochemical reactions are used with the view of accurately predicting the behaviour of biochemical systems and explaining the behaviour through the mechanisms of signal transduction within the system [172]. However, the ability to do this is highly dependent on the accuracy of the rate laws used to describe enzymatic reactions in the model. The use of rate laws outside of their scenarios where they are valid can lead to incorrect predictions of the mechanisms relevant in the system [139, 173].

In order to avoid these follow on effects, it is imperative that newly designed rate laws be rigorously tested to verify the areas in which the model is accurate, and also determine its behaviour when applied to more complex networks that are more typical of the biological systems that are studied. Thus in this chapter, we apply the dQSSA model developed in the last chapter to *in silico* and *in vitro* validation tests to evaluate its accuracy and performance compared to the standard Michaelis Menten and mass action models.

## 4.2 METHOD

### 4.2.1 In Silico Validation in a hypothetical complex network

In order to validate the dQSSA rate law, it is necessary to test it both in a simple reaction as well as in the context of a larger complex network with complex reactions. In order to perform both functions as well, a single artificial network was designed which would initially test the performance of the dQSSA as an isolated enzymatic reaction but can be altered into a complex network through an activating switch built into the network. The network that will be used is shown in Fig. 18.

The initial test of the dQSSA in a simple network is performed when the switch I is absent from the network. In this scenario, which will be referred to later as the "I-off" network, the only active reactions are a unimolecular, a bimolecular and a reversible enzyme reaction. This allows us to isolate the performance of the dQSSA within a controlled simple enzymatic system. A separate



arm also tests the generalised dQSSA model's ability to model fundamental reactions.

After the basic tests of the dQSSA model as completed, activator I is added to activate more complex reactions in the system. This scenario, which will be referred to later as the "I-on" network, will assess the ability of the dQSSA to handle networks with complex coupling between reactions. More specifically the activator I will create species that are enzymes to multiple reactions and cause some substrates to become targets of multiple enzymes. To summarise, the network will contain:

- Unidirectional reactions (reaction 2, 4, 6, 8, 11)
- Bimolecular binding reactions (reaction 3)
- Enzyme reactions (reaction 1, 5, 7, 9, 10)
- A single substrate enzyme (species I)
- A multiple substrate enzyme (species p1B, pCD)
- A single enzyme substrate (species CD)
- A multiple enzyme substrate (species B)
- A substrate that is itself an enzyme for a different species (species A).

All species in the network start at a concentration of zero except the species A, B, C and D. I are added to the system with an input velocity profile that follows a gaussian function with peak at  $t_{1/2}$  and width  $t_w$ :

$$I(t) = I_0 \frac{1}{0.2\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \frac{(t - t_{1/2})^2}{t_w^2} \right] \quad (96)$$

The Gaussian function was chosen because it is a smooth function which will reduce the likelihood of residual errors during the numerical integration process. For species A, B, C and D, the following injection profile (given by  $f(t)$  in Eqn. 78) was used:

$$f(t, x) = x_0 \frac{1}{0.2\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \frac{(t)^2}{0.5^2} \right] \quad (97)$$

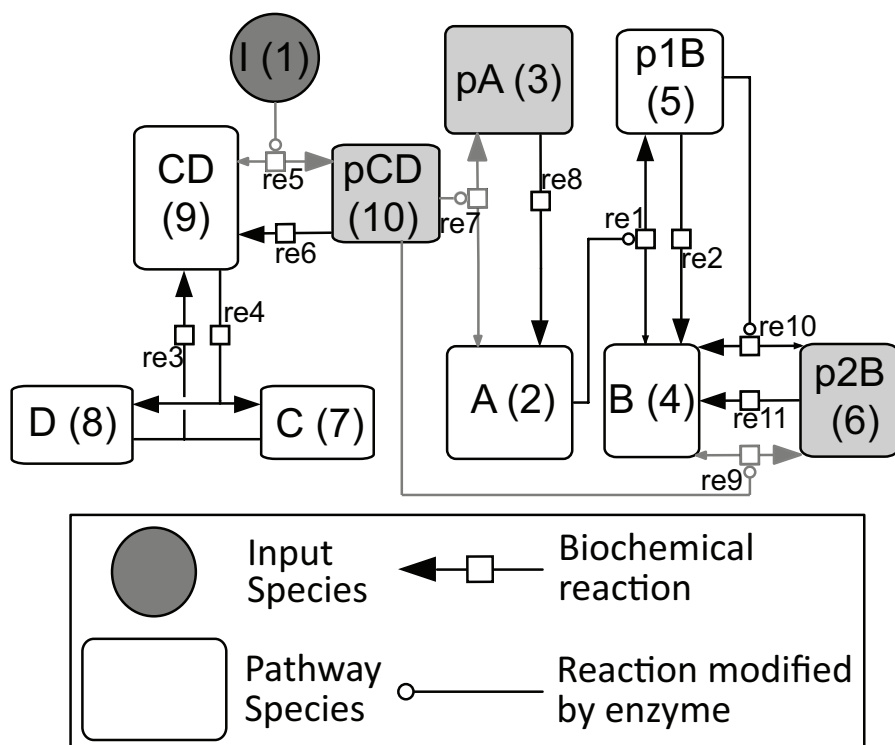


Figure 18: Network topology of the hypothetical network used to test the accuracy of the dQSSA model compared to the mass action model of enzyme kinetics. Numbers inside species refers to their index in the model equations. Re refers to reactions and their numbers as they are discussed in the text. The network is designed to operate in two modes. In the first mode only reactions with black lines are active. This is because enzymes required for those reactions (more specifically species with grey backgrounds) are unavailable then. In the second mode, all reactions become active. This second mode is triggered by the present of the species I.

Where  $x_0$  is the initial concentrations of A, B, C and D. The parameters of the Gaussian function was chosen such that the injection will be definition be complete within 0.5 seconds, with 0.5 additional seconds for the system to stabilise. Thus the initial condition is solved as described in section 3.4.3 over a time interval of  $t = [0, 1]$  in order for the simulation initial conditions to be solved and stabilised.

The network was constructed using three different rate laws: a mass action based model, a Michaelis-Menten based model and a dQSSA based model. The mass action model was constructed using mass action kinetics for fundamental reactions, and the full mass action form of enzyme kinetic reactions given by Eqn. 9-11 for enzymatic reactions. Each enzymatic reaction contains 6 rate parameters. All other reactions contain one. As such, this model required 41 parameters in total: 36 rate parameters and 5 concentration parameters. Equations for this model can be found in the computer code in section C.1.1 in the appendix.

For the dQSSA and Michaelis-Menten models, the 6 enzyme reaction parameters are simplified into 4 as per the quasi-steady state assumption. This resulted in 31 parameters: 26 rate parameters and 5 concentration parameters. The dQSSA model was constructed as given in Eqn. 78. The tensors were assigned as per the rules given in Eqn. 60-61. The 41 parameters for the mass action model were generated using Eqn. 98. The probability generating function was chosen such that their logarithm is uniformly distributed between  $[-1, 1]$ .

$$\log(\phi) = \text{rand}[-1, 1] \quad (98)$$

Where  $\phi$  represents all parameters in the model. These parameter were then modified in order to satisfy the rapid equilibrium conditions. The modifications applied are:

- The two generated catalytic rate of enzyme reactions are assigned such that the favoured direction has the larger catalytic rate.
- The dissociation rate (of the forward and reverse reaction) is made 100 times larger than the catalytic rate of the favoured direction
- All association rates are made 100 times larger in order to reduce the size of Michaelis constants.

This is necessary because in the previous chapter it was demonstrated that the quasi-steady reversible enzyme kinetic model is only valid under the rapid equilibrium condition. Because the dQSSA and Michaelis-Menten models are derived under this assumption, these models will both become inaccurate if the condition is violated. However, in such scenarios, it would be necessary to use all six parameters to fully define the system, which means the mass action model would be the simplest model that can accurately simulate the reaction.

The resulting network was simulated using Matlab's ODE15s function, with a relative tolerance of  $1 \times 10^{-5}$  with the following steps:

1. **Determining initial conditions.** For the dQSSA model, the initial conditions were found by solving Eqn. 78 with tensors  $\mathbf{Y}$  and  $\mathbf{Z}$  set to zero only at this step and with  $\sigma(t)$  set to add the initial concentrations of A, B, C and D with a Gaussian time profile centred at  $t = 0.5\text{s}$  and width  $t_w = 0.01\text{s}$  to the system. The time course was solved for  $t = [0, 1]\text{s}$ .
2. **Determining time course of the  $I_{\text{off}}$  phase.** The mass action, Michaelis-Menten and dQSSA models were then solved for  $t = [0, t_{1/2} - 10t_w]\text{s}$ .
3. **Addition of input I.** The three models were then solved for  $t = [t_{1/2} - 10t_w, t_{1/2} + 10t_w]\text{s}$ , using the final concentration from the previous run as the initial condition. This is to solve the addition of I into the system with high time resolution.
4. **Determining time course of the  $I_{\text{on}}$  phase.** The three models were then solved for  $t = [t_{1/2} + 10t_w, t_{\text{end}}]\text{s}$ , using the final concentration from the previous run as the initial condition.
5. The results were then stitched together.

Simulations were performed in these three stages because of limitations involved with the ODE15s routine. Average error and average computing time was calculated over 1000 different combinations of system parameters to test accuracy under all parameter conditions except those already specified. The average

computing time was determined using the "tic" and "toc" function in Matlab, with the total time summed from the simulation time of each phase. The average error for the dQSSA was calculated using all free and complex states compared to their mass action counterpart at equilibrium. For the Michaelis-Menten model, the error was compared to the equivalent state in the mass action model after all complexes are dissociated (e.g.  $[A_T] = [A] + [A - B] + [A - p1B]$ ) again at equilibrium. The following equation was then applied.

$$\text{err}_i = \begin{cases} 0 & , \text{ if } [dQSSA_i] = [MA_i] \\ \log_{10} \left[ \frac{|[dQSSA_i] - [MA_i]|}{\min([dQSSA_i], [MA_i])} \right] & , \text{ else} \end{cases} \quad (99)$$

$$\text{error}_{\text{tot}} = 10^{\left[ \frac{1}{n} \sum_{i=1}^n \text{err}_i \right]} \quad (100)$$

where  $i$  is the model states to be summed over, excluding states where  $[dQSSA_i] = [MA_i]$  because the logarithm of this value gives an undefined value. This gives a measure of the scale of the difference across all states, rather than the absolute differences, which can become dominated by states with large differences.

#### 4.2.2 In Vitro Validation Using a Biological Enzyme

##### 4.2.2.1 Mathematical Model

Since the Michaelis-Menten model is a single substrate enzyme kinetic models, the ordered bi-bi nature of LDH was separated into the enzyme and coenzyme binding step (Eqn. 101 and 102), followed by the enzymatic step (Eqn. 101 and 102). This is illustrated in Fig. 19.

The Michaelis-Menten implementation produces the following reaction velocities:

$$v_1 = 1000 / K_{NH}^i [NADH][LDH] - 1000 [NADH-LDH] \quad (101)$$

$$v_2 = 1000 / K_{NH^+}^i [NAD^+][LDH] - 1000 [NAD^+-LDH] \quad (102)$$

$$v_3 = k_{Py}^c \frac{[Py][NADH-LDH]}{[Py] + K_{Py}^m} \quad (103)$$

$$v_4 = k_{Lac}^c \frac{[Lac][NAD^+-LDH]}{[Lac] + K_{Lac}^m} \quad (104)$$

to form the following rate equations for each of the species:

$$d[\text{NADH}]/dt = -v_1 - v_3 + v_4 \quad (105)$$

$$d[\text{NAD}^+]/dt = -v_2 + v_3 - v_4 \quad (106)$$

$$d[\text{LDH}]/dt = -v_1 - v_2 \quad (107)$$

$$d[\text{Py}]/dt = -v_3 + v_4 \quad (108)$$

$$d[\text{Lac}]/dt = v_3 - v_4 \quad (109)$$

$$d[\text{NADH-LDH}]/dt = v_1 \quad (110)$$

$$d[\text{NAD}^+\text{-LDH}]/dt = v_2 \quad (111)$$

The dQSSA implementation produces the following tensors that are populated as per Eqn. 112-114, using the coupled irreversible enzyme reaction form. The rate parameters used are the  $K^i$  (dissociation constant) of LDH-NADH and LDH-NAD<sup>+</sup> and the  $K^m$  and  $k^c$  of the lactate and pyruvate reactions. The readout from

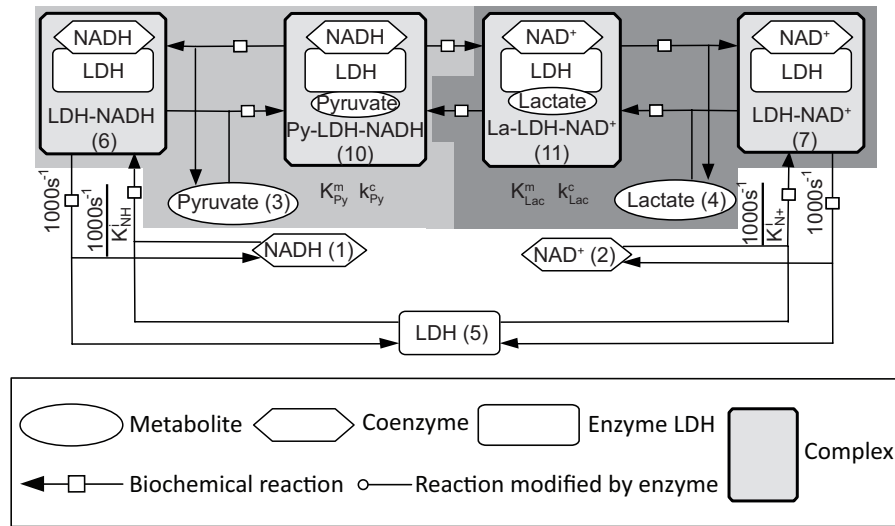


Figure 19: Illustration of the reaction network used to simulate the LDH reversible conversion of pyruvate to lactate. The numbers within the species indicates the species index produced by the simulation used in the study. Rate parameters are as defined at the beginning of chapter 3.1 except for  $K^i$  which are dissociation constants for their corresponding reactions. Numbers inside species indicate their index used in the mathematical model in the models shown in Appendix C.2

the models is the total NADH concentration (both free and bound forms).

$$\mathbf{G}_{ij} = \begin{cases} \frac{1}{K_{Py}^m}, & \text{if } [i, j, k] = [3, 3, 6], [3, 6, 3], [6, 3, 6], [6, 6, 3] \\ -\frac{1}{K_{Py}^m}, & \text{if } [i, j, k] = [10, 3, 6], [10, 6, 3] \\ \frac{1}{K_{Lac}^m}, & \text{if } [i, j, k] = [4, 4, 7], [4, 7, 4], [7, 4, 7], [7, 7, 4] \\ -\frac{1}{K_{Lac}^m}, & \text{if } [i, j, k] = [11, 4, 7], [11, 7, 4] \\ 0, & \text{else} \end{cases} \quad (112)$$

which is the dQSSA tensor that contains the Michaelis constants.

$$\mathbf{Y}_{ijk} = \begin{cases} -1000/K_{NH}^i, & \text{if } [i, j, k] = [1, 1, 5], [5, 1, 5] \\ 1000/K_{NH}^i, & \text{if } [i, j, k] = [6, 1, 5] \\ -1000/K_{NH+}^i, & \text{if } [i, j, k] = [2, 2, 5], [5, 2, 5] \\ 1000/K_{NH+}^i, & \text{if } [i, j, k] = [7, 2, 5] \\ 0, & \text{else} \end{cases} \quad (113)$$

which is the coenzyme and LDH dissociation reactions. And finally

$$\mathbf{Z}_{il} = \begin{cases} -1000, & \text{if } [i, l] = [6, 6], [7, 7] \\ 1000, & \text{if } [i, l] = [1, 6], [5, 6], [2, 7], [5, 7] \\ -k_{Py}^c, & \text{if } [i, l] = [1, 10], [3, 10] \\ k_{Py}^c, & \text{if } [i, l] = [2, 10], [4, 10] \\ -k_{Lac}^c, & \text{if } [i, l] = [2, 11], [4, 11] \\ k_{Lac}^c, & \text{if } [i, l] = [1, 11], [3, 11] \\ 0, & \text{else} \end{cases} \quad (114)$$

where the first two are the coenzyme and LDH association reactions, and the remaining are the enzyme kinetic reaction catalytic steps.

Since both Michaelis-Menten and dQSSA implementations of the system contain the initial enzyme-coenzyme binding phase

which we are not interested in, the simulation is run in two phases. The first phase sets the catalytic rates for both enzyme reactions to zero. This enables the enzyme-coenzyme binding reactions to equilibrate before the reaction begins. The catalytic rates are then reset to their required value and the time courses captured. The simulation is then run for 10 s before determining the initial velocity of NADH consumption/production, since the Michaelis-Menten system needs to some time to settle into its new transient quasi-steady state first, giving a more accurate representation of the predicted initial velocity.

The first parameters to be fitted were the  $K^m$  and  $k^c$  of the catalytic reactions. This was done by fitting the regression line of the Lineweaver-Burk plot of the initial velocities of the reactions. The residual  $\epsilon$  for this fitting routine is:

$$\epsilon = \sum_i (y - \mu_i)^2 \quad (115)$$

This allows the enzymatic parameters  $K^m$  and  $k^c$  to be determined independently of the  $K^i$  of LDH-coenzyme formation because the concentration of the competitive coenzyme was zero at the beginning of the reaction. Once the enzymatic parameters were fitted, the  $K^i$  of LDH-coenzyme formation was fitted using the Nelder-Mead simplex method. The residual  $\epsilon$  in this case is:

$$\epsilon = \sum_i \left( \frac{y - \mu_i}{\sigma_i} \right)^2 \quad (116)$$

The odds ratio was then used to compare the models in their ability to predict the data. This was calculated using:

$$O = \frac{P(M_{dQSSA}|D)}{P(M_{MM}|D)} \quad (117)$$

Where  $O$  is the odds ratio,  $M$  is the model and,

$$P(M_i|D) = \prod_i \exp \left( -\frac{1}{2} \left( \frac{y - \mu_i}{\sigma_i} \right)^2 \right) \quad (118)$$

Where,  $y$  is the model prediction,  $\mu_i$  is the mean of the experimental data and  $\sigma_i$  is the standard deviation of the experimental data.



#### 4.2.2.2 *Experimental Materials*

Rabbit muscle L-LDH with a concentration of 5 mg/mL (101278 76001) was purchased from Roche Diagnostics. NADH (43420), NAD<sup>+</sup> (N0632), sodium pyruvate (P4562), sodium lactate (L7022) and Corning polystyrene black-bottom microtitre plates (3915) were purchased from Sigma-Aldrich. Phosphate buffered solution (PBS) was used as the buffer due to its activating effects on LDH [46]. It was made internally using 0.36 (w/v)% Na<sub>2</sub>HPO<sub>4</sub>, 0.02 (w/v)% KCl, 0.024 (w/v)% KH<sub>2</sub>PO<sub>4</sub> and 0.8 (w/v)% NaCl.

#### 4.2.2.3 *In Vitro Experiments*

To derive the kinetic constants, equimolar solutions of substrate and cofactor (NADH and pyruvate, NAD<sup>+</sup> and lactate) were prepared in PBS. 50  $\mu$ L of each solution was added to the 96 well plate in 4 replicates. LDH solution (4 U/L, 100 nM) was similarly prepared in PBS diluted and 50  $\mu$ L injected into each well. Experiments were run in polystyrene black flat bottom 96 well plates (Corning 3915). Enzyme kinetics were measured by the fluorescence of NADH in a Fluostar Omega plate reader ( $\lambda_{\text{Ex}} = 355$  nm,  $\lambda_{\text{Em}} = 450$  nm, 5 flashes per well), with 250 reads per well recorded over 300 seconds. The gain was calibrated using an NADH standard curve (in PBS, without other substrates or LDH), which was prepared in each experiments in 4 replicates from 0-5 mM (100  $\mu$ L total volume). It was also found that presence of NAD<sup>+</sup> significantly reduced the fluorescence of the NADH. This decay was measured using a standard curve of 0 to 50 mM of NAD<sup>+</sup> in the presence of 0.3 mM of pyruvate and NADH. The observed fraction drop in fluorescence is then fitted to an exponential decay to model the absorbance of NAD<sup>+</sup>. The resulting model is the same as that shown previously in Fig. 19. Acquired fluorescence were converted to [NADH] using this standard curve and model of absorbance by [NAD<sup>+</sup>].

The initial velocity was calculated by fitting a straight line through the first 5 linear points of each time course, averaged over each technical replicate. We assume that the products of the reaction (or the reactants of the reverse reaction) are negligible at this time, allowing us to analyse each half reaction individually. The kinetic parameters,  $K^{\text{m}}$  and  $k^{\text{c}}$ , were determined by fitting the initial reaction velocities to the initial concentrations on a Lineweaver-Burk

plot. The NADH and LDH dissociation constant and  $\text{NAD}^+$  and LDH dissociation constants were the parameters remaining to be determined. Since these are quick processes, the dissociation rate was fixed at  $1000 \text{ s}^{-1}$ . From there it was found that, given the fast rate of association and dissociation between the enzyme and the coenzymes, the important parameter determine how NADH and  $\text{NAD}^+$  competes for binding with LDH was in fact the ratio between the NADH-LDH dissociation constant and the  $\text{NAD}^+$ -LDH dissociation constant. As such, the  $\text{NAD}^+$  and LDH dissociation constant was fixed at  $4.0 \times 10^{-7} \text{ M}$  based on literature [174]. From there the NADH-LDH dissociation constant was identified using the full time course of the reactions used to identify the enzymatic rate parameters. We did this by taking advantage of the fact that the four reactants become present at the reactions evolve beyond the initial regime where only two reactants are present.

To investigate the inhibitory effects of  $\text{NAD}^+$  on the pyruvate to lactate reaction, we measured the initial velocity of the reaction initiated using  $100 \text{ }\mu\text{L}$  solutions of pyruvate and NADH ( $0.3 \text{ mM}$ ) in the presence of increasing concentrations of  $\text{NAD}^+$  ( $0$  to  $10 \text{ mM}$ ).

### 4.2.3 In Silico Validation of Two Substrate dQSSA

The reversible LDH reaction shown in figure 19 was implemented using the two substrate dQSSA model (Eqn 94-95) as follows:

$$\mathbf{G}_{ij} = \begin{cases} \frac{1}{K_{NH}^i}, & \text{if } [i, j, k] = [1, 1, 5], [1, 5, 1], [5, 1, 5], [5, 5, 1] \\ -\frac{1}{K_{NH}^i}, & \text{if } [i, j, k] = [6, 1, 5], [6, 1, 2] \\ \frac{1}{K_{NH^+}^i}, & \text{if } [i, j, k] = [2, 2, 5], [2, 5, 2], [5, 2, 5], [5, 5, 2] \\ -\frac{1}{K_{NH^+}^i}, & \text{if } [i, j, k] = [7, 2, 5], [7, 5, 2] \\ \frac{1}{K_{Py}^i}, & \text{if } [i, j, k] = [1, 3, 6], [1, 6, 3], [3, 3, 6], \\ & [3, 6, 3], [5, 3, 6], [5, 6, 3] \\ -\frac{1}{K_{Py}^i}, & \text{if } [i, j, k] = [10, 3, 6], [10, 6, 3] \\ \frac{1}{K_{Lac}^i}, & \text{if } [i, j, k] = [2, 4, 7], [2, 7, 4], [4, 4, 7], \\ & [4, 7, 4], [5, 4, 7], [5, 7, 4] \\ -\frac{1}{K_{Lac}^i}, & \text{if } [i, j, k] = [11, 4, 7], [11, 7, 4] \\ 0, & \text{else} \end{cases} \quad (119)$$

which is the dQSSA tensor that contains the dissociation constants.

$$\mathbf{Z}_{ij} = \begin{cases} -k_{Py}^c, & \text{if } [i, j] = [1, 10], [3, 10] \\ k_{Py}^c, & \text{if } [i, j] = [2, 10], [4, 10] \\ -k_{Lac}^c, & \text{if } [i, j] = [2, 11], [4, 11] \\ k_{Lac}^c, & \text{if } [i, j] = [1, 11], [3, 11] \\ 0, & \text{else} \end{cases} \quad (120)$$

where the first two are the coenzyme and LDH association reactions, and the remaining are the enzyme kinetic reaction catalytic steps.

In this case, as per the expression that was derived in section 3.5, the constants used in the dQSSA here are simply the dissociation constants given in Eqn. 89 and 90. These are then compared

to the mass action implementation of the LDH reaction as shown in Appendix C.3.

This model was then evaluated in two different ways. Firstly, the parameters of the two substrate LDH dQSSA model was randomised using the same distribution described previously in Eqn. 98 (used in the hypothetical network to validate the one substrate dQSSA *in silico*). The error was evaluated using the same equation for evaluating the error for the one substrate dQSSA previously. (Eqn. 100). Secondly, the parameters achieved for the LDH system was applied to the two substrate dQSSA model to see how it compares to the mass action model. In both cases, the exact temporal profiles of all nodes are generated and compared to their corresponding nodes in the mass action model.

### 4.3 RESULT

#### 4.3.1 In Silico Validation of dQSSA

In the one thousand randomly generated parameter sets, the dQSSA model closely matched the time course of the mass action models, both in the I-off network and I-on network. At equilibrium, the average percentage difference between the two models was 0.5% over 1000 simulations (Fig. 20a). This I-off network verified that the dQSSA can replicate the mass action based reversible enzyme model while the I-on network showed that the dQSSA can accurately model the system behaviour with complex couplings in the network. We also investigated the performance of the Michaelis-Menten model in this scenario. We found that in many cases, the Michaelis-Menten model was inconsistent with the mass action and dQSSA model results, with an average difference of 9.7% over 1000 simulations (Fig. 20b). Fig. 21 shows a representative time course of one of these simulations, with the relevant parameters shown in Table 3. It can be seen that the Michaelien Menten model diverges from the mass action model prediction even in the "I off" regime, where the enzymatic reaction is completely isolated.

As seen in the theoretical analysis, this is most likely due to a violation of the low enzyme assumption of Eqn. 28. Therefore, we further investigated whether this inconsistency requires the low

enzyme assumption to be violated throughout the whole network, or whether a single instance is sufficient to cause inaccuracies in the whole network. To investigate this, we tested parameters such that the low enzyme assumption in all reactions except the forward direction of reaction 9. More specifically, the  $K^m$  of reaction 9 was set to unity while all other  $K^m$  were set to greater than 100. The specifics of the parameters used are given in Table 4. As such, the Michaelis-Menten model should be consistent with the other two models during the I-off phase but inconsistent in the I-on phase. We found that all three models were consistent during the I-off phase (Fig. 22). However, in the I-on phase, the dQSSA and mass action model remained consistent with each other while the Michaelis-Menten model became inconsistent and varied between 7% and 45% consistency at equilibrium. The A and pA states had a difference of approximately 20% even though these are not directly related to the species involved with the high enzyme concentration reaction. Hence, this result demonstrates that violation of the low enzyme assumption even in just one reaction can lead

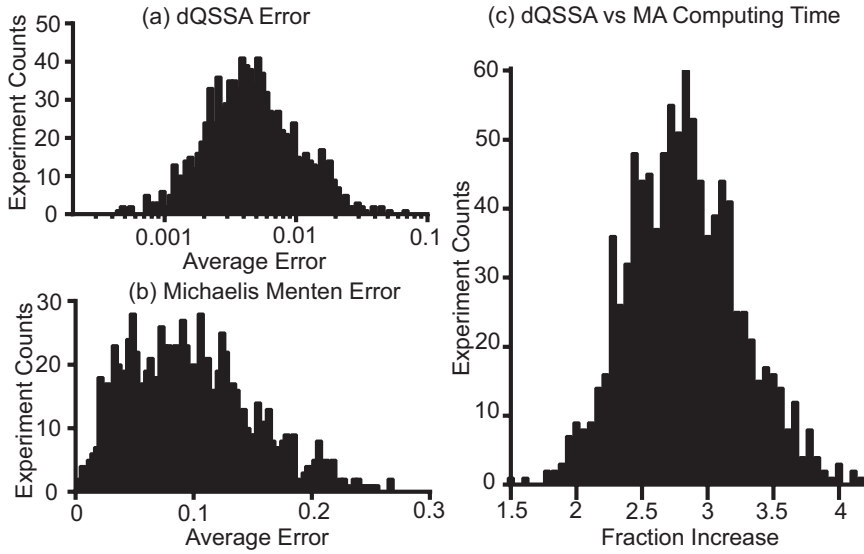


Figure 20: Accuracy of the (A) dQSSA and (B) Michaelis Menten model compared with the mass action model with 1000 different generated parameter sets. (C) Fraction increase in the computation time of each of the runs when compared with the mass action model. The errors are calculated averaged between the outcomes from the "I off" and "I on" simulations.

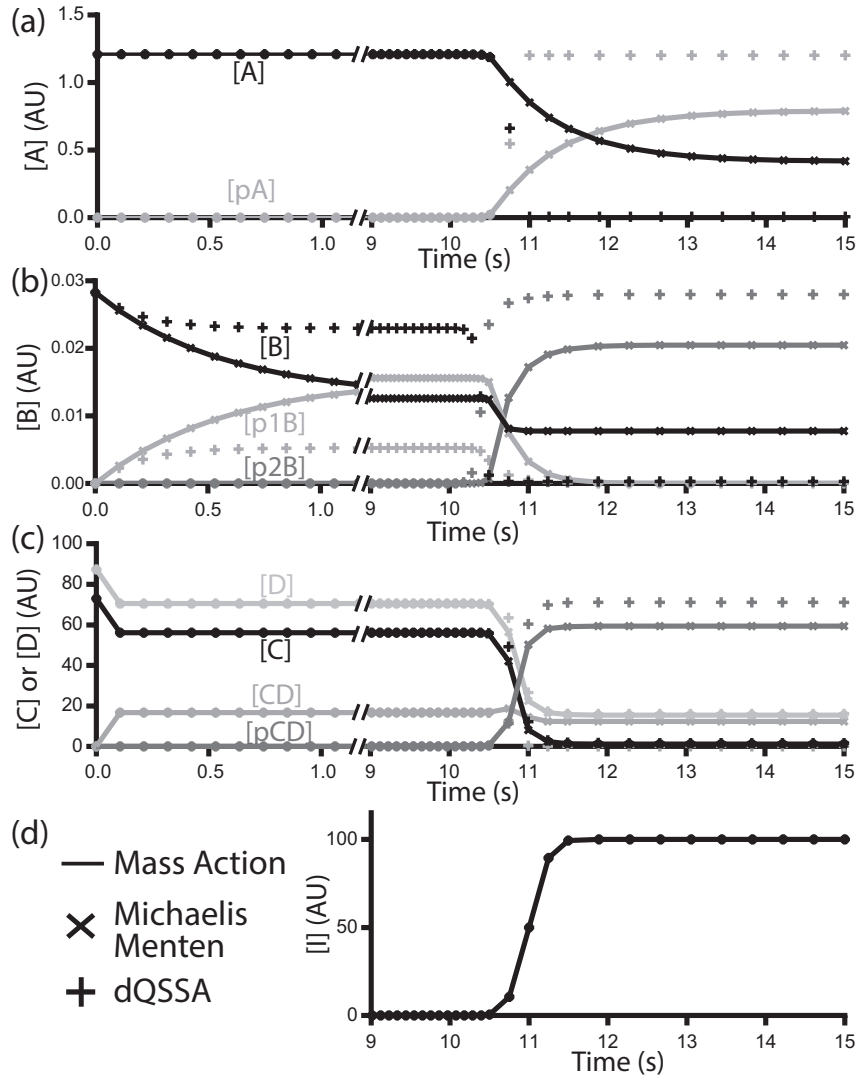


Figure 21: Representative time course of the reaction with the parameters used shown in table 3 where the Michaelis-Menten model does not agree with the mass action or dQSSA models in both the "I off" and "I on" regime. Between  $t = [0, 10)$  I is not present and the system operates in the "I off" regime. Between  $t = [10, 12)$  I is added which initiates some changes in the system (shown in subpanel d). When  $t > 12$  the system is in the "I on" phase. The solid line indicates the mass action prediction which is used as the gold standard. Crosses indicates the dQSSA model while pluses indicates the Michaelis-Menten model.

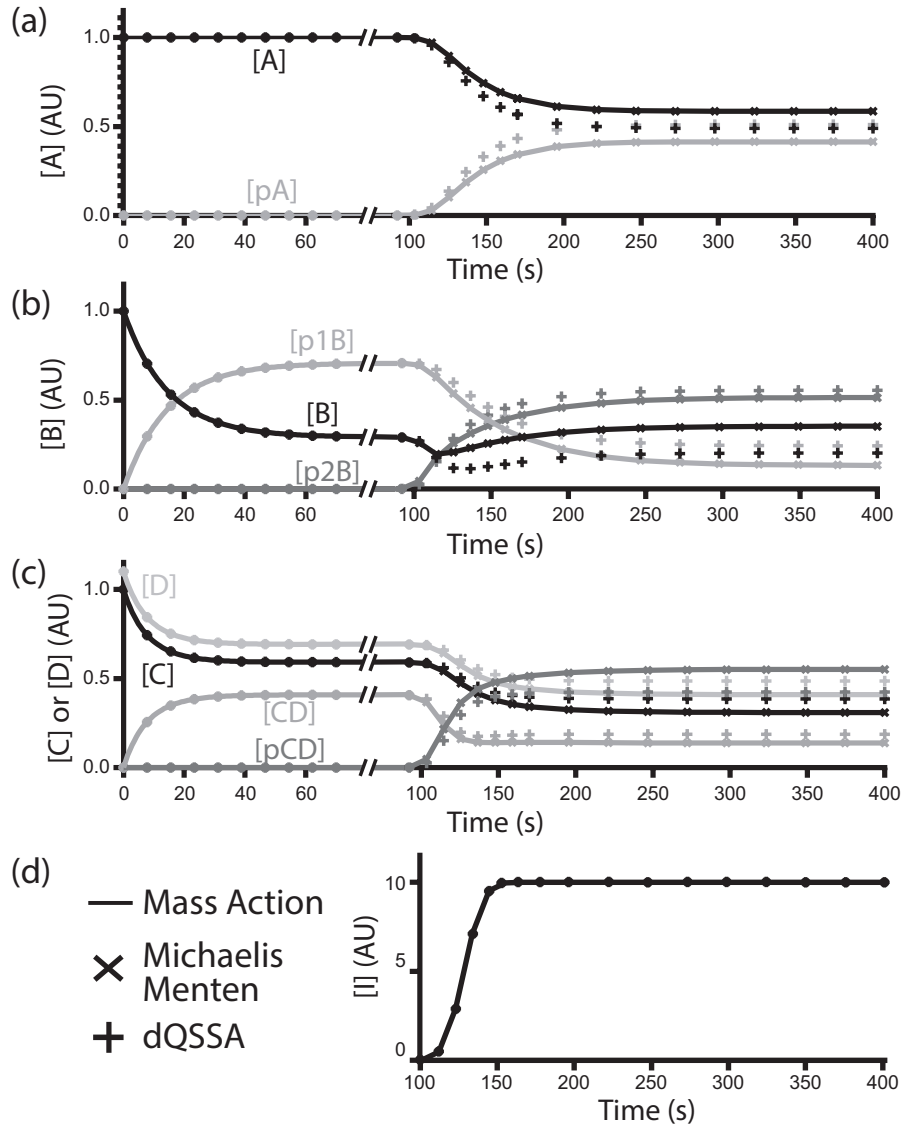


Figure 22: Representative time course of the reaction with the parameters used shown in table 4. Between  $t = [0, 100)$   $I$  is not present and the system operates in the "I off" regime. Between  $t = [100, 120]$   $I$  is added which initiates some changes in the system (shown in subpanel d). When  $t > 120$  the system is in the "I on" phase. The solid line indicates the mass action prediction which is used as the gold standard. Crosses indicates the dQSSA model while pluses indicates the Michaelis Menten model.

to non-trivial discrepancies in the Michaelis-Menten model's predictions.

While the dQSSA was found to be significantly better than the Michaelis-Menten model at matching the mass action model's predictions, the linearisation of the dynamic equation and the inclusion of a matrix inversion increases the model's computational cost. For example, in our hypothetical network, we found that the dQSSA model (including solving for the initial conditions) required approximately 3 times longer to solve when compared to the mass action model (Fig. 20c). The accuracy of the dQSSA for estimating kinetic parameters was also tested by fitting the dQSSA model to the mass action generated time course with artificial noise with signal to noise ratio of 5 was added. It was found that fitted parameters are at most different by an average of 20% and the uncertainty of the fitted parameters covers the true parameter value as shown in Fig 23. This demonstrates that the dQSSA is useful in parameter estimation against experimental time course data.

Table 3: Parameters used to generate the time course shown in Fig. 21.

Mass Action						dQSSA and Michaelis-Menten					
$k_{1f}^a$	84.0	$k_{5r}^a$	45.9	$k_{9f}^d$	884					$K_{9f}^m$	0.0378
$k_{1f}^d$	2750	$k_{5r}^d$	1340	$k_{9f}^c$	8.84	$K_{1f}^m$	33.1	$K_{5r}^m$	29.3	$K_{9f}^m$	8.84
$k_{1f}^c$	27.5	$k_{5r}^c$	3.66	$k_{9r}^a$	658	$k_{1f}^c$	27.5	$k_{5r}^c$	3.66		
$k_{1r}^a$	10800	$k_6$	1.39	$k_{9r}^d$	884			$k_6$	1.39	$K_{9r}^m$	1.35
$k_{1r}^d$	2750	$k_{7f}^a$	8480	$k_{9r}^c$	3.36	$K_{1r}^m$	0.26			$k_{9r}^c$	3.36
$k_{1r}^c$	0.229	$k_{7f}^d$	77.6	$k_{10f}^a$	0.343	$k_{1r}^c$	0.229	$K_{7f}^m$	0.00925		
$k_2$	3.32	$k_{7f}^c$	0.776	$k_{10f}^d$	2550	$k_2$	3.315	$k_{7f}^c$	0.776	$K_{10f}^m$	7510
$k_3$	0.416	$k_{7r}^a$	235	$k_{10f}^c$	25.5	$k_3$	0.416			$k_{10f}^c$	25.5
$k_4$	98.4	$k_{7r}^a$	77.6	$k_{10r}^a$	181	$k_4$	98.37	$K_{7r}^m$	0.332		
$k_{5f}^a$	4320	$k_{7r}^c$	0.4	$k_{10r}^d$	2550			$k_{7r}^c$	0.4	$K_{10r}^m$	14.1
$k_{5f}^d$	1340	$k_8$	0.821	$k_{10r}^c$	3.25	$K_{5f}^m$	0.31	$k_8$	0.821	$k_{10r}^c$	3.25
$k_{5f}^c$	13.4	$k_{9f}^a$	23600	$k_{11}$	3.14	$k_{5f}^c$	13.4			$k_{11}$	3.14
Concentrations											
$[A]_0$		$[B]_0$		$[C]_0$		$[D]_0$		$[I]_0$			
1.21		0.0283		73.0		87.3		100.0			



### 4.3.2 In Vitro Validation of dQSSA

Thus far, the in silico validation demonstrated good agreement between the mass action model and the dQSSA while highlighting that the Michaelis-Menten model is not applicable for this type of modelling, particularly when the reactant stationary assumption is violated. To determine its practical significance, we extended our comparison to an in vitro setting, modelling the action of LDH.

LDH is a well-characterised enzyme which reversibly converts pyruvate and reduced nicotinamide adenine dinucleotide (NADH) to lactate and oxidised nicotinamide adenine dinucleotide (NAD<sup>+</sup>). The reaction mechanism for LDH (Fig. 19) involves the ordered binding of the coenzymes (NADH or NAD<sup>+</sup>) to LDH followed by the subsequent binding of its corresponding substrate (pyruvate or lactate, respectively) [175]. The transfer of an electron between the coenzyme and the substrate then reversibly occurs in

Table 4: Parameters used to generate the time course shown in Fig. 22.

