

# Mathematical models of signalling through G proteins and phospholipids

*Robert John Stanley*

A thesis submitted in partial fulfilment of the  
requirements for the degree of:

*Doctor of Philosophy of  
University College London*

2015

Primary supervisor:

*Prof. Geraint M.H. Thomas*

Secondary supervisor:

*Dr. Kevin Bryson*



*I, Robert John Stanley, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.*



# Abstract

*G proteins and phospholipids are two major classes of signalling molecule. They are each—independently and together—involved in diverse ‘signalling pathways’ – biochemical networks through which cells maintain healthy responses to stimuli.*

*A unique ‘cross-talk motif’ is formed by regulation of the phospholipid-modifying enzymes phospholipase D (PLD) and phosphatidylinositol 4-phosphate 5-kinase (PI4P5K) by the Arf family of small G proteins and—curiously—each other’s product.*

*Here, understanding of this inherently complex motif has been strengthened by the development and analysis of mathematical models, specifically systems of ordinary differential equations (ODEs).*

*Construction of simple empirical models suggests asymmetry in the mechanisms of regulation of the enzymes is responsible for production of two distinct outgoing signals from a single input signal, one displaying threshold activation behaviour.*

*Additionally, well-defined quasi-steady-state (QSS) mechanistic models (à la Michaelis-Menten) have been developed for each of: PLD; PI4P5K; and G protein/Arf regulation. During this process—due to insufficient pre-existing descriptions—biochemically-plausible assumptions were required for certain regulatory and catalytic interactions.*

*Analysis of the G protein regulation models establishes that—contrary to previous representations—this regulation is best described by a balance/unbalance mechanism, where observed activation absolutely requires the presence of the inactivator.*

*Together, the QSS models can be combined to form a complete model of the Arf/PLD/PI4P5K motif suitable for computational simulation – preliminary parameters show that this model is capable of displaying physiologically-plausible behaviours*

*These results give a better understanding of the signalling role of the Arf/PLD/PI4P5K motif; lead to novel biological hypotheses amenable to later experimental validation; highlight where our current biological understanding of the system is insufficient; and suggest novel methods for the therapeutic control of G proteins.*



# Contents

List of Figures . . . . .	13
List of Tables . . . . .	19
List of Abbreviations . . . . .	21
<b>1 Introduction . . . . .</b>	<b>29</b>
1.1 Models in biochemistry . . . . .	31
1.2 Mathematical proof versus numerical simulation . . . . .	32
1.3 Empirical versus mechanistic models . . . . .	33
1.4 The scale of models . . . . .	34
1.5 Thesis outline . . . . .	35
<b>2 Signalling through G proteins and phospholipids . . . . .</b>	<b>37</b>
2.1 G proteins . . . . .	39
2.2 Arf family G proteins . . . . .	41
2.3 Phospholipids . . . . .	43
2.3.1 Phosphatidylcholine . . . . .	44
2.3.2 Phosphatidic acid . . . . .	44
2.3.3 Phosphoinositides . . . . .	45
2.4 Phospholipase D . . . . .	49
2.5 Phosphatidylinositol 4-phosphate 5-kinase . . . . .	50
2.6 The Arf/PLD/PI4P5K phospholipid signalling motif . . . . .	51
2.7 Related pathways . . . . .	53
2.7.1 Regulation of Arf by phospholipids . . . . .	53
2.7.2 The PI cycle . . . . .	53
2.7.3 The PI3K/AKT/mTOR pathway . . . . .	54
2.8 Summary . . . . .	54

3	Methods and derivations . . . . .	55
3.1	Methods for ordinary differential equations (ODEs) . . . . .	57
3.1.1	Simulation . . . . .	57
3.1.2	Phase plane analysis . . . . .	58
3.1.3	Steady-states . . . . .	58
3.1.4	Stability of steady-states . . . . .	58
3.2	Mass-action models . . . . .	59
3.2.1	Notation . . . . .	59
3.2.2	The law of mass-action . . . . .	62
3.2.3	Time-scale separation . . . . .	65
3.2.4	Rapid equilibrium . . . . .	65
3.2.5	Quasi-steady-state . . . . .	66
3.3	Linear framework for time-scale separation . . . . .	66
3.3.1	Example: the Michaelis-Menten equation . . . . .	69
3.3.2	Example: the Michaelis-Menten with product inhibition equation . . . . .	71
3.4	Python module: enzymegraph . . . . .	73
3.4.1	Algorithm for enumeration of spanning trees . . . . .	73
3.4.2	Using the enzymegraph module . . . . .	75
3.5	Cytosol-membrane transitions . . . . .	77
3.5.1	Notation . . . . .	77
3.5.2	Surface adsorption . . . . .	79
3.5.3	Langmuir's model . . . . .	81
3.5.4	Random sequential adsorption model . . . . .	81
3.6	Summary . . . . .	82
4	Empirical models . . . . .	83
4.1	Arf/PLD/PI4P5K: pseudo-mass-action model . . . . .	85
4.1.1	Implicit solution for the pseudo-mass-action model . . . . .	88
4.1.2	Implicit solution for the pseudo-mass-action model: beha- viour when $t \gg 0$ . . . . .	90
4.2	Theoretical motifs . . . . .	90
4.2.1	Logical biochemistry . . . . .	91
4.2.2	Pseudo-mass-action models: construction . . . . .	93
4.2.3	Pseudo-mass-action models: analysis . . . . .	93
4.3	Bounding growth using the Hill equation . . . . .	99
4.3.1	Hill models: construction . . . . .	99



4.3.2	Hill models: analysis . . . . .	101
4.4	Bounding growth using conservation of mass . . . . .	108
4.4.1	Conservation of mass models: construction . . . . .	108
4.4.2	Conservation of mass models: analysis . . . . .	109
4.5	Discussion . . . . .	118
5	PLD & PI4P5K: mechanisms and models . . . . .	123
5.1	Di-allosteric mechanism: QSS model . . . . .	125
5.2	PLD: mechanism . . . . .	130
5.3	PLD: QSS model . . . . .	137
5.4	PI4P5K: mechanism . . . . .	144
5.5	PI4P5K: QSS models . . . . .	151
5.5.1	3-site QSS model . . . . .	153
5.5.2	2-site QSS model . . . . .	163
5.6	Discussion . . . . .	172
6	Arf family G proteins: mechanisms and models . . . . .	175
6.1	GEFs: mechanism and QSS model . . . . .	181
6.2	GAPs: mechanism and QSS model . . . . .	185
6.3	Simulations of G protein regulation . . . . .	185
6.4	Notes on the regulation of G proteins by GEFs and GAPs . . . . .	191
6.4.1	GEFs act to attain an equilibrium between active and inactive G protein . . . . .	191
6.4.2	GEFs can be inhibitory . . . . .	193
6.4.3	GTPase activity has a crucial role in the observed activation of G protein . . . . .	198
6.4.4	Sequestration is not sufficient to restore the activation of G proteins by GEFs . . . . .	203
6.4.5	Regulation of G protein regulation by GEFs and GAPs is by a balance/imbalance mechanism . . . . .	204
6.5	Regulation of Arf: specific mechanisms . . . . .	207
6.6	Regulation of Arf: QSS model . . . . .	209
6.7	Discussion . . . . .	212
7	The Arf/PLD/PI4P5K motif: mechanistic models . . . . .	215
7.1	Estimates of protein & lipid concentrations . . . . .	217
7.2	Notes on adsorption processes . . . . .	218

## 10 CONTENTS

7.2.1	A form for the fractional surface coverage . . . . .	218
7.2.2	Estimating membrane surface area . . . . .	219
7.2.3	Estimating maximal surface concentration of adsorbate . . . . .	221
7.3	PLD model . . . . .	221
7.3.1	Parameters for the PLD model . . . . .	223
7.4	PI4P5K models . . . . .	223
7.4.1	PI4P5K 3-site model . . . . .	224
7.4.2	PI4P5K 2-site model . . . . .	224
7.4.3	Parameters for the PI4P5K models . . . . .	224
7.4.4	Equality of models when $[PA] = 0$ . . . . .	225
7.4.5	Steady-state behaviour of the 3-site model . . . . .	225
7.4.6	PI4P5K as a ‘competitive activator’ . . . . .	227
7.5	Arf regulation model . . . . .	231
7.5.1	Parameters for the Arf regulation model . . . . .	231
7.5.2	Perturbing the complete model via the GAP concentration . . . . .	233
7.6	Arf/PLD/PI4P5K motif: complete models . . . . .	233
7.6.1	Preliminary analysis of the complete models . . . . .	235
7.7	Discussion . . . . .	235
8	Highlights & perspectives . . . . .	239
8.1	Summary & highlights . . . . .	239
8.1.1	Semi-automation of model construction using the Python module enzymegraph . . . . .	240
8.1.2	Construction of both empirical and mechanistic models of the Arf/PLD/PI4P5K signalling motif . . . . .	240
8.1.3	Comparison with theoretical systems . . . . .	241
8.1.4	Empirical models of the Arf/PLD/PI4P5K signalling motif . . . . .	241
8.1.5	Empirical models: hypothesised mechanism for the Arf/PLD/PI4P5K signalling motif . . . . .	242
8.1.6	Mechanistic models of PLD & PI4P5K . . . . .	242
8.1.7	A lack of well-defined biochemistry . . . . .	243
8.1.8	Complexity of mechanistic ( $n$ -site) models . . . . .	243
8.1.9	Membrane-recruitment of PI4P5K by Arf . . . . .	244
8.1.10	Mechanistic model of Arf regulation by GEFs and GAPs . . . . .	244
8.1.11	Regulation of G proteins by GEFs and GTPase activity is via a balance/imbalance mechanism . . . . .	245
8.1.12	Mechanistic models of the Arf/PLD/PI4P5K signalling motif . . . . .	245

8.2	Experimental approaches to validation . . . . .	247
8.2.1	Biochemical assays . . . . .	247
8.2.2	G protein assays using fluorescent GTP . . . . .	249
8.2.3	Surface-adsorption assays using surface plasmon resonance . . . . .	250
8.3	Proposed extensions . . . . .	250
8.3.1	Feedback of phospholipids on GAPs and GEFs . . . . .	250
8.3.2	Inclusion of ATP/ADP dynamics in kinase models . . . . .	250
8.3.3	$n$ -site models . . . . .	251
8.4	Perspectives . . . . .	251
8.4.1	Theoretical & synthetic biology . . . . .	251
8.4.2	Evolutionary constraints on enzyme kinetics, product inhibition . . . . .	253
8.4.3	Perturbing intracellular signalling through nucleotides . . . . .	253
8.4.4	Topology of membranes . . . . .	255
8.5	Conclusion . . . . .	256
A	Python module enzymegraph: source code . . . . .	259
B	Derivation of the PMA model . . . . .	267
C	Complete model: source code and additional figures . . . . .	271
	Bibliography . . . . .	279



## List of Figures

1.1	Examples of alternative descriptions of intracellular signalling pathways.	30
1.2	Two different scales for biochemical systems. . . . .	30
2.1	G protein regulation. . . . .	38
2.2	GEFs and GPCRs are capable of activating (and inactivating) G proteins.	38
2.3	The differential localisation of active and inactive Arf. . . . .	40
2.4	Molecular structures of the discussed phospholipids. . . . .	42
2.5	There are eight possible phosphoinositides, formed by the combinatorial phosphorylation of phosphatidylinositol (PI) on the 3, 4, and 5 positions on the inositol head group. . . . .	46
2.6	Subset of the routes for phosphorylation and dephosphorylation (from the full graph in figure 2.5) that are present in each cellular membrane. Modified from Figure 1c in Di Paolo & De Camilli (2006). . . . .	47
2.7	Structures for Myristoylated yeast Arf1-GDP; Human Arf6-GTP; Phospholipase D2; Phosphatidylinositol 4-phosphate 5-kinase. . . . .	48
2.8	The complete Arf/PLD/PI4P5K phospholipid signalling motif. . . . .	52
2.9	The PI cycle formed by the breakdown of PI(4,5)P <sub>2</sub> by phospholipase C (PLC) and the re-synthesis of phosphatidylinositol (PI). . . . .	52
3.1	Plots of the dynamics of the Van der Pol oscillator. . . . .	56
3.2	Stability classes of two dimensional steady-states, illustrated by the nearby trajectories. . . . .	60
3.3	Simulation of the Michaelis-Menten system with the ODEs in equations 3.5 to 3.8. . . . .	64
3.4	Illustration of the concept of time-scale separation. . . . .	64
3.5	Graph and spanning trees for the Michaelis-Menten reaction scheme. . . . .	68
3.6	Graph and spanning trees for the Michaelis-Menten with product inhibition reaction scheme. . . . .	70

14 LIST OF FIGURES

3.7	Application of the <code>SpanningTrees(G)</code> algorithm to find all of the directed spanning trees of an example graph. . . . .	74
3.8	Simple reversible surface adsorption of an enzyme, $E$ , binding to a membrane. . . . .	78
3.9	Illustrations of the physical meaning of adsorption model parameters. . . . .	78
3.10	Example jammed coverages for the RSA model on a (bounded) surface with different radiuses of adsorbate. . . . .	78
3.11	Langmuir's model. . . . .	80
3.12	Random sequential adsorption (RSA) model. . . . .	80
4.1	Cartoon phase planes. The stability and location of the single steady-state in the pseudo-mass-action model depends on the choice of parameters and the value of $z$ (the activation by Arf). . . . .	86
4.2	Requirements for the activation of PLD and PI4P5K in respect to the presence of two regulatory molecules. . . . .	92
4.3	Logic gates and their inputs ( $A, B$ ) and outputs. . . . .	92
4.4	Three theoretically possible signalling motifs, in analogy to the Arf/PLD/PI4P5K signalling motif. . . . .	92
4.5	Steady-states values of $x$ and $y$ (the products of thw two enzymes) as a function of $z$ for the three pseudo-mass-action models listed in table 4.1. . . . .	96
4.6	Eigenvalues as a function of $z$ for the three pseudo-mass-action models listed in table 4.1. . . . .	97
4.7	The Hill function. . . . .	100
4.8	Steady-states values of $x$ and $y$ (the products of thw two enzymes) as a function of $z$ for the three Hill models listed in table 4.2. . . . .	102
4.9	Transcritical bifurcation in the Hill model of the AND-AND motif. . . . .	104
4.10	Transcritical bifurcation in the Hill model of the AND-OR motif. . . . .	106
4.11	Steady-states values of $x$ and $y$ (the products of thw two enzymes) as a function of $z$ for the seven conservation of mass models listed in table 4.3. . . . .	113
4.12	Transcritical bifurcation in the conservation of mass model of the AND-OR-Y motif. . . . .	116
5.1	The di-allosteric mechanism, where a generic 3-site enzyme has two allosteric binding sites for regulatory molecules, in addition to its substrate binding site; and unrestricted order of binding. . . . .	124
5.2	The graph on the enzyme complexes for the di-allosteric mechanism shown in figure 5.1 – an enzyme $E$ with two allosteric regulators $A$ and $R$ . . . . .	126

5.3	The 16 spanning trees of the graph on the enzyme complexes for the di-allosteric mechanism, shown in figure 5.2. . . . .	128
5.4	The 3-site assumption for PLD, with a substrate binding pocket and distinct, allosteric binding sites for the two regulators, Arf and PI(4,5)P <sub>2</sub> . . . . .	132
5.5	The 3-site mechanism for PLD including regulation by Arf and PI(4,5)P <sub>2</sub> , and catalytic action. . . . .	132
5.6	Graph on the enzyme complexes for the 3-site model for PLD. . . . .	136
5.7	A small proportion of the 3072 spanning trees of the 3 site PLD graph shown in figure 5.6. . . . .	138
5.8	Subgraphs of the 3-site PLD graph. . . . .	140
5.9	Simple reversible surface adsorption of an enzyme binding to a membrane. . . . .	146
5.10	The 3-site and 2-site hypotheses for PI4P5K. . . . .	146
5.11	The 3-site and 2-site mechanisms for PI4P5K including regulation by Arf and PI(4,5)P <sub>2</sub> , and catalytic action. . . . .	148
5.12	Graph on the enzyme complexes for the 3-site model of PI4P5K. . . . .	152
5.13	A small proportion of the 1,612,800 spanning trees of the graph on the enzyme complexes for the mechanism shown in figure 5.12. . . . .	156
5.14	Subgraphs of the 3-site PI4P5K graph. . . . .	158
5.15	Graph on the enzyme complexes for the 2-site model of PI4P5K. . . . .	162
5.16	Possible approaches for the simplification of the graph shown in figure 5.15. . . . .	162
5.17	The catalytic components, subgraphs of the 2-site PI4P5K graph. . . . .	166
5.18	The 16 spanning trees of the graph on the enzyme complexes for the catalytic component of the 2 site mechanism, shown in figure 5.17. . . . .	168
6.1	G protein regulation. . . . .	176
6.2	The molecular mechanism for the guanine nucleotide exchange activity of GEFs. . . . .	178
6.3	Graph on the enzyme complexes for the GEF mechanism. . . . .	180
6.4	The 16 spanning trees of the graph on the enzyme complexes for the GEF mechanism, shown as figure 6.3. . . . .	182
6.5	Simulation of the system described by equations 6.15 to 6.20. . . . .	188
6.6	Enlargement of figure 6.5A with the total concentration of GEF complexes also drawn. . . . .	192
6.7	The steady-state proportion of active G protein as a function of the total concentration of GEF. . . . .	196
6.8	The steady-state proportion of active G protein as a function of the total concentration of GEF divided by the total rate of GTPase activity. . . . .	202

16 LIST OF FIGURES

6.9	The proposed balance/imbalance mechanism by which GEFs and GAPs have distinct and crucial roles, and collaborate to regulate G proteins. . . . .	206
6.10	Regulation of Arf by GEFs, GAPs, and an adsorption process. . . . .	208
7.1	Illustration of lipid surface and volume concentrations. . . . .	220
7.2	Composition of Hill functions. Illustration of the effect on the rate of production of PA by the concentrations of Arf and PI(4,5)P <sub>2</sub> . . . . .	220
7.3	The effect of the concentration of PA on the steady-state proportion of cytosolic PI4P4K (concentration divided by the total concentration of PI4P5K) and the steady-state concentration of PI(4,5)P <sub>2</sub> for the 3-site model of PI4P5K. . . . .	226
7.4	The effect of varying the parameters $\mu_{rel}$ and $\mu_{exc}$ on the rate of the 2-site PI4P5K model. . . . .	228
7.5	The effect of the concentration of PA on the steady-state proportion of cytosolic PI4P4K (concentration divided by the total PI4P5K concentration) and the steady-state concentration of PI(4,5)P <sub>2</sub> for the 3-site model and the 2-site model. . . . .	229
7.6	The effect of varying the active concentrations of GEF and GAP on the steady-state concentration of active Arf. . . . .	230
7.7	Diagram showing the components of the complete model of the Arf/PLD/PI4P5K motif. . . . .	232
7.8	Simulation of the complete mechanistic model of the Arf/PLD/PI4P5K motif. . . . .	234
7.9	Steady-state concentrations of active Arf, PA, and PI(4,5)P <sub>2</sub> as a function of the concentration of active GAP, assuming the 3-site model of PI4P5K. . . . .	236
8.1	Two approaches for the production of lipid vesicles with varying concentrations of PIP. . . . .	246
8.2	Outline of surface plasmon resonance using sensor chip with captured lipid vesicles. . . . .	248
8.3	Demonstrating a possible trade-off between specificity to substrate orientation and product-inhibition, geometrically. . . . .	252
8.4	Proposed effect of altering nucleotide levels for a cascade of G protein and kinase signalling. . . . .	254
8.5	Proposed mechanism for which the choline-headgroup of PC is an inhibitors of PI4P5K activity by restricting access of the kinase domain of PI4P5K to the PIP. . . . .	254



8.6	Key themes and connections between themes within this thesis. . . . .	258
B.1	Phase portraits of equation B.2. . . . .	268
C.1	Simulations of the complete mechanistic model of the Arf/PLD/PI4P5K motif. . . . .	274
C.2	Steady-state concentrations of active Arf, PA, and PI(4,5)P <sub>2</sub> as a function of the concentration of active GAP, assuming the 2-site model of PI4P5K. . . . .	277



# List of Tables

3.1	ODEs generated by application of the law of mass-action for the five common reaction schemes introduced in section 3.2.1. . . . .	60
4.1	Pseudo-mass-action models and characterisation of steady-states for the three motifs (AND-AND; AND-OR; OR-OR). . . . .	94
4.2	Hill models for the three motifs (AND-AND; AND-OR; OR-OR). . . . .	100
4.3	Models of the motifs with one or more reverse reactions. . . . .	110
4.4	Inputs, in the form of concentrations of Arf, and associated outputs, in the form of PA and PI(4,5)P <sub>2</sub> production, as implied by the Hill and conservation of mass models. . . . .	120
6.1	Kinetic rate parameters ( $k_1, \dots, k_8$ ) and summary rate constants for the mass-action and QSS GEF models, used for illustrative purposes. . . . .	186
7.1	Relative abundances of the species included in the models of PLD, PI4P5K, and the regulation of Arf. . . . .	216
7.2	Provisional parameters for the PLD model. . . . .	222
7.3	Provisional parameters for the 3-site and 2-site PI4P5K models. . . . .	222
7.4	Provisional parameters for the regulation of Arf model. . . . .	230



## List of Abbreviations

Arf	<b>Arf</b> / ■	Arf, <i>ADP ribosylation factor (historic)</i> <i>monomeric G protein</i>
Arf•GTP	<b>Arf•GTP</b>	Arf with GTP bound <i>active form of Arf</i>
Arf•GDP	<b>Arf•GDP</b>	Arf with GDP bound <i>inactive form of Arf</i>
GEF	<b>GEF</b>	guanine nucleotide exchange factor <i>activator of G proteins</i>
GAP	<b>GAP</b>	GTPase activating protein <i>inactivator of G proteins</i>
PLD	<b>PLD</b>	phospholipase D <i>lipid-modifying enzyme</i>
PC	<b>PC</b> / □	phosphatidylcholine <i>substrate of PLD</i>
PA	<b>PA</b> / ■	phosphatidic acid <i>product of PLD</i>
PI4P5K	<b>PIP<sub>2</sub></b>	phosphatidylinositol 4-phosphate 5-kinase <i>lipid-modifying enzyme</i>
PI4P	<b>PIP</b> / □	phosphatidylinositol 4-phosphate <i>substrate of PI4P5K</i>
PI(4,5)P <sub>2</sub>	<b>PIP<sub>2</sub></b> / ■	phosphatidylinositol (4,5)-bisphosphate <i>product of PI4P5K</i>
GTP	<b>GTP</b>	guanosine triphosphate
GDP	<b>GDP</b>	guanosine diphosphate
ATP	<b>ATP</b>	adenosine triphosphate
ADP	<b>ADP</b>	adenosine diphosphate
Cho	<b>Cho</b>	choline
P <sub>i</sub>	<b>P<sub>i</sub></b>	inorganic phosphate



## Acknowledgements

I will forever appreciate the support, guidance, and encouragement of my supervisors: Geraint Thomas and Kevin Bryson. They have allowed me the freedom to tread my own path, but have always been there to explain the scenery.

I would like to acknowledge the camaraderie of the entire Computational Systems and Synthetic Biology (CSSB) group and associated people—of who were few when I joined and now are many: Gerold Baier, Chris Barnes, Alex Fedorec, Riana Gai-fulina, Hannah Lawrence, Miriam Leon, Aaran Lewis, Phil Lewis, Andrew Maher, Helina Marshall, Nadine Mogford, Tanel Ozdemir, Lourdes Sriraja, Gordon Walsh, Marc Williams, Bethan Wolfenden, and Mae Woods. Special thanks goes to Janine Symonds, who started in the CSSB group at the same time as me and who has been a constant friend throughout. I would also like to thank all the administrative and support staff at UCL who have supported me and my projects, and Sandra Martelli for being a source of useful advice and moral support.

I am thankful for funding from the British Heart Foundation.

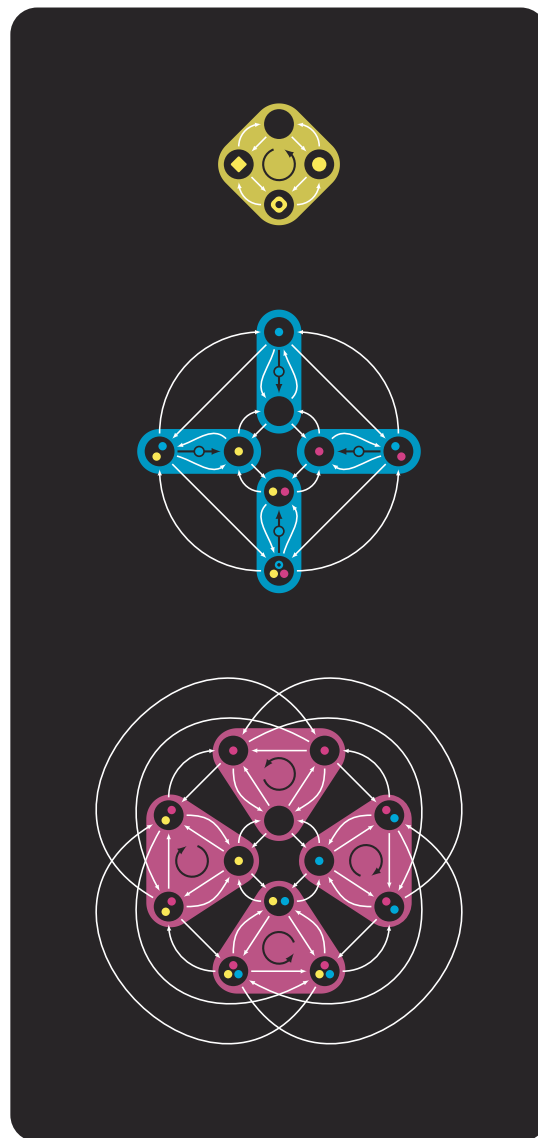
I am especially grateful to researchers from other institutions who welcomed me for visits during my Ph.D.: Jeremy Gunawardena (Harvard Medical School, MA, USA) whose work is a source of great inspiration – this visit was made possible by the support of a Bogue Research Fellowship; and Oliver Ebenhöf (HHU, Düsseldorf, Germany) who supplied me with the phrase ‘oligo-enzyme’.

I greatly appreciate the 28 years of encouragement, advice, and love from my family: my parents and my Gran. I would like to thank Steve and Irene Brown for looking after me, and my friends for their support, certainly including Juliet (Juby) Chippindale and Ryan Dee.

Finally—most importantly—I would like to thank Melissa for the mountain of emotional support which I have been so lucky to have had throughout it all.







*Enzymes are biological molecules that catalyse a reaction, converting one molecule (the substrate) into another (the product), a process which may also require the presence of further regulatory molecules. At each stage of the reaction the enzyme will be bound to a different set of molecules. Each of these is called an 'enzyme complex' (the black circles). When the transitions between the enzyme complexes are drawn (the white arrows) we get 'the graph on the enzyme complexes'. (This usage of 'graph' is synonymous with 'network'.) Different enzymes will have different behaviours and so have different graphs. Here I have displayed the graphs of three highly inter-regulated enzymes (from top-to-bottom): an Arf-GEF; PLD; and PI4P5K.*

UCL Graduate School, Research as Art competition 2014 submission



*If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.* John von Neumann (Alt 1972)

*If the difficulty of a physiological problem is mathematical in essence, ten physiologists ignorant of mathematics will get precisely as far as one physiologist ignorant of mathematics, and no further. If a physiologist who knows no mathematics works together with a mathematician who knows no physiology, the one will be unable to state his problem in terms that the other can manipulate, and the second will be unable to put the answers in any form that the first can understand.*

Norbert Wiener (Wiener 1965)



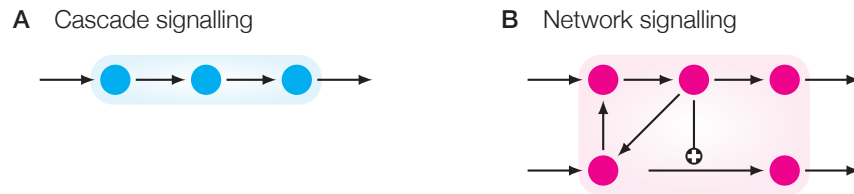
# 1 Introduction

In order to survive, a cell must be able to sense, transmit, and interpret information about its own internal state and external environment. This information is internally transmitted through *signalling pathways* – complex machinery consisting of networks of biochemical interactions. Signalling pathways work—independently and together—to integrate multiple sources of information and relay messages to appropriate intracellular locations. These pathways control behaviours not just related to unforeseen and unexpected external changes but also those behaviours that require a consistent presence, efficiently and reliably maintaining homeostasis.

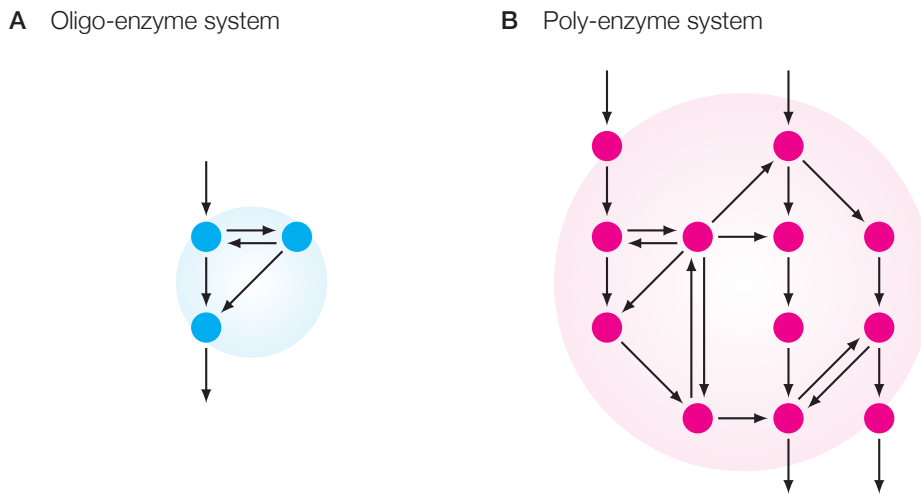
Faults in signalling pathways can lead to abnormal, spontaneous, or dissipated signals, and so cause inappropriate, unwanted, or non-existent responses. Diseases that involve faults in signalling are some of the largest issues in 21<sup>st</sup> century medicine including cancer, heart disease, and diabetes (Berridge 2014). An understanding of how signals are processed in healthy and diseased states is a pressing requirement for the development of novel therapeutic and pharmaceutical approaches in order to correct these faults. This in turn requires an understanding of both individual biochemical mechanisms, and their relationships within entire signalling pathways.

Signalling pathways can be thought of as connected systems of messages (interactions) and messengers (the molecules or medium through which the message propagates). Many different substances and properties are used by cells as messengers including (but not limited to): proteins; elemental ions; small molecule second messengers; and physical characteristics of the cell (e.g. membrane curvature). The mechanisms by which the message is transmitted by interactions of the messengers are also many and varied, including: activation of an enzyme's catalytic activity; localisation of a protein; and, generation of new second messenger molecules.

Signalling pathways are often much more complex than a simple *cascade* of molecular interactions. More frequently they comprise *networks* of interactions, containing divergent and convergent processes, and positive and negative feedback. This distinction is illustrated by figure 1.1. As the size of these complex, network-like signalling



**Figure 1.1** Examples of alternative descriptions of intracellular signalling pathways described by reactions ( $\rightarrow$ ) and regulations ( $\rightarrow\oplus$ ).



**Figure 1.2** Two different scales for biochemical systems, and the terms I use to reference these.

**A** An *oligo-enzyme* (few enzyme) system comprises only a few enzymes and the interactions between them.

**B** A *poly-enzyme* (many enzyme) system contains many enzymes and interactions.

pathways increases the number of interactions and regulations they contain increase combinatorially. This is important because systems of many interacting components are capable of displaying complex, emergent behaviour (Bhalla & Iyengar 1999), which transcends the behaviour of any individual component. For an organism, emergence is a powerful tool for processing intracellular signals; but as an observer, it can confound our understanding of individual mechanisms and processes. Therefore, as the size of a system increases it quickly becomes impractical, inappropriate, or even impossible to continue our investigation using solely traditional biochemical methods.

## 1.1 Models in biochemistry

A better approach for the exploration of complex intracellular signalling pathways is supplementing traditional biochemical investigation with *mathematical* and/or *computational* models. It is crucial to understand that the important words in the last sentence are ‘mathematical’ and ‘computational’ not ‘model’ as we already need to maintain *mental* models of the system of interest.

The primary role of these mathematical and computational models is to mean that we are no longer solely reliant on human mental capacity to understand and predict the outcome of complex systems of interactions.

Additionally, dissemination of mental models primarily occurs through written description and illustration, the clarity of which often depends on the skills of the original author. In comparison, a properly annotated mathematical or computational model should be inherently less ambiguous and so, arguably, a better means of communication.

The study of biochemistry has historically incorporated a mathematical component—for over a century and a half the law of mass-action (see section 3.2.2) has given a rigorous method for the conversion of biochemical reaction schemes into systems of ordinary differential equations (ODEs). Also just over 100 years old is the Michaelis-Menten equation, a general model for irreversible enzymes (Michaelis & Menten 1913; Johnson & Goody 2011). With the later quasi-steady-state (QSS) derivation of this equation by Briggs & Haldane (1925), we begin to have a complete toolbox for deriving models of molecular and enzymatic interactions in a precise manner. More recent methods make the derivation of QSS models from known biochemical mechanisms even more standardised (Gunawardena 2012).

Mass-action and QSS models of molecular processes can be brought together as

component parts of models of whole intracellular signalling pathways. In this way, we begin to create models that may be able to predict a cell's response to complex stimuli, and explore how that response can be modulated for our benefit, for example, to restore healthy behaviours. Usually, these models are sufficiently complex so that computational—rather than mathematical—analysis and simulation is required, see section 1.2.

Mathematical and computational modelling should never be used to completely replace experiments as our primary source of information about the world, as the validity of our predictions can only ultimately be determined through experimental observation. Rather, the best outcome for a model is to suggest novel hypotheses for later experimental validation, ideally identifying experiments that require only simple, inexpensive, and reproducible procedures. In this way, we can move some of the scientific process away from the expensive laboratory work towards the (relatively) inexpensive mathematical/computational modelling. Models also allow us to pose hypotheses about the internal structures of biological systems that otherwise—for many reasons—we may find difficult or impossible to probe.

## 1.2 Mathematical proof versus numerical simulation

Mathematical and computational models inherently contain a number of parameters, values specific to the system of interest. Typically for biochemical models these will be kinetic rates and (initial or basal) concentrations of reactants, products, and enzymes.

With relevant and accurate experimental evidence it becomes possible to *parameterise* a model using statistical fitting methods to produce (what is often seen as) a canonical set of parameters. Alternatively, we can sometimes estimate these parameters from related results or other data.

However, issues arise when referring to a single set of parameters as the canonical set. This is because this set will be inherently specific to the experimental system that was used to generate the originating data. Yet we know that even cells of the same type will have different internal states depending on a large number of intrinsic and extrinsic factors, for example: temperature, pH, and history. So unless its behaviour is trivial it is unlikely that one set of parameters can capture the entire spectrum of behaviours of the biological system under all circumstances.

A better approach—whenever possible—is to use mathematical analysis to *prove things* about models. In this manner, we can attempt to explore all possible qualitative



behaviours of the system for all possible sets of parameters. As always, experiments must be designed for the validation or invalidation of the resulting hypotheses.

Unfortunately, as the complexity of the biological system increases or the accuracy of the model increases, it can quickly become implausible or impossible to use (existing, routine) mathematical methods to analyse a model. So the approach of mathematical proof is generally only possible where the complexity of the model is low, for example in small, uncoupled systems.

For larger, more complicated systems we must rely on computational techniques, most commonly numerical simulation. In this case we must rely on specific sets of parameters (for each simulation), and so additional care must be taken against making definitive statements about the behaviour of the system under all circumstances. Still, there is much information we can obtain about a system in using this approach. Additionally, simulation has a clear role in illustrative purposes, even for systems we can mathematically analyse.

It is also possible to extend this numerical approach to explore arbitrarily more of the behaviour ‘space’ of the model. Using Monte-Carlo methods and modern Bayesian tools it is possible to find distributions of sets of parameters that map to particular behaviours (Toni et al. 2009; Liepe et al. 2010).

In this thesis I will use a combination of mathematical analysis and computational simulation when each is appropriate, to investigate the properties of models.

### 1.3 Empirical versus mechanistic models

There are two distinct approaches that we could use when constructing models of biochemical systems.

An *empirical* approach allows for the construction of models based on the observed behaviours/phenomena of the system. Given an appropriate evidence base (for instance dose-response curves or observed causal responses), we *choose* appropriate mathematical functions and expressions that replicate this behaviour in (at least) a qualitative manner. This can be considered a *top-down* approach, where the (emergent) behaviour of the system is directly modelled. This is appropriate when we want to quickly build simple models for a system, and when we have little understanding of the underlying molecular interactions. However, this approach can lead to models that are biased towards the behaviours we expect to see, and so which may not correctly describe the system under all circumstances – for instance

in diseased states. These models are also, by design, implicitly arbitrary with respect to the choices of the modeller.

A *mechanistic* approach allows for the construction of models by the application of a translational framework (e.g. the law of mass-action) to a well-defined set of individual molecular interactions. This is a *bottom-up* approach and results in a model constructed from *first-principles*. The observed (emergent) behaviour of the system should then be displayed by this model. If not, the model should be rejected, and the underlying biological knowledge and assumptions reconsidered.

As a general rule, mechanistic models are preferable to empirical models as they should be less subjective and implicitly include additional checks as to our understanding of the modelled biochemical process. However, empirical models can often be far simpler to construct, analyse, and simulate.

In this thesis I will consider models built under both paradigms, and attempt to make use of their individual advantages.

## 1.4 The scale of models

The construction of models of whole signalling and regulatory pathways—and even entire cells—has recently become a credible proposition, see for instance Karr et al. (2012). Whole-pathway and whole-cell models combine the totality of our understanding of the individual components of the system into a larger structure, with the aim of predicting higher-order emergent behaviours. However, we often lack sufficiently well-investigated (mechanistic) models of the component parts of these pathways and must often assume generic (empirical) models instead – for example, assuming all enzymes follow the standard Michaelis-Menten mechanism (without specific information that that is the case). In order to have faith in the results of the model we must then assume that the system (both biological and theoretical) is inherently robust to small differences (errors) in its internal interactions. It is entirely possible that biological systems have evolved such robustness, as it is desirable that small differences due to genotype, genetic mutation, and other biological sources of noise do not change the fundamental signalling properties of a system.

Comparatively, it is not likely that this robustness (to small differences) will hold when we consider diseased states, as any evolutionary pressure is more than likely in the opposite direction, towards ensuring that the system does not display this behaviour. This could suggest that models that are good at predicting healthy behaviour, may not always be quite so useful in understanding diseased states.

Altogether, this leaves plenty of scope for improving the predictions of these large models (in health and disease) by improving the physiological-realism of the component parts (e.g. models of individual enzymes).

We must also consider the relationship between the size of the system and our need for accuracy. Consider the two illustrative systems in figure 1.2. One, which could be termed an *oligo-enzyme* system (few enzymes) contains only three elements and six interactions. The other, termed an *poly-enzyme* system (many enzymes) contains 11 elements and 23 interactions. So each interaction is a larger proportion of the oligo-enzyme system than the poly-enzyme system. Therefore, if any one of the interactions is incorrectly modelled (or incorrectly described in the literature) it is more likely to have a greater impact on our predictions in the smaller system. This means that extra care should be taken in the investigation of small systems of few interacting components, in terms of both the biological understanding, and the assumptions made during the modelling process.

In this thesis I will be considering a system of (relatively) few interacting elements, and so these considerations provide extra motivation for ensuring the models produced are particularly physiologically well-founded. The final mechanistic models produced will be ideal for inclusion into larger models of signalling pathways.

## 1.5 Thesis outline

This thesis will focus on the consequences on the signal propagation of the structure and internal mechanisms of a signalling motif that contains regulation of G proteins and phospholipids – the Arf/PLD/PI4P5K motif. This motif is defined by the activation of the enzymes PLD and PI4P5K, by members of the Arf family of G proteins. I will use a theoretical approach towards furthering the understanding of this motif, leading to the construction of a series of mathematical models.

This thesis will continue as follows:

**Chapter 2** contains a review of the current understanding of the biology of the constituent parts of the Arf/PLD/PI4P5K motif; and discussion of related pathways.

**Chapter 3** contains a review of the mathematical methods that will be required for the remainder of the thesis, including an introduction to the two major references for the construction of models: the linear framework for time-scale separation, used for the derivation of quasi-steady state (QSS) models; and

a framework for the construction of models which include physiologically-relevant cytosol-membrane translocations. This chapter will also include: a note on the mathematical notation that I will use throughout the thesis; the re-derivation of specific well-established models; and a description of my own module for the Python programming language which allows the semi-automation of the derivation of QSS models of enzymes.

**Chapter 4** contains an expanded discussion of a previously described empirical model for the Arf/PLD/PI4P5K phospholipid signalling motif; and continues the development of further empirical models. Specifically, I discuss the idea of investigating the properties of the original motif in the light of related but theoretical motifs; and produce a series of models of increasing complexity.

**Chapter 5** contains the development of physiologically-realistic, well-defined QSS mechanistic models of the enzymes PLD and PI4P5K based (wherever possible) on existing biochemical references.

**Chapter 6** contains a exploration of the consequences of the specific manner in which G proteins (of which Arf is a member) are controlled by guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs). Mechanistic models are constructed for both of these regulators, and are used to show that the current understanding of the regulation of G proteins is perhaps often incorrect. Finally, a model of Arf is developed with the addition of the differential localisation of the active and inactive forms.

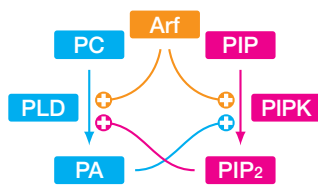
**Chapter 7** collects together the models from the previous two chapters and compiles a preliminary set of parameters in order to finalise a biochemically-plausible, well-defined mechanistic model of the entire Arf/PLD/PI4P5K motif. I demonstrate that this is capable of producing physiologically plausible behaviour.

**Chapter 8** concludes the thesis with a discussion of the material presented within; and a discussion of related, but currently unexplored, ideas.

A brief note on style. Throughout the thesis I shall use 'I' whenever the text discusses my own thoughts and decisions. I shall use 'we' to refer to either: *the scientific community*, particularly during expository remarks; or *the reader and I*, particularly during mathematical derivations.

## 2 Signalling through G proteins and phospholipids

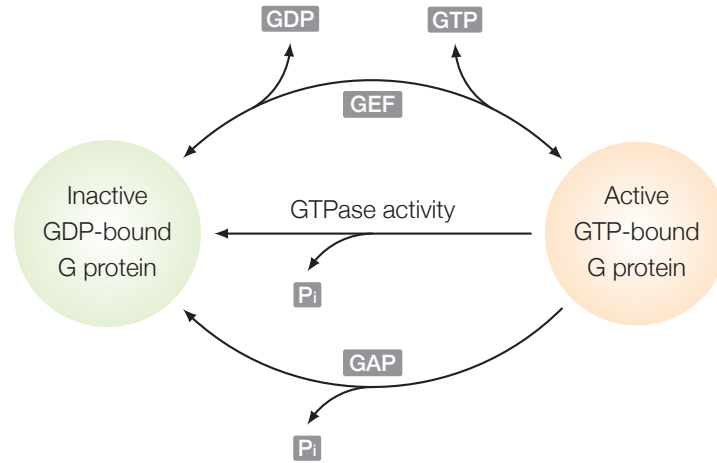
A unique phospholipid signalling motif is comprised of the activation of the phospholipid modifying enzymes phospholipase D (PLD) and phosphatidylinositol 4-phosphate 5-kinase (PI4P5K) by members of the Arf family of G proteins, and their mutual cross-regulation via their products.



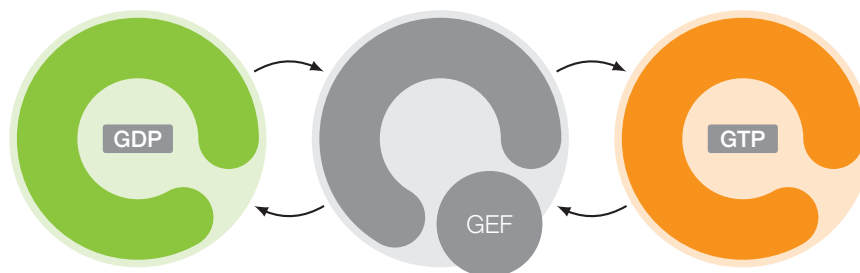
In order to better understand the component process of this signalling motif, in this chapter I will review the current understanding of

- Arfs, and their super-family of G proteins;
- GEFs and GAPs, the molecules which regulate G proteins;
- phospholipids, specifically phosphatidylcholine, phosphatidic acid, and the phosphoinositides;
- phospholipase D (PLD);
- phosphatidylinositol 4-phosphate 5-kinase (PI4P5K);
- and related signalling pathways.

Throughout I will use consistent abbreviations and symbols, defined in the front matter.



**Figure 2.1** G protein regulation. G proteins exist in one of two states depending on the bound guanine nucleotide, traditionally referred to as active or inactive. The switch between states is regulated by GTPase activity, both intrinsic and GAP-mediated; and (reversible) nucleotide exchange catalysed by: GEFs for monomeric G proteins; GPCRs for heterotrimeric G proteins.



**Figure 2.2** GEFs and GPCRs are capable of activating (and inactivating) G proteins by mediating a nucleotide exchange process. This is a sequential process: the GEF/GPCR acts to open the guanine nucleotide binding pocket of the G protein; this allows the release of the bound nucleotide and the formation of a stable nucleotide-free G protein:GEF complex; subsequently another nucleotide (either GTP or GDP) can bind, and the GEF can disassociate.

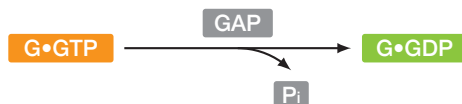
## 2.1 G proteins

G proteins (from *guanine nucleotide-binding proteins*, also sometimes known as *GTPases*) are a major family of intracellular signalling proteins that have wide and diverse roles. There are two major subcategories of G protein: the heterotrimeric G proteins that consist of triples of  $\alpha$ ,  $\beta$ , and  $\gamma$  subunits; and the small monomeric G proteins of the Ras superfamily. In general both monomeric G proteins and the  $\alpha$ -subunit of heterotrimeric G proteins contain a highly conserved (both in sequence and structure) nucleotide-binding pocket which is ordinarily filled by either GDP (guanosine diphosphate) or GTP (guanosine triphosphate) (Simon, Strathmann, & Gautam 1991; Valencia et al. 1991; Rojas et al. 2012).

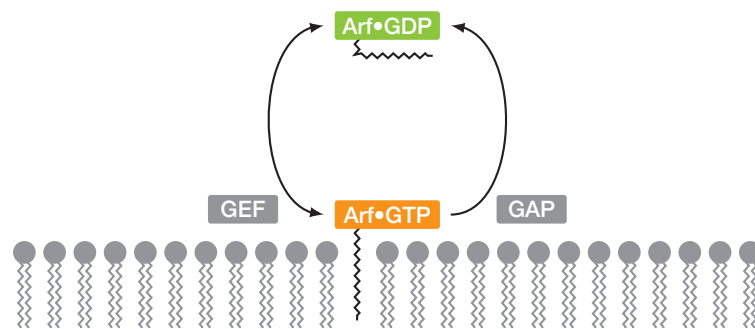
Specific structural changes enforced by the bound nucleotide cause the GDP- and GTP-bound forms of the G protein to have differential activities and be capable of changing their subcellular localisation. In this way G proteins are commonly referred to as *molecular switches* with the GDP-bound form typically referred to as the *inactive* or *off* state; and the GTP-bound form as the *active* or *on* state. This reflects that—in most systems—we are interested in the downstream signals that result from the activity of the GTP-bound form (Vetter & Wittinghofer 2001; Oldham & Hamm 2008).

Regulation of G protein activation state is primarily controlled through two competing mechanisms: *GTPase activity* and *nucleotide exchange*, illustrated in figure 2.1. The correct regulation of G proteins by these mechanisms is important. Incorrect regulation is known to cause disease, for instance: cancer (Young et al. 2009; Vigil et al. 2010; O’Hayre et al. 2013); cardiovascular disease (Loirand, Sauzeau, & Pacaud 2013); and genetic disorders (Seixas et al. 2013).

G proteins are inactivated by GTPase activity—the hydrolysis of the bound GTP molecule to form GDP. This process can either be entirely *intrinsic* to the G protein or catalysed in collaboration with external *GTPase activating proteins (GAPs)*.



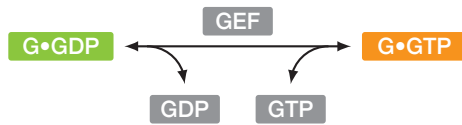
G proteins are also regulated by *nucleotide exchange* – the sequential disassociation and association of guanine nucleotides from the G protein, illustrated in figure 2.2, with the G protein passing through a intermediate, stable nucleotide-free state. This process is known to be completely reversible, but allows for the activation of the G protein if GDP is replaced by GTP. Nucleotide exchange is mediated by *guanine*



**Figure 2.3** The differential localisation of active and inactive Arf. Arf contains a myristoylated N-terminus which is capable of insertion into lipid membranes, thus tethering the protein to cytosol-facing membrane surfaces. This myristoyl group is largely hidden from the bulk solvent when GDP is bound, and so inactive Arf is ordinarily cytosolic. Only when Arf approaches the membrane, and becomes activated through nucleotide exchange by an Arf GEF, does the myristoyl group become exposed. Subsequent hydrolysis of the bound GTP causes the inactivation of the protein and so the loss of specific membrane-localisation.



*nucleotide exchange factors* (GEFs) for the small G proteins, and *G protein coupled receptors* (GPCRs) for the heterotrimeric G proteins.



## 2.2 Arf family G proteins

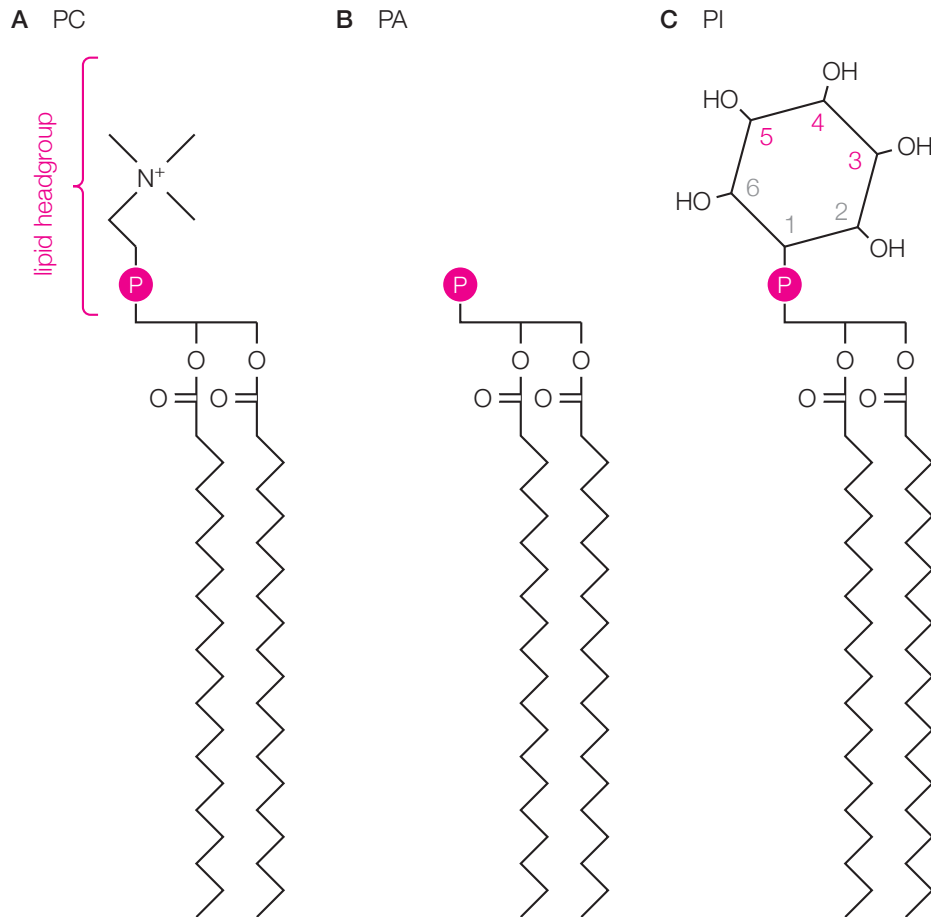
Arf G proteins (from *ADP ribosylation factor*, the activity for which Arfs were first identified) are members of the Ras superfamily of small monomeric G proteins (Schleifer et al. 1982; Kahn & Gilman 1984). Arfs (in their many isoforms) are ubiquitously expressed, with amino-acid sequences that are well conserved throughout eukaryotes (D'Souza-Schorey & Chavrier 2006).

There are six mammalian Arfs of three classes:

- class I contains Arf1, Arf2, and Arf3;
- class II contains Arf4, and Arf5;
- and class III contains Arf6 (Donaldson & Jackson 2011).

All Arfs contain a post-translational myristoylation on their N-terminus. This, along with the adjacent N-terminal amphipathic  $\alpha$ -subunit, is capable of insertion into lipid membranes, and so acts to tether the rest of the protein to the surface of the membrane (Liu, Kahn, & Prestegard 2009). This domain is only exposed when GTP is bound to Arf (Antonny et al. 1997). Hence the activation of Arf is correlated with specific changes in localisation: in general the inactive Arf is cytosolic and the active Arf is membrane-tethered. This relationship is illustrated in figure 2.3.

The regulation of Arf is typical for its classification as a monomeric G protein. It has no intrinsic GTPase activity, and so is reliant on inactivation via GAPs. Nucleotide exchange—thus their activation—is mediated by GEFs, which contain a conserved Sec7 domain, and which are themselves membrane-recruited (Kolanus 2007; Casanova 2007; Donaldson & Jackson 2011). Thus Arf activation and deactivation is a membrane associated process. Many different GAPs and GEFs exist, each more or less specific to different members of the family of Arfs and located in different sub-cellular compartments or expressed differentially in tissues. (Donaldson & Jackson 2011).



**Figure 2.4** Molecular structures of the discussed phospholipids. For illustration purposes, all fatty acyl chains are shown as fully saturated forms.

**A** Phosphatidylcholine (PC).

**B** Phosphatidic acid (PA).

**C** Phosphatidylinositol (PI). PI can be phosphorylated at positions 3, 4, and 5 on its inositol ring leading to eight combinatorial phosphatidylinositol phosphate derivatives, as shown in figure 2.5.

The incorrect regulation of Arfs has been implicated in a number of diseases, including: genetic diseases (Seixas et al. 2013); and viral replication (Bui, Golinelli-Cohen, & Jackson 2009).

In this thesis I will specifically refer to Arf1 and Arf6, structures for which are shown in figures 2.7A and 2.7B. These two Arf isoforms are structurally similar when GTP-bound but dissimilar when GDP-bound (Pasqualato et al. 2001). They are both known to regulate PLD and PI4P5K, though each is possibly selective to different isoforms of the enzymes (Perez-Mansilla et al. 2006).

## 2.3 Phospholipids

Phospholipids are a major subclass of lipid that are present as a large proportion of all cellular membranes. The family includes *phosphatidylethanolamine (PE)*, *phosphatidylserine (PS)*, *phosphatidylcholine (PC)*, *phosphatidic acid (PA)*, and *phosphatidylinositol (PI)*. Structurally, the hydrophobic portion of each of these molecules is *diacylglycerol (DAG)*, however each has a distinct lipid headgroup. In this thesis I will only need to consider PC, PA, and PI derivatives, the structures of which are shown in figure 2.4.

Phospholipids have multiple signalling and regulatory roles. They are an integral component of membranes and so contribute to the segregation of the cell interior from its surrounding environment and the compartmentalisation of cellular functions and activities within different organelles (van Meer, Voelker, & Feigenson 2008; van Meer & de Kroon 2011).

Different phospholipids have different geometry, and so are able to affect the physical properties of the membrane including its curvature (McMahon & Gallop 2005). In this way, phospholipids have an important role in the many crucial intracellular processes such as endocytosis and exocytosis.

The composition of phospholipids varies between different cellular membranes – they are dispersed inhomogeneously. As such, they are a major determinant of membrane identity and so can be used by intracellular processes to positively identify target membranes (Di Paolo & De Camilli 2006; van Meer, Voelker, & Feigenson 2008; van Meer & de Kroon 2011)

Many specific lipid binding domains have been observed within the structures of proteins. Fluctuations in the concentrations of phospholipids can therefore control the temporal regulation of the recruitment of proteins to membranes. Phospholipids are also capable of acting in concert with other membrane-recruitment processes,

increasing the affinity with which the target protein becomes membrane-associated. As a result, phospholipids can play a part in coincidence detection, or the integration of multiple signals from different sources (Di Paolo & De Camilli 2006).

A final—and critical—role for phospholipids is as mediators of signals in their own right. Particularly, the production of cytosolic second-messengers from their degradation, for instance *inositol trisphosphate* ( $IP_3$ ) from the degradation of phosphatidylinositol (4,5)-bisphosphate ( $PI(4,5)P_2$ ) by *phospholipase C* (*PLC*).

### 2.3.1 Phosphatidylcholine

Phosphatidylcholine (PC), shown in figure 2.4A, is the most abundant phospholipid in all cellular membranes, in some reaching up to and above 50% of total phospholipids. PC occupies an almost cylindrical volume as the dimensions of the choline headgroup are similar to the DAG base, and so PC can spontaneously combine to form flat lipid bilayers (van Meer, Voelker, & Feigenson 2008; van Meer & de Kroon 2011).

### 2.3.2 Phosphatidic acid

Phosphatidic acid (PA), shown in figure 2.4B, is the simplest phospholipid, but is present in much smaller proportions than PC. It can be formed by several reactions, including when the choline group is cleaved from PC (via the action of PLD, see section 2.4), or alternatively from the phosphorylation of DAG (via diacylglycerol kinases, see section 2.7.2) (Cai et al. 2009). PA produced by diacylglycerol kinases and by PLD can be distinguished by their distinct fatty acid compositions (Pettitt et al. 1997).

PA has many varied and crucial signalling roles (Wang et al. 2006; Stace & Ktistakis 2006), however unlike other phospholipids there appears to be no general conserved amino acid sequence that describes a binding domain for PA in proteins. However, PA has the potential of carrying more negative charge than other phospholipids, such as PI. It is likely therefore that electrostatic interactions have a major role in the recruitment of proteins to membranes by PA (Kooijman & Burger 2009).

Physically the head group of PA is smaller than its tail and so it occupies a conical volume providing an ability to form membranes with negative curvature (Wang et al. 2006).

### 2.3.3 Phosphoinositides

The family of phosphoinositides are generated from the phosphorylation of phosphatidylinositol (PI), shown in figure 2.4C, on the 3, 4, and 5 positions of the inositol ring on its head group. Every combination of phosphorylations is possible – leading to seven derivative phospholipids:

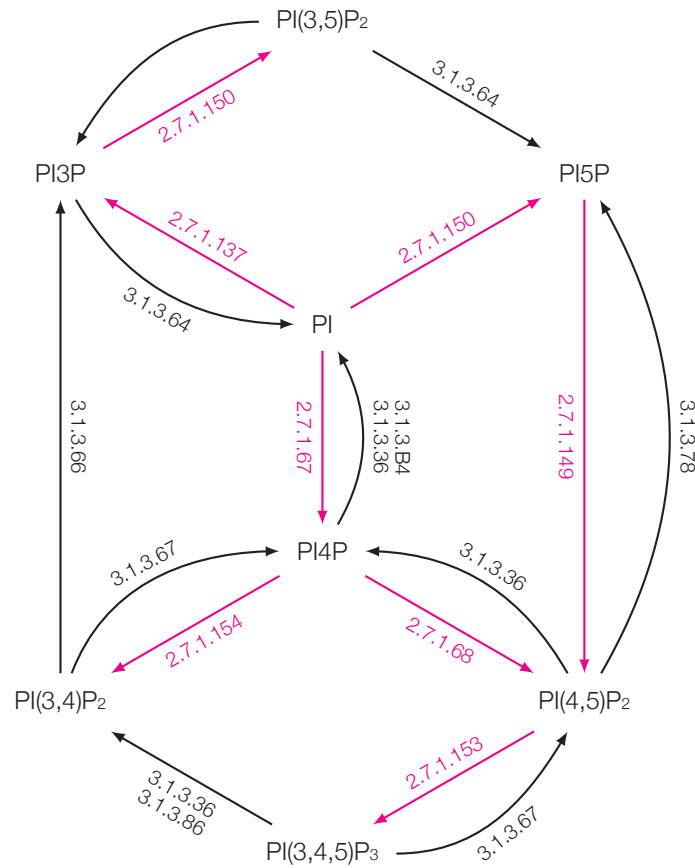
- phosphatidylinositol 3-phosphate (PI3P),
- phosphatidylinositol 4-phosphate (PI4P),
- phosphatidylinositol 5-phosphate (PI5P),
- phosphatidylinositol (3,4)-bisphosphate (PI(3,4)P<sub>2</sub>),
- phosphatidylinositol (3,5)-bisphosphate (PI(3,5)P<sub>2</sub>),
- phosphatidylinositol (4,5)-bisphosphate (PI(4,5)P<sub>2</sub>),
- and phosphatidylinositol (3,4,5)-trisphosphate (PI(3,4,5)P<sub>3</sub>).

PI constitutes approximately 15% of total phospholipids in the cell, the phosphorylated forms are generally less abundant by an order of magnitude, of which PI4P and PI(4,5)P<sub>2</sub> are the most abundant (Di Paolo & De Camilli 2006).

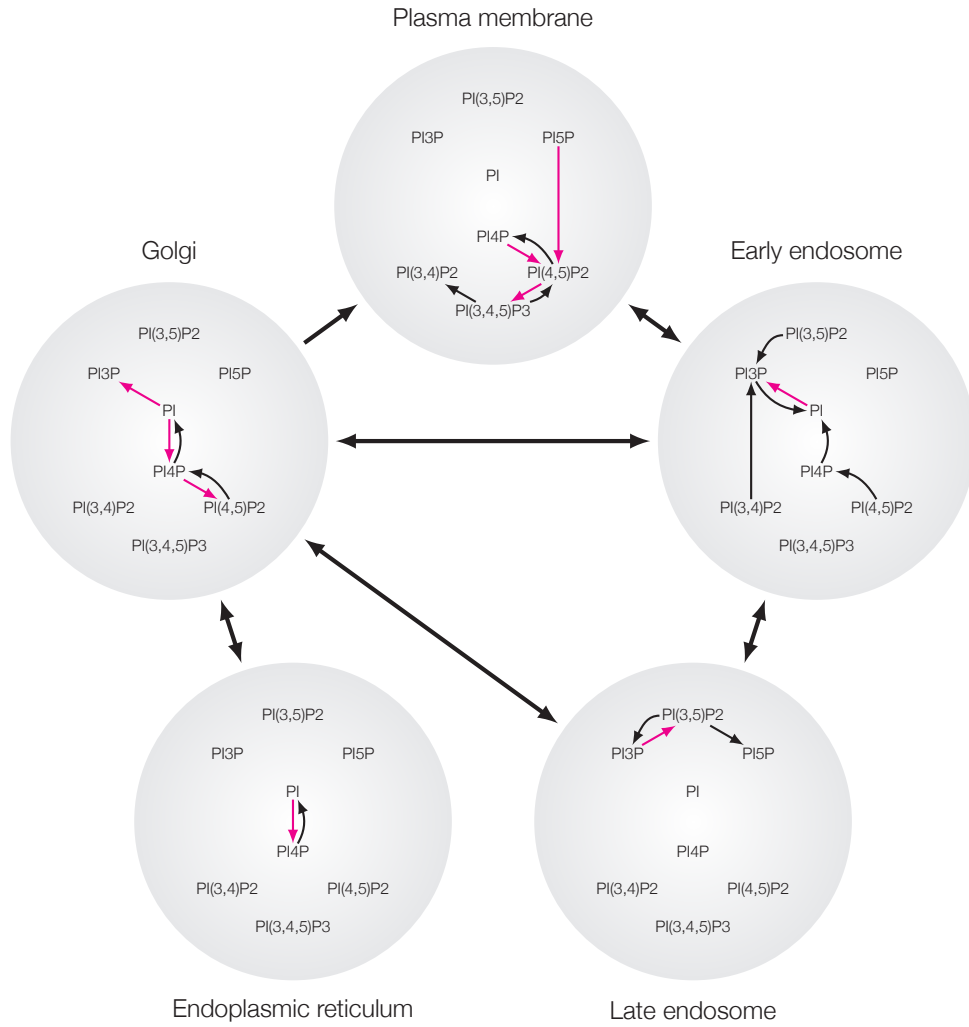
A variety of different kinases and phosphatases exist that interconvert the phosphoinositides, each adding or removing a single phosphate group (Sasaki et al. 2009; Jean & Kiger 2012). Not all of the possible conversions have been observed *in vivo*, only the subset shown in figure 2.5 appear to be physiologically relevant. Additionally, these are not all present in each intracellular compartment – only specific interconversions will occur in each, as shown in figure 2.6.

Each of these phosphoinositides has distinct signalling properties (Cauvin & Echard 2014). Many different conserved binding domains exist that correspond to one or more of the phosphoinositides (Di Paolo & De Camilli 2006). For instance, the pleckstrin homology (PH) domain, and the phox homology (PX) domain (Itoh & Takenawa 2002; Narayan & Lemmon 2006). By providing binding sites for proteins on the surface of the membrane (and so facilitating membrane-recruitment) the phosphoinositides are major components in many intracellular signalling processes, such as the regulation of the actin cytoskeleton (Yin & Janmey 2003), and membrane trafficking (Martin 2001).

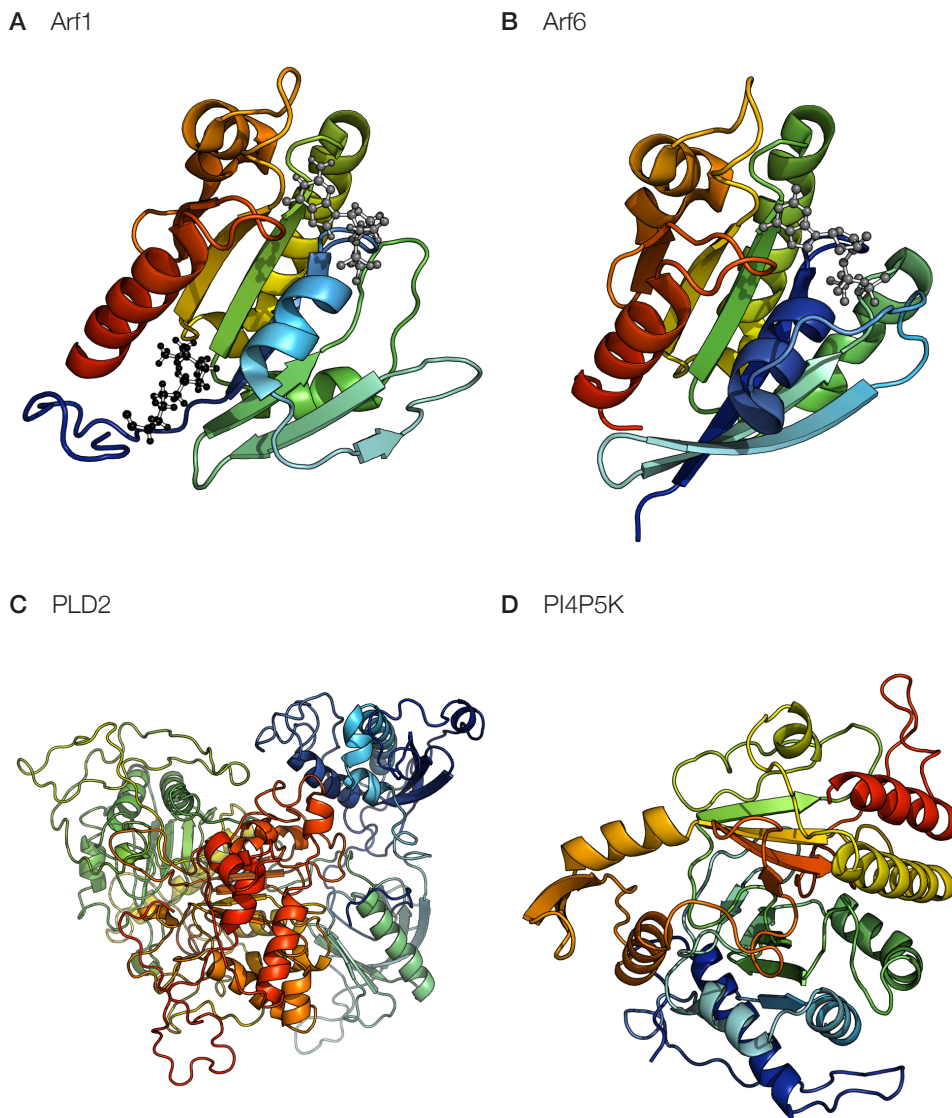
The phosphoinositides can also directly affect the curvature of the membrane (Rusinova et al. 2013).



**Figure 2.5** There are eight possible phosphoinositides, formed by the combinatorial phosphorylation of phosphatidylinositol (PI) on the 3, 4, and 5 positions on the inositol head group. Interconversion between the phosphoinositides occurs via kinases (drawn in magenta) and phosphatases (drawn in black) which are capable of adding or removing a single phosphate group, respectively. Not all of the possible interconversions are known to occur *in vitro*, here I have plotted those kinases and phosphatases in the BRENDA database (Schomburg et al. 2013) with their associated EC Numbers. Additionally, phosphatidylinositol (3,5)-bisphosphate 5-phosphatase does not appear to be listed in the BRENDA database but is known to exist (Kong et al. 2000; Duex et al. 2006; Cauvin & Echard 2014). Other sources suggest slightly different allowable interconversions (Di Paolo & De Camilli 2006; Jean & Kiger 2012).



**Figure 2.6** Subset of the routes for phosphorylation and dephosphorylation (from the full graph in figure 2.5) that are present in each cellular membrane. Modified from Figure 1c in Di Paolo & De Camilli (2006).



**Figure 2.7** Structures for:

**A** Myristoylated yeast Arf1-GDP, bound nucleotide drawn in grey; myristoylation drawn in black, PDB\* ID: 2K5U (Liu, Kahn, & Prestegard 2009).

**B** Human Arf6-GTP, bound nucleotide drawn in grey, PDB\* ID: 2J5X (Pasqualato et al. 2001).

**C** Phospholipase D2 (Mahankali, Alter, & Gomez-Cambronero 2014; personal communication).

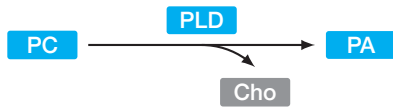
**D** Phosphatidylinositol 4-phosphate 5-kinase, predicted structure using SWISS-MODEL (Arnold et al. 2006; Guex, Peitsch, & Schwede 2009; Kiefer et al. 2009; Biasini et al. 2014) using a known structure of PI5P4K $\beta$ , PDB\* ID: 1B01 (Rao et al. 1998).

\* Protein database ID (Berman et al. 2000).



## 2.4 Phospholipase D

*Phospholipase D (PLD)* is an enzyme which catalyses the hydrolysis of phosphatidylcholine (PC) to produce phosphatidic acid (PA) and choline (Cho) (Hanahan & Chaikoff 1948).



Enzymes with phospholipase D activity (EC number 3.1.4.4) have been found in eukaryotes, prokaryotes, and viruses. They can be divided into two classes: those with one or more HKD domains, conserved amino acid sequence containing a well-conserved histine (H), lysine (K), and aspartic acid (D) which are responsible for their catalytic activity; and those with an alternative mechanism (Selvy et al. 2011).

PLDs are required for the healthy maintenance of membranes (Gomez-Cambronero 2014) and the regulation of the cytoskeleton (Rudge & Wakelam 2009). They have also been implicated in disease (Peng & Frohman 2012), including: cancer (Su, Chen, & Frohman 2009); fetal alcohol syndrome; Alzheimer's disease (Klein 2005; Burkhardt et al. 2014); and defective platelet aggregation (Elvers et al. 2012).

There are six known human PLDs: PLD1, PLD2, PLD3, PLD4, PLD5, and PLD6. The best characterised are PLD1 (Hammond et al. 1995; Hammond et al. 1997; Sung et al. 1999b) and PLD2 (Colley et al. 1997; Lopez, Arnold, & Lambeth 1998; Sung et al. 1999a); the existence of PLDs 3–6 has only recently been discovered (Gomez-Cambronero 2014). PLD1–5 contain two HKD domains, while PLD6 only contains a single HKD domain (Gomez-Cambronero 2014). In this thesis I will only consider PLD1 and PLD2, due to their observed regulation by Arfs.

PLD1 and PLD2 have approximately 50% sequence identity (Gomez-Cambronero 2014). They are the only isoforms to contain a PH (pleckstrin homology) domain, a PX (phox homology) domain, and a phosphoinositide binding motif (Sciorra et al. 1999; Liscovitch et al. 2000; Hodgkin et al. 2000). These domains contribute to their recruitment to lipid membranes, specifically: the Golgi apparatus (Freyberg et al. 2001; Freyberg 2002; Riebeling, Morris, & Shields 2009); and the plasma membrane (during cellular stimulation for PLD1) (Colley et al. 1997; Morgan et al. 1997; Brown et al. 1998).

There are known differences in the regulation of PLD1 and PLD2. First of all—PLD2 has a greater basal activity than PLD1. This is known to be due to regulatory

regions unique to each enzyme which decrease the activity of PLD1, and increase the activity of PLD2 (Sung et al. 1999b; Sung et al. 1999a).

Both are activated by both Arf1 and Arf6. However PLD1 appears to be preferentially activated by Arf1 (Hammond et al. 1995; Hammond et al. 1997; Perez-Mansilla et al. 2006) and PLD2 by Arf6 (Hiroyama & Exton 2005). Additionally, the specific mechanism by which this activation occurs is still unknown. This is possibly due to the fact that a potential structure (for human PLD2) has only recently been reported (Mahankali, Alter, & Gomez-Cambronero 2014), see figure 2.7C, and this computational model is the current limit of structural insight. However, a region on Arf1 is known to be essential for its interaction with PLD1 (Liang et al. 1997).

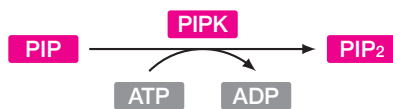
It is known that the activation of PLD by Arf is absolutely dependent on the presence of PI(4,5)P<sub>2</sub> – the product of PI4P5K (see section 2.5 below). That is, PI(4,5)P<sub>2</sub> is a *cofactor* for the Arf-activation of PLD (Liscovitch et al. 1994; Pertile et al. 1995; Divecha et al. 2000). PI(4,5)P<sub>2</sub> is capable of interacting with both PLD1 and PLD2 via their PH domain (Hodgkin et al. 2000). This interaction also appears to have a role in the membrane localisation of PLD (Sciorra 2002; Du 2003).

Other activators of PLD are: protein kinase C (Chen & Exton 2004; Henage, Exton, & Brown 2006); other monomeric G proteins (Malcolm, Elliott, & Exten 1996; Jiang et al. 1995; Zhang & Du 2009); and (PLD2) by oleic acid (Sarri et al. 2003).

PLD has also recently been categorised as having GEF activity for the monomeric G proteins Rac2 and RhoA (Mahankali et al. 2011; Jeon et al. 2011; Mahankali et al. 2012).

## 2.5 Phosphatidylinositol 4-phosphate 5-kinase

Phosphatidylinositol 4-phosphate 5-kinase (PI4P5K) is an enzyme which phosphorylates PI4P to form PI(4,5)P<sub>2</sub> (Van Dongen, Zwiers, & Gispen 1984; Cochet & Chambaz 1986; Ling, Schulz, & Cantley 1989).



PI4P5Ks were previously categorised as either type I or type II, but the type IIs are now known to be phosphatidylinositol 5-phosphate 4-kinases (PI5P4Ks) (Rameh et al. 1997). There are three mammalian PI4P5K isoforms: I $\alpha$ , I $\beta$ , and I $\gamma$  (Loijens & Anderson 1996; Ishihara et al. 1996; Ishihara et al. 1998). The human and mouse

terminology is reversed for I $\alpha$  and I $\beta$ . (I will use the mouse terminology.) No other kinases are similar to PI4P5Ks and PI5P4Ks (Oude Weernink, Schmidt, & Jakobs 2004). No structure for PI4P5K has been reported, however it is possible to use known structures of PI5P4Ks (based on their sequence similarity) as a template, see figure 2.7D.

All three isoforms of PI4P5K have a conserved kinase core domain (Ishihara et al. 1998), and an activation loop that controls substrate specificity and localisation (Kunz et al. 2000).

PI4P5Ks are ordinarily cytosolic, and are recruited by Arf to the plasma membrane (Honda et al. 1999) and the Golgi (Godi et al. 1999). (This can be observed by the presence of these reaction routes in figure 2.6.) Arf1 and Arf6 have both been reported as activators of PI4P5K (Honda et al. 1999; Jones et al. 2000; Skippen et al. 2002), and it has been proposed that the mechanism of activation is via membrane-recruitment (Perez-Mansilla et al. 2006). Phosphorylation of PI4P5K may also be an important component in its regulation (Itoh 2000; Park, Itoh, & Takenawa 2001; Funakoshi, Hasegawa, & Kanaho 2011).

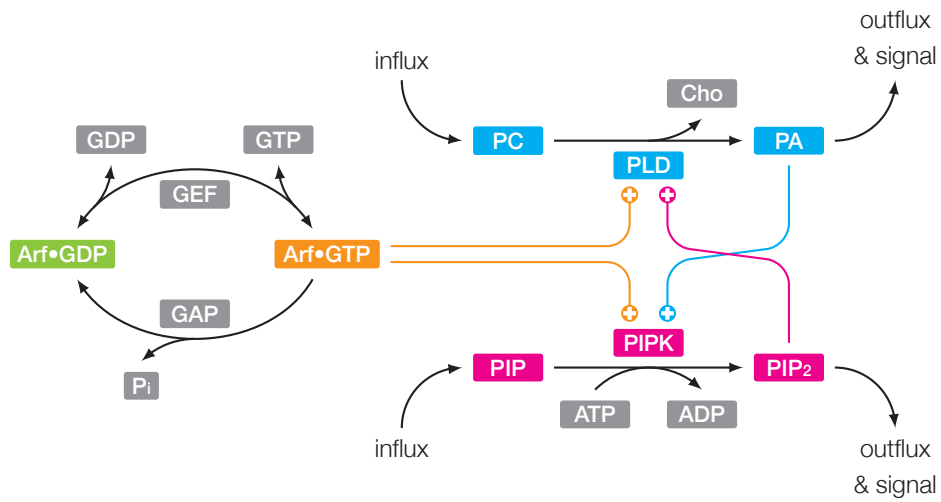
PA, the product of PLDs (see section 2.4 above), is also able to activate PI4P5K (Moritz et al. 1992; Jenkins, Fiset, & Anderson 1994; Ishihara et al. 1998; Cockcroft 2009). There is some evidence that suggests a strict requirement for the presence PA for PI4P5K activity in some circumstances (Honda et al. 1999), but this does not appear to be the majority view. No single region of PI4P5K is correlated with the ability to bind PA. Instead this is likely due to electrostatic and hydrophobic interactions on specific regions of the 3-dimensional surface of the enzyme (Jarquin-Pardo et al. 2007; Stace et al. 2008).

Arf and PA appear to have at least an additive—and possibly synergistic—effect on the activation of PI4P5K (Honda et al. 1999; Cockcroft 2009). Product inhibition by PI(4,5)P<sub>2</sub> has also been observed (Ling, Schulz, & Cantley 1989; Moritz et al. 1992).

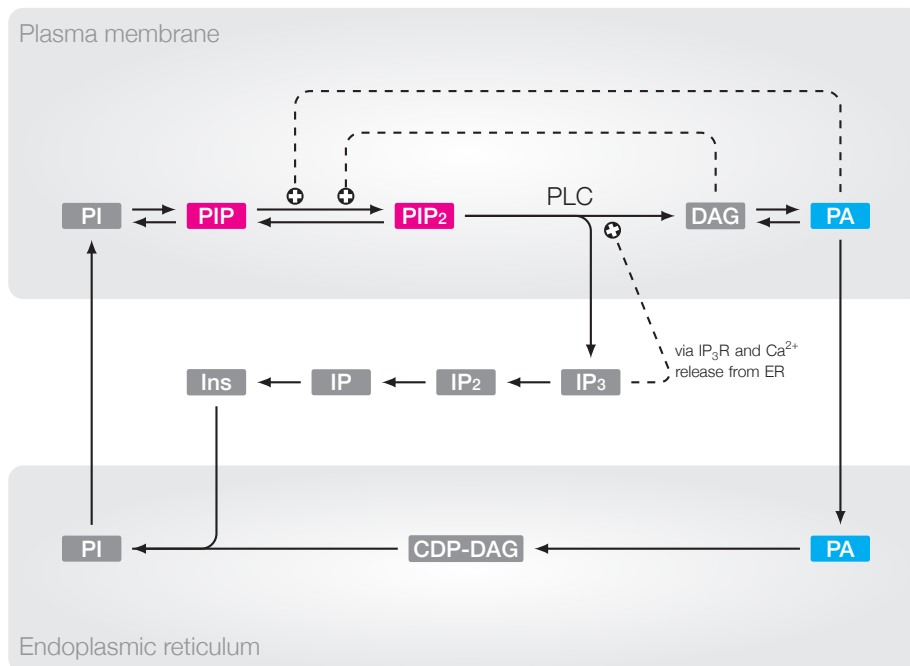
## 2.6 The Arf/PLD/PI4P5K phospholipid signalling motif

Together, Arf, PLD, and PI4P5K form a triple of proteins connected in a signalling motif that is known to be important in regulating many intracellular processes. This motif is illustrated in figure 2.8.

From the above discussion it should be evident that each component of this motif is complex in its own right, and that the behaviour and mechanism of each is distinct. While there is some superficial symmetry in the illustration in figure 2.8 between PLD



**Figure 2.8** The complete Arf/PLD/PI4P5K phospholipid signalling motif. The motif comprises of the activation of phospholipase D (PLD) and phosphatidylinositol 4-phosphate 5-kinase (PI4P5K) by active Arf family G proteins, and the cross-talk of the enzymes' products.



**Figure 2.9** The PI cycle formed by the breakdown of PI(4,5)P<sub>2</sub> by phospholipase C (PLC) and the re-synthesis of phosphatidylinositol (PI).

and PI4P5K, this cannot be said to reflect biochemical reality. For instance PLD and PI4P5K have different mechanisms of activation: PLD requires both PI(4,5)P<sub>2</sub> and Arf; PI4P5K is independently activated by PA and Arf. Therefore, distinct mathematical models will be needed for each enzyme.

The motif appears to be present at both the Golgi and the plasma membrane, though constructed from different triples of isoforms: Arf1, PLD1, PI4P5K $\alpha$  at the Golgi; and Arf6, PLD2, PI4P5K $\gamma$  at the plasma membrane (Perez-Mansilla et al. 2006).

## 2.7 Related pathways

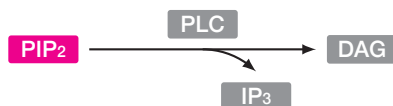
A number of other important pathways interact with one or more of the components of the Arf/PLD/PI4P5K phospholipid signalling motif. This section is not intended to be an exhaustive list of these interactions—the complexity of intracellular signalling is such that this would be an almost impossible task—but to comprehensively demonstrate that important interactions exist.

### 2.7.1 Regulation of Arf by phospholipids

There is evidence for regulation of ARF GAPs by PI(4,5)P<sub>2</sub> and PA (Randazzo & Kahn 1994; Randazzo 1997) and ARF GEFs by PI(4,5)P<sub>2</sub> (Terui, Kahn, & Randazzo 1994; Paris et al. 1997). This would constitute a negative or positive feedback within the system, but will not be included within the models in this thesis.

### 2.7.2 The PI cycle

Phosphoinositides are tightly regulated by the PI cycle. Phospholipase C (PLC) cleaves PI(4,5)P<sub>2</sub> into diacylglycerol (DAG) and inositol (1,4,5) trisphosphate (IP<sub>3</sub>).



IP<sub>3</sub> is a cytosolic second messenger, and is capable of interacting with intracellular IP<sub>3</sub> receptors (IP<sub>3</sub>R) and so influence calcium signalling (Foskett et al. 2007). DAG remains as part of the membrane, and is converted into PA by specific kinases. Both DAG (GMH Thomas and G Walsh, unpublished results) and this PA (Cockcroft 2009) are capable of activating PI4P5K.

Independently, the metabolites of DAG and IP<sub>3</sub> are transported or diffuse (respectively) to the endoplasmic reticulum where they can combine to reform PI. This PI can then relocate to the plasma membrane giving rise to the PI cycle, illustrated in full in figure 2.9.

### 2.7.3 The PI3K/AKT/mTOR pathway

PI(4,5)P<sub>2</sub> is the substrate for phosphatidylinositol (4,5)-bisphosphate 3-kinase (PI3K), a component part of the PI3K/AKT/mTOR pathway highly implicated in cancer (Ocana et al. 2014).

## 2.8 Summary

In this chapter, I have summarised much of the current understanding regarding the phospholipids (PC, PA, PIs) and the proteins (Arf, PLD, PI4P5K) that will be discussed and investigated in the following chapters.

Together, these form the component parts of the Arf/PLD/PI4P5K phospholipid signalling motif. Different triples of isoforms of the three proteins have been suggested as operating at the Golgi and at the plasma membrane (Perez-Mansilla et al. 2006). To simplify discussion, I will in general consider the motif in an abstract sense – one isoform each of Arf, PLD, and PI4P5K operating on a lipid membrane containing PC, PA, PI4P, PI(4,5)P<sub>2</sub>, and other unspecified (non-interacting) lipids.

### 3 Methods and derivations

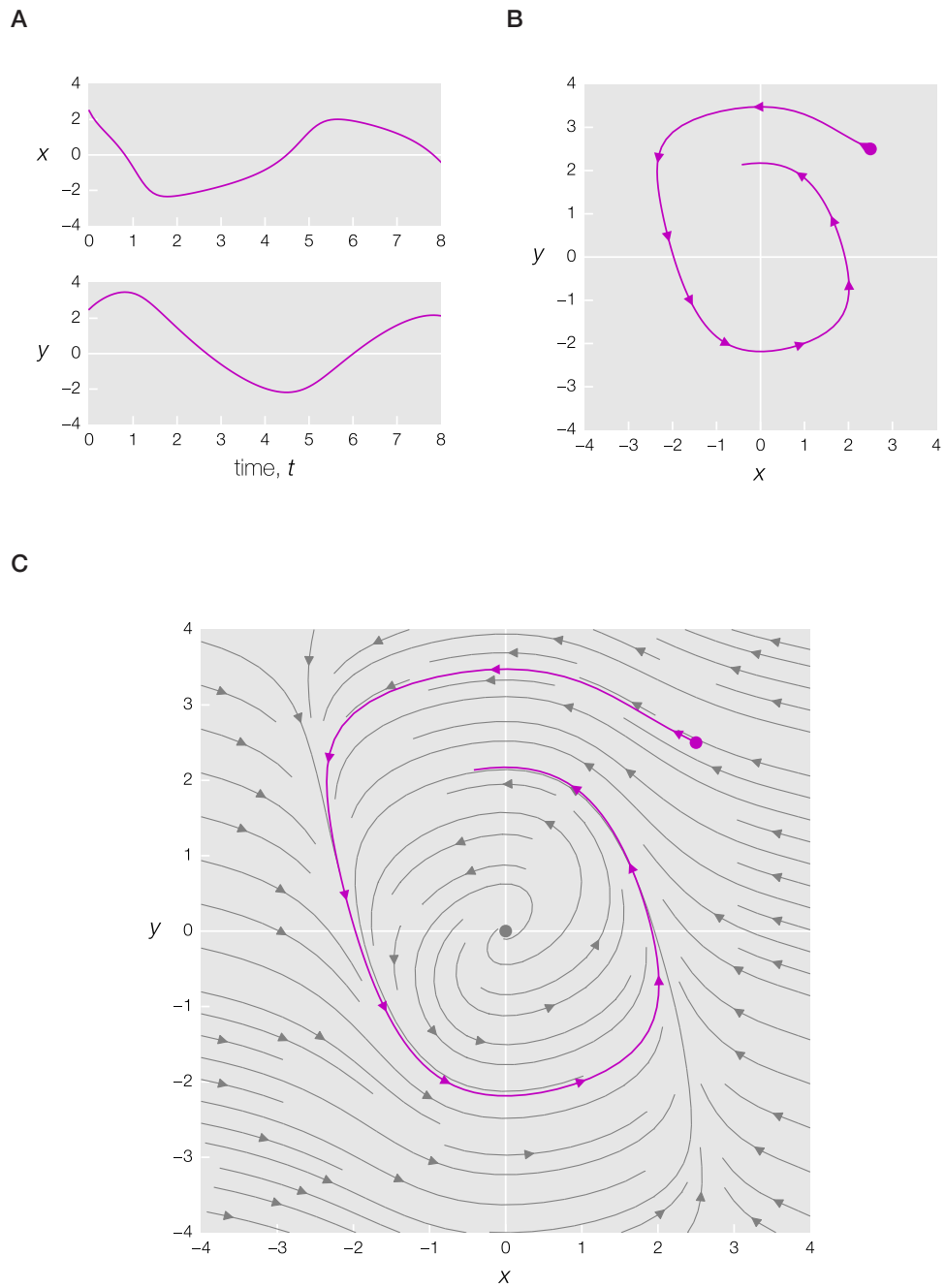
Most biological systems, studied at most scales, are highly time-dependent. Such systems—and their component parts—can be described in terms of *rates of change* which can be written mathematically as *differential equations*. Here, I will consider only *ordinary* differential equations (ODEs), rather than *stochastic* or *partial* differential equations. I will focus on ODEs largely because the law of mass-action—which will be my base assumption for models in this thesis—generates systems of ODEs. But also because ODE models are simple to simulate and analyse, using methods that I will describe in this chapter.