Mass Action						dQSSA and Michaelis-Menten					
$k_{1f}^a$	5	$k_{5r}^a$	0.1	$k_{9f}^d$	100					$K_{9f}^m$	0.2
$k_{1f}^d$	1000	$k_{5r}^d$	100	$k_{9f}^c$	1	$K_{1f}^m$	202	$K_{5r}^m$	1001	$K_{9f}^m$	0.2
$k_{1f}^c$	10	$k_{5r}^c$	0.1	$k_{9r}^a$	0.5	$k_{1f}^c$	10	$k_{5r}^c$	0.1		
$k_{1r}^a$	0.5	$k_6$	0.4	$k_{9r}^d$	1000			$k_6$	0.4	$K_{9r}^m$	1
$k_{1r}^d$	1000	$k_{7f}^a$	5	$k_{9r}^c$	0.5	$K_{1r}^m$	2001			$k_{9r}^c$	0.5
$k_{1r}^c$	0.5	$k_{7f}^d$	1000	$k_{10f}^a$	5	$k_{1r}^c$	0.5	$K_{7f}^m$	202		
$k_2$	0.02	$k_{7f}^c$	10	$k_{10f}^d$	1000	$k_2$	0.02	$k_{7f}^c$	10	$K_{10f}^m$	202
$k_3$	0.05	$k_{7r}^a$	0.5	$k_{10f}^c$	10	$k_3$	0.05			$k_{10f}^c$	10
$k_4$	0.05	$k_{7r}^a$	1000	$k_{10r}^a$	0.5	$k_4$	0.05	$K_{7r}^m$	2001		
$k_{5f}^a$	1	$k_{7r}^c$	0.5	$k_{10r}^d$	1000			$k_{7r}^c$	0.5	$K_{10r}^m$	2001
$k_{5f}^d$	100	$k_8$	0.02	$k_{10r}^c$	0.5	$K_{5f}^m$	100.5	$k_8$	0.02	$k_{10r}^c$	0.5
$k_{5f}^c$	10	$k_{9f}^a$	1000	$k_{11}$	0.5	$k_{5f}^c$	10			$k_{11}$	0.5
Concentrations											
[A] <sub>0</sub>		[B] <sub>0</sub>		[C] <sub>0</sub>		[D] <sub>0</sub>		[I] <sub>0</sub>			
1.0		1.0		1.0		1.1		10.0			

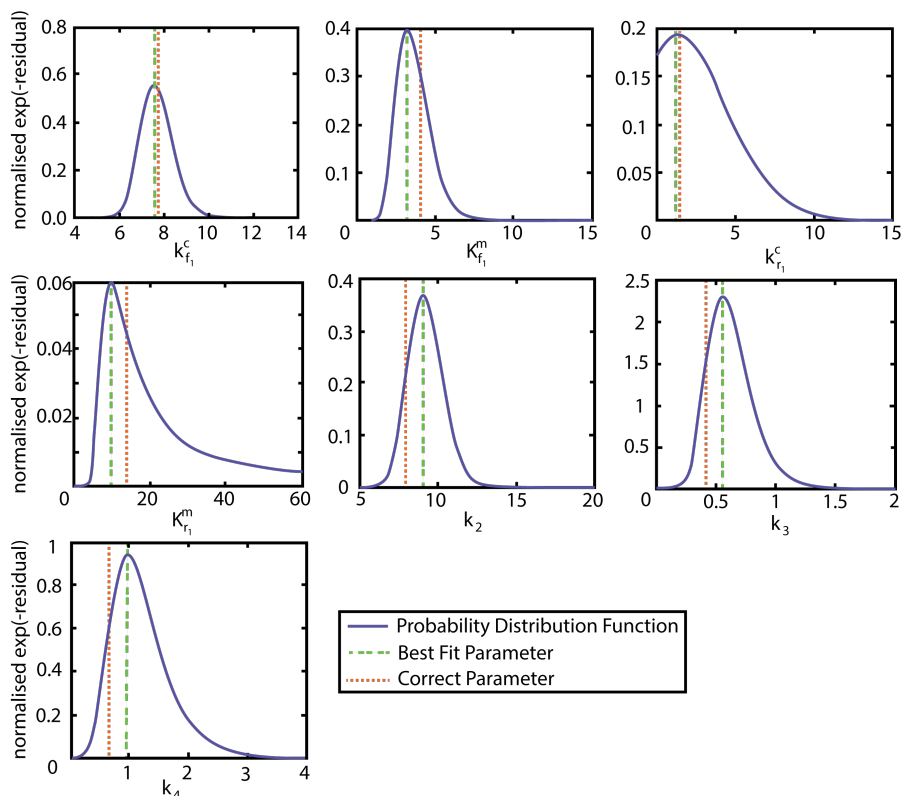


Figure 23: Inference of the parameter when Gaussian noise with a signal to noise ratio of 5 is added. The correct parameter (dotted vertical line) is contained within the blue curve (the likelihood distribution of the parameter)

the ternary complex as part of the catalytic process. As LDH appears to be an enzyme that satisfies the "rapid equilibrium" assumption, this reversible catalysis can be viewed as two distinct reactions as described in section 3.3 [176]. Thus, it is ideal for verifying whether the *in silico* difference found between the dQSSA model and the Michaelis-Menten model seen in the previous *in silico* experiment translates *in vitro*.

As this is a real system, the first step involved characterising the kinetic constants in both directions. The reactions were run in the irreversible regime by adding equimolar concentrations of the reactant and coenzyme of the relevant reaction direction. The constants were then obtained by least squares fitting of the initial concentration and reaction velocity using a Lineweaver Burk plot (Fig. 24a,b). The time course of the reaction was run with a low time resolution to smooth out the kinetics of the transient phase,

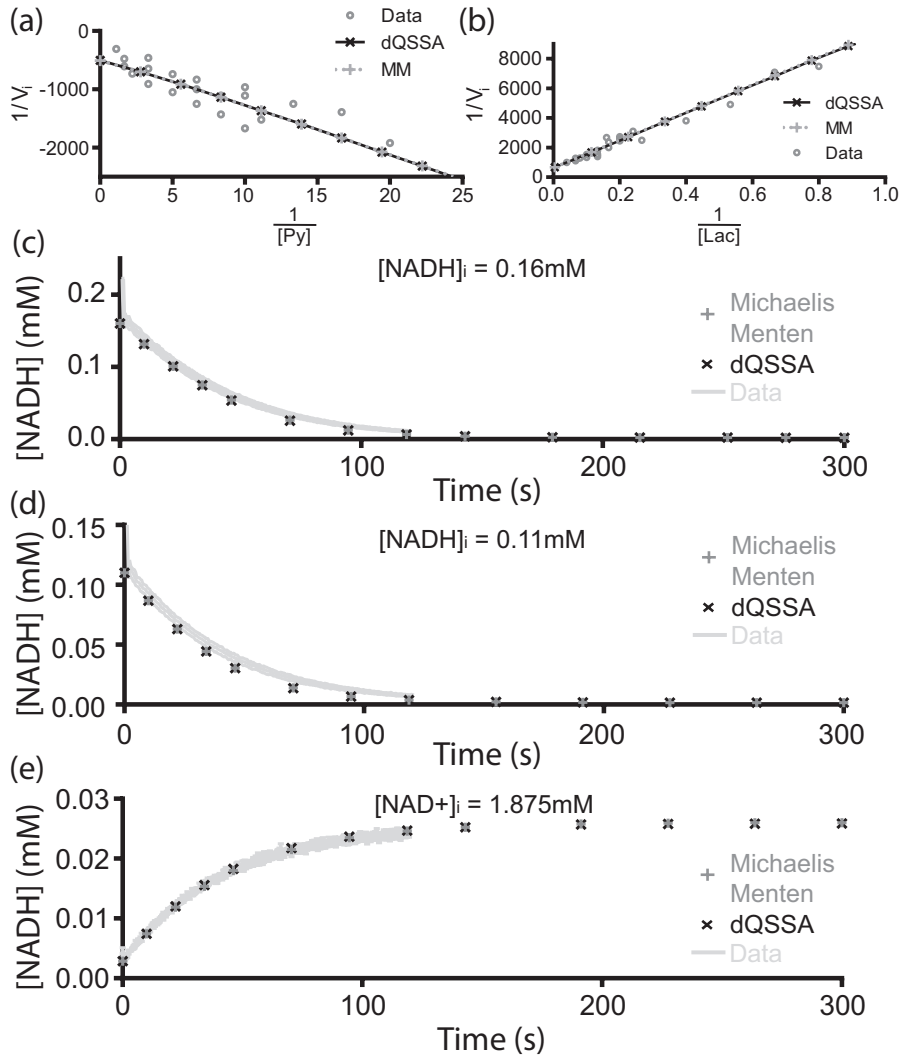


Figure 24: Illustration of the reaction network used to simulate the LDH reversible conversion of pyruvate to lactate. (a) and (b) shows the Lineweaver-Burke plots of the inverse initial reaction rate against the inverse concentrations of pyruvate and lactate respectively. Circles represent the experimental result while solid black and dashed grey lines represent the dQSSA and Michaelis-Menten model predictions respectively. (c) - (e) shows a representative time course of the experiment with the two model predictions overlaid. In these panels, solid line represents the experimental data while crosses and pluses represent the dQSSA and Michaelis-Menten model respectively.

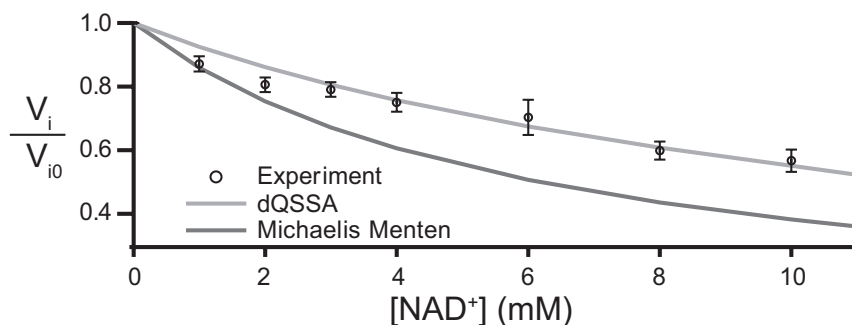


Figure 25: Validation result of the LDH model. The pyruvate to lactate reaction was tested with varying amount of  $\text{NAD}^+$ . The initial velocity of the reaction ( $V_i$ ) was then measured as a fraction of the initial velocity in the absence of  $\text{NAD}^+$  ( $V_{i0}$ ). This is shown as open circles. Error bars show the SEM. The dQSSA model (light grey line) and the Michaelis Menten Model (dark grey line) of this scenario was simulated and their results plotted.

and the reaction velocity was calculated using the first 5 time points to smooth out other experimental noise that can skew this parameter. The dissociation constants of the coenzymes were then determined by fitting the predicted time course with the experimental time course, which describes the kinetics of the system when all four reactants are present. Fits to three representative measured time courses for three different initial concentrations of  $[\text{NAD}^+]$  (all with the same parameters values) are shown in Fig. 24c-e. This approach was taken as the progression of the reaction caused all four reactants to become present. We reasoned that this causes the effect of coenzyme competition for the LDH to become present, which can be used to determine the dissociation constants. The resulting parameters are shown in Table 5 and were found to be in good agreement with that found in the literature [174, 177].

At this point, it was necessary to validate the prediction potential of the dQSSA and Michaelis-Menten models. It is expected that when a reaction is initiated in a single direction, the presence of the opposing coenzyme will cause inhibition as some enzyme is bound with the wrong coenzyme. Given the quantitative nature of the both models, they should be able to correctly predict the degree of inhibition this produces. As such, prediction of

the change in initial reaction velocity of the pyruvate to lactate reaction, under varying concentration of  $\text{NAD}^+$  was used to test the model's predictions. A good agreement was found between the dQSSA model's prediction and the observed result (Fig. 25). On the other hand, the Michaelis-Menten model gave a different prediction from the dQSSA model which was a poorer fit to the experimental results (Fig. 25). Using an odds ratio quantification of the goodness of fit, the dQSSA is the better model with an odds ratio  $O = 2.3 \times 10^{28}$  in favour of the dQSSA model. This shows that the dQSSA is able to make accurate temporal predictions for enzyme reactions under physiological conditions, and that the Michaelis-Menten model produces inaccuracies beyond intrinsic inaccuracies originating from finite precision in numerical methods.

#### 4.3.3 In Silico Validation of Two Substrate Model

Using 1000 different randomly generated parameters, the error of the two substrate dQSSA was found to be on average thirty times lower than the one substrate dQSSA model when compared to the mass action models (Fig. 26). It should be noted that the average error in the one substrate dQSSA model in this model is about 0.05%, which is already extremely small.

Table 5: Rate parameters of the LDH system that are achieved when fitted to the Michaelis-Menten and dQSSA models. The residuals of the best fits are also given.

	Michaelis-Menten		dQSSA	
	Value	Residual	value	Residual
$\text{NADH } K_{\text{NH}}^i \text{ (M)}$	$9.23 \times 10^{-9}$	$4.2 \times 10^4$	$9.83 \times 10^{-9}$	$3.0 \times 10^4$
$\text{NAD}^+ K_{\text{NH}}^i \text{ (M)}$	$3.79 \times 10^{-7}$		$3.79 \times 10^{-7}$	
$K_{\text{Py}}^m \text{ (M)}$	$1.25 \times 10^{-4}$	$9.5 \times 10^5$	$1.27 \times 10^{-4}$	$9.5 \times 10^5$
$k_{\text{Py}}^c \text{ (s}^{-1}\text{)}$	78.9		79.3	
$K_{\text{Lac}}^m \text{ (M)}$	$1.30 \times 10^{-2}$	$2.0 \times 10^6$	$1.30 \times 10^{-2}$	$2.0 \times 10^6$
$k_{\text{Lac}}^c \text{ (s}^{-1}\text{)}$	62.8		63.3	

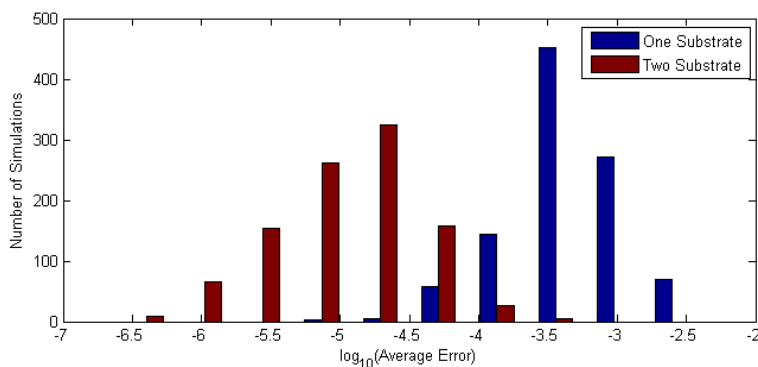


Figure 26: Distribution of errors in 1000 simulations with randomly generated parameters in the LDH model between the one substrate and two substrate dQSSA.

In general the time course for all states were very consistent between the three models. The use of the parameters measured for LDH demonstrated this point. The time course derived for the three models, using an LDH concentration of  $2.5\mu\text{M}$  and NADH, Pyruvate and  $\text{NAD}^+$  concentrations of  $1\text{ mM}$ , is shown in Fig. 27. The three models produce the same kinetic behaviour so illustrated in the stable portion of Fig. 27. However, under some circumstances, particularly with low enzyme concentrations, the numerical integration of the one substrate dQSSA is prone to instabilities. This is illustrated by the unstable region in Fig. 27. In this case, the three models produce a very similar time course, but the one substrate model slowly deviates, leading to an error of some 5% at the 2000 s mark. After approximately 2500 s, the one substrate model becomes unstable, which can be seen by the large fluctuations in the time course of the complex states. It should be noted that the instability does not always occur. This example was chosen as it consisted of a stable and unstable region and can be used to illustrate both behaviours concisely. It is possible that this is associated with the inversion of near singular matrices which becomes sources a large numerical errors in the simulation.

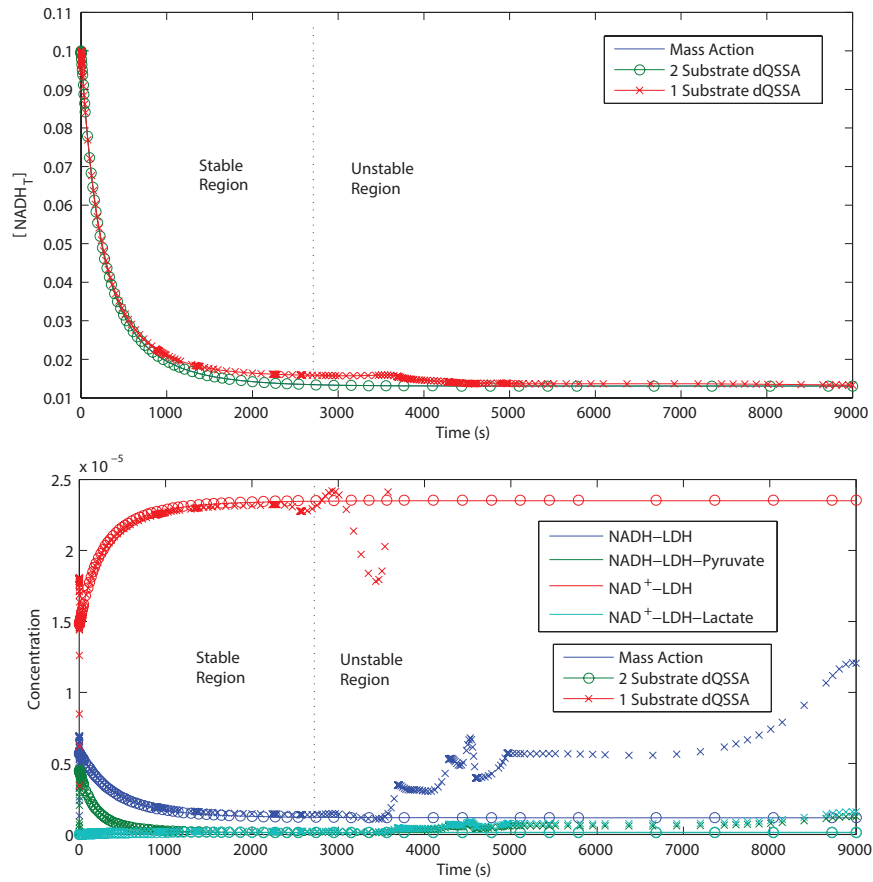


Figure 27: Representative time course showing instability in the one substrate dQSSA model showing instability which does not occur in the two substrate model. In all panels, the mass action, one substrate dQSSA and two substrate dQSSA models are shown as a solid line, crosses and circles respectively. The upper panel shows the time course of total NADH in the three models, showing they produce similar predictions. The lower panel shows the concentrations of the various complexes of LDH, which reveals instability in the one substrate model.

#### 4.4 DISCUSSION

In this study, we set out to test and validate the dQSSA model derived in chapter 3. The one substrate dQSSA model was thoroughly tested both *in silico* and *in vitro*. In the *in silico* test, the dQSSA was found to computationally match the accuracy of the

fully reversible enzyme kinetic model both in a single cyclic reaction, and in a multienzyme complex network. While these were promising results, the dQSSA is more computationally expensive as there is a three fold increase in computation time associated with it. In spite of this, the use of the less computationally efficient dQSSA model can be justified for a number of reasons. Firstly, the dQSSA is a simplified model compared to the mass action model as it requires one less rate parameters out of the three in the mass action model to fully define, which in itself reduces computational burden originating from parameter fitting [178]. Secondly, the dQSSA is accurate a wide range of scenarios. The equally simple Michaelis-Menten model becomes inaccurate when the reactant stationary assumption is violated, but the dQSSA remains accurate in these scenarios. And finally, the dQSSA is applicable in biologically relevant scenarios. The dQSSA model was able to correctly predict the kinetic behaviour of LDH in a situation where both substrate and product are initially present, which is a more physiologically relevant scenario than experiments where products are not presents when the reaction is initialised.

While the dQSSA does not account for all physical mechanisms involved in the interchange between intermediate complexes, its focus is on the general existence of intermediate complexes and simulating conditions that more closely resemble the *in vivo* context. Although the initial quantity of the intermediate complex cannot be measured in our *in vitro* assay, the model can be further validated by investigating this using high-throughput, highly-sensitive spectrometric techniques [14]. The excellent agreement between our model prediction and *in vitro* experimental data in the LDH example, even using a predicted rather than measured initial [ES], shows that indeed this additional detail can greatly improve the dQSSA's performance as compared to the Michaelis-Menten model, which justifies its use in modelling physiological systems.

In spite of the successes of the one substrate dQSSA model, further testing on the LDH system revealed that it is prone to computational instabilities. This was solved by the two substrate dQSSA model, which we have shown is for all intents and purposes as accurate as the one substrate model. However, it is unclear how the two substrate model resolves the instability issue. This would be an interesting avenue for further theoretical re-



search into the dQSSA models. The two substrate model also has the advantage of simplifying the system more by further reducing the dimensionality by one. Thus a two substrate enzyme kinetic model can be fully described with three parameter rather than the five that would be required with the equivalent mass action model. Although the two substrate model has not been tested in an *in silico* network as complex as the one tested for the one substrate model, the similarity of the two substrate model to the one substrate model in modelling the LDH system suggests that their performances would universally be the same. Aside from the instabilities caused by modelling processes with vastly different time scales, the model is able to produce the same prediction whether reactions are modelled in quasi-steady state or as dynamic reactions.

There are other advantages in using either forms of the dQSSA. As mentioned in the introduction, systems modellers must currently choose from a plethora of enzyme kinetic models of different complexities, such as the inclusion of product inhibition. As the dQSSA was derived using only the rapid equilibrium assumption, it retains most features of the mass action based model, encompassing all simplified models of the reversible enzyme kinetic model. Thus, the dQSSA is able to model a wide variety of enzyme mediated biological processes from post translational modifications in signalling, to metabolic processes.

Overall, we have shown that the dQSSA can act as a faithful substitute for the reversible enzyme kinetic model in cases where the rapid equilibrium assumption is valid, and allows all single substrate enzyme kinetic models to be collapsed into a single model. Since the dQSSA requires the rapid equilibrium assumption, this model would be less accurate for modelling enzymes which do not satisfy this, for example, carbonic anhydrase or acetylcholinesterase. In these cases, the full mass action model would be required. Nonetheless, our model merges the enzymology understanding of enzyme kinetics based on the reversible enzyme kinetics model, with the systems biology application of enzyme kinetics, typically based on irreversible Michaelis-Menten kinetics.

The dQSSA model can also improve the way models are constructed and communicated. Conventionally, systems models are constructed and communicated using non-linearised rate equa-

tions. While this is an unambiguous way of describing a model, it can become verbose for large models - it is not uncommon to find equations spanning over many pages [103]. In the case of the dQSSA, the rate equation remains in the same form regardless of the system studied. Instead, the model topology is changed by varying the element values within the tensors in the rate equation. Whilst the population of the dQSSA tensors elements is complex and unintuitive, this can be automated using computer algorithms by creating rules relevant to different reaction schemes. The shift in focus from rate equation to topology in describing biochemical models has some support in the literature, with the Systems Biology Market Language (SBML) project attempting to overcome model ambiguity and verbosity in a similar way [179, 180]. The dQSSA is well suited to the SBML approach since the rules governing how reactions are implemented are independent of the topology of the network and hence universal. As such, topology can be easily created using the dQSSA's framework.

Conversely, the model topology of a dQSSA model can in principle be inferred from the tensor elements, which means dQSSA models can be communicated by providing the tensor structures. This also enables system topology to be inferred by fitting tensor values to experimental time courses and dose response data. In practice, this is not currently possible as the size of the tensor scales as  $n^3$  for  $n$  number of species in the model. Nonetheless, with continued improvements in computational software (e.g. optimisation techniques) and hardware (e.g. memory), there remains potential for this approach to become practical in the future.

Now that we have validated our novel enzyme kinetics model, we move onto the development of tools for applying this model to arbitrary descriptions of biological networks in order to streamline the model design process.

## DEVELOPMENT OF MODELLING TOOLS

---

### **Abstract**

Many modelling tools exist for the construction of biochemical models and generating the final differential equation needed to make quantitatively simulate. However, these tools are not compatible with the dQSSA model. In order to overcome this, a new modelling tool was developed in the MATLAB environment. Coined SigMat, the tool incorporates other features such as modelling of protein localisation. It is shown to be intuitive to use and correctly reproduces the simulation results of a manually constructed model. The capability of the algorithm is also demonstrated by rapidly generating two network motifs, a signalling merger and a sustained responder to transient stimulations. These were analysed through parameter perturbation to understand their behaviour.

The computer code associated with this chapter is publicly available at [the SigMat GitHub repository](#).

## 5.1 INTRODUCTION

At this stage we have established that the mathematical formalism behind the dQSSA remains consistent when applied to single reactions or multiple reactions in a system. The next stage in the modelling process, is to apply the dQSSA to signalling pathways. This immediately presents a challenge because signalling pathways are often large and involve hundreds of reactions. This quickly leads to a mathematically complex set of differential equations that can span dozens of pages [104]. Additionally, different rate laws may be used for the similar reaction types (e.g. mechanism and empirical rate laws may be used for different enzyme kinetic reactions). Together, these factors can lead to difficulties in producing the model correctly and consistently, which can make published models difficult to independently verify [179]. Even with the dQSSA formalism, this remains a challenge because the large lists of ODEs are replaced with large arrays of matrices.

To overcome this difficulty, and to standardise the working blueprint of biological systems in the scientific community, a modelling standard (the SBML format) was established using the universal XML format. This is a cross platform, compact, human readable formalism for describing biochemical systems as a collection of attributes which are needed to unambiguously simulate the system [89, 181, 182]. Unfortunately the way SBML has been designed assumes the system differential equation can be constructed by simple addition of individual rate equations. This is not the case for the dQSSA due to the additional prefactor it introduces. Thus the dQSSA model cannot be defined using the SBML format and is not compatible with existing modelling tools.

Here, a new tool is created to parse models with a structure similar to the SBML format but compatible with the dQSSA, mass action and hill models as well as protein compartmentalisation. These models are integrated into the modelling algorithm. Finally, we use this framework to explore a number of theoretical pathways that can produce biologically interesting behaviour as a demonstration of the tools.

## 5.2 IMPLEMENTING COMPARTMENTALISATION

### 5.2.1 Discretised model of compartmentalisation

Compartmentalisation models generally start with a number of assumptions. Firstly, compartments are internally well mixed with constant concentration throughout the compartment, allowing it to be modelled as a deterministic process. Secondly, transport between compartments is slow compared to the mixing rate within compartments [102]. When both assumptions are satisfied, transport between compartments can be modelled as elementary reactions [102].

There are many computational tools in the field that incorporate compartmentalisation. However, a study by Alves showed that the results generated by many of these tools were non-intuitive [183]. This is because many are poorly documented and use inconsistent units (e.g. in one case the rate parameter used for the transport process is scaled to the absolute size of the compartment while in another it is not). To address this, we will be aiming to implement compartment transport such that the unimolecular rate parameter is defined based on the relative geometry between the relative compartments.

Transport between compartments can be modelled by discretising Fick's law of diffusion. The net flux moving from compartment 2 to compartment 1 is described by the following expressions, following the system illustrated in Fig. 28 [183]:

$$J = D \frac{\partial [X]}{\partial x} \quad (121)$$

$$\frac{1}{A} \frac{dX_1}{dt} = P_s ([X_2] - [X_1]) \quad (122)$$

$$\frac{dX_1}{dt} = P_s A ([X_2] - [X_1]) \quad (123)$$

where  $J$  is the flux into compartment 1,  $D$  is the diffusion constant,  $X$  is the absolute number of molecules,  $[X]$  is the concentration of  $X$  and their subscripts indicating their originating compartment,  $x$  is the spatial dimension along which diffusion occurs,  $P_s$  is the permeability coefficient and  $A$  is the interface area between the two compartments.

These equations show that diffusion determines the number of molecules transported as a function of the concentration. In

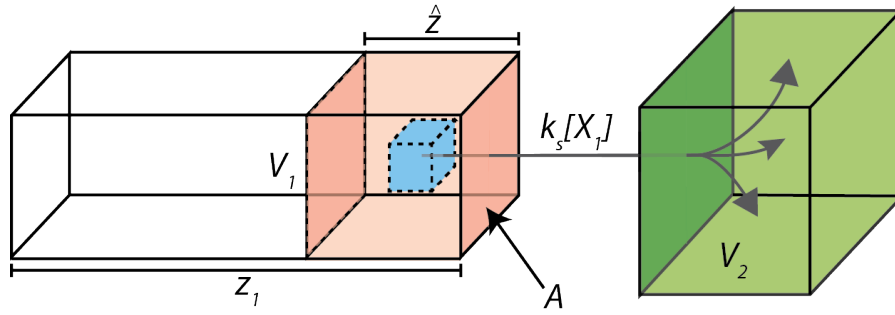


Figure 28: Figure showing the physical interpretation of the discretised model of transport between  $V_1$  to  $V_2$ .  $X_1$  within a unit volume (dash bordered cube) is transported to  $V_2$  at a rate of  $k_s^{ap}$  and disperses throughout  $V_2$ . All  $X_1$  within the red volume of  $V_1$  (with a depth of  $\hat{z}$  undergoes this transport process to form the total transported amount. The red volume is only a subset of the total volume of  $V_1$  because the clear parts of  $V_1$  are too deep within the compartment to be transported to  $V_2$ . This affects the accessibility of  $X_1$  available to be transported.

order to determine the rate of change of concentration in each compartment due to transport, the expressions need to be further manipulated by normalising to the volume of the compartment of interest. In other words:

$$\frac{d[X_1]}{dt} = \frac{P_s A}{V_1} ([X_2] - [X_1]) \quad (124)$$

$$\frac{d[X_2]}{dt} = \frac{P_s A}{V_2} ([X_1] - [X_2]) \quad (125)$$

Where  $V_1$  is the size of the compartment which  $X_1$  is in. Similarly for  $V_2$ .

We note that  $P_s[X]$  is similar to the rate equation of a unimolecular reaction. The difference is  $P_s$  has units  $\text{ms}^{-1}$  while the unimolecular rate constant  $k$  has units  $\text{s}^{-1}$ , then we can rewrite Eqns. 124 as:

$$\frac{d[X_1]}{dt} = \frac{1}{V_1} \left( \frac{k^{ap} \hat{z} A}{V_2} V_2 [X_2] - \frac{k^{ap} \hat{z} A}{V_1} V_1 [X_1] \right) \quad (126)$$

This generates the apparent rate constant  $k^{ap}$  and a combination of geometry terms. We can break up  $V_1$  into  $z_1 A_1$  where  $z_1$  is a characteristic depth of compartment  $V_1$ .

$$\frac{d[X_1]}{dt} = \frac{1}{V_1} \left( \frac{k^{ap} \hat{z} A}{z_2 A} V_2 [X_2] - \frac{k^{ap} \hat{z} A}{z_1 A} V_1 [X_1] \right) \quad (127)$$

This then allows  $A_1$  to be cancelled out from both the numerator and denominator. Subsequently,  $\hat{z}$ , the transport depth, can be combined with  $z_1$  to form the geometry term  $r$  that is the ratio of  $\hat{z}$  and  $z_i$ , which describes the percentage suppression of the reaction due to suboptimal interface area to compartment volume ratio. In this model,  $k^{ap}$  is then the transport rate when the interface area to volume ratio is maximised (for instance if a mitochondria is in its most elongated form).  $r$  is then the percentage reduction of the interface area in another scenario compared to the previously described idealised scenario.

$$\frac{d[X_1]}{dt} = \frac{1}{V_1} (r V_2 k^{ap} [X_2] - r V_1 k^{ap} [X_1]) \quad (128)$$

This gives the rate of change due to transport between compartments modelled as a unimolecular reaction. The interpretation of this is illustrated in Fig. 28. Using this interpretation,  $k^{ap}$  can be considered the diffusion rate per unit volume under a geometric arrangement that maximises molecular transfer between compartments.

### 5.2.2 Elementary Reactions

It is possible to expand the concept of transport to other rate laws using the following general expression:

$$\frac{d[X_j]}{dt} = \frac{\sum_i \dot{N}_i}{V_j} \quad (129)$$

Where  $\dot{N}_i$  is the rate equation in terms of absolute molecules number from reaction  $i$ :

$$\dot{N}_i = r_i V_i v_i \quad (130)$$

and can be applied to Eqn. 129 to construct the rate equation for the species concentration, where  $v_i$  is the standard rate equation as discussed in section 2.3.3.

An example of a cross compartment unimolecular reaction is the dissociation of the plasma membrane localised insulin receptor - insulin receptor substrate complex. The elementary rate equation would look like Eqn. 128 with  $k_s$  instead being the unimolecular rate constant,  $r_i$  would be one (as the reaction is never limited by geometry) and  $V_1$  would be the volume where the complex resides. The insulin receptor would remain in the plasma membrane while the insulin receptor substrate would dissociate into the cytosol. In this case,  $V_i$  is always the volume of the reactant.

The reverse case of complex association deserves some additional attention due to the choice required for the characteristic parameters. The two reactants insulin receptor and insulin receptor substrate originate from different compartments with different volumes. So what is the appropriate choice for  $V_i$  and  $r_i$  for this association reaction? Considering the nature of bimolecular reaction, the reaction can only occur in some volume where both species are present, and this is physically a thin layer surrounding the interface between the two compartments. From this consideration,  $r_i V_i$  could define the thin volume surrounding the interface. Using the definition of  $r$  we introduced previously, we can define the volume  $V_i$  to be the reacting volume when the geometric layout of the system is optimised, i.e.  $r = 1$ . This means the entire volume of one of the two compartments is in the overlapping region. This can only be the smaller of the two compartment volumes. Thus for bimolecular reactions,  $V_i = \min(V_1, V_2)$  and  $0 \leq r \leq 1$ .

### 5.2.3 Compartmentalising the dQSSA

To determine the correct implementation of compartmentalisation with the single substrate dQSSA enzyme kinetic model, we will briefly reanalyse the derivation of the dQSSA (Eqn. 31 and 52-53)



using the formalism presented in Eqn. 129 and 130. The original quasi-steady state is:

$$[\dot{E}S] = \frac{1}{V_c} \left( r_a V_0 k^a [E][S] - V_c (k^c + k^d) [ES] \right) = 0 \quad (131)$$

$$[ES]_Q = \frac{V_0 [E][S]}{V_c K^m} \quad (132)$$

$$[\dot{E}S]_Q = \frac{1}{V_c} \left( \frac{V_0 r_a}{K^m} [S][\dot{E}] + \frac{V_0 r_a}{K^m} [E][\dot{S}] \right) \quad (133)$$

$$[\dot{S}] = \frac{1}{V_s} \left( -\frac{V_0 r_a}{K^m} [S][\dot{E}] - \frac{V_0 r_a}{K^m} [E][\dot{S}] + k_2 V_p [P] \right) \quad (134)$$

$$[\dot{S}] + \frac{V_0 r_a}{K^m V_s} [S][\dot{E}] + \frac{V_0 r_a}{K^m V_s} [E][\dot{S}] = \frac{k_2 V_p [P]}{V_s} \quad (135)$$

$$[\dot{E}] + \frac{V_0 r_a}{K^m V_E} [S][\dot{E}] + \frac{V_0 r_a}{K^m V_E} [E][\dot{S}] = 0 \quad (136)$$

Where  $V_0 = \min(V_s, V_E)$  as per the bimolecular implementation of compartmentalisation. From this analysis it can be seen that the effect of compartmentalisation is that  $K^m$  is modified by the relative size of the compartments involved. However, what is interesting is the modified  $K^m$  is no longer symmetric between the enzyme, substrate or complex rate equation.

### 5.3 CREATION OF MODELLING ALGORITHM

With the framework for compartmentalisation established, it is now possible to develop the algorithm for solving arbitrarily large biochemical models. The algorithm we created, called SigMat, was developed in the MATLAB environment, and is based on the dQSSA model presented in Eqn. 78. The dQSSA with compartmentalisation is:

$$\dot{x}_i = (\delta_{ij} + \mathbf{G}_{ijk} x_k)^{-1} [\mathbf{V}_{jlm} x_l x_m + \mathbf{W}_{jli} x_l + \sigma_j(t) + f_j(t, \vec{x})] / V_i \quad (137)$$

The goal of the algorithm is to parse a model that is written in a similar way to an SBML model file, then dynamically parse it to construct an ODE in the form of Eqn. 137 and finally solve it correctly by implicitly including any necessary initialisation steps (as describe in section 3.4.3). As part of the model parsing and matrix

generation steps, it is necessary to also determine the components of the matrix elements that are defined as free parameters in the model. Thus these need to be efficiently isolated from the matrix as a whole. In this section we will describe the SigMat model description convention. This will be subsequently referred to as the SigMat format with models described using the convention called SigMat models).

Fig. 29 outlines the components of a SigMat model and how information about system architecture and rate laws used is transmitted in the algorithm. In this section, we discuss the motivation and implementation of these various components and how they are used to construct the differential equations governing the input model, with an accompany wrapping script that enables the differential equation to be solved with a varied of ODE solvers.

### 5.3.1 Layout of SigMat Model Files

The first thing we will discuss in this section is the SigMat format. The SigMat format is largely motivated by the SBML format. This is because the SigMat algorithm is ultimately designed to be compatible with SBML models used by other biochemical modelling algorithms that does not currently utilises the dQSSA model. So in this section we discuss the fundamental components of the SBML format and then discuss the similarities and differences with the SigMat format.

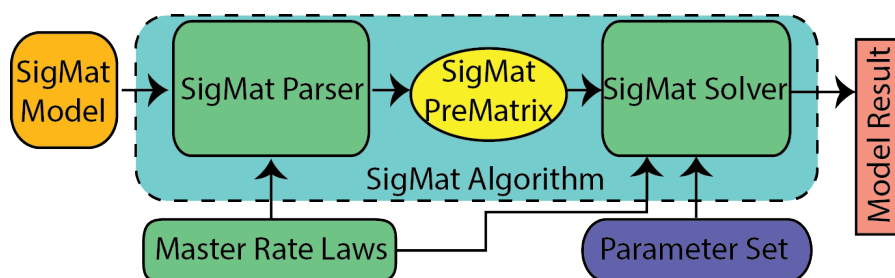


Figure 29: Outline of the SigMat algorithm. The components within the dashed box are components within the algorithm. Components outside are passed as inputs to the algorithm and can be customised.

#### 5.3.1.1 *Overview of the SBML format*

The SBML defines some key attributes that describes the physical environment, interactions, and metrics that are simulated by the model [181]. Some of the key ones are:

- Species - These are the biochemical species used in the model.
- Parameters - These list and define the parameters in the model and can either be global to the whole model or localised to single reactions.
- Compartments - These are places species can reside in. This can either be a physical compartment or a region species can become localised in.
- Reactions - These define how species interact with each other.
- Initial Assignment - Defines the initial conditions of the model in terms of parameters and concentrations.
- Rules - Mathematical relationships between parameters. E.g. parameters may be multiplicative factors of one another. Enables modelling of dynamic changes to compartment size and rate parameters.
- Constraints - Boundaries for parameters in the model for fitting purposes.

The features included are very comprehensive with some of them fulfilling more niche applications than others. As such, for our algorithm, we focused on implementing the key descriptions. These were initially species, parameters, compartments and reactions.

Species are all distinct molecular species (e.g. proteins or complexes) and have the additional attributes of compartment ID (defines where it belongs) and either initial amount or concentration (which is converted to amounts using the compartment size).

Parameters are any other variables such as rate constants or binding/dissociation constants. This contains no other attributes except the parameter's value.

Compartments define distinct locations in the cell and have the attribute dimensions (e.g. 3 dimensional compartments like

the cytosol or 2 dimensional compartments like the plasma membrane) and size (in terms of its dimension, e.g. the volume or the area).

Finally, reactions defines interactions between species using the following attributes:

- List of reactants. Defined by a numerical ID and stoichiometry that is consumed.
- List of products. Defined by a numerical ID and stoichiometry that is produced.
- List of modifiers. Species that affect reactions but are not consumed, such as enzymes and inhibitors. Defined by a numerical ID and stoichiometry when used.
- KineticLaw. Rate law used for the reaction.
- Reversibility. Boolean. Defines if rate parameters can be negative. Note: does not create the opposing reaction explicitly.
- Compartment ID. Compartment where the reaction occurs.

More detailed information on the SBML format is available through the technical manual [181].

#### 5.3.1.2 Overview of the SigMat model structure

The model structure shown in Fig. 30 outlines the MATLAB adaptation of the SBML format used in the SigMat algorithm. The algorithm needs to perform a few functions:

- determine and define the free parameters required based on the model structure
- convert the structure into a kinetic equation dynamically
- enable assignment of these free parameters back into the model when a parameter set is given as a vector
- solve the parameterised time course

There are a number of simplifications between SBML and SigMat model definitions arising from different design motivations. While SBML is designed with completeness in mind, SigMat was

<b>Species</b> Type: Cell Array Syntax <pre>modSpc={string, string ,vector;         string, string ,vector;}</pre> <hr/> Meaning Species    Compartment    Initial Name       Name       Conc <hr/> Example <pre>modSpc={'A','PM' ,NaN;         'B','Cytosol',0.0;}</pre>																													
<b>Compartments</b> Type: Cell Array Syntax <pre>modComp={ string , vector ;            string , vector ;}</pre> <hr/> Meaning Compartment    Compartment Name           Size <hr/> Example <pre>modComp={'Cytosol',1.0;           'PM' ,0.2;}</pre>																													
<b>Reaction</b> Type: Structure <table> <tr> <th>Syntax</th><th>Meaning</th><th>Example</th></tr> <tr> <td><code>rxn(end+1).label = string</code></td><td>Reaction Description</td><td>'A → φ B'</td></tr> <tr> <td><code>rxn(end).sub = cells of strings</code></td><td>List of Substrates</td><td>{'A'}</td></tr> <tr> <td><code>rxn(end).prod = cells of strings</code></td><td>List of Products</td><td>{''}</td></tr> <tr> <td><code>rxn(end).enz = string</code></td><td>Enzymes/Modifiers</td><td>'B'</td></tr> <tr> <td><code>rxn(end).k = vector</code></td><td>Rate constant</td><td>0.1</td></tr> <tr> <td><code>rxn(end).Km = vector</code></td><td>Equilibrium constant</td><td>1x10<sup>2</sup></td></tr> <tr> <td><code>rxn(end).n = vector</code></td><td>Hill coefficient</td><td>[NaN 0.1 10]</td></tr> <tr> <td><code>rxn(end).r = vector</code></td><td>Interface coefficient</td><td>Blank</td></tr> </table>			Syntax	Meaning	Example	<code>rxn(end+1).label = string</code>	Reaction Description	'A → φ B'	<code>rxn(end).sub = cells of strings</code>	List of Substrates	{'A'}	<code>rxn(end).prod = cells of strings</code>	List of Products	{''}	<code>rxn(end).enz = string</code>	Enzymes/Modifiers	'B'	<code>rxn(end).k = vector</code>	Rate constant	0.1	<code>rxn(end).Km = vector</code>	Equilibrium constant	1x10 <sup>2</sup>	<code>rxn(end).n = vector</code>	Hill coefficient	[NaN 0.1 10]	<code>rxn(end).r = vector</code>	Interface coefficient	Blank
Syntax	Meaning	Example																											
<code>rxn(end+1).label = string</code>	Reaction Description	'A → φ B'																											
<code>rxn(end).sub = cells of strings</code>	List of Substrates	{'A'}																											
<code>rxn(end).prod = cells of strings</code>	List of Products	{''}																											
<code>rxn(end).enz = string</code>	Enzymes/Modifiers	'B'																											
<code>rxn(end).k = vector</code>	Rate constant	0.1																											
<code>rxn(end).Km = vector</code>	Equilibrium constant	1x10 <sup>2</sup>																											
<code>rxn(end).n = vector</code>	Hill coefficient	[NaN 0.1 10]																											
<code>rxn(end).r = vector</code>	Interface coefficient	Blank																											
<b>Rate Law</b> Type: String <table> <tr> <th>Syntax</th><th>Meaning</th><th>Example</th></tr> <tr> <td><code>rateLaw = string</code></td><td>Function name of rate law file</td><td><code>rateLaw = odeKinetic</code></td></tr> </table>			Syntax	Meaning	Example	<code>rateLaw = string</code>	Function name of rate law file	<code>rateLaw = odeKinetic</code>																					
Syntax	Meaning	Example																											
<code>rateLaw = string</code>	Function name of rate law file	<code>rateLaw = odeKinetic</code>																											

Figure 30: Outline of the way models are constructed in SigMat.

designed with simplicity and streamlining in mind. The most significant change is in the way system parameters are handled. In the SBML format, parameters, rules, initial assignments and constraints are defined separately. These are streamlined in SigMat and handled together. SigMat has retained the SBML approach of classifying properties such as concentration, compartment size and kinetic parameters all as parameters. This enables them to all be treated identically. In SigMat, parameters are assigned as vectors that follow a set of rules. These define whether the parameter is fixed or free, and their constraints and rules. The specific syntax for this is described in table 6.

Following that, the concept of IDs have been completely removed because index associate with each system attribute can be used as an identifier, making the inclusion of a separate identifier redundant. Reversibility is also removed because reactions should be defined as elementary reactions, which means reversibility can be explicitly implemented. The dimensionality and units attributes were also removed as it has no functional significance on the mathematics. In SigMat, the assumption is units are self consistent within a model.

Based on these design decisions, the final SigMat framework that was adopted is shown in Fig. 30. Species and compartments are created using cells arrays where each row in the array corresponds to a distinct species or compartment. The first column is the name of the species/compartment and each subsequent column encodes the various associated attributes. Compartments, as with the SBML structure, has the additional numerical attribute size. Species on the other hand, analogous to SBML, has the additional compartment attribute (linked to the compartment name), and the initial assignment numerical attribute.

Reactions are encoded using a structure instead. This is because the content of the structure is more complex and requires more flexibility. As there are no fields that are universally required for all reactions (e.g. a synthesis reaction has no substrate, a degradation reaction has no product and elementary reactions do not have enzymes). In order to cover all possible reactions, all fields are initialised for all reactions as part of the variable structure. For the reaction parameters, the available fields are, these are:

- Reaction rates (from elementary reactions such as  $k^a$ ,  $k^d$  of Eqn. 9)
- Equilibrium constant (such as dissociation constants  $K^i$  of Eqn. 101 or Michaelis constants  $K^m$  of Eqn. 15)
- Hill coefficient (as part of the hill function)
- Interface coefficient ( $r$  are defined in the compartmentalisation formalism derived in section 5.2.1)

The inclusion of all fields for all reactions are important in the dynamic construction of the system differential equation. The user is only required to complete the fields that are relevant for

Table 6: Rules relating to how parameters should be assigned and how they are interpreted by SigMat.

Elements in vector	1	2	3	4
Parameter Interpretation	Unconstrained free or fixed	Grouped parameters related by a factor.	Ungrouped free but constrained	Constrained reference parameter of a group
Syntax	- NaN or positive	- First value is NaN or a number - Second value must be a positive integer	- First value is NaN - Second and third values must be a positive	- First value is NaN - Second value must be a positive integers - Third and fourth values must be a positive
Syntax Definition	- If NaN, parameter is free and unconstrained. - If number, then fixed parameter	- If first value NaN, it is the group's reference parameter and is unconstrained. Else the number is the multiplicative factor that parameter is to the reference. - The second value defines the group the parameters is in.	- Second and third values defines the range the parameter is constrained within.	- The second value defines the group of parameters it is in. - Third and fourth values defines the range the parameter is constrained within.

that particular reaction after which SigMat handles the rest. This will be discussed in the following section 5.3.2.2.

And finally, a link to the rate law support function is included for users that wish to compile their model with a custom dynamic equation inference algorithm. This is implemented through the  $\vec{f}(t, \vec{x})$  term in the dQSSA equation (Eqn. 78).

### 5.3.2 Dynamic Inference of Rate Equation and its Matrix Composition

To generate the system of governing differential equations from a model file, the final algorithm performs the following steps:

1. Converting the human readable components of the model, more specifically reaction equations, into a mathematical reference appropriate for computation
2. Parsing the model file to a model structure that can dynamically determine the rate equations necessary for each reaction
3. Populating the matrices in Eqn. 137 based on the rate equations inferred from the model
4. Placing the matrices within the generalised differential equation
5. Numerically solving the differential equation

The following sections will describe each of these steps in turn.

#### 5.3.2.1 Model Interpretation and Parsing

Model interpretation is the conversion of the model from human readable elements (such as compartment name or reaction equations) to a mathematical reference (such as the compartment index in its cell array). This enables species and compartments to be referenced by their indices rather than by multiple string comparisons which is less computationally efficient.

Compartments are the first attributes to be processed as they have no other dependencies. The conversion is simply performed by isolating the first column of the array and identifying as the



compartment name array. The Second column is then converted to a numerical array.

Species are processed next as they depend only on the compartment attribute. In the model file, compartments are referred by name. Using the array of compartment names, the second row of the species array is converted to a compartment index. This results in a cell array structure that has the species name as its first column, compartment index in the second column, and initial concentration in the third column. Each of these arrays were then separated out and the latter two converted to numerical arrays. The names array will be referred to as the master species list.

Reactions are processed in a different way. This is because interpretation of the reaction involves classifying the rate equation. Each reaction is processed separately. The first step is to convert all species to their respective species index by comparing it to the master species list. Then the reaction type is inferred based on the follow attributes within each reaction:

- The number of substrates
- The number of modifiers such as enzymes or inhibitors
- Whether rate/equilibrium/Hill coefficients/other parameters are defined as necessary parameters

From there, the reaction velocity is determined based on the reaction type and its associated rate law in the rate law file. It should be noted that users are able to create their own rate law file. The description will only cover the default rule set prepacked with the SigMat algorithm, also known as the odeKinetic rule set. Currently, the reaction types covered in the odeKinetic rule set are the elementary reactions synthesis, degradation (zeroth order), unimolecular (first order), bimolecular (second order) and elementary reaction like simplifications of enzyme kinetic reactions, and more complex enzyme kinetic reactions: dQSSA model and Hill functions. Table 7 summarises the inference logic and the resulting reaction velocities that are relevant for each reaction type.

From this reaction velocity, the reaction rates relevant to the system can be compiled by subtracting and adding the reaction velocity to all substrates and products respectively. This gives a list of reaction velocities that need to be combined to form the

Table 7: Rules relating to reaction rates are inferred from the structure of the reaction that is parsed by SigMat. Dashed line separates groups of reactions that use rate laws of the same form.

	Substrate	Enzymes	Products	k	K <sub>m</sub>	n	rate law
Synthesis	✗	✗	> 1	✓	✗	✗	$v = k$
Enzymatic synthesis High K <sub>m</sub>	✗	✗	> 1	✓	✗	✗	$v = k[E]$
Degradation/ Conversion/ Dissociation	1	✗	✗ 1 > 1	✓	✗	✗	$v = k[S]$
General enzymatic	1	✓	> 0	✓	✓	✗	$v = \text{dQSSA}$
Hill function	1	✓	> 0	✓	✓	✓	$v = \frac{k[E]^n}{[E]^n + [K_m]^n}$
High K <sub>m</sub> enzymatic	1	✓	> 0	✓	✗	✗	$v = k[S][E]$
Association	2	✗	1	✓	✗	✗	$v = k[S_1][S_2]$

system rate equation. In SigMat, instead of compiling the list of reactions velocities as functions, they are converted to matrix notation, with each rate parameter fitting inside a tensor for each reaction type. The advantage of this approach is instead of storing the reaction velocities as explicit equations which would need to be parsed as text using a complex algorithm, the reactions can now be stored as arrays and constructed using matrix operations. The general vectorised equation is implemented as below using tensor notation, and incorporating the modifications required to account for compartmentalisation:

$$V_j \dot{x}_j = \overbrace{(\delta_{jk} + \mathbf{G}_{ijko} x_k / V_o)^{-1}}^{\text{dQSSA}} \left( \overbrace{\mathbf{Y}_{ilmo} x_l x_m V_o}^{\text{2nd order}} + \underbrace{\mathbf{Z}_{ilo} x_l V_o}_{\text{1st order}} + \underbrace{\sigma_{io} V_o}_{\text{0th order}} + \underbrace{\mathbf{H}_{ilo} x_l V_o}_{\text{Hill}} + \underbrace{f_i(t, x)}_{\text{custom}} \right) \quad (138)$$

where  $V$  is the vector of compartment size for each species. The remaining are as defined in Eqn. 78.

Section 3.4.2 has already covered how the  $G$  tensor (which covers the quasi-steady state component) is populated for reactions using the dQSSA model. The general rules for the elements of the remaining tensors are:

Table 8: Index assignment for matrices within Eqn. 138 for each reaction type

Reaction type	index for l	index for m	index for o	value	Matrix
Zeroth order	$\times$	$\times$	product	$k_0$	$\sigma_j$
First order	reactant	$\times$	substrate	$k_1$	$\mathbf{V}_{j\text{lm}o}$
Second order	substrate 1	reactant 2	$\min(V_1, V_2)^*$	$k_2$	$\mathbf{W}_{j\text{lo}}$
Hill function	substrate	$\times$	$\min(V_1, V_2)^*$	$\frac{k[E]^n}{[E]^n + K_m}$	$H_{i\text{lo}}$
dQSSA	complex	$\times$	$\min(V_1, V_2)^*$	$k_{\text{cat}}$	$\mathbf{W}_{j\text{lo}}$

\*smaller volume of reactants if from different compartments

- within a term (e.g.  $\mathbf{N}_{j\text{lo}}x_lV_o$ ), the  $j$  index is the species for which this reaction actually acts on. It is not  $i$  because  $i$  is the quasi-steady state rate of change, after the non-quasi-steady state rate change is modified by the dQSSA prefactor.
- the sign of the element is positive for products and negative for reactants.

Beyond this, there are specific rules regarding the other indices that are different for different reaction types. These are outlined in table 8. Once the matrices are constructed in accordance to the specified rules, matrices are non-dimensionalised for time. More specifically this includes but is not limited to the following transformation for the first and second order rate parameters (such as  $k^a$  and  $k^d$ ):

$$\hat{k} = k(t_{\text{end}} - t_0) \quad (139)$$

where  $\hat{k}$  indicates the non-dimensional version of their non-hat counterpart. In terms of Eqn. 138, this implies that elements within tensors  $\mathbf{Y}$ ,  $\mathbf{Z}$ ,  $\sigma$  and  $\vec{f}(t, x)$  need to be non-dimensionalised. Non-dimensionalisation is performed to prevent floating point errors that may be generated in the numerical simulation process.

The dynamic equation is solved in three stages. First, all complexes from the initial condition are dissociated. This needs to be done manually. For example, if the biological initial condition is  $[S] = 1$ ,  $[E] = 2$  and  $[ES] = 0.3$ , where  $S$  is the substrate,  $E$  is its enzyme and  $ES$  is the enzyme-substrate complex they form, then  $[ES]$  must be manually added to  $[S]$  and  $[E]$  to form the

equilibration initial condition for the simulation, which is then  $[S] = 1.3$ ,  $[E] = 2.3$  and  $[ES] = 0$ . The equilibration initial condition is then entered into SigMat. Next, the simulation initial condition is solved by setting all matrices except  $G$  to zero and adding the equilibration initial conditions to the simulation with a time profile shown in Eqns. 97, and solving between the times  $t = [0, 1]$  (because as explained in section 4.2.1, this time interval allows the simulation to fully equilibrate, given the definition of the time profile used to add the equilibration initial condition into the system (Eqns. 97). Finally, all matrices are set to their nominal values and the simulation run from  $\hat{t} = [0, 1]$  because the system parameters have already been non-dimensionalised within this interval.

### 5.3.2.2 *Dynamic Adaptation of Matrices to Free Parameters*

The previous section covered the construction of the matrices from the system's list of reactions. However, we have not covered free parameters within the model. To do this requires a few additional steps in the matrix compilation process. These are:

1. Determining the free parameters in the model
2. Constructing matrices without committing a value by creating pre-matrices
3. Efficiently apply the values into the pre-matrices in a computationally efficient way

Free parameters can be determined using the criteria given in Table 6. When the free parameters are identified, the model parsing stage is able to determine which element within which matrix the parameter belongs. This is all done by referencing the matrix element to the index of the vector of free parameter, rather than directly assigning the parameter. This enables the pre-matrices to be pre-generated without requiring an explicit parameter set to be known. This means SigMat models only need to be parsed once rather than everytime a new parameter set is required, leading to improved computational efficiency. Thus, the pre-matrix needs to include

Table 9: Overview of the layout of the Matrix value pre-matrix. Cells marked as Target Rate, Reactant, Substrate and Enzyme indicate their indices. Crosses indicate unused columns.

Reaction type	MV pre-matrix column index						
	1	2	3	4	5	6	7
Zeroth order	Target Rate	$k_0$	$\times$	$\times$	$\times$	$\times$	$\times$
First order	Target Rate	Reactant	$r$	$k_1$	$\times$	$\times$	$\times$
Second order	Target Rate	Reactant 1	Reactant 2	$r$	$k_2$	$\times$	$\times$
Hill function	Target Rate	Substrate	Enzyme	$r$	$k$	$K_m$	$n$
dQSSA	indices as explained previously			$r$	$K_m$	$\times$	$\times$

1. Matrix value (MV) pre-matrix: a pre-matrix that encodes the element reference for the parameter and information relating to the actual parameter value.
2. Parameter Index (PI) pre-matrix: a pre-matrix that encodes a reference to an index of the free parameter vector

The matrix value pre-matrix is arranged such that columns indicate different features of a single reaction, while rows encode different reactions (or different targets of a single reaction). In this case, features refers to things like the indices  $i$ ,  $l$  or  $m$  ( $o$  is not included as it can be inferred from the other indices), and parameter values. More specifically, the layout of the pre-matrices are shown in table 9. The definition of the number in the parameter values field in the Matrix value pre-matrix varies depending on whether the parameter is free or not. If the parameter is fixed, the value in the pre-matrix is the final value used. On the other hand if the parameter is free, then the value in this matrix is the multiplicative factor (which is relevant if the parameter is a multiplicative factor of another parameter, as discussed previously).

The Parameter Index pre-matrix mirrors the Matrix Value pre-matrix. For elements in the MV that represent state indices, the equivalent element in the PI is NaN. This indicates the equivalent MV element is not linked to a value in the vector of free parameters. On the other hand, where for elements in the MV that represent parameters, the corresponding element in the PI element contains the index in the free parameter vector which it is linked

to. If the element is zero, the parameter is fixed (since a vector index in MATLAB cannot be zero). With this convention, free parameters can be replaced inside the MV, and the multiplicative factors for grouped parameters be accounted for using the efficient logical indexing capability of MATLAB. The MATLAB code used to perform this can be summarised as:

```
mask = PI~=NaN|PI~=0
MV_Sub[mask] = MV[mask] .* p[PI[mask]]
```

Where  $MV_{Sub}$  is the MV matrix with the required parameter vector values substituted in, and  $p$  is the set of input parameters. The mask only selects elements that contain free parameters for replacement. Basically what this operation does is only for free parameters, select the right parameter from the input list, and then multiply it by the necessary multiplicative factor.

The ODE can then be then compiled by either constructing a real matrix from the MV prematrix, and then using matrix operations to calculate the differential equation, for example in the case of first order reactions the following MATLAB code might be used:

```
dxdt = sparse(MV_Sub[:,1],MV_Sub[:,2],MV_Sub[:,3])*(x.*V)
```

Where  $MV_{Sub}$  is the first order reaction matrix value prematrix. The sparse function constructs the pre-matrix into a true matrix. Alternatively they can be compiled using the columns in the MV pre-matrix as reference indices (e.g. species concentrations or compartment size). For example a second order rate velocity would be compiled by the following MATLAB code:

```
dxdt = sparse(MV_Sub[:,1],ones(size(MV_Sub[:,1])),...
             min(V[MV_Sub[:,2]],V[MV_Sub[:,3]]) *...
             MV_Sub[:,4] * x[MV_sub[:,2]] * x[MV_sub[:,3]])
```

The sparse function in this case generates the vector representing the ODE directly.

### 5.3.3 Validation of the SigMat Algorithm

To ensure that the algorithm parses, compiles and simulates the model correctly, a series of test cases were created. From chapter 4 we have established that the degree of accuracy the dQSSA

can produce compared to a mass action implementation and the parameter conditions required to reproduce this performance. Thus, if the SigMat algorithm correctly implements the dQSSA, then we would expect the SigMat algorithm to have the same performance. Afterall, the SigMat algorithm is simple a combination of the mass action model and dQSSA for larger networks. Thus a number of test cases were generated where the SigMat model was constructed in parallel with a manually written equivalent. The two models were simulated together and then compared. One test case was designed for each of the reactions described in Table 7 except for the enzyme kinetic model, where an irreversible case and a reversible case was designed to test the two facets of enzyme kinetic models. This was designed to ensure that each reaction is modelled by the algorithm correctly.

Finally, the complex model used to validate the dQSSA in section 4.3.1 was converted to a SigMat model and compared to the manually constructed mass action based series of differential equations. This was designed to verify that the algorithm correctly combines multiple reactions together into a interconnected biochemical system.

In each of these test cases, the models were designed to pass in parameters sets as an input. These parameters sets were then randomised in order to test the performance of the algorithm under all parameter conditions subject to the parameter limits for the dQSSA as described in section 4.2.1. Rate parameters and concentration parameters were generated using the random generator:

$$\log(\phi) = \text{rand}[-2, 2] \quad (140)$$

because rate parameters are scale parameters, and a four order of magnitude range is typically sufficient to cover various extremes in system behaviour, particularly in the location of the system equilibrium. The Hill coefficient was generating using the generator:

$$n = \text{rand}[1, 4] \quad (141)$$

and volumes were generated using the generator:

$$V = \text{rand}[0, 1] \quad (142)$$

Table 10: Spread of the log transformed, normalised error for each reaction type when simulated with sigMat compared to when modelled with a manually compiled mass action model. A value of -2 equates to an error of 1%, with lower values equating to lower errors.

Reaction	$\min_{\text{err}}$	$\mu_{\text{err}} - \sigma_{\text{err}}$	$\mu_{\text{err}}$	$\mu_{\text{err}} + \sigma_{\text{err}}$	$\max_{\text{err}}$
Synthesis	-Inf	-15.9	-15.5	-15.3	-14.9503
Conversion	-3.35	-2.84	-2.58	-2.32	-2.02
Dissociation	-3.49	-2.95	-2.69	-2.43	-2.18
Degradation	-3.42	-2.72	-2.55	-2.39	-2.30
MM Synthesis	-6.77	-4.42	-3.83	-3.23	-2.78
Association	-3.48	-3.08	-2.80	-2.52	-2.18
High $K_m$ Enzymatic	-15.78	-7.05	-3.76	-0.48	-2.21
High $K_m$ Enzymatic degradation	-5.17	-3.12	-2.68	-2.23	-2.23
dQSSA irreversible	-6.00	-4.80	-4.24	-3.67	-2.76
dQSSA reversible	-6.02	-4.80	-4.18	-3.58	-2.38
Hill function	-5.42	-3.31	-2.78	-2.24	-1.77
Complex Model	-2.88	-2.66	-2.46	-2.26	-1.98

The differential equations were then solved using the algorithm options described in section 4.2.1 such that the previous results would be directly comparable to these results.

The accuracy for all models except for the dQSSA model was evaluated using the following equation:

$$\text{err}_{\text{tot}} = \log_{10} \left( \sqrt{\sum_i \sum_t \left( \frac{Y_{\text{MA}} - Y_{\text{Alg}}}{\max_t(Y_{\text{MA}})} \right)^2} \right) \quad (143)$$

where  $t$  is across the time points and  $i$  is across the states. Subscript Alg indicates the output generated by the SigMat algorithm. The accuracy for the dQSSA models are generated using the equation given in Eqn. 100, i.e. the same equation used to evaluate the accuracy of the dQSSA when it was validated *in silico*.

The results of these test cases are shown in Table 10. From these results, it can be seen that the models generated by the SigMat algorithm has negligible error compared to their manually written



counter parts. This indicates that the SigMat algorithm is able to automatically generate system equations without error.

## 5.4 NETWORK MOTIFS IN SIGNALLING

With the SigMat toolbox completed, it is possible to efficiently utilise the dQSSA framework to investigate signalling pathways. Although the main procedure for investigating signalling pathways is through the cyclical process of model construction, model analysis, experimental validation and biological analysis, it is possible to perform smaller theoretical studies of distinct network motifs [184]. Investigation of these are interesting as they aid in the interpretation and prediction of network behaviours through modularisation of pathways into these discrete motifs. From there, the system wide behaviour would be constructed by combining the effects of the discrete motifs.

In this section, we will describe two motifs that have been encountered in this thesis and explore their behaviours as they are of interest to signalling pathways in general.

### 5.4.1 Merging of Signalling Inputs

A common feature of signalling pathways appears to be their hour glass like structure which we described in Fig. 4[16]. One of the key component of this behaviour is the existence of various distinct receptors (with their respective ligand) whose downstream targets are identical [185, 186]. This is unusual because the distinction between receptor imply that they should perform different functions, even though they activate the same pathways [185, 186]. In the case of receptor tyrosine kinases (RTKs), their downstream targets commonly activate the PI3K and Akt pathways [185, 186]. Furthermore, most RTKs autophosphorylate (and thus activate) itself upon binding to their ligand. However, the insulin receptor, a member of the RTK family, is unique in is separated into two components with the receptor tyrosine kinase component residing within the Insulin Receptor Substrate (IRS) protein [185].

To investigate the difference between having a separate and embedded tyrosine kinase component from the receptor itself, we

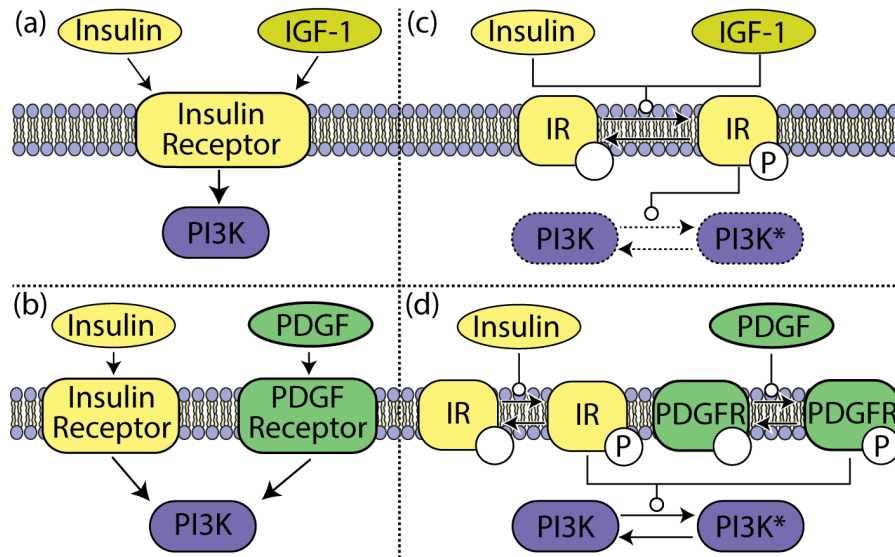


Figure 31: Model layout of (a) a single receptor that is the target of two ligands (b) two receptors with distinct ligands that target a common substrate. Panels (c) and (d) presents the reaction schemes used to model the network motif. In (c) the PI3K components (dashed line components) are not explicitly modelled because it is a direct substrate of pIR so is assumed to simply transfer the signal of pIR forward. All enzymatic reactions are modelled using the dQSSA raw law. All other reactions are modelled with first order mass action rate law.

constructed two models. One with two ligands binding to a single receptor, and a second with two ligands binding to two separate receptors which then activate the same downstream target. We then tweaked the parameters and ligand concentration to investigate the downstream effects of these perturbations. Biological examples of these two models are illustrated in Fig. 31. The SigMat model used for these two cases are:

<pre> Model: Single Receptor modSpc ={'Insulin'      , 'Cytosol', 0 ;         'IGF-1'        , 'Cytosol', 0 ;         'IR'           , 'Cytosol', 1 ;         'pIR'          , 'Cytosol', 0 ;         'Phosphatase'  , 'Cytosol', NaN}; modComp = {'Cytosol', 1}; </pre>	<pre> Model: Double Receptor modSpc ={'Insulin'      , 'Cytosol', NaN;         'IR'           , 'Cytosol', 1 ;         'pIR'          , 'Cytosol', 0 ;         'PDGF'         , 'Cytosol', NaN;         'PDGFR'        , 'Cytosol', 1 ;         'pPDGFR'       , 'Cytosol', 0 ;         'PI3K'         , 'Cytosol', 1 ;         'pPI3K'        , 'Cytosol', 0 ;         'Phosphatase'  , 'Cytosol', 1 }; modComp = {'Cytosol', 1}; </pre>
--	---

```

rxn(end+1).label = 'IR -> pIR';      | rxn(end+1).label = 'IR -> pIR | Insulin';
rxn(end).sub = 'IR';                  | rxn(end).sub = 'IR';
rxn(end).prod= 'pIR';                 | rxn(end).prod= 'pIR';
rxn(end).enz = 'Insulin';             | rxn(end).enz = 'Insulin';
rxnv(end).k = NaN;                   | rxn(end).k = NaN;
rxn(end).Km = NaN;                   | rxn(end).Km = NaN;
rxn(end+1).label = 'IR -> pIR';      | rxn(end+1).label = 'pIR -> IR';
rxn(end).sub = 'IR';                  | rxn(end).sub = 'pIR';
rxn(end).prod= 'pIR';                 | rxn(end).prod= 'IR';
rxn(end).enz = 'IGF-1';               | rxn(end).enz = 'Phosphatase';
rxn(end).k = NaN;                    | rxn(end).k = NaN;
rxn(end).Km = NaN;                   |
rxn(end+1).label = 'pIR -> IR';      | rxn(end+1).label = 'PDGFR -> pPDGFR | PDGF';
rxn(end).sub = 'pIR';                 | rxn(end).sub = 'PDGFR';
rxn(end).prod= 'IR';                 | rxn(end).prod= 'pPDGFR';
rxn(end).enz = 'Phosphatase';         | rxn(end).enz = 'PDGF';
rxn(end).k = NaN;                    | rxn(end).k = NaN;
                                     | rxn(end).Km = NaN;
                                     | rxn(end+1).label = 'pPDGFR -> PDGFR';
                                     | rxn(end).sub = 'pPDGFR';
                                     | rxn(end).prod= 'PDGFR';
                                     | rxn(end).enz = 'Phosphatase';
                                     | rxn(end).k = NaN;
                                     |
                                     | rxn(end+1).label = 'PI3K -> pPI3K | pPDGFR';
                                     | rxn(end).sub = 'PI3K';
                                     | rxn(end).prod= 'pPI3K';
                                     | rxn(end).enz = 'pPDGFR';
                                     | rxn(end).k = NaN;
                                     | rxn(end).Km = NaN;
                                     | rxn(end+1).label = 'PI3K -> pPI3K | pIR';
                                     | rxn(end).sub = 'PI3K';
                                     | rxn(end).prod= 'pPI3K';
                                     | rxn(end).enz = 'pIR';
                                     | rxn(end).k = NaN;
                                     | rxn(end).Km = NaN;
                                     | rxn(end+1).label = 'pPI3K -> PI3K';
                                     | rxn(end).sub = 'pPI3K';
                                     | rxn(end).prod= 'PI3K';
                                     | rxn(end).enz = 'Phosphatase';
                                     | rxn(end).k = NaN;

```

The mechanism of receptor activation involves a ligand - receptor binding and dissociation steps, and an autophosphorylation step [149, 154]. Mechanistically, this is very similar to an enzyme kinetic reaction with the ligand performing like an enzyme and so some models have taken to modelling receptor activation as an enzymatic process [127]. The same enzyme kinetic assumption were made in our receptor model. To understand the general behaviours of the system independent of the actual receptor concentration, the system is non-dimensionalised to the receptor concentration. In the two receptor model, the two receptors were made the same concentration, and this concentration was set as the reference concentration for the whole system. This means

$[IR] = [PDGFR] = 1 \mu\text{M}$ . All dephosphorylation reactions are also modelled as enzymatic reaction with a large  $K_m$ . This assumes that the phosphatases have low concentration, are unregulated and constitutively active, and uniquely targets its phosphoprotein monogamously. While the study of phosphatase are still in their infancy, some phosphatase (such as PP2A) have been found to have a high concentration of few catalytic subunit [14] which combine with many lower concentration [14] regulatory subunits that confer specificity to the phosphatase's action [20]. This property of phosphatases suggests that this assumption could be true in vivo. This assumption allows the dephosphorylation reactions to be modelled as first order mass action reactions.

From there, we start with the first order approximation that the  $K_m$  of all reactions are very large (i.e.  $K_m > 100\mu\text{M}$ ). The other kinetic parameters are chosen such that effective dose at 50% activity ( $ED_{50}$ ) for the receptor is  $1 \mu\text{M}$  (this is because changes in the catalytic rate/phosphatase concentration only shifts the  $ED_{50}$  of the reaction). Finally the concentration of PI3K is made 50 times larger than the concentration of the receptors/receptor substrate such that substrate limitation effects do not become present downstream. This assumption appears to hold in some in vivo circumstances. Proteomic study of 3T3-L1 adipocytes, where signal transmission is through insulin receptor and IRS1, showed that PI3K and IRS1 abundance is equal. However, since PI3K is a membrane bound protein and IRS1 is cytosolic, it could be inferred that PI3K concentration in the membrane is much later than IRS1 in the cytosol [14].

Using this setup, the output concentration (that is active PI3K concentration) at steady state is measured under different doses of the two ligands in both models. The result of this is shown in Fig. 32. The simple addition of two separate receptors for each ligand changes the behaviour of the dose response of the downstream substrate. In the case of the one receptor system, the downstream substrate no longer becomes responsive if any of the ligands saturate it. On the other hand, the two receptor model actually sums the results of the two inputs, even if one ligand has reached a maximum level of stimulation. It must be noted that this model assumed equal concentrations of the two receptors. Although this is not expect to be true in a physiological system, changes in the concentrations on a particular receptor will simply

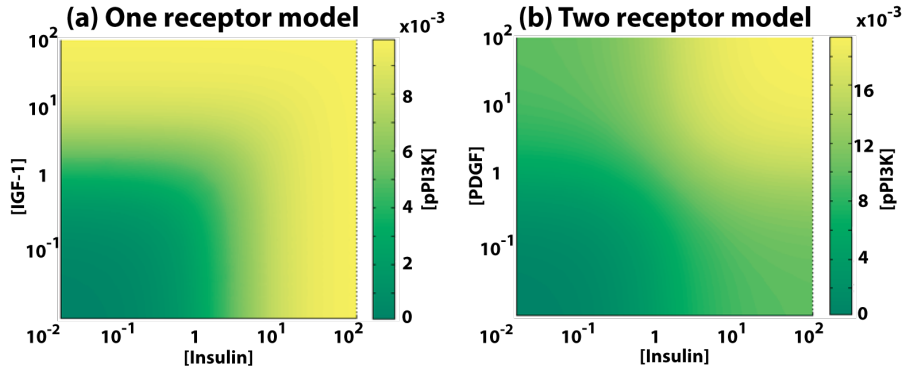


Figure 32: Result of the pPI3K output at steady state for the (a) single receptor model, and (b) two receptors model, under different doses of the their respective ligands.

shift the  $ED_{50}$ , or the sensitivity of the system to that particular ligand. Further more a high receptor concentration could cause it to shift its maximum PI3K activation capacity which can be seen as a lever that is available to the cell in deciding which activation arm is prioritised in downstream activation, such that the two inputs are not summed with equal priority. This result suggests that when multiple ligands (such as Insulin and EGF) directly target a receptor, they compete to trigger a downstream signal. On the other hand, when multiple ligands transmit signals through their own dedicated receptor (Insulin and PDGF), their individual signal will always be transmitted downstream when the signalling pathway merges. This may be an indication of how cells can transmit different signals through a common pathway without loss of fidelity.

#### 5.4.2 Sustained Response from Transient Stimulation

More recent studies of signalling pathways have explored the impact of temporal control, such as frequency modulation, on phenotypic responses [74, 187, 188]. Some of these models demonstrate that certain arms in the signalling pathway are fast acting and respond to all stimulations. Others, are slow acting and become blind to acute stimulations [74]. This is offered as a way that cells exhibit different behaviours while stimulated by the same signalling protein. An *in vivo* example of this is the stimulation of MAPK in PC12 cells. A sustained MAPK activity leads to differen-

tiation, which is expected as a sustained stimulation necessarily leads to sustained responses. However, transient response MAPK activity on the other hand leads to some sustained response that results in cell proliferation [189, 190]. In most mathematical models of these signalling systems, substrates activated by a transient stimulation would themselves respond transiently. Alternatively a sustained but binary output can be produced using a positive feedback mechanism. However, this *in vivo* example shows that there are biological circumstances where a transient stimulation can lead to a sustained response with a range of output strengths. Thus, there must be a mechanism for a system to activate and sustain the downstream components in a dose dependent manner.

Using theoretical arguments, we designed the following signalling motif as illustrated in Fig. 33, which we coin the "transient stimulation responder". In a typical enzymatic cyclic reaction, the enzymatic activation of protein is only dependent on the dose of the activating enzyme if there is a balance between the activating enzyme and a deactivating enzyme. When the stimulus is removed, the balance shifts such that the deactivating enzyme is

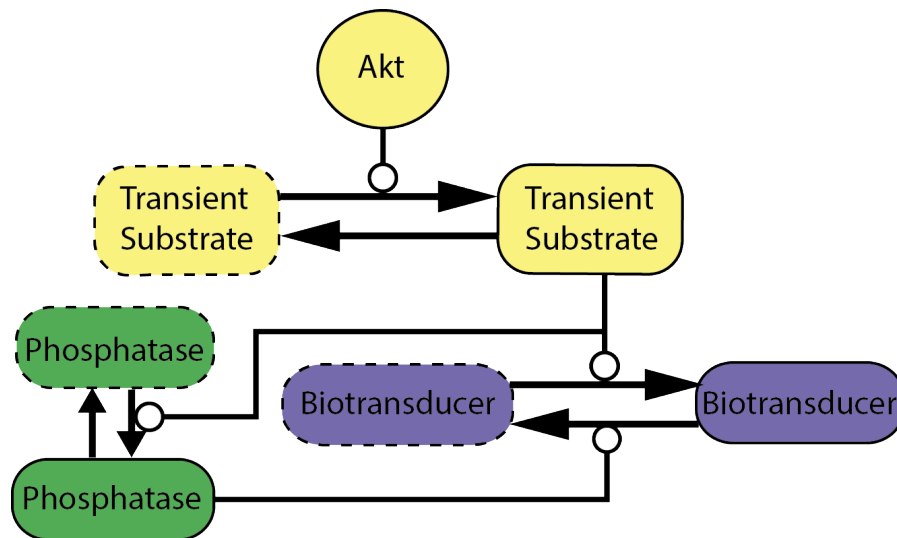


Figure 33: Illustration of the proposed signalling motif that will lead to a sustained signal upon transient stimulation. All enzymatic reactions are modelled using second order mass action rate laws. All other reactions are modelled with first order mass action rate law. Species with dashed outlines are inactive with species with solid outlines are active.

the only active enzyme in the system. Thus the final equilibrium results in all biotransducer in the system to become deactivated. For the biotransducer to remain active given a transient stimulation of the activating enzyme, the deactivating enzyme must itself be deactivated after the stimulus is removed. In the specific case illustrated by Fig. 33, the phosphatase itself must thus be regulated by the stimulus Akt.