In this chapter I will also discuss methods of reducing the complexity of models that were built using the law of mass-action through application of time-scale separation and the quasi-steady-state approximation. Furthermore, I will describe the semi-automation of the derivation of these models using a new module written for the Python programming language.

Finally, I will introduce a method for modelling species that undergo cytosol-membrane transitions, as this will be crucial in being able to construct models of PI4P5K and Arf.

The majority of code included in this thesis has been written and tested for version 3.5 of the Python programming language, where appropriate with the packages: `numpy` (version 1.10.1) (van der Walt, Colbert, & Varoquaux 2011), `scipy` (version 0.16.1), `sympy` (version 0.7.6.1), and `matplotlib` (version 1.5.0) (Hunter 2007).

For additional background information I refer the reader to the following texts: Strogatz (2014) for mathematical methods related to analysis of nonlinear systems of ordinary differential equations; and Cornish-Bowden (2012) for in-depth discussion of chemical kinetics.



**Figure 3.1** Plots of the dynamics of the Van der Pol oscillator, equation 3.1.

**A** Simulation,  $x$  and  $y$  against time  $t$ .

**B** A single trajectory,  $x$  against  $y$ . The initial condition has been marked with a circle. Time is now implicit and progresses in the direction of the marked arrows.

**C** Phase plane, all trajectories of the system (in the plotted region). There is a single steady-state at the origin.



## 3.1 Methods for ordinary differential equations (ODEs)

There are several basic techniques I will use in the analysis of my mathematical models, including: simulation; phase plane analysis; steady-state analysis; and determination of the stability of steady-states. I will briefly recap these methods here.

### 3.1.1 Simulation

Any system of ODEs can be numerically simulated *in silico*. For example, for illustrative purposes, the Van Der Pol oscillator is described by the equations

$$\begin{aligned}\frac{dx}{dt} &= \mu \left( x - \frac{x^3}{3} - y \right) \\ \frac{dy}{dt} &= \frac{x}{\mu}.\end{aligned}\tag{3.1}$$

where  $x, y$  are variables; and  $\mu$  is a parameter, which needs to be set to some value for the duration of an individual simulation. Here I will choose  $\mu = 1$ .

This system of ODEs can be simulated in the Python programming language with a script that uses the `scipy.integrate.odeint` function. We also need to specify the time-period (e.g.  $0 \leq t \leq 8$ ), and the initial conditions (e.g.  $x = 2.5, y = 2.5$  at  $t = 0$ ).

```

1 # setup environment
2 import numpy as np
3 from scipy.integrate import odeint
4
5 # Van der Pol oscillator as a function
6 def vanderpol(xy, t, mu = 1):
7     x, y = xy
8     dx = mu*(x - x**3/3.0 - y)
9     dy = x/mu
10    return (dx, dy)
11
12 t = np.linspace(0, 8, 100) # 100 equally spaced time points
13 ic = [2.5, 2.5]           # initial conditions
14
15 # use odeint to simulate the system
16 x, y = zip(*odeint(vanderpol, ic, t))

```

The output of this script can then be plotted, for example with  $x$  and  $y$  as a function of time  $t$  as in figure 3.1A.

### 3.1.2 Phase plane analysis

An alternative plot of the dynamics of a system can be obtained by plotting the dependent variables  $x$  and  $y$  against each other, rather than against time. For example, the simulation of the Van der Pol equation, shown in figure 3.1A, can be plotted with  $y$  as a function of  $x$ , to give figure 3.1B. This plot now describes a *trajectory* through the *phase plane* of the system. Time is now implicit and increases as this trajectory is followed.

For two dimensional systems it is possible to extend this idea further. If the value of the differential equations (that describe the system of ODEs) are found at each point on the plane, we can draw all of the possible trajectories of the system. (In practical terms we find the values on a finite, discrete grid.) This results in the *phase plane portrait* of the system, shown in figure 3.1C. Now we can follow the time-evolution of all initial conditions (within the plotted region) just by following their trajectory through the plane, and also can determine certain characteristics of the dynamics of the system as a whole. For example, we can observe that the Van der Pol oscillator displays a stable oscillation (for which it is named).

### 3.1.3 Steady-states

Many physiologically relevant models have the property that any initial condition will eventually tend towards one of a finite number of *steady-states* – states where the dynamics of the system will no longer change with respect to increasing time. If  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  is a vector of our system variables, then the steady-states of the system

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

are the solutions to the equations

$$f(\mathbf{x}) = 0.$$

For the Van der Pol oscillator, setting equation 3.1 equal to zero gives a single steady-state at the origin  $(0, 0)$ .

### 3.1.4 Stability of steady-states

Steady-states can be either *stable* or *unstable*, as illustrated in figure 3.2. A stable steady-state is able to *attract* a region surrounding itself such that all trajectories in

this region tend towards the steady-state. This does not hold for unstable steady-states where, in general, a small perturbation will create a trajectory that tends away from the steady-state. For this reason, unstable steady-states tend not to be experimentally observable, and correspond to particularly non-robust states.

The stability of a steady-state can be determined by the sign of the *eigenvalues* of the *Jacobian* at that point. Given a two-dimensional system

$$\begin{aligned}\frac{dx}{dt} &= f_x(x, y) \\ \frac{dy}{dt} &= f_y(x, y)\end{aligned}\tag{3.2}$$

then the *Jacobian matrix* is given by

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f_x}{\partial x} & \frac{\partial f_x}{\partial y} \\ \frac{\partial f_y}{\partial x} & \frac{\partial f_y}{\partial y} \end{bmatrix}\tag{3.3}$$

and the two eigenvalues,  $\lambda$ , are the solutions to the equation

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0\tag{3.4}$$

where  $\det$  is the determinant function and  $\mathbf{I}$  is the identity matrix.

The steady-state is a *stable node* if both eigenvalues are negative, an *unstable node* if both are positive, and a (unstable) *saddle node* if one is positive and one is negative, illustrated in figure 3.2. If the sign of an eigenvalue changes as a function of one or more of the parameters, then at the critical value the system is said to undergo a *bifurcation*. These ideas extend to three or more dimensions, but only two dimensional systems will be analysed in this thesis.

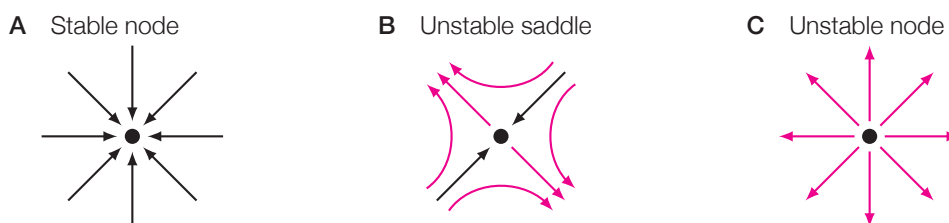
## 3.2 Mass-action models

Since the late 19<sup>th</sup> century there has been a well-accepted method for the translation of biochemical mechanisms into mathematics called the *law of mass-action*.

### 3.2.1 Notation

Here, and where appropriate through the rest of the thesis, I will use the following notation. The biochemical systems I will consider consist of:

- biochemical *species* ( $\mathcal{S}$ );



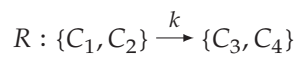
**Figure 3.2** Stability classes of two dimensional steady-states, illustrated by the nearby trajectories.

Reaction type	Scheme	ODEs	Dim. of $k$
conversion	$A \xrightarrow{k} B$	$\frac{d[A]}{dt} = -k[A]$ $\frac{d[B]}{dt} = k[A]$	$s^{-1}$
formation	$A + B \xrightarrow{k} C$	$\frac{d[A]}{dt} = -k[A][B]$ $\frac{d[B]}{dt} = -k[A][B]$ $\frac{d[C]}{dt} = k[A][B]$	$M^{-1} s^{-1}$
degradation	$A \xrightarrow{k} B + C$	$\frac{d[A]}{dt} = -k[A]$ $\frac{d[B]}{dt} = k[A]$ $\frac{d[C]}{dt} = k[A]$	$s^{-1}$
constant source	$\emptyset \xrightarrow{k} A$	$\frac{d[A]}{dt} = k$	$M s^{-1}$
constant sink	$A \xrightarrow{k} \emptyset$	$\frac{d[A]}{dt} = -k[A]$	$s^{-1}$

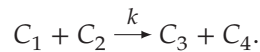
**Table 3.1** ODEs generated by application of the law of mass-action for the five common reaction schemes introduced in section 3.2.1. For each scheme the rate constant,  $k$ , will have a different dimension. Units:  $s$ , seconds;  $M$ , molar concentration.

- *complexes* ( $\mathcal{C}$ ), where each complex  $C \in \mathcal{C}$  is a subset of the species  $C \subset \mathcal{S}$  including the trivial complexes consisting of a single species  $\{S\} \equiv S$  for  $S \in \mathcal{S}$  (I will not need to allow for complexes that contain multiples of the same species);
- and, mono- or bi-molecular *reactions* ( $\mathcal{R}$ ).

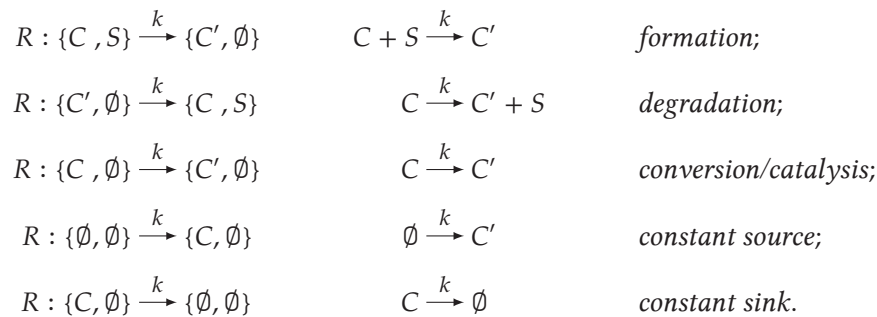
In general, a reaction  $R \in \mathcal{R}$  can be described as a function converting one pair of complexes (possibly including the empty complex,  $\emptyset$ ) into another pair at some rate  $k$ . That is,



or written in standard chemical notation



Most commonly, reactions will be one of:



For formation and degradation we have:

$$C \cup S = C'$$

that is, the species on the left hand side equal the species on the right hand side of the reaction.

Constant sources and sinks will be required when considering entry and exit points of mass (respectively) into and out of the system.

Species will be denoted by either:

- a specific name (e.g. Arf);
- a single capital letter (e.g. A);
- or, the graphical symbol listed in the list of abbreviations (e.g. Arf).

Complexes (e.g.  $C = \{S_1, S_2, \dots, S_n\}$ ) will be denoted by either:

- a single capital letter (e.g.  $C$ );
- the (dot) product of the component species (e.g.  $C \equiv S_1 \cdot S_2 \cdot \dots \cdot S_n$ );
- or, the product omitting the dots (e.g.  $C \equiv S_1 S_2 \dots S_n$ ).

For *enzyme complexes*, complexes that contain an enzyme, then the enzyme will tend to be listed first. I will sometimes denote an enzyme-complex with the additional notation:

- the enzyme (e.g.  $E$ ) with the remainder of the species in the complex ( $\sigma = C \setminus \{E\}$ ) as a lowercase subscript (e.g. if  $C \equiv ES_1 S_2 \dots S_n$  then  $C \equiv E_\sigma \equiv E_{s_1 s_2 \dots s_n}$ ).

The notation  $A \setminus B$  gives the elements of the set  $A$  removing any elements in the set  $B$ .

Specific notation will also be used for the rate constants for reactions forming or degrading enzyme complexes by a single species:



Empty sets will be dropped, for example:  $E_\emptyset \equiv E$ ,  $k_{\emptyset;s} \equiv k_s$ , and  $k_{\emptyset;-s} \equiv k_{-s}$ .

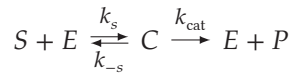
As per convention, the *concentration per unit volume* of a complex,  $C$ , will be denoted using square brackets,  $[C]$  (this has units  $\text{mol dm}^{-3} = \text{mol l}^{-1} = \text{M}$ ).

### 3.2.2 The law of mass-action

The law of mass-action is a proposition that states that the rate of a reaction is proportional to the concentrations of the molecular species that participate in that reaction. It is a continuous approximation of the discrete, stochastic dynamics that occur at the scale of individual molecular interactions. For the five types of reaction listed in section 3.2.1 the law of mass-action can be applied to give the systems of ODE listed in table 3.1.

Equations for more complicated reaction schemes can be generated by the summation of the rates given by individual reactions. For example, the classical Michaelis-Menten (Michaelis & Menten 1913; Johnson & Goody 2011) system describes an

enzyme that converts a substrate into a product via the reaction scheme



for which the law of mass-action gives the equations

$$\frac{d[E]}{dt} = -k_s[E][S] + (k_{-s} + k_{\text{cat}})[C] \quad (3.5)$$

$$\frac{d[C]}{dt} = k_s[E][S] - (k_{-s} + k_{\text{cat}})[C] \quad (3.6)$$

$$\frac{d[S]}{dt} = -k_s[E][S] + k_{-s}[C] \quad (3.7)$$

$$\frac{d[P]}{dt} = k_{-s}[C]. \quad (3.8)$$

Equations 3.5 to 3.8 are nonlinear as they contain multiples of variables of order at least two. Given a set of parameters ( $k_s$ ,  $k_{-s}$ , and  $k_{\text{cat}}$ ) and a set of initial conditions ( $[E]$ ,  $[C]$ ,  $[S]$ , and  $[P]$  at  $t = 0$ ), it is possible to simulate these. For example, with the following script written in the Python programming language:

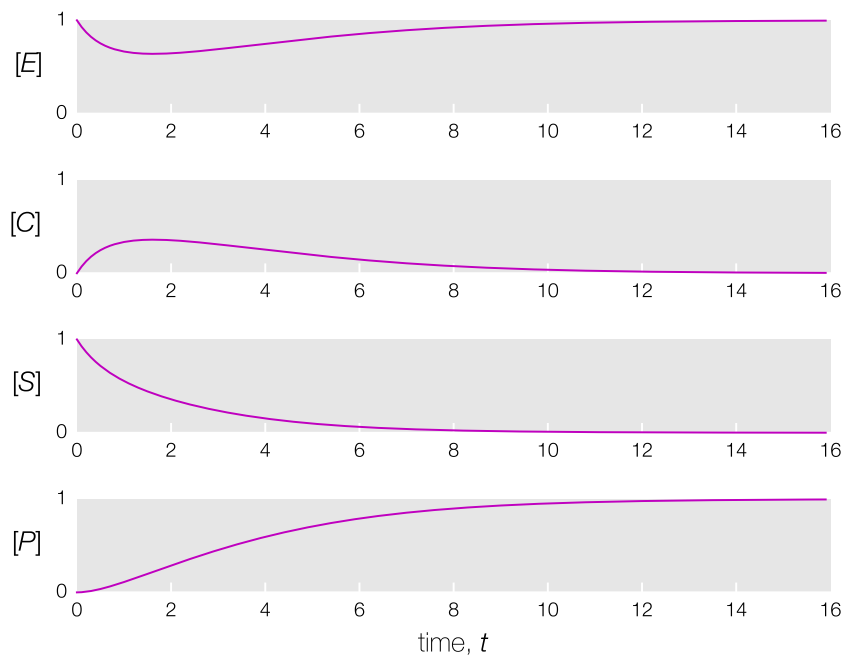
```

1 from scipy.integrate import odeint
2 import numpy as np
3
4 def michaelis_menten(x, t, k1=0.8, k2=0.5, km1=0.1):
5     e, c, s, p = x
6
7     de = -k1*s*e + (km1 + k2)*c
8     dc = k1*s*e - (km1 + k2)*c
9     ds = -k1*s*e + km1*c
10    dp = k2*c
11
12    return [de, dc, ds, dp]
13
14 t = np.arange(0, 16, 0.1)
15 ic = [1, 0, 1, 0]
16
17 x = zip(*odeint(michaelis_menten, ic, t))

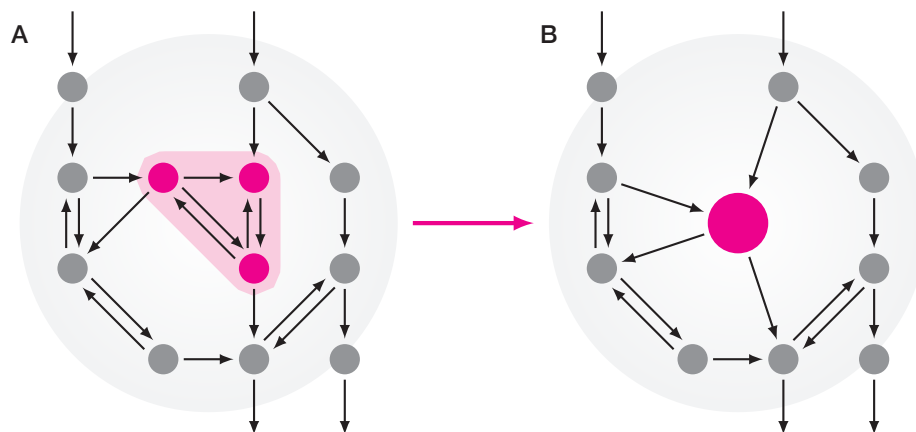
```

The results of this simulation are plotted in figure 3.3.

The law of mass-action applied to a biochemical system will give an equation for each species and complex in the system. Although it should be noted that some of these equations may be linear combinations of others. Often, we would like to simplify models generated by the law of mass-action, to reduce the number of equations (possibly with a resulting complexity cost).



**Figure 3.3** Simulation of the Michaelis-Menten system with the ODEs in equations 3.5 to 3.8 with  $k_s = 0.8$ ,  $k_{-s} = 0.1$ ,  $k_{-c} = 0.5$ , and initial conditions  $[E] = 1$ ,  $[C] = 0$ ,  $[S] = 1$ ,  $[P] = 0$ . Here  $E$  is the free enzyme;  $C$  is the enzyme-substrate complex;  $S$  is the substrate; and,  $P$  is the product.



**Figure 3.4** Illustration of the concept of time-scale separation.

**A** The highlighted region and complexes are assumed to operate on a much faster time scale than the rest of the system and so the transient dynamics of the fast system are assumed to complete before the slow surrounding processes can react.

**B** Then if we are only interested in the behaviour of the system at the slower time-scale we can simplify the model of the fast sub-system.



### 3.2.3 Time-scale separation

The idea behind *time-scale separation* is that different biological processes—even those involved within a single system—can operate at different time-scales, possibly across multiple orders of magnitude. Assuming that we are interested in the dynamics of the system at slow time-scales, then we are only really interested in the properties of the fast components of the system that become evident over the long-term. Thus, we can ignore any *transient* effects that arise from the fast components. We can use this fact to simplify the system by—in some sense—reducing the complexity of the fast component, as illustrated in figure 3.4. The use of time-scale separation is not limited to systems where we have good evidence for slow and fast processes, but also can act as a framework (with additional assumptions) for the derivation of simplified models more generally (Gunawardena 2014).

Historically, two related assumptions have been made to reduce the complexity of biochemical systems: the *rapid equilibrium* assumption; and the *quasi-steady-state* assumption. Incidentally, the original derivation of the Michaelis-Menten equation used the rapid equilibrium assumption (Michaelis & Menten 1913; Johnson & Goody 2011), but now more commonly an equivalent quasi-steady-state argument is used (Briggs & Haldane 1925).

### 3.2.4 Rapid equilibrium

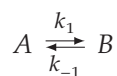
Under the rapid equilibrium (RE) assumption, we conclude that two complexes related by reversible reactions are in instantaneous equilibrium. That is, the net flux between those two complexes is zero. For example, if



which can be described by the system of ODEs

$$\begin{aligned} \frac{d[A]}{dt} &= -k_1[A] + k_{-1}[B] + \dots \\ \frac{d[B]}{dt} &= k_1[A] - (k_{-1} + k_2)[B] \\ \frac{d[C]}{dt} &= k_2[B] + \dots \end{aligned}$$

Then if complexes  $A$  and  $B$  are assumed to be in rapid equilibrium from the reaction scheme we have



and if the net flux between these species is zero, this means

$$0 = k_1[A] - k_{-1}[B]$$

$$\Rightarrow [B] = \frac{k_1}{k_{-1}}[A].$$

Under the rapid equilibrium assumption we explicitly ignore fluxes originating to and from complexes that we are not currently considering.

### 3.2.5 Quasi-steady-state

Under the quasi-steady-state (QSS) assumption, we assume that the rate of change of the concentration of a specific set of complexes is zero. For example, again assuming the reaction scheme in equation 3.9, then if species  $B$  is assumed to follow the quasi-steady-state assumption then setting the above equation for  $\frac{d[B]}{dt}$  equal to zero gives

$$0 = k_1[A] - (k_{-1} + k_2)[B]$$

$$\Rightarrow [B] = \frac{k_1}{k_{-1} + k_2}[A].$$

The quasi-steady-state assumption is arguably more biologically appropriate, as we are not artificially segregating sets of reactions, but can give more mathematically complicated solutions than the rapid equilibrium assumption.

## 3.3 Linear framework for time-scale separation

Throughout this thesis I will need to derive simplified models of enzyme-kinetic mechanisms. I will assume that each of these mechanisms forms a component of an eventual larger model, and that each is amenable to reduction in complexity (given the above assumptions) via the approach of time-scale separation. For the derivation of these models I will use the linear framework for time-scale separation provided by Gunawardena (2012). (This procedure is equivalent to the King-Altman method.) Here, I will restate the steps required in the derivation of such models, for further discussion of the mathematical proof behind these steps see Gunawardena (2012) or Gunawardena (2014). This framework consists of the sequential steps:

1. The *graph on the enzyme complexes*, henceforth  $G$ , should be drawn from the original biochemical mechanism. In this graph: the vertices are given by the unique list of enzyme complexes (complexes that include the enzyme) involved

in the mechanism; and the edges correspond to reactions that convert between the relevant complexes. The edges should be labelled by the kinetic rates multiplied by the concentration of any other species involved in the reaction, so that every label has dimension  $s^{-1}$ . Given a vector of concentrations of the enzyme complexes,  $\mathbf{x}$ , the original system of ODEs can be reconstituted as

$$\frac{d\mathbf{x}}{dt} = \mathcal{L}(G)\mathbf{x}$$

where  $\mathcal{L}(G)$  is the *Laplacian matrix* of  $G$ .

2. All of the *directed spanning trees* should then be found from the graph on the enzyme complexes, where: *directed* implies that arrow directions should be retained; *tree* implies the directions should be consistent, so that following the directions will always lead to a single *root* vertex; and *spanning* implies that every vertex remains connected by at least one edge.
3. The *kernel* of  $\mathcal{L}(G)$  describes all  $\mathbf{x}$  such that

$$\mathcal{L}(G)\mathbf{x} = 0.$$

Elements of the kernel of  $\mathcal{L}(G)$  are therefore steady-state solutions to the original system of ODEs. A *basis element* of the kernel of  $G$ ,  $\rho \in \ker \mathcal{L}(G)$  can be directly constructed from the directed spanning trees, using what is known as the Matrix Tree Theorem. Each component of the basis element,  $\rho_i$ , is given by the sum over the directed spanning trees rooted at vertex  $i$  ( $T \in T^i(G)$ ), of the product over the labels ( $l$  such that  $(A \xrightarrow{l} B) \in T$ ).

$$\rho_i = \sum_{T \in T^i(G)} \left( \prod_{(A \xrightarrow{l} B) \in T} l \right)$$

4. If  $G$  is *strongly connected* then we have  $\dim \ker \mathcal{L}(G) = 1$ , and so we know that the steady-state solutions must be of the form

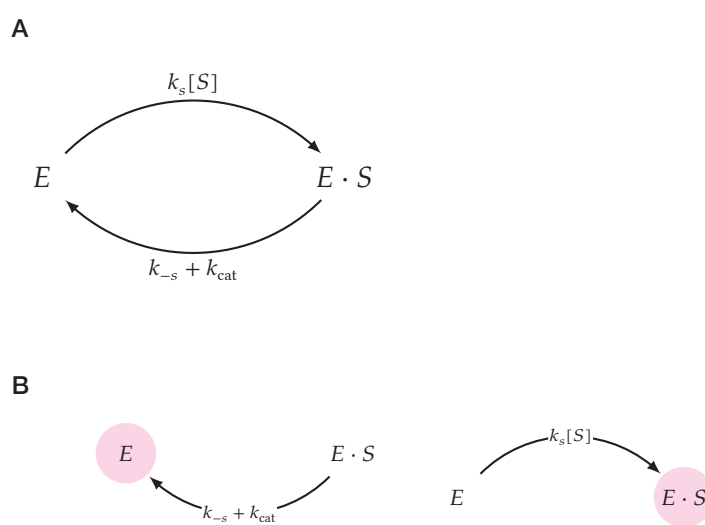
$$\mathbf{x} = \lambda \rho$$

where  $\lambda \in \mathbb{R}$  (i.e.  $\mathbf{x}$  is a scalar multiple of the basis element). Therefore

$$\frac{x_i}{\rho_i} = \lambda \text{ and } \frac{\sum_i x_i}{\sum_i \rho_i} = \lambda$$

and so we can write

$$x_i = \frac{\rho_i}{\sum_i \rho_i} \sum_i x_i.$$



**Figure 3.5** Graph and spanning trees for the Michaelis-Menten reaction scheme, equation 3.10.

**A** The graph on the enzyme complexes.

**B** The directed spanning trees of the graph on the enzyme complexes, each with the root node highlighted.

This gives the quasi-steady-state value of  $x_i$  in terms of the basis element and  $\sum_i x_i$ . Recall that the  $x_i$  described the (steady-state) concentrations of the enzyme complexes – now it only remains to substitute these values into the relevant equations in the original system of ODEs.

The models generated by the linear framework do not take into consideration the additional mass of substrate, product or regulator held in complex with the enzyme. Therefore the dynamics are only a good approximation when the concentrations of the intermediate complexes are small compared to the concentration of these species – which tends to be true when the total concentration of enzyme is small, and for enzymes with abundant small molecules substrates. This requirement is typical for quasi-steady-state approximations of enzyme-kinetic models, and implies a constraint on validity of the model in regards to experimental systems.

### 3.3.1 Example: the Michaelis-Menten equation

The best known example of a model that utilises the concept of time-scale separation and the quasi-steady-state assumption is the derivation of the Michaelis-Menten equation (Michaelis & Menten 1913; Johnson & Goody 2011) by Briggs & Haldane (1925). A derivation using the linear framework for time-scale separation—though not novel—is included here for illustrative purposes and later reference.

The Michaelis-Menten mechanism is described by the reaction scheme



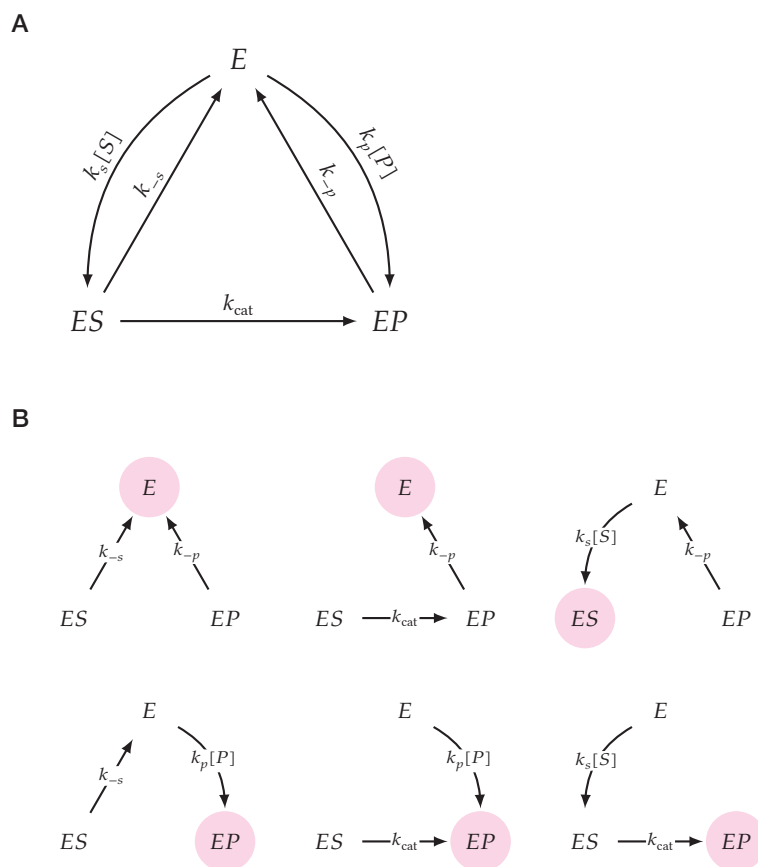
where  $E$  is an enzyme that converts a substrate,  $S$ , into a product,  $P$ , in an irreversible fashion.

The graph on the enzyme complexes and the directed spanning trees of this graph are given in figure 3.5. These spanning trees imply a basis element

$$\rho = \begin{bmatrix} k_{-s} + k_{\text{cat}} \\ k_s[S] \end{bmatrix}$$

in the order  $(E, E \cdot S)$ , and so

$$\begin{aligned} [E] &= \frac{k_{-s} + k_{\text{cat}}}{k_{-s} + k_{\text{cat}} + k_s[S]} e_0 \\ [E \cdot S] &= \frac{k_s[S]}{k_{-s} + k_{\text{cat}} + k_s[S]} e_0 \end{aligned}$$



**Figure 3.6** Graph and spanning trees for the Michaelis-Menten with product inhibition reaction scheme, equation 3.11.

**A** The graph on the enzyme complexes.

**B** The directed spanning trees of the graph on the enzyme complexes, each with the root node highlighted.

where

$$e_0 = [E] + [E \cdot S].$$

The rate of production of product is given by the equation

$$\frac{d[P]}{dt} = k_{\text{cat}}[E \cdot S].$$

Substituting in the quasi-steady-state value of  $[E \cdot S]$  gives

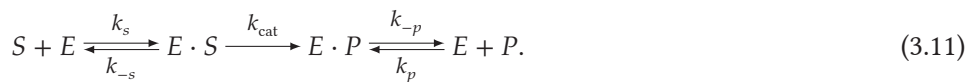
$$\begin{aligned} \frac{d[P]}{dt} &= \frac{k_{\text{cat}}k_s[S]}{k_{-s} + k_{\text{cat}} + k_s[S]}e_0 \\ &= \frac{k_{\text{cat}}[S]}{K_m + [S]}e_0 \end{aligned}$$

where  $K_m = \frac{k_{-s} + k_{\text{cat}}}{k_s}$  is known as the *Michaelis constant*.

### 3.3.2 Example: the Michaelis-Menten with product inhibition equation

An extension to the Michaelis-Menten mechanism is given by the inclusion of product inhibition. Again, a derivation using the linear framework for time-scale separation—though not novel—is included here for illustrative purposes and later reference.

This mechanism is described by the reaction scheme



The graph on the enzyme complexes and the directed spanning trees of this graph are given in figure 3.6. These spanning trees imply a basis element

$$\rho = \begin{bmatrix} (k_{-s} + k_{\text{cat}})k_{-p} \\ k_s k_{-p} [S] \\ (k_{-s} + k_{\text{cat}})k_p [P] + k_s k_{\text{cat}} [S] \end{bmatrix}$$

in the order  $(E, E \cdot S, E \cdot P)$ . It is possible to generate an equivalent basis element by multiplying each row by a scalar factor. By choosing an appropriate scalar factor we can simplify the final solution.

Dividing  $\rho$  by  $(k_{\text{cat}} + k_{-p})k_s$  gives

$$\rho' = \left[ \begin{array}{c} K_m \\ \left(1 - \frac{k'_{\text{cat}}}{k_{-p}}\right) [S] \\ \frac{K_m}{K_{\text{ic}}} [P] + \frac{k'_{\text{cat}}}{k_{-p}} [S] \end{array} \right]$$

where

$$K_m = \frac{(k_{-s} + k_{\text{cat}})k_{-p}}{(k_{\text{cat}} + k_{-p})k_s} \text{ is the apparent Michaelis constant;}$$

$$K_{\text{ic}} = \frac{k_{-p}}{k_p} \text{ is known as the constant of inhibition;}$$

$$k'_{\text{cat}} = \frac{k_{\text{cat}}k_{-p}}{k_{\text{cat}} + k_{-p}}.$$

From  $\rho'$  we obtain

$$[E] = \left( \frac{K_m}{K_m \left(1 + \frac{[P]}{K_{\text{ic}}}\right) + [S]} \right) e_0 \quad (3.12)$$

$$[E \cdot S] = \left( \frac{\left(1 - \frac{k'_{\text{cat}}}{k_{-p}}\right) [S]}{K_m \left(1 + \frac{[P]}{K_{\text{ic}}}\right) + [S]} \right) e_0 \quad (3.13)$$

$$[E \cdot P] = \left( \frac{\frac{K_m}{K_{\text{ic}}} [P] + \frac{k'_{\text{cat}}}{k_{-p}} [S]}{K_m \left(1 + \frac{[P]}{K_{\text{ic}}}\right) + [S]} \right) e_0 \quad (3.14)$$

where

$$e_0 = [E] + [E \cdot S] + [E \cdot P].$$

The rate of production of product is given by the equation

$$\frac{d[P]}{dt} = k_{-p}[E \cdot P] - k_p[P][E].$$

Substituting in the quasi-steady-state values of  $[E]$  and  $[E \cdot P]$  gives

$$\begin{aligned} \frac{d[P]}{dt} &= \left( \frac{k_{-p} \frac{k_s}{K_{\text{ic}}} [P] + k'_{\text{cat}} [S] - k_s k_p [P]}{K_m \left(1 + \frac{[P]}{K_{\text{ic}}}\right) + [S]} \right) e_0 \\ &= \frac{k'_{\text{cat}} [S] e_0}{K_m \left(1 + \frac{[P]}{K_{\text{ic}}}\right) + [S]}. \end{aligned} \quad (3.15)$$



This is the standard textbook definition of the quasi-steady-state equation for the Michaelis-Menten mechanism with product inhibition. This demonstrates that the generalised method of the linear framework is able to produce equations consistent with previous derivations.

## 3.4 Python module: enzymegraph

In order to simplify and semi-automate the derivation of quasi-steady-state models of enzymes using the linear framework for time-scale separation, I have written a module for the Python programming language. The source code for this module has been included in appendix A.

### 3.4.1 Algorithm for enumeration of spanning trees

In order to enumerate the directed spanning trees of a given graph I have modified an algorithm from Gabow & Myers (1978), such that:

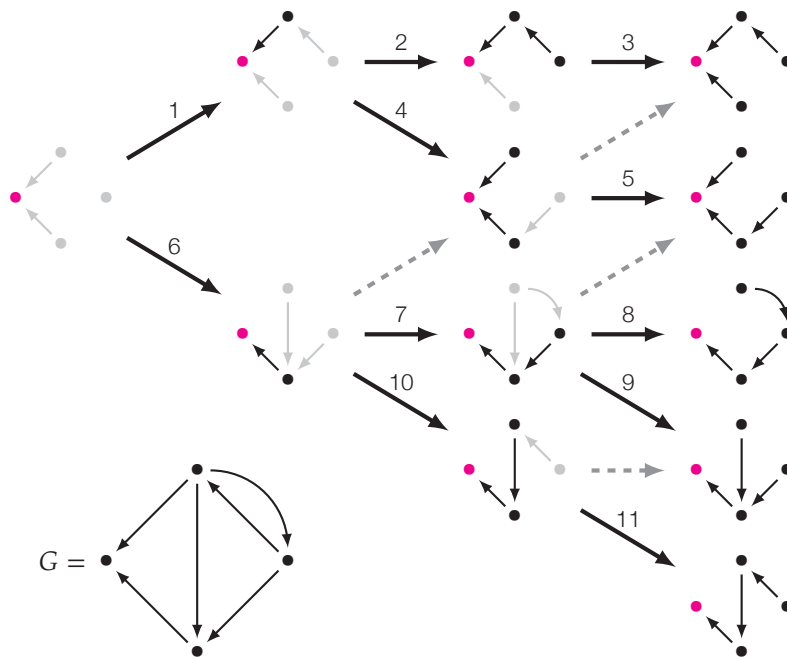
- Rather than maintaining a single graph object and growing a tree object in place, these are copied and modified appropriately on each recursion.
- Rather than maintaining a single list of inward edges for the growing tree, these are calculated newly on each recursion.
- The direction of the growing tree is inverted, so that all directions point *towards*, rather than away from, the root vertex.
- So that all spanning trees rooted at all vertices are returned, rather than those corresponding to a specified vertex.

This modified algorithm can be described by the pseudo-code:

```

procedure SpanningTrees(G)                                # G = graph
  procedure Grow(G, T)                                    # T = tree
    if T spans G then
      output T
    else
      G' ← G
      F ← all edges (w, v), w ∉ T and v ∈ T           # inward edges
      for each edge e ∈ F do
        T' ← T + e

```



**Figure 3.7** Application of the SpanningTrees( $G$ ) algorithm to find all of the directed spanning trees (here rooted at a single vertex) of an example graph,  $G$ . Each numbered (thick) arrow denotes a sequential, iterated recursion of the algorithm, each generating a new partial tree,  $T$ . For each partial tree, the chosen edges are drawn in black, and the inward edges,  $F$ , are drawn in grey. The algorithm outputs each  $T$  when the vertices of  $G$  have been spanned. Within each level of recursion, the edge that is added to  $T$  is removed from  $G$ , dashed (thick) lines denote duplicated trees that would be possible without this step.

```

    Grow( $G'$ ,  $T'$ )
     $G' \leftarrow G' - e$  # disallow choosing this edge again
  end for
end if
end procedure

for each vertex  $v$  in  $G$  do # main procedure loop
   $T \leftarrow$  vertex  $v$ 
  Grow( $G$ ,  $T$ )
end for
end procedure

```

The algorithm is illustrated by application, starting at a vertex of an arbitrary graph, in figure 3.7.

### 3.4.2 Using the enzyme graph module

The module `enzyme graph` is imported as a standard Python module. A graph object can be created from a dictionary of edges of the form  $(v, w) : l$ , which represents an edge from vertex  $v$  to vertex  $w$  with label  $l$ . Alternatively it can be created from a list of edges, where each edge is assumed to have label equal to 1; or from a  $n \times n$  matrix representing the Laplacian matrix of the graph with a list of vertex labels. The `enzyme graph` object can then be queried to return: the system of ODEs; the directed spanning trees (as `enzyme graph` objects); the basis element; and, the steady-state solution. For example, the complete derivation of the Michaelis-Menten with product inhibition mathematical model (previously derived manually in section 3.3.2) can be automated as:

```

1 from enzyme graph import *
2 from sympy import *
3
4 # symbols
5 e, s, p, es, ep = symbols('[E] [S] [P] [ES] [EP]', \
6     positive = True)
7 et = symbols('e_t', positive = True)
8 ks, kms, kp, kmp, kcat = symbols('k_s k_-s k_p, k_-p k_cat', \
9     positive = True)
10
11 # edges defined as a dictionary
12 edges = {(e, es): ks*s, (es, e): kms,
13     (e, ep): kp*p, (ep, e): kmp,

```

```

14         (es,ep): kcat, }
15
16 # create enzymegraph object
17 graph = enzymegraph(edges)
18
19 # print results
20 print('ODEs:')
21 for var, ode in graph.ode_model().items():
22     print('d%4s/dt = %s' % (var, ode))
23
24 print('\nSpanning trees:')
25 for span_tree in graph.spanning_trees():
26     print(span_tree.edges)
27
28 print('\nBasis element:')
29 for var, basis_elem in graph.basis_element().items():
30     print('%4s = %s' % (var, collect(basis_elem,[s,p] )))
31
32 print('\nSteady-state solution:')
33 for var, rep in graph.qssa_replacements(total_value=et).items():
34     print('%4s = %s' % (var, collect(rep,[s,p] )))

```

This gives the output:

```

ODEs:
d[EP]/dt = -[EP]*k_-p + [ES]*k_cat + [E]*[P]*k_p
d[ES]/dt = -[ES]*k_-s - [ES]*k_cat + [E]*[S]*k_s
d [E]/dt = [EP]*k_-p + [ES]*k_-s - [E]*[P]*k_p - [E]*[S]*k_s

Spanning trees:
{([E], [ES]): [S]*k_s, ([ES], [EP]): k_cat}
{([ES], [EP]): k_cat, ([E], [EP]): [P]*k_p}
{([ES], [E]): k_-s, ([E], [EP]): [P]*k_p}
{([E], [ES]): [S]*k_s, ([EP], [E]): k_-p}
{([EP], [E]): k_-p, ([ES], [EP]): k_cat}
{([EP], [E]): k_-p, ([ES], [E]): k_-s}

Basis element:
[EP] = [P]*(k_-s*k_p + k_cat*k_p) + [S]*k_cat*k_s
[ES] = [S]*k_-p*k_s
[E] = k_-p*k_-s + k_-p*k_cat

Steady-state solution:
[EP] = e_t*([P]*(k_-s*k_p + k_cat*k_p) + [S]*k_cat*k_s)/([P]*(k_-s
*k_p + k_cat*k_p) + [S]*(k_-p*k_s + k_cat*k_s) + k_-p*k_-s + k_
-p*k_cat)

```

$$\begin{aligned}
[ES] &= [S] * e_t * k_{-p} * k_s / ([P] * (k_{-s} * k_p + k_{cat} * k_p) + [S] * (k_{-p} * k_s + k_{cat} * k_s) + k_{-p} * k_{-s} + k_{-p} * k_{cat}) \\
[E] &= e_t * (k_{-p} * k_{-s} + k_{-p} * k_{cat}) / ([P] * (k_{-s} * k_p + k_{cat} * k_p) + [S] * (k_{-p} * k_s + k_{cat} * k_s) + k_{-p} * k_{-s} + k_{-p} * k_{cat})
\end{aligned}$$

These results are identical to those in section 3.3.2.

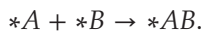
## 3.5 Cytosol-membrane transitions

In the models in this thesis I will need to consider the adsorption of species and complexes onto membranes, and be able to represent intrinsic (e.g. lipids) and transient (e.g. proteins) membrane-associated species. I require a careful mathematical/physical description of species that can move between the 3-dimensional cytosol and 2-dimensional membranes. Among these species are *surface-active* enzymes – enzymes that are ordinarily cytosolic but which have membrane-associated substrates.

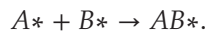
### 3.5.1 Notation

In addition to the notation described in section 3.2.1, specific notation relates to volume-surface interactions. *Concentrations per unit area* (as opposed to per unit volume) will be denoted by angular brackets  $\langle C \rangle$  (units  $\text{mol dm}^{-2}$ ). Membrane-associated species will be pre- or post-labelled with an asterisk, \*, such that intrinsic components of the membrane (e.g. lipids) have a pre-labelling, \*C, and extrinsic components of the membrane (e.g. membrane-tethered proteins) have a post-labelling, C\* (Kartal & Ebenhöf 2013). Reaction rules on labelled species are:

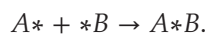
- Two intrinsic species combine to give an intrinsic complex,



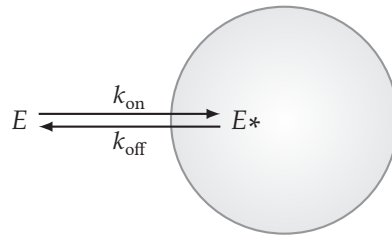
- Two extrinsic species combine to produce an extrinsic complex,



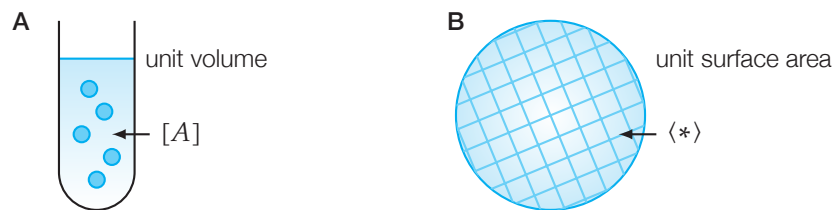
- An intrinsic and an extrinsic species combine to produce a mixed complex,



For shorthand, membrane-associated enzyme complexes will be denoted as previously described but with a superscript asterisk (e.g.  $C* \equiv E_{\sigma}^* \equiv E_{s_1 s_2 \dots s_n}^*$ ).