When this is attempted in simulation, it was discovered that this needs to happen quickly such that the phosphatase cannot deactivate the target protein. Additionally, the deactivation of the first target (in the case of Fig. 33 the transient substrate) also needs to happen quickly to prevent it hyperactivating the downstream transducer while the phosphatase is deactivating. Thus this network motif exists in two regimes. The first regime is the activation and deactivation of the upstream enzymes and phosphatase. After this has happened quickly, the activation of the downstream target (the biotransducer in this case) happens at a more moderate rate that only detects the time averaged effects of the upstream components. During this phase, the concentration of the activated phosphatase needs to be near constant irrespective of the strength of the upstream signal. In other words, phosphatase activation must be saturated. This implies that the  $ED_{50}$  of the phosphatase must be far higher than the  $ED_{50}$  of the transducer. The SigMat code used to generate this model is:

```
modComp = {'Cyto', 1};
modSpc = {'I'      , 'Cyto', 0;
          'A'      , 'Cyto', 1;
          'pA'     , 'Cyto', 0;
          'B'      , 'Cyto', 1;
          'pB'     , 'Cyto', 0;
          'Phos'   , 'Cyto', 1;
          'ActPhos', 'Cyto', 0;};
rxn(end+1).desc = 'A -> pA | I';
rxn(end).sub = 'A';
rxn(end).prod = 'pA';
rxn(end).enz = 'I';
rxn(end).k = 100;
rxn(end+1).desc = 'pA -> A';
rxn(end).sub = 'pA';
rxn(end).prod = 'A';
rxn(end).k = 100;
rxn(end+1).desc = 'B -> pB | A';
rxn(end).sub = 'B';
rxn(end).prod = 'pB';
rxn(end).enz = 'pA';
rxn(end).k = 0.1;
rxn(end+1).desc = 'pB -> B';
rxn(end).sub = 'pB';
```

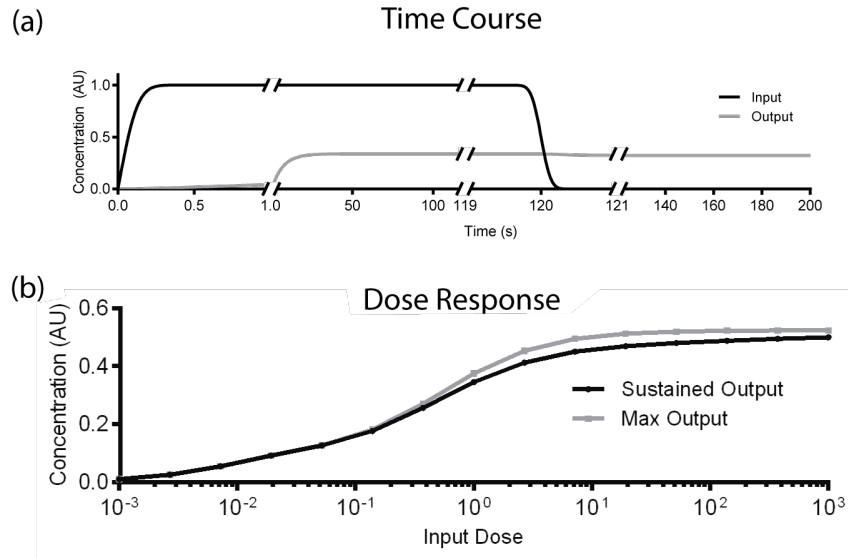


Figure 34: Time course and dose response of the transient stimulation responder network motif.

```

rxn(end).prod= 'B';
rxn(end).enz = 'ActPhos';
rxn(end).k    = 0.1;
rxn(end+1).desc = 'Phos -> ActPhos | pA';
rxn(end).sub = 'Phos';
rxn(end).prod= 'ActPhos';
rxn(end).enz = 'pA';
rxn(end).k    = 1000;
rxn(end+1).desc = 'ActPhos -> Phos';
rxn(end).sub = 'ActPhos';
rxn(end).prod= 'Phos';
rxn(end).k    = 100;

```

In this model (shown in Fig. 34) with input as Akt and output as the active biotransducer, large  $K^m$ 's were assumed to be much larger than all enzyme concentrations in order to reduce the parameter range of the model. This generally assumes a low concentration of enzyme-substrate complexes. This enables the enzymatic reactions to be second order mass action reactions. Thus, all concentrations are also set to one because the rate velocity in a second order rate law is proportional to both the catalytic rate and enzyme concentration, so changing the former is equivalent to changing the latter. As a test, the  $ED_{50}$  of the phosphatase activation was set at 0.1 by varying the reaction rate of the forward and reverse reactions. The  $ED_{50}$  of the other reactions were set at 1 using a similar method. The ability for the motif to sustain a



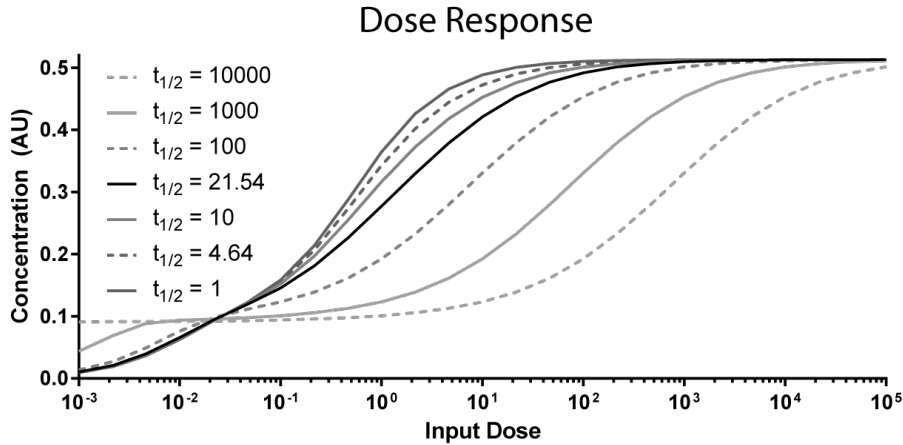


Figure 35: Dose response of the transient stimulation responder network motif with different rate at which the stimulus is removed (shown in this figure as  $t_{1/2}$ ). At small stimulus removal times ( $t_{1/2} \approx 1$  s), the dose response is a typical sigmoidal curve. At an intermediate stimulus removal half time ( $t_{1/2} \approx 20$  s), the output behaves like a log-linear converted (shown in the black line) with a broadened dose response. At large stimulus removal times ( $t_{1/2} \approx 10,000$  s), the dose response becomes binary at low dose.

transient signal was tested by making a step like stimulation for 120 s with an initial transient phase of 0.1 s. After that the stimulation was quickly removed at a constant rate with a half time of 0.1 s. The system is then allowed to equilibrium for another 80 s when the steady state measurement was obtained. Initially all species have a concentration of 1 in their inactive forms. The resulting time course, when applied with a dose of 1 A.U., is shown in Fig. 34a. This was simulated for a range of stimulation doses. The maximum output, which is the output at steady state during stimulation, and the sustained output, which is the output after removal of the stimulus, were recorded and compared to each other in Fig. 34b.

Although this network motif is able to convert a transient stimulation into a sustained output, it currently requires a step-like addition and removal conditions. The system is less able to sustain the signal if the stimulation is removed slowly. In Fig. 35 the stimulation is removed at a constant rate with a half time of between 1 s and 10,000 s (approximately 1 day). As removal half

life increases, two effects are observed. Firstly, at lower doses, the system spends longer without the phosphatase being fully activated. This has the effect of increasing the biotransducer activity at steady state. Secondly, the amount of time available for the phosphatase to attenuate the signal increases at large doses, and this leads to a decrease in the sustained output. Both of these combined means that at an intermediate half time (about 20 s), the steady state biotransducer activity is boosted at low doses but suppressed at high doses. This overall leads to a broader biotransducer active range with respect to the stimulus. As the half time increases beyond this, the dose response eventually flattens to become like a binary dose response curve. Thus, this network motif produces a number of emergent behaviour that can be summarised as:

- At low stimulus removal half times, the active biotransducer exhibits a standard sigmoidal dose response even when the signal is completely removed.
- At intermediate stimulus removal half times, the active range of the transducer to the stimulant is greatly broadened.
- At large stimulus removal half time, the active biotransducer begins to act as a binary switch.

## 5.5 DISCUSSION

Although a large number of biochemical modelling tools are available, none have the capacity for incorporating an additional divisive term to a system of rate equations that is necessary for the dQSSA model [183, 191]. As the size of studied biochemical networks increase, there is also an increased need to automate the generation of mathematical models of these networks. This necessitated the development of a new tool for modelling complex signalling pathways that can automatically generate models that use the dQSSA enzyme kinetic model. Additionally, common features among biological modelling tools, beyond modelling of kinetic pathways, is the modelling of multiple compartments and compatibility with the SBML format [183]. Given these design requirements, the dQSSA was further developed to include com-

partmentalisation. This theoretical framework was then used as the basis of the algorithm.

The result of the work in this chapter is the SigMat MATLAB toolbox. Using a set of test scenarios, we showed that the toolbox accurately reproduces the simulation results achieved by a pure mass action model. This shows that the implementation of the code not only faithfully and accurately compiles the rate equation of a model written using the SigMat format, but also correctly incorporates compartmentalisation, particularly in the dQSSA model. Furthermore, the ability for the SigMat algorithm to reproduce the complex network demonstrates the reliability and robustness of the algorithm.

The key motivation for the creation of the SBML format is to address inconsistency between mathematical models implemented by different researchers by offering a compact standardised format [179]. The SigMat model format can achieve the same goal as its models can be fully defined with only two modules of the program: the model file and the rate definitions. Beyond this, the SigMat toolbox also contains three additional advantages compared to manual generation and curation of models of biochemical networks.

Firstly, the goal is that the software describes the reaction network in question as a list of reaction equations. This enables the reaction structure of the network to be available at a glance, rather than as a series of differential equation. This is similar to the SBML format which only includes a list of system compartments, a list of system species, and a list of system reactions [179, 181]. This has advantages over algorithms such as Sim Biology, which requires the modellers to manually enter the differential equation for all system states.

Secondly, it compactly and unambiguously defines rate laws by inferring reaction types through an unambiguous series of criteria and then assigning each reaction type with a corresponding rate law that is automatically generated. This process is also designed to be flexible as modellers can define their own inference criteria and assign rate laws by writing a customised rate law definition script that can be selected as an option in the algorithm. This is designed to be simpler than the SBML implementations where the rate law needs to be included manually in each reaction as a MathML string [181]

Finally, SigMat aims to automatically distinguish between free parameters and fixed parameters, and their boundaries purely through a single list of parameter definitions. Additionally, free parameter descriptions are dynamically generated and labelled such that they can be easily identified in downstream processes such as parameter fitting. This aims to be more intuitive than the SBML implementation, where parameter related functionality is spread over three branches of the SBML structure: parameters, rules and constraints [179, 181].

As a demonstration of the SigMat algorithm, and exploration of the theoretical network motifs encountered in this thesis, two network motifs were analysed. The transient stimulation responder and the single and multiple receptor system are relevant to different parts of the signalling networks. The former is relevant for the effector components while the latter is related to the input part of signalling pathways. For the transient stimulation responder, our analysis demonstrated that it is possible to design a network motif that generates a sustained signal even from a short transient signal. This may explain how a proliferative response can be maintained from an initially transient signal [189, 190]. Additionally, the analysis showed that the motif has the behaviour of acting as a log-linear converter, though it is currently unclear what the biological application of this behaviour is.

In the case of the receptor systems, what our analysis demonstrated was a multiple receptor system essentially shields input arms from saturating effects of parallel arms. This prevents signals from becoming "ignored" by the system at their point of convergence. This is a powerful method of combining signals because the cell is able to prioritise inputs arms into a pathway by controlling the level at which it converges to the main branch. A likely reason for the need of such prioritisation is that signalling pathways are mapped such that inputs with more similar functions converge higher up the signalling pathway. An example of this is receptor tyrosine kinases generally contributes either to proliferation or growth like functions [185]. These typically merge at the PI3K node. However, there are two classes of PI3K isoforms which converge to target PIP2, with Class IA associated with RTKs and Class IB associated exclusively with G Protein Coupled Receptors (GPCRs) [31, 192]. There is some evidence for this where Akt activity in MEFs and NIH 3T3 cells were insensitive

to a class 1B PI3K inhibitor when stimulated by RTK ligands, but sensitive when stimulated by GPCR ligands [193]. Overall, what this analysis has shown is that small network motifs can generate unexpected network wide behaviour through their emergent properties. For example, the difference between single and multiple receptors can change the system from being competitive to cooperative. Also regulation of cyclic enzymatic reaction at different time scales can lead to sustained signalling outputs even when the stimulus is removed under one mode of operation, or expand the active range of the system in another mode of operation.

SigMat is thus a highly valuable tool as it enables rapid prototyping, *in silico* testing, and validation of hypothetical biochemical motifs and networks that may explain unusual behaviour in signalling pathways. Notwithstanding these advantages, there are a number of shortcomings with the SigMat algorithm. Firstly, the algorithm is not strictly SBML compatible yet as it can only read SigMat designed models written as a MATLAB .m file. However, since SigMat was inspired by the SBML format, it should be relatively simple to create a helper function that can convert an SBML model into a post-processed SigMat model given their relatively similar structures. Secondly, there is an emerging interest in the effects of signalling dependent changes in compartment volumes/surface areas. This requires volume parameters to not be modelled as a constant but rather a state variable with an associated differential equation [180, 194]. This is currently not available, but should in principle be possible with minor modifications of the SigMat algorithm.

In the next chapter, we will apply the framework developed in this chapter to the insulin signalling pathway. There we will investigate the ability for the network map that is currently known to predict experimental data.



## INFERRING MECHANISM OF AKT ACTIVATION WITH MODELLING

---

**Abstract** Although there are many models of insulin signalling in the literature, they are constructed using data from different cell models. This may be problematic as it is not yet clear if insulin signalling is conserved between different cell types. Therefore it is important to reconstruct pathway models for each individual cell type investigated. We have constructed the core insulin signalling pathway of the 3T3-L1 adipocyte using temporal data generated from the cell model. One key observed feature is an overshoot in the phosphorylation kinetics of Akt but a monotonic increase in the phosphorylation of Akt substrates. *In silico* simulation revealed that the pathway model cannot reproduce both the phosphorylation kinetics of Akt and its substrates simultaneously. This suggested that model was missing crucial mechanism, hypothesised to be located in the process of Akt activation. To investigate this, a an Akt focused model was developed which incorporated translocation of Akt to the plasma membrane, and the subsequent sequential phosphorylation of Akt at the 308 and 473 sites. This model revealed that only a small fraction of the total Akt pool (at most 20%) is phosphorylated. Under some circumstances the overshoot behaviour in Akt is not observed in the cytosolic fraction in the model. Thus it is possible for Akt substrates to be phosphorylate monotonically if it is phosphorylated only by the cytosolic fraction of Akt. An attempt was made to validate these predictions by immunoprecipitation and measurement of Akt activity. Immunoprecipitation of phosphorylated Akt supports the model prediction that only a small fraction of Akt is phosphorylated under insulin stimulation. The Akt activity assay was found to be unable to quantify the activity with sufficient resolution and specificity to verify the hypothesis.

The following pieces of data referred to in this chapter were generated by Dougall Norris.

- PIP3 time profile used as the input of the Akt model that was measured using TIRF microscopy.
- Immunoprecipitation of Akt under insulin stimulation.



## 6.1 INTRODUCTION

Following on from the work of the previous chapters we have arrived at a point where we can accurately model enzyme kinetic action and automatically apply models to networks of any size. Thus, we are in a position to begin building a model of the insulin signalling pathway. There is currently a large body of work describing the proteins and mechanisms involved in this signalling pathway [70, 195, 196]. However, many specifics about the mechanisms involved are missing. Thus, there is still some distance between the known mechanisms and what is required to build a full mechanistic model of the system.

There is also a large body of work describing different mathematical models of the insulin signalling pathway, as we have described previously in section 2.4 [74, 86, 124, 127, 150, 153, 155]. However, these models are based on different cell types. As we have explained previously, the response of the insulin pathway varies greatly between cell models. For example, the insulin dependent increase in the phosphorylation response of Akt at the Ser473 site under insulin stimulation, features an overshoot which peaks at 20 s in Fao cells, but 5 mins in human primary adipocytes [74, 124]. However, HeLa cells do not display the overshoot behaviour [127]. There are number of reasons leading to these discrepant results. For example, the expression profile of signalling molecules may differ between the cell models. Alternatively, the insulin signalling pathway itself may not be completely conserved.

Experimental data obtained using the 3T3-L1 adipocyte cell model displays an overshoot in Akt at the Ser473 site that peaks at 1.5 mins [169]. Furthermore, the Akt substrate temporal profile does not have the overshoot feature which its kinase possesses, unlike the Fao, HeLa or human adipocyte models where the kinase and substrate models are similar [74, 124, 127, 169]. These two unique temporal features in the 3T3-L1 adipocyte enables these more subtle regulatory features in the insulin signalling pathway to be investigated in this cell type.

In order to investigate this temporal feature and infer the regulatory mechanisms that it implies, a new model of the insulin signalling pathway, along with new parameter sets, was developed to study this cell type which has not been previously modelled

*in silico*. In this chapter, we firstly built the model from the well characterised core components of the insulin signalling pathway. The core components will be used to identify nodes which do not agree with experimental data. It is then hypothesised that these nodes contain missing mechanisms that are important to the regulation of the system. From there, new mechanisms will be added and tested to evaluate their ability to resolve the conflicting *in silico* results. The updated model can be verified and validated with further experimentation and finally integrated into the core pathway model. These models will be validated using the parameter free core prediction model proposed by Brannmark [149]. This involves deep exploration of the parameter spaces of unidentifiable models (models with large numbers of parameters but few data), and identifying features that are predicted by all parameter sets as the core predictions of the model. These predictions will then be tested and can be used to validate or reject the entire model.

## 6.2 METHODS

### 6.2.1 Mathematical Modelling

Generation of the mathematical model and simulation of the model was performed using the modelling software described in chapter 5. All association/dissociation reactions are modelled with mass action kinetics while enzyme kinetic reactions are modelled with dQSSA kinetics except where specified. Parameters were then varied using a customised Markov Chain Monte Carlo (MCMC) which finds all parameter sets with likelihood above a predefined acceptance probability threshold of  $L \geq 0.01$  given a likelihood function and an initial test parameter set. This optimisation algorithm was chosen because it produces multiple solutions through the generated Markov Chain, and is stochastic which allows the algorithm to evade local minima in the parameter space. These together allow the algorithm to obtain a more complete picture of the model's parameter space and obtain the multiple possible solutions.

The likelihood function is defined as the likelihood of the simulated data being observed. The likelihood function used in this

thesis assumes that the experimental data follows a normal distribution is expressed by [197]:

$$\epsilon(\theta) = \sum_d \sum_i \sum_t a_{it} \frac{b_i y_{idt} - \mu_{idt}}{\sigma_{idt}} \quad (144)$$

$$L(\theta) = \exp \left[ -\epsilon^2 / (2T) \right] \quad (145)$$

Subscripts  $d$ ,  $i$  and  $t$  indicates the stimulation dose, system species and time point respectively.  $\mu_{it}$  and  $\sigma_{it}$  are the estimated mean and standard deviation of the experimental data respectively.  $y$  is the simulation result and  $a_{it}$  is the weight of the data point. In general the weighting is set to 1 except for when specified.  $T$  is the annealing temperature.  $b_{it}$  is the scaling factor between the simulation and the experimental data and is calculated by:

$$b_i = \mu_i(t_{\text{end}}) / y_i(t_{\text{end}}) \quad (146)$$

The scaling factor is determined using the simulation and experimental endpoint because the study is concerned with fitting both the shape and the magnitude of the simulation to the experiment, it is necessary to prevent the model from compensating for an inability to fit the shape by altering its magnitude (which can occur if the simulation and experimental data is normalised to the mean or median of the time profile). Thus anchoring the simulation data to the end point of the experimental data forces the fitting algorithm to achieve a good fit of the simulation time profile.

The customised MCMC algorithm used essentially follows the algorithm outlined in the excellent textbooks by Gregory and Brooks [197, 198]. The algorithm is varied in that a boundary is added to the algorithm in order to limit the search space and prevent divergence of the search algorithm. This is important because the goal of using MCMC is to perform a detailed search within some parameter space. The efficiency of this will naturally reduce as the parameter space increases. Parameter bounds were chosen by first randomly selecting parameters until all species in the model reproduces the time scale of the experimental data (without regard for the actual shape of the temporal profile). The upper and lower bounds of all parameters are then set two orders of magnitudes above and below the parameter set identified as described in the previous sentence respectively.

Because proposal distributions become non-symmetric in the boundary, the acceptance probability is also modified to account for the asymmetric. Additionally, an annealing procedure is introduced in order to prevent the algorithm from becoming trapped in local minimas. The algorithm was first run at an annealing temperature of  $T = 31$ . The algorithm was set to store 20,000 unique parameter sets before termination. The annealing temperature was then dropped to the next temperature in the annealing schedule. The algorithm is then re-run but using the parameter sets stored from the previous temperature as seeds point for the new simulation. This process was repeated iteratively until the annealing temperature was dropped to  $T = 1$ . The annealing schedule used in this thesis was  $T = [31, 10, 3, 1]$ .

### 6.2.2 Immunoprecipitation

15  $\mu\text{L}$  of Protein G Sepharose Beads (P3296 Sigma Aldrich) were washed by adding 500  $\mu\text{L}$  of ice cold lysis buffer and then centrifuged at  $2000\times g$  for two minutes to pellet the beads. This was performed three times. The supernatant is then discarded. 500  $\mu\text{L}$  of lysis buffer was then added to the beads with 7.5  $\mu\text{L}$  of the specified antibody (see Appendix A.3 for antibodies product codes) was added and incubated on a rotator at  $4^\circ\text{C}$  for 2 hours. The beads were again pelleted by centrifugation and washed three more times. Cells were treated, harvested and quantified as described in section A.2. 277  $\mu\text{g}$  of cell lysate, diluted up to 500  $\mu\text{L}$  with lysis buffer supplemented with protease inhibitor and phosphatase inhibited, was then loaded onto the beads and incubated on a rotator at  $4^\circ\text{C}$  for 2 hours. The supernatant was captured as the flow through, and the beads were washed three times. The beads were then dried by aspiration and resuspended in 50  $\mu\text{L}$  of  $2\times$  sample buffer (100 mM Tris [pH 6.8], 4% SDS, 20% glycerol, 0.4mg/mL bromophenol blue). The sample was then boiled at  $95^\circ\text{C}$  for 10 mins to elute the proteins. The samples were then reduced by adding  $10\times$  TCEP (ThermoFisher Scientific 77720). 25  $\mu\text{L}$  was then separated by SDS-PAGE and analysed with Western Blot.

### 6.2.3 In-vitro Kinase Assay of Akt activity

Akt *in vitro* kinase assay was performed using the Omnia ST1 Kinase Assay (ThermoFisher Scientific KNZ1011, discontinued). Cells were harvested in 1% Triton X-100, 1% Octyl- $\beta$ -glucoside (BOG) or RIPA buffer (see appendix for recipe). For controls 0.4  $\mu$ g of recombinant Akt2 phosphorylated at the Thr-308 and Ser-473 sites were used (Active Motif, 31146). The reaction was performed in a reaction volume of 20  $\mu$ L in 384 well plates (Falcon 3962). The reaction buffer was made of 20 mM HEPES-KOH (pH 7.5), 10 mM  $MgCl_2$  (Sigma Aldrich, M2670), 2 mM EGTA (Sigma Aldrich E3887), 1 mM DTT (with kit), 1 mM ATP (with kit) and 1 mM substrate peptide unless otherwise specified. Inhibitors used were 100 nM GDC0068 (Selleckchem, S2808, Akt inhibitor), 200 nM CRT0066101 (Abcam, ab144637, PKD inhibitor), 200 nM PKI 6-22 (Sigma Aldrich, P6062, PKA inhibitor) 10  $\mu$ M GSK650394 (Sigma Aldrich, SML0773, GSK inhibitor), 400 nM LY2584702 (Selleckchem, S7698, S6K inhibitor) and 300 nM Sotrastaurin (Selleckchem, S2791, PKC Inhibitor) where indicated. Reactions were first made up to 15  $\mu$ L on ice and then initiated with addition of cell lysates or purified Akt (as indicated) in a Tecan Infinite M1000 plate reader in fluorescence mode with excitation and emission wavelengths of 360nm and 485nm respectively. Fluorescence intensity was measured for 6 hours with at 30 second intervals. Recombinant Akt (Active Motif 31146) was used as specified in the experimental figures using the manufacturer defined units of activity.

## 6.3 MODELLING OF THE INSULIN SIGNALLING PATHWAY

### 6.3.1 Modelling the core pathway

To create a new model of the upstream insulin signalling pathway 3T3-L1 adipocyte, we constructed a map of the insulin signalling curated from the literature. This was then adapted into a kinetic model which is shown in Fig 36.

In this simplified model, the sequential activation of IR-IRS-PI3K-PIP3 has been condensed into a single activation step, and

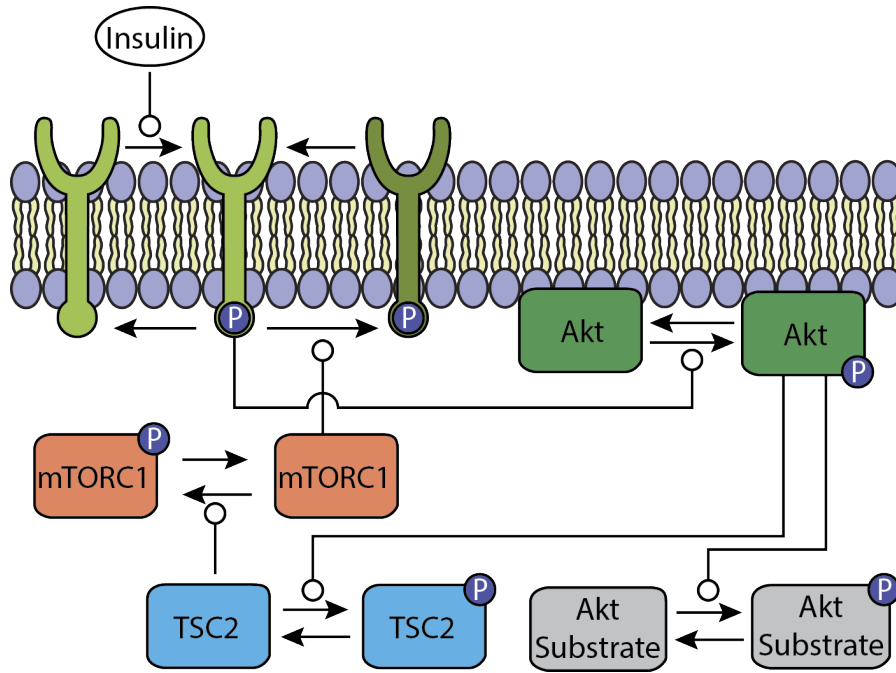


Figure 36: Setup of the model of the Insulin Signalling Pathway.

Akt activation has been simplified to be a one step process and directly activated by the IR. These simplifications are similar to those made by Kubota et al. and Dalle Pezze et al. [74, 127]. Whilst it is commonly asserted that an S6K mediated negative feedback onto IRS1 is a crucial regulatory mechanism of the pathway, it was not included in this model. This is because this mechanism is not expected to be important until around 15 minutes after Insulin stimulation, which is when S6K phosphorylation becomes appreciable at a 100 nM dose. Since the model is investigating the response of the system at about 1-5 minutes after stimulation, it is assumed that this negative feedback will not yet be active and can thus be safely omitted.

For this model, the data generated by Tan et al. was used extensively for the initial model calibration [169]. This data set contains temporal data at a sub-maximal (1 nM) and maximal (100 nM) dose of Insulin. This provides two dimensions of information (temporal and dose dependence) which is useful for both the extrapolation of general dose dependence of the system, and the exact temporal response at each dose. Understanding the dependence between both dimensions can present features that can

be used to identify novel regulatory mechanisms in the system. The key features investigated in this chapter is the overshoot feature, defined as an initial increase in activity to some maximal, and then a subsequent decrease into a lower steady state activity. The size of the overshoot is defined as the difference between the maximal and steady state activity. In this experimental data, the overshoot is observed in Akt phosphorylation under 1 nM of Insulin stimulation. However at the same time, there is a lack of this overshoot feature in Akt substrates under the same Insulin dose as identified by Tan et al. [169]. At 100 nM Insulin stimulation, the overshoot behaviour in Akt becomes lots and substrate phosphorylation reaches its maximum sooner.

To train the model, the phosphorylation kinetics of both the T308 and S473 residues of Akt and all phosphorylation sites of Akt substrates were used. Since the model as shown in Fig. 36 only contains a single Akt phosphorylation site (indicative of the "active" form of Akt), and a single Akt substrate, it is necessary to choose how the experimental data will be adapted to the model. To begin with, we assumed that both T308 and S473 phosphorylation are indicative of Akt activity. Thus, Akt activity is approximated as a mean aggregate of the phosphorylation of the T308 and S473 sites. In the case of the Akt substrates, we assume that the all Akt substrates respond identically, thus the Akt substrate in the model reflects the average response of all individual Akt substrates. Thus the phosphorylation time profiles for all Akt substrates were also mean aggregated. The temporal response of the phosphorylation level of each residue and the previously described aggregated time profiles are shown in Fig. 37.

In general, the response of the aggregated temporal profiles appear to be identical in both path and relative magnitude to the responses of the individual phosphorylation sites. The only exceptions to this are the two phosphorylation sites on Akt at 1 nM of Insulin. While both responses have the same general shape as the aggregated response, T308 phosphorylation appears to be far smaller than the aggregate. The reverse is true for S473 phosphorylation. Similarly GSK3 at 1 nM has a similar shape to the aggregate response, but is far smaller in magnitude. Since the model will be fitted to the aggregated time courses rather than the individual phosphorylation sites, the model will not be able



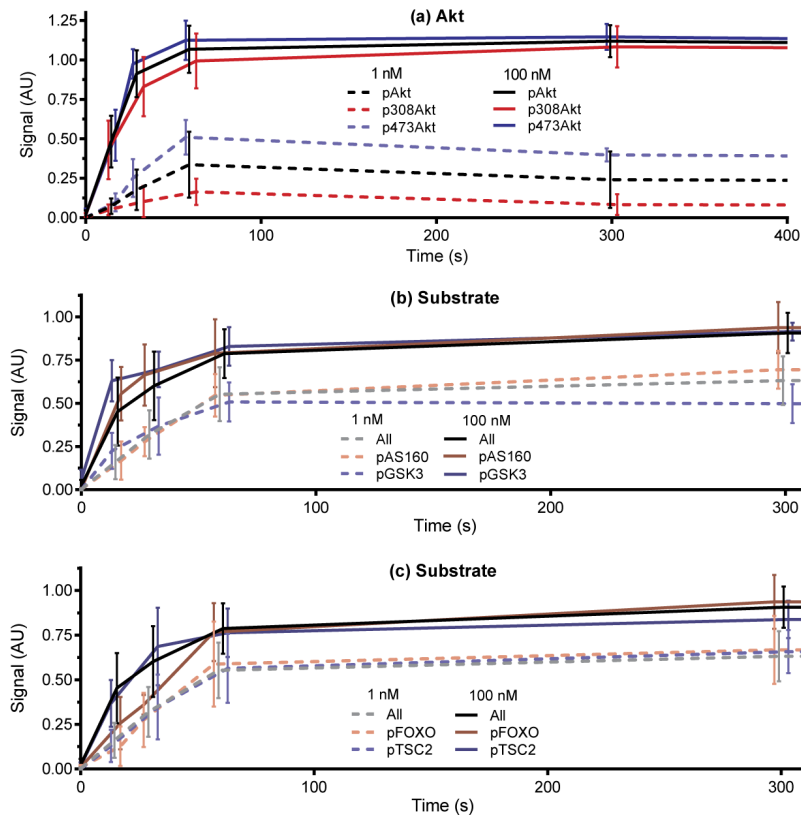


Figure 37: Time course data of (a) phosphorylated Akt under 1 nM insulin stimulation (dashed lines) and 100 nM insulin stimulation (solid lines), and (b) and (c) some Akt substrates under 1 nM insulin stimulation (dashed lines) and 100 nM insulin stimulation (solid lines). In both figures, coloured lines correspond to the phosphorylation of individual sites as indicated in the figure. Black lines are the mean aggregate of all phosphorylation site shown within the subpanels. Error bars are standard deviations. [169]

to capture the subtle differences between the responses of the individual phosphorylation sites.

From this starting point, the model was first fitted to both the average phosphorylation data (black lines) in both Akt and Akt substrates using Markov Chain Monte Carlo. This resulted in 20,000 parameter sets that fits with a residual less than 2. The resulting fits were divided into three clusters using k-means clustering of the temporal response of Akt and its substrates, nor-



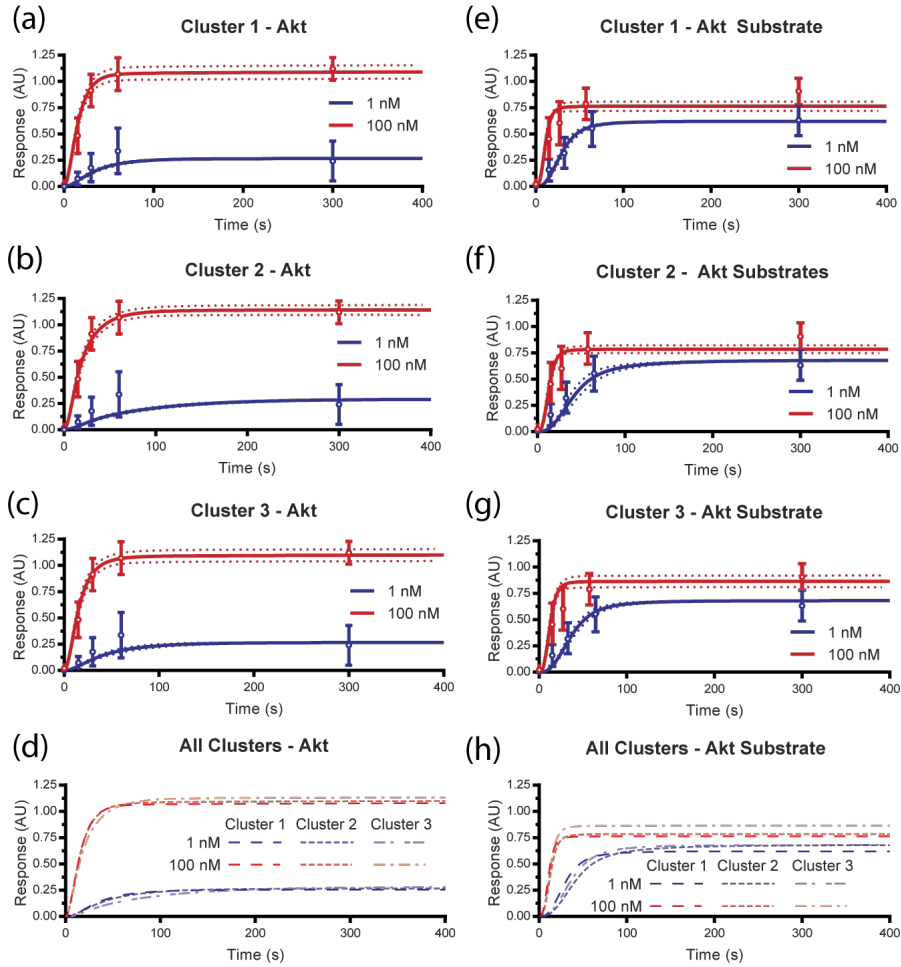


Figure 38: Result of calibrated model (solid line shows mean of best parameter sets, dashed lines shows the  $\pm$  one standard deviation spread from optimal parameters, defined by the parameter set with the largest likelihood) compared to the experimental data (open circle for mean, triangles for  $\pm$  on standard deviation from the mean). (a)-(c) Clusters 1-3 of Akt response to insulin. (d) shows the means of the Akt responses of the three clusters plotted together. (e)-(g) Clusters 1-3 of Akt substrate response to insulin. (h) shows the means of the Akt substrate responses of the three clusters plotted together.

malised to the median of the response of each node at 100 nM. The mean and standard deviation of the response of each cluster are shown in Fig. 38. While the overall shape of the response captured by the three clusters look remarkably similar, there are some subtle differences the three clusters. Firstly, the steady state

response (final time point) of Akt substrates at 100 nM insulin in cluster 3 ( $0.86 \pm 0.57$ ) is higher than clusters 1 and 2 ( $0.76 \pm 0.04$ ). Furthermore, cluster 2's Akt response at 1 nM (0.15 at 60 s) is slower than the other two clusters (0.2 at 60 s). Apart from these subtle differences, the three clusters are equally capable of replicating the experimental data. For example, the predicted 100 nM Akt substrate response is modelled to be far faster than the 1 nM response, which is in line with the experimental data.

On the other hand, the model did not replicate the overshoot behaviour in Akt with 1 nM insulin stimulation. This raised concerns as the overshoot feature is robustly observed between experimental replicates. We argued that the fitting procedure likely missed this feature due to the error bars being artificially increased by aggregating the time courses of the 308 and 473 sites on Akt. Recall in Fig. 37 that individually, the error bars for each site are much smaller and forced non-overshoot responses to be rejected.

Taking this into consideration, a second round of fitting was attempted where the model was fitted to the time courses of the 308 and 473 sites on Akt individually. Additionally, we hypothesised that the temporal feature of Akt and its substrates were mechanistically related. Because the temporal features of the experimentally observed time course of Akt and its substrates are different, and that Akt is upstream of its substrates, it is possible that the substrate time profile is affected by the upstream Akt time profile. Thus it does not make sense to fit them both with equal weighting. To test this hypothesis, the model was additionally fitted to the full Akt time profile but only the end point of the substrate time profile simultaneously. This would enable us to assess the range of responses the substrates can generate with a given Akt time profile while retaining the same dose response. Thus to be more specific, the second round of fitting differed from the first round by:

- Fitting the simulated data to the experimental data for only one of the two Akt phosphorylation sites.
- Fitting only the end point (1800 s) of the Akt substrates to ensure the substrate dose response is reproduced but its temporal profile is unconstrained.

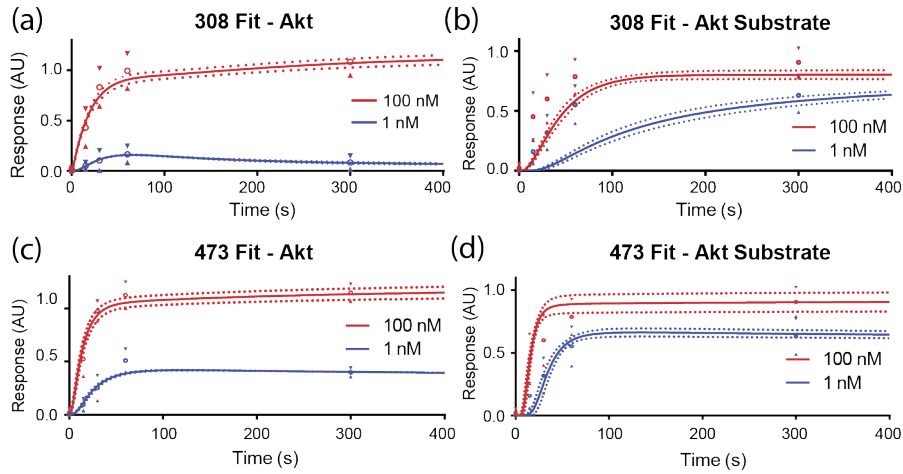


Figure 39: Plots of the (a) model fit to the pT308 Akt phosphorylation time profile and (b) resulting substrate phosphorylation fit. (c) model fit to the pS473 Akt phosphorylation time profile and (d) its resulting substrate phosphorylation fit. Solid line represents the mean of all accepted parameter sets while dotted lines show one standard deviation of the range of simulated kinetics across accepted parameter sets.

The result of this alternative fitting method is shown in Fig. 39. Due to the smaller error bars associated with each individual phosphorylation site, the overshoot feature of the time course became an important constraint in the fitting procedure with the new set of parameters now reproducing it. It is interesting to note that the 308 site is shown to have a larger overshoot at 1 nM of Insulin compared to the 473 site.

Another noteworthy result of the simulation is the Akt substrates responses. The simulations fitted the end points reasonable well (as required by the fitting criteria). It also produced no overshoot in the substrate time course. However, it showed a preference for parameters where the substrate response lags behind the Akt response. This lag was much slower than that observed experimentally. So this raises a number of questions. Why is the algorithm favouring slower responses for the Akt substrates, and what would happen if we forced the Akt substrates to be phosphorylated at a faster rate?

To investigate this, a third round of simulations was carried out. We took a parameter set fitted to the Akt 308 site, and fixed all parameters except for the phosphorylation catalytic rate and the

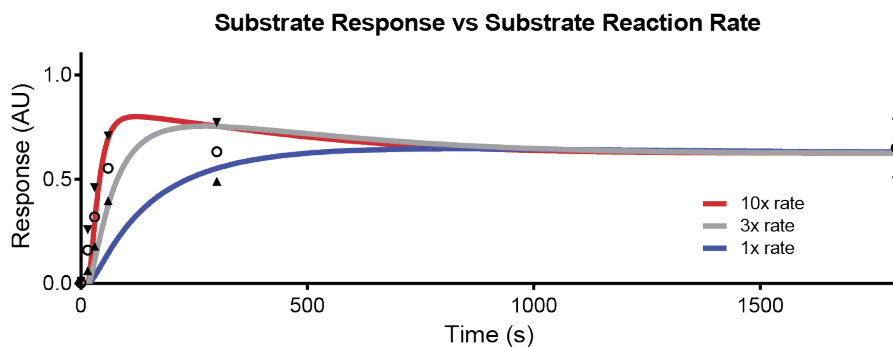


Figure 40: The resulting substrate temporal profiles that can be achieved by varying the reaction rate of the substrate reactions based on the model fitted to the pT308 phosphorylation time profile.

dephosphorylation rate of the substrates. We sought to test the effect the reaction rate has on the time course of the Akt substrates when the time profile of Akt featured an overshoot behaviour. In this case, the reaction rate can be interpreted as a number of physical processes. Firstly it can indicate the actual speed of the phosphorylation reaction and secondly it includes delays caused by the dissociation and translocation of Akt to its substrates before phosphorylation can occur.