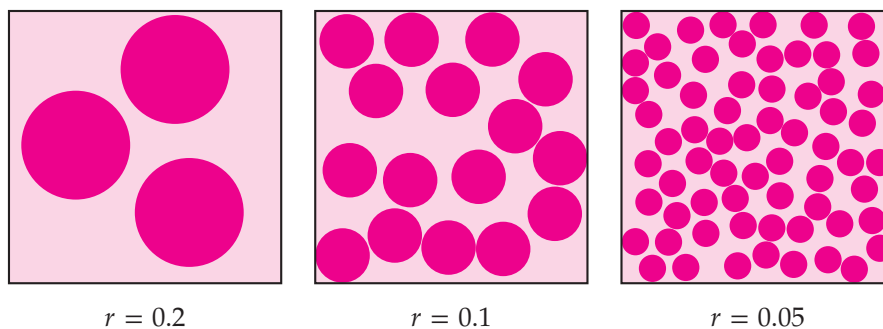
**Figure 3.8** Simple reversible surface adsorption of an enzyme,  $E$ , binding to a membrane.



**Figure 3.9** Illustrations of the physical meaning of:

**A**  $[A]$ , the concentration of surface area per unit volume. Given a unit volume, and knowing the (membrane) surface area contained within, we can calculate  $[A]$ .

**B**  $\langle * \rangle$ , the concentration of available elementary spaces per unit surface area. Given a unit surface area, and knowing the total number of particles that can fit within, we can calculate  $\langle * \rangle$ .

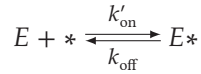


**Figure 3.10** Example jammed coverages for the RSA model on a (bounded) surface with different radiuses of adsorbate.

### 3.5.2 Surface adsorption

A simple adsorption process consists of a single species or complex which can bind or unbind from a membrane, as illustrated in figure 3.8. In this section I will reproduce the derivation by Kartal & Ebenhöh (2013) of a mathematical model of this process.

Adsorption is modelled as a bimolecular reaction between the adsorbate (e.g. an enzyme),  $E$ , and an elementary space,  $*$ , on a membrane



where  $k'_{\text{on}}$  is a second-order rate constant. In terms of volume concentrations, using the law of mass-action we can write

$$\frac{d[E]}{dt} = -k'_{\text{on}}[*][E] + k_{\text{off}}[E*] \quad (3.16)$$

where  $[*]$  is the concentration of elementary spaces available (on membranes) per unit volume.

The value of  $[*]$  is not immediately self-evident, but we can calculate it from

$$[*] = [A]\langle * \rangle$$

where  $[A]$  is the surface area per unit volume; and  $\langle * \rangle$  is the concentration of available elementary spaces per unit surface area. These two variables are illustrated in figure 3.9. In most systems  $[A]$  can be assumed to be constant. Now equation 3.16 can be written as

$$\frac{d[E]}{dt} = -k'_{\text{on}}\langle * \rangle[A][E] + k_{\text{off}}[E*].$$

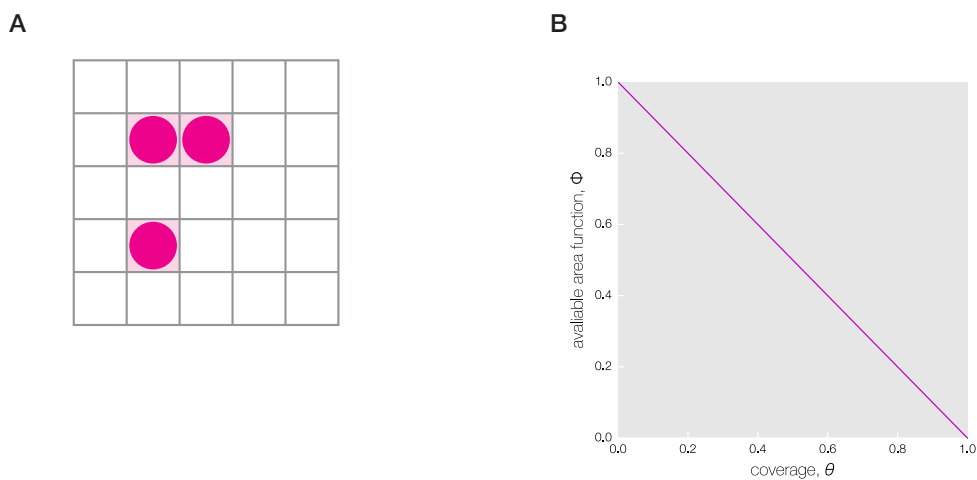
We need to calculate a value for  $\langle * \rangle$ , in general we can write

$$\langle * \rangle = n_0\Phi$$

where  $0 \leq \Phi \leq 1$  is the *available area function*, the proportion of surface that is still available for adsorbate to bind; and  $n_0$  is the maximum surface concentration of the adsorbate. Now equation 3.16 can be written as

$$\frac{d[E]}{dt} = -k_{\text{on}}\Phi[A][E] + k_{\text{off}}[E*] \quad (3.17)$$

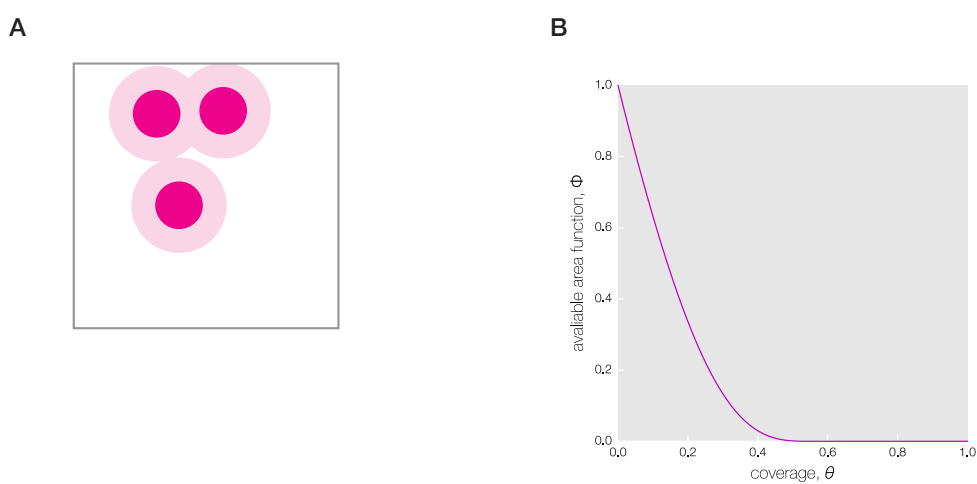
where  $k_{\text{on}} = n_0k'_{\text{on}}$ . This equation describes the mathematical model for simple surface adsorption.



**Figure 3.11** Langmuir's model.

**A** Adsorption sites are assumed to be discrete and independent. Each site is only inaccessible if a molecule is currently bound in that specific site.

**B** The available area function.



**Figure 3.12** Random sequential adsorption (RSA) model.

**A** Occluded area includes any point less than two times the radius from the centre of any existing discs.

**B** The available area function, approximated by equation 3.20.



It remains to determine a form for  $\Phi$ . First we need to introduce the theoretical inverse, the *fractional surface coverage* or *coverage*,  $0 \leq \theta \leq 1$ . This can be described by the equation

$$\theta = \frac{\langle E^* \rangle}{n_0} = \frac{[E^*]}{[A]n_0}. \quad (3.18)$$

In general the available area function,  $\Phi$ , may not be a linear function of the coverage,  $\theta$ . Instead, this relation, while monotonically decreasing, will depend on the physical model assumed for the adsorption process. In this thesis, following Kartal & Ebenhöh (2013), this will be assumed to be either *Langmuir's model* or the *Random sequential adsorption* (RSA) model.

### 3.5.3 Langmuir's model

*Langmuir's model* (Langmuir 1918) is the simplest model for the available area function. It assumes that the adsorption sites are discrete and independent and that adsorbed molecules do not interact, see figure 3.11. It is described by the (linear) relation

$$\Phi(\theta) = 1 - \theta. \quad (3.19)$$

### 3.5.4 Random sequential adsorption model

*Random sequential adsorption* (RSA) (Feder 1980) is a more complex model for the available area function. It describes a continuous physical model, where all positions are equally likely for an adsorbate to bind, unless already covered or partially covered by a previous adsorbate, see figure 3.12. Under this model, *jamming* can occur where enough total available area could exist for another molecule to bind, but where this area is not contiguous, shown in figure 3.10. (Jamming is a well established phenomena exhibited in general in other particle systems (van Hecke 2010).)

The RSA model can be approximated by the equation

$$\Phi(\theta) = (\theta_c - \theta)^3 (a_1 + a_2\theta + a_3\theta^2 + a_4\theta^3) \quad (3.20)$$

with  $\theta_c = 0.547$ ,  $a_1 = 6.107$ ,  $a_2 = 9.059$ ,  $a_3 = 8.669$ , and  $a_4 = -28.56$  (Manciu & Ruckenstein 2004). This equation is only valid for  $0 \leq \theta \leq \theta_c$ . ( $\Phi(\theta) = 0$  for  $\theta_c < \theta \leq 1$ .)

## 3.6 Summary

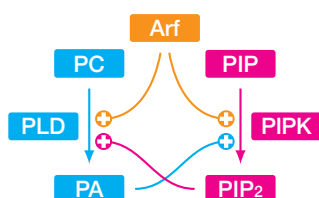
In this chapter I have summarised the mathematical modelling and analysis techniques that I will require for the remainder of this thesis.

The analytical techniques for steady-state analysis of systems of ODEs will be required in chapter 4 to determine the possible behaviours for whole classes of (two-dimensional, qualitative, empirical) models of the Arf/PLD/PI4P5K motif and other related (theoretical) motifs.

Also, I have introduced two major techniques in this chapter: time-scale separation using the quasi-steady-state approximation, via a new Python module that implements the framework of Gunawardena (2014); and, a framework for modelling cytosol-membrane transitions. These will be used to construct mechanistic models of PLD and PI4P5K in chapter 5, and Arf in chapter 6.

## 4 Empirical models

Consider again the Arf/PLD/PI4P5K signalling motif. This can be drawn as follows, described by reactions ( $\rightarrow$ ) and (positive, activating) regulations ( $\rightarrow\oplus$ ).



As a first approach I would like to construct *empirical* models of this system – models based on coarse-grained qualitative descriptions of the regulatory and catalytic processes of PLD and PI4P5K. My intention is to develop small, simple, and fully analysable models which are sufficient to allow some exploration of the response to stimuli of the Arf/PLD/PI4P5K motif caused by the cross-talk between the products.

These empirical models will be simplified representations of the signalling motif, where I will *choose* suitable mathematical functions to reproduce the observed behaviours of the cross-talk and other processes within the motif. The main advantage of this approach is that I can construct small models (of two variables), which I hope to be able to mathematically analyse using the techniques listed in section 3.1. This would provide a complete characterisation of the responses to stimuli (for a model) of the system.

Empirical models are inherently biased and incomplete compared to mechanistic models – see the previous discussion in section 1.3. The main advantages of these empirical models will lie in their simplicity and their mathematical analysability. If an empirical model becomes too complicated to analyse easily then it has lost much of its advantage over a mechanistic model. Therefore, one measure of success for the following empirical models is whether they are easily analysable.

To build empirical models I will need to relax some of the mathematical rules introduced in chapter 3. While I will still consider rates as proportional to concentrations, I will no longer use a strict application of the law of mass-action. For this reason I will refer to these empirical models as *pseudo-mass-action* models. In general, I will choose suitable mathematical functions to reproduce the experimentally observed regulation of the two enzymes.

First, I will re-introduce an empirical model of the Arf/PLD/PI4P5K motif taken from earlier work, and expand significantly on those results. This model is simple enough that its qualitative behaviours can be completely characterised using the techniques introduced in section 3.1.2. However, this model displays physiologically-unrealistic behaviour – given sufficient stimulation, the concentration of the products are unbounded and increase exponentially. This behaviour is physically impossible, as an (absolute) upper bound for the amount of PA and PI(4,5)P<sub>2</sub> is given by the (finite) total mass of a cell!

Here, I will make two attempts to modify this first model in order to bound its growth, either:

- replacement of the linear cross-talk activation by a non-linear Hill function, widely used in modelling biochemical systems;
- or, adding conservation of mass to the system.

I will show that both of these approaches add sufficiently to the complexity of the models, such that they are less amenable to mathematical characterisation.

I will also introduce a complementary *comparative* approach which will help me explore the effect of the asymmetry in the structure of Arf/PLD/PI4P5K motif. This involves the construction of models of *theoretical* motifs that are—in some way—related to the original biological motif. If each of these models is constructed in the same manner, we can expect each of them to have roughly similar advantages and disadvantages (in terms of their mathematical complexity and physiological realism). The point of this approach is that it allows us to ask questions such as: given that all of the motifs are equally plausible (in some evolutionary sense), are there any specific advantages of the biologically observed motif? In this way, we can ignore some of the specific issues of the models of the biological system, instead focusing on a comparison with models of the theoretical motifs. Here, these alternative motifs will be generated by mirroring the regulatory description of either one of the enzymes (PLD and PI4P5K).

This chapter will make use of a small custom Python module `empirical.py` which defines a function `characterise()` which can calculate symbolic functions or numerical values for the steady-states and eigenvalues of a system:

```

1 from sympy import *
2
3 # symbolic setup
4 x, y, z = symbols('x y z')
5 a0, a1, a2, ar = symbols('a0 a1 a2 ar', positive = True)
6 b0, b1, b2, br = symbols('b0 b1 b2 br', positive = True)
7
8 # characterise steady-states and eigenvalues
9 def characterise(name, sys, p={}):
10     # get list of steady-states
11     ss = solve(sys, [x,y], dict=True)
12     # get list of eigenvalues
13     ev = list(Matrix(sys).jacobian([x,y]).eigenvals().keys())
14
15     # output
16     print("%s: (%d solutions)" % (name, len(ss)))
17     for ssi in ss:
18         # parameterise steady-state, eigenvalues
19         ssi = {k:v.subs(p) for k,v in ssi.items()}
20         evi = [v.subs(ssi).subs(p) for v in ev]
21
22         try: # try to print numerical expressions
23             print(" x=%+.2f, y=%+.2f, ev1=%+.2f, ev2=%+.2f" \
24                   % (ssi[x], ssi[y], evi[0], evi[1]))
25         except: # print string expressions
26             print(" x=%s\n y=%s\n ev1=%s\n ev2=%s" \
27                   % (ssi[x], ssi[y], evi[0], evi[1]))
28
29     print('')

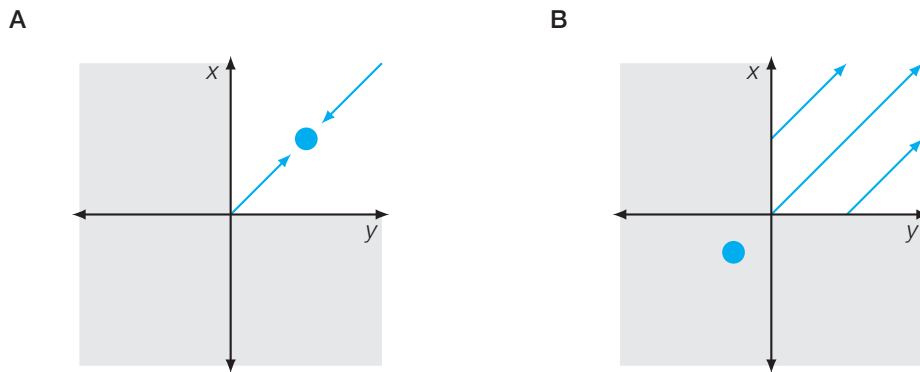
```

## 4.1 Arf/PLD/PI4P5K: pseudo-mass-action model

As part of earlier work, I constructed a simple empirical model that attempted to describe the behaviour of the Arf/PLD/PI4P5K motif (Stanley 2011). The important derivations have been included here in appendix B. I will (continue) to refer to this as the pseudo-mass-action model for the Arf/PLD/PI4P5K motif. The model is given by the 2-dimensional system of ODEs

$$\frac{dx}{dt} = a_0 + a_1 yz - a_r x \quad (4.1)$$

$$\frac{dy}{dt} = b_0 + b_1 x + b_2 z - b_r y \quad (4.2)$$



**Figure 4.1** Cartoon phase planes. The stability and location of the single steady-state (blue dot) in the pseudo-mass-action model depends on the choice of parameters and the value of  $z$  (the activation by Arf). Example trajectories are drawn in blue.

- A** Stable steady-state in positive quadrant when Arf activation ( $z$ ) is small.
- B** Unstable steady-state in negative quadrant when Arf activation ( $z$ ) is large.

where  $x = [\text{PA}]$ ,  $y = [\text{PI}(4,5)\text{P}_2]$ , and  $z = [\text{Arf}]$ . These equations describe the net rate of change of the products of PLD and PI4P5K, PA and PI(4,5)P<sub>2</sub> respectively. The positive terms in these equations describe the catalytic activity of PLD and PI4P5K, which will be described shortly; and the negative terms describe first-order degradation ( $a_r x$ ,  $b_r y$ ). I have not included regulation of Arf by Arf-GEFs and Arf-GAPs, so the system is perturbed by varying the parameter  $z$  manually.

Both equations include a basal rate of production ( $a_0$ ,  $b_0$ ). However, they differ in the description of their regulation by Arf and the other product:

- In equation 4.1, PLD is assumed to require both Arf and PI(4,5)P<sub>2</sub> to be active, and so this is modelled using multiplication of the two regulator concentrations, scaled by a rate constant ( $a_1 y z$ ).
- In equation 4.2, PI4P5K is assumed to be activated independently by Arf and PA, and so this is modelled using addition of the two regulator concentrations, each scaled by a rate constant ( $b_1 x + b_2 z$ ).

It is possible to use mathematical analysis to completely categorise all of the possible behaviours of this pseudo-mass-action model. This analysis has been included in appendix B. The system can be shown to have a single steady-state, the stability and location of which is determined by the system's parameters, but the behaviour of which is always consistent, following the form illustrated in figure 4.1. This behaviour, under control by  $z$ , is switch-like – a small increase in  $z$  past a threshold value (a point of bifurcation) massively changes the qualitative dynamics of the system; and decreasing  $z$  past this threshold switches the system back to the original behaviour.

When  $z$  is small, the system has a positive, stable steady-state that will attract all (physiologically-relevant) initial conditions,

$$\lim_{t \rightarrow \infty} x = x_{ss}$$

$$\lim_{t \rightarrow \infty} y = y_{ss}$$

where  $(x_{ss}, y_{ss})$  is the coordinate of the steady-state. When  $z$  is large enough, the system has a negative, unstable steady-state, which means that all (physiologically-relevant) initial conditions will give trajectories that increase without bound,

$$\lim_{t \rightarrow \infty} x = \infty$$

$$\lim_{t \rightarrow \infty} y = \infty.$$

I will call systems of the former type *bounded* and the latter type *unbounded*.

### 4.1.1 Implicit solution for the pseudo-mass-action model

I will now extend my previous analysis of the pseudo-mass-action model of the Arf/PLD/PI4P5K motif by finding an implicit solution for equations 4.1 and 4.2. This will give an implicit equation for each of  $x$  and  $y$ .

I will start by finding a solution to the *homogenous* system (without constant part) formed from the change of variables,

$$u(t) = x(t) - x_{ss}$$

$$v(t) = y(t) - y_{ss}$$

where  $(x_{ss}, y_{ss})$  is the coordinate of the steady-state. This system can be written as

$$\frac{du}{dt} = a_1 v z - a_r u \quad (4.3)$$

$$\frac{dv}{dt} = b_1 u - b_r v. \quad (4.4)$$

In general, a 2-dimensional homogenous linear system of ODEs will have a solution of the form

$$u(t) = c_1 e^{\lambda_+ t} + c_2 e^{\lambda_- t} \quad (4.5)$$

$$v(t) = c_3 e^{\lambda_+ t} + c_4 e^{\lambda_- t} \quad (4.6)$$

where  $\lambda_{\pm}$  are the eigenvalues of the system.

We want to find a value for each of the constants  $c_i$ , in terms of the parameters of the system. From equations 4.5 and 4.6 we know that at  $t = 0$

$$u(0) = c_1 + c_2$$

$$v(0) = c_3 + c_4.$$

We can also differentiate equations 4.5 and 4.6, and again substitute  $t = 0$ , to get

$$\frac{du}{dt}(0) = \lambda_+ c_1 + \lambda_- c_2$$

$$\frac{dv}{dt}(0) = \lambda_+ c_3 + \lambda_- c_4.$$

However, we have alternative descriptions for these initial rates from equations 4.3 and 4.4,

$$\frac{du}{dt}(0) = a_1 z v(0) - a_r u(0)$$

$$\frac{dv}{dt}(0) = b_1 u(0) - b_r v(0).$$



Setting values for  $\frac{du}{dt}(0)$  equal, substituting  $c_2 = u(0) - c_1$ , and rearranging gives

$$c_1 = -\frac{\lambda_- + a_r}{\lambda_+ - \lambda_-} u(0) + \frac{a_1 z}{\lambda_+ - \lambda_-} v(0).$$

and (using  $c_2 = u(0) - c_1$ )

$$c_2 = \frac{\lambda_+ + a_r}{\lambda_+ - \lambda_-} u(0) - \frac{a_1 z}{\lambda_+ - \lambda_-} v(0).$$

Similarly, setting values for  $\frac{dv}{dt}(0)$  equal, substituting  $c_4 = v(0) - c_3$ , and rearranging gives

$$c_3 = \frac{b_1}{\lambda_+ - \lambda_-} u(0) - \frac{\lambda_- + b_r}{\lambda_+ - \lambda_-} v(0)$$

$$c_4 = -\frac{b_1}{\lambda_+ - \lambda_-} u(0) + \frac{\lambda_+ + b_r}{\lambda_+ - \lambda_-} v(0).$$

From appendix B we already know the form of the eigenvalues, which are

$$\lambda_{\pm} = \frac{1}{2} \left[ -a_r - b_r \pm \sqrt{(a_r - b_r)^2 + 4a_1 b_1 z} \right]$$

which we can substitute into the  $c_i$  to obtain

$$c_1 = \frac{1}{C} \left[ \left( b_r - a_r + \sqrt{(a_r - b_r)^2 + 4a_1 b_1 z} \right) u(0) + 2a_1 z v(0) \right]$$

$$c_2 = \frac{1}{C} \left[ \left( a_r - b_r + \sqrt{(a_r - b_r)^2 + 4a_1 b_1 z} \right) u(0) - 2a_1 z v(0) \right]$$

$$c_3 = \frac{1}{C} \left[ b_1 u(0) + \left( a_r - b_r + \sqrt{(a_r - b_r)^2 + 4a_1 b_1 z} \right) v(0) \right]$$

$$c_4 = \frac{1}{C} \left[ -b_1 u(0) + \left( b_r - a_r + \sqrt{(a_r - b_r)^2 + 4a_1 b_1 z} \right) v(0) \right]$$

where

$$C = 2\sqrt{(a_r - b_r)^2 + 4a_1 b_1 z}.$$

The values for the constants  $c_i$ , together with equations 4.5 and 4.6 give an implicit solution for the system of ODEs given by equations 4.3 and 4.4, and we can now follow any trajectory of the system without requiring numerical integration of the original ODEs.

As the homogeneous  $(u, v)$  system is just a linear shift of the non-homogeneous  $(x, y)$  system, we can use this solution to derive a solution to the original pseudo-mass-action model of the Arf/PLD/PI4P5K motif,

$$x(t) = c_1 e^{\lambda_+ t} + c_2 e^{\lambda_- t} + x_{ss}$$

$$y(t) = c_3 e^{\lambda_+ t} + c_4 e^{\lambda_- t} + y_{ss}$$

where the  $c_i$ ,  $\lambda_{\pm}$  are defined above, and  $x_{ss}$  and  $y_{ss}$  are given in appendix B.

### 4.1.2 Implicit solution for the pseudo-mass-action model: behaviour when $t \gg 0$

It is possible to use the implicit solution for the pseudo-mass-action model of the Arf/PLD/PI4P5K motif to verify the behaviour of the system under the constraint that  $z$  is large – specifically when

$$z > \frac{a_r b_r}{a_1 b_1}$$

which corresponds to the unstable, negative steady-state illustrated in figure 4.1B, see appendix B.

The equations for  $u$  and  $v$  in equations 4.5 and 4.6 are the sums of two exponential terms. We know that

$$\begin{aligned} \lim_{t \rightarrow \infty} e^{\lambda_- t} &= 0 \\ \lim_{t \rightarrow \infty} e^{\lambda_+ t} &\gg 0 \end{aligned}$$

as  $\lambda_-$  is always negative, and  $\lambda_+$  is always positive under the above constraint. So when  $t = t_\infty \gg 0$  the positive exponential term dominates, and we have

$$\begin{aligned} u(t_\infty) &\approx c_1 e^{\lambda_+ t_\infty} \\ v(t_\infty) &\approx c_3 e^{\lambda_+ t_\infty} \end{aligned}$$

where  $c_1$  and  $c_3$  are strictly positive as

$$b_r - a_r < \sqrt{(a_r - b_r)^2 + 4a_1 b_1 z}.$$

So we confirm that with a large enough value of  $z$  both  $u$  and  $v$  (and so  $x$  and  $y$ ) increase exponentially, and are unbounded.

## 4.2 Theoretical motifs

The unbounded growth which characterises the highly stimulated state of the pseudo-mass-action model of the Arf/PLD/PI4P5K motif, limits its applicability as a predictive tool because this corresponds to a physically impossible behaviour. However, the analytic solution to this model remains an attractive result, so for the moment I would like to continue to discuss linear models of approximately this complexity.

From previous description, we know that PLD and PI4P5K are regulated distinctly: PLD is only activated when both Arf and PI(4,5)P<sub>2</sub> are present; whereas PI4P5K

is activated by the presence of either Arf or PA. So the motif is not symmetric in regards to the regulation of the enzymes.

But what about a hypothetical motif that was symmetric? What behaviours would this motif display, that would distinguish it from the Arf/PLD/PI4P5K motif? More broadly, given a set of equally plausible motifs: does the Arf/PLD/PI4P5K motif have any particular advantages over the others in terms of signal propagation? This could also be considered through an evolutionary perspective: are there any advantages that may have led to the evolution of the structure of this specific motif?

The remainder of this chapters will be driven by these questions. A natural way to generate two new theoretical motifs is by mirroring the regulatory behaviour of each of PLD and PI4P5K. I will attempt to answer the above questions by constructing a series of models of all three motifs (the original motif and the two symmetric motifs).

#### 4.2.1 Logical biochemistry

The names that I will give to the theoretical models that I will discuss in this chapter are derived from my observation that the regulation of the enzymes PLD and PI4P5K are analogous to the inputs and outputs of specific *logic gates*. Logic gates are mathematical constructs that when given two inputs ( $A$  and  $B$ ), both of which can be either *on* or *off*, give an associated output.

We can describe the activation requirements for PLD and PI4P5K using the tables shown in figure 4.2, and I note these appear identical to the logic gates shown in figure 4.3. Therefore, I will consider the activation of PLD to be analogous to an *AND gate*, and the activation of PI4P5K to be analogous to an *OR gate*. For this reason, in this chapter, I will henceforth refer to the Arf/PLD/PI4P5K motif as the *AND-OR motif*.

I can then define two theoretical motifs by mirroring each enzyme in the AND-OR motif. These are:

- an *AND-AND motif*, where both enzymes are regulated like PLD – the activation of the enzymes requires the presence of both of their regulators;
- and, an *OR-OR motif* where both enzymes are regulated like PI4P5K – the enzymes are independently activated by their two regulators.

Together, the three motifs are illustrated in figure 4.4.

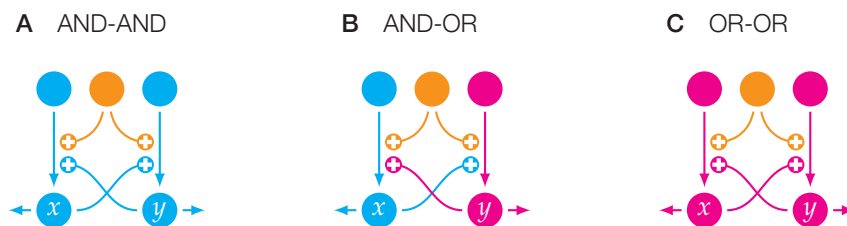
The models I will develop will expect inputs and outputs in terms of concentrations (real, positive numbers) and so logic gates, which are binary functions, are not

A PLD			B PI4P5K		
Presence of Arf	PI(4,5)P <sub>2</sub>	Activation of PLD	Presence of Arf	PA	Activation of PI4P5K
no	no	no	no	no	no
yes	no	no	yes	no	yes
no	yes	no	no	yes	yes
yes	yes	yes	yes	yes	yes

**Figure 4.2** Requirements for the activation of PLD and PI4P5K in respect to the presence of two regulatory molecules.

A AND gate			B OR gate		
Input A	Input B	Output A AND B	Input A	Input B	Output A OR B
0	0	0	0	0	0
1	0	0	1	0	1
0	1	0	0	1	1
1	1	1	1	1	1

**Figure 4.3** Logic gates and their inputs ( $A, B$ ) and outputs. Here 1 corresponds to a *on* signal, and 0 to an *off* signal.



**Figure 4.4** Three theoretically possible signalling motifs, in analogy to the Arf/PLD/PI4P5K signalling motif.

technically appropriate. Instead, corresponding to the previous description of the pseudo-mass-action model, qualitatively similar, continuous behaviour is attained by using multiplication instead of an AND gate, and addition instead of an OR gate.

#### 4.2.2 Pseudo-mass-action models: construction

I will construct a first set of models of the theoretical motifs in direct analogy to the pseudo-mass-action model of the Arf/PLD/PI4P5K motif described by equations 4.1 and 4.2. Each model will consist of a two dimensional system of ODEs, describing the net rate of change in the products of two enzymes ( $x$  and  $y$ ), controlled by the parameter  $z$ . For the AND-OR (the Arf/PLD/PI4P5K) motif, we will continue to have  $x = [PA]$ ,  $y = [PI(4,5)P_2]$ , and  $z = [Arf]$ . Otherwise these variables should be understood to represent arbitrary species. Other assumptions in the following models (unless otherwise specified) are:

- Substrates are well buffered, and are continuously available from a constant source.
- Products are removed via a sink.
- There is a (possibly zero) rate of basal activity for each enzyme.

Table 4.1 lists the pseudo-mass-action models of the three motifs: one is equivalent to the previously discussed model of the Arf/PLD/PI4P5K motif; and an equivalent model each for the AND-AND and OR-OR motifs, formed by duplicating the corresponding equation.

#### 4.2.3 Pseudo-mass-action models: analysis

The steady-state solutions and eigenvalues for the pseudo-mass-action models given in table 4.1 can be validated using the Python script:

```

1 from sympy import *
2 from empirical import *
3
4 # model setup
5 sys_aa = [a0 + a1*y*z - ar*x,
6           b0 + b1*x*z - br*y]
7 sys_ao = [a0 + a1*y*z - ar*x,
8           b0 + b1*x + b2*z - br*y]
9 sys_oo = [a0 + a1*y + a2*z - ar*x,
10          b0 + b1*x + b2*z - br*y]

```

Motif	ODEs	Steady-state	EV*	Stable
<p>A</p>	$\frac{dx}{dt} = a_0 + a_1 yz - a_r x$ $\frac{dy}{dt} = b_0 + b_1 xz - b_r y$	$x = \frac{a_0 b_r + a_1 b_0 z}{a_r b_r - a_1 b_1 z^2}$ $y = \frac{a_r b_0 + a_0 b_1 z}{a_r b_r - a_1 b_1 z^2}$	$\lambda_+(z^2)$ $\lambda_-(z^2)$	$\frac{a_r b_r}{a_1 b_1} > z^2$
<p>B</p>	$\frac{dx}{dt} = a_0 + a_1 yz - a_r x$ $\frac{dy}{dt} = b_0 + b_1 x + b_2 z - b_r y$	$x = \frac{a_0 b_r + a_1 z(b_0 + b_2 z)}{a_r b_r - a_1 b_1 z}$ $y = \frac{a_0 b_1 + a_r(b_0 + b_2 z)}{a_r b_r - a_1 b_1 z}$	$\lambda_+(z)$ $\lambda_-(z)$	$\frac{a_r b_r}{a_1 b_1} > z$
<p>C</p>	$\frac{dx}{dt} = a_0 + a_1 y + a_2 z - a_r x$ $\frac{dy}{dt} = b_0 + b_1 x + b_2 z - b_r y$	$x = \frac{(a_1(b_0 + b_2 z) + b_r(a_0 + a_2 z))}{a_r b_r - a_1 b_1}$ $y = \frac{(a_r(b_0 + b_2 z) + b_1(a_0 + a_2 z))}{a_r b_r - a_1 b_1}$	$\lambda_+(1)$ $\lambda_-(1)$	$\frac{a_r b_r}{a_1 b_1} > 1$

\*Eigenvalues,  $\lambda_{\pm}(\zeta) = \frac{1}{2} \left[ -(a_r + b_r) \pm \sqrt{(a_r - b_r)^2 + 4a_1 b_1 \zeta} \right]$

**Table 4.1** Pseudo-mass-action models and characterisation of steady-states for the three motifs (AND-AND; AND-OR; OR-OR). The AND-OR model and its solution have been reported previously (Stanley 2011), see appendix B.

```

11
12 # characterise
13 characterise('AND-AND', sys_aa)
14 characterise('AND-OR', sys_ao)
15 characterise('OR-OR', sys_oo)

```

This gives the output:

```

AND-AND: (1 solutions)
  x=-(a0*br + a1*b0*z)/(a1*b1*z**2 - ar*br)
  y=-(a0*b1*z + ar*b0)/(a1*b1*z**2 - ar*br)
  ev1=-ar/2 - br/2 + sqrt(4*a1*b1*z**2 + ar**2 - 2*ar*br + br**2)/2
  ev2=-ar/2 - br/2 - sqrt(4*a1*b1*z**2 + ar**2 - 2*ar*br + br**2)/2

AND-OR: (1 solutions)
  x=-(a0*br + a1*z*(b0 + b2*z))/(a1*b1*z - ar*br)
  y=-(a0*b1 + ar*(b0 + b2*z))/(a1*b1*z - ar*br)
  ev1=-ar/2 - br/2 - sqrt(4*a1*b1*z + ar**2 - 2*ar*br + br**2)/2
  ev2=-ar/2 - br/2 + sqrt(4*a1*b1*z + ar**2 - 2*ar*br + br**2)/2

OR-OR: (1 solutions)
  x=-(a1*(b0 + b2*z) + br*(a0 + a2*z))/(a1*b1 - ar*br)
  y=-(ar*(b0 + b2*z) + b1*(a0 + a2*z))/(a1*b1 - ar*br)
  ev1=-ar/2 - br/2 - sqrt(4*a1*b1 + ar**2 - 2*ar*br + br**2)/2
  ev2=-ar/2 - br/2 + sqrt(4*a1*b1 + ar**2 - 2*ar*br + br**2)/2

```

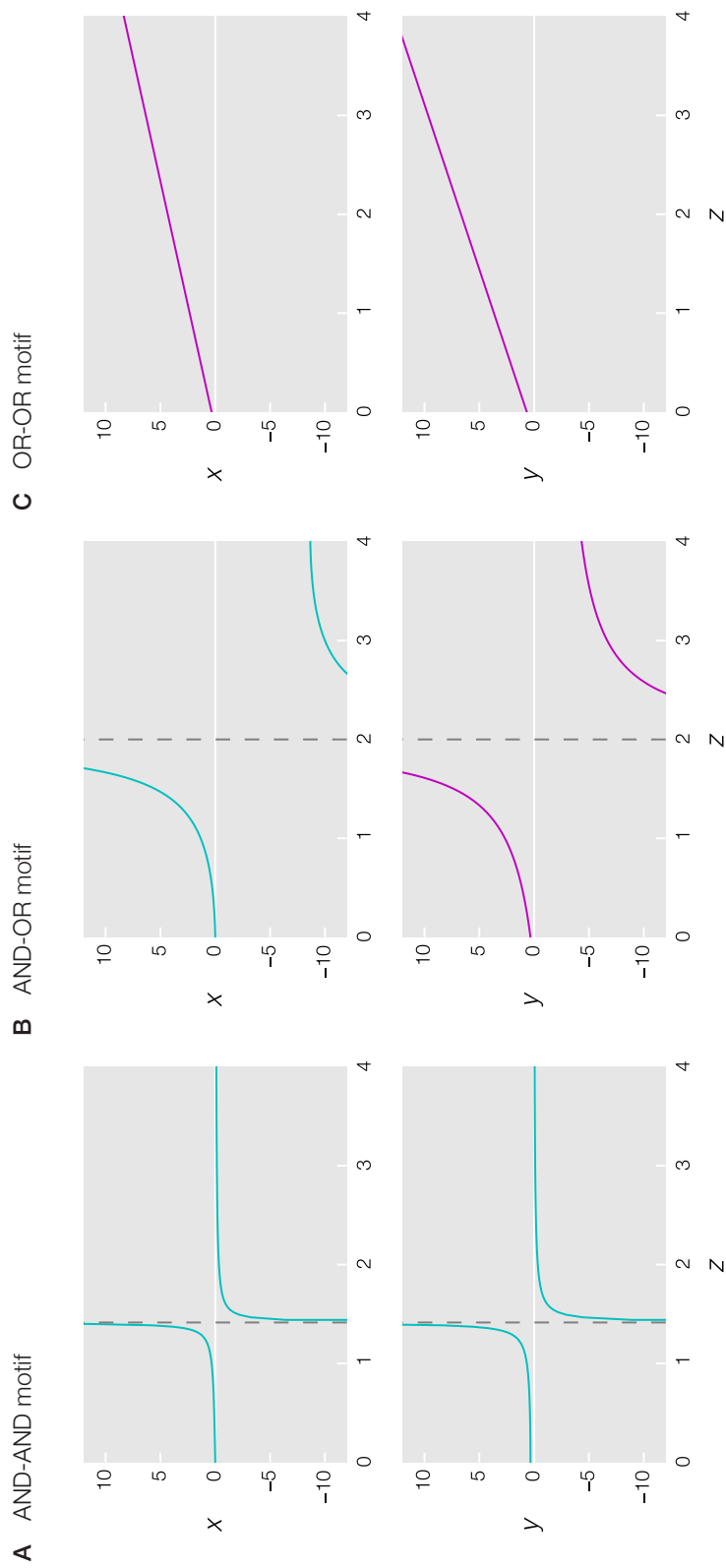
All three have a single steady-state, and eigenvalues of a comparable form,

$$\lambda_{\pm}(\zeta) = \frac{1}{2} \left[ -(a_r + b_r) \pm \sqrt{(a_r - b_r)^2 + 4a_1 b_1 \zeta} \right]$$

where  $\zeta = z^2$  for the model of the AND-AND motif;  $\zeta = z$  for the model of the AND-OR motif; and  $\zeta = 1$  for the model of the OR-OR motif.

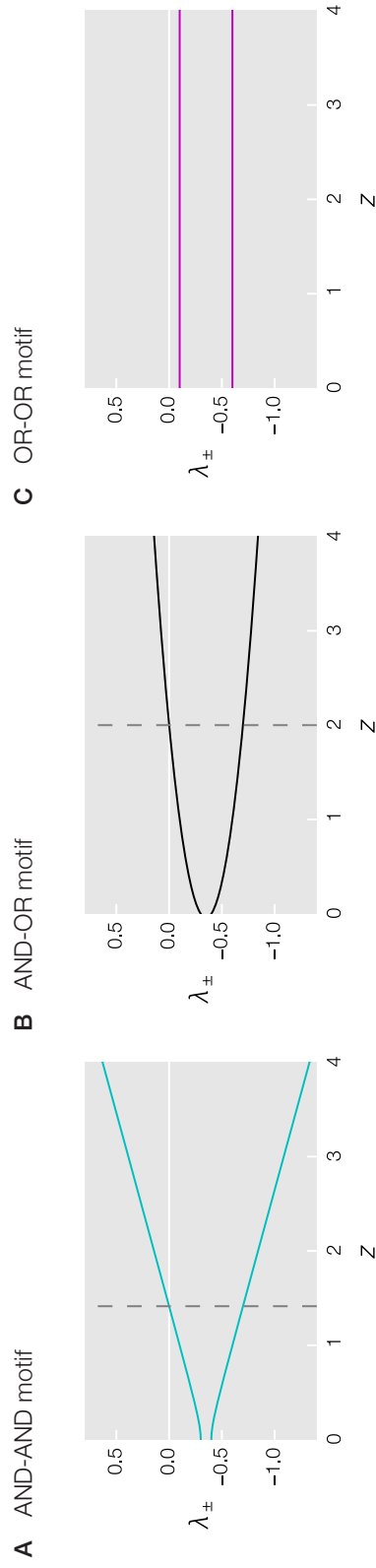
One eigenvalue,  $\lambda_-(\zeta)$ , is always negative as it is the sum of negative components. So the stability of the steady-state depends on the sign of  $\lambda_+(\zeta)$ . This is negative if and only if

$$\begin{aligned}
 a_r + b_r &> \sqrt{(a_r - b_r)^2 + 4a_1 b_1 \zeta} \\
 (a_r + b_r)^2 &> (a_r - b_r)^2 + 4a_1 b_1 \zeta \\
 (a_r + b_r)^2 - (a_r - b_r)^2 &> 4a_1 b_1 \zeta \\
 4a_r b_r &> 4a_1 b_1 \zeta \\
 a_r b_r &> a_1 b_1 \zeta
 \end{aligned}$$



**Figure 4.5** Steady-state values of  $x$  and  $y$  (the products of the two enzymes) as a function of  $z$  for the three pseudo-mass-action models listed in table 4.1, with  $a_0 = 0$ ,  $a_1 = a_2 = 0.2$ ,  $a_r = 0.4$ ,  $b_0 = 0.1$  and  $b_1 = b_2 = 0.3$ . Points of bifurcation, as  $z$  varies, are illustrated with a vertical, dashed, grey line.





**Figure 4.6** Eigenvalues as a function of  $z$  for the three pseudo-mass-action models listed in table 4.1, with  $a_0 = 0$ ,  $a_1 = a_2 = 0.2$ ,  $a_r = 0.4$ ,  $b_0 = 0.1$  and  $b_1 = b_2 = b_r = 0.3$ . Points of bifurcation, as  $z$  varies, are illustrated with a vertical, dashed, grey line.

Additionally, we can see by the denominator of the steady-state that, for all three models, the steady-state is stable if and only if the steady-state is positive (in both  $x$  and  $y$ ), meaning that positive initial conditions will give trajectories that always remain positive. (Recall that this was illustrated for the AND-OR motif in figure 4.1.)

The values of  $x$  and  $y$  as a function of  $z$  (recall—which denotes the relative state of activation of the system) have been plotted for the three models with arbitrary parameters in figure 4.5. The AND-AND and AND-OR models, but not the OR-OR model, undergo a bifurcation (manifesting as a discontinuity in figure 4.5) when  $z$  is increased or decreased causing the qualitative behaviour of the system to change. This occurs at

$$z = \sqrt{\frac{a_r b_r}{a_1 b_1}} \quad \text{for the AND-AND model;}$$

$$z = \frac{a_r b_r}{a_1 b_1} \quad \text{for the AND-OR model.}$$

The relationship between the sign of the eigenvalues and the value of  $z$  are illustrated for the same arbitrary parameters in figure 4.6.

All three models are capable of the unbounded growth caused by a negative, unstable steady-state – this was previously discussed as a major limitation of the original pseudo-mass-action model for the Arf/PLD/PI4P5K motif (the AND-OR motif). For the models of the AND-AND and AND-OR motifs this behaviour is accessible whenever the value of  $z$  is sufficiently large. For the model of the OR-OR motif this behaviour is only accessible if  $a_r b_r < a_1 b_1$ . Otherwise, all three models are capable of behaviour that is effectively bounded at large time, with a stable, positive steady-state which will attract all trajectories. It should be noted that while an attractive steady-state implies that the system is bounded, in practice we can choose  $z$  such that the steady-state is as far from the origin as we like.

Due to the bifurcation, the behaviour of the models of the AND-AND and AND-OR motifs, but not that of the OR-OR motif, could be described as switch-like, as the former are capable of qualitatively different behaviours depending on the relative strength of activation. (Note, however, that this behaviour is qualitatively distinct from hysteresis or bistability, which is also often described as ‘switch-like’.)