The results for different simulations were shown in Fig. 40. The blue curve shows the reaction rate from the original parameter set (i.e. it is the same as Fig. 39b), which is slow (does not match the first three experimental time points) and it does not feature an overshoot. If we triple the base reaction rate (grey curve in Fig. 40), the substrate time course is now fast enough to be within one standard deviation of the mean of the third experimental time point ( $t = 180$  s). However, the temporal profile now begins to exhibit an overshoot behavior. This causes the simulated result to be just inside one standard deviation of the mean of the fourth experimental time point ( $t = 300$  s). If we triple the rate again (red curve in Fig. 40), we find the simulation fits the second ( $t = 30$  s) and final time point ( $t = 1800$  s) well, but the remaining time points are at a difference of at least one standard deviation between the experimental mean and simulation. What this demonstrates is for this particular parameter set, while the Akt time profile is fitted well, the Akt substrate time profile fits poorly. Thus if this parameter set is evolved to a more optimal fit (quantified by Eqn. 145),

it will likely compromise the goodness of the Akt time profile fit in order to improve the fit to the Akt substrate time profile. This phenomenon is a likely explanation of the difference between the quality of the free fits shown in Fig. 38 where the model was fitted to both the Akt and Akt substrate time profiles simultaneously, and the fits shown in Fig. 39 where the model was fitted only to the Akt time profile.

This shows that with this network architecture, the simulated Akt substrate temporal response at 1 nM insulin stimulation is unable to fit more than two of the experimental time point. In other words, the model can explain Akt's temporal profile but has a poor ability to explain both the Akt substrate time profile simultaneously. This strongly implies that the network topology in this model is incorrect and there are likely to be as yet unaccounted for mechanisms behind Akt's activation of its substrates beyond Akt translocation and kinase activation.

### 6.3.2 Modelling of the Akt Signalling Pathway

#### 6.3.2.1 *Calibration of the Model*

From the results of the previous simulations, it is clear that there are missing mechanisms between Akt becoming phosphorylated and Akt subsequently phosphorylating its substrates. This is noteworthy as it is there are two commonly held assumptions within the literature regarding the mechanisms of Akt signalling. Firstly, it is assumed that Akt phosphorylation is synonymous with Akt activity [199]. Secondly, it is assumed that Akt directly phosphorylates its substrates [200]. Our results imply that at least one of these assumptions is wrong. In this chapter, we will investigate the validity of the first assumption. If we assume that Akt phosphorylation does not imply activation, then this implies additional mechanisms are required before Akt becomes active. Here we test two possibilities.

The first possibility is that Akt must first be dually phosphorylated before it is fully active [58, 201]. The phosphorylation data used in the previous section were only measures of total phosphorylation at a single site, regardless of the phosphorylation states of the other sites [38]. Thus the resulting measure is a mixture of the singly phosphorylated form of Akt as well as the doubly

phosphorylated form. It is possible that only doubly phosphorylated Akt is active and the second phosphorylation is able to filter out the overshoot behaviour.

The second possibility is that localisation of Akt is important. It is known that Akt translocation to the plasma membrane (PM) is required and sufficient for its phosphorylation and activation [60]. However, it is unclear whether Akt phosphorylates its substrates at the PM or in the cytosol, although there is data in the literature to suggest that substrate phosphorylation occurs at the plasma membrane [169]. It is possible that individual fractions of phosphorylated Akt can occur without an overshoot and that those fractions are where Akt is able to phosphorylate its substrates.

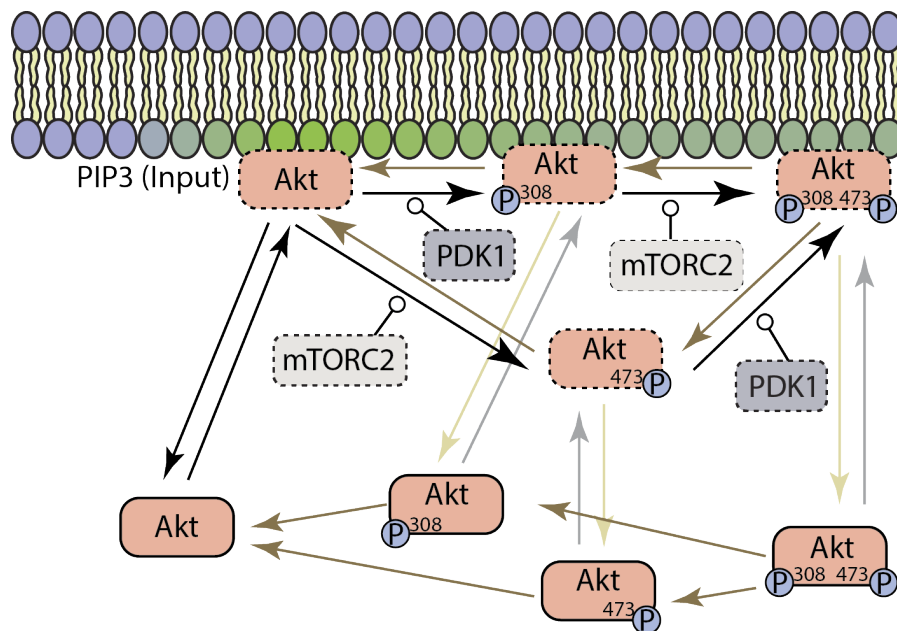


Figure 41: Setup of the model for Akt phosphorylation and activation. Nodes with dotted outlines lie within the plasma membrane and nodes with solid outlines lie in the cytosol. Reactions can only occur when reactants are in the same compartment (e.g. PDK1 can only phosphorylate Akt in the plasma membrane). PIP3 dimerizes with Akt when Akt is recruited to the plasma membrane. PIP3 is released back into the plasma membrane when Akt dissociates from the plasma membrane. The input for the model is PIP3, where its total concentration is controlled external to the model.

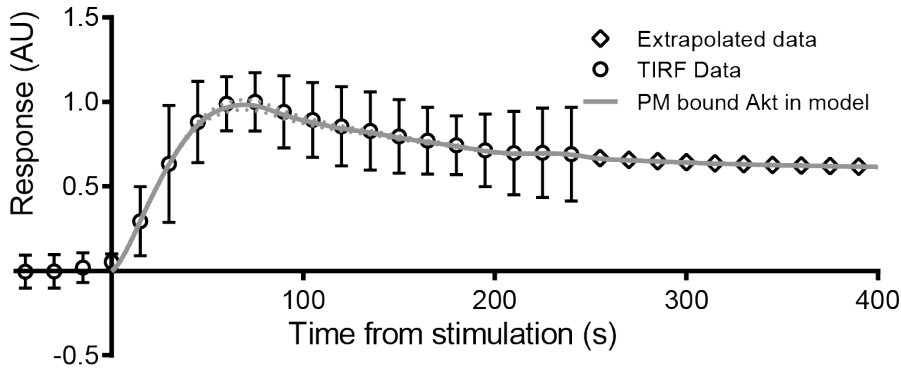


Figure 42: Time course of Akt translocation at 1 nM Insulin. Open circles and their associated error bars are TIRF microscopy results ( $n=5$ ). The open diamonds are extrapolated from the post-overshoot decay using an exponential decay function. The grey line shows the average PM bound Akt response predicted by the model across the accepted parameter sets. [Experiment performed by Dougall Norris]

In this section, we investigate both of these hypotheses by isolating the Akt signalling pathway and modelling the mechanisms of Akt phosphorylation in more detail. The model used is shown in Fig. 41.

This model increases the complexity of the Akt components of the insulin pathway in two ways. Firstly it incorporates compartmentalisation, which is a vital process in the activation of Akt. Secondly, the phosphorylation sites of Akt have been explicitly defined, and their respective kinases are also explicitly included. In order to simplify the model though, the upstream components of the model is truncated at the PIP3 generation stage. This removes other signalling components such as the mTORC1 or other negative feedback mechanisms that target nodes upstream of PIP3. Instead, the effects of these negative feedback pathways are accounted for by directly replicating the *in vitro* dynamics of PIP3 generation. By capturing the *in vitro* dynamics of PIP3 generation under a step increase in insulin, the measured kinetics of PIP3 generation will implicitly include any feedback mechanisms that act on it, or nodes upstream of it. This is because any feedback affects will necessarily alter the PIP3 generation dynamics, which is then transmitted downstream via Akt. Thus, replicating the PIP3 generation dynamics will account for any feedbacks that apply to



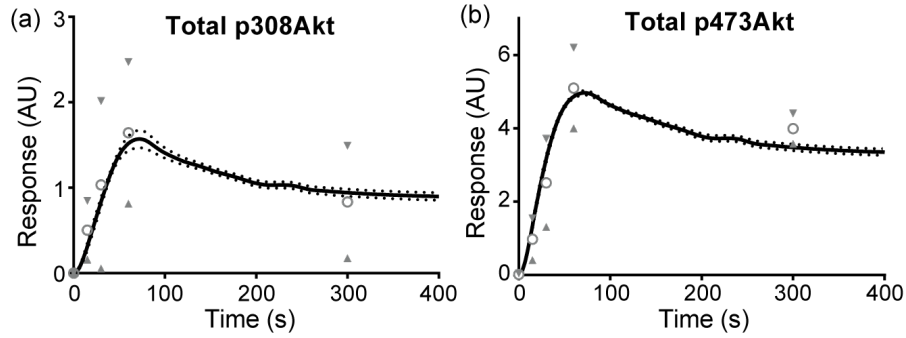


Figure 43: Time course of Akt phosphorylation at 1 nM Insulin. Open circles and the associated error bars are the experimental mean and standard deviation of  $n = 5$  [38]. The black line shows the average of the Akt phosphorylation on the (a) 308 site and (b) 473 site across the accepted parameter sets.

PIP3 or nodes upstream of it. Thus, the Akt focused model would then be able to ignore some feedback connections without jeopardising the accuracy of the model. However, it should be noted that this method only accounts for feedbacks that target nodes at or upstream of PIP3. Feedbacks that target components downstream of PIP3 are not implicitly accounted for and will require explicit modelling. This will be one of the subjects of investigation of this model.

To model the system in this way, the temporal dynamics of PIP3 generation must be acquired first. This is possible using TIRF microscopy, and using the translocation of various PH domain containing proteins as a measure of the PIP3 generation rate. The obtained temporal profile (data generated by Dougall Norris) is shown in Fig. 42. Since the TIRF microscopy time course was only measured up to 210 seconds, and the phosphorylation data exists up to 300 seconds, the TIRF microscopy data was extrapolated to the 300 second time point by fitting an exponential decay curve to the PIP3 time course between  $t = [75, 240]$ s using the Nelder-Mead method using the cost function shown in Eqn. 145.

The experimental and extrapolated data were then used as the inputs of the Akt model. This is done by finding the time derivative of the piecewise polynomial spline of PIP3 generation time profile shown in Fig. 42. First a piecewise cubic spline of the PIP3 generation time profile is constructed. Next the coefficients of the



spline is modified to represent differential of the spline using the operation:

$$C'_{n-1} = nC_n \quad (147)$$

Where  $C'$  is the coefficient of the derivative spline and  $C$  is the coefficient for the PIP3 generation time profile spline.  $C_n$  indicates the coefficient of the  $n$ th polynomial term and this convention applies for both  $C$  and  $C'$ . As the operation is a derivative,  $C'_3 = 0$ . The derivative piecewise polynomial spline is then added to the rate of change of PIP3 in the model.

The system is also normalised to the concentration of PIP3 as per the TIRF microscopy time course, then the Western Blot data was fitted to the total amount of Akt phosphorylated at the 308 and 473 sites (the same data as in the Insulin signalling model, i.e. Fig. 37). This is shown in Fig. 43. The simulated translocation time course is shown in Fig. 42 by the solid line. In both of these figures, it can be seen that the simulated data fits within the uncertainty in the data. Thus the model structure as it stands is able to explain the data.

#### 6.3.2.2 Partitioning of the forms of phosphorylated Akt

Using this model, we move towards some predictions. We first investigated the range of time courses of each species of Akt that are produced. This is shown as a distribution of the response across the accepted parameter sets in Fig. 44. Although This shows that in most cases, all species of Akt displays an overshoot regardless of its location in the cell.

Looking at the steady state, it was found that only a small percentage of Akt was phosphorylated with up to 25% being phosphorylated only at the 473 site, 10% phosphorylated at only the 308 site and only 5% phosphorylated at both the 308 and 473 sites Fig. 45. However, these numbers only represent the maximum fractions that can be generated by the model. It is more likely that a smaller amount of Akt is phosphorylated. Nonetheless, the upper limit of the phosphorylated percentage present a core prediction of this Akt model that can be tested experimentally.

The distribution of the fractions phosphorylated throughout the parameter sets generated is shown in Fig. 45. What this fig-

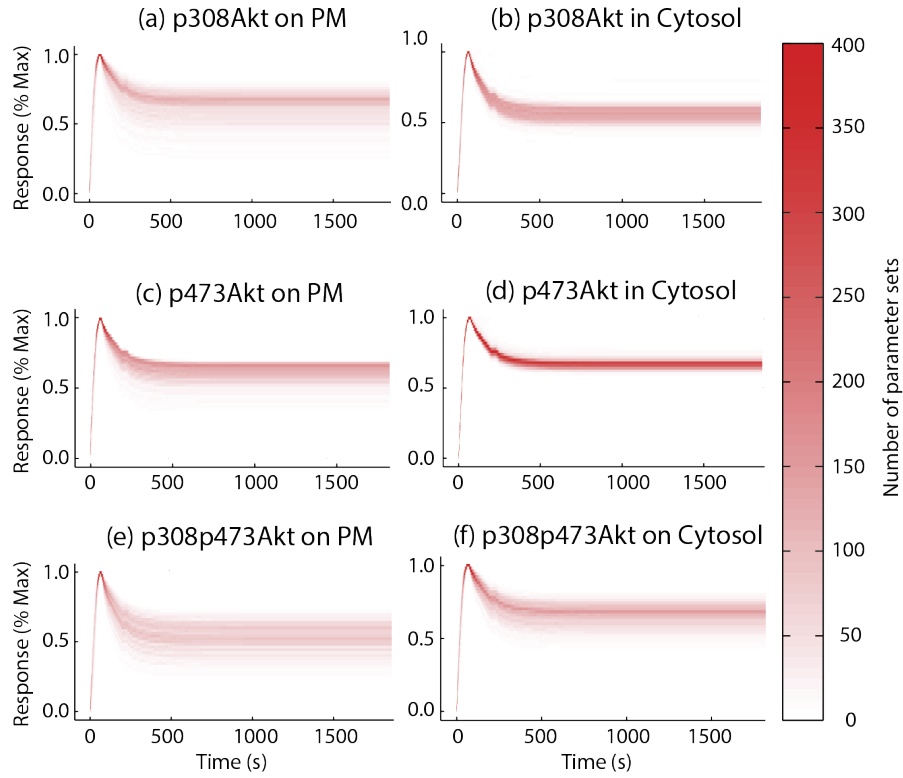


Figure 44: Time course of Akt phosphorylation at 1 nM Insulin at the (a) 308 site only at the plasma membrane, (b) 308 site only at the cytosol, (c) 473 site only at the plasma membrane, (d) 473 site only at the cytosol, (e) 308 and 473 site only at the plasma membrane, (f) 308 and 473 site only at the cytosol. The Deepness of the red at point shows the number of parameter sets that shows that level of response at that particular time point. This shows aggregated results from 1000 parameter sets. Distribution of the parameter sets in parameter space is shown in Fig. ?? in Appendix D.3

ure illustrates is the model suffers from strong issues with non-identifiability. With only two experimental time courses to fit to, the model have sufficient degrees of freedom to exhibit a large number of system behaviours whilst still replicating the experimental data. The large variation in the percentage of Akt in various physical location with different levels of phosphorylation is an indicator of this. Thus what this model attempts to achieve is not to provide specific predictions. Instead, this model aims to understand the broader, more general behaviour of the model as implied by the experimental data. An example of this, again using Fig. 45 as an example, is the clear trends in the relative abundance of the different phosphorylated forms of Akt. Singly phosphorylated Akt at the 473 site is most abundant, followed by phosphorylation only at the 308 site. Dually phosphorylated Akt is least abundant.

A second general trend that was observed across all parameters sets, was that the system has a strong preference for achieving double phosphorylation of Akt when Thr308 is phosphorylated first. This was verified *in silico* by deleting the Thr308 phosphorylation step of Akt already that has already been phosphorylated at Ser473 (i.e. the catalytic rate of this step was set to zero while the reaction rates of the Thr308 phosphorylation from unphosphorylated Akt is left unchanged). It was found that the amount of doubly phosphorylated Akt remained unchanged under this perturbation. However, there is still a preference for the system to achieve singly phosphorylated Akt on the 473 site. This appears to suggest that the initial phosphorylation of Akt is not ordered. However, to achieve double phosphorylation, Akt must first be phosphorylated on the 308 site.

Beyond the distribution of the various phosphorylated forms, it is also of interest to look at the distribution of Akt between the cytosol and the plasma membrane. Currently, the model assumes that the affinity of Akt for PIP3 does not vary with its phosphorylation state (this is reasonable to assume because the phosphorylation sites in question are not in the PH domain). Given this assumption, it was found that most of the accepted parameter sets predicts a small amount of Akt in the plasma membrane.

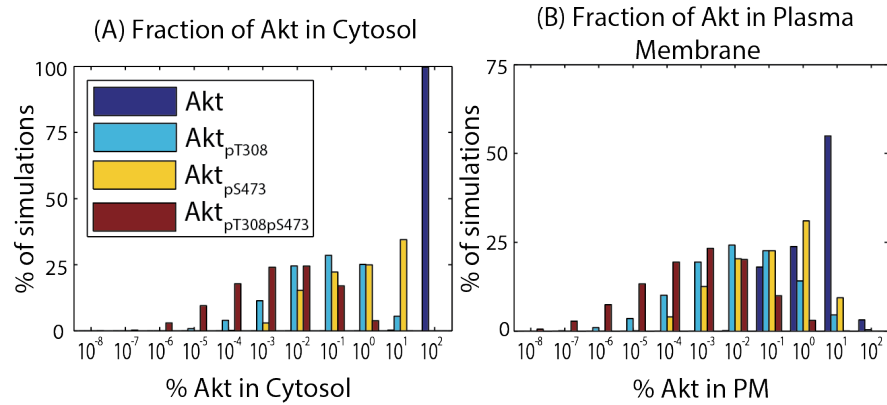


Figure 45: Distribution of the predicted partition of phosphorylated Akt between its different phosphorylated forms as well as their locations at steady state. (a) Indicates the fraction in the cytosol (b) indicates the fractions in the plasma membrane.

### 6.3.2.3 Transduction of Akt Signal

Having explored some of the more interesting aspects of this model, we now focus our investigation of the Akt model to our original hypothesis regarding the mechanism of Akt phosphorylation of its substrates (in the same manner as the investigation illustrated in Fig. 40). In spite of the result that most parameters sets will show an overshoot behaviour in all states (shown in Fig. 44), there are some parameter sets that show an exception to this rule. One of these is the parameter set shown in Table. 11. With this parameter set, the total 308 and 473 phosphorylation time course fits within the clusters that are constrained by the

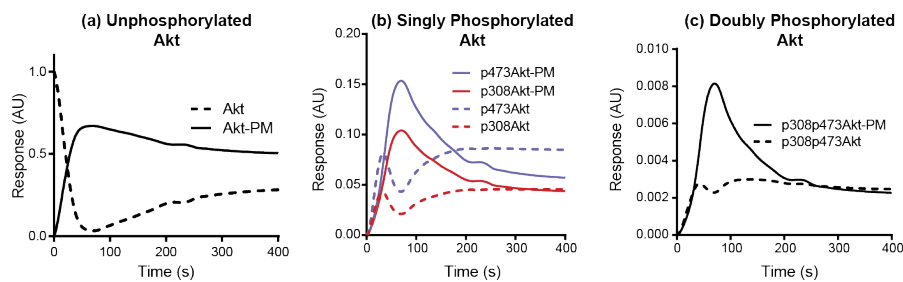


Figure 46: Temporal profiles of the different forms of Akt produced by the non-overshooting parameters set listed in Table. 11. Note there is no distribution shown because this is the result of a single parameter set. States without any suffixes are cytosolic.

Table 11: List of the parameters used to achieve the model with a cytosolic phosphorylated Akt temporal profile that has a suppressed overshoot. Non-conventional units are used for the concentration and volume. The model is non-dimensionalised with respect to the PIP3 concentration and volume of the cytoplasm, thus N indicates concentration as a fraction of PIP3 concentration, and V indicates volume as fraction of cytosol volume.

$[\text{PIP3}]_{\text{max}}$ (N)	1.0	$k_{\text{cat}} \text{ (s}^{-1}\text{)}$ $\text{Akt} \rightarrow \text{Akt}_{\text{p308}}$	0.14	$K_{\text{m}} \text{ (N)}$ $\text{Akt} \rightarrow \text{Akt}_{\text{p308}}$	0.15
$[\text{Akt}]_0$ (N)	1.0	$k_{\text{cat}} \text{ (s}^{-1}\text{)}$ $\text{Akt} \rightarrow \text{Akt}_{\text{p473}}$	0.43	$K_{\text{m}} \text{ (N)}$ $\text{Akt} \rightarrow \text{Akt}_{\text{p473}}$	0.29
$[\text{PDK1}]_0$ (N)	1.0	$k_{\text{cat}} \text{ (s}^{-1}\text{)}$ $\text{Akt}_{\text{p473}} \rightarrow \text{Akt}_{\text{p308p473}}$	0.31	$K_{\text{m}} \text{ (N)}$ $\text{Akt}_{\text{p473}} \rightarrow \text{Akt}_{\text{p308p473}}$	800
$[\text{mTORC2}]_0$ (N)	1.0	$k_{\text{cat}} \text{ (s}^{-1}\text{)}$ $\text{Akt}_{\text{p308}} \rightarrow \text{Akt}_{\text{p308p473}}$	0.34	$K_{\text{m}} \text{ (N)}$ $\text{Akt}_{\text{p308}} \rightarrow \text{Akt}_{\text{p308p473}}$	1.86
$[\text{Substrate}]_0$ (N)	1.0	$k_{\text{cat}} \text{ (s}^{-1}\text{)}$ $\text{Substrate} \rightarrow \text{pSubstrate}$	160	$K_{\text{m}} \text{ (N)}$ $\text{Substrate} \rightarrow \text{pSubstrate}$	10
		$k \text{ (s}^{-1}\text{)}$ $\text{pSubstrate} \rightarrow \text{Substrate}$	0.02		
PM Size (V)	0.96	Dephosphorylation <sub>Akt</sub> (s <sup>-1</sup> )	0.49		
Association Rate (s <sup>-1</sup> )	72.53	Dissociations rate (s <sup>-1</sup> )	1.58		

experimental data (i.e. it fits within the range shown in Fig. 43). However, when separating out the individual species, there is a stark difference between the time courses of the plasma membrane bound fractions and the cytosol localised fractions. The time course of each of these fractions are shown in Fig. 46. It shows that the plasma membrane fractions follow the time course of the PIP3 generation. The cytosolic fraction, however, actually undergoes a transient suppression in phosphorylation which recovers some time later.

This particular parameter set is interesting because it presents a fraction of Akt that does not closely follow the total 308 and 473 phosphorylation temporal dynamics that was shown in Fig. 43. On the whole, the cytosolic fractions do not exhibit an overshoot behaviour. Thus, this can lead to substrate phosphorylation that

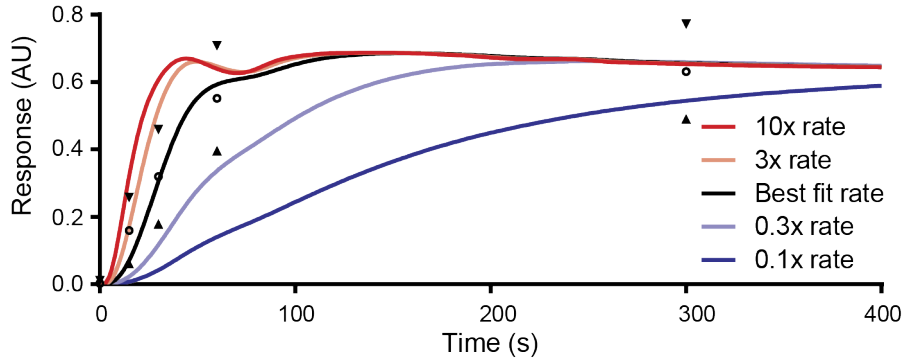


Figure 47: Substrate temporal profiles that can be achieved by varying  $k_{\text{cat}}$  of the  $\text{Substrate} \rightarrow \text{pSubstrate}$  and  $k$  of the  $\text{pSubstrate} \rightarrow \text{Substrate}$  reactions based on the model using parameters of the model that produces a non-overshooting cytosolic phosphorylated Akt (listed in Table. 11). This is compared to experimental data of Akt substrates aggregated in the same way as in the investigation of the insulin signalling model in section 6.3.

does not exhibit an overshoot. Additionally, the added transient repression may actually be ignored by substrates which are phosphorylated at a slow rate because this transient drop in cytosolic phosphorylated Akt occurs on a much faster time scale than substrate phosphorylation. This presents a possible model that can simultaneously fit the overshoot feature of the Akt phosphorylation time profile as well as the flat phosphorylation time profile of Akt's substrates.

In order to test this hypothesis, the model was modified such that only the cytosolic portion was allowed to phosphorylate a generic substrate that represents the aggregate of the experimentally observed Akt substrates (in the same manner as performed in the investigation of the insulin signalling model in section 6.3). The speed of this reaction was then modified (by varying both the catalytic and dephosphorylation rates by the same multiplicative factor) to investigate the range of temporal profiles that the substrate can produce under this system. This is shown in Fig. 47. It can be seen that the substrate temporal profile fits the experimental data extremely well in this case.

It should be mentioned that in this particular case, the percentage of Akt at the plasma membrane is about 50%. This is different to the general trend that is found across the range of accepted parameter sets as shown in Fig. 44. It is likely that the different behaviour of this model is attributed to the different mechanism by which Akt achieve its dose response. In the case of the low

percentage plasma membrane localised models, the limiting factor in Akt phosphorylation is the PIP3 generation. Here, the rate limiting step is the the total Akt available. As Akt is recruited to the PM by PIP3, it becomes depleted from the cytosol. Eventually, this process leads to PIP3 locking Akt to the membrane when PIP3 concentration is sufficiently high. Thus, the overshoot in the cytosolic fraction is suppressed because Akt becomes significantly depleted from the cytosol. The sudden burst of PIP3 (from the overshoot) is in fact a negative feedback on the Akt available in the cytosol because the PIP3 acts as a sponge that sequesters excess Akt from the cytosol. This presents one explanation for the absence of the overshoot behaviour in the phosphorylation of Akt substrates. Whilst this parameter set presents an interesting hypothesis, it is unclear whether this behaviour is physiologically relevant. In particular the saturation of Akt at the height of the PIP3 overshoot is a bold prediction of this model. Thus this provides a experimentally tractable validation of this model that can be interrogated using further TIRF microscopy experiments at different doses of Insulin. However, given the rare occurrence of this behaviour amongst the identified parameters sets, it is possible that the behaviour produced by this parameter set is merely an exception that is not physiological.

#### 6.4 BIOLOGICAL VALIDATION

Two key predictions were determined from the mathematical modelling. First is that only a small fraction of Akt is phosphorylated, and an even smaller fraction is doubly phosphorylated. Second is that only the cytosolic fraction of Akt is active and its time course follows that of the substrates (overshoot is suppressed by the excess PIP3 generated).

In this section, this will be tested in one of two ways. The first prediction was tested by immunoprecipitation of the various phosphorylated forms of Akt. The second prediction was testing using an *in vitro* kinase assays to determine the variation in Akt activity with time profile upon Insulin stimulation.

#### 6.4.1 Verification of separate pools of phosphorylated Akt

Immunoprecipitation (IP) was used to determine the population of the various phosphorylated forms of Akt. This experiment (performed by Dougall Norris), involved performing a four point time course under insulin stimulation, and then pulling down the Akt with antibody targeting one phosphorylation site, and then performing western blot analysis with an antibody targeting the opposing phosphorylation site. Cross reactivity of the Akt and the IP antibody was mitigated by eluting the IP in non-reducing sample buffer (i.e. sample buffer without TCEP). This prevents the IP antibody from dissociating and causing the heavy chain band from obscuring the Akt band, which will separate at a similar rate on an SDS-PAGE gel based on their mass.

Preliminary results suggest that only a small fraction of the total pT308 Akt is pulled down by the pS473 antibody, that is the dually phosphorylated fraction (Fig. 48). To quantify the fraction of dually phosphorylated Akt, we first calculated the enrichment of each individual pull down of Akt. Fig. 48 shows that there is significant enrichment in pS473 Akt by the IP. Using the fractions defined in Fig. 49, this is the ratio:

$$\% \text{ T308 dually phosphorylated} = \frac{ST_{\text{wcl}}}{T_{\text{wcl}} + ST_{\text{wcl}}} \quad (148)$$

The S473 blotted lanes can be used to relate the concentration of some of the S473 fractions. Firstly, we need to determine the an-

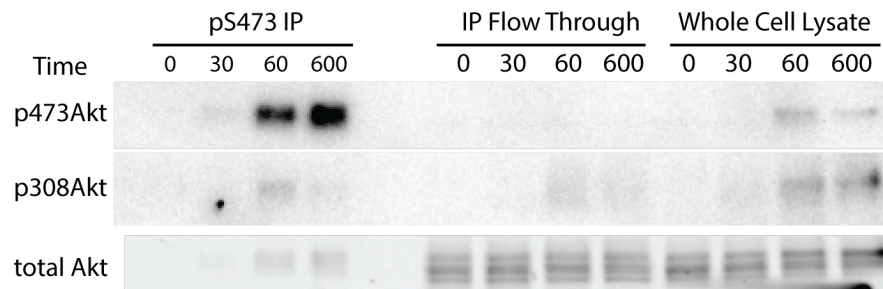


Figure 48: Immunoprecipitation result of Akt after stimulation with 1 nM of Insulin. The analysis of the two phosphorylation sites were performed on the same membrane.



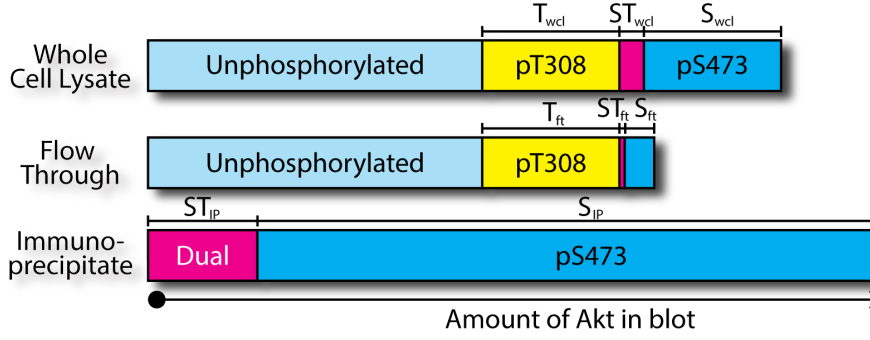


Figure 49: Illustration of the various fractions of phosphorylated Akt that are in the various Western blotting samples shown in Fig. 48. The whole cell lysate and flow through samples are loaded to the same total protein amount, which means unaltered protein contents stay the same between those two samples. The IP sample only contains samples extracted by the IP process, but they become highly enriched. However, the concentration ratios between extracted species are the same as their ratio in their original whole cell lysate samples assuming the antibody binding efficiency is constant between species.

tibody binding efficiency of the IP process. The binding efficiency can be expressed as:

$$A = 1 - \frac{S_{FT} + ST_{FT}}{S_{wcl} + ST_{wcl}} = 1 - \frac{S_{FT}}{S_{wcl}} = 1 - \frac{ST_{FT}}{ST_{wcl}} \quad (149)$$

Which is one minus the fraction of S473 phosphorylated Akt left behind after the IP process. Now the ratio of the band intensity between the flow through to whole cell lysate is equivalent to:

$$\frac{\text{FT Band}}{\text{WCL Band}} = \gamma_{\text{Ser}} \frac{S_{FT} + ST_{FT}}{S_{wcl} + ST_{wcl}} \quad (150)$$

Where  $\gamma_{\text{Ser}}$  is an additional factor to account for loss of signal in the IP and flow through lanes due to longer processing times. The band intensity ratio is measured to be less than 1% (below the sensitivity of the experimental process). Thus  $A \approx 1$  and the  $\gamma_{\text{Ser}}$  indeterminate. Next, we can determine the fold enrichment

of pS473 from the IP process by quantifying the ratio. Firstly enrichment is defined and rearranged as:

$$E = \frac{ST_{IP} + S_{IP}}{ST_{wcl} - ST_{FT} + S_{wcl} - S_{FT}} \quad (151)$$

$$= \frac{ST_{IP} + S_{IP}}{(ST_{wcl} + S_{wcl}) \left(1 - \frac{ST_{FT} + S_{FT}}{ST_{wcl} + S_{wcl}}\right)} \quad (152)$$

$$= \frac{ST_{IP} + S_{IP}}{(ST_{wcl} + S_{wcl}) A} \quad (153)$$

$$E \times A = \frac{ST_{IP} + S_{IP}}{ST_{wcl} + S_{wcl}} \quad (154)$$

And similarly:

$$E \times A = \frac{ST_{IP}}{ST_{wcl}} = \frac{S_{IP}}{S_{wcl}} \quad (155)$$

Now the ratio of the band intensity between the IP lanes and the whole cell lysate lanes measures the ratio of the quantities:

$$\frac{\text{IP Band}}{\text{WCL Band}} = \gamma_{\text{Ser}} \frac{S_{IP} + ST_{IP}}{S_{wcl} + ST_{wcl}} = E \times A \quad (156)$$

Thus the enrichment ratio  $E$  can be calculated using this relationship. This ratio of the band intensity was measured to be approximately 5-25 (there is some variation in enrichment between the time points). Since the antibody binding efficiency is approximately 1, the enrichment ratio is thus also approximately 5-25, possibly higher if there is any processing time related intensity loss in the IP lane.

Turning our attention to the pT308 bands, we can take the ratio of the intensity of the flow through lane and the whole cell lysate lane. This can be expressed as:

$$\frac{\text{FT Band}}{\text{WCL Band}} = \gamma_{\text{Thr}} \frac{T_{FT} + ST_{FT}}{T_{wcl} + ST_{wcl}} \quad (157)$$

The factor  $\gamma_{\text{Thr}}$  is for the loss of phosphorylation of the T308 flow through and IP lanes, again due to longer processing times. This expression can be rearranged and reexpressed by noting a few things. Firstly  $T_{FT} = T_{wcl}$  because the singly threonine phosphorylated fraction is not pulled down by the S473 antibody. And:

$$ST_{FT} = ST_{wcl} (1 - A) \approx 0 \quad (158)$$

Since  $A \approx 1$ . Further more, the intensity ratio between the T308 flow through and whole cell lysate bands were measured to be 0.5. So Eqn. 157 can be reexpressed as:

$$\gamma_{\text{Thr}} \frac{T_{\text{wcl}}}{T_{\text{wcl}} + ST_{\text{wcl}}} = \frac{1}{2} \quad (159)$$

For the ratio between the IP and whole cell lysate band, the expression is:

$$\frac{\text{IP Band}}{\text{WCL Band}} = \gamma_{\text{Thr}} \frac{ST_{\text{IP}}}{T_{\text{wcl}} + ST_{\text{wcl}}} \quad (160)$$

The ratio of the intensity of these two bands were measured to be between 0.25 - 0.5. Using this information and the enrichment of the serine fractions between IP and whole cell lysate (Eqn. 155), we obtain:

$$[0.25 - 0.5] = \gamma_{\text{Thr}} \frac{ST_{\text{wcl}}[5 - 25]\gamma_{\text{Ser}}}{T_{\text{wcl}} + ST_{\text{wcl}}} \quad (161)$$

$$[0.01 - 0.1] = \gamma_{\text{Thr}}\gamma_{\text{Ser}} \frac{ST_{\text{wcl}}}{T_{\text{wcl}} + ST_{\text{wcl}}} \quad (162)$$

Now combining Eqn. 159 and 162, we can isolate  $\gamma_{\text{Thr}}$ .

$$[0.01 - 0.1] + \frac{\gamma_{\text{Ser}}}{2} = \frac{\gamma_{\text{Thr}}\gamma_{\text{Ser}}ST_{\text{wcl}}}{T_{\text{wcl}} + ST_{\text{wcl}}} + \frac{\gamma_{\text{Thr}}\gamma_{\text{Ser}}T_{\text{wcl}}}{T_{\text{wcl}} + ST_{\text{wcl}}} \quad (163)$$

$$[0.01 - 0.1] + \frac{\gamma_{\text{Ser}}}{2} = \gamma_{\text{Thr}}\gamma_{\text{Ser}} \frac{T_{\text{wcl}} + ST_{\text{wcl}}}{T_{\text{wcl}} + ST_{\text{wcl}}} \quad (164)$$

$$\gamma_{\text{Thr}} = \gamma_{\text{Ser}}[0.01 - 0.1] + 0.5 \quad (165)$$

Since  $\gamma_{\text{Ser}} \leq 1$  and  $0 \leq [0.01 - 0.1] \leq 0.1$ , we can infer that the fraction of T308 phosphorylation loss due to processing is:

$$0.5 \leq \gamma_{\text{Thr}} \leq 0.6 \quad (166)$$

Finally substituting Eqn. 166 back into 162 leads to:

$$[0.01 - 0.1] = [0.5 - 0.6]\gamma_{\text{Ser}} \frac{ST_{\text{wcl}}}{T_{\text{wcl}} + ST_{\text{wcl}}} \quad (167)$$

$$\frac{[0.017 - 0.2]}{\gamma_{\text{Ser}}} = \frac{ST_{\text{wcl}}}{T_{\text{wcl}} + ST_{\text{wcl}}} \quad (168)$$

Thus the fraction of T308 Akt that is dually phosphorylated is at most 20%. However, it can be as low as about 2%. This is also subjected to the motility of the phosphorylated S473. If the IP process involved some loss of S473 phosphorylation in the IP and flow through conditions, then the fraction of T308 that is dually phosphorylated becomes even lower. Although the *in silico* model produced a wide range of predictions for the fraction of Akt at various states of phosphorylation, they all agree that dually phosphorylated Akt appears in far lower abundances than both of the singly phosphorylated states. This seems to be in agreement with the preliminary immunoprecipitation results. However, more work will need to be performed in future to further validate this computational prediction. For example, this experiment will need to be repeated, and the reverse IP will need to be performed (i.e. immunoprecipitating our the T308 phosphorylation site and then blotting the IP for S473) Furthermore, this experiment does not reveal the relative amount of phosphorylation between S473 and T308 phosphorylation. This ratio is also predicted by the model (there is more S473 phosphorylation than T308) and can be used to validate the model's behaviour.

#### 6.4.2 Determination of AKT activity using *in Vitro* Kinase Assays

*In vitro* kinase assays are a common technique to investigate protein activity [55, 62, 202, 203]. The conventional technique involves the immunoprecipitation of overexpressed, tagged proteins that are used to catalyse a reaction using radioactive ATP (specifically phosphorus-32 substituted into the phosphate group) as the reaction's cofactor. The activity of the immunoprecipitate is then measured by the amount of radioactive  $^{32}\text{P}$  incorporated into the substrate as determined by counting the radioactivity of the substrate band after separation by SDS-PAGE [204].

This technique has two key disadvantages. Firstly, proteins need to be isolated by immunoprecipitation, which means they need to be overexpressed, which has the potential to cause phenotypic changes to the cells. Secondly, the progress curve of the kinase assay can be partially captured because the detection method is destructive. which greatly reduces the temporal resolution of the kinase assay and the accuracy of the activity measurement.

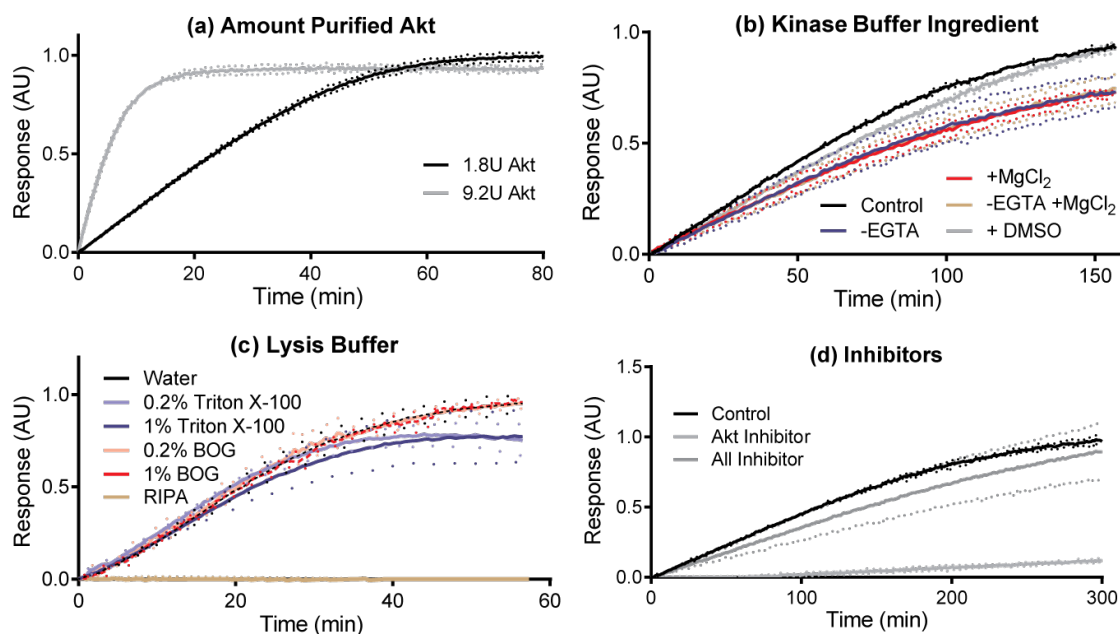


Figure 50: Results of optimisation *In vitro* kinase assay experiments performed using purified active Akt. The y-axis normalised to the initial and maximal fluorescence with recombinant Akt. Dotted lines indicate  $\pm$  one standard deviation of the experimental mean with  $n = 3$ . (a) Optimising the amount of Akt required. (b) Optimising the buffer components. (c) Optimising the lysis buffer used and observing their effects (d) Validation of the inhibitors used and their effects on assay behaviour.

Some of these disadvantages appear to be overcome by an *in vitro* kinase assay that is based on the fluorescent artificial SOX amino acid [205]. When SOX chelates with  $\text{Mg}^{2+}$ , the SOX amino acid increases in fluorescence.  $\text{Mg}^{2+}$  on the other hand, has a higher affinity for phosphopeptides than unphosphorylated substrate peptides. This brings the  $\text{Mg}^{2+}$  close to the SOX amino acid that is adjacent to the phosphorylate residue. Thus, the degree of phosphorylation of a substrate peptide can be measured by the change in fluorescence as the reaction progresses. It is also claimed that this assay can be performed using whole cell lysate, which avoids the need to immunoprecipitate tagged proteins [205]. After considering the advantages, we decided to perform our Akt activity assay using this method.

A number of optimisation experiments were first performed to ascertain the Akt activity required for a detectable SOX-peptide

phosphorylation rate and also variation of the components of the kinase and lysis buffer to maximise the output signal. In a reaction volume of 20  $\mu\text{L}$  it was determined that the reaction can progress up to 3 h before evaporation begins to affect the kinetics of the reaction. Thus, the amount of purified Akt used for optimisation reactions were such that the reaction will reach completion by 3 h. By varying the amount of recombinant Akt (shown in Fig. 50a), it was found that 1.8 U of Akt was the optimal amount of Akt to use for future assays. Using this baseline, the composition of the kinase buffer was tested to maximise the fluorescence signal produced by the assay. Starting at a control composition of 2 mM EGTA and 10 mM  $\text{MgCl}_2$ , the effect of removing EGTA and increasing  $\text{MgCl}_2$  concentration to 20 mM was tested by varying the two concentrations. And finally the effect of adding 0.3% DMSO to the buffer was tested as it is a solvent for many inhibitors and will necessarily be part of the kinase buffer when inhibitors are used in subsequent experiments. The results, shown in Fig. 50b, shows that 2 mM EGTA and 10 mM  $\text{MgCl}_2$  are optimal for fluorescence and the effect of DMSO on Akt activity in the *in vitro* kinase assay was minimal. From this point on, 0.3% DMSO is added to the kinase buffer.

Next the effect of lysis buffer addition is tested in order to determine the optimal lysis buffer to use. More explicitly this is minimal autofluorescence of the buffer, minimal loss of fluorescence over time due to sequestration of SOX-peptides within micelles and minimisation of denaturation of kinase. The common ones we tested were RIPA buffer, Triton X-100, and Octyl- $\beta$ -glucoside (BOG). RIPA buffer is a robust lysis buffer but has the risk of denaturing proteins due to its 0.1% SDS content. Triton X-100 on the other hand does not as readily denature proteins. However, it does have a critical micelle concentration below the concentrations typically used to lyse cells, which can also affect the *in vitro* kinase assay in unfavourable ways. Finally BOG does not have either of these disadvantages but is an expensive detergent. The results (shown in Fig. 50c) show that as predicted, RIPA buffer completely inhibits Akt kinase activity. Triton X-100 and BOG operate in a similar way. However, Triton X-100 appears to cause a reduction in fluorescence that is not concentration dependent. This is potentially due to the eventual sequestration of the SOX-peptide which inhibits its ability to chelate and hence leading to

quenching of the fluorescence signal. The assay with BOG addition responds the same as a reaction without detergent.

Finally, we set out to test the effect of the various inhibitors that is necessary for testing Akt activity with whole cell lysate. In this case we are optimising for an Akt inhibitor that full inhibits Akt activity as well as other AGC kinase inhibitors that minimises off target inhibition of Akt. Because AGC kinases have similar consensus motifs, it is likely a cell lysate lacking in biochemical context will allow other kinases to phosphorylate the Akt SOX-peptide [69]. As such, Akt activity must be isolated by inhibiting other kinases that are not of interest. Alternatively, the degree of non-specific activity can be gauged by quantifying the lysate activity after inhibiting Akt activity. Before this can be done we need to verify that the inhibitors that target Akt efficiently suppresses its activity, and the inhibitors that target other kinases do not significantly inhibit Akt. When this was tested (see Fig. 50d) it was found that the Akt inhibitor GDC-0068 suppresses at least 90% of its activity while the inhibitor cocktail for non-Akt kinases has only a 10% effect on Akt. Thus these strategies for quantification of Akt activity are feasible.

With the optimisation and preliminary validation experiments completed, we moved onto measuring the Akt activity in whole cell lysate (3T3-L1 adipocytes). The following stimulation conditions were tested:

- Unstimulated cell lysate
- 1 nM Insulin stimulation
- 100 nM Insulin stimulation (maximum insulin stimulation)
- 100 nM Insulin stimulation with the PI3K inhibitor Wortmannin (inhibited maximal insulin stimulation).
- 100 nM Insulin stimulation with PP2A phosphatase inhibitor Calyculin A (beyond maximal insulin stimulation)

and for the Calyculin A (Cal A) inhibitor, the kinase assay reaction was run with either the Akt inhibitor or the non-Akt inhibitor cocktail (Fig. 51a). These studies revealed a number of interesting observations. Firstly, there was a significant increase in Akt kinase activity with 100 nM insulin and with Cal A. However, were

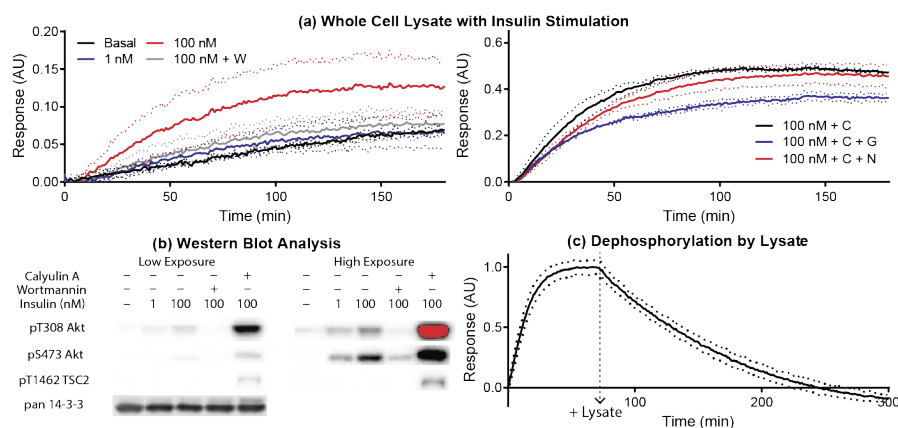


Figure 51: (a) *In vitro* kinase assay results with cell lysates under different conditions, with y-axis normalised to the initial and maximal fluorescence with recombinant Akt. Dotted lines indicate  $\pm$  one standard deviation from the experimental mean with  $n = 3$ . Cells are stimulated with insulin and other drugs. W is Wortmannin during stimulation, and C is Calyculin A during stimulation. G is GDC0068 and is added during the kinase assay only. N is the non-Akt inhibitor cocktail (see method for full list), also added during the kinase assay only. (b) Western blot analysis of the whole cell lysates used in the kinase assays to verify the change in phosphorylation state of Akt and TSC2 and AS160 as an indication of Akt activity. (c) Kinase assay performed with SOX-peptide already phosphorylated by recombinant Akt to demonstrate the phosphatase activity of whole cell lysate stimulated with 100 nM insulin.



was no significant increase in kinase activity at 1 nM insulin probably due to the limited dynamic range of the assay. This is likely due at least in part to the high background level of kinase activity that was evident in the lysate from PI3K inhibited cells, which should not have Akt activity. However, when the Akt inhibitor (marked as G in Fig. 51a) was used in the 100 nM insulin + Cal A condition, the kinase activity was reduced by 40%. However, the non-Akt inhibitor cocktail (marked as N in Fig. 51a) only reduced the lysate activity by 10%. This suggests that 50% of non-specific kinase activity is unaccounted for. This leads to two observations. First, there is significantly more non-specific kinase activity than anticipated, some of which exists in the unstimulated state. Secondly, the small reduction in kinase activity that represents Akt activity would be difficult to measure under physiological doses of insulin. So while the optimisation controls appeared to show a potentially clean *in-vitro* kinase assay that can be performed using whole cell lysate, our results demonstrate that this is experimentally impractical.

In spite of these results, the non-maximal plateau that was achieved by the assay when initiated with whole cell lysate was of interest. In previous optimisation experiments, the fluorescence eventually reached the maximum possible as determined by the concentration of the SOX-peptide. However, when initiated with whole cell lysate, the reaction reaches a non-maximal plateau that was dependent on the kinase activity as measured by the initial rate of increase in fluorescence. This suggests there is cycling in the phosphorylation of the SOX-peptide, which indicates there is phosphatase activity. This was tested by first phosphorylating the SOX-peptide using purified Akt, and then testing the dephosphorylating capacity of whole cell lysate by adding it to the SOX-peptide with the Akt inhibitor. It was found that indeed the fluorescence of the phosphorylated SOX-peptide begins to decrease upon lysate addition eventually reaching a value significantly below zero (which can be accounted for by baseline drift in the instrument and loss of kinase activity after the long reaction time (Fig. 51c).

In light of these results, we set out to develop an alternative strategy for measuring Akt activity in cell lysates. Given the ability of purified Akt to phosphorylate the SOX-peptide, we next looked to purify Akt from cell lysates using immunoprecipitation

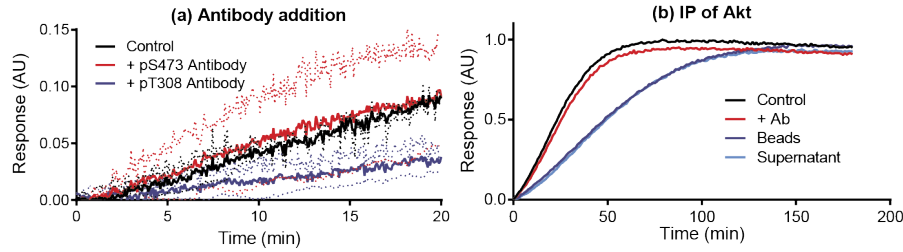


Figure 52: Verification of the compatibility of immunoprecipitation conditions with *in vitro* kinase assay. The y-axis normalised to the initial and maximal fluorescence with recombinant Akt. Dotted lines indicate  $\pm$  one standard deviation from the mean with  $n = 3$ . (a) shows the change in activity of recombinant Akt with antibody addition. (b) shows the efficiency of immunoprecipitation of recombinant Akt where control indicates reaction run with just recombinant Akt, beads indicates addition of empty beads to the control reaction, Ab indicates addition of beads bound with recombinant Akt (without free recombinant Akt) and supernatant is from the immunoprecipitation process (again without further addition of recombinant Akt).

(IP). To determine if Akt can be immunoprecipitated without loss of activity, we first tested the effect of incubating Akt with only the antibodies. Interestingly, the pT308 Akt antibody which target the Akt kinase domain reduced Akt activity (Fig. 52a). This suggests that Akt purified using this antibody would in general not retain its activity. On the other hand, the pS473 antibody (which targets the hydrophobic domain) does not significantly alter Akt kinase activity. With these results, we moved to test the IP efficiency of Akt. In this experiment, the pan-Akt antibody was used instead of the pS473 antibody to capture all Akt possible without regard for its post translational modification. Starting with  $1.7 \mu\text{g}$  of recombinant Akt, we were able to recapture half of the Akt on the protein G beads using the pan-Akt antibody as shown by the kinase assay (shown in Fig. 52b). This is equivalent to a 50% efficiency using the manufacturer recommended antibody concentration.

Based on these results, it can be concluded that the pan-Akt antibody is appropriate for immunoprecipitation quantification of Akt activity. However, we need to determine the feasibility of this using whole cell lysate. To do this, we first verified the amount

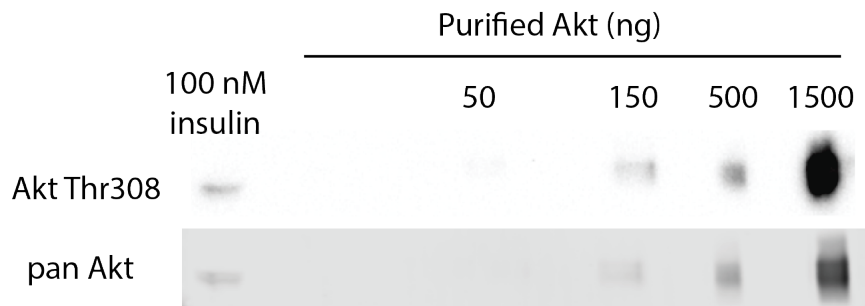


Figure 53: Western blot analysis showing amount of Akt phosphorylated at the Thr308 site and total Akt from 15  $\mu$ g of cell lysate from 3T3-L1 adipocytes stimulated with 100 nM insulin for 10 minutes, and various amount of purified Akt.

of phosphorylation in the purified Akt in order to benchmark the amount of Akt that would need to be immunoprecipitated for use in the *in vitro* kinase assay. We compared the amount of total Akt and Akt phosphorylated at the Thr308 site from 15  $\mu$ g of cell, and various amounts of purified Akt (Fig. 53). Quantification of these lanes band intensity of the Akt originating from the cell lysate is equivalent to the band intensity of about 200 ng of purified Akt, and the phosphorylated fraction is similar between endogenous Akt and the purified Akt. Given that we previously utilised 1.8 U of purified Akt for the *in vitro* kinase assay, this also equates to around 200 ng of purified Akt. Thus, the amount of Akt available in cell lysates is sufficient for use with the *in vitro* kinase assay subject to efficiency of the immunoprecipitation process. However, it should be noted that this was compared to Akt from cells stimulated with 100 nM insulin. The activity of Akt under 1 nM insulin will be far lower, thus more Akt will be required to achieve the required activity for use with the *in vitro* kinase assay.

## 6.5 DISCUSSION

In this chapter, we have developed a basic model of the Insulin Signalling pathway that is specific to the 3T3-L1 adipocyte cell model. By focusing on the overshoot feature in the temporal behaviour of phosphorylated Akt, we identified that the phosphorylation time profile of Akt substrates could not be correctly sim-

ulated in the *in silico* model. This component of the system was then studied in more detail by constructing and investigating an Akt and Akt substrate only model. By focusing the study to a small component of the system, it was possible to incorporate and investigate the effects of more complex physicochemical mechanisms, such as translocation and independent and unordered phosphorylation of Akt.

By performing our *in silico* investigation in this way, we were able to identify translocation of Akt to the membrane as a possible mechanism for blunting the overshoot observed by Akt substrates. Our original hypothesis to explain this disconnect between Akt and its substrates was that translocation of Akt delays the phosphorylation time profile between the plasma membrane portion and the cytosol portion of Akt. However, this is equivalent to a delay in the phosphorylation of the Akt substrates caused by a slower reaction rate, and the delay required was shown to be significantly slower than the phosphorylation rate observed experimentally. What was found instead, was that translocation served as a sequestration mechanism of limited cytosolic Akt. Although total Akt translocation, and hence Akt phosphorylation, exhibits an overshoot behaviour, this behaviour is dominated by the plasma membrane portion.

In the cytosol, the abundance of the phosphorylated form of Akt is governed by two mechanism. Firstly, the rate of phosphorylation at the plasma membrane. An equilibrium exists between plasma membrane bound total Akt and cytosolic total Akt, thus the amount of phosphorylated Akt in the cytosol will increase as phosphorylated Akt increases in the plasma membrane. Secondly, the equilibrium between the plasma membrane bound total Akt and cytosolic Akt can vary as the amount of PIP3 generated varies. In this system, the two mechanisms act as opposing systems. The phosphorylation necessarily increases cytosolic phosphorylated Akt abundance while PIP3 generation decreases it. In this way, the overshoot in PIP3 generation actually causes an undershoot in cytosolic phosphorylated Akt if it is sufficiently abundant.

With this result, the implication would be that active Akt is only able to access its substrates in the cytosol and that Akt activity in the cytosol does not exhibit the overshoot behaviour. Although this would also suggest that substrate phosphorylation would ex-

hibit an undershoot behaviour, the undershoot can be overcome in one of two ways. Firstly, the system can strike a balance between the degree of PIP3 overshoot and the Akt phosphorylation rate at the plasma membrane. Secondly, the substrates can roughly ignore the undershoot by having a slower phosphorylation rate, which is effective in blunting the signaling in this case because the undershoot occurs in a shorter time scale than the overshoot that causes it.

So the current understanding of the Akt system is an overshoot in PIP3 generation has the potential to suppress the abundance of Akt phosphorylation in the cytosol, and the majority of substrate phosphorylation occurs in the cytosol. If we consider a stronger dose of insulin which will cause a greater PIP3 generation rate, this will lead to a faster Akt phosphorylation rate simply because there is higher concentration of Akt in the plasma membrane leading to a higher phosphorylation probability. This also leads to more sequestration of the cytosolic Akt. It is likely the system has been optimised to ensure the sequestration effects of PIP3 generation does not overcome the increased phosphorylation. However, if this system was hyper stimulated such that PIP3 abundance is increased to sufficiently high levels, then the degree of Akt sequestration could become high enough such that cytosolic active Akt abundance could actually decrease below normal levels. Hence, this could represent a form of insulin resistance.

This possibility would make an interesting hypothesis for further study. However, a potential problem with the Akt focused model is it requires specifically PIP3 generation. However, the current data used is Akt translocation as opposed to PIP3 generation directly, and the former is only a surrogate measure of the latter. Under well behaved conditions, this is not a concern as Akt and PIP3 generation should have correlating temporal profiles. However, under less ideal conditions, such as maximal insulin dose or even insulin resistance conditions, the two temporal profiles can potentially become difference due to changes in regulation. Thus the true temporal profile of PIP3 generation would need to be measured for use with the Akt model. Alternatively, the Akt model could be reincorporated into the insulin Signalling pathway model, in which case the kinetics of PIP3 generation and its associated overshoot behaviour then becomes dependent on the mechanisms of the negative feedbacks in the system.

There were a number of ways to validate this model. One of the implications of this mechanism is that a small amount of total Akt is phosphorylated, and an even smaller fraction is doubly phosphorylated. Our preliminary results using immunoprecipitation suggests that this also occurs *in vitro*. Secondly we wished to determine the change in Akt activity over time. Unfortunately we were not able to verify this because of the non-specific nature of the *in vitro* kinase assay that was chosen. However, in this experiment we would expect the activity of unfractionated Akt to also overshoot. This would present the first stage of further experimentation that would arise from this work.

Although we have adopted the most widely accepted negative feedback mechanism in our model (i.e. mTORC1 mediated feedback), there are alternative mechanisms that have been proposed in the literature [74, 206, 207]. One prominent case is the endocytosis of the IR-IRS1 complex, and modelled *in silico* by Brannmark et al [124]. It would be expected that the two different models would result in differences in kinetic behaviour, which would form the basis of further validating experiments. Beyond this there are other proposed feedback mechanisms that can be added to the model. An example is the proposed positive feedback mechanism on Akt, by Akt activation of mTORC2, which is itself an activator of Akt [14]. There is also the translocation of PDK1 to the plasma membrane with PIP3 generation, which can also further increase complexity in the Akt focused model [73].

Finally, we would like comment on the specific finding that this study has achieved, that is a mechanism by which an overshoot in Akt phosphorylation can be subsequently blunted at the substrate level. This is not behaviour that has been explored in other models of insulin signalling. One reason for this is alternative models are based on data originating from different cell types, which exhibit correspondingly different behaviour. An example of this is the Kubota et. al. model, which were performed in Fao cells. This cell type in fact exhibits overshoot behaviour at the substrate level, specifically in pGSK3 [74]. This highlights the difference model cell types can cause in the mechanisms that can be revealed in mechanistic models. Secondly, the effect was seen in other models, but were not investigated as the focus of their study. An example of this is the Brannmark model, which focused on the nodes that cause T2D, although their model produced a

similar behaviour, specifically the overshoot in Akt phosphorylation, and concurrent overshoot in substrate phosphorylation. This highlights the utility of mechanistic models and their breadth of potential research directions they can uncover.





## CONCLUSIONS AND FUTURE WORK

---

### 7.1 CONCLUSIONS

In Chapter 1 of this thesis, we posed two questions:

1. Are existing modelling methods sufficiently accurate, and if they are not, can they be improved?
2. What does a quantitative model of the common cellular model system for studying insulin signalling and insulin resistance, the 3T3-L1 adipocyte, look like?

To answer the first question, we investigated the commonly used Michaelis Menten model of enzyme kinetics. This was firstly because enzyme kinetics covers a large proportion of biochemical reactions that are relevant to the insulin signalling pathway in adipocytes. Secondly, the Michaelis Menten model is the most commonly accepted model used to simulation enzyme kinetic reactions. In our investigation we found that the model's assumption that the enzyme-substrate complex is negligible leads to inaccuracies in simulated results. In fact, this drawback with the Michaelis Menten model has already been discussed in the literature [139]. From this motivation, we sought to improve the Michaelis Menten model by explicitly accounting for the complex concentration in quasi-steady state. In doing so we developed the differential quasi-steady state assumption (dQSSA) enzyme kinetic model, which was found to be more accurate than the Michaelis Menten model when tested *in silico* and *in vitro*.

This is a potentially useful result because mathematical modelling is a complex exercise. The main aim of performing this type of study is to better understand the way information flows through the biological system and the regulatory mechanisms involved [172]. However, in practice parameter identifiability and

the accuracy of the underlying mathematical model needs to also be considered and can confound the result of the model [139, 147, 171]. The development of the dQSSA allows at least some portion of these confounding factors to be solved. Furthermore, the existing models of enzyme kinetics, and in fact biochemical models, are focused on single substrate enzyme kinetic reactions. The dQSSA we developed has both a single substrate and two substrate form. This enables biochemical models to be expanded such that they begin to include the contribution of cofactors such as global ATP levels in the cell. These components of the model up until now have been largely abstract if they are even considered at all [102].

To answer the next question, we needed to develop a mathematical model of the insulin signalling pathway. However, construction of the mathematical model from first principles is a tedious and difficult task and is not ideal there is a high risk of introducing errors when constructing the model [89, 182]. This has been recognised in the literature and led to the development of the SBML format [179]. Unfortunately, the algorithms that are currently available do not have the flexibility required to encode the dQSSA model. Thus, an analogous modelling algorithm was developed: the SigMat algorithm.

Using the SigMat algorithm, we have been able to apply the dQSSA model on all enzyme kinetic type reactions in the insulin signalling pathway in a relatively simple yet consistent way. The confounding effects of parameter and kinetic model selection were mitigated by coupling this algorithm with the more exploratory Markov Chain Monte Carlo method of parameter fitting when fitting the data. Thus, it was possible to focus on the ability of the topology of the model studied to reproduce the experimental data. It was using this strategy that we were able to discover that the detailed mechanistic detail involving Akt activation is in fact important. The inclusion of this detail revealed that translocation has a potential sequestration effect on cytosolic Akt, which allows it to shield the overshoot behaviour it experiences from its substrates. Although translocation as a necessary mechanism for Akt activation is well known, the effect that translocation has on the behaviour of the signalling pathway is new insight that could not have otherwise been revealed without the use of accurate kinetic modelling [78, 208]. The development of this model has also

let to other new insights, such as the ratio in which Akt is divided between its various combinations of phosphorylation states. Preliminary data presented in this thesis suggests this insight may be true in vivo, which would be one validating evidence for the Aky model developed in this thesis.

Comparison between our model of insulin signalling to other models of insulin signalling actually reveals some interesting differences in regard to the differences between cell types. The overshoot in Akt is observed in the Kubota model but not the Brannmark or Dalle Pezze models [74, 124, 127]. And in the case of the Kubota model, the overshoot is transmitted downstream, unlike what is observed in the 3T3-L1 adipocyte data. This shows that these cell types do exhibit different behaviours even when the same pathway is investigated. Based on this observation it is unclear if core pathways, such as the insulin signalling pathway, are truly conserved between different cell types. In the Kubota, Brannmark and Dalle Pezze models, the Akt modules following the widely held assumption that Akt phosphorylation directly relates to Akt activity [199, 200], and Akt directly activates its substrates. In all of these models, these assumptions do not lead to disagreement between the model and the experimental data. However, in our model, these two assumptions resulting in conflicting simulation and experimental data. This implies that additional mechanisms must be involved in the insulin signalling pathway within an 3T3-L1 adipocyte and is yet to be discovered. On top of this however, our model result would also question whether or not these putative additional mechanisms also exist in the cell types studied by Kubota, Brannmark and Dalle Pezze, but simply played no regulatory roles in those cell types. These questions would lead to some interesting directions for future research into the workings of the insulin and Akt signalling pathways.

## 7.2 FUTURE WORK

Looking forward beyond this thesis, there are a number of questions that the insulin signalling model developed in this thesis has raised. Firstly, the model was designed with the mTORC1-IRS1 negative feedback mechanism. This was necessary in order

to produce the overshoot behaviour observed during Akt phosphorylation. There are a number of studies that suggest this occurs [53, 73]. However, this negative feedback is also controversial as other studies have suggested that the negative feedback is IRS1 independent [209]. In fact, the Brannmark model of the insulin signalling pathway does not incorporate the mTORC1-IRS1 negative feedback. It utilises endocytosis of the IR-IRS1 complex as the mechanism for generating an overshoot behaviour downstream [124, 149]. These vastly different hypotheses demand further investigation using mathematical modelling. With two competing models of the underlying mechanism, it would be highly possible to compare the predictions between the two models and use them as the basis for model validation and comparison. There is the added benefit that both our model and the Brannmark model were constructed using a parameter free core prediction method, which means any differences in predicted outcome can be attributed directly to model topology rather than parameter choice.

An additional avenue for extending our insulin signalling model is through expanding the downstream components of the pathway. A particularly important path is extension of the AS160 arm, which leads to GLUT4 translocation and ultimately glucose uptake [26]. This is a vital component of the pathway, particularly in relation to diabetes because it is related to the ability of adipocytes to maintain glucose homeostasis in the body. Although knowledge of this component of the pathway, that is the mechanism of GLUT4 translocation, docking and fusion with the plasma membrane is still largely incomplete, current advances in microscopy techniques will be pushing the frontiers of this field in the near future.

Also, the use of mass spectrometry could be used to generate the vast amounts of data required to tune the network. This would enable large portion of the network to be mechanistically interrogated at once. Firstly, the ability for quantitative proteomic techniques to determine the abundance of all proteins within a sample. This can be used to define many of the concentration parameters in the kinetic model. Secondly, the use of temporal phospho-proteomics data at multiple doses would be an excellent replacement for Western Blot analysis data due to its high throughput nature. Although phospho-proteomic profile of the

3T3-L1 adipocyte has already been performed, the key requirement is this data at multiple doses, as we have done in investigating the original insulin signalling pathway model [14]. Finally, these two approaches can be combined to interrogate the causes of insulin resistance building kinetic models of cell models of insulin resistance. By combining proteomics and phosphoproteomic data obtained of each model of insulin resistance, it would be possible to construct kinetic models of insulin resistance under the assumption that the protein abundance is as found in the proteomic study and rate parameters remain the same across the different cell models. In doing so, it would be possible to determine if insulin resistance is purely due to a change in the proteome of the cells, or if the regulatory mechanisms within the cell are intrinsically changed.

In terms of the dQSSA and the theoretical aspects of kinetic modelling, more work can be performed to better understand the wider power of this new kinetic modelling, and validation of its use within an *in vivo* context. While this thesis has provided evidence with *in vitro* experiments, it is unclear under which biological systems it will provide different predictions compared to a Michaelis Menten or tQSSA implementation. While the use of relative protein abundances obtained from proteomic data can offer some hints into systems that may violate the reactant stationary assumption and thus require use of the dQSSA to accurately model, it is still necessary to validate this in a controlled experiment.

Finally, further development of the SigMat would aid in simplifying the use of the dQSSA and be accessible for the wider scientific community. The most important extension of the SigMat model would be to achieve compatibility with the SBML format. This is crucial because the SBML format has been a standard in the systems biology field for over a decade, and thus many biological models have been designed and communicated using this format. Compatibility with this format would enable the dQSSA to be made available to the vast number of SBML models in the literature. Furthermore, the SigMat model, currently written in the MATLAB environment, is not very accessible to the research community because MATLAB is proprietary software. If the software can be migrated to an open source programming language, such as Python, then the accessibility of the algorithm would in-

crease. The combination of SBML compatibility and migration of the SigMat code can allow the dQSSA method to become more widely adopted in the research community.

Here we have listed only a small sample of the potential applications of the foundational insulin signalling pathway model that was developed in this thesis. Although some of these applications are likely to be ambitious and long term projects, the scope of expansion of these models are limitless and this can be realised through the modelling approaches applied and demonstrated in this thesis.

## ADDITIONAL METHODS

---

### a.1 CELL CULTURING

3T3-L1 fibroblasts were cultured in 2 mL of culture medium which is made of high glucose Dulbecco's Modified Eagle Medium (DMEM, Gibco, 11965), 10% FCS (Gibco, 16000-044), and 100X Glutamax (Gibco, 35050) in 6 well plates (Costar 3516) or 3cm petri dishes (Falcon 3001) at 37°C and 10% CO<sub>2</sub>. Fibroblasts were passaged at 50-70% confluence (around every 48 hours).

3T3-L1 fibroblasts were prepared for differentiation into adipocytes by reseeding into fresh plates and then cultured until confluence is reached. Differentiation is then initiated by replacing the culture medium with differentiation medium which is made of DMEM, 0.22  $\mu$ M Dexamethasone (Sigma Aldrich, D4902), 100 ng/mL Biotin (Sigma Aldrich, B4639), 2  $\mu$ g/mL Insulin (Calbiochem) and 500  $\mu$ M Isobutyl-1-methyl-xanthine (IBMX, Sigma Aldrich, I5879). After 72 hours the differentiation medium is replaced with post-differentiation medium which is made up of DMEM, and 2  $\mu$ g/mL Insulin. After a subsequent 72 hours, the post-differentiation medium is replaced with culture medium. This is replaced every 48 hours. After 10-12 days adipocytes were used.

### a.2 TIME COURSE EXPERIMENT

A water bath was preheated to 37°C for 1 hour to allow its temperature to stabilise and leveled to ensure its platform is flat. Cell culture medium was replaced with basal medium which is made up of 0.22  $\mu$ m filter sterilised Krebs-Ringer-Phosphate buffer (KRP, 12.5 mM HEPES [Millipore 391340], 0.4 mM NaH<sub>2</sub>PO<sub>4</sub> [Sigma Aldrich 71505], 0.6 mM Na<sub>2</sub>HPO<sub>4</sub> [Sigma Aldrich S7907], 6 mM KCl [Univar 383-500g], 120 mM NaCl [Sigma Aldrich S9888], 1.2 mM MgSO<sub>4</sub> [Sigma Aldrich M9397], 1 mM CaCl<sub>2</sub> [Sigma Aldrich C5080]) and 0.2% Bovine Serum Albumin (BSA, Bovostar

BSAS1.0). This was done by removing cell culture media, then performing three washing with Phosphate buffered solution (PBS, 150 mM NaCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2.5 mM KCl, 1.75 mM KH<sub>2</sub>PO<sub>4</sub> [Sigma Aldrich 795488]), and then adding 2 mL of basal medium. The plates were then transferred to the preheated water bath. The water level of the water bath it adjusted such that it reaches the same water level as the medium in the cell culture. 6 mL of 3 nM Insulin in KRP was also prepared and placed in the water bath in 6 well plates. Adipocytes were serum-starved in this way for 2 hours.

After serum starvation, cells were stimulated by addition of 1 mL of 3 nM Insulin while in the water bath and then incubated for the required amount of time. Immediately after the incubation the media was removed and the stimulation was quenched by placing the cells on ice and washing the cells three times with ice cold PBS. Cells were then harvested by either adding 400  $\mu$  L of 1% Triton X-100 (50 mM Tris [Univar 2311] pH 7.5 [neutralised with NaOH], 150 mM NaCl, 1% Triton X-100 [Sigma Aldrich X100-500mL]) or 1% Octyl- $\beta$ -glucoside (50 mM Tris [Univar 2311] pH 7.5 [neutralised with NaOH], 150 mM NaCl, 1% Octyl- $\beta$ -glucoside [Sigma Aldrich O8001]) for in vitro kinase assay experiments, or Radio-Immuno-Precipitation-Assay (RIPA, 50 mM Tris pH 7.5 [neutralised with NaOH], 150 mM NaCl, 1% NP40 [Sigma 13021], 0.5% Sodium Deoxycholate [D6750], 0.1% SDS [Sigma Aldrich L4509], 1 mM [Amresco 0105], 1% Glycerol [Sigma G5516]) for immunoprecipitation and western blotting experiments, both with protease inhibitors (Roche, 11873580001) and phosphatase inhibitors (2 mM Sodium Orthovanadate [Sigma Aldrich S6508, first activated at 95°C then neutralised with HCl], 1 mM Sodium Pyrophosphate [Sigma Aldrich S6422], 1 mM Ammonium Molybdate [Sigma Aldrich A7302], 10 mM Sodium Fluoride [Fluka 71522]). Cells were then scrapped down and transferred to a 1.5 mL centrifuge tube (Axygen MCT-150-C).

Cell lysates were then homogenised by sonicating the samples for 6 s, made up of a 1 s pulse followed by 3 s interval for a total time of 21 s (total energy of 100 J deposited). Insoluble fractions and lipids separated by centrifuging the samples at 21,000  $\times$  g for 30 minutes. Lipids were then removed by carefully aspirating the resulting milky suspension. The supernatant is then transferred to a clean centrifuge tube. The samples were then quantified



using the Pierce bicinchroninic acid assay (ThermoFisher Scientific 23224).

### a.3 WESTERN BLOT ANALYSIS

Proteins were separated by SDS-PAGE then transferred to polyvinylidene difluoride membranes (Merck Millipore, IPVH00010). Membranes were blocked in blocking buffer (5% skim milk [coles], TBST: Tris-buffered saline with 0.1

The following antibodies were used in the analysis. pan-14-3-3 (Rabbit, Santa Cruz Biotechnology, sc-629), pan-Akt (Rabbit, Cell Signalling Technology, 4685), phospho-Akt Ser473 (Mouse, Cell Signalling Technology, 4051), phospho-Akt Thr308 (Rabbit, Cell Signalling Technology, 9275), phospho-AS160 Thr642 (Rabbit, Cell Signalling Technology, 4288).

### a.4 TOTAL INTERNAL REFLECTION MICROSCOPY (TIRF)

TagRFPt-Akt2 was cloned using Gibson assembly from pDEST53-eGFP-Akt2 [210]. The pDEST53-Akt2 fragment was cloned using the primers TCTTCGCCCTTAGACACCATGTCTCCCTATAGTGAGTC and TAATGGCATGGACGAGCTGTACAAGAATGAGGTGTCTGT-CATCAAAG. TagRFPt fragment was cloned from pGEM-T-tagRFPt using the primers GACTCACTATAGGGAGACATGGTGTCTAAGGGC-GAAGA and CTTTGATGACAGACACCTCATTCTTGTACAGCTCGTC-CATGCCATTA.

pDEST53-eGFP-Akt2 was cloned using Gibson assembly from pDEST53-Cycle3\_GFP-Akt2. The pDEST53-Akt2 fragment was cloned using the same primers as previously described. The eGFP fragment was cloned from pEGFP-C1 (Clontech) using the primers CCAAGCTGGCTAGACACCATGGTGTGAGCAAGGGCGAGGA and CTTTGATGACAGACACCTCATTCTTGTACAGCTCGTCCATGCCATTA.

pDEST53-Cycle3\_GFP-Akt2 was cloned using the gateway technique where human Akt2 in the pDONR221 backbone was inserted into pcDNA-DEST53 vector.

Differentiated cells were transfected with the tagRFPt-Akt2 fluorescent construct, seeded onto coverslips, and imaged with Total Internal Reflection Fluorescence (TIRF) microscopy as described

in [211]. Imaging was performing with the excitation wavelength at 555 nm and emission wavelength at 584 nm.