In the extreme case when there is no basal rate of production of either enzyme,  $a_0 = b_0 = 0$  then the steady-state for the pseudo-mass-action model of the AND-AND motif remains at the origin for all values of  $z$ . Furthermore, a trajectory that begins at the steady-state will remain there for all time, even if  $z$  later increases such that the stability of the steady-state changes. So the system can (at least

theoretically) get stuck in an unresponsive off state – a state that is not possible for the pseudo-mass-action models of the AND-OR and OR-OR motif. It should be understood that this state is not likely to be particularly physiologically relevant, as in a biological system there are likely always other sources of the product from other intracellular processes. But it does serve to highlight that while figure 4.6 suggests that AND-AND motif is the better (more tightly controlled) switch, it may have disadvantages that mean the AND-OR motif is more appropriate for a robust signal propagation.

Of course, these conclusions are based on very simple models, which all display unbounded behaviours. However, these results begin to suggest that the AND-OR motif may capture positive, useful characteristics of both of the two symmetric motifs. Next I would like to continue to use this comparative approach while altering the assumptions behind the construction of the models so as to attempt to develop empirical models that are inherently incapable of unbounded growth, and so remain physiologically plausible under variable stimulation.

### 4.3 Bounding growth using the Hill equation

One option for bounding growth is to replace the cross-regulation of each enzyme by some function that has a bounded output. Commonly used in biochemical modelling is the Hill function,

$$f_{\text{Hill}}(x, n) = \frac{x^n}{k^n + x^n}$$

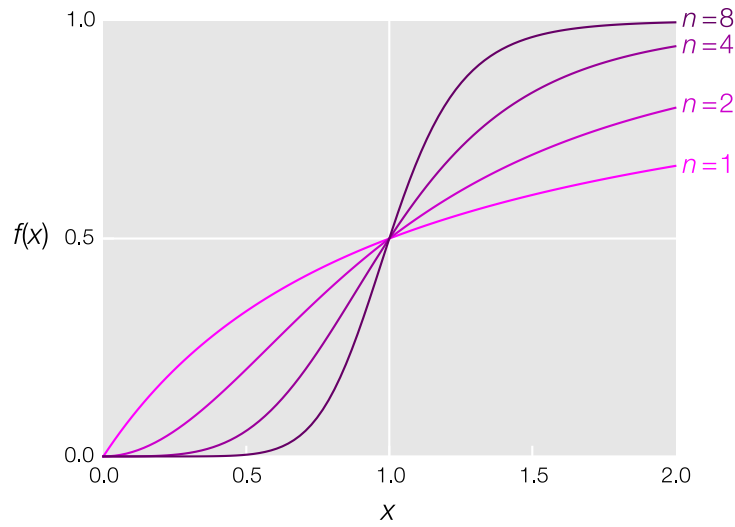
where  $k$  is a new parameter. This function is shown in figure 4.7 for different values of  $n$ . Observe that the Hill function is bounded as  $f_{\text{Hill}}(x) < 1$  for all values of  $x$ .

#### 4.3.1 Hill models: construction

Models using the Hill function, one for each of the AND-AND, AND-OR, and OR-OR motifs illustrated in figure 4.4, are listed in table 4.2. These models were generated from the systems of ODEs listed in table 4.1 by:

- for  $\frac{dx}{dt}$ , replacing  $y$  by  $f_{\text{Hill}}(y, n)$ ;
- for  $\frac{dy}{dt}$ , replacing  $x$  by  $f_{\text{Hill}}(x, m)$ .

(It is unnecessary to replace  $z$  by  $f_{\text{Hill}}(z)$ , as this remains the parameter by which the system can be manually stimulated.)



**Figure 4.7** The Hill function,  $f_{\text{Hill}}(x) = \frac{x^n}{k^n + x^n}$ , for different values of  $n$ , with  $k = 1$ .

Motif	ODEs
<b>A</b> AND-AND 	$\frac{dx}{dt} = a_0 + a_1 z \frac{y^n}{a_k^n + y^n} - a_r x$ $\frac{dy}{dt} = b_0 + b_1 z \frac{x^m}{b_k^m + x^m} - b_r y$
<b>B</b> AND-OR 	$\frac{dx}{dt} = a_0 + a_1 z \frac{y^n}{a_k^n + y^n} - a_r x$ $\frac{dy}{dt} = b_0 + b_1 \frac{x^m}{b_k^m + x^m} + b_2 z - b_r y$
<b>C</b> OR-OR 	$\frac{dx}{dt} = a_0 + a_1 \frac{y^n}{a_k^n + y^n} + a_2 z - a_r x$ $\frac{dy}{dt} = b_0 + b_1 \frac{x^m}{b_k^m + x^m} + b_2 z - b_r y$

**Table 4.2** Hill models for the three motifs (AND-AND; AND-OR; OR-OR). The variables  $x$  and  $y$  represent the concentrations of the two products (PA and PI(4,5)P<sub>2</sub>, respectively, for the AND-OR model). The variable  $z$  represents the concentration of active Arf.

In order to analyse this system we must choose values for  $n$  and  $m$ . Here I will only discuss the system with  $n = m = 1$  for which analytical solutions can be found. Analytical solutions for larger  $n$  and  $m$  may not exist, due to the greater nonlinearity of the system.

### 4.3.2 Hill models: analysis

The steady-states of the Hill model of the AND-OR motif are given by values of  $x$  and  $y$  that satisfy the equations

$$x = \frac{1}{a_r (a_k + y)} (a_0 a_k + a_0 y + a_1 y z)$$

$$0 = c_2 y^2 - c_1 y - c_0$$

where

$$c_2 = b_r (a_0 + a_1 z + a_r b_k)$$

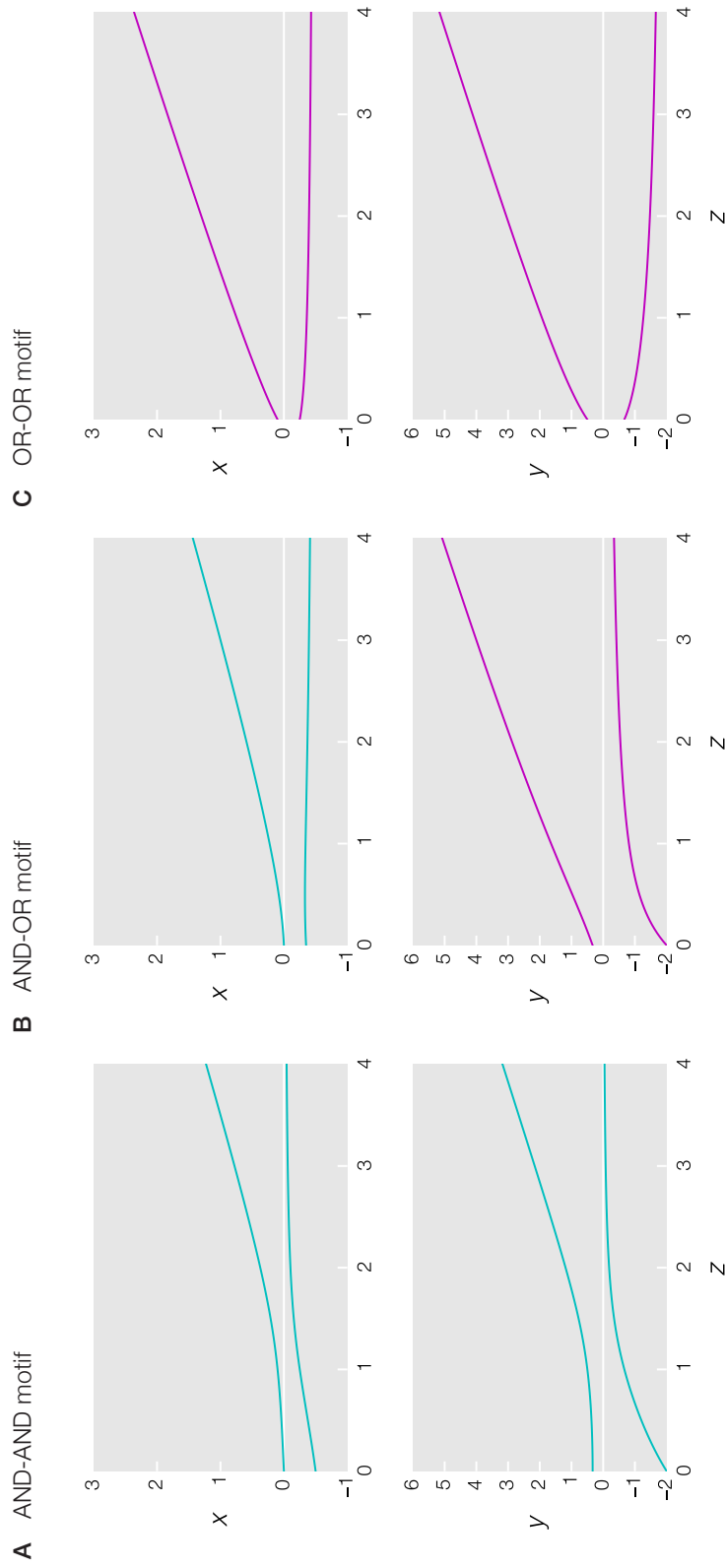
$$c_1 = a_1 b_2 z^2 + (a_0 b_2 + a_1 (b_0 + b_1) + a_r b_2 b_k) z + a_0 (b_0 + b_1 - a_k b_r) - a_r b_k (a_k b_r - b_0)$$

$$c_0 = a_k (a_r b_k) z + a_0 (b_0 + b_1) + a_r b_0 b_k + b_2 (a_0).$$

The steady-state solutions of the Hill models of the AND-AND and OR-OR motifs can each be described by equivalent equations:  $x$  as a function of  $y$ ; and  $y$  as the solutions to a quadratic equation with coefficients that are complicated functions of the parameters. So each of the models listed in table 4.2 have two steady-states, and so we can already see that these systems are qualitatively distinct from the previous iteration of the models. Due to the complexity of the coefficients, the full steady-state solutions to the Hill models are long and complicated functions of the parameters which are impractical to include in print, and which have limited mathematical tractability. Therefore, in general these systems will require a numerical investigation.

Numerical solutions for the steady-states of the models incorporating the Hill functions are illustrated with arbitrary parameters in figure 4.8. Note that for each model one of the two steady-states is always positive and one is always negative. To check the stability of these steady-states, we need to calculate the eigenvalues of the system – these are also complicated functions of the parameters. The steady-states of the system can be characterised numerically using the Python script:

```
1 from sympy import *
2 from empirical import *
```



**Figure 4.8** Steady-states values of  $x$  and  $y$  (the products of thw two enzymes) as a function of  $z$  for the three Hill models listed in table 4.2, with  $a_0 = 0$ ,  $a_1 = a_2 = 0.2$ ,  $a_r = 0.4$ ,  $a_k = 2$ ,  $b_0 = 0.1$ ,  $b_1 = b_2 = b_r = 0.3$ ,  $b_k = 0.5$ , and  $n = m = 1$ .

```

3
4 # symbolic setup
5 ak, bk = symbols('ak bk', positive = True)
6
7 # model setup
8 hill_aa = [a0 + a1*y*z/(ak + y) - ar*x,
9            b0 + b1*x*z/(bk + x) - br*y]
10 hill_ao = [a0 + a1*y*z/(ak + y) - ar*x,
11            b0 + b1*x/(bk + x) + b2*z - br*y]
12 hill_oo = [a0 + a1*y/(ak + y) + a2*z - ar*x,
13            b0 + b1*x/(bk + x) + b2*z - br*y]
14
15 param_hill = {z:1, a0:0, a1:0.2, a2:0.2, ar:0.4, ak:2,
16              b0:0.1, b1:0.3, b2:0.3, br:0.3, bk:0.5}
17
18 # characterise
19 characterise('AND-AND', hill_aa, param_hill)
20 characterise('AND-OR', hill_ao, param_hill)
21 characterise('OR-OR', hill_oo, param_hill)

```

This gives the output:

```

AND-AND: (2 solutions)
  x=-0.25, y=-0.67, ev1=-1.09, ev2=+0.39
  x=+0.10, y=+0.50, ev1=-0.52, ev2=-0.18

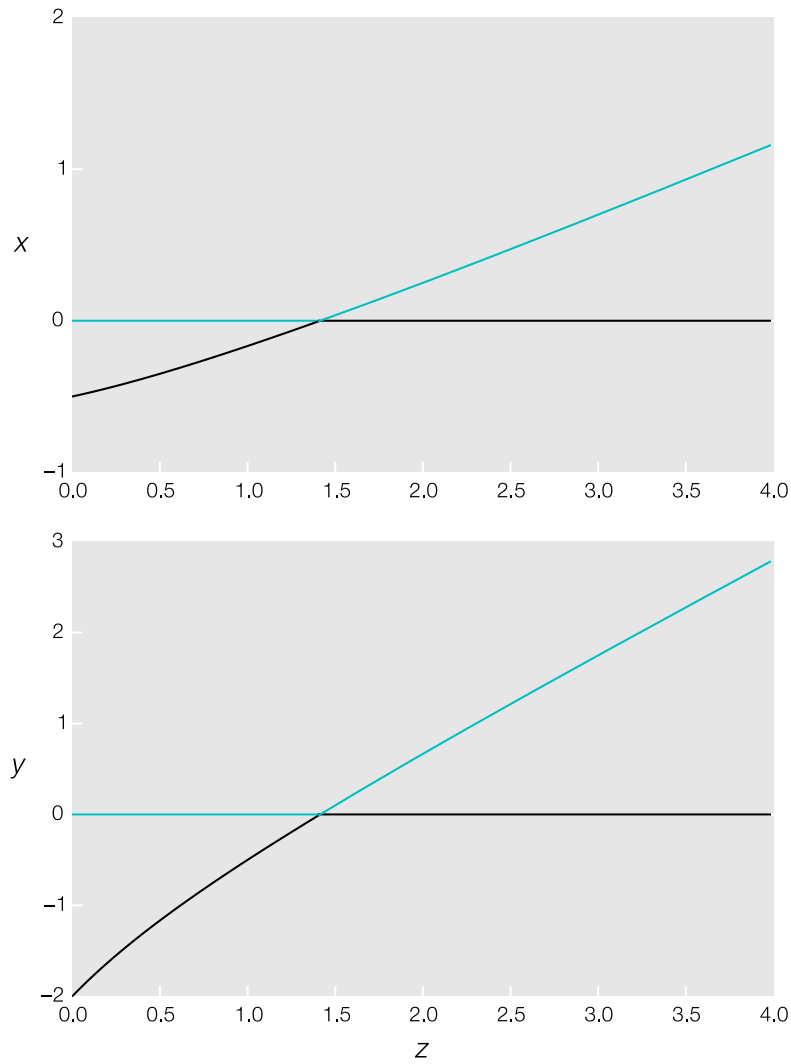
AND-OR: (2 solutions)
  x=-0.34, y=-0.81, ev1=+0.95, ev2=-1.65
  x=+0.23, y=+1.64, ev1=-0.24, ev2=-0.46

OR-OR: (2 solutions)
  x=-0.36, y=-1.27, ev1=+2.05, ev2=-2.75
  x=+0.75, y=+1.93, ev1=-0.28, ev2=-0.42

```

The negative steady-state is unstable and the positive steady-state is stable in all three models when  $z = 1$ . As there are no bifurcations shown in figure 4.8, these stabilities remain for  $0 \leq z \leq 4$ . As the steady-states appear to be either bounded by the lines  $x = 0$  and  $y = 0$  or diverge to infinity these stabilities also likely hold for all  $z$ . With my previous definition, a positive stable steady-state gives dynamics which are bounded at large time. So we see it is possible to choose parameters so that these models are incapable of unbounded growth.

I would like to again consider whether the two theoretical motifs can help me to understand the real Arf/PLD/PI4P5K motif. First—considering the AND-AND motif. In figure 4.8A, the positive steady-state has a nonlinear response to increases in the value of  $z$ : it is relatively insensitive to changes in the value of  $z$ , if  $z$  remains small;



**Figure 4.9** Transcritical bifurcation in the Hill model of the AND-AND motif, with  $a_0 = b_0 = 0$ ,  $a_1 = 0.2$ ,  $a_r = 0.4$ ,  $a_k = 2$ ,  $b_1 = b_r = 0.3$ , and  $b_k = 0.5$ . The unstable steady-state is drawn in black, the stable steady-state in blue.



and sensitive (a near linear response) to changes in the value of  $z$ , if  $z$  takes larger values.

It is possible to show this nonlinearity more clearly for the system with zero basal activity,  $a_0 = b_0 = 0$ . In this case the steady-states take a relatively simple form,

$$(0, 0) \text{ and } \left( \frac{a_1 b_1 z^2 - a_r b_r a_k b_k}{a_r (a_k b_r + b_1 z)}, \frac{a_1 b_1 z^2 - a_r b_r a_k b_k}{b_r (a_1 z + a_r b_k)} \right).$$

Depending on the parameters, the second steady-state can be positive or negative. It is positive whenever

$$a_r b_r a_k b_k < a_1 b_1 z^2.$$

(Note that this has some similarity with the point of bifurcation of the pseudo-mass-action model of the AND-AND motif.)

The eigenvalues associated with the zero steady-state are given by

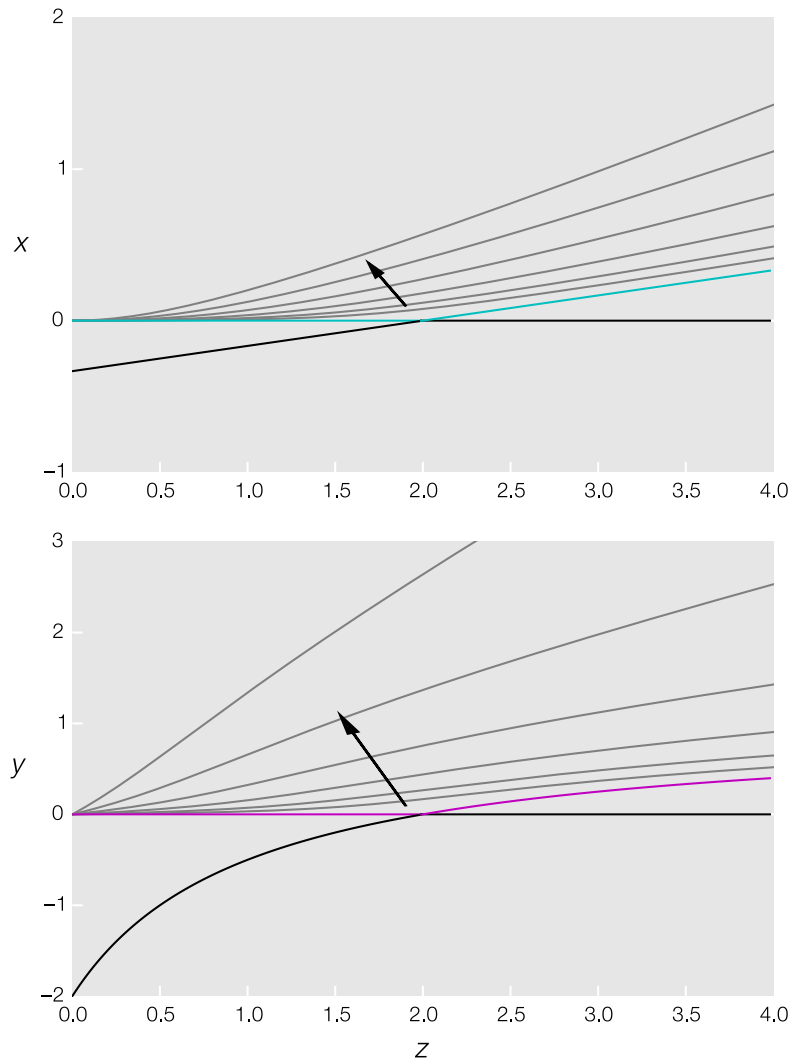
$$\lambda_{\pm} = \frac{1}{2} \left[ -(a_r + b_r) \pm \sqrt{(a_r - b_r)^2 + 4 \frac{a_1 b_1}{a_k b_k} z^2} \right].$$

The eigenvalue  $\lambda_-$  is always negative. The eigenvalue  $\lambda_+$  is positive whenever

$$\begin{aligned} a_r + b_r &> \sqrt{(a_r - b_r)^2 + 4 \frac{a_1 b_1}{a_k b_k} z^2} \\ (a_r + b_r)^2 &> (a_r - b_r)^2 + 4 \frac{a_1 b_1}{a_k b_k} z^2 \\ (a_r + b_r)^2 - (a_r - b_r)^2 &> 4 \frac{a_1 b_1}{a_k b_k} z^2 \\ 4a_r b_r &> 4 \frac{a_1 b_1}{a_k b_k} z^2 \\ a_r b_r a_k b_k &> a_1 b_1 z^2. \end{aligned}$$

So the zero steady-state switches between stability and instability as it passes through a point of bifurcation. This also corresponds to the point at which the non-zero steady-state becomes positive. This is characteristic of a *transcritical* bifurcation, and implies that the non-zero steady-state also changes stability at this point (in the other direction). So, depending on the value of  $z$ , the system either has a zero or a positive stable steady-state. This can be observed in figure 4.9.

Thus this model suggests that the AND-AND motif can be characterised by a mechanism that is insensitive to small concentrations of the activator, while remaining sensitive to high concentrations. This mechanism implies a system which



**Figure 4.10** Transcritical bifurcation in the Hill model of the AND-OR motif, with  $a_0 = b_0 = b_2 = 0$ ,  $a_1 = a_2 = 0.2$ ,  $a_r = 0.4$ ,  $a_k = 2$ ,  $b_1 = b_r = 0.3$ , and  $b_k = 0.5$ . The unstable steady-state is drawn in black, the stable steady-state in blue or magenta. Grey lines show the positive steady-state under increasing values of  $b_2$  in the direction shown.

requires sufficient stimulation before it will propagate the signal – the system can either be in an ‘off’ state or, if the activation is above a threshold value, in a (graded) ‘on’ state. This could be useful for filtering low levels of noise arising upstream in the signalling pathway.

Unfortunately, the Hill model of the OR-OR motif retains a complicated solution even with  $a_0 = b_0 = 0$ . So instead I will only comment on numerical results shown in figure 4.8C, where the values of both  $x$  and  $y$  have a near linear response to an increase in  $z$ . This could be explained by the OR-OR motif acting to propagate an upstream signal with little or no modulation, though possibly with an increase or decrease in amplitude.

I note that the numerical solutions for the Hill model of the AND-OR motif, as shown in figure 4.8B, has characteristics of both theoretical motifs. The variable  $y$ , corresponding to the ‘OR’ enzyme, certainly has a similar, linear response to changes in the value of  $z$  as displayed by both variables in the OR-OR model. The variable  $x$ , corresponding to the ‘AND’ enzyme, has a less of a clear-cut resemblance to the variables in the AND-AND model, but some nonlinearity can still be observed.

The Hill model of the AND-OR motif can be mathematically analysed under the constraints of no basal activation,  $a_0 = b_0 = 0$ , and no activation of the rate of production of  $y$  by  $z$ , so  $b_2 = 0$ . The steady-states of this system are

$$(0, 0) \text{ and } \left( \frac{a_1 b_1 z - a_r b_r a_k b_k}{a_r (a_k b_r + b_1)}, \frac{a_1 b_1 z - a_r b_r a_k b_k}{b_r (a_1 z + a_r b_k)} \right).$$

The eigenvalues associated with the *zero* steady-state are

$$\lambda_{\pm} = \frac{1}{2} \left[ -(a_r + b_r) \pm \sqrt{(a_r - b_r)^2 + 4 \frac{a_1 b_1}{a_k b_k} z} \right]$$

one of which ( $\lambda_-$ ) is always negative, the other ( $\lambda_+$ ) is negative whenever

$$a_r + b_r > \sqrt{(a_r - b_r)^2 + 4 \frac{a_1 b_1}{a_k b_k} z}$$

$$\Rightarrow a_r b_r a_k b_k > a_1 b_1 z$$

which corresponds exactly to the requirement for the non-zero steady-state to be negative. So this system is also able to undergo a transcritical bifurcation. This is shown in figure 4.10.

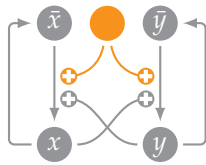
Recall that this bifurcation occurs when  $b_2 = 0$ , yet I remain primarily interested in the case when  $b_2 > 0$ , and so in figure 4.10 I have illustrated the effect of increasing  $b_2$ . Recall (from table 4.2) that the term  $b_2 z$  describes the effect of the concentration

of activator on the rate of change of the variable  $y$ . So as  $b_2$  increases, this linear response to increasing  $z$  begins to dominate, over the nonlinear response to the value of  $x$ . This can be seen in the lower plot in figure 4.10. Comparatively, as it is not directly reliant on this parameter, increasing the value of  $b_2$  has a smaller effect on the magnitude and qualitative behaviour of  $x$ , which maintains at least part of the nonlinear response distinctive of the bifurcation occurring when  $b_2 = 0$ .

So we can begin to see evidence for two distinct responses to stimulation in the behaviour of the two enzyme, hence concentration of the two products, of the Arf/PLD/PI4P5K signalling motif. Next, I will show that equivalent results can be observed when a different mechanism of bounding the growth of the system is assumed.

## 4.4 Bounding growth using conservation of mass

A second, alternative approach to bounding the growth of the system is to limit the total amount of substrate and product that can be present, using conservation of mass terms. I introduce new variables  $\bar{x}$  and  $\bar{y}$  corresponding to the substrates of the two modelled enzymes. Consider either or both of the forward reactions to be paired with a reverse reaction.



### 4.4.1 Conservation of mass models: construction

The equations that describe the systems with added conservation of mass, are in general

$$\frac{dx}{dt} = f_x(y, z) \times \bar{x} - a_r x \quad (4.7)$$

$$\frac{dy}{dt} = f_y(x, z) \times \bar{y} - b_r y \quad (4.8)$$

where

$$f_x(y, z) = \begin{cases} a_0 + a_1 xz & \text{when } x \text{ corresponds to an 'AND' enzyme} \\ a_0 + a_1 x + a_2 z & \text{when } x \text{ corresponds to an 'OR' enzyme} \end{cases}$$

$$f_y(x, z) = \begin{cases} b_0 + b_1 y z & \text{when } y \text{ corresponds to an 'AND' enzyme} \\ b_0 + b_1 y + b_2 z & \text{when } y \text{ corresponds to an 'OR' enzyme} \end{cases}$$

Equations 4.7 and 4.8 reduce to the original pseudo-mass-action models in table 4.1 when  $\bar{x} = 1$  and  $\bar{y} = 1$ ,

$$\frac{dx}{dt} = f_x(y, z) - a_r x \quad (4.9)$$

$$\frac{dy}{dt} = f_y(x, z) - b_r y. \quad (4.10)$$

To introduce conservation of mass, we set

$$x_t = x + \bar{x}$$

$$y_t = y + \bar{y}.$$

Using these to replace  $\bar{x}$  and  $\bar{y}$  in equations 4.7 and 4.8 gives

$$\frac{dx}{dt} = f_x(y, z) \times (x_t - x) - a_r x \quad (4.11)$$

$$\frac{dy}{dt} = f_y(x, z) \times (y_t - y) - b_r y. \quad (4.12)$$

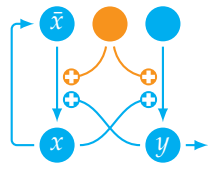
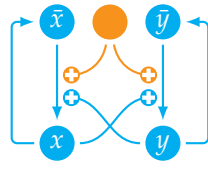
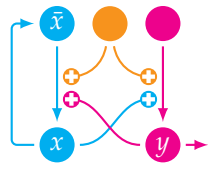
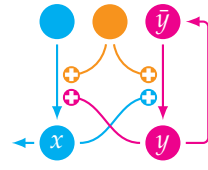
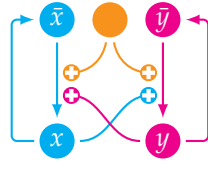
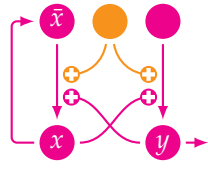
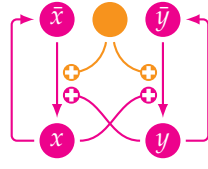
I will consider systems in which either one or both of the enzymes are restricted by conservation of mass: one or two of equations 4.11 and 4.12, the remainder from equations 4.9 and 4.10. With the choice of ‘AND’ (PLD-like) and ‘OR’ (PI4P5K-like) behaviours for each enzyme—and ignoring duplicate (symmetric, with  $x \leftrightarrow y$ ) motifs—we can generate the seven motifs shown in table 4.3.

#### 4.4.2 Conservation of mass models: analysis

It is possible to find analytical solutions for the steady-states and eigenvalues of the systems in table 4.3, however these are once more long and complicated functions of the parameters, and difficult to analyse mathematically. Figure 4.11 shows numerical solutions for these steady-states (with arbitrary parameters), which can be characterised with the Python script:

```

1 from sympy import *
2 from empirical import *
3
4 xt, yt = symbols('xt yt', positive = True)
5
6 # motif setup
    
```

Motif	ODEs	Example phase plane ( $x$ vs. $y$ )
A AND-AND-X		$\frac{dx}{dt} = (a_0 + a_1 y z)(x_t - x) - a_r x$ $\frac{dy}{dt} = b_0 + b_1 x z - b_r y$
B AND-AND-XY		$\frac{dx}{dt} = (a_0 + a_1 y z)(x_t - x) - a_r x$ $\frac{dy}{dt} = (b_0 + b_1 x z)(y_t - y) - b_r y$
C AND-OR-X		$\frac{dx}{dt} = (a_0 + a_1 y z)(x_t - x) - a_r x$ $\frac{dy}{dt} = b_0 + b_1 x + b_2 z - b_r y$
D AND-OR-Y		$\frac{dx}{dt} = a_0 + a_1 y z - a_r x$ $\frac{dy}{dt} = (b_0 + b_1 x + b_2 z)(y_t - y) - b_r y$
E AND-OR-XY		$\frac{dx}{dt} = (a_0 + a_1 y z)(x_t - x) - a_r x$ $\frac{dy}{dt} = (b_0 + b_1 x + b_2 z)(y_t - y) - b_r y$
F OR-OR-X		$\frac{dx}{dt} = (a_0 + a_1 y + a_2 z)(x_t - x) - a_r x$ $\frac{dy}{dt} = b_0 + b_1 x + b_2 z - b_r y$
G OR-OR-XY		$\frac{dx}{dt} = (a_0 + a_1 y + a_2 z)(x_t - x) - a_r x$ $\frac{dy}{dt} = (b_0 + b_1 x + b_2 z)(y_t - y) - b_r y$

**Table 4.3** Models of the motifs with one or more reverse reactions. Phase portraits drawn using the parameters  $a_0 = 0$ ,  $a_1 = a_2 = 0.2$ ,  $a_r = 0.4$ ,  $b_0 = 0.1$ ,  $b_1 = b_2 = b_r = 0.3$ ,  $z = 1$ , and  $x_t = y_t = 1$  whenever applicable. Note that the maximum values  $x_t$  and  $y_t$  only constrain systems in which they are present. The variables  $x$  and  $y$  represents the concentrations of the two enzymes (PA and PI(4,5)P<sub>2</sub>, respectively, for the AND-OR models). The variable  $z$  represents the concentration of active Arf.

```

7 fa, fb = symbols('f_x f_y', positive=True)
8 f_opts = [{"name":"AND-AND", fa : a0 + a1*y*z,
9           fb : b0 + b1*x*z},
10          {"name":"AND-OR", fa : a0 + a1*y*z,
11           fb : b0 + b1*x + b2*z},
12          {"name":"OR-OR", fa : a0 + a1*y + a2*z,
13           fb : b0 + b1*x + b2*z}]
14
15 # conservation of mass setup
16 xp, yp = symbols('x_p y_p', positive=True)
17 p_opts = [{"name":"-X", xp : xt-x, yp : 1},
18          {"name":"-Y", xp : 1, yp : yt-y},
19          {"name":"-XY", xp : xt-x, yp : yt-y}]
20
21 # generate all combinations
22 # (including symmetric duplicates)
23 com_sys = {(f_opt['name'] + p_opt['name']) : \
24            [eq.subs(p_opt).subs(f_opt)
25             for eq in [fa*xp - ar*x, fb*yp - br*y]]
26            for p_opt in p_opts for f_opt in f_opts}
27
28 # arbitrary parameters
29 param = {z:1, xt:1, yt:1,
30          a0:0, a1:0.2, a2:0.2, ar:0.4,
31          b0:0.1, b1:0.3, b2:0.3, br:0.3}
32
33 # output
34 sys_names = ['AND-AND-X', 'AND-AND-XY', 'AND-OR-X', \
35             'AND-OR-Y', 'AND-OR-XY', 'OR-OR-X', 'OR-OR-XY']
36 for name in sys_names:
37     characterise(name, com_sys[name], param)

```

This gives the output:

```

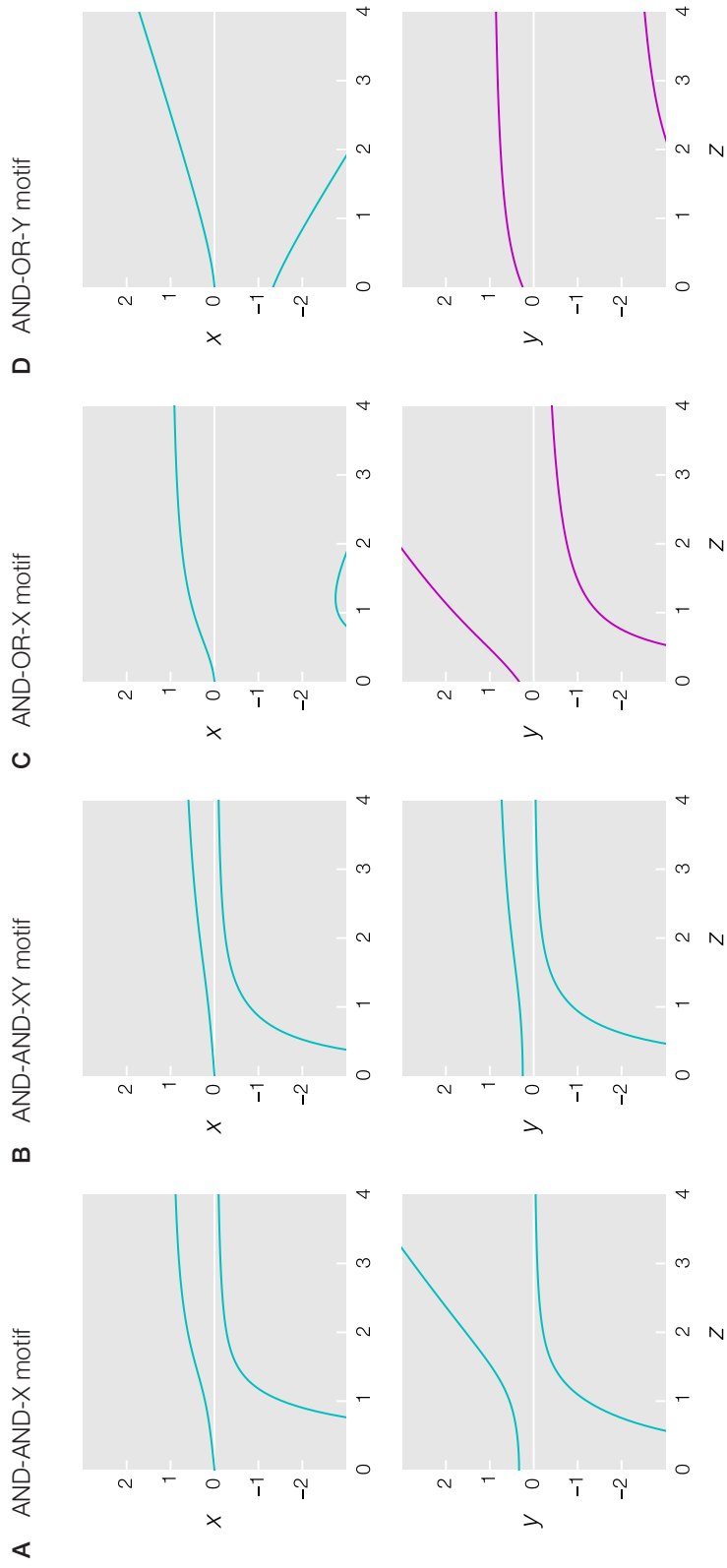
AND-AND-X: (2 solutions)
  x=-1.55, y=-1.22, ev1=+0.17, ev2=-0.63
  x=+0.22, y=+0.55, ev1=-0.16, ev2=-0.65

AND-AND-XY: (2 solutions)
  x=-0.80, y=-0.89, ev1=-0.64, ev2=+0.26
  x=+0.14, y=+0.32, ev1=-0.64, ev2=-0.26

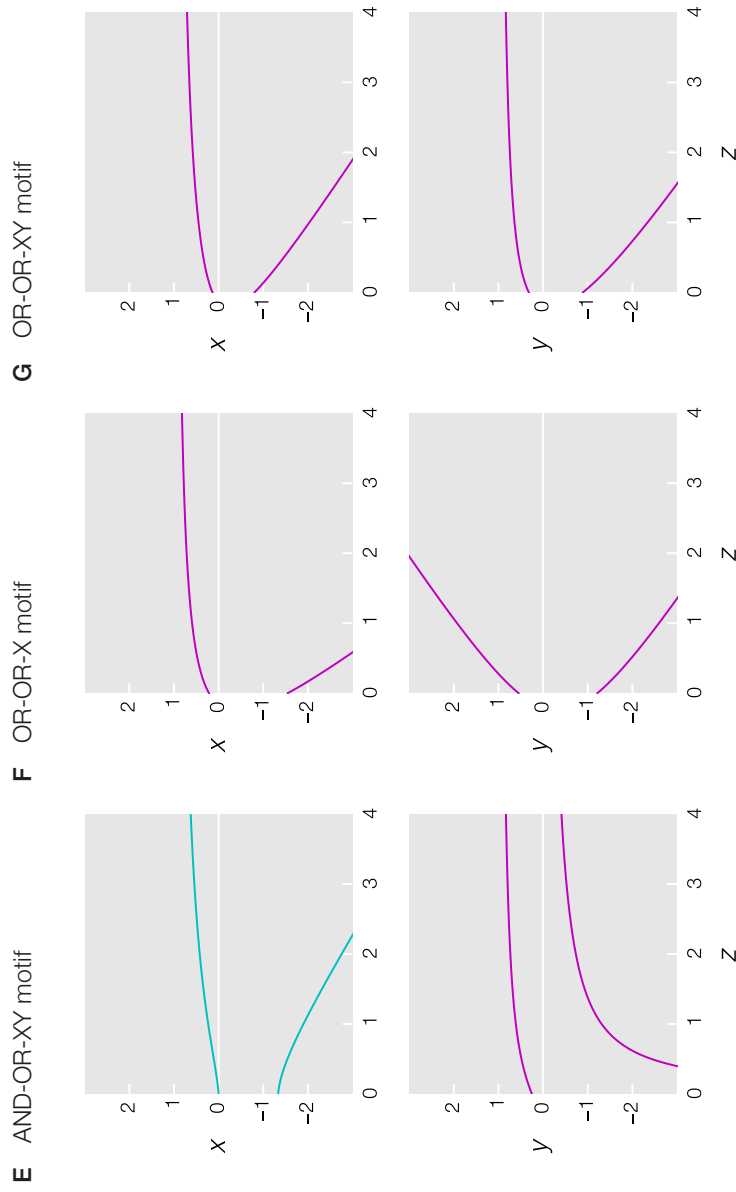
AND-OR-X: (2 solutions)
  x=+0.47, y=+1.81, ev1=-0.82, ev2=-0.24
  x=-2.81, y=-1.47, ev1=-0.69, ev2=+0.29

AND-OR-Y: (2 solutions)
  x=-2.14, y=-4.29, ev1=-0.82, ev2=+0.36

```







**Figure 4.11** Steady-states values of  $x$  and  $y$  (the products of thw two enzymes) as a function of  $z$  for the seven conservation of mass models listed in table 4.3, with  $a_0 = 0$ ,  $a_1 = a_2 = 0.2$ ,  $a_r = 0.4$ ,  $b_0 = 0.1$ ,  $b_1 = b_2 = b_r = 0.3$ , and  $x_t = y_t = 1$ .

```

x=+0.31, y=+0.62, ev1=-0.84, ev2=-0.35

AND-OR-XY: (2 solutions)
x=-1.90, y=-1.31, ev1=+0.50, ev2=-0.77
x=+0.23, y=+0.61, ev1=-0.46, ev2=-0.83

OR-OR-X: (2 solutions)
x=+0.59, y=+1.93, ev1=-1.02, ev2=-0.27
x=-3.93, y=-2.59, ev1=-0.75, ev2=+0.36

OR-OR-XY: (2 solutions)
x=-2.03, y=-2.34, ev1=-0.89, ev2=+0.67
x=+0.45, y=+0.64, ev1=-0.90, ev2=-0.66

```

So—exactly as for the Hill models—we have a series of models which display a negative, unstable steady-state and a positive, stable steady-state, which (as we do not observe any bifurcations in figure 4.11) appear to maintain this stability as  $z$  varies (given these parameters).

By construction, we know that the steady-state solution will be bounded for a variable whenever we subject that variable to conservation of mass, as then  $x < x_t$  and/or  $y < y_t$ . These constraints can be observed (where included) for the numerical solutions in figure 4.11. Increasing  $z$ : ‘AND’ enzymes show an approximately sigmoidal response; ‘OR’ enzymes show an approximately hyperbolic response.

Again from figure 4.11, variables not subjected to conservation of mass display similar behaviours to those displayed for the Hill models. Increasing  $z$ :

- ‘AND’ enzymes show the nonlinear response to an increasing signal, insensitive to small concentrations of the activator, and remaining sensitive to high concentrations.
- ‘OR’ enzymes show a near linear response.

So I have evidence that both of the methods I have used to bound the growth of the systems can result in similar (steady-state) characteristics. This implies I may be able to make hypotheses about the underlying biochemistry of the cross-talk motif that are somewhat robust to the choice of model.

From the simulations in figure 4.11, we can again observe that the models of the AND-OR motif display behaviours consistent with the symmetric models, both when conservation of mass is present and when it is absent. I have already noted that the complexity of the steady-state solutions means that they are in general

not amenable to mathematical analysis, but I would like to replicate the analysis performed on the Hill model of the AND-OR motif on the conservation of mass model of the AND-OR-Y motif.

I have chosen to focus on this model as I believe that this best represents the biochemistry of the Arf/PLD/PI4P5K motif. This motif has an additional reversible reaction on the variable  $y$ , corresponding to the concentration of PI(4,5)P<sub>2</sub> in the system. From chapter 1 we know that this reverse reaction does exist: either mediated by PI(4,5)P<sub>2</sub> 5-phosphatases (figure 2.5), or through the slower multiple steps involved in the PI cycle (figure 2.9). Comparatively, PC concentrations are known to be high in all cellular membranes (van Meer, Voelker, & Feigenson 2008; van Meer & de Kroon 2011), and so PC is less likely to be rate limiting.