## ADDITIONAL DQSSA DERIVATIONS

---

### b.1 TWO ENZYME ONE SUBSTRATE MODEL

Beginning with:

$$[S_T]_0 = [S] + [E_1S] + [E_2S] + [P_1] + [P_2] \quad (169)$$

$$[E_{1T}]_0 = [E_1] + [E_1S] \quad (170)$$

$$[\dot{P}_1] = k_1^c [E_1S] \quad (171)$$

$$[E_1\dot{S}] = k_1^a [S][E_1] - (k_1^d + k_1^c) [E_1S] \quad (172)$$

$$[E_{2T}]_0 = [E_2] + [E_2S] \quad (173)$$

$$[\dot{P}_2] = k_2^c [E_2S] \quad (174)$$

$$[E_2\dot{S}] = k_2^a [S][E_2] - (k_2^d + k_2^c) [E_2S] \quad (175)$$

We find the functional form for the complex concentrations under the quasi-steady state assumption. Setting equations 172 and 175 to zero we obtain:

$$[E_1S] = \frac{[S][E_1]}{K_1^m} \quad (176)$$

$$[E_2S] = \frac{[S][E_2]}{K_2^m} \quad (177)$$

The rate of change of the complex concentrations under the quasi-steady state assumption are then:

$$[E_1\dot{S}] = \frac{\partial [E_1S]}{\partial [E_1]} [\dot{E}_1] + \frac{\partial [E_1S]}{\partial [S]} [\dot{S}] = \frac{[S]}{K_1^m} [\dot{E}_1] + \frac{[E_1]}{K_1^m} [\dot{S}] \quad (178)$$

$$[E_2\dot{S}] = \frac{\partial [E_2S]}{\partial [E_2]} [\dot{E}_2] + \frac{\partial [E_2S]}{\partial [S]} [\dot{S}] = \frac{[S]}{K_2^m} [\dot{E}_2] + \frac{[E_2]}{K_2^m} [\dot{S}] \quad (179)$$

Then differentiating Eqns 169, 170 and 173 and substituting the above (Eqns 190 and 193) leads to:

$$\begin{aligned} [\dot{S}] &= -[E_1 \dot{S}] - [E_2 \dot{S}] - [\dot{P}_1] - [\dot{P}_2] \\ &= -\frac{[S]}{K_1^m} [\dot{E}_1] - \frac{[E_1]}{K_1^m} [\dot{S}] - k_1^c [E_1 S] - \frac{[S]}{K_2^m} [\dot{E}_2] - \frac{[E_2]}{K_2^m} [\dot{S}] - k_2^c [E_2 S] \end{aligned} \quad (180)$$

$$\begin{aligned} [\dot{E}_1] &= -[E_1 \dot{S}] \\ &= -\frac{[S]}{K_1^m} [\dot{E}_1] - \frac{[E_1]}{K_1^m} [\dot{S}] \end{aligned} \quad (181)$$

$$[\dot{P}_1] = k_1^c [E_1 S] \quad (182)$$

$$[E_1 \dot{S}] = \frac{[S]}{K_1^m} [\dot{E}_1] + \frac{[E_1]}{K_1^m} [\dot{S}] \quad (183)$$

$$\begin{aligned} [\dot{E}_2] &= -[E_2 \dot{S}] \\ &= -\frac{[S]}{K_2^m} [\dot{E}_2] - \frac{[E_2]}{K_2^m} [\dot{S}] \end{aligned} \quad (184)$$

$$[\dot{P}_2] = k_2^c [E_2 S] \quad (185)$$

$$[E_2 \dot{S}] = \frac{[S]}{K_2^m} [\dot{E}_2] + \frac{[E_2]}{K_2^m} [\dot{S}] \quad (186)$$

Placing all derivative terms on the left hand side:

$$\frac{[S]}{K_1^m} [\dot{E}_1] + \frac{[S]}{K_2^m} [\dot{E}_2] + \frac{[E_1]}{K_1^m} [\dot{S}] + \frac{[E_2]}{K_2^m} [\dot{S}] + [\dot{S}] = -k_1^c [E_1 S] - k_2^c [E_2 S] \quad (187)$$

$$\frac{[S]}{K_1^m} [\dot{E}_1] + [\dot{E}_1] + \frac{[E_1]}{K_1^m} [\dot{S}] = 0 \quad (188)$$

$$[\dot{P}_1] = k_1^c [E_1 S] \quad (189)$$

$$-\frac{[S]}{K_1^m} [\dot{E}_1] - \frac{[E_1]}{K_1^m} [\dot{S}] + [E_1 \dot{S}] = 0 \quad (190)$$

$$\frac{[S]}{K_2^m} [\dot{E}_2] + [\dot{E}_2] + \frac{[E_2]}{K_2^m} [\dot{S}] = 0 \quad (191)$$

$$[\dot{P}_2] = k_2^c [E_2 S] \quad (192)$$

$$-\frac{[S]}{K_2^m} [\dot{E}_2] - \frac{[E_2]}{K_2^m} [\dot{S}] + [E_2 \dot{S}] = 0 \quad (193)$$

As a matrix equation:

$$\begin{bmatrix}
 1 + \frac{[E_1]}{K_1^m} + \frac{[E_2]}{K_2^m} & \frac{[S]}{K_1^m} & 0 & 0 & \frac{[S]}{K_2^m} & 0 & 0 \\
 \frac{[E_1]}{K_1^m} & 1 + \frac{[S]}{K_1^m} & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 -\frac{[E_1]}{K_1^m} & -\frac{[S]}{K_1^m} & 0 & 1 & 0 & 0 & 0 \\
 \frac{[E_2]}{K_2^m} & 0 & 0 & 0 & 1 + \frac{[S]}{K_2^m} & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 -\frac{[E_2]}{K_2^m} & 0 & 0 & 0 & -\frac{[S]}{K_2^m} & 0 & 1
 \end{bmatrix}
 \begin{bmatrix}
 [\dot{S}] \\
 [\dot{E}_1] \\
 [\dot{P}_1] \\
 [\dot{E}_1S] \\
 [\dot{E}_2] \\
 [\dot{P}_2] \\
 [\dot{E}_2S]
 \end{bmatrix}
 =
 \begin{bmatrix}
 0 & 0 & 0 & -k_1^c & 0 & 0 & -k_2^c \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & k_1^c & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & k_2^c \\
 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{bmatrix}
 \begin{bmatrix}
 [S] \\
 [E_1] \\
 [P_1] \\
 [E_1S] \\
 [E_2] \\
 [P_2] \\
 [E_2S]
 \end{bmatrix}
 \quad (194)$$

The matrices can be broken up into:

$$\begin{aligned}
 & \begin{bmatrix} 1 + \frac{[E_1]}{K_1^m} + \frac{[E_2]}{K_2^m} & \frac{[S]}{K_1^m} & 0 & 0 & \frac{[S]}{K_2^m} & 0 & 0 \\ \frac{[E_1]}{K_1^m} & 1 + \frac{[S]}{K_1^m} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{[E_1]}{K_1^m} & -\frac{[S]}{K_1^m} & 0 & 1 & 0 & 0 & 0 \\ \frac{[E_2]}{K_2^m} & 0 & 0 & 0 & 1 + \frac{[S]}{K_2^m} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -\frac{[E_2]}{K_2^m} & 0 & 0 & 0 & -\frac{[S]}{K_2^m} & 0 & 1 \end{bmatrix} = \delta_{ij} + \\
 & \hspace{15em} (195) \\
 & \begin{bmatrix} \frac{[E_1]}{K_1^m} & \frac{[S]}{K_1^m} & 0 & 0 & 0 & 0 & 0 \\ \frac{[E_1]}{K_1^m} & \frac{[S]}{K_1^m} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{[E_1]}{K_1^m} & -\frac{[S]}{K_1^m} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} \frac{[E_2]}{K_2^m} & 0 & 0 & 0 & \frac{[S]}{K_2^m} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{[E_2]}{K_2^m} & 0 & 0 & 0 & \frac{[S]}{K_2^m} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{[E_2]}{K_2^m} & 0 & 0 & 0 & -\frac{[S]}{K_2^m} & 0 & 0 \end{bmatrix}
 \end{aligned}$$

The first term is the same as Eqn. 60 . The second term is the same as Eqn. 60 with the  $[E_1]$ ,  $[E_1S]$  and  $[P_1]$  indices replaced with the

$[E_2]$ ,  $[E_2S]$  and  $[P_2]$  indices. And similarly for the right hand side matrix:

$$\begin{aligned}
 & \begin{bmatrix} 0 & 0 & 0 & -k_1^c & 0 & 0 & -k_2^c \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_1^c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k_2^c \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \\
 & \begin{bmatrix} 0 & 0 & 0 & -k_1^c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_1^c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -k_2^c \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k_2^c \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (196)
 \end{aligned}$$

The first term is the same as Eqn. 61. The second term is the same as Eqn. 61 with the  $[P_1]$  indices replaced with the  $[P_2]$  indices. And similarly for the right hand side matrix:





# DQSSA VALIDATION MODEL DEFINITIONS

---

## c.1 IN SILICO VALIDATION

### c.1.1 dQSSA Model

```
function dx_dt = validatedQSSA(t,x,k,f)
x(x<0) = 0;

% Select variation method to use for dQSSA calculation.
% vartn = 1 explicitedly models the complex
% vartn = 2 implicitly models the complexes
vartn = 1;

%% Definition of states
%x(1) = I
%x(2) = A
%x(3) = pA
%x(4) = B
%x(5) = p1B
%x(6) = p2B
%x(7) = C
%x(8) = D
%x(9) = CD
%x(10) = pCD
%x(11) = A-B
%x(12) = A-p1B
%x(13) = I-CD
%x(14) = I-pCD
%x(15) = pCD-A
%x(16) = pCD-pA
%x(17) = pCD-B
%x(18) = pCD-p2B
%x(19) = p1B-B
%x(20) = p1B-p2B

%% Initialisation
B = zeros(length(x));
M = 0*B;
L = M;

%% Creating Tensors
if vartn ==1
    % Variation 1. Complex explicitly modelled
    G = [11 2 4 -1/k( 2); %Enzyme Reaction 1f
        11 4 2 -1/k( 2);
        4 2 4 1/k( 2);
        4 4 2 1/k( 2);
```

```

2 2 4 1/k( 2);
2 4 2 1/k( 2);
12 2 5 -1/k( 4); %Enzyme Reaction 1r
12 5 2 -1/k( 4);
5 2 5 1/k( 4);
5 5 2 1/k( 4);
2 2 5 1/k( 4);
2 5 2 1/k( 4);
13 1 9 -1/k( 9); %Enzyme Reaction 5f
13 9 1 -1/k( 9);
9 1 9 1/k( 9);
9 9 1 1/k( 9);
1 1 9 1/k( 9);
1 9 1 1/k( 9);
14 1 10 -1/k(11); %Enzyme Reaction 5r
14 10 1 -1/k(11);
10 1 10 1/k(11);
10 10 1 1/k(11);
1 1 10 1/k(11);
1 10 1 1/k(11);
15 10 2 -1/k(14); %Enzyme Reaction 7f
15 2 10 -1/k(14);
2 10 2 1/k(14);
2 2 10 1/k(14);
10 2 10 1/k(14);
10 10 2 1/k(14);
16 10 3 -1/k(16); %Enzyme Reaction 7r
16 3 10 -1/k(16);
3 10 3 1/k(16);
3 3 10 1/k(16);
10 3 10 1/k(16);
10 10 3 1/k(16);
17 10 4 -1/k(19); %Enzyme Reaction 9f
17 4 10 -1/k(19);
4 10 4 1/k(19);
4 4 10 1/k(19);
10 4 10 1/k(19);
10 10 4 1/k(19);
18 10 6 -1/k(21); %Enzyme Reaction 9r
18 6 10 -1/k(21);
6 10 6 1/k(21);
6 6 10 1/k(21);
10 6 10 1/k(21);
10 10 6 1/k(21);
19 5 6 -1/k(23); %Enzyme Reaction 10f
19 6 5 -1/k(23);
5 5 6 1/k(23);
5 6 5 1/k(23);
6 5 6 1/k(23);
6 6 5 1/k(23);
20 5 4 -1/k(25); %Enzyme Reaction 10r
20 4 5 -1/k(25);
5 5 4 1/k(25);
5 4 5 1/k(25);
4 5 4 1/k(25);
4 4 5 1/k(25);
];

G(isnan(G(:,4)),4) = 0;

```

```

A = [7 7 8 -k(6) ; %Reaction 3
      8 7 8 -k(6) ;
      9 7 8 k(6)];

B(5,11) = k(1); %Reaction 1f
B(4,11) = -k(1);
B(5,12) = -k(3); %Reaction 1r
B(4,12) = k(3);
B(4,5) = k(5); %Reaction 2
B(5,5) = -k(5);
B(7,9) = k(7); %Reaction 4
B(8,9) = k(7);
B(9,9) = -k(7);
B(10,13) = k(8); %Reaction 5f
B(9,13) = -k(8);
B(10,14) = -k(10); %Reaction 5r
B(9,14) = k(10);
B(9,10) = k(12); %Reaction 6
B(10,10) = -k(12);
B(3,15) = k(13); %Reaction 7f
B(2,15) = -k(13);
B(3,16) = -k(15); %Reaction 7r
B(2,16) = k(15);
B(2,3) = k(17); %Reaction 8
B(3,3) = -k(17);
B(6,17) = k(18); %Reaction 9f
B(4,17) = -k(18);
B(6,18) = -k(20); %Reaction 9r
B(4,18) = k(20);
B(4,19) = k(22); %Reaction 10f
B(6,19) = -k(22);
B(4,20) = -k(24); %Reaction 10r
B(6,20) = k(24);
B(4,6) = k(26); %Reaction 11
B(6,6) = -k(26);
end

%% Solving the tensor equation
MTmp = sparse(G(:,1),G(:,3),G(:,4).*(G(:,2)));
[a,b] = size(MTmp);
M(1:a,1:b) = MTmp;

LTmp = sparse(A(:,1),A(:,2),A(:,4).*(A(:,3)));
[a,b] = size(LTmp);
L(1:a,1:b) = LTmp;

dx_dt = (eye(length(x))+M)\(L*x+B*x+f(t));
dx_dt(x<0) = 0;

```

## c.1.2 Mass Action Model

```

function dx_dt = validateMA(t,x,k,f)

%% Definition of states
%x(1) = I
%x(2) = A
%x(3) = pA
%x(4) = B

```

```

%x(5) = p1B
%x(6) = p2B
%x(7) = C
%x(8) = D
%x(9) = CD
%x(10) = pCD
%x(11) = A-B
%x(12) = A-p1B
%x(13) = I-CD
%x(14) = I-pCD
%x(15) = pCD-A
%x(16) = pCD-pA
%x(17) = pCD-B
%x(18) = pCD-pB
%x(19) = p1B-B
%x(20) = p1B-p2B

%% Reaction velocities
rxn1_1f = k(1)*x(2)*x(4); %A + B -> A-B
rxn1_2f = k(2)*x(11); %A-B -> A + B
rxn1_3f = k(3)*x(11); %A-B -> A-p1B
rxn1_1r = k(4)*x(2)*x(5); %A + p1B -> A-p1B
rxn1_2r = k(5)*x(12); %A-p1B -> p1B + A
rxn1_3r = k(6)*x(12); %A-p1B -> B + A

rxn2 = k(7)*x(5); %p1B -> B
rxn3 = k(8)*x(7)*x(8); %C+D -> CD
rxn4 = k(9)*x(9); %CD -> C + D

rxn5_1f = k(10)*x(1)*x(9); %CD + I -> I-CD
rxn5_2f = k(11)*x(13); %I-CD -> CD + I
rxn5_3f = k(12)*x(13); %I-CD -> I-pCD
rxn5_1r = k(13)*x(1)*x(10); %pCD + I -> I-pCD
rxn5_2r = k(14)*x(14); %I-pCD -> pCD + I
rxn5_3r = k(15)*x(14); %I-pCD -> I-CD

rxn6 = k(16)*x(10); %pCD -> CD

rxn7_1f = k(17)*x(2)*x(10); %pCD + A -> pCD-A
rxn7_2f = k(18)*x(15); %pCD-A -> A + pCD
rxn7_3f = k(19)*x(15); %pCD-A -> pCD-pA
rxn7_1r = k(20)*x(3)*x(10); %pCD + pA -> pCD-pA
rxn7_2r = k(21)*x(16); %pCD-pA -> pCD + pA
rxn7_3r = k(22)*x(16); %pCD-pA -> pCD-A

rxn8 = k(23)*x(3); %pA -> A

rxn9_1f = k(24)*x(4)*x(10); %pCD + B -> pCD-B
rxn9_2f = k(25)*x(17); %pCD-B -> B + pCD
rxn9_3f = k(26)*x(17); %pCD-B -> pCD-pB
rxn9_1r = k(27)*x(6)*x(10); %pCD + pB -> pCD-pB
rxn9_2r = k(28)*x(18); %pCD-pB -> pCD + pB
rxn9_3r = k(29)*x(18); %pCD-pB -> pCD-B

rxn10_1f = k(30)*x(5)*x(6); %p1B + p2B -> p1B-p2B
rxn10_2f = k(31)*x(19); %p1B-p2B -> p1B + p2B
rxn10_3f = k(32)*x(19); %p1B-p2B -> p1B-B
rxn10_1r = k(33)*x(5)*x(4); %p1B + B -> p1B-B
rxn10_2r = k(34)*x(20); %p1B-B -> p1B + B
rxn10_3r = k(35)*x(20); %p1B-B -> p1B-p2B

```

```

rxn11 = k(36)*x(6); %p2B -> pB

% Constructing ODEs
dx_dt(1) = -rxn5_1f +rxn5_2f -rxn5_1r +rxn5_2r;
dx_dt(2) = -rxn1_1f +rxn1_2f -rxn1_1r +rxn1_2r -rxn7_1f +rxn7_2f +rxn8;
dx_dt(3) = -rxn7_1r +rxn7_2r -rxn8;

dx_dt(4) = -rxn1_1f +rxn1_2f +rxn2 -rxn9_1f +rxn9_2f -rxn10_1r +rxn10_2r +rxn11;
dx_dt(5) = -rxn1_1r +rxn1_2r -rxn2 -rxn10_1f +rxn10_2f -rxn10_1r +rxn10_2r;
dx_dt(6) = -rxn9_1r +rxn9_2r -rxn11 -rxn10_1f +rxn10_2f;
dx_dt(7) = -rxn3 +rxn4;
dx_dt(8) = -rxn3 +rxn4;
dx_dt(9) = rxn3 -rxn4 -rxn5_1f +rxn5_2f +rxn6;
dx_dt(10)= -rxn5_1r +rxn5_2r -rxn6 -rxn9_1f +rxn9_2f -rxn9_1r +...
            rxn9_2r -rxn7_1f + rxn7_2f -rxn7_1r +rxn7_2r;

dx_dt(11)= rxn1_1f -rxn1_2f -rxn1_3f +rxn1_3r;
dx_dt(12)= rxn1_1r -rxn1_2r -rxn1_3r +rxn1_3f;
dx_dt(13)= rxn5_1f -rxn5_2f -rxn5_3f +rxn5_3r;
dx_dt(14)= rxn5_1r -rxn5_2r -rxn5_3r +rxn5_3f;
dx_dt(15)= rxn7_1f -rxn7_2f -rxn7_3f +rxn7_3r;
dx_dt(16)= rxn7_1r -rxn7_2r -rxn7_3r +rxn7_3f;
dx_dt(17)= rxn9_1f -rxn9_2f -rxn9_3f +rxn9_3r;
dx_dt(18)= rxn9_1r -rxn9_2r -rxn9_3r +rxn9_3f;
dx_dt(19)= rxn10_1f -rxn10_2f -rxn10_3f +rxn10_3r;
dx_dt(20)= rxn10_1r -rxn10_2r -rxn10_3r +rxn10_3f;

dx_dt = dx_dt'+f(t);

```

### c.1.3 Michaelis Menten Model

```

function dx_dt = validateMM(t,x,k,f)

%% Definition of states
%x(1) = I
%x(2) = A
%x(3) = pA
%x(4) = B
%x(5) = p1B
%x(6) = p2B
%x(7) = C
%x(8) = D
%x(9) = CD
%x(10) = pCD

%% Reaction velocities
rxn1f = k(1)*x(2)*x(4)/(k(2)+x(4)); % B -> p1B | A
rxn1r = k(3)*x(2)*x(5)/(k(4)+x(5)); % p1B -> B | A
rxn2 = k(5)*x(5); % p1B -> B
rxn3 = k(6)*x(7)*x(8); % C+D -> CD
rxn4 = k(7)*x(9); % CD -> C + D
rxn5f = k(8)*x(1)*x(9)/(k(9)+x(9)); % CD -> pCD | I
rxn5r = k(10)*x(1)*x(10)/(k(11)+x(10)); % pCD -> CD | I
rxn6 = k(12)*x(10); % pCD -> CD
rxn7f = k(13)*x(10)*x(2)/(k(14)+x(2)); % A -> pA | pCD
rxn7r = k(15)*x(10)*x(3)/(k(16)+x(3)); % pA -> A | pCD

```

```

rxn8  = k(17)*x(3); % pA -> A
rxn9f = k(18)*x(10)*x(4)/(k(19)+x(4)); % B -> p2B | pCD
rxn9r = k(20)*x(10)*x(6)/(k(21)+x(6)); % p2B -> B | pCD
rxn10f = k(22)*x(5)*x(6)/(k(23)+x(6)); % p2B -> B | p1B
rxn10r = k(24)*x(5)*x(4)/(k(25)+x(4)); % B -> p2B | p1B
rxn11 = k(26)*x(6); % p2B -> B

%% Constructing ODEs
dx_dt(1) = 0;
dx_dt(2) = -rxn7f + rxn7r + rxn8;
dx_dt(3) = rxn7f - rxn7r - rxn8;
dx_dt(4) = -rxn1f + rxn1r + rxn2 - rxn9f + rxn9r + rxn10f - rxn10r + rxn11;
dx_dt(5) = rxn1f - rxn1r - rxn2;
dx_dt(6) = rxn9f - rxn9r - rxn10f + rxn10r - rxn11;
dx_dt(7) = -rxn3 + rxn4;
dx_dt(8) = -rxn3 + rxn4;
dx_dt(9) = rxn3 - rxn4 - rxn5f + rxn5r + rxn6;
dx_dt(10) = rxn5f - rxn5r - rxn6;

fTrunc = f(t);
dx_dt = dx_dt' + fTrunc(1:10);

```

## c.2 IN VITRO VALIDATION ONE SUBSTRATE MODEL

### c.2.1 dQSSA Model

```

function dx_dt = LDHQSSA(t,x,k,f)

%% Species Definition
%X(1) = NADH
%X(2) = NAD+
%X(3) = Py
%X(4) = Lac
%X(5) = LDH
%X(6) = NADH-LDH
%X(7) = NAD+-LDH
%X(8) = NADH-LDH-Py
%X(9) = NAD+-LDH-Lac

B = zeros(length(x));
M = 0*B;
L = M;

G = [%%%Pyruvate Side Reactions%%%%%%%%
      8 3 6 -1/k(5); %NADH-LDH-Py
        8 6 3 -1/k(5);
        3 3 6 1/k(5);
        3 6 3 1/k(5);
      6 3 6 1/k(5);
        6 6 3 1/k(5);
      %%%Lactate Side Reactions%%%%%%%%
        9 4 7 -1/k(6); %NAD-LDH-Lac
        9 7 4 -1/k(6);
      4 4 7 1/k(6);
        4 7 4 1/k(6);
        7 4 7 1/k(6);

```

```

    7 7 4 1/k(6);
];

A = [1 5 1 -k(1);
     5 5 1 -k(1);
     6 5 1  k(1);
     2 5 2 -k(3);
     5 5 2 -k(3);
     7 5 2  k(3);
     ];

B(6,6) = -k(2);
B(1,6) =  k(2);
B(5,6) =  k(2);
B(7,7) = -k(4);
B(2,7) =  k(4);
B(5,7) =  k(4);

B(2,8) =  k(7);      %Py-NADH-LDH -> Lac + NAD+ + LDH
B(4,8) =  k(7);
B(1,8) = -k(7);
B(3,8) = -k(7);
B(1,9) =  k(8);      %Lac-NAD+-LDH -> Py + NADH + LDH
B(3,9) =  k(8);
B(2,9) = -k(8);
B(4,9) = -k(8);

% Solving the linear equation
MTmp = sparse(G(:,1),G(:,3),G(:,4).*(G(:,2))));
[a,b] = size(MTmp);
M(1:a,1:b) = MTmp;

LTmp = sparse(A(:,1),A(:,2),A(:,4).*(A(:,3))));
[a,b] = size(LTmp);
L(1:a,1:b) = LTmp;

dx_dt = (eye(length(x))+M)\(L*x+B*x+f(t));

```

## C.2.2 Michaelis Menten Model

```

function dx_dt = LDHMM(t,x,k)

% x(1) = NADH
% x(2) = NAD+
% x(3) = Py
% x(4) = Lac
% x(5) = LDH
% x(6) = NADH-LDH
% x(7) = NAD+-LDH

v1 = k(1)*x(1)*x(5)-k(2)*x(6); %NADH + LDH <-> NADH-LDH
v2 = k(3)*x(2)*x(5)-k(4)*x(7); %NAD + LDH <-> NAD-LDH
v3 = k(7)*x(3)*x(6)/(k(5)+x(3)); %py -> Lac | LDH-NADH
v4 = k(8)*x(4)*x(7)/(k(6)+x(4)); %Lac -> Py | LDH-NAD

dx_dt(1) = -v1-v3+v4;
dx_dt(2) = -v2+v3-v4;
dx_dt(3) = -v3+v4;

```

```

dx_dt(4) = v3-v4;
dx_dt(5) = -v1-v2;
dx_dt(6) = v1;
dx_dt(7) = v2;

dx_dt = dx_dt';

```

### c.3 IN VITRO VALIDATION TWO SUBSTRATE MODEL

#### c.3.1 dQSSA Model (One Substrate and Two Substrate)

```

function dx_dt = dQSSA(t,x,k,f,var,kMA)
%      1 2 3 4 5 6 7 8 9
%x = [ [LDH],[NADH],[Py],[NADH-LDH],[NADH-LDH-Py],[Lac],[NAD+],[NAD+-LDH],[NAD+-LDH-Lac] ]

A = zeros(0,4);
G = zeros(0,4);
B = zeros(length(x));
M = 0*B;
N = 0*B;

if var == 1
    %% Two substrate dQSSA reversible reaction
    B = zeros(length(x));
    B(6,5) = k(3);
    B(7,5) = k(3);
    B(2,5) = -k(3);
    B(3,5) = -k(3);

    B(2,9) = k(6);
    B(3,9) = k(6);
    B(6,9) = -k(6);
    B(7,9) = -k(6);

    M = 0*B;
    N = 0*B;

    G = [%QSS of E+0
        4 1 2 -k(1)
        4 2 1 -k(1)
        1 1 2 k(1)
        1 2 1 k(1)
        2 1 2 k(1)
        2 2 1 k(1)
        %QSSA of E0+S
        5 4 3 -k(2)
        5 3 4 -k(2)
        1 4 3 k(2)
        1 3 4 k(2)
        2 4 3 k(2)
        2 3 4 k(2)
        3 4 3 k(2)
        3 3 4 k(2)
        %QSS of E+0p
        8 1 7 -k(4)
        8 7 1 -k(4)
        1 1 7 k(4)

```



```

1 7 1 k(4)
7 1 7 k(4)
7 7 1 k(4)
%QSSA of E0p+P
9 6 8 -k(5)
9 8 6 -k(5)
1 6 8 k(5)
1 8 6 k(5)
6 6 8 k(5)
6 8 6 k(5)
7 6 8 k(5)
7 8 6 k(5)
];

elseif var == 2
    % One substrate dQSSA reversible reaction (first substrate binding is
    % mass action
    G = [
        %%%Pyruvate Side Reactions%%%%%%%%
5   3   4 -k(1); %NADH-LDH-Py
        5   4   3 -k(1);
        3   3   4 k(1);
        3   4   3 k(1);
        4   3   4 k(1);
        4   4   3 k(1);
        %%%Lactate Side Reactions%%%%%%%%
        9   6   8 -k(3); %NAD-LDH-Lac
        9   8   6 -k(3);
6   6   8 k(3);
        6   8   6 k(3);
        8   6   8 k(3);
        8   8   6 k(3);
    ];

    A = [1 2 1 -k(5); %-> NADH + LDH
        2 2 1 -k(5);
        4 2 1 k(5);
        1 1 7 -kMA(6); %-> NAD+ + LDH
        7 1 7 -kMA(6);
        8 1 7 kMA(6);

        3 4 3 -kMA(3); %-> NADH-LDH + Py
        4 4 3 -kMA(3);
        5 4 3 kMA(3);

        6 6 8 -kMA(8); %-> NAD+-LDH + Lac
        8 6 8 -kMA(8);
        9 6 8 kMA(8);
    ];

    B(5,5) = B(5,5) - kMA(4); %NADH-LDH + Py ->
    B(3,5) = B(3,5) + kMA(4);
    B(4,5) = B(4,5) + kMA(4);

    B(9,9) = B(9,9) - kMA(9); %NAD+-LDH + Lac ->
    B(8,9) = B(8,9) + kMA(9);
    B(6,9) = B(6,9) + kMA(9);

    B(4,4) = B(4,4) - kMA(2);
    B(1,4) = B(1,4) + kMA(2);
    B(2,4) = B(2,4) + kMA(2);

```

```

B(8,8) = B(8,8) - kMA(7);
B(7,8) = B(7,8) + kMA(7);
B(1,8) = B(1,8) + kMA(7);

B(6,5) = B(6,5) + k(2);      %Py-NADH-LDH -> Lac + NAD+ + LDH
B(7,5) = B(7,5) + k(2);
B(2,5) = B(2,5) - k(2);
B(3,5) = B(3,5) - k(2);
B(2,9) = B(2,9) + kMA(10);   %Lac-NAD+-LDH -> Py + NADH + LDH
B(3,9) = B(3,9) + kMA(10);
B(6,9) = B(6,9) - kMA(10);
B(7,9) = B(7,9) - kMA(10);
end

MTmp = sparse(G(:,1),G(:,2),G(:,4).*(G(:,3))));
[a,b] = size(MTmp);
M(1:a,1:b) = MTmp;

NTmp = sparse(A(:,1),A(:,2),A(:,4).*(A(:,3))));
[a,b] = size(NTmp);
N(1:a,1:b) = NTmp;

dx_dt = (eye(length(x))+M)\(N*x+B*x+f(t));

```

### c.3.2 Mass Action Model

```

function varargout = MA(t,x,k)

% Ordered bi-bi Binding of 3 species into a complex
%1 2 3 4 5 6 7 8 9
%x = [ [LDH], [NADH], [Py], [NADH-LDH], [NADH-LDH-Py], [Lac], [NAD+], [NAD+-LDH], [NAD+-LDH-Lac] ]

% Reversible Enz Kinetic
v(1) = -k( 1)*x(1)*x(2) + k( 2)*x(4); % [LDH] + [NADH] <-> [NADH-LDH]
v(2) = -k( 3)*x(3)*x(4) + k( 4)*x(5); % [NADH-LDH] + [Py] <-> [NADH-LDH-Py]
v(3) = -k( 5)*x(5) + k(10)*x(9); % [NADH-LDH-Py] -> [NAD+-LDH-Lac]
v(4) = -k( 8)*x(8)*x(6) + k( 9)*x(9); % [NAD+-LDH-Lac] <-> [NAD+-LDH] + [Lac]
v(5) = -k( 6)*x(1)*x(7) + k( 7)*x(8); % [NAD+-LDH] <-> [LDH] + [NAD+]

dx_dt(1) = v(1) + v(5);
dx_dt(2) = v(1);
dx_dt(3) = v(2);
dx_dt(4) = - v(1) + v(2);
dx_dt(5) = - v(2) + v(3);
dx_dt(6) = v(4);
dx_dt(7) = v(5);
dx_dt(8) = - v(5) + v(4);
dx_dt(9) = - v(4) - v(3);

varargout{1} = dx_dt';

```

# BIOLOGICAL MODEL DEFINITIONS AND EXTRA DATA

---

## d.1 INSULINS SIGNALLING MODEL

```

%% Define postProc states
modSpc = { 'IR', 'PM', NaN ; %1
'pIR', 'PM', 0 ; %2
'piIR', 'PM', 0 ; %3
'AKT', 'PM', NaN ; %4
'pAKT', 'PM', 0 ; %5
'AKTSub', 'Cytosol', NaN ; %6
'pAKTSub', 'Cytosol', 0 ; %7
'TSC2', 'Cytosol', NaN ; %8
'piTSC2', 'Cytosol', 0 ; %9
'mTORC1', 'Cytosol', NaN ; %10
'pimTORC1', 'Cytosol', 0 ; %11
'Insulin', 'Outside', 0 ; %12
};

modComp = {'Cytosol', 1;
'PM', NaN;
'Outside', 1000};

% Features of default parameters
Bnd.k0 = [1e01 1e04];
Bnd.k1 = [5e-5 5e-1];
Bnd.k2 = [5e-5 5e-1];
Bnd.Km = [1e-2 1e02];
Bnd.Conc = [1e-1 1e1];
Bnd.Comp = [0 1];
Bnd.n = [1 4];
Bnd.r = [0 1];

% IR
rxn(end+1).desc = 'IR -> pIR | Insulin';
rxn(end).sub = {'IR'};
rxn(end).prod= 'pIR';
rxn(end).enz = 'Insulin';
rxn(end).k = [NaN 1e-2 1e1];
rxn(end).Km = [NaN 1-4 1e-1];
rxn(end+1).desc = 'pIR -> IR';
rxn(end).sub='pIR';
rxn(end).prod='IR';
rxn(end).k = [NaN 1e-4 1e0];
rxn(end+1).desc = 'IR -> piIR | mTORC1 (2)';
rxn(end).sub = 'IR';
rxn(end).prod= 'piIR';
rxn(end).enz = 'mTORC1';

```

```

    rxn(end).k = [NaN 1e-2 1e2];
    rxn(end).Km = NaN;
    rxn(end+1).desc = 'piIR -> IR';
    rxn(end).sub = 'piIR';
    rxn(end).prod= 'IR';
    rxn(end).k = [NaN 1e-4 1e0];

%AKT
    rxn(end+1).desc = 'AKT -> pAKT(1)';
    rxn(end).sub = 'AKT';
    rxn(end).prod= 'pAKT';
    rxn(end).enz = 'pIR';
    rxn(end).k = [NaN 1e-2 1e2];
    rxn(end).Km = NaN;
    rxn(end+1).desc = 'pAKT -> AKT';
    rxn(end).sub = 'pAKT';
    rxn(end).prod= 'AKT';
    rxn(end).k = [NaN 1e-4 1e0];

%Akt Substrates
    rxn(end+1).desc = 'AKTSub->pAKTSub | pAKT (4)';
    rxn(end).sub = 'AKTSub';
    rxn(end).prod= 'pAKTSub';
    rxn(end).enz = 'pAKT';
    rxn(end).k = [NaN 5e0 1e4];
    rxn(end).Km = [NaN 5e0 1e4];
    rxn(end+1).desc = 'pAS160 -> AS160';
    rxn(end).sub = 'pAKTSub';
    rxn(end).prod= 'AKTSub';
    rxn(end).k = [NaN 1e-4 1e0];

%TSC2
    rxn(end+1).desc = 'TSC2->piTSC2 | pAKT (7)';
    rxn(end).sub = 'TSC2';
    rxn(end).prod= 'piTSC2';
    rxn(end).enz = 'pAKT';
    rxn(end).k = [NaN 1e-2 1e2];
    rxn(end).Km = [NaN 1e-4 1e0];
    rxn(end+1).desc = 'piTSC2 -> TSC2';
    rxn(end).sub = 'piTSC2';
    rxn(end).prod= 'TSC2';
    rxn(end).k = [NaN 1e-4 1e0];

%mTORC1
    rxn(end+1).desc = 'mTORC1->pimTORC1 | TSC2 (8)';
    rxn(end).sub = 'mTORC1';
    rxn(end).prod = 'pimTORC1';
    rxn(end).enz = 'TSC2';
    rxn(end).k = [NaN 1e-2 1e2];
    rxn(end).Km = [NaN 1e-4 1e0];
    rxn(end+1).desc = 'pimTORC1 -> mTORC1';
    rxn(end).sub = 'pimTORC1';
    rxn(end).prod= 'mTORC1';
    rxn(end).k = [NaN 1e-4 1e0];

```

## d.2 AKT MODEL

```
%% Compartment definition
```

```

modComp = {'Cyto', 1;
           'PM' , [NaN 1e-3 1];
           'ES' , Inf;
           };

%% Model species definition
modSpc = {'PIP3'      , 'PM' , 0;
          'AKT'       , 'Cyto', 1;
          'mAKT'      , 'PM' , 0;
          'p308mAKT'  , 'PM' , 0;
          'p473mAKT'  , 'PM' , 0;
          'p308p473mAKT' , 'PM' , 0;
          'p308AKT'   , 'Cyto', 0;
          'p473AKT'   , 'Cyto', 0;
          'p308p473AKT' , 'Cyto', 0;
          'mTORC2'    , 'PM' , 1;
          'PDK1'      , 'PM' , 1};

%% Features of default parameters
Bnd.k0 = [1e01 1e04];
Bnd.k1 = [5e-5 0.5];
Bnd.k2 = [5e-5 5e-1];
Bnd.Km = [1e-2 1e02];
Bnd.Conc = [1e-1 1e1];
Bnd.n = [1 4];
Bnd.r = [0 1];
Bnd.Comp = [0 1];

%% Reactions

%% AKT Translocation Mechanics
rxn(end+1).desc = 'PIP3 + AKT -> mAKT';
rxn(end).sub = {'PIP3', 'AKT'};
rxn(end).prod= 'mAKT';
rxn(end).k = [NaN 1 1 1e3];
rxn(end+1).desc = 'mAKT -> PIP3 + AKT';
rxn(end).sub = 'mAKT';
rxn(end).prod= {'PIP3', 'AKT'};
rxn(end).k = [NaN 2 0.3 30];

rxn(end+1).desc = 'PIP3 + p308AKT -> p308mAKT';
rxn(end).sub = {'PIP3', 'p308AKT'};
rxn(end).prod= 'p308mAKT';
rxn(end).k = [NaN 1];
rxn(end+1).desc = 'p308mAKT -> PIP3 + p308AKT';
rxn(end).sub = 'p308mAKT';
rxn(end).prod= {'PIP3', 'p308AKT'};
rxn(end).k = [NaN 2];

rxn(end+1).desc = 'PIP3 + p473AKT -> p473mAKT';
rxn(end).sub = {'PIP3', 'p473AKT'};
rxn(end).prod= 'p473mAKT';
rxn(end).k = [NaN 1];
rxn(end+1).desc = 'p473mAKT -> PIP3 + p473AKT';
rxn(end).sub = 'p473mAKT';
rxn(end).prod= {'PIP3', 'p473AKT'};
rxn(end).k = [NaN 2];

rxn(end+1).desc = 'PIP3 + p308p473AKT -> p308p473mAKT';
rxn(end).sub = {'PIP3', 'p308p473AKT'};

```

```

rxn(end).prod= 'p308p473mAKT';
rxn(end).k    = [NaN 1];
rxn(end+1).desc = 'p308p473mAKT -> PIP3 + p308p473AKT';
rxn(end).sub = 'p308p473mAKT';
rxn(end).prod= {'PIP3','p308p473AKT'};
rxn(end).k    = [NaN 2];

%% 308 phosphorylation and dephosphorylation mechanics
rxn(end+1).desc = 'mAKT -> p308mAKT';
rxn(end).sub = 'mAKT';
rxn(end).prod= 'p308mAKT';
rxn(end).enz = 'PDK1';
rxn(end).k    = NaN;
rxn(end).Km   = NaN;
rxn(end+1).desc = 'p473mAKT -> p308p473mAKT';
rxn(end).sub = 'p473mAKT';
rxn(end).prod= 'p308p473mAKT';
rxn(end).enz = 'PDK1';
rxn(end).k    = NaN;
rxn(end).Km   = [NaN 1 1e3];
rxn(end+1).desc = 'p308AKT -> AKT';
rxn(end).sub = 'p308AKT';
rxn(end).prod= 'AKT';
rxn(end).k    = [NaN 3 0.1 10];
rxn(end+1).desc = 'p308p473AKT -> p473AKT';
rxn(end).sub = 'p308p473AKT';
rxn(end).prod= 'p473AKT';
rxn(end).k    = [NaN 3];
rxn(end+1).desc = 'p308mAKT -> mAKT';
rxn(end).sub = 'p308mAKT';
rxn(end).prod= 'mAKT';
rxn(end).k    = [NaN 3];
rxn(end+1).desc = 'p308p473mAKT -> p473mAKT';
rxn(end).sub = 'p308p473mAKT';
rxn(end).prod= 'p473mAKT';
rxn(end).k    = [NaN 3];

%% 473 Phosphorylation and dephosphorylation mechanics
rxn(end+1).desc = 'mAKT -> p473mAKT';
rxn(end).sub = 'mAKT';
rxn(end).prod= 'p473mAKT';
rxn(end).enz = 'mTORC2';
rxn(end).k    = NaN;
rxn(end).Km   = [NaN 0.3 1e-3];
rxn(end+1).desc = 'p308mAKT -> p308p473mAKT';
rxn(end).sub = 'p308mAKT';
rxn(end).prod= 'p308p473mAKT';
rxn(end).enz = 'mTORC2';
rxn(end).k    = NaN;
rxn(end).Km   = NaN;
rxn(end+1).desc = 'p473AKT -> AKT';
rxn(end).sub = 'p473AKT';
rxn(end).prod= 'AKT';
rxn(end).k    = [NaN 3];
rxn(end+1).desc = 'p308p473AKT -> p308AKT';
rxn(end).sub = 'p308p473AKT';
rxn(end).prod= 'p308AKT';
rxn(end).k    = [NaN 3];
rxn(end+1).desc = 'p473mAKT -> mAKT';
rxn(end).sub = 'p473mAKT';

```

```

rxn(end).prod= 'mAKT';
rxn(end).k    = [NaN 3];
rxn(end+1).desc = 'p308p473mAKT -> p308mAKT';
rxn(end).sub  = 'p308p473mAKT';
rxn(end).prod= 'p308mAKT';
rxn(end).k    = [NaN 3];

```

### d.3 ADDITIONAL DATA

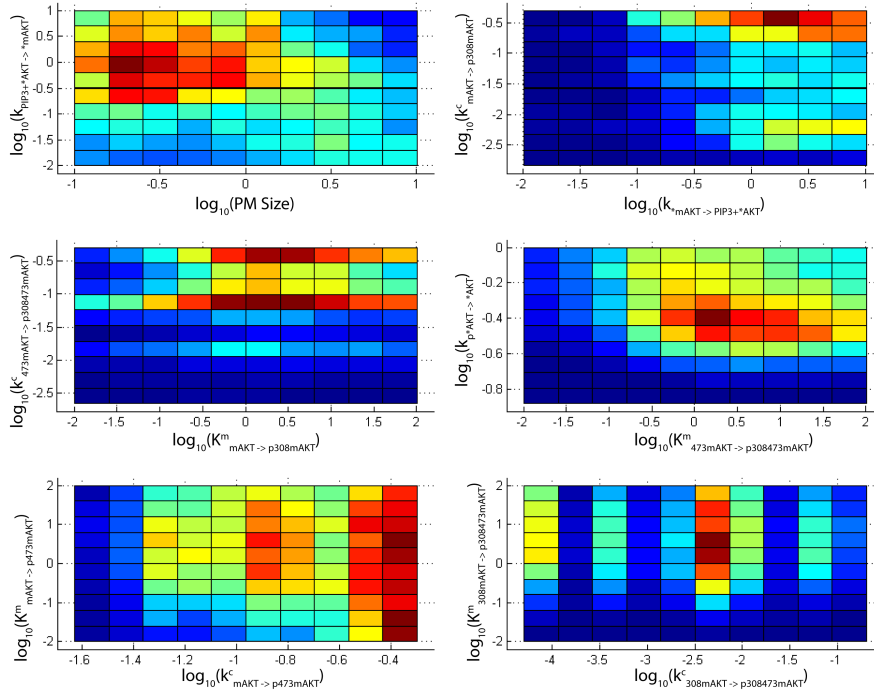


Figure 54: Visualisation of the distribution of 1000 parameter sets achieved in the Akt-Akt substrate model using 2D heat maps. Red indicates high density of accepted parameters while blue shows a low density (deep blue is zero density).