From table 4.3, this motif is described by the equations

$$\frac{dx}{dt} = a_0 + a_1 y z - a_r x \quad (4.13)$$

$$\frac{dy}{dt} = (b_0 + b_1 x + b_2 z)(y_t - y) - b_r y. \quad (4.14)$$

At steady-state equation 4.13 implies

$$x = \frac{a_0 + a_1 y z}{a_r} \quad (4.15)$$

which can be substituted into equation 4.14, and rearranged, to give a quadratic equation. Without loss of generality (as it is only a scaling factor, in arbitrary units) we can assume  $y_t = 1$ . So the  $y$  component of two steady-states are described by solutions to

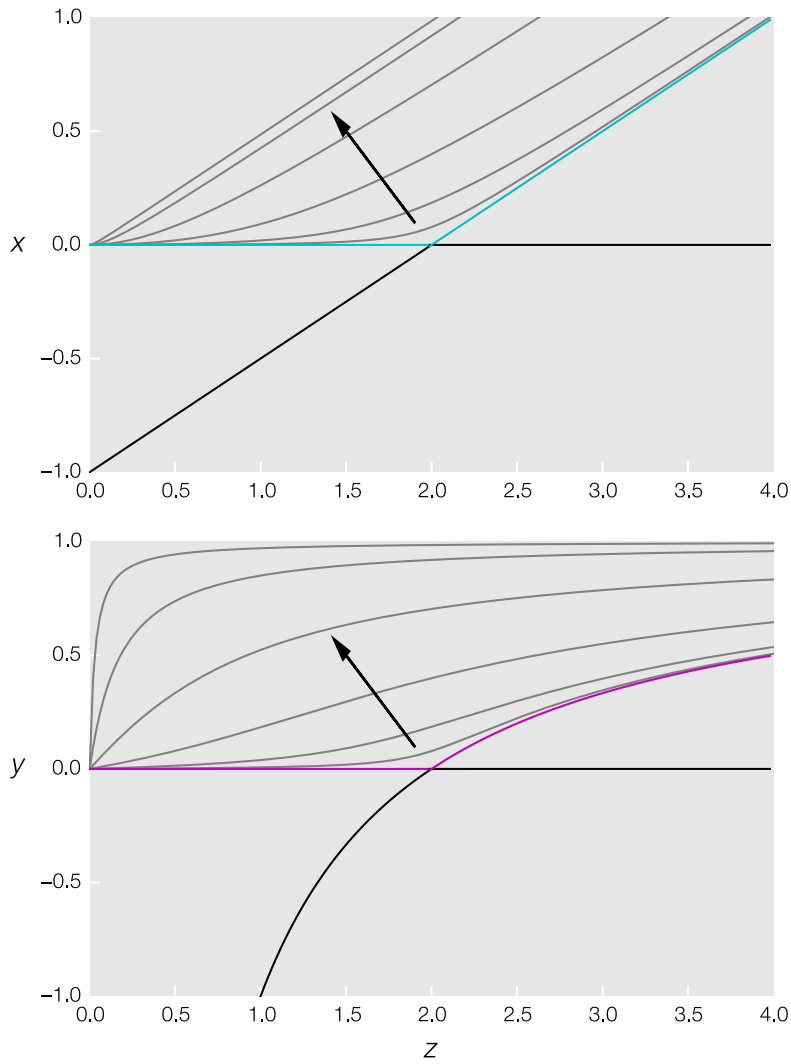
$$0 = a_1 b_1 z y^2 + (a_0 b_1 + (b_0 + b_r) a_r + (a_r b_2 - a_1 b_1) z) y - (a_0 b_1 + b_0 + b_2 z) a_r.$$

Solutions to this quadratic remain complicated functions of the parameters. However, exactly as for the Hill model, it is possible to mathematically characterise the system with no basal rate of production of either product,  $a_0 = b_0 = 0$ , and no activation of the rate of production of  $y$  by  $z$ ,  $b_2 = 0$ . In this case the steady-states are

$$(0, 0) \text{ and } \left( \frac{a_1}{a_r} z - \frac{b_r}{b_1}, 1 - \frac{a_r b_r}{a_1 b_1 z} \right).$$

The eigenvalues associated with the zero steady-state are given by

$$\lambda_{\pm} = \frac{1}{2} \left[ -(a_r + b_r) \pm \sqrt{(a_r - b_r)^2 + 4a_1 b_1 z^2} \right]$$



**Figure 4.12** Transcritical bifurcation in the conservation of mass model of the AND-OR-Y motif, with  $a_0 = b_0 = b_2 = 0$ ,  $a_1 = a_2 = 0.2$ ,  $b_1 = b_r = 0.3$ , and  $y_t = 1$ . The unstable steady-state is drawn in black, the stable steady-state in blue or magenta. Grey lines show the positive steady-state under increasing values of  $b_2$  in the direction shown.

which are identical to the eigenvalues for the original pseudo-mass-action model of the AND-OR motif, so the zero steady-state is stable for small  $z$ , unstable for large  $z$ , and has a point of bifurcation at

$$z = \frac{a_r b_r}{a_1 b_1}.$$

This point of bifurcation corresponds to the point at which the non-zero steady-state changes stability, and so again we observe a transcritical bifurcation, as shown in figure 4.12.

The effect of increasing  $b_2$  is demonstrated in figure 4.12. Whenever  $b_2 > 0$  there is no bifurcation, but rather a stable steady-state that remains positive and a negative unstable steady-state (that has not been drawn). This figure is equivalent to that for the Hill model in figure 4.10, however note that now  $y$  respects conservation of mass ( $y \leq y_t = 1$ ). The behaviours shown in figure 4.12 can be categorised as follows:

- For the variable  $x$ : when  $b_2$  is small it maintains much of the behaviour characteristic of the bifurcation (insensitive to small, sensitive to large values of  $z$ ); and when  $b_2$  is large the original signal is propagated with little modulation. In this model there is no limit to the size we can make the  $x$  component of the steady-state.
- For the variable  $y$ : for moderate values of  $b_2$  the original signal is propagated with little modulation; and when  $b_2$  is large, changes in small values of  $z$  are amplified, before reaching a maximum defined by conservation of mass.

So I hypothesise that this motif acts to propagate two distinct signals from a single initial input signal:

- The concentration of PA ( $x$ ) only increases significantly (and so propagates a signal) once a threshold concentration of Arf ( $z$ ) is reached. Thus PLD would be able to filter noise resulting from fluctuations in small concentrations of Arf.
- The concentration of PI(4,5)P<sub>2</sub> ( $y$ ) is sensitive to the signal originating from Arf ( $z$ ). Thus PI4P5K is sensitive to even small changes in the strength of the upstream signal.

It would be possible to analyse the other models in table 4.3 in a similar way to this model, but these would require individual consideration, and this will not be attempted here. Finally, note once more that these conclusions are currently based

on a single parameterisation of the system, and so it remains possible that given other parameterisations the model will display alternative behaviours. However, this provides sufficient evidence to suggest the hypothesis of differential signal propagation by the two products of the Arf/PLD/PI4P5K motif.

## 4.5 Discussion

I have constructed a series of empirical models of the Arf/PLD/PI4P5K signalling motif, and related theoretical motifs, in order to begin to explore behaviours which result from the differential mechanisms of regulation of the two enzymes, and cross-talk from the products. Furthermore, I have suggested novel hypotheses concerning the signalling properties of the system.

The models were based on qualitative descriptions of the regulation of the enzymes PLD and PI4P5K, and my observation that these have relation to logic gates: ‘AND’ for PLD; and ‘OR’ for PI4P5K. Three motifs were considered: the Arf/PLD/PI4P5K motif, also termed the AND-OR motif; an AND-AND motif, where both enzymes were regulated like PLD; and an OR-OR motif, where both enzymes were regulated like PI4P5K. Each model consists of two equations describing the net rate of change of the two products (PA and PI(4,5)P<sub>2</sub>; or arbitrary species). Production of the products is due to the catalytic rate of the enzymes, in part specified by the concentration of the activators, and removal of the products is via a sink.

A first series of models was termed the *pseudo-mass-action* models (section 4.2.2). These were based on a previously described model of the Arf/PLD/PI4P5K signalling motif (Stanley 2011), which here I have used an alternative approach to showing that it is capable of unbounded growth. Here:

- The catalytic rate of PLD, and equivalent hypothetical enzymes, was modelled using multiplication of the two regulator concentrations, scaled by a rate constant.
- The catalytic rate of PI4P5K, and equivalent hypothetical enzymes, was modelled using addition of the two regulator concentrations, each scaled by a rate constant.

Subsequent series of models were constructed in an attempt to ensure the system remained bounded:

- The Hill models (section 4.3.1), where the concentration of the products regulating the enzymes was replaced by a suitable Hill function.

- The conservation of mass models (section 4.4.1), where the catalytic rates of the enzymes are bounded by the total amount of product and substrate in the system – due to non-symmetric conservation of mass this generated seven models.

Where possible mathematical analysis (as per section 3.1) was used to investigate these models.

With consideration of the hypothetical motifs, plausible hypotheses of the signalling behaviour of the Arf/PLD/PI4P5K motif were discussed. These hypothetical motifs have allowed me to utilise an additional perspective and so comment on plausible advantages for the asymmetry in the Arf/PLD/PI4P5K motif. Together, this has led to novel hypotheses about the signalling properties of the motif using simple empirical models that are sometimes amenable to complete mathematical characterisation.

The pseudo-mass-action model of the AND-AND and AND-OR motifs—but not the model of the OR-OR motif—were found to undergo a bifurcation as the concentration of the activator is varied. Thus these models suggest switch-like behaviour for these two motifs, controlled in relation to a critical threshold value: above which the system is turned on; and below which the system is turned off. While the model of the AND-AND motif is the better (more precise) switch there is evidence that it could get stuck in an unresponsive off state. This gives evidence for a hypothesis that the AND-OR motif is a compromise – displaying less precise switch-like behaviour than the model of the AND-AND motif, but more robust.

A different story—suggesting an alternative hypothesis for the signalling properties of the system—arises from the analysis of the Hill models and the conservation of mass models. (The difference between these results is likely indicative of the relatively arbitrary nature of the empirical modelling approach.) In these models, all ‘AND’ (PLD-like) enzymes have similar characteristics, and all ‘OR’ (PI4PK-like) enzymes have similar characteristic. These behaviours appear to mostly hold for both symmetric and asymmetric models. In moderate and non-zero parameterisations, these models imply the following behaviour for these motifs.

PLD and other ‘AND’ enzymes seem to be characterised by an ability to produce a nonlinear response to the concentration of the activator – initially relatively insensitive to changes in the concentration of activator until sufficient activator is present. In the most extreme case, the signal is only propagated once a threshold concentration is reached. This behaviour has noise-reduction properties – fluctuations in the concentration of the activator at low activation levels are not propagated

Arf concentration	zero	low	high
PA production	no	no	yes
PI(4,5)P <sub>2</sub> production	no	yes	yes

**Table 4.4** Inputs, in the form of concentrations of Arf, and associated outputs, in the form of PA and PI(4,5)P<sub>2</sub> production rates, as implied by the Hill and conservation of mass models of the Arf/PLD/PI4P5K motif (the AND-OR) motif. Thus a single input signal (the Arf concentration) will give two distinct output signals (the concentrations of PA and PI(4,5)P<sub>2</sub>).



downstream.

PI4P5K and other 'OR' enzymes seem to be characterised by an ability to propagate signals *as-is*, with little or no modulation. Though it should be noted that modulation of a signal can still occur in extremes (for example, no activation of the enzyme by the activator), or when conservation of mass is present – in which case the concentration of the product has a strict upper bound.

The Arf/PLD/PI4P5K motif (as the AND-OR motif) therefore has an asymmetry in the outgoing signals resulting from the production of PA and PI(4,5)P<sub>2</sub>. The system can take a single input and propagate two distinct signals: PI(4,5)P<sub>2</sub> reports the signal with little modulation; and PA reports only whenever the signal is sufficiently large. This behaviour is illustrated in table 4.4. So downstream processes could rely on either of these signals, with their distinct advantages. In comparison, the symmetric AND-AND and OR-OR motifs effectively only output a single signal via their two products.

From the steady-state solutions of the models, it is apparent that the pseudo-mass-action models are close to an upper limit of complexity, above which much of the mathematical tractability is lost. The addition of Hill functions or conservation of mass give models where the steady-state solutions can be found analytically (computationally), but where these solutions are long and complicated functions of the parameters that are difficult to investigate through algebraic manipulation.

If I must rely on numerical solutions in order to discuss the behaviours of the system, then a major advantage of empirical models, above mechanistic models, has been lost. Also, empirical models are inherently biased towards behaviours that have been previously classified, and so may well be biased towards our (my) expectations. (See the discussion in section 1.3.) So I consider that a better approach for the continuing investigation of the Arf/PLD/PI4P5K motif is the development of mechanistic models from first principles, based on the application of the law of mass-action to well researched biochemical reaction schemes. This approach is what will follow in the subsequent chapters for each of PLD, PI4P5K, and Arf.

In total, I have constructed a series of simple empirical models which have been sufficient to suggest novel hypotheses about the Arf/PLD/PI4P5K motif. Attempts to improve the realism of the initial pseudo-mass-action model caused an increase in complexity, meaning that the resulting models are no longer amenable to mathematical analysis. This led to the decision that a further, more rigorous, investigation into these models was less important than developing more carefully constructed mechanistic models – for which there is the possibility of later returning to similar

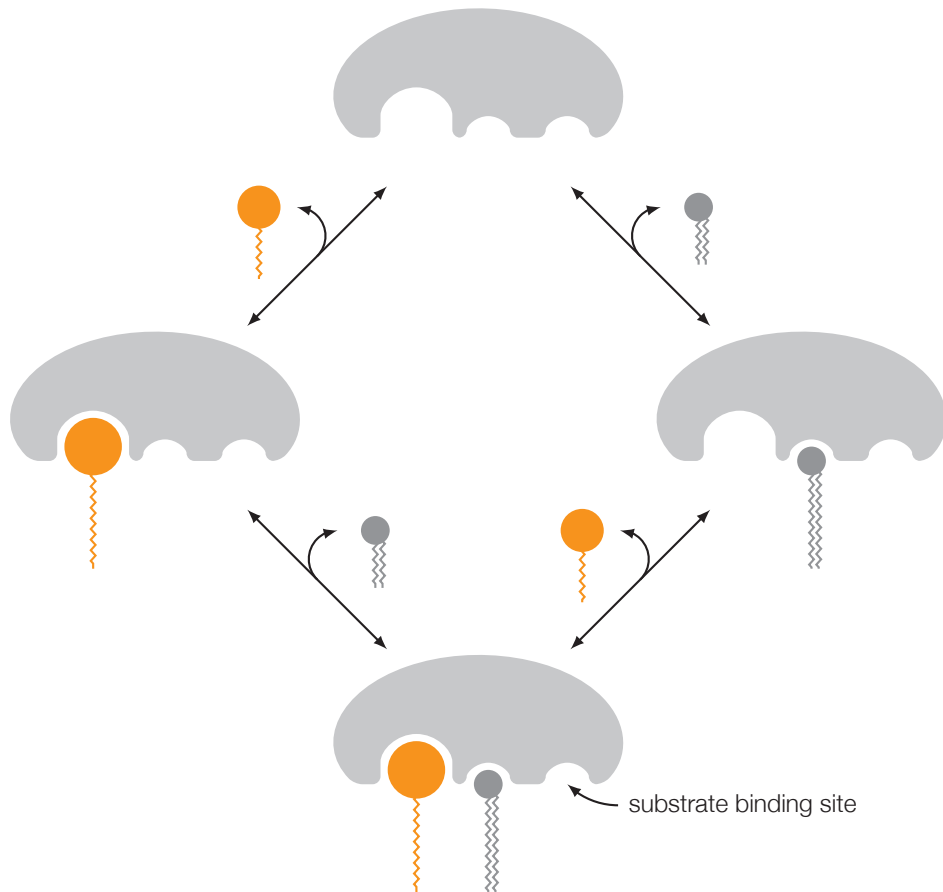
analysis. Additionally, the introduced method by which a system is compared to related but theoretical motifs shows promise as an additional tool for understanding biochemical systems.

## 5 PLD & PI4P5K: mechanisms and models

To construct a well-defined mechanistic model of an enzyme-catalysed reaction, or other biochemical process, it is crucial that the biochemistry that underlies the process is itself well-defined. This does not preclude the development of models based on theoretical mechanisms, so long as it is understood that these models are reliant on a specific set of biochemical assumptions. In fact, for many processes for which only a coarse-grained description is currently available, it will be essential to supplement an ill-defined biochemical process with a series of assumptions and hypotheses. We should also prefer that the majority of assumptions are described as part of the biochemical mechanism, rather than as part of the mathematical derivation. This, perhaps, will let us more easily develop experiments that could validate or invalidate our chosen assumptions and hypotheses.

In this chapter, I intend to construct new, suitably physiologically-realistic models of PLD and PI4P5K. This will require the consideration of the biochemical mechanisms and the specific molecular interactions that occur in the regulatory and catalytic activities of the two enzymes. Wherever possible, these interactions will be supported by primary sources. However, for some (mainly regulatory) interactions, the pre-existing descriptions are insufficient for the immediate construction of mathematical models. In these cases, sensible assumptions will have to be taken, either within the broader context of physical constraints on the regulatory and catalytic interactions of enzymes in general, or using specific alternative arguments relying on information that is present in alternative primary or secondary sources. I will generally consider only processes which can be said to form the main functional processes of the enzymes.

From these—now well-defined—mechanisms I will develop quasi-steady-state mathematical models of both PLD and PI4P5K using the linear framework of Gunawardena (2012), described in section 3.3. These models will later—in chapter 7—be



**Figure 5.1** The di-allosteric mechanism, where a generic 3-site enzyme has two allosteric binding sites for regulatory molecules, in addition to its substrate binding site; and unrestricted order of binding.

brought together to form models of the complete Arf/PLD/PI4P5K motif.

I previously noted, in chapter 4, that PLD and PI4P5K appear to have fundamentally distinct regulatory and catalytic mechanisms. The mechanisms and models that I will develop in this chapter will later allow me to explore these differences further. However, later in this chapter I will find that—under a specific set of assumptions—it is possible that the regulation of both enzymes can be described using the same mechanism. This mechanism arises from an assumption which I will call the *3-site* assumption where the enzyme is assumed to have two independent *allosteric* binding sites – sites distinct from the catalytic site which allow regulation of the enzyme by an effector molecule. I will refer to this as the *di-allosteric* mechanism. I will pre-empt the requirement for the quasi-steady-state solution of this mechanism by deriving it here first. This solution will not consider other regulatory processes such as localisation, substrate binding, nor the catalytic mechanism itself – these will be dealt with independently in complete models of both PLD and PI4P5K.

Additionally, I will also develop a *2-site* model for PI4P5K which does not follow this di-allosteric mechanism, but which relies on an alternative description for the regulation of PI4P5K by PA.

## 5.1 Di-allosteric mechanism: QSS model

Consider an enzyme with two independent, allosteric binding sites for two distinct regulatory molecules – subsequently referred to as the *regulators*. If the order of binding of the regulators is unrestricted, then this mechanism can be described by the diagram in figure 5.1. The quasi-steady-state solutions for the di-allosteric mechanism will be required during the construction of models for both PLD and PI4P5K.

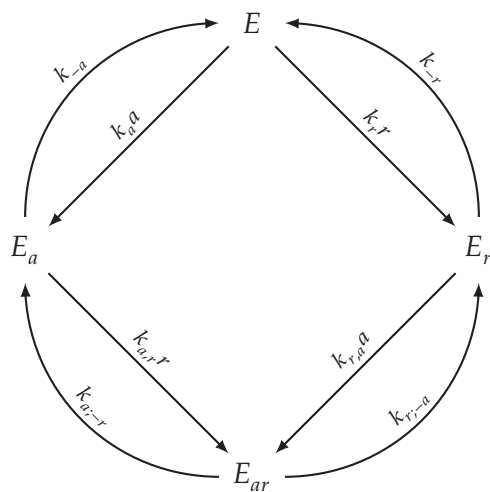
In the following mathematics I will use the shorthand:



( $R$  and  $r$  will correspond to PI(4,5)P<sub>2</sub> in discussion of PLD and PA in discussion of PI4P5K.) I will also use the notation for rates and complexes defined in section 3.2.1.

The di-allosteric mechanism can be described using the reaction scheme:





**Figure 5.2** The graph on the enzyme complexes for the di-allosteric mechanism shown in figure 5.1 – an enzyme  $E$  with two allosteric regulators  $A$  and  $R$ .



Using the law of mass-action, the corresponding system of ODEs is

$$\frac{da}{dt} = - (k_a e + k_{r;a} e_r) a + k_{-a} e_a + k_{r;-a} e_{ar} \quad (5.1)$$

$$\frac{dr}{dt} = - (k_r e + k_{a;r} e_a) r + k_{-r} e_r + k_{a;-r} e_{ar} \quad (5.2)$$

$$\frac{de}{dt} = - (k_a a + k_r r) e + k_{-a} e_a + k_{-r} e_r \quad (5.3)$$

$$\frac{de_a}{dt} = - (k_{-a} + k_{a;r} r) e_a + k_a a e + k_{-r} e_{ar} \quad (5.4)$$

$$\frac{de_r}{dt} = - (k_{r;a} a + k_{-r}) e_r + k_r r e + k_{-a} e_{ar} \quad (5.5)$$

$$\frac{de_{ar}}{dt} = - (k_{r;-a} + k_{a;-r}) e_{ar} + k_{a;r} r e_a + k_{r;a} a e_r. \quad (5.6)$$

To obtain the quasi-steady-state solutions for  $e$ ,  $e_a$ ,  $e_r$ , and  $e_{ar}$  I will use the linear framework of Gunawardena (2012), described in section 3.3. The graph on the enzyme complexes for the di-allosteric mechanism is shown in figure 5.2. The spanning trees of this graph are shown in figure 5.3. These can then be used to derive a basis element

$$\rho = \begin{bmatrix} k_{-a} k_{-r} (k_{r;-a} + k_{a;-r}) + k_{-a} k_{a;-r} k_{r;a} a + k_{-r} k_{r;-a} k_{a;r} r \\ (k_{-r} k_a (k_{r;-a} + k_{a;-r}) + k_{a;-r} k_{r;a} (k_a a + k_r r)) a \\ (k_{-a} k_r (k_{r;-a} + k_{a;-r}) + k_{r;-a} k_{a;r} (k_a a + k_r r)) r \\ (k_{-a} k_r k_{r;a} + k_{-r} k_a k_{a;r} + k_{a;r} k_{r;a} (k_a a + k_r r)) ar \end{bmatrix}$$

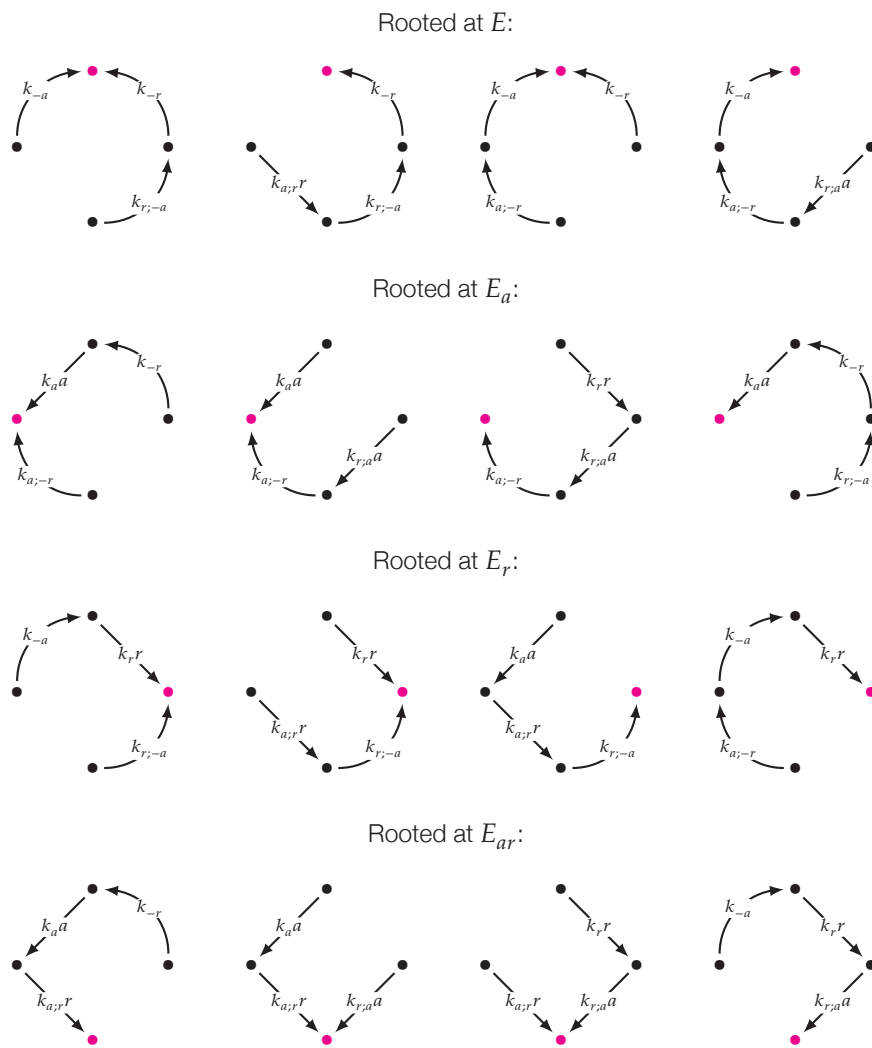
in the order  $(E, E_a, E_r, E_{ar})$ .

It is possible to produce an equivalent, alternative basis element by multiplying or dividing each row by a scalar factor. Dividing  $\rho$  by  $k_a k_r k_{a;r} k_{r;a}$  gives

$$\rho' = \begin{bmatrix} \frac{K_a K_r K_{r;a}}{k_{a;r}} + \frac{K_a K_r K_{a;r}}{k_{r;a}} + \frac{K_a K_{a;r}}{k_r} a + \frac{K_r K_{r;a}}{k_a} r \\ \left( \frac{K_r K_{r;a}}{k_{a;r}} + \frac{k_{r;a}}{K_r K_{a;r}} + \frac{K_{a;r}}{k_r} a + \frac{K_{a;r}}{k_a} r \right) a \\ \left( \frac{K_a K_{r;a}}{k_{a;r}} + \frac{K_a K_{a;r}}{k_{r;a}} + \frac{K_{r;a}}{k_r} a + \frac{K_{r;a}}{k_a} r \right) r \\ \left( \frac{K_r}{k_{r;a}} + \frac{K_a}{k_{a;r}} + \frac{a}{k_r} a + \frac{a}{k_a} r \right) ar \end{bmatrix}$$

where I have introduced equilibrium constants, defined as

$$K_{\sigma;i} = \frac{k_{\sigma;-i}}{k_{\sigma;i}}.$$



**Figure 5.3** The 16 spanning trees of the graph on the enzyme complexes for the di-allosteric mechanism, shown in figure 5.2. The highlighted nodes denote the root of each spanning tree.



To simplify this basis element further, I will assume that the cyclical graph shown in figure 5.2 is in thermodynamic equilibrium and so the *principle of microscopic reversibility* will hold (Colquhoun et al. 2004). This means that the product of the rates clockwise around the cycle will equal the product of the rates anticlockwise around the cycle,

$$k_a a k_{a;r} r k_{r;-a} k_{-r} = k_r r k_{r;a} a k_{a;-r} k_{-a} \implies k_a k_{a;r} k_{r;-a} k_{-r} = k_r k_{r;a} k_{a;-r} k_{-a}.$$

Dividing by the forward rate constants gives

$$\frac{k_{-r} k_{r;-a}}{k_r k_{r;a}} = \frac{k_{-a} k_{a;-r}}{k_a k_{a;r}} \implies K_r K_{r;a} = K_a K_{a;r}.$$

This is equivalent to the statement that the free energy of the complex  $E_{ar}$  must be the same whether  $A$  is added then  $R$ , or  $R$  is added then  $A$ . (This result was used in the derivation of the *general modifier mechanism* by Botts & Morales (1953).)

Using this substitution gives

$$\rho' = \begin{bmatrix} \left( \frac{K_a}{k_{a;r}} + \frac{K_r}{k_{r;a}} + \frac{a}{k_r} a + \frac{a}{k_a} r \right) K_r K_{r;a} \\ \left( \frac{K_a}{k_{a;r}} + \frac{K_r}{k_{r;a}} + \frac{a}{k_r} a + \frac{a}{k_a} r \right) K_{a;r} a \\ \left( \frac{K_a}{k_{a;r}} + \frac{K_r}{k_{r;a}} + \frac{a}{k_r} a + \frac{a}{k_a} r \right) K_{r;a} r \\ \left( \frac{K_a}{k_{a;r}} + \frac{K_r}{k_{r;a}} + \frac{a}{k_r} a + \frac{a}{k_a} r \right) ar \end{bmatrix}.$$

Then, dividing  $\rho'$  by  $\left( \frac{K_a}{k_{a;r}} + \frac{K_r}{k_{r;a}} + \frac{a}{k_r} a + \frac{a}{k_a} r \right)$  gives

$$\rho'' = \begin{bmatrix} K_r K_{r;a} \\ K_{a;r} a \\ K_{r;a} r \\ ar \end{bmatrix}.$$

We know that the steady-state solutions are of the form  $x_i = \frac{\rho_i}{\sum_i \rho_i} \sum_i x_i$ . So from the basis vector  $\rho''$  we can write down the quasi-steady-state concentrations of the intermediate enzymes

$$e = \left( \frac{K_r K_{r;a}}{K_r K_{r;a} + K_{a;r} a + K_{r;a} r + ar} \right) e_0 \quad (5.7)$$

$$e_a = \left( \frac{K_{a;r} a}{K_r K_{r;a} + K_{a;r} a + K_{r;a} r + ar} \right) e_0 \quad (5.8)$$

$$e_r = \left( \frac{K_{r;a}r}{K_r K_{r;a} + K_{a;r}a + K_{r;a}r + ar} \right) e_0 \quad (5.9)$$

$$e_{ar} = \left( \frac{ar}{K_r K_{r;a} + K_{a;r}a + K_{r;a}r + ar} \right) e_0 \quad (5.10)$$

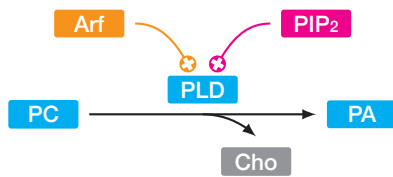
where  $e_0$  is the total concentration of enzyme

$$e_0 = e + e_a + e_r + e_{ar}.$$

I will use the derived values for  $e$ ,  $e_a$ ,  $e_r$ , and  $e_{ar}$  later in this chapter.

## 5.2 PLD: mechanism

Recall that phospholipase D (PLD) is an enzyme that hydrolyses phosphatidylcholine (PC) to produce phosphatidic acid (PA) and choline. Also recall that PLD is regulated by Arf and PI(4,5)P<sub>2</sub>.



To construct a physiologically-realistic model of PLD I require a better understanding of the molecular interactions that determine its regulatory and catalytic activities. I require well-defined (at best based on primary literature; at worst based on assumptions) descriptions for:

- the mechanism of regulation by Arf;
- the mechanism of regulation by PI(4,5)P<sub>2</sub>;
- the localisation of PLD;
- the method of catalytic action;
- and, possible competition for binding-sites (within the considered species).

I will explicitly ignore any effects or interactions that might arise due to other species, complexes, or intracellular processes.

There is limited published literature that discusses the enzyme in sufficient detail to allow the immediate construction of a mechanistic model. So I will have to

supplement those descriptions that do exist with arguments from other, indirect sources.

First—regulation. We know that PLD is activated by Arf only in the presence of PI(4,5)P<sub>2</sub>. I have been unable to find a published description of the mechanism for this regulation. Therefore, in the absence of contradictory (or confirmatory) evidence, I will need to make sensible assumptions in order to progress to a plausible model. I will assume that PLD has three independent, distinct binding sites. One for each of: active Arf; PI(4,5)P<sub>2</sub>; and substrate PC. Thus Arf and PI(4,5)P<sub>2</sub> will behave as allosteric activators. This is illustrated in figure 5.4. There is evidence that there are as many as three distinct binding sites for PI(4,5)P<sub>2</sub> on PLD (Mahankali, Alter, & Gomez-Cambronero 2014) – for simplicity I will assume that (at any one time) only one of these is important for the dynamics of this regulation. (I will discuss the idea of multisite binding mechanisms further in section 8.3.3.) I will assume any other protein or lipid binding sites are inconsequential in the context of the Arf/PLD/PI4P5K motif, and so I can refer to this as the 3-site assumption for PLD.

In text and diagrams, I will illustrate the 3-site assumption using the (compound) symbol:

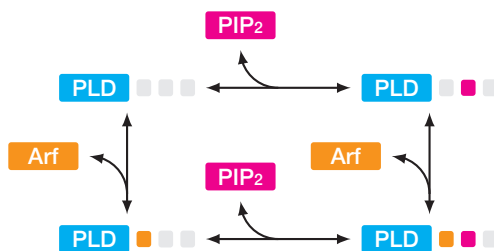


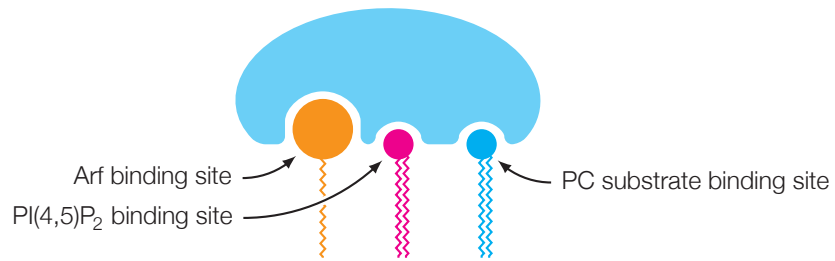
where the grey rectangles denote empty binding sites. Correspondingly, I will illustrate the complete quaternary complex—with substrate and both regulators bound—using the symbol:



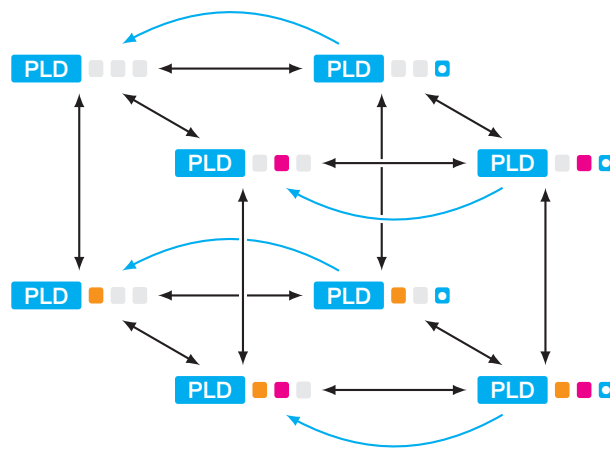
Intermediate complexes will be illustrated equivalently and appropriately.

I will assume that binding of the regulatory molecules, Arf and PI(4,5)P<sub>2</sub>, to PLD will follow the di-allosteric mechanism, shown in figure 5.1. Principally, because this is a generally plausible hypothesis for an otherwise unknown regulatory system containing two distinct regulatory molecules. I also know of no contradictory evidence for this hypothesis. So I will assume a fully reversible scheme with unrestricted order for PLD binding to Arf and PI(4,5)P<sub>2</sub>.





**Figure 5.4** The 3-site assumption for PLD, with a substrate binding pocket and distinct, allosteric binding sites for the two regulators, Arf and PI(4,5)P<sub>2</sub>.



**Figure 5.5** The 3-site mechanism for PLD including regulation by Arf and PI(4,5)P<sub>2</sub>, and catalytic action (blue arrows). For clarity, added species have not been drawn.

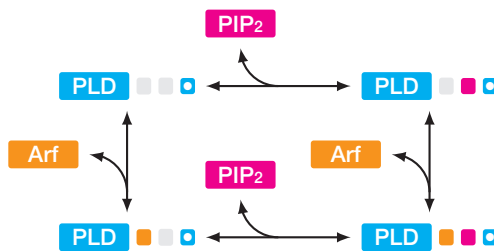
Before I can complete the discussion of the regulatory mechanism I need to consider substrate binding. There are (at least) two plausible mechanisms for the binding of PC to PLD:

1. substrate may bind to any intermediate complex so long as the substrate-binding site is free, **PLD** □ □ □ □, **PLD** ■ □ □ □, **PLD** □ ■ □ □, or **PLD** ■ ■ □ □;
2. or, substrate may only bind to the complex with both Arf and PI(4,5)P<sub>2</sub> already bound, **PLD** ■ ■ ■ □.

The second hypothesis immediately suggests a plausible mechanism of activation – a restriction on substrate binding, explicitly requiring the two regulators to be present. This would be true if the regulators operate to modulate the structure of the enzyme so that the substrate binding site becomes exposed.

However, by restricting substrate binding we also restrict catalysis, and so the second hypothesis disallows a basal rate of catalysis in the absence of the regulators. This appears to be false in at least one set of experimental results – see for instance figure 11 in Perez-Mansilla et al. (2006). This shows a small—but non-zero—basal rate of PLD activity in the absence of Arf. Therefore I will assume the first hypothesis to be true.

With PC able to bind to any intermediate complex, I will also assume that Arf and PI(4,5)P<sub>2</sub> themselves can also bind to any intermediate complex (so long as the correct binding site is free). This implies an equivalent application of the di-allosteric mechanism, now in respects to the enzyme-substrate complex.



Within the framework of this regulatory scheme, it remains to state exactly how Arf and PI(4,5)P<sub>2</sub> regulate PLD. Activation could be due to either:

- an increase in the rate of substrate binding;
- a decrease in the rate of substrate unbinding;
- or, an increase in the total catalytic rate of the quaternary complex.

In the full model that will first be derived, any of these hypotheses could be investigated. However, for simplicity I will assume that the last of these is true, and that substrate binding/unbinding is completely independent of the bound regulators.

Next I will investigate the catalytic mechanism of PLD – for which a proposed mechanism is available (Mahankali, Alter, & Gomez-Cambronero 2014). This mechanism is described as irreversible and is characterised by the ordered release of the products of the reaction – first choline, then PA.



As we are only interested in the production of PA—and not choline—the final model would be simplified if it is sensible to assume that the products are released in a single catalytic step. Effectively, we need to assume that the **PLD** **PA** complex is only very short-lived. Additionally, this mechanism can be modelled as irreversible only in the absence of product inhibition by either PA or choline.

Inhibition by choline would likely be the less impactful inhibition as soluble and so that produced by PLD will be free to diffuse away from the membrane. So the surface concentration of the (produced) choline will be lower than the surface concentration of the (produced) PA. In fact, there is no evidence for inhibition of PLD by choline listed in the BRENDA database (Schomburg et al. 2013). Nor is there any evidence in the BRENDA database for inhibition by other lipids which contain choline-headgroups such as sphingomyelin.

From a review of the existing literature and a search on the BRENDA database (Schomburg et al. 2013) I have found only two references to the existence or non-existence of inhibition of PLD by PA.

- Okamura & Yamashita (1994) states:

... lyso-PC, PE, PI, PA, and DAG were inhibitory to the enzyme ...

- Hirano et al. (2012) states:

... typical bacterial PLD from *Streptomyces* sp. was found to hydrolyze all the PC molecules at the outer surface of LUVs [large unilamellar vesicles] suggesting that this enzyme is free from product inhibition.

Also:

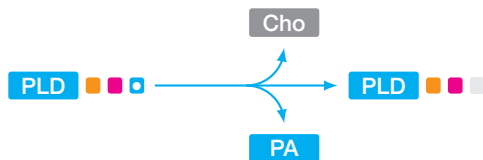
Compared with PLD from *S. chromofuscus*, PLD from *Streptomyces* sp. showed no product inhibition.

Note that the PLD from *Streptomyces* sp. catalyses the reaction via a HKD domain, as per mammalian PLDs, while the PLD from *Streptomyces chromofuscus* lacks a HKD group and so catalyses the reaction using an alternative mechanism. (It is perhaps interesting to note that the PLD from *S. chromofuscus* does hydrolyse sphingomyelin (Imamura & Horiuti 1979).)

Neither description is particularly satisfying, and neither article includes any further evidence supporting these statements.

I have chosen to contradict the first, older report, and so will assume that PLD is not inhibited by PA (specifically mammalian isoforms that operate through HKD domains). Primarily, I make this choice because it will result in a simpler (biochemical) model of the regulation of the enzyme. But also, the lack of further reporting of inhibition of PLD by PA suggests that it is not a significant modifier to the activity of the enzyme in most systems.

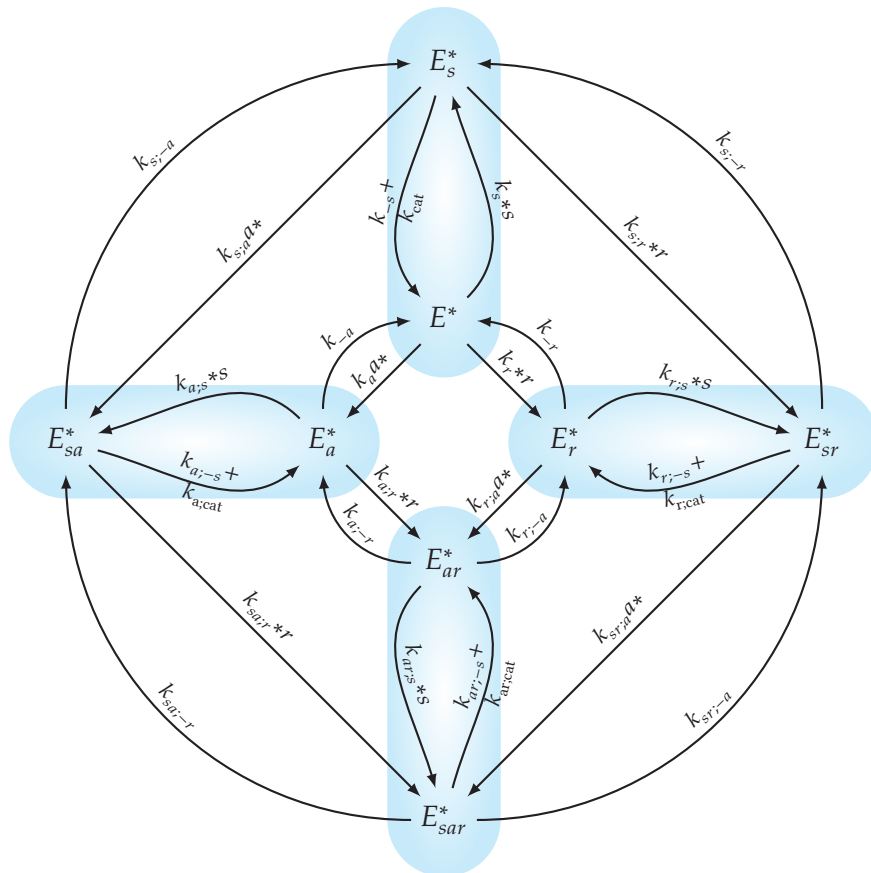
In total, I will assume that PLD does not display product inhibition by either PA or choline. Additionally, I will assume that there is no long-lived PLD-PA complex and so hydrolysis of PC and release of PA/choline will occur in a single irreversible step.



As I have mentioned above, I would like to allow for the possibility of a basal rate of catalysis in the absence of the regulators. So I will (initially) assume that this catalytic reaction can occur for any PLD complex so long as substrate is bound.

Together, the mechanisms of regulation, substrate binding/unbinding, and substrate catalysis are illustrated in figure 5.5.

Finally, I will assume that all PLD is constitutively membrane-associated, and that the membrane-association is independent of the concentration of Arf and PI(4,5)P<sub>2</sub> (otherwise, I will leave this particular mechanism unspecified). This is possibly an over-simplification – as stated in section 2.4 the phospholipid-binding domains of PLD have been characterised as having a role in membrane-recruitment. However, numerous reports do characterise PLD as having predominantly membrane-localised



**Figure 5.6** Graph on the enzyme complexes for the 3-site model for PLD. The highlighted regions correspond to pools of complexes that are connected by substrate binding, unbinding, and catalysis.



distributions (PLD1 to the Golgi and PLD2 to the plasma membrane) – see the introduction and results of Hiroyama & Exton (2005).

### 5.3 PLD: QSS model

Now that I have described a plausible mechanism for PLD I am able to construct a suitable quasi-steady-state model based on this mechanism.

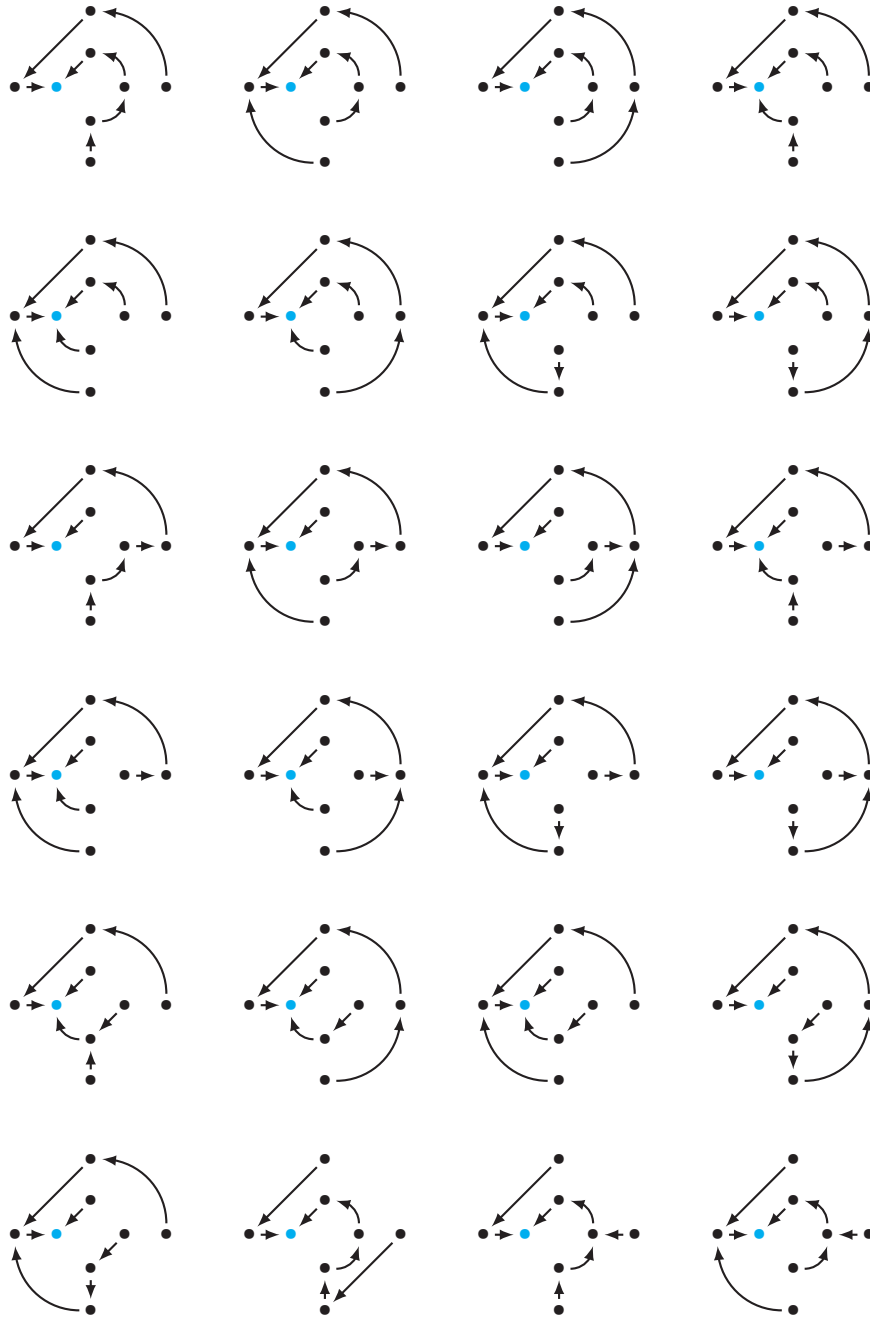
In this section I will use the shorthand:

$$\begin{array}{ll}
 \text{PLD} \rightarrow E^*, & [\text{PLD}] \rightarrow e^*; \\
 \text{PC} \rightarrow *S, & [\text{PC}] \rightarrow *s; \\
 \text{PA} \rightarrow *P, & [\text{PA}] \rightarrow *p; \\
 \text{Arf} \rightarrow A^*, & [\text{Arf}] \rightarrow a^*; \\
 \text{PI(4,5)P}_2 \rightarrow *R, & [\text{PI(4,5)P}_2] \rightarrow *r.
 \end{array}$$

I will also use the notation for rates and complexes defined in section 3.2.1. As PLD operates on the membrane, and is itself membrane-associated, all species/complexes have been pre/appended by an asterisk (\*) according to the rules in section 3.5.1.

The 3-site mechanism for PLD was illustrated in figure 5.5. This can be drawn as the graph on the enzyme complexes shown in figure 5.6. The corresponding system of ODEs is

$$\begin{aligned}
 \frac{de^*}{dt} &= - (k_s *s + k_a a^* + k_r *r) e^* + (k_{-s} + k_{\text{cat}}) e_s^* + k_{-a} e_a^* + c_{-r} e_r^* \\
 \frac{de_s^*}{dt} &= - (k_{-s} + k_{\text{cat}} + k_{a;s} a + k_{r;s} r) e_s^* + k_s *s e^* + k_{s;-a} e_{sa}^* + k_{s;-r} e_{sr}^* \\
 \frac{de_a^*}{dt} &= - (k_{a;s} *s + k_{-a} + k_{a;r} *r) e_a^* + (k_{a;-s} + k_{a;\text{cat}}) e_{sa}^* + k_a a^* e^* + k_{-r} e_r^* \\
 \frac{de_{sa}^*}{dt} &= - (k_{a;-s} + k_{a;\text{cat}} + k_{s;-a} + k_{sa;r} *r) e_{sa}^* + k_{a;s} *s e_a^* + k_{s;a} a^* e_s^* + k_{sa;-r} e_{sar}^* \\
 \frac{de_r^*}{dt} &= - (k_{r;s} *s + k_{r;a} a^* + k_{-r}) e_r^* + (k_{r;-s} e_{sr}^* + k_{r;\text{cat}}) + k_{-a} e_a^* + k_r *r e^* \\
 \frac{de_{sr}^*}{dt} &= - (k_{r;-s} + k_{r;\text{cat}} + k_{sr;a} a^* + k_{s;-r}) e_{sr}^* + k_{r;s} *s e_r^* + k_{sr;-a} e_{sar}^* + k_{s;r} *r e_s^* \\
 \frac{de_{ar}^*}{dt} &= - (k_{a;r} *s + k_{r;-a} + k_{a;-r}) e_{ar}^* + (k_{a;-s} + k_{a;\text{cat}}) e_{sar}^* + k_{a;r} *r e_a^* + k_{r;a} a^* e_r^* \\
 \frac{de_{sar}^*}{dt} &= - (k_{a;-s} + k_{a;\text{cat}} + k_{sr;-a} + k_{sa;-r}) e_{sar}^* + k_{a;r} *s e_{ar}^* + k_{sa;r} *r e_{sa}^* + k_{sr;a} a^* e_{sr}^* \\
 \frac{da^*}{dt} &= - (k_a e^* + k_{s;a} e_s^* + k_{r;a} e_r^* + k_{sr;a} e_{sr}^*) a^* + k_{-a} e_a^* + k_{s;-a} e_{sa}^* + k_{r;-a} e_{ar}^* + k_{sr;-a} e_{sar}^*
 \end{aligned}$$



**Figure 5.7** A small proportion of the 3072 spanning trees of the 3 site PLD graph shown in figure 5.6.

$$\begin{aligned}\frac{d*r}{dt} &= - (k_r e^* + k_{s;r} e_s^* + k_{a;r} e_a^* + k_{sa;r} e_{sa}^*) * r + k_{-r} e_r^* + k_{s;-r} e_{sr}^* + k_{a;-r} e_{ar}^* + k_{sa;-r} e_{sar}^* \\ \frac{d*s}{dt} &= - (k_s e^* + k_{a;s} e_a^* + k_{r;s} e_r^* + k_{ar;s} e_{ar}^*) * s + k_{-s} e_s^* + k_{a;-s} e_{sa}^* + k_{r;-s} e_{sr}^* + k_{ar;-s} e_{sar}^* \\ \frac{d*p}{dt} &= k_{cat} e_s^* + k_{a;cat} e_{sa}^* + k_{r;cat} e_{sr}^* + k_{ar;cat} e_{sar}^*.\end{aligned}$$

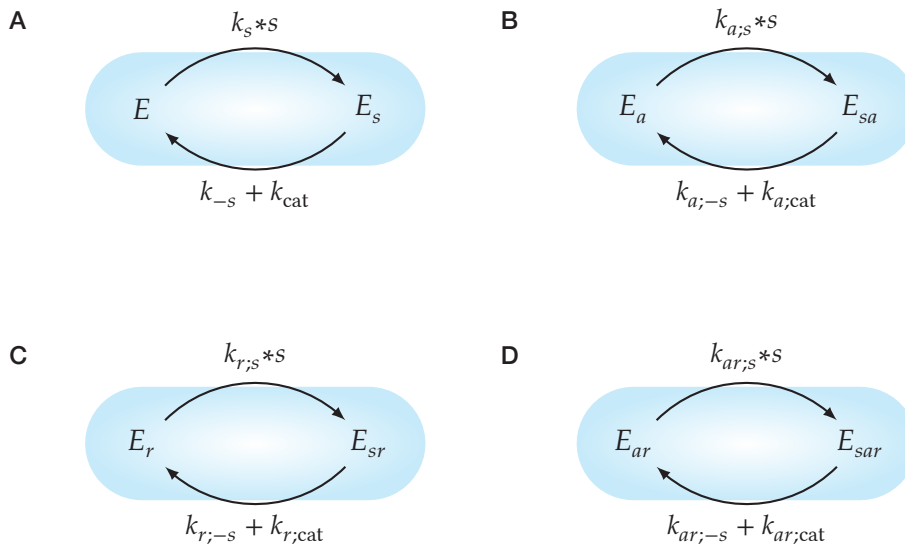
This is a system of 12 variables and 28 parameters.

I would like to use the linear framework of Gunawardena (2012), described in section 3.3, to construct a reduced model from this system. The number of spanning trees was found—using the algorithm in section 3.4.1—with the following Python code:

```

1 from sympy import *
2 from enzymegraph import *
3
4 # setup symbols
5 s, p, a, r = symbols('s p a r', positive=True)
6 e, ea, er, ear = symbols('E E_a E_r E_ar', positive=True)
7 es, esa, esr, esar = symbols('E_s E_sa E_sr E_sar', positive=True)
8
9 # list of edges, with arbitrary parameter labels
10 k = parameter_generator()
11 edges = {
12     # no substrate bound
13     (e, ea) : next(k), (ea, e) : next(k),
14     (e, er) : next(k), (er, e) : next(k),
15     (er, ear) : next(k), (ear, er) : next(k),
16     (ea, ear) : next(k), (ear, ea) : next(k),
17     # substrate bound
18     (es, esa) : next(k), (esa, es) : next(k),
19     (es, esr) : next(k), (esr, es) : next(k),
20     (esr, esar) : next(k), (esar, esr) : next(k),
21     (esa, esar) : next(k), (esar, esa) : next(k),
22     # add/remove substrate
23     (e, es) : next(k), (es, e) : next(k),
24     (ea, esa) : next(k), (esa, ea) : next(k),
25     (er, esr) : next(k), (esr, er) : next(k),
26     (ear, esar) : next(k), (esar, ear) : next(k),
27 }
28
29 # generate graph
30 graph = enzymegraph(edges)
31
32 # count spanning trees
33 print(len([1 for _ in graph.spanning_trees()])))

```



**Figure 5.8** Subgraphs of the 3-site PLD graph, the highlighted regions in figure 5.6 corresponding to pools of complexes that have interconversions comprising of only substrate binding, unbinding, and catalysis.

Thus the graph on the enzyme complexes can be shown to have 3072 distinct spanning trees. This is clearly too many to sensibly enumerate or manipulate. A small proportion of these have been drawn in figure 5.7.

I need to amend my approach so as to introduce sufficient simplification, so that I can sensibly develop a reduced model. For this reason I will assume that dynamics related to interactions with the substrate/product can—in some sense—be separated from dynamics related to interactions with the regulators. I will assume that the time scales of these two sets of interactions are such that they can be uncoupled. Specifically, I will assume that substrate binding, unbinding, and catalysis follow the rapid equilibrium assumption, as specified in section 3.2.4. This is supported by the following quote from Hammes (1978):

*The rates of regulatory conformational changes are generally somewhat slower than those associated with catalysis and in some cases are extremely slow (min and hr).*

These interactions have been highlighted on the graph in figure 5.6, defining *pools* of related complexes.

I intend to introduce new variables which correspond to total concentrations of these pools of complexes. By making additional assumptions on the rate constants of the system, I will be able to generate a new system. This system can be described by an associated graph where the vertices corresponding to the pools. In a (very loose) sense I will be *collapsing* the graph in figure 5.6 along the highlighted, substrate-interaction, dimension.

The four independent (sub-)graphs shown in figure 5.8 correspond to the highlighted regions in the graph in figure 5.6. To each of these I can apply the rapid equilibrium assumption, or equivalently the linear framework of Gunawardena (2012) – in which case the general basis element is

$$\rho = \begin{bmatrix} k_{\sigma;-s} + k_{\sigma;\text{cat}} \\ k_{\sigma;s} * s \end{bmatrix}$$

in the order  $(e_{\sigma}, e_{\sigma \cup \{s\}}^*)$ , where  $\sigma \subset \{a, r\}$  (including the empty set).

Therefore, the general forms of the rapid equilibrium solutions for each complex are

$$e_{\sigma}^* = \frac{K_{\sigma;s}}{K_{\sigma;s} + *s} \zeta_{\sigma} \quad (5.11)$$

$$e_{\sigma \cup \{s\}}^* = \frac{*s}{K_{\sigma;s} + *s} \zeta_{\sigma} \quad (5.12)$$

where

$$K_{\sigma;s} = \frac{k_{\sigma;-s} + k_{\sigma;\text{cat}}}{k_{\sigma;s}}$$

and

$$\zeta_{\sigma} = e_{\sigma} + e_{\sigma \cup \{s\}}^*$$

The  $\zeta$  are the new variables mentioned previously, which correspond to the total concentration of the pools of complexes (which are partial total concentrations of the whole system).

We can now also write down the rates of change of each  $\zeta_{\sigma}$ . For example

$$\begin{aligned} \frac{d\zeta}{dt} &= \frac{de}{dt} + \frac{de_s}{dt} \\ &= - (k_a a^* + k_r^* r) e^* - (k_{s;a} + k_{s;r}) e_s^* + k_{-a} e_a^* + k_{s;-a} e_{sa}^* + k_{-r} e_r^* + k_{s;-r} e_{sr}^* \\ &= - \frac{(k_a a^* + k_r^* r) K_s + (k_{s;a} + k_{s;r}) *s}{K_s + *s} \zeta + \frac{k_{-a} K_{a;s} + k_{s;-a} *s}{K_{a;s} + s} \zeta_a + \\ &\quad \frac{k_{-r} K_{r;s} + k_{s;-r} *s}{K_{r;s} + s} \zeta_r. \end{aligned}$$

In order to simplify this further I need to assume that the rates of binding and unbinding of Arf and PI(4,5)P<sub>2</sub> are independent of whether substrate is bound. That is,

$$\begin{array}{cccc} k_{s;a} = k_a & k_{s;-a} = k_{-a} & k_{s;r} = k_r & k_{s;-r} = k_{-r} \\ k_{sr;a} = k_{r;a} & k_{sr;-a} = k_{r;-a} & k_{sa;r} = k_{a;r} & k_{sa;-r} = k_{a;-r} \end{array}$$

Then the rate of change of  $\zeta$  can be written as

$$\frac{d\zeta}{dt} = - (k_a a^* + k_r^* r) \zeta + k_{-a} \zeta_a + k_{-r} \zeta_r.$$

Similarly, for the other variables,

$$\begin{aligned} \frac{d\zeta_a}{dt} &= - (k_{-a} + k_{a;r}^* r) \zeta_a + k_a a^* \zeta + k_{-r} \zeta_{ar} \\ \frac{d\zeta_r}{dt} &= - (k_{r;a} a^* + k_{-r}) \zeta_r + k_{-a} \zeta_{ar} + k_r^* r \zeta \\ \frac{d\zeta_{ar}}{dt} &= - (k_{r;-a} + k_{a;r}^* r) \zeta_a + k_{r;a} a^* \zeta_r + k_{a;-r} \zeta_{ar}. \end{aligned}$$

Excluding the variable names, the equations for the rate of change between the pools of complexes exactly match those that describe the di-allosteric mechanism,

equations 5.1 to 5.6. So this system has been reduced to the di-allosteric mechanism, and can also be described using the graph in figure 5.2 by appropriate relabelling of the vertices. From the original derivation I can immediately write down the quasi-steady-state values of the  $\zeta$ . These were originally given as equations 5.7 to 5.10, and so

$$\zeta = \left( \frac{K_r K_{r;a}}{K_r K_{r;a} + K_{a;r} a^* + K_{r;a}^* r + a^* r} \right) e_0 \quad (5.13)$$

$$\zeta_a = \left( \frac{K_{a;r} a^*}{K_r K_{r;a} + K_{a;r} a^* + K_{r;a}^* r + a^* r} \right) e_0 \quad (5.14)$$

$$\zeta_r = \left( \frac{K_{r;a}^* r}{K_r K_{r;a} + K_{a;r} a^* + K_{r;a}^* r + a^* r} \right) e_0 \quad (5.15)$$

$$\zeta_{ar} = \left( \frac{a^* r}{K_r K_{r;a} + K_{a;r} a^* + K_{r;a}^* r + a^* r} \right) e_0 \quad (5.16)$$

where

$$e_0 = \zeta + \zeta_a + \zeta_r + \zeta_{ar} = e^* + e_s^* + e_a^* + e_{sa}^* + e_r^* + e_{sr}^* + e_{ar}^* + e_{sar}^*.$$

I now have enough information to construct a model for the 3-site mechanism for PLD. From the original system of ODEs, we know that

$$\begin{aligned} \frac{d^*p}{dt} &= k_{\text{cat}} e_s^* + k_{a;\text{cat}} e_{sa}^* + k_{r;\text{cat}} e_{sr}^* + k_{ar;\text{cat}} e_{sar}^* \\ &= \sum_{\sigma \subset \{a,r\}} k_{\sigma;\text{cat}} e_{\sigma \cup \{s\}}^* \end{aligned}$$

Substituting in the value of  $e_{\sigma \cup \{s\}}^*$  from equation 5.12 gives

$$\frac{d^*p}{dt} = \sum_{\sigma \subset \{a,r\}} k_{\sigma;\text{cat}} \left( \frac{*s}{K_{\sigma;s} + *s} \right) \zeta_{\sigma}.$$

Substituting in the values of  $\zeta_{\sigma}$  from equations 5.13 to 5.16, and rearranging, gives

$$\begin{aligned} \frac{d^*p}{dt} &= \left( \frac{k_{\text{cat}} K_r K_{r;a}}{K_s + *s} + \frac{k_{a;\text{cat}} K_{a;r} a^*}{K_{a;s} + *s} + \frac{k_{r;\text{cat}} K_{r;a}^* r}{K_{r;s} + *s} + \frac{k_{ar;\text{cat}} a^* r}{K_{ar;s} + *s} \right) \\ &\quad \times \left( \frac{*s e_0}{K_r K_{r;a} + K_{a;r} a^* + K_{r;a}^* r + a^* r} \right). \end{aligned} \quad (5.17)$$

Equation 5.17 describes the most general version of the model, which has 3 variables and 12 parameters. This equation is useful if we wish to investigate the effect on substrate binding/unbinding rates of different combinations of bound regulators. This full form of the model is very verbose, and—given that the PLD model will comprise part of a model of the Arf/PLD/PI4P5K motif, which itself is intended to be included in still larger models—a simplified form is desirable. So I will make the following additional assumptions:

1. That the catalytic rates for the three non-quaternary complexes are equal, there is no partial activation by either regulator,

$$k_{\text{cat}} = k_{a;\text{cat}} = k_{r;\text{cat}}.$$

2. That the equilibrium constants for the three non-quaternary complexes are equal,

$$K_s = K_{a;s} = K_{r;s}.$$

3. That the equilibrium constant for PI(4,5)P<sub>2</sub> binding/unbinding is independent of whether Arf is bound,

$$K_{a;r} = K_r.$$

With these assumptions, the model reduces to

$$\frac{d^*p}{dt} = \left( \frac{K_r K_{r;a} + K_r a^* + K_{r;a}^* r}{K_s + *s} + \frac{k_{ar;\text{act}} a^* *r}{K_{ar;s} + *s} \right) \left( \frac{k_{\text{cat}} *s e_0}{(K_{r;a} + a^*)(K_r + *r)} \right) \quad (5.18)$$

where

$$k_{ar;\text{act}} = \frac{k_{ar;\text{cat}}}{k_{\text{cat}}}$$

is the fold-change in the catalytic rate due to the presence of both Arf and PI(4,5)P<sub>2</sub>.

The model can be simplified further by assuming that there is no basal rate of catalysis (in the absence of either regulator). Given the previous evidence this is probably not true, but depending on the relative rates of activated and unactivated enzymes it may be an appropriate approximation. That is  $k_{\text{cat}} = 0$  ( $k_{ar;\text{cat}} \neq 0$ ), in which case

$$\frac{d^*p}{dt} = \frac{k_{ar;\text{cat}} *s a^* *r e_0}{(K_{ar;s} + *s)(K_{r;a} + a^*)(K_r + *r)}. \quad (5.19)$$

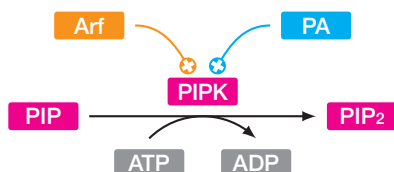
I will analyse the behaviour of the model given by equation 5.19 in chapter 7.

## 5.4 PI4P5K: mechanism

Recall that phosphatidylinositol 4-phosphate 5-kinase (PI4P5K) is an enzyme that phosphorylates PI4P to produce PI(4,5)P<sub>2</sub>. The additional phosphate group is obtained



from ATP. Also recall that PI4P5K is activated by Arf and PA.

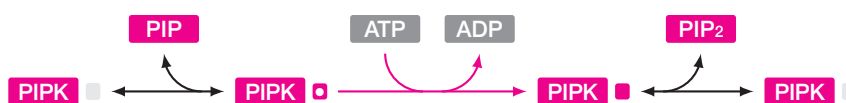


As for PLD, to construct a physiologically-realistic model of PI4P5K I require a better understanding of the molecular interactions that determine its catalytic action and mechanism of activation. Based on arguments from primary and secondary sources and additional sensible assumptions I require well-defined descriptions for:

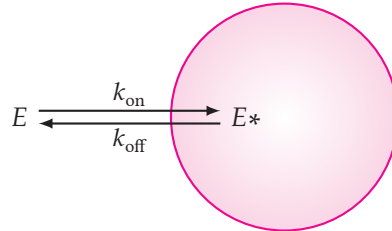
- the mechanism of regulation by Arf;
- the mechanism of regulation by PA;
- the localisation of PI4P5K;
- the method of catalytic action;
- and, possible competition for binding-sites (within the considered species).

I will explicitly ignore any interactions with species outside the limits of the model I intend to construct. In particular, I will ignore any dynamics associated with the binding and unbinding of ATP and ADP – I assume that the interaction with ATP/ADP is inconsequential and the requirement for ATP is never rate limiting. (I will discuss the idea of modelling interactions with ATP/ADP in section 8.3.2.)

First—for the catalytic action of PI4P5K—I will assume that catalysis of PIP to  $\text{PI}(4,5)\text{P}_2$  proceeds as follows:

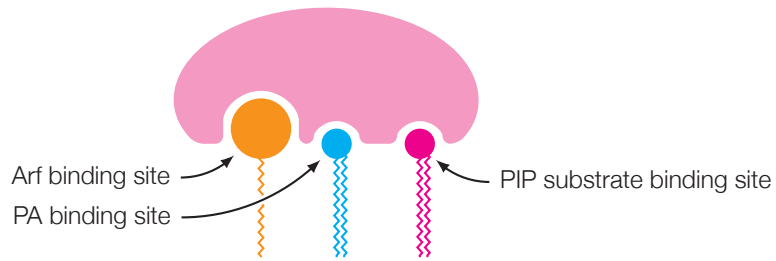


(Only the substrate binding site has been considered in this diagram.) This reaction scheme specifies a reversible binding of substrate; followed by irreversible catalysis; followed by a reversible unbinding of product. The reaction scheme includes product inhibition as this has been experimentally observed (Ling, Schulz, & Cantley 1989; Moritz et al. 1992). Product inhibition may result from the structural similarity between the substrate and product, and the fact that the substrate plus the additional phosphate group must fit inside the catalytic site.

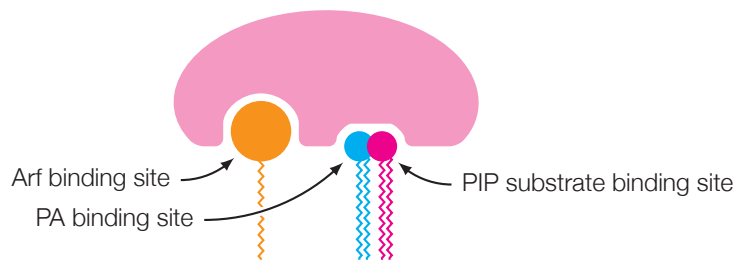


**Figure 5.9** Simple reversible surface adsorption of an enzyme binding to a membrane.

**A** 3-site



**B** 2-site



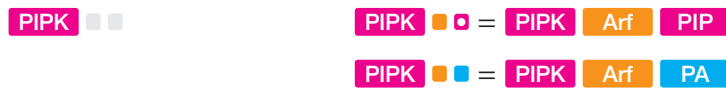
**Figure 5.10** The 3-site and 2-site hypotheses for PI4P5K. **A** a substrate binding pocket and distinct, allosteric binding sites for the two regulators, Arf and PI(4,5)P<sub>2</sub>; **B** a substrate binding pocket which overlaps a binding site for PI(4,5)P<sub>2</sub> and a distinct, allosteric binding site for Arf.

Not all PI4P5K is located near to its substrate – it must first localise to the lipid membrane by diffusion or some other process. Once membrane-localised it is plausible that small inter-molecular forces will keep an individual molecule of PI4P5K near the membrane for short periods of time. Mathematically, I will consider PI4P5K to act as a surface-active enzyme, illustrated in figure 5.9, as was described in section 3.5.

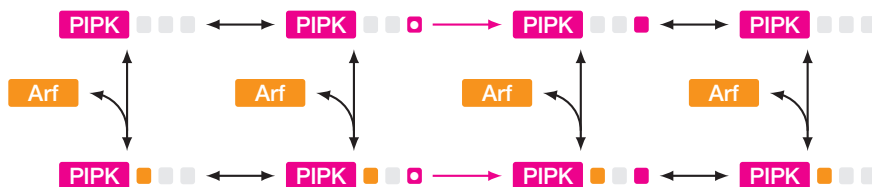
Regulation (activation) by Arf and PA has been reported to be at least additive, and perhaps synergistic (Honda et al. 1999; Cockcroft 2009). I will consider regulation by each to be independent processes. Two different hypotheses for the mechanism of regulation by PA will lead to two models: a *3 site* hypothesis, equivalent to the derived model for PLD; and a *2 site* hypothesis. I will illustrate the 3-site hypothesis using the (compound) symbols:



I will illustrate the 2-site hypothesis using the symbols:



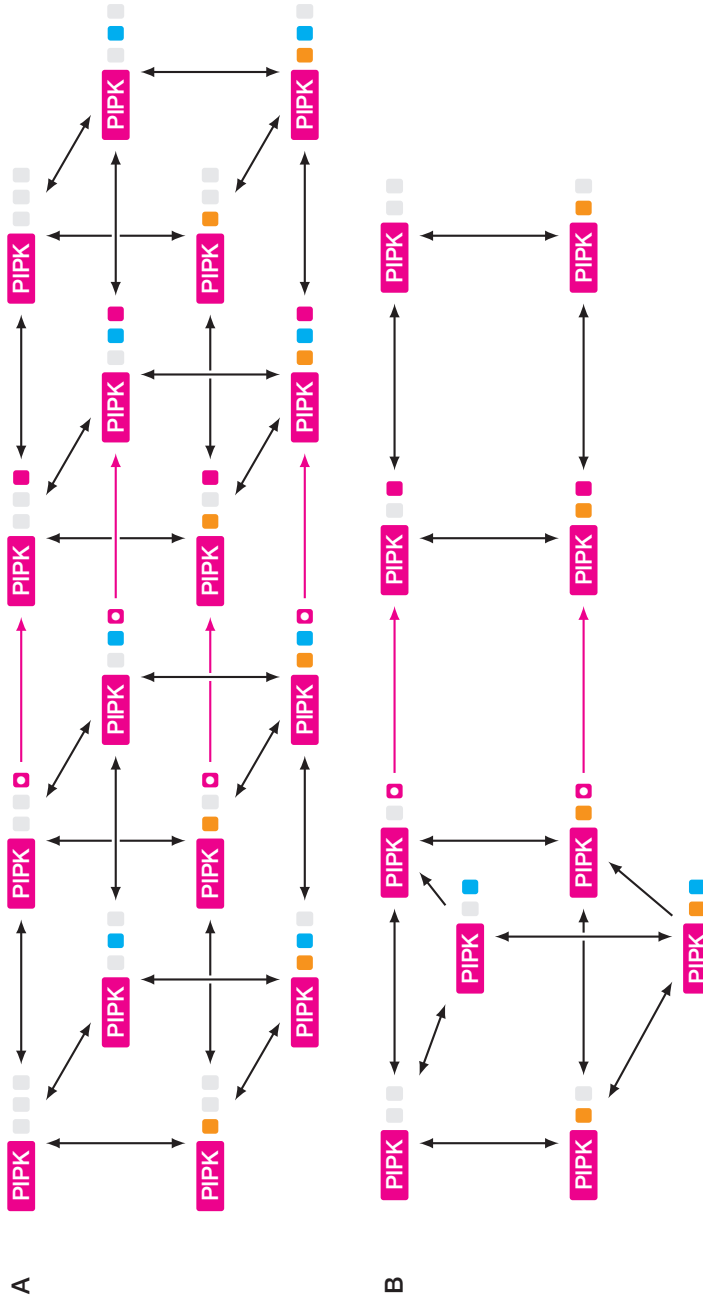
For both the 3-site and 2-site hypotheses, I will assume the same mechanism of regulation by Arf – via an allosteric binding site. If binding and unbinding can occur at any stage of the catalytic process, this can be described using the reaction scheme:



(Illustrated for the 3-site hypothesis.)

Within this framework, there are two plausible mechanisms of activation:

1. PI4P5K has a conformational change when bound to Arf that in some way improves its total catalytic activity, either by:
  - an increase in the rate of substrate binding;
  - a decrease in the rate of substrate unbinding;
  - or, an increase in the total catalytic rate of the complete complex.



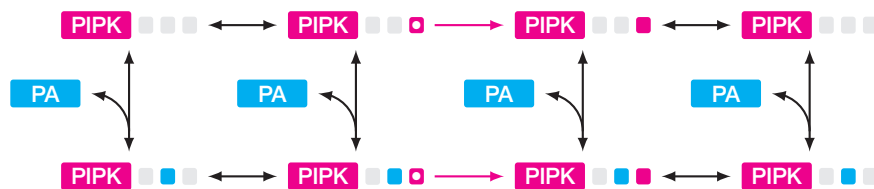
**Figure 5.11** The 3-site and 2-site mechanisms for PI4P5K including regulation by Arf and PI(4,5)P<sub>2</sub>, and catalytic action (magenta arrows). For clarity, added species have not been drawn. Also note the duplication of the left-most and right-most complexes.

This can be described as activation via improving the efficacy of an individual molecule of PI4P5K. These mechanisms correspond to those regulatory processes considered for PLD.

2. When bound to active Arf (itself membrane bound) PI4P5K has a stronger membrane association – it must first unbind Arf before it can leave the membrane. This is activation via increasing the effective surface concentration of PI4P5K, or activation by membrane-tethering/-recruitment.

These two mechanisms are not contradictory – it is entirely possible that action of PI4P5K by Arf is by a combination of both. However, the second is arguably more compelling and raises an additional question as to whether this membrane-tethering is sufficient to explain observed levels of activation. Therefore, I choose to primarily investigate this hypothesis. However, a complete non-simplified model, that will be an intermediate step in the model derivation, could be used to investigate any of these hypotheses.

The 3-site hypothesis for PI4P5K is then generated by the assumption that there is a second, independent, allosteric binding site in which PA operates in a manner equivalently to Arf, as illustrated in figure 5.10A. I will assume that PA can bind to any intermediate complex.



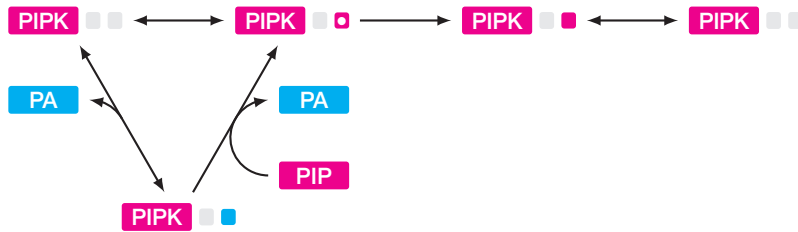
(Equivalently for the PI4P5K-Arf complex.) In this case, either or both of two mechanisms of activation described for Arf could also hold for PA. This does not necessarily have to be the same mechanism, but here I will again assume the second hypothesis. In total, the reaction scheme for the 3-site mechanism is illustrated in figure 5.11A.

The 2-site hypothesis for PI4P5K derives from an alternative mechanism proposed by Stace et al. (2008):

*We hypothesize a mechanism whereby binding to PA may allow the substrate binding site to exist in optimal conformation, ready for substrate binding. Once substrate is available, it could replace PA—due to higher affinity—and be phosphorylated to  $PI(4,5)P_2$ . This model is consistent with*

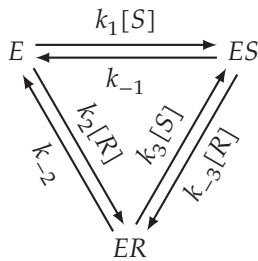
very recent work from Jarquin-Pardo et al. (2007) who suggested that PA stimulates PIP5K by increasing its affinity for the substrate.

This is illustrated in figure 5.10B. The description of this mechanism specifically argues against the formation of a quaternary complex. It can be represented alternatively using the reaction scheme:



(Equivalently for the PI4P5K-Arf complex.) This mechanism is supported by results which state that in the presence of strong product inhibition, activation by PA is completely suppressed (Moritz et al. 1992) – suggesting that PI(4,5)P<sub>2</sub> and PA are in some way competitive.

Regarding the drawn irreversibility of the exchange reaction – briefly consider the thermodynamically-complete system:



Here  $E$  is enzyme (PI4P5K),  $S$  is substrate (PIP), and  $R$  is regulator (PA). This system is equivalent to the sub-system formed by the substrate-binding and PA-regulatory dynamics of the 2-site PI4P5K mechanism, with the addition of a reverse reaction opposite to the exchange mechanism ( $ES \rightarrow ER$ ).

At thermodynamic equilibrium, the product of the clockwise rates around this cyclical graph will equal the product of the anticlockwise rates, so

$$k_1[S] k_{-3}[R] k_{-2} = k_2[R] k_3[S] k_{-1}$$

$$\frac{k_1}{k_{-1}} = \frac{k_2 k_3}{k_{-2} k_{-3}}$$

$$K_1 = K_2 K_3$$

where  $K_i = \frac{k_i}{k_{-i}}$ . This means that the change in total free energy (for the creation of the enzyme-substrate complex) must be the same via the direct route as via the enzyme-regulator complex.

This also implies a constraint on the parameters of the system. For example, to approximate the 2-site mechanism for PI4P5K we would like  $k_{-3}$  to be (very) small. But then the following combination of parameters must be equally small:

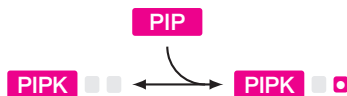
$$k_{-3} = \frac{k_{-1} k_2 k_3}{k_1 k_{-2}}.$$

Hence this completely reversible system, as stated, is incompatible with (steady-state) activation. Although this system could instead mediate transient activation (plausibly by the initial absence of substrate and a build up of the enzyme-regulator complex) this is not consistent with previous description of the regulation of PI4P5K by PA. Effectively, this could be seen as a thermodynamic counter-argument against the mechanism of regulation suggested by Stace et al. (2008).

So to continue with this model I will have to make the (thermodynamically questionable) assumption that the exchange reaction is strictly irreversible. (Note that this is a common assumption for catalytic reactions, and so is not completely implausible.) Under this scheme, to observe activation, the rate of formation of the PI4P5K-substrate complex must be greater via the PI4P5K-PA complex than the direct route. That is, the total rate of the two reactions:



must be faster than the single reaction:

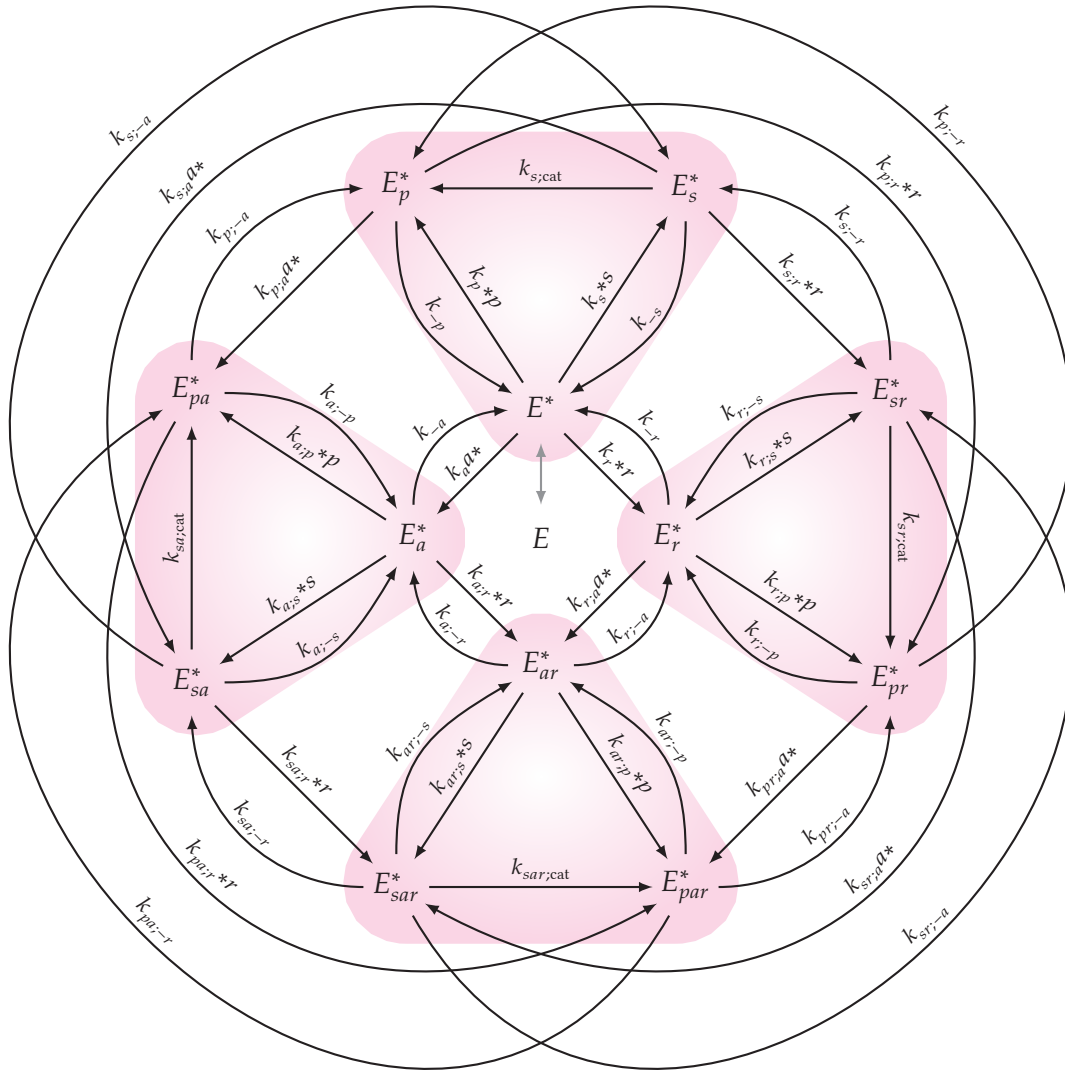


I will confirm that this mechanism leads to (steady-state) activation in section 7.4.6.

In total, the reaction scheme for the 2-site mechanism is drawn as figure 5.11B.

## 5.5 PI4P5K: QSS models

Now that I have described two plausible mechanisms for PI4P5K, I am able to construct suitable quasi-steady-state models based on these mechanisms.



**Figure 5.12** Graph on the enzyme complexes for the 3-site model of PI4P5K. The highlighted regions correspond to pools of complexes that have interconversions comprising of only substrate/product binding, unbinding, and catalysis. The adsorption process ( $E \rightleftharpoons E^*$ ) has been indicated on this graph, however this has not been included in the derivation of the quasi-steady-state models, and is considered separately.



In this section I will use the shorthand:

$$\begin{array}{ll}
 \text{PI4P5K} \rightarrow E, & [\text{PI4P5K}] \rightarrow e; \\
 \text{PIP} \rightarrow *S, & [\text{PIP}] \rightarrow *s; \\
 \text{PI(4,5)P}_2 \rightarrow *P, & [\text{PI(4,5)P}_2] \rightarrow *p; \\
 \text{Arf} \rightarrow A*, & [\text{Arf}] \rightarrow a*; \\
 \text{PI(4,5)P}_2 \rightarrow *R, & [\text{PI(4,5)P}_2] \rightarrow *r.
 \end{array}$$

I will also use the notation for rates and complexes defined in section 3.2.1. Species/complexes will be pre/appended by an asterisk (\*) according to the rules in section 3.5.1.

I will assume that PI4P5K follows an adsorption process as discussed in section 3.5. This process will be ignored in the derivations of quasi-steady-state models for the enzyme, and instead will be modelled separately using the equation

$$f_{\text{adsorp}}(e, e^*) = k_{\text{on}}\Phi\alpha e - k_{\text{off}}e^*$$

where  $\Phi$  is the available area function, and  $\alpha$  is the surface area per unit volume. (I cannot continue to use the notation  $[A]$  as this would conflict with the notation for Arf.)