## BIBLIOGRAPHY

---

- Zimmet, Paul Z. et al. (2014). "Diabetes: A 21st century challenge." In: *The Lancet Diabetes and Endocrinology* 2.1, pp. 56–64.
- International Diabetes Federation (2013). *IDF Diabetes Atlas*. 6th.
- Danaei, Goodarz et al. (2011). "National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants." In: *Lancet* 378.9785, pp. 31–40.
- World Health Organisation (2012). *Health statistics and information | Topics | Casues of death estimates*.
- Australia Diabetes (2014). *Understanding Diabetes - Diabetes Australia*.
- Shaw, Jonathan and Stephanie Tanamas (2012). *Diabetes : the silent pandemic and its impact on Australia*. Diabetes Australia.
- Pickup, John C and Gareth Williams (2003). *Textbook of Diabetes* 1. Third. Oxford: Blackwell Publishing.
- ADS, Americal Diabetes Association (2014). "Diagnosis and clas-sification of diabetes mellitus." In: *Diabetes Care* 37.SUPPL.1, pp. 81–90.
- Taylor, S I (1999). "Deconstructing type 2 diabetes." In: *Cell* 97.1, pp. 9–12.
- Nussey, S and S Whitehead (2001). *Endocrinology: An Integrated Approached*. Oxford: BIOS Scientific Publishers.
- Samuel, Varman T. and Gerald I. Shulman (2012). "Mechanisms for insulin resistance: Common threads and missing links." In: *Cell* 148.5, pp. 852–871. arXiv: [176](#).
- Marieb, E. N. and K Hoehn (2005). *Human Anatomy and Physiology*. 7th ed. San Francisco: Pearson Benjamin Cummings.
- Skyler, Jay (2012). *Atlas of Diabetes*. New York: Springer.
- Humphrey, Sean J et al. (2013). "Dynamic adipocyte phosphopro-teome reveals that Akt directly regulates mTORC2." In: *Cell Metabolism* 17.6, pp. 1009–20.

- Yugi, Katsuyuki et al. (2014). "Reconstruction of Insulin Signal Flow from Phosphoproteome and Metabolome Data." In: *Cell Reports* 8.4, pp. 1171–1183.
- Citri, Ami and Yosef Yarden (2006). "EGF-ERBB signalling: towards the systems level." In: *Nature reviews. Molecular cell biology* 7.7, pp. 505–516.
- Kitano, Hiroaki (2004). "Biological Robustness." In: *Nature Reviews, Genetics* 64.November, pp. 239–263.
- Alberts, Bruce et al. (2002). *Molecular Biology of the Cell*. Garland Science.
- Goltsov, Alexey et al. (2011). "Compensatory effects in the PI3K/PTEN/AKT signaling network following receptor tyrosine kinase inhibition." In: *Cellular Signalling* 23.2, pp. 407–16.
- Kiely, Maeve and Patrick Kiely (2015). "PP2A: The Wolf in Sheep's Clothing?" In: *Cancers* 7.2, pp. 648–669.
- Parker, Benjamin et al. (2015). "Terminal galactosylation and sialylation switching on membrane glycoproteins upon TNF-alpha-induced insulin resistance in adipocytes." In: *Molecular & cellular proteomics* In Press.
- Risso, Guillermo et al. (2015). "Akt/PKB: one kinase, many modifications." In: *Biochemical Journal* 468.2, pp. 203–214.
- Parker, Carol E. et al. (2010). "Mass Spectrometry for Post-Translational Modifications." In: *Neuroproteomics*. Boca Raton: CRC Press. Chap. 6.
- Prior, Ian a. and John F. Hancock (2012). "Ras trafficking, localization and compartmentalized signalling." In: *Seminars in Cell and Developmental Biology* 23.2, pp. 145–153.
- Taniguchi, Cullen M, Brice Emanuelli, and C Ronald Kahn (2006). "Critical nodes in signalling pathways: insights into insulin action." In: *Nature reviews. Molecular cell biology* 7.2, pp. 85–96.
- Rowland, Alexander F, Daniel J Fazakerley, and David E James (2011). "Mapping insulin/GLUT4 circuitry." In: *Traffic (Copenhagen, Denmark)* 12.6, pp. 672–81.
- Stenkula, Karin G. et al. (2007). "Human, but not rat, IRS1 targets to the plasma membrane in both human and rat adipocytes." In: *Biochemical and Biophysical Research Communications* 363.3, pp. 840–845.
- Backer, J M et al. (1992). "Phosphatidylinositol 3'-kinase is activated by association with IRS-1 during insulin stimulation." In: *The EMBO journal* 11.9, pp. 3469–79.

- Klippel, A et al. (1996). "Membrane localization of phosphatidylinositol 3-kinase is sufficient to activate multiple signal-transducing kinase pathways." In: *Molecular and Cellular Biology* 16.8, pp. 4117–4127.
- Cantley, Lewis C (2002). "The phosphoinositide 3-kinase pathway." In: *Science (New York, N.Y.)* 296.5573, pp. 1655–7.
- Kriplani, Nisha et al. (2015). "Class I PI 3-kinases: Function and evolution." In: *Advances in Biological Regulation*, pp. 1–12.
- Rodríguez-Escudero, Isabel et al. (2005). "Reconstitution of the mammalian PI3K/PTEN/Akt pathway in yeast." In: *The Biochemical journal* 390.Pt 2, pp. 613–23.
- Erneux, Christophe et al. (2011). "SHIP2 multiple functions: a balance between a negative control of PtdIns(3,4,5)P3 level, a positive control of PtdIns(3,4)P3 production, and intrinsic docking properties." In: *Journal of cellular biochemistry* 112.9, pp. 2203–9.
- James, Stephen R et al. (1996). "Specific binding of the Akt-1 protein kinase to phosphatidylinositol 3,4,5-trisphosphate without subsequent activation." In: *Biochemical Journal* 315, pp. 709–713.
- Andjelković, M et al. (1997). "Role of translocation in the activation and function of protein kinase B." In: *The Journal of biological chemistry* 272.50, pp. 31515–31524.
- Alessi, D R et al. (1997). "Characterization of a 3-phosphoinositide-dependent protein kinase which phosphorylates and activates protein kinase Balpha." In: *Current biology : CB* 7.4, pp. 261–269.
- Sarbassov, D D et al. (2005). "Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex." In: *Science* 307.5712, pp. 1098–101.
- Tan, Shi-Xiong et al. (2011). "Amplification and demultiplexing in the insulin regulated Akt pathway in adipocytes." In: *The Journal of biological chemistry* 287.9, pp. 6128–6138.
- Siddle, Kenneth (2011). "Signalling by insulin and IGF receptors: Supporting acts and new players." In: *Journal of Molecular Endocrinology* 47.1.
- Cohen, Philip (2006). "The twentieth century struggle to decipher insulin signalling." In: *Nature reviews. Molecular cell biology* 7.11, pp. 867–873.
- Manning, Brendan D and Lewis C Cantley (2007). "AKT/PKB signaling: navigating downstream." In: *Cell* 129.7, pp. 1261–74.
- Foti, Michelangelo et al. (2007). "The neck of caveolae is a distinct plasma membrane subdomain that concentrates insulin

- receptors in 3T3-L1 adipocytes." In: *Proceedings of the National Academy of Sciences of the United States of America* 104.4, pp. 1242–1247.
- Karlsson, Margareta et al. (2004). "Colocalization of insulin receptor and insulin receptor substrate-1 to caveolae in primary human adipocytes. Cholesterol depletion blocks insulin signalling for metabolic and mitogenic control." In: *European Journal of Biochemistry* 271.12, pp. 2471–2479.
- Fagerholm, Siri et al. (2009). "Rapid Insulin-Dependent Endocytosis of the Insulin Receptor by Caveolae in Primary Adipocytes." In: *PLoS ONE* 4.6, e5985.
- Wang, Hong et al. (2015). "Caveolin-1 phosphorylation regulates vascular endothelial insulin uptake and is impaired by insulin resistance in rats." In: *Diabetologia* 58.6, pp. 1344–53.
- Ceresa, B P et al. (1998). "Inhibition of clathrin-mediated endocytosis selectively attenuates specific insulin receptor signal transduction pathways." In: *Molecular and cellular biology* 18.7, pp. 3862–3870.
- Tseng, Linda Tzu Ling et al. (2013). "LMBD1 protein serves as a specific adaptor for insulin receptor internalization." In: *Journal of Biological Chemistry* 288.45, pp. 32424–32432.
- Haruta, T et al. (2000). "A rapamycin-sensitive pathway down-regulates insulin signaling via phosphorylation and proteasomal degradation of insulin receptor substrate-1." In: *Molecular endocrinology (Baltimore, Md.)* 14.6, pp. 783–94.
- Harrington, Laura S et al. (2004b). "The TSC1-2 tumor suppressor controls insulin-PI3K signaling via regulation of IRS proteins - Supplementary." In: *The Journal of cell biology* 166.2, pp. 213–23.
- Destefano, Michael a and Estela Jacinto (2013). "Regulation of insulin receptor substrate-1 by mTORC2 (mammalian target of rapamycin complex 2)." In: *Biochemical Society transactions* 41.4, pp. 896–901.
- Tremblay, F and a Marette (2001). "Amino acid and insulin signaling via the mTOR/p70 S6 kinase pathway. A negative feedback mechanism leading to insulin resistance in skeletal muscle cells." In: *The Journal of biological chemistry* 276.41, pp. 38052–60.
- Nyman, Elin et al. (2012). "Mechanistic explanations for counter-intuitive phosphorylation dynamics of the insulin receptor and insulin receptor substrate-1 in response to insulin in murine adipocytes." In: *FEBS journal* 279, pp. 987–999.

- Giraud, Jodel et al. (2007). "Phosphorylation of Irs1 at SER-522 inhibits insulin signaling." In: *Molecular endocrinology* (Baltimore, Md.) 21.9, pp. 2294–2302.
- Williams, M R et al. (2000). "The role of 3-phosphoinositide-dependent protein kinase 1 in activating AGC kinases defined in embryonic stem cells." In: *Current biology : CB* 10.8, pp. 439–48.
- Hresko, Richard C and Mike Mueckler (2005). "mTOR.RICTOR is the Ser473 kinase for Akt/protein kinase B in 3T3-L1 adipocytes." In: *The Journal of biological chemistry* 280.49, pp. 40406–16.
- Shiota, Chiyo et al. (2006). "Multiallelic Disruption of the rictor Gene in Mice Reveals that mTOR Complex 2 Is Essential for Fetal Growth and Viability." In: *Developmental Cell* 11.4, pp. 583–589.
- Jacinto, Estela et al. (2006). "SIN1/MIP1 maintains rictor-mTOR complex integrity and regulates Akt phosphorylation and substrate specificity." In: *Cell* 127.1, pp. 125–37.
- Vanhaesebroeck, B and D R Alessi (2000). "The PI3K-PDK1 connection: more than just a road to PKB." In: *The Biochemical journal* 346 Pt 3, pp. 561–76.
- Bayascas, Jose R. (2008). "Dissecting the role of the 3-phosphoinositide-dependent protein kinase-1 (PDK1) signalling pathways." In: *Cell Cycle* 7.19, pp. 2978–2982.
- Ng, Yvonne et al. (2008). "Rapid activation of Akt2 is sufficient to stimulate GLUT4 translocation in 3T3-L1 adipocytes." In: *Cell metabolism* 7.4, pp. 348–56.
- Carpten, John D et al. (2007). "A transforming mutation in the pleckstrin homology domain of AKT1 in cancer." In: *Nature* 448.7152, pp. 439–444.
- Kohn, Aimee D, Fumito Takeuchi, and Richard A Roth (1996). "Akt, a Pleckstrin Homology Domain Containing Kinase, Is Activated Primarily by Phosphorylation." In: *Journal of Biological Chemistry* 271.6, pp. 21920–21926.
- Najafov, Ayaz, Natalia Shpiro, and Dario Alessi (2012). "Akt is efficiently activated by PIF-pocket- and PtdIns(3,4,5)P3-dependent mechanisms leading to resistance to PDK1 inhibitors." In: *Biochemical Journal* 295, pp. 285–295.
- Huang, Bill X. et al. (2011). "Phosphatidylserine is a critical modulator for Akt activation." In: *Journal of Cell Biology* 192.6, pp. 979–992.

- Lucas, Nathan and Wonhwa Cho (2011). "Phosphatidylserine binding is essential for plasma membrane recruitment and signaling function of 3-phosphoinositide-dependent kinase-1." In: *Journal of Biological Chemistry* 286.48, pp. 41265–41272.
- Park, H. et al. (2015). "AMIGO2, a novel membrane anchor of PDK1, controls cell survival and angiogenesis via Akt activation." In: *The Journal of Cell Biology* 211.3, pp. 619–637.
- Bracho-Valdés, Ismael et al. (2011). "mTORC1- and mTORC2-interacting proteins keep their multifunctional partners focused." In: *IUBMB Life* 63.10, pp. 880–898.
- Obata, T. et al. (2000). "Peptide and protein library screening defines optimal substrate motifs for AKT/PKB." In: *Journal of Biological Chemistry* 275.46, pp. 36108–36115.
- Ubersax, Jeffrey a and James E Ferrell (2007). "Mechanisms of specificity in protein phosphorylation." In: *Nature reviews. Molecular cell biology* 8.7, pp. 530–41.
- Schultze, Simon M et al. (2012). "PI3K/AKT, MAPK and AMPK signalling: protein kinases in glucose homeostasis." In: *Expert reviews in molecular medicine* 14.January, e1.
- Emamian, Effat S (2012). "AKT/GSK3 signaling pathway and schizophrenia." In: *Frontiers in molecular neuroscience* 5.March, p. 33.
- Copps, K D and M F White (2012). "Regulation of insulin sensitivity by serine/threonine phosphorylation of insulin receptor substrate proteins IRS1 and IRS2." In: *Diabetologia* 55.10, pp. 2565–82.
- Huang, Jingxiang and Brendan Manning (2009). "A complex interplay between AKT, TSC2, and the two mTOR complexes." In: *Biochem Soc Trans* 37, pp. 217–222.
- Kubota, Hiroyuki et al. (2012). "Temporal coding of insulin action through multiplexing of the AKT pathway." In: *Molecular Cell* 46, pp. 1–13.
- Schultze, Simon M et al. (2011). "Promiscuous affairs of PKB / AKT isoforms in metabolism." In: *Archives of Physiology and Biochemistry* 117.2, pp. 70–77.
- Yang, Z.-Z. et al. (2003). "Protein Kinase B / Akt1 Regulates Placental Development and Fetal Growth." In: *Journal of Biological Chemistry* 278.34, pp. 32124–32131.
- Tschopp, Oliver et al. (2005). "Essential role of protein kinase Bgamma (PKBgamma/Akt3) in postnatal brain development

- but not in glucose homeostasis." In: *Development* 132.13, pp. 2943–2954.
- Osorio-fuentealba, Cesar and Amira Klip (2015). "Dissecting signalling by individual Akt / PKB isoforms , three steps at once." In: *Biochemical Journal* 470, e13–16.
- Dummler, B and B a Hemmings (2007). "Physiological roles of PKB/ Akt isoforms in development and disease." In: *Biochemical Society transactions* 35.Renin 1, pp. 231–235.
- Tan, Shixiong, Yvonne Ng, and David E James (2011). "Next-generation Akt inhibitors provide greater specificity: effects on glucose metabolism in adipocytes." In: *The Biochemical journal* 435.2, pp. 539–44.
- Kajno, Esi, Timothy E. McGraw, and Eva Gonzalez (2015). "Development of a new model system to dissect isoform specific Akt signaling in adipocytes." In: *Biochemical Journal*, pp. 425–434.
- Lazebnik, Iu (2002). "Can a biologist fix a radio, or what I learned while studying apoptosis." In: *Cancer Cell* 12.September, pp. 166–71.
- Saez-Rodriguez, Julio et al. (2007). "A Logical Model Provides Insights into T Cell Receptor Signaling." In: *PLoS Computational Biology* 3.8, e163.
- Jong, Hidde de (2002). "Modeling and simulation of genetic regulatory systems: a literature review." In: *Journal of computational biology* 9.1, pp. 67–103.
- Bornholdt, Stefan (2008). "Boolean network models of cellular regulation: prospects and limitations." In: *Journal of the Royal Society, Interface / the Royal Society* 5 Suppl 1, S85–94.
- Aldridge, Bree B. et al. (2009). "Fuzzy Logic Analysis of Kinase Pathway Crosstalk in TNF/EGF/Insulin-Induced Signaling." In: *PLoS Computational Biology* 5.4.
- Naldi, a. et al. (2009). "Logical modelling of regulatory networks with GINsim 2.3." In: *BioSystems* 97, pp. 134–139.
- Albert, Réka and Rui-Sheng Wang (2009). "Discrete dynamic modeling of cellular signaling networks." In: *Methods in enzymology* 467.09, pp. 281–306.
- Steffen, Klamt, Saez-Rodriguez J., and Gilles Ernst D. (2007). "Structural and functional analysis of cellular networks with CellNetAnalyzer." In: *Journal of Biotechnology* 129, pp. 329–351.
- Morris, Melody K., Ioannis Melas, and Julio Saez-Rodriguez (2012). "Construction of Cell Type-Specific Logic Models of Signaling

- Networks Using CellNOpt." In: *Methods in Molecular Biology* 930, pp. 179–214.
- Bock, Matthias et al. (2014). "BooleSim: An interactive Boolean network simulator." In: *Bioinformatics* 30.1, pp. 131–132.
- Chaouiya, Claudine et al. (2013). "SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools." In: *BMC systems biology* 7, p. 135. arXiv: [1309.1910](#).
- Cecil, J a, K Srihari, and C R Emerson (1992). "A Review of Petri Net Applications in Manufacturing." In: *International Journal of Advanced Manufacturing Technology* 177, pp. 168–177.
- Salimifard, Khodakaram and Mike Wright (2001). "Petri net-based modelling of workflow systems: An overview." In: *European Journal of Operational Research* 134, pp. 664–676.
- Hardy, Simon and Pierre N Robillard (2004). "Modeling And Simulation Of Molecular Biology Systems Using Petri Nets: Modeling Goals Of Various Approaches." In: *Journal of Bioinformatics and Computational Biology* 2.4, pp. 619–637.
- Reddy, V N, M N Liebman, and M L Mavrovouniotis (1996). "Qualitative analysis of biochemical reaction systems." In: *Computers in biology and medicine* 26.1, pp. 9–24.
- Lee, Dong-Yup et al. (2006). "Colored Petri net modeling and simulation of signal transduction pathways." In: *Metabolic engineering* 8.2, pp. 112–22.
- Chaouiya, Claudine (2007). "Petri net modelling of biological networks." In: *Briefings in bioinformatics* 8.4, pp. 210–9.
- Chellaboina, Vijaysekhar (2009). "Modeling and Analysis of Mass-Action Kinetics." In: *IEEE Control Systems Magazine* 29 (4).August, pp. 60 –78.
- Gillespie, Daniel T. (1992). "A rigorous derivation of the chemical master equation." In: *Physica A: Statistical Mechanics and its Applications* 188.1-3, pp. 404–425.
- Cangiani, a and R Natalini (2010). "A spatial model of cellular molecular trafficking including active transport along microtubules." In: *Journal of theoretical biology* 267.4, pp. 614–25.
- Aldridge, Bree B et al. (2006). "Physicochemical modelling of cell signalling pathways." In: *Nature cell biology* 8.11, pp. 1195–203.
- Chen, William W, Mario Niepel, and Peter K Sorger (2010). "Classic and contemporary approaches to modeling biochemical reactions." In: *Genes & Development* 24.17, pp. 1861–75.



- Chen, William W et al. (2009). "Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data." In: *Molecular systems biology* 5.239, pp. 1–19.
- Gillespie, Daniel T. (2000). "The chemical Langevin equation." In: *Journal of Chemical Physics* 113.1, p. 297.
- Smadbeck, Patrick and Yiannis N. Kaznessis (2014). *Chemical Master Equation Closure for Computer-Aided Synthetic Biology*. New York: Springer.
- Anderson, David F, Desmond J Higham, and Yu Sun (2014). "Complexity of Multilevel Monte Carlo Tau-Leaping." In: *SIAM Journal on Numerical Analysis* 52.6, pp. 3106–3127.
- Goldstein, Raymond E, Jan-willem Van De Meent, and Raymond E Goldstein (2015). "A physical perspective on cytoplasmic streaming." In:
- Terry, Alan J and Mark a J Chaplain (2011). "Spatio-temporal modelling of the NF- $\kappa$ B intracellular signalling pathway: the roles of diffusion, active transport, and cell geometry." In: *Journal of theoretical biology* 290, pp. 7–26.
- Muñoz-García, Javier, Zoltan Neufeld, and Boris N. Kholodenko (2009). "Positional information generated by spatially distributed signaling cascades." In: *PLoS Computational Biology* 5.3.
- Kolch, Walter, Muffy Calder, and David Gilbert (2005). "When kinases meet mathematics: the systems biology of MAPK signalling." In: *FEBS letters* 579.8, pp. 1891–5.
- Tolle, Dominic P and Nicolas Le Novère (2006). "Particle-Based Stochastic Simulation in Systems Biology." In: *Current Bioinformatics* 1.3, pp. 1–6.
- McGuffee, Sean R and Adrian H Elcock (2010). "Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm." In: *PLoS computational biology* 6.3, e1000694.
- Wheatley, D N and P C Malone (1993). "Heat conductance, diffusion theory and intracellular metabolic regulation." In: *Biology of the cell* 79.1, pp. 1–5.
- Loverdo, C. et al. (2008). "Enhanced reaction kinetics in biological cells." In: *Nature Physics* 4.2, pp. 134–137.
- Bartel, Soren (2015). *Numerical Methods for Nonlinear Partial Differential Equations*. New York: Springer.

- Johnson, Claes (1987). *Numerical Solutions of Partial Differential Equations By The Finite Element Method*. Cambridge: Cambridge University Press.
- Kee, Robert J, Michael E Coltrin, and Peter Glarborg (2003). *Chemically Reacting Flow: Theory and Practice*. New Jersey: John Wiley & Sons.
- Schnell, S. and T. E. Turner (2004). "Reaction kinetics in intracellular environments with macromolecular crowding: Simulations and rate laws." In: *Progress in Biophysics and Molecular Biology* 85.2-3, pp. 235–260.
- Moore, John W. and Ralph G. Pearson (1961). *Kinetics and Mechanisms*. John Wiley & Sons.
- Wilhelm, Thomas (2000). "Chemical systems consisting only of elementary steps - a paradigm for nonlinear behavior." In: *Journal of Mathematical Chemistry* 27.1-2, pp. 71–88.
- Frenzen, C. L. and P. K. Maini (1988). "Enzyme kinetics for a two-step enzymic reaction with comparable initial enzyme-substrate ratios." In: *Journal of Mathematical Biology* 26.6, pp. 689–703.
- Min, Wei, X Sunney Xie, and Biman Bagchi (2008). "Two-dimensional reaction free energy surfaces of catalytic reaction: effects of protein conformational dynamics on enzyme catalysis." In: *Journal of Physical Chemistry B* 112.2, pp. 454–66.
- Brännmark, Cecilia et al. (2013). "Insulin Signalling in Type 2 Diabetes: Experimental and modeling analyses reveal mechanisms of insulin resistance in human adipocytes." In: *Journal of Biological Chemistry* 288.14, pp. 9867–9880.
- Briggs, George Edward and John Burdon Sanderson Haldane (1925). "A note on the kinetics of enzyme action." In: *Journal of Biochemistry* 19.2, pp. 338–339.
- Chen, K C et al. (2000). "Kinetic analysis of a molecular model of the budding yeast cell cycle." In: *Molecular Biology of the Cell* 11.1, pp. 369–91.
- Dalle Pezze, P. et al. (2012). "A Dynamic Network Model of mTOR Signaling Reveals TSC-Independent mTORC2 Regulation." In: *Science Signaling* 5.217, pp. 1–17.
- Fujita, Kazuhiro A et al. (2010). "Decoupling of receptor and downstream signals in the Akt pathway by its low-pass filter characteristics." In: *Science signaling* 3.132, ra56.
- Hao, Nan et al. (2012). "Combined computational and experimental analysis reveals mitogen-activated protein kinase-mediated

- feedback phosphorylation as a mechanism for signaling specificity." In: *Molecular biology of the cell* 23.19, pp. 3899–910.
- Moles, Carmen G, Pedro Mendes, and Julio R Banga (2003). "Parameter Estimation in Biochemical Pathways : A Comparison of Global Optimization Methods." In: *Genome Research* 13.11, pp. 2467–2474.
- Michaelis, Leonor et al. (2011). "The original Michaelis constant: translation of the 1913 Michaelis-Menten paper." In: *Biochemistry* 50.39, pp. 8264–9.
- Schnell, Santiago (2014). "Validity of the Michaelis-Menten equation. Steady-state, or reactant stationary assumption: that is the question." In: *FEBS Journal* 281.2, pp. 464–472.
- Bajzer, Zeljko and Emanuel E Strehler (2012). "About and beyond the Henri-Michaelis-Menten rate equation for single-substrate enzyme kinetics." In: *Biochemical and Biophysical Research Communications* 417.3, pp. 982–5.
- Borghans, Josè A M, Rob J De Boer, and Lee A Segel (1996). "Extending the Quasi-Steady State Approximation by Changing Variables." In: *Bulletin of mathematical biology* 58.1, pp. 43–63.
- Kargi, Fikret (2009). "Generalized rate equation for single-substrate enzyme catalyzed reactions." In: *Biochemical and biophysical research communications* 382.1, pp. 157–9.
- Tzafiriri, A Rami (2003). "Michaelis-Menten kinetics at high enzyme concentrations." In: *Bulletin of Mathematical Biology* 65.6, pp. 1111–29.
- Lim, Henry C. (1973). "On kinetic behavior at high enzyme concentrations." In: *AIChE Journal* 19.3, pp. 659–661.
- Tzafiriri, a. R. and E. R. Edelman (2004). "The total quasi-steady-state approximation is valid for reversible enzyme kinetics." In: *Journal of Theoretical Biology* 226.3, pp. 303–313.
- Pedersen, Morten Gram, Alberto Maria Bersani, and Enrico Bersani (2007). "Quasi steady-state approximations in complex intracellular signal transduction networks - a word of caution." In: *Journal of Mathematical Chemistry* 43.4, pp. 1318–1344.
- Nyman, Elin et al. (2011). "A hierarchical whole-body modeling approach elucidates the link between in Vitro insulin signaling and in Vivo glucose homeostasis." In: *Journal of Biological Chemistry* 286.29, pp. 26028–41.
- Nim, Tri Hieu et al. (2015). "Non-canonical Activation of Akt in Serum-Stimulated Fibroblasts, Revealed by Comparative Mod-

- eling of Pathway Dynamics." In: *PLOS Computational Biology* 11.11, e1004505.
- Xing, Jianhua and Jing Chen (2008). "The Goldbeter-Koshland switch in the first-order region and its response to dynamic disorder." In: *PloS one* 3.5, e2140.
- Press, William H et al. (1988). *Numerical Recipes in C*.
- Weise, Thomas (2009). *Global Optimisation and Algorithms, Theory and Application*.
- Olufsen, Mette S and Johnny T Ottesen (2012). "A practical approach to parameter estimation applied to model predicting heart rate regulation." In: *Journal of Mathematical Biology* 67.1, pp. 39–68.
- He, Xin et al. (2010). "Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression." In: *PLoS computational biology* 6.9.
- Apgar, Joshua F et al. (2010). "Sloppy models, parameter uncertainty, and the role of experimental design." In: *Molecular BioSystems* 6.10, pp. 1890–900.
- Erguler, Kamil and Michael P H Stumpf (2011). "Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models." In: *Molecular BioSystems* 7.5, pp. 1593–602.
- Brännmark, Cecilia et al. (2010). "Mass and information feedbacks through receptor endocytosis govern insulin signaling as revealed using a parameter-free modeling framework." In: *Journal of Biological Chemistry* 285.26, pp. 20171–9.
- Wu, Ming, Xuerui Yang, and Christina Chan (2009). "A dynamic analysis of IRS-PKR signaling in liver cells: a discrete modeling approach." In: *PloS one* 4.12, e8040.
- Quon, M J and L a Campfield (1991a). "A mathematical model and computer simulation study of insulin receptor regulation." In: *Journal of theoretical biology* 150.1, pp. 59–72.
- Wanant, Sumanas and Michael J Quon (2000). "Insulin Receptor Binding Kinetics: Modeling and Simulation Studies." In: *Journal of Theoretical Biology* 205.3, pp. 355–364.
- Sedaghat, Ahmad R, Arthur Sherman, and Michael J Quon (2002). "A mathematical model of metabolic insulin signaling pathways." In: *American journal of physiology. Endocrinology and metabolism* 283.5, E1084–101.

- Nyman, Elin, Gunnar Cedersund, and Peter Strålfors (2012). "Insulin signaling - mathematical modeling comes of age." In: *Trends in endocrinology and metabolism: TEM* 23.3, pp. 107–115.
- Shearn, Colin T et al. (2011). "Phosphatase and tensin homolog deleted on chromosome 10 (PTEN) inhibition by 4-hydroxynonenal leads to increased Akt activation in hepatocytes." In: *Molecular pharmacology* 79.6, pp. 941–952.
- Quon, M J and L a Campfield (1991b). "A mathematical model and computer simulation study of insulin sensitive glucose transporter regulation." In: *Journal of theoretical biology* 150.1, pp. 93–107.
- Kholodenko, B., M. B. Yaffe, and W. Kolch (2012). "Computational Approaches for Analyzing Information Flow in Biological Networks." In: *Science Signaling* 5.220, re1.
- Bachmann, J et al. (2012). "Predictive mathematical models of cancer signalling pathways." In: *Journal of Internal Medicine* 271.2, pp. 155–65.
- Cornish-Bowden, Athel (1995). *Fundamentals of Enzyme Kinetics*. Revised. London: Portland Press.
- Schenter, Gregory K et al. (1999). "Statistical Analyses and Theoretical Models of Single-Molecule Enzymatic Dynamics." In: *Journal of Physical Chemistry A*, pp. 10477–10488.
- Bersani, Alberto Maria and Guido Acqua (2010). "Is there anything left to say on enzyme kinetic constants and quasi-steady state approximation?" In: *Journal of Mathematical Chemistry* 50.2, pp. 335–344.
- Blüthgen, Nils et al. (2006). "Effects of sequestration on signal transduction cascades." In: *FEBS Journal* 273.5, pp. 895–906.
- Kawai, Shigeyuki and Kousaku Murata (2008). "Structure and Function of NAD Kinase and NADP Phosphatase: Key Enzymes That Regulate the Intracellular Balance of NAD(H) and NADP(H)." In: *Bioscience, Biotechnology and Biochemistry* 72.4, pp. 919–930.
- Noor, Elad et al. (2013). "A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects." In: *FEBS letters* 587.17, pp. 2772–7.
- Schnell, S (2003). "A Century of Enzyme Kinetics : Reliability of the  $K_M$  and  $v_{max}$  Estimates." In: *Comments on Theoretical Biology* 8, pp. 169–187.

- Segel, Lee A (1988). "On the validity of the steady state assumption of enzyme kinetics." In: *Bulletin of Mathematical Biology* 50.6, pp. 579–593.
- Tzafriri, A Rami and Elazer R Edelman (2007). "Quasi-steady-state kinetics at enzyme and substrate concentrations in excess of the Michaelis-Menten constant." In: *Journal of Theoretical Biology* 245.4, pp. 737–48.
- Cha, Sungman (1970). "Kinetic Behaviour at High Enzyme Concentrations." In: *Journal of Biological Chemistry* 245.18.
- Tan, Shi-Xiong et al. (2012). "Amplification and Demultiplexing in Insulin-regulated Akt Protein Kinase Pathway in Adipocytes." In: *The Journal of biological chemistry* 287.9, pp. 6128–38.
- Marangoni, Alejandro G (2003). "Two-Substrate Reactions." In: *Enzyme Kinetics A Modern Approach*. Vol. 90. New Jersey: John Wiley & Sons, pp. 90–101.
- Berthoumieux, Sara et al. (2012). "On the identifiability of metabolic network models." In: *Journal of Mathematical Biology*.
- Cedersund, Gunnar and Jacob Roll (2009). "Systems biology: model based evaluation and comparison of potential explanations for given biological data." In: *FEBS journal* 276.4, pp. 903–22.
- Ferrell, James E and Sang Hoon Ha (2014). "Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity." In: *Trends in biochemical sciences* 39.10, pp. 496–503.
- Zewe, Virginia and Herbert J Fromm (1962). "Kinetic Studies of Rabbit Muscle Lactate Dehydrogenase." In: *The Journal of Biological Chemistry* 237.5, pp. 1668–1675.
- Gutfreund, H et al. (1968). "The Kinetics of the Reversible Inhibition of Heart Lactate Dehydrogenase through the Formation of the Enzyme-Oxidized Nicotinamide-Adenin Dinucleotide-Pyruvate Compound." In: *Journal of Biochemistry* 106, pp. 683–687.
- McClendon, Sebastian, Nick Zhadin, and Robert Callender (2005). "The approach to the Michaelis complex in lactate dehydrogenase: the substrate binding pathway." In: *Biophysical journal* 89.3, pp. 2024–32.
- Yancey, Paul H and George N Somero (1978). "Temperature Dependence of Intracellular pH: Its Role in the Conservation of Pyruvate Apparent Km Values of Vertebrate Lactate Dehydrogenase." In: *Journal of Comparative Physiology B* 125, pp. 129–134.

- Klinke, David J (2009). "An empirical Bayesian approach for model-based inference of cellular signaling networks." In: *BMC Bioinformatics* 10, p. 371.
- Hucka, M. et al. (2003). "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models." In: *Bioinformatics* 19.4, pp. 524–531.
- Novere, Nicolas Le et al. (2009). "The Systems Biology Graphical Notation." In: *Nature Biotechnology* 27.8, pp. 735–742.
- Hucka, Michael et al. (2010). *The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core*. Tech. rep.
- Funahashi, Akira et al. (2003). "CellDesigner: a process diagram editor for gene-regulatory and biochemical networks." In: *Biosilico* 1.5, pp. 159–162.
- Alves, Rui, Fernando Antunes, and Armindo Salvador (2006). "Tools for kinetic modeling of biochemical networks." In: *Nature biotechnology* 24.6, pp. 667–672.
- Tyson, John J, Katherine C Chen, and Bela Novak (2003). "Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell." In: *Current Opinion in Cell Biology* 15.2, pp. 221–231.
- Lemmon, Mark a and Joseph Schlessinger (2010). "Cell signaling by receptor tyrosine kinases." In: *Cell* 141.7, pp. 1117–34.
- Locascio, Lauren E. and Daniel J. Donoghue (2013). "KIDs rule: regulatory phosphorylation of RTKs." In: *Trends in Biochemical Sciences* 38.2, pp. 75–84.
- Mitchell, Amir, Ping Wei, and Wendell A Lim (2015). "Oscillatory stress stimulation uncovers an Achilles' heel of the yeast MAPK signaling network." In: *Science* 350.November, pp. 1–7.
- Ryu, Hyunryul et al. (2015). "Frequency modulation of ERK activation dynamics rewires cell fate." In: *Molecular systems biology* 11, p. 838.
- Kholodenko, Boris N. (2006). "Cell Signalling Dynamics in Time and Space." In: *Nature Reviews Molecular Cell Biology* 7.3, pp. 165–176. arXiv: [NIHMS150003](#).
- Housden, Benjamin E. and Norbert Perrimon (2014). "Spatial and temporal organization of signaling pathways." In: *Trends in Biochemical Sciences* 39.10, pp. 457–464.

- Wong, Martin Kin Lok et al. (2015). "A generalised enzyme kinetic model for predicting the behaviour of complex biochemical systems." In: *Febs Open Bio* 5, pp. 226–239.
- Thorpe, Lauren M., Haluk Yuzugullu, and Jean J. Zhao (2015). "PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting." In: *Nature Reviews Cancer* 15.1, pp. 7–24.
- Guillermet-Guibert, J. et al. (2008). "The p110 $\alpha$  isoform of phosphoinositide 3-kinase signals downstream of G protein-coupled receptors and is functionally redundant with p110 $\beta$ ." In: *Proceedings of the National Academy of Sciences* 105.24, pp. 8292–8297.
- John, Mathias, Cedric Lhoussaine, and Joachim Niehren (2009). "Dynamic Compartments in the Imperative  $\pi$ -Calculus." In: *Computational Methods in Systems Biology*. John2009, pp. 235–250.
- Fröjdö, Sara, Hubert Vidal, and Luciano Pirola (2009). "Alterations of insulin signaling in type 2 diabetes : A review of the current evidence from humans." In: *Biochimica et Biophysica Acta* 1792.2, pp. 83–92.
- Bogan, Jonathan S (2012). "Regulation of Glucose Transporter Translocation in Health and Diabetes." In: *Annual review of biochemistry* March.
- Gregory, Phillip (2005). *Bayesian Logical Data Analysis for the Physical Sciences*.
- Brooks, Steve et al. (2011). *Handbook of Markov Chain Monte Carlo*. Boca Raton: CRC Press.
- Kumar, Neil et al. (2007). "Quantitative analysis of Akt phosphorylation and activity in response to EGF and insulin treatment." In: *Biochemical and Biophysical Research Communications* 354.1, pp. 14–20.
- Zhang, Xiaoling et al. (2006). "Kinetic mechanism of AKT/PKB enzyme family." In: *Journal of Biological Chemistry* 281.20, pp. 13949–13956.
- Alessi, Dario R et al. (1996). "Mechanism of activation of protein kinase B by insulin and IGF-1." In: *The EMBO journal* 15.23, pp. 6541–51.
- Potter, Christopher J, Laura G Pedraza, and Tian Xu (2002). "Akt regulates growth by directly phosphorylating Tsc2." In: *Nature cell biology* 4.9, pp. 658–65.



- Harrington, Laura S et al. (2004a). "The TSC1-2 tumor suppressor controls insulin-PI3K signaling via regulation of IRS proteins." In: *The Journal of cell biology* 166.2, pp. 213–23.
- Kohn, a D, K S Kovacina, and R a Roth (1995). "Insulin stimulates the kinase activity of RAC-PK, a pleckstrin homology domain containing ser/thr kinase." In: *The EMBO journal* 14.17, pp. 4288–95.
- Shults, Melissa D et al. (2005). "A multiplexed homogeneous fluorescence-based assay for protein kinase activity in cell lysates." In: *Nature methods* 2.4, pp. 277–283.
- Vinod, Palakkad Krishnan Unni and Kareenhalli Viswanath Venkatesh (2009). "Quantification of the effect of amino acids on an integrated mTOR and insulin signaling pathway." In: *Molecular bioSystems* 5.10, pp. 1163–73.
- Du, Yipeng and Taotao Wei (2014). "Inputs and outputs of insulin receptor." In: *Protein & cell*.
- Ng, Yvonne, Georg Ramm, and David E James (2010). "Dissecting the mechanism of insulin resistance using a novel heterodimerization strategy to activate Akt." In: *The Journal of biological chemistry* 285.8, pp. 5232–9.
- Hoehn, Kyle L et al. (2008). "IRS1-independent defects define major nodes of insulin resistance." In: *Cell metabolism* 7.5, pp. 421–33.
- Gibson, Daniel G et al. (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." In: *Nature methods* 6.5, pp. 343–5.
- Burchfield, James G. et al. (2013). "Novel systems for dynamically assessing insulin action in live cells reveals heterogeneity in the insulin response." In: *Traffic* 14.3, pp. 259–273.