### 5.5.1 3-site QSS model

The 3-site mechanism for PI4P5K is illustrated in figure 5.11A. This can be drawn as the graph on the enzyme complexes shown in figure 5.12. The corresponding system of ODEs is

$$\begin{aligned}
 \frac{de}{dt} &= -f_{\text{adsorp}}(e, e^*) \\
 \frac{de^*}{dt} &= +f_{\text{adsorp}}(e, e^*) - (k_a a^* + k_r^* r + k_s^* s + k_p^* p)e_a^* + k_{-s}e_s^* + k_{-p}e_p^* + k_{-a}e_a^* + \\
 &\quad k_{-r}e_r^* \\
 \frac{de_s^*}{dt} &= - (k_{-s} + k_{\text{cat}} + k_{s;a}a^* + k_{s;r}^* r)e_s^* + k_s e^* + k_{s;-a}e_{sa}^* + k_{s;-r}e_{sr}^* \\
 \frac{de_p^*}{dt} &= - (k_{-p} + k_{p;a}a^* + k_{p;r}^* r)e_p^* + k_{\text{cat}}e_s^* + k_p^* p e^* + k_{p;-a}e_{pa}^* + k_{p;-r}e_{pr}^* \\
 \frac{de_a^*}{dt} &= - (k_{a;s}^* s + k_{a;p}^* p + k_{-a} + k_{a;r}^* r)e_a^* + k_{a;-s}e_{sa}^* + k_{a;-p}e_{pa}^* + k_a a^* e^* + \\
 &\quad k_{a;-r}e_{ar}^* \\
 \frac{de_{sa}^*}{dt} &= - (k_{a;-s} + k_{a;\text{cat}} + k_{s;-a} + k_{sa;r}^* r)e_{sa}^* + k_{ar;s}e_a^* + k_{s;a}a^* e_s^* + k_{sa;-r}e_{sar}^*
 \end{aligned}$$

$$\begin{aligned}
\frac{de_{pa}^*}{dt} &= - (k_{a;-p} + k_{p;-a} + k_{pa;r} * r) e_{pa}^* + k_{a;cat} e_{sa}^* + k_{a;p} * p e_a^* + k_{p;a} a * e_p^* + k_{pa;-r} e_{par}^* \\
\frac{de_r^*}{dt} &= - (k_{r;s} * s + k_{r;p} * p + k_{r;a} a * + k_{-r}) e_r^* + k_{r;-s} e_{sr}^* + k_{r;-p} e_{pr}^* + k_{r;-a} e_{ar}^* + k_r * r e^* \\
\frac{de_{sr}^*}{dt} &= - (k_{r;-s} + k_{r;cat} + k_{sr;a} a * + k_{s;-r}) e_{sr}^* + k_{s;r} * r e_s^* + k_{sr;-a} e_{sar}^* + k_{r;s} * s e_r^* \\
\frac{de_{pr}^*}{dt} &= - (k_{r;-p} + k_{pr;a} a * + k_{p;-r}) e_{pr}^* + k_{r;cat} e_{ra}^* + k_{r;p} * p e_r^* + k_{pr;-a} e_{par}^* + k_{p;r} * r e_p^* \\
\frac{de_{ar}^*}{dt} &= - (k_{r;-a} + k_{a;-r} + k_{ar;s} * s + k_{ar;p} * p) e_{ar}^* + k_{r;a} a * e_r^* + k_{a;r} * r e_a^* + k_{ar;-s} e_{sar}^* + \\
&\quad k_{ar;-p} e_{par}^* \\
\frac{de_{sar}^*}{dt} &= - (k_{ar;-s} + k_{ar;cat} + k_{sr;-a} + k_{sa;-r}) e_{sar}^* + k_{ar;s} * s e_{ar}^* + k_{sr;a} a * e_{sr}^* + k_{sa;r} * r e_{sa}^* \\
\frac{de_{par}^*}{dt} &= - (k_{ar;-p} + k_{pr;-a} + k_{pa;-r}) e_{par}^* + k_{ar;cat} e_{sar}^* + k_{ar;p} * p e_{ar}^* + k_{pr;a} a * e_{pr}^* + \\
&\quad k_{pa;r} * r e_{pa}^* \\
\frac{da^*}{dt} &= - (k_a e^* + k_{s;a} e_s^* + k_{p;a} e_p^* + k_{r;a} e_r^* + k_{sr;a} e_{sr}^* + k_{pr;a} e_{pr}^*) a^* + k_{-a} e_a^* + k_{s;-a} e_{sa}^* \\
&\quad + k_{p;-a} e_{pa}^* + k_{r;-a} e_{ar}^* + k_{sr;-a} e_{sar}^* + k_{pr;-a} e_{par}^* \\
\frac{d*r}{dt} &= - (k_r e^* + k_{s;r} e_s^* + k_{p;r} e_p^* + k_{a;r} e_a^* + k_{sa;r} e_{sa}^* + k_{pa;r} e_{pa}^*) * r + k_{-r} e_r^* + k_{s;-r} e_{sr}^* \\
&\quad + k_{p;-r} e_{pr}^* + k_{a;-r} e_{ar}^* + k_{sa;-r} e_{sar}^* + k_{pa;-r} e_{par}^* \\
\frac{d*s}{dt} &= - (k_s e^* + k_{a;s} e_a^* + k_{r;s} e_r^* + k_{ar;s} e_{ar}^*) * s + k_{-s} e_s^* + k_{a;-s} e_{sa}^* + k_{r;-s} e_{sr}^* + \\
&\quad k_{ar;-s} e_{sar}^* \\
\frac{d*p}{dt} &= - (k_p e^* + k_{a;p} e_a^* + k_{r;p} e_r^* + k_{ar;p} e_{ar}^*) * p + k_{-p} e_p^* + k_{a;-p} e_{pa}^* + k_{r;-p} e_{pr}^* + \\
&\quad k_{ar;-p} e_{par}^* .
\end{aligned}$$

I would like to use the linear framework of Gunawardena (2012), described in section 3.3, to construct a reduced model based on this system (ignoring the adsorption process). The number of spanning trees was found—using the algorithm in section 3.4.1—with the following Python code:

```

1 from sympy import *
2 from enzymegraph import *
3
4 # setup symbols
5 s, p, a, r = symbols('s p a r', positive=True)
6 e, ea, er, ear = symbols('E E_a E_r E_ar', positive=True)
7 es, esa, esr, esar = symbols('E_s E_sa E_sr E_sar', positive=True)
8 ep, epa, epr, epar = symbols('E_p E_pa E_pr E_par', positive=True)
9
10 # list of edges, with arbitrary parameter labels
11 k = parameter_generator()
12

```

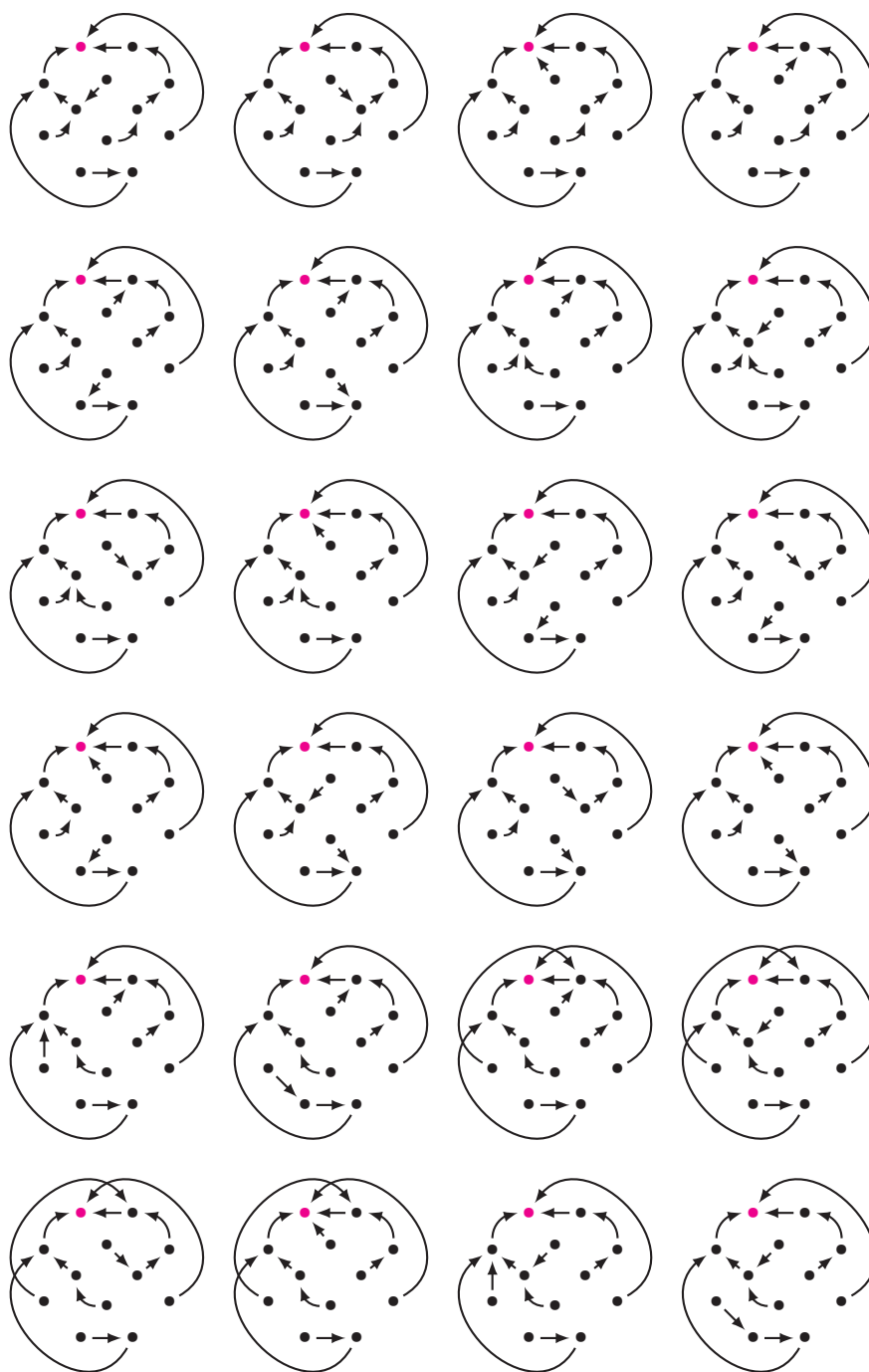
```

13 edges = {
14     # no substrate/product bound
15     (e, ea) : next(k), (ea, e) : next(k),
16     (e, er) : next(k), (er, e) : next(k),
17     (er, ear) : next(k), (ear, er) : next(k),
18     (ea, ear) : next(k), (ear, ea) : next(k),
19     # substrate bound
20     (es, esa) : next(k), (esa, es) : next(k),
21     (es, esr) : next(k), (esr, es) : next(k),
22     (esr, esar) : next(k), (esar, esr) : next(k),
23     (esa, esar) : next(k), (esar, esa) : next(k),
24     # add/remove substrate
25     (e, es) : next(k), (es, e) : next(k),
26     (ea, esa) : next(k), (esa, ea) : next(k),
27     (er, esr) : next(k), (esr, er) : next(k),
28     (ear, esar) : next(k), (esar, ear) : next(k),
29     # product bound
30     (ep, epa) : next(k), (epa, ep) : next(k),
31     (ep, epr) : next(k), (epr, ep) : next(k),
32     (epr, epar) : next(k), (epar, epr) : next(k),
33     (epa, epar) : next(k), (epar, epa) : next(k),
34     # add/remove product
35     (e, ep) : next(k), (ep, e) : next(k),
36     (ea, epa) : next(k), (epa, ea) : next(k),
37     (er, epr) : next(k), (epr, er) : next(k),
38     (ear, epar) : next(k), (epar, ear) : next(k),
39     # catalysis
40     (es, ep) : next(k),
41     (esa, epa) : next(k),
42     (esr, epr) : next(k),
43     (esar, epar) : next(k),
44 }
45
46 # generate graph
47 graph = enzymegraph(edges)
48
49 # count spanning trees
50 print(len([1 for _ in graph.spanning_trees()]))

```

Thus this graph can be shown to have 1,612,800 distinct spanning trees. This is many hundreds of times the number of spanning trees obtained for the PLD mechanism, which I already determined was too many to sensibly enumerate or manipulate. A (very) small proportion of these have been drawn in figure 5.13.

Again, I need to amend my approach so as to introduce sufficient simplification, so that I can sensibly develop a reduced model. I will take the same approach as I took for the PLD model. That is, I assume that dynamics related to interactions with the substrate, and now also product, can—in some sense—be separated from



**Figure 5.13** A small proportion of the 1,612,800 spanning trees of the graph on the enzyme complexes for the mechanism shown in figure 5.12.

dynamics related to interactions with the regulators. Specifically, I will assume that substrate/product binding, unbinding, and catalysis follow the rapid equilibrium assumption, as specified in section 3.2.4. These interactions have been highlighted on the graph in figure 5.12. These again define pools of related complexes, for which I intend to introduce new variables corresponding to the total concentrations of the complexes in these pools. With additional assumptions on the rate constants of the system, I will again be able to collapse the graph in figure 5.12 and produce a simplified system with a smaller graph.

The four independent (sub-)graphs shown in figure 5.14 correspond to the highlighted regions in the graph in figure 5.12. Each of these is equivalent to the Michaelis-Menten mechanism with product inhibition, as was described in section 3.3.2. Therefore, we can immediately write down the quasi-steady-state solutions for the enzyme complexes from equations 3.12 to 3.14. In general—for  $\sigma \subset \{a, r\}$ —these are

$$e_{\sigma}^* = \left( \frac{K_{\sigma;s}}{K_{\sigma;s} \left( 1 + \frac{*p}{K_{\sigma;ic}} \right) + *s} \right) \tilde{\zeta}_{\sigma} \quad (5.20)$$

$$e_{\sigma \cup \{s\}}^* = \left( \frac{\left( 1 - \frac{k'_{\sigma,cat}}{k_{\sigma;-p}} \right) *s}{K_{\sigma;s} \left( 1 + \frac{*p}{K_{\sigma;ic}} \right) + *s} \right) \tilde{\zeta}_{\sigma} \quad (5.21)$$

$$e_{\sigma \cup \{p\}}^* = \left( \frac{\frac{K_{\sigma;s}}{K_{\sigma;ic}} p + \frac{k'_{\sigma,cat}}{k_{\sigma;-p}} *s}{K_{\sigma;s} \left( 1 + \frac{*p}{K_{\sigma;ic}} \right) + *s} \right) \tilde{\zeta}_{\sigma} \quad (5.22)$$

where

$$\tilde{\zeta}_{\sigma} = e_{\sigma}^* + e_{\sigma \cup \{s\}}^* + e_{\sigma \cup \{p\}}^*.$$

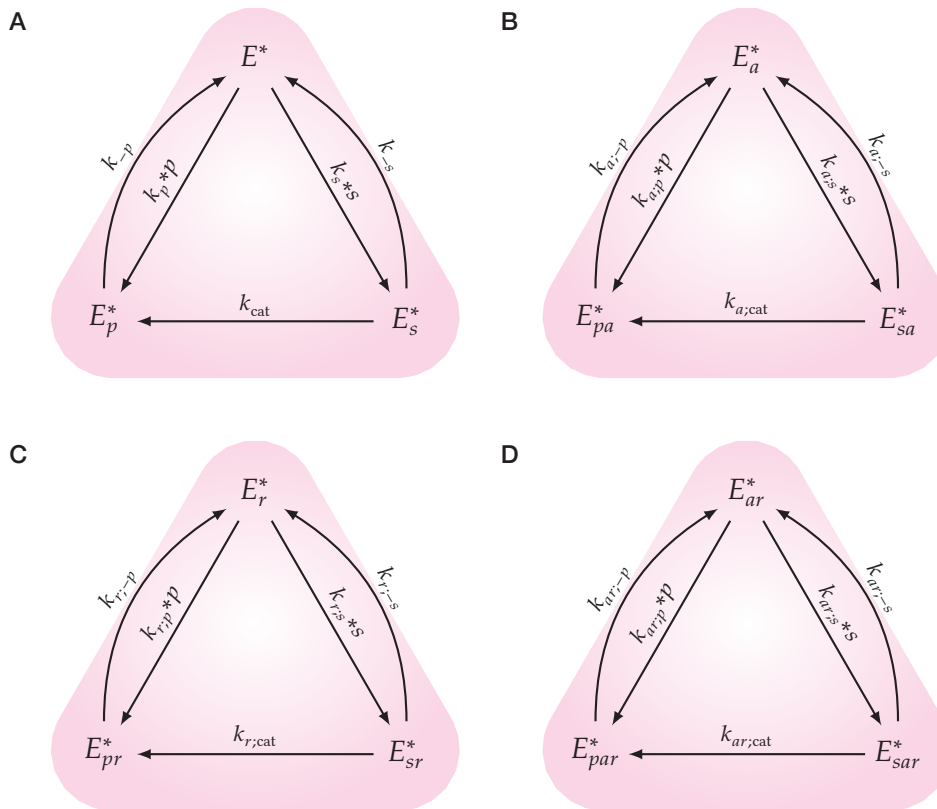
The  $\tilde{\zeta}_{\sigma}$  are again variables which correspond to the total concentrations in the pool of complexes.

We can also write down the rates of change of each  $\tilde{\zeta}_{\sigma}$ , for example

$$\begin{aligned} \frac{d\tilde{\zeta}}{dt} &= \frac{de^*}{dt} + \frac{de_s^*}{dt} + \frac{de_p^*}{dt} \\ &= f_{\text{adsorp}}(e, e^*) \\ &\quad - (k_a a^* + k_r *r) e^* - (k_{s;a} a^* + k_{s;r} *r) e_s^* - (k_{p;a} a^* + k_{p;r} *r) e_p^* + k_{-a} e_a^* \\ &\quad \quad \quad + k_{s;-a} e_{sa}^* + k_{p;-a} e_{pa}^* + k_{-r} e_r^* + k_{s;-r} e_{sr}^* + k_{p;-r} e_{pr}^* \\ &= f'_{\text{adsorp}}(e, \tilde{\zeta}) - \frac{f}{g} \tilde{\zeta} + \frac{f_a}{g_a} \tilde{\zeta}_a + \frac{f_r}{g_r} \tilde{\zeta}_r \end{aligned}$$

where

$$f'_{\text{adsorp}}(e, \tilde{\zeta}) = k_{\text{on}} \Phi a e - k_{\text{off}} \left( \frac{K_s}{g} \right) \tilde{\zeta}$$



**Figure 5.14** Subgraphs of the 3-site PI4P5K graph, the highlighted regions in figure 5.12 corresponding to pools of complexes that have interconversions comprising of only substrate/product binding, unbinding, and catalysis.

and

$$\begin{aligned}
g_\sigma &= K_{\sigma;s} \left( 1 + \frac{*p}{K_{\sigma;ic}} \right) + *s \\
f &= (k_a a * + k_r * r) K_s + (k_{s;a} a * + k_{s;r} * r) \left( 1 - \frac{k'_{cat}}{k_{-p}} \right) *s + \\
&\quad (k_{p;a} a * + k_{p;r} * r) \left( \frac{K_s}{K_{ic}} *p + \frac{k'_{cat}}{k_{-p}} *s \right) \\
f_a &= k_{-a} K_{a;s} + k_{s;-a} \left( 1 - \frac{k'_{a;cat}}{k_{a;-p}} \right) *s + k_{p;-a} \left( \frac{K_{a;s}}{K_{a;ic}} *p + \frac{k'_{a;cat}}{k_{a;-p}} *s \right) \\
f_r &= k_{-r} K_{r;s} + k_{s;-r} \left( 1 - \frac{k'_{r;cat}}{k_{r;-p}} \right) *s + k_{p;-r} \left( \frac{K_{r;s}}{K_{r;ic}} *p + \frac{k'_{r;cat}}{k_{r;-p}} *s \right).
\end{aligned}$$

This can be simplified in the same manner as the PLD model, by assuming that the rates of binding and unbinding of Arf and PA are independent of whether substrate/product is bound,

$$\begin{aligned}
k_a &= k_{s;a} = k_{p;a}, & k_{-a} &= k_{s;-a} = k_{p;-a}, \\
k_r &= k_{s;r} = k_{p;r}, & k_{-r} &= k_{s;-r} = k_{p;-r}, \\
k_{r;a} &= k_{sr;a} = k_{pr;a}, & k_{r;-a} &= k_{sr;-a} = k_{pr;-a}, \\
k_{a;r} &= k_{sa;r} = k_{pa;r}, & k_{a;-r} &= k_{sa;-r} = k_{pa;-r}.
\end{aligned}$$

Using these replacements gives

$$\frac{f}{g} = (k_a a * + k_r * r); \quad \frac{f_a}{g_a} = k_{-a}; \quad \frac{f_r}{g_r} = k_{-r}$$

and so the above equation simplifies to

$$\frac{d\zeta}{dt} = f'_{adsorp}(e, \zeta) - (k_a a * + k_r * r) \zeta + k_{-a} \zeta_a + k_{-r} \zeta_r.$$

Similarly,

$$\begin{aligned}
\frac{d\zeta_a}{dt} &= -(k_{-a} + k_{a;r} * r) \zeta_a + k_a a \zeta + k_{-r} \zeta_{ar} \\
\frac{d\zeta_r}{dt} &= -(k_{r;a} a * + k_{-r}) \zeta_r + k_r r \zeta + k_{-a} \zeta_{ar} \\
\frac{d\zeta_{ar}}{dt} &= -(k_{r;-a} + k_{a;-r}) \zeta_{ar} + k_{a;r} r \zeta_a + k_{r;a} a \zeta_r.
\end{aligned}$$

Ignoring the adsorption process described by  $f'_{adsorp}(e, \zeta)$ , these equations again match the equations that describe the di-allosteric mechanism, equations 5.1 to 5.6. So the membrane-associated part of the system has been reduced to the di-allosteric mechanism, and can also be described using the graph in figure 5.2 by appropriate relabelling of the vertices. From the original derivation I can immediately write down

the quasi-steady-state values of the  $\zeta$ . These were originally given as equations 5.7 to 5.10. We now need to allow for the mass of enzyme which remains cytosolic, which is not included in the di-allosteric mechanism, and so replace  $e_0$  by  $(e_0 - e)$ .

This means

$$\zeta = \left( \frac{K_r K_{r;a}}{K_r K_{r;a} + K_{a;r} a^* + K_{r;a}^* r + a^* r} \right) (e_0 - e) \quad (5.23)$$

$$\zeta_a = \left( \frac{K_{a;r} a^*}{K_r K_{r;a} + K_{a;r} a^* + K_{r;a}^* r + a^* r} \right) (e_0 - e) \quad (5.24)$$

$$\zeta_r = \left( \frac{K_{r;a}^* r}{K_r K_{r;a} + K_{a;r} a^* + K_{r;a}^* r + a^* r} \right) (e_0 - e) \quad (5.25)$$

$$\zeta_{ar} = \left( \frac{a^* r}{K_r K_{r;a} + K_{a;r} a^* + K_{r;a}^* r + a^* r} \right) (e_0 - e) \quad (5.26)$$

where

$$e_0 = \zeta + \zeta_a + \zeta_r + \zeta_{ar} + e = e + e_s + e_a + e_{sa} + e_r + e_{sr} + e_{ar} + e_{sar} + e.$$

We now have enough information to construct a model for the 3-site mechanism for PI4P5K. We need to start with the net rate of production of free PI(4,5)P<sub>2</sub> by the system. In this case, this is given by the difference between the forwards and reverse reactions between the complexes  $E_{\sigma \cup \{p\}}^*$  and  $E_\sigma^*$  for each  $\sigma \subset \{a, r\}$ . From the original system of ODEs we know that

$$\frac{d^*p}{dt} = \sum_{\sigma \subset \{a, r\}} (k_{\sigma; -p} e_{\sigma \cup \{p\}}^* - k_{\sigma; p}^* p e_\sigma^*)$$

Substituting in values of  $e_\sigma^*$  and  $e_{\sigma \cup \{p\}}^*$  from equations 5.20 to 5.22 gives

$$\begin{aligned} \frac{d^*p}{dt} &= \sum_{\sigma \subset \{a, r\}} \left( \frac{k_{\sigma; -p} \left( \frac{k_{\sigma; s}}{K_{\sigma; ic}} *p + \frac{k'_{\sigma, cat}}{k_{\sigma; -p}} *s \right) - k_{\sigma; p}^* k_{\sigma; s} *p}{k_{\sigma; s} \left( 1 + \frac{*p}{K_{\sigma; ic}} \right) + *s} \right) \zeta_\sigma \\ &= \sum_{\sigma \subset \{a, r\}} \left( \frac{k'_{\sigma, cat} *s}{k_{\sigma; s} \left( 1 + \frac{*p}{K_{\sigma; ic}} \right) + *s} \right) \zeta_\sigma \end{aligned}$$

Substituting in the values of  $\zeta_\sigma$  from equations 5.23 to 5.26, and rearranging, gives

$$\begin{aligned} \frac{d^*p}{dt} &= \left( \frac{k'_{cat} K_r K_{r;a}}{K_s \left( 1 + \frac{*p}{K_{ic}} \right) + *s} + \frac{k'_{a, cat} K_{a;r} a^*}{K_{a; s} \left( 1 + \frac{*p}{K_{a; ic}} \right) + *s} + \frac{k'_{r, cat} K_{r;a}^* r}{K_{r; s} \left( 1 + \frac{*p}{K_{r; ic}} \right) + *s} \right. \\ &\quad \left. + \frac{k'_{ar, cat} a^* r}{K_{ar; s} \left( 1 + \frac{*p}{K_{ar; ic}} \right) + *s} \right) \left( \frac{*s(e_0 - e)}{K_r K_{r;a} + K_{a;r} a^* + K_{r;a}^* r + a^* r} \right) \end{aligned} \quad (5.27)$$



Equation 5.27 describes the most general version of this model, with 4 variables and 17 parameters. This equation would be useful if we wanted to investigate the effects of substrate/product binding or catalytic rates that depended on the presence of either or both regulator. However, this full form is verbose, and a simplified form of this model is desirable to comprise part of a model of the the Arf/PLD/PI4P5K motif.

I will make the following additional assumptions:

1. That catalytic rates are equal,

$$k_{\text{cat}} = k_{a;\text{cat}} = k_{r;\text{cat}} = k_{ar;\text{cat}}.$$

2. That the equilibrium constants for substrate binding/unbinding are equal,

$$K_s = K_{a;s} = K_{r;s} = K_{ar;s}.$$

3. That the (product) inhibition constants are equal,

$$K_{\text{ic}} = K_{a;\text{ic}} = K_{r;\text{ic}} = K_{ar;\text{ic}}.$$

With these assumptions the model reduces to

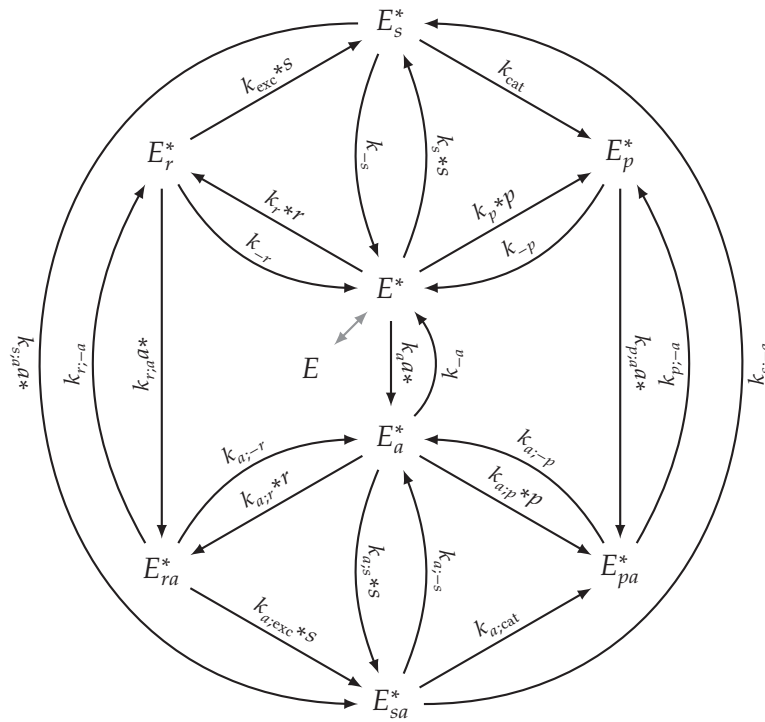
$$\frac{d^*p}{dt} = \frac{k_{\text{cat}}^*s (e_0 - e)}{K_s \left(1 + \frac{^*p}{K_{\text{ic}}}\right) + ^*s}. \quad (5.28)$$

This is in the form of the original equation for Michaelis-Menten with product inhibition given by equation 3.15, except with the total concentration is now given by  $(e_0 - e)$ . Perhaps surprisingly, this equation has no reliance on the concentrations of the regulators,  $a^*$  and  $r^*$ . Instead the concentrations of the regulators affect the maximum rate ( $V_{\text{max}}$ ) by influencing the amount of enzyme present (on the membrane). So how do the Arf and PA effect PI4P5K? We still require an additional equation for the concentration cytosolic PI4P5K,  $e$ . From the original system of ODEs we know that

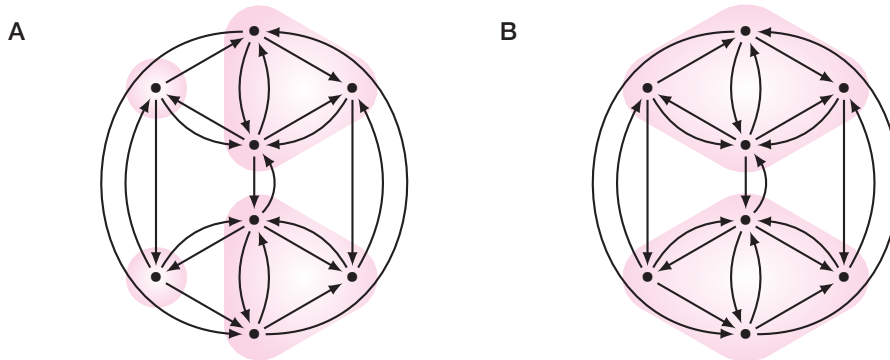
$$\begin{aligned} \frac{de}{dt} &= -f_{\text{adsorp}}(e, e^*) \\ &= -k_{\text{on}}\Phi ae + k_{\text{off}}e^* \end{aligned}$$

Substituting in the value of  $e^*$  from equation 5.20 gives

$$\frac{de}{dt} = -k_{\text{on}}\Phi ae + k_{\text{off}} \left( \frac{K_s}{K_s \left(1 + \frac{^*p}{K_{\text{ic}}}\right) + ^*s} \right) \zeta$$



**Figure 5.15** Graph on the enzyme complexes for the 2-site model of PI4P5K. The adsorption process ( $E \rightleftharpoons E^*$ ) has been indicated on this graph, however this has not been included in the derivation of the quasi-steady-state models, and is considered separately.



**Figure 5.16** Possible approaches for the simplification of the graph shown in figure 5.15. The highlighted regions correspond to pools of complexes that can be separated out of the full graph.

- A** PA is considered part of the (independent) regulatory processes.
- B** PA is considered part of the catalytic process.

Substituting in the value of  $\zeta$  from equation 5.23 gives

$$\frac{de}{dt} = -k_{\text{on}}\Phi ae + k_{\text{off}} \left( \frac{k_{\text{off}}K_s}{K_s \left(1 + \frac{*p}{K_{\text{ic}}}\right) + *s} \right) \left( \frac{K_r K_{r;a} (e_0 - e)}{K_r K_{r;a} + K_{a;r}a + K_{r;a}r + ar} \right)$$

To simplify this I will introduce one further assumption:

4. That the equilibrium constant for PA binding/unbinding is independent of whether Arf is bound,

$$K_{a;r} = K_r.$$

With this, the equation reduces to

$$\frac{de}{dt} = -k_{\text{on}}\Phi ae + \frac{k_{\text{off}}K_s K_r K_{r;a} (e_0 - e)}{\left(K_s \left(1 + \frac{*p}{K_{\text{ic}}}\right) + *s\right) (K_{r;a} + a*) (K_r + *r)}. \quad (5.29)$$

Together, equations 5.28 and 5.29 constitute what I will call the 3-site model of PI4P5K. In this model, Arf and PA only contribute to the amount of enzyme on the membrane and do not contribute to the actual kinetic rate of catalysis. We can see that as the concentration of either Arf ( $a*$ ) or PA ( $*r$ ) increases, the amount of cytosolic enzyme ( $e$ ) decreases. I will analyse the behaviours of this model further in chapter 7.

## 5.5.2 2-site QSS model

The 2-site mechanism for PI4P5K is illustrated in figure 5.11B. This can be drawn as the graph on the enzyme complexes shown in figure 5.15. The corresponding system of ODEs is

$$\begin{aligned} \frac{de^*}{dt} &= -(k_r *r + k_s *s + k_p *p + k_a a*)e^* + k_{-r}e_r^* + k_{-s}e_s^* + k_{-p}e_p^* + k_{-a}e_a^* \\ \frac{de_r^*}{dt} &= -(k_{\text{exc}} *s + k_{-r} + k_{r;a} a*)e_r^* + k_r *r e^* + k_{r;-a} e_{ra}^* \\ \frac{de_s^*}{dt} &= -(k_{\text{cat}} + k_{-s} + k_{r;a} a*)e_s^* + k_{\text{exc}} *s e_r^* + k_s *s e^* + k_{s;-a} e_{sa}^* \\ \frac{de_p^*}{dt} &= -(k_{-p} + k_{p;a} a*)e_p^* + k_{\text{cat}} e_s^* + k_p *p e^* + k_{p;-a} e_{pa}^* \\ \frac{de_a^*}{dt} &= -(k_{a;r} *r + k_{a;s} *s + k_{a;p} *p + k_{-a})e_a^* + k_{a;-r} e_{ar}^* + k_{a;-s} e_{as}^* + k_{a;-p} e_{ap}^* + k_a a *e^* \\ \frac{de_{ra}^*}{dt} &= -(k_{a;\text{exc}} *s + k_{a;-r} + k_{r;-a})e_{ra}^* + k_{a;r} *r e_a^* + k_{r;a} a *e_r^* \end{aligned}$$

$$\begin{aligned} \frac{de_{sa}^*}{dt} &= -(k_{a;cat} + k_{a;-s} + k_{s;-a})e_{sa}^* + k_{a;exc} * se_{ar}^* + k_{a;s} * se_a^* + k_{s;a} a * e_s^* \\ \frac{de_{pa}^*}{dt} &= -(k_{a;-p} + k_{p;-a})e_{pa}^* + k_{a;cat} e_{sa}^* + k_{a;p} * pe_a^* + k_{p;a} a * e_p^* \\ \frac{da^*}{dt} &= -(k_a e^* + k_{r;a} e_r^* + k_{s;a} e_s^* + k_{p;a} e_p^*) a^* + k_{-a} e_a^* + k_{r;-a} e_{ra}^* + k_{s;-a} e_{sa}^* + k_{p;-a} e_{pa}^* \\ \frac{d*r}{dt} &= -(k_r e^* + k_{a;r} e_a^*) * r + (k_{exc} e_r^* + k_{a;exc} e_{ar}^*) * s + k_{-r} e_r^* + k_{a;-r} e_{ra}^* \\ \frac{d*s}{dt} &= -(k_s e^* + k_{a;s} e_a^* + k_{exc} e_r^* + k_{a;exc} e_{ra}^*) * s + k_{-s} e_s^* + k_{a;-s} e_{sa}^* \\ \frac{d*p}{dt} &= -(k_p e^* + k_{a;p} e_a^*) * p + k_{-p} e_p^* + k_{a;-p} e_{pa}^* \end{aligned}$$

This is a system of 12 variables and 24 parameters.

Of particular interest are the *exchange* reactions, which are



These model the release of PA and the binding of PI(4,5)P<sub>2</sub> in a single reaction step. (Therefore, free PA is generated at the same rate.)

Once more, I would like to use the linear framework of Gunawardena (2012), described in section 3.3, to construct a reduced model based on this system (again minus the adsorption process). The number of spanning trees was found—using the algorithm in section 3.4.1—with the following Python code:

```

1 from sympy import *
2 from enzymegraph import *
3
4 # setup symbols
5 s, p, a, r = symbols('s p a r', positive=True)
6 e, ea, er, ear = symbols('E E_a E_r E_ar', positive=True)
7 er, era = symbols('E_r E_ra', positive=True)
8 es, esa = symbols('E_s E_sa', positive=True)
9 ep, epa = symbols('E_p E_pa', positive=True)
10
11 # list of edges, with arbitrary parameter labels
12 k = parameter_generator()
13 edges = {
14     # on/off substrate/product/regulator
15     (e, es) : next(k), (es, e) : next(k),
16     (e, ep) : next(k), (ep, e) : next(k),
17     (e, er) : next(k), (er, e) : next(k),
18     (ea, esa) : next(k), (esa, ea) : next(k),
19     (ea, epa) : next(k), (epa, ea) : next(k),
20     (ea, era) : next(k), (era, ea) : next(k),

```

```

21 # on/off arf
22 (e, ea) : next(k), (ea, e) : next(k),
23 (es, esa) : next(k), (esa, es) : next(k),
24 (ep, epa) : next(k), (epa, ep) : next(k),
25 (er, era) : next(k), (era, er) : next(k),
26 # exchange/catalysis
27 (er, es) : next(k), (era, esa) : next(k),
28 (es, ep) : next(k), (esa, epa) : next(k),
29 }
30
31 # generate graph
32 graph = enzymegraph(edges)
33
34 # count spanning trees
35 print(len([1 for _ in graph.spanning_trees()]))

```

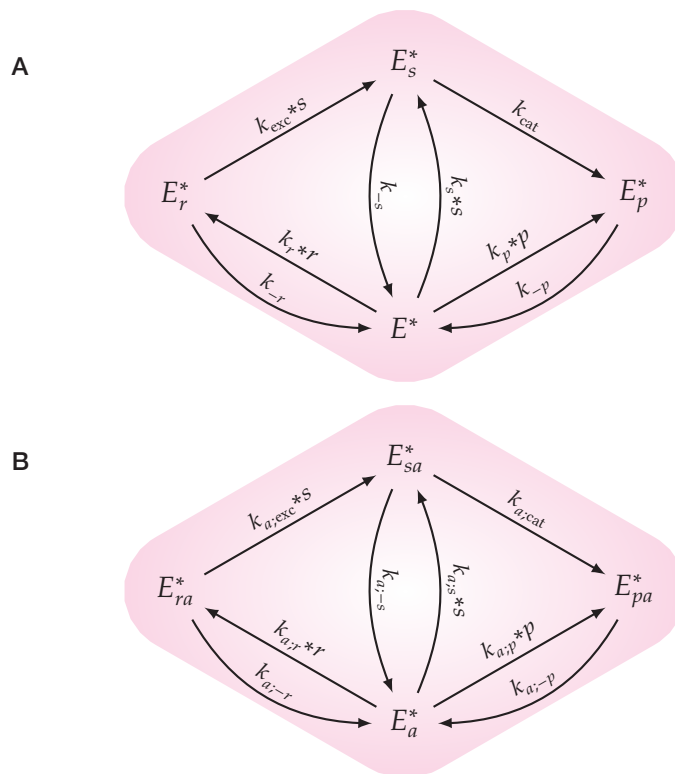
This shows we again have too many (2240) spanning trees to sensibly enumerate and manipulate. So I will take the same approach as before, and separate the dynamics related to each of catalysis and regulation into distinct components. However, I now have a choice:

- If I include PA in the regulation component, then this is equivalent to the choice made for the previous 3-site models. This choice is illustrated in figure 5.16A. We already know the rapid equilibrium values of the pools of complexes:  $e_r^* = \zeta_r$ ;  $e_{ar}^* = \zeta_{ar}$ ; and the others are defined by the quasi-steady-state solutions for Michaelis-Menten with product inhibition as per the 3-site model, given by equations 5.20 to 5.22. However, we cannot directly use the results of the di-allosteric mechanism, due to a loss of symmetry between the pools of complexes. In this system, the rates of exchange between the pools will not match those of the di-allosteric mechanism. For example

$$\begin{aligned} \frac{d\zeta_r}{dt} = \frac{de_r}{dt} &= -(k_{\text{exc}}*s + k_{-r} + k_{r,a}a*)e_r^* + k_r*re^* + k_{r,-a}e_{ra}^* \\ &= -(k_{\text{exc}}*s + k_{-r} + k_{r,a}a*)\zeta_{ra} + k_r*r \left( \frac{K_{*s}}{K_s \left( 1 + \frac{*p}{K_{ic}} \right) + *s} \right) \zeta + k_{r,-a}\zeta_r \end{aligned}$$

which, even with similar equality constraints on the rate constants as before, does not simplify to the relevant equation for the di-allosteric mechanism. So the labels on the reduced graph would be different, and so the di-allosteric mechanism would have to be re-derived using these new (complicated, fractional) rates as edge labels.

- If I include PA in the catalytic component, then I will need to derive new models for both components. This choice is illustrated in figure 5.16B. However,



**Figure 5.17** The catalytic components, subgraphs of the 2-site PI4P5K graph, the highlighted regions in figure 5.16B, corresponding to pools of complexes that have interconversions comprising of substrate/product binding, unbinding, and catalysis; and PA binding, unbinding, and exchange.

this is arguably more satisfactory as it means that all dynamics related to the substrate binding site are considered together and, as we shall see, the regulatory component is described by a simple reversible binding process. In this case it then only remains to derive a new model of the catalytic component.

I will choose the second of these assumptions, and include the PA interactions within the catalytic component. This gives the two independent (sub-)graphs in figure 5.17 which correspond to the highlighted regions of the graph in figure 5.16B. I will consider the general graph where  $\sigma \subset \{a\}$  (so  $\sigma = \emptyset$  or  $\sigma = \{a\}$ ). The 16 spanning trees for this graph are shown in figure 5.18. This gives the basis element

$$\rho = \begin{bmatrix} k_{\sigma;-p}(k_{\sigma;-s} + k_{\sigma;\text{cat}})(k_{\sigma;-r} + k_{\sigma;\text{exc}}*s) \\ k_{\sigma;-p}k_{\sigma;r}(k_{\sigma;-s} + k_{\sigma;\text{cat}})*r \\ k_{\sigma;-p}(k_{\sigma;\text{exc}}k_{\sigma;r}*r + k_{\sigma;s}(k_{\sigma;-r} + k_{\sigma;\text{exc}}*s))*s \\ (k_{\sigma;-r} + k_{\sigma;\text{exc}}*s)((k_{\sigma;-s} + k_{\sigma;\text{cat}})k_{\sigma;p}*p + k_{\sigma;\text{cat}}k_{\sigma;s}*s) + k_{\sigma;\text{cat}}k_{\sigma;\text{exc}}k_{\sigma;r}*r*s \end{bmatrix}$$

in the order  $(E^*, E_r^*, E_s^*, E_p^*)$ .

It will be helpful to simplify this basis element further. We can first do this in analogy with the method in which the quasi-steady-state equation for Michaelis-Menten was derived in section 3.3.2. Dividing by  $(k_{\sigma;\text{cat}} + k_{\sigma;-p})k_{\sigma;s}$  gives

$$\rho = \begin{bmatrix} (k_{\sigma;-r} + k_{\sigma;\text{exc}}*s)K_{\sigma;m} \\ k_{\sigma;r}K_{\sigma;m}*r \\ \left(1 - \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}}\right)\left(\frac{k_{\sigma;\text{exc}}k_{\sigma;r}}{\sigma;s}*r + k_{\sigma;-r} + k_{\sigma;\text{exc}}*s\right)*s \\ (k_{\sigma;-r} + k_{\sigma;\text{exc}}*s)\left(\frac{K_{\sigma;m}}{K_{\sigma;\text{ic}}}*p + \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}}*s\right) + \frac{k_{\sigma;\text{exc}}k_{\sigma;r}}{\sigma;s}\frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}}*r*s \end{bmatrix}$$

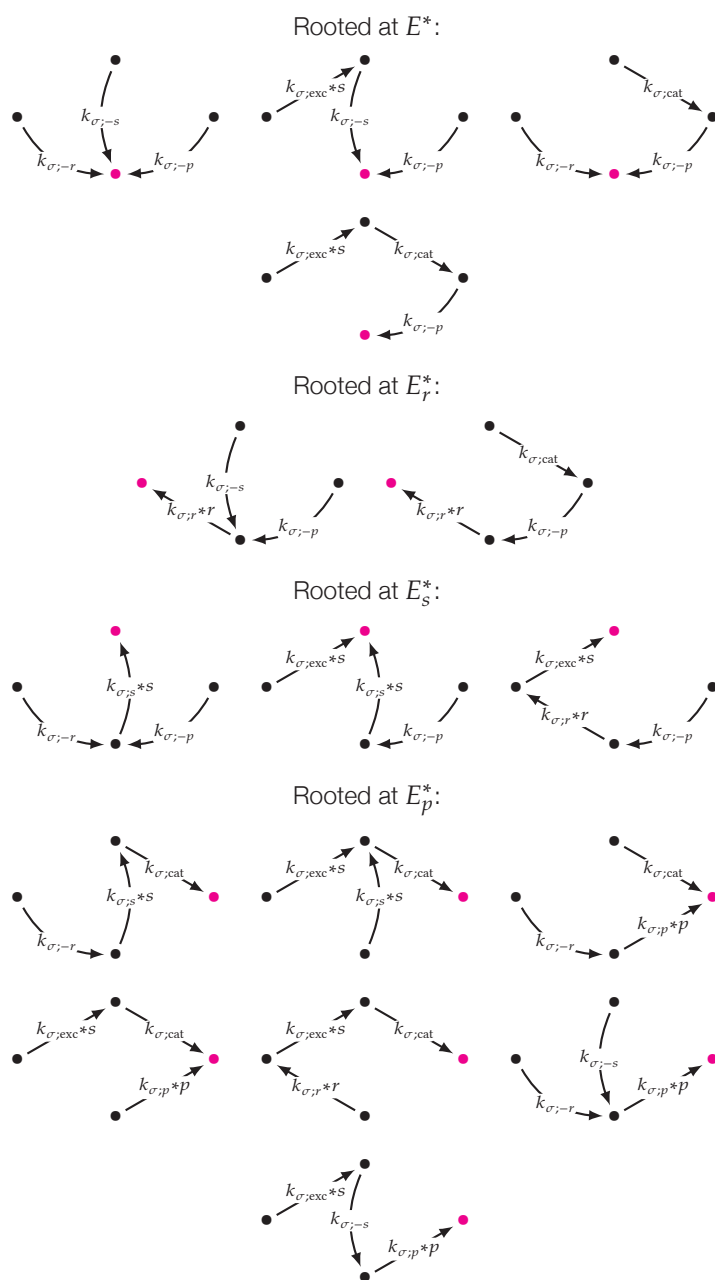
where I have used the relation

$$\frac{k_{\sigma;-p}}{k_{\sigma;\text{cat}} + k_{\sigma;-p}} = 1 - \frac{k_{\sigma;\text{cat}}}{k_{\sigma;\text{cat}} + k_{\sigma;-p}} = 1 - \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}}$$

and where (as before)

$$K_{\sigma;m} = \frac{(k_{\sigma;-s} + k_{\sigma;\text{cat}})k_{\sigma;-p}}{(k_{\sigma;\text{cat}} + k_{\sigma;-p})k_{\sigma;s}}, \quad K_{\sigma;\text{ic}} = \frac{k_{\sigma;-p}}{k_{\sigma;p}}, \quad k'_{\sigma;\text{cat}} = \frac{k_{\sigma;\text{cat}}k_{\sigma;-p}}{k_{\sigma;\text{cat}} + k_{\sigma;-p}}.$$

Next I desire to simplify the components related to PA ( $*r$ ) addition/removal. Note that this portion of the mechanism is similar to the catalysis/product inhibition portion of the mechanism, this may suggest the division by a similar factor to the last step. This would be dividing by  $(k_{\sigma;\text{exc}} + k_{\sigma;-s})k_{\sigma;r}$ . However, now the total rate of exchange is given by the parameter  $k_{\sigma;\text{exc}}$  multiplied by  $*s$  (compared with the



**Figure 5.18** The 16 spanning trees of the graph on the enzyme complexes for the catalytic component of the 2 site mechanism, shown in figure 5.17. The highlighted nodes denote the root of each spanning tree.



total rate of catalysis which is just given by the rate  $k_{\sigma;\text{cat}}$ ). So the system can not be simplified in the same way, as I would like to keep the variable  $*s$  accessible.

Instead, dividing by  $k_{\sigma;r}$  gives

$$\rho = \left[ \begin{array}{c} (1 + k'_{\sigma;\text{exc}} *s) K_{\sigma;r} K_{\sigma;m} \\ K_{\sigma;m} *r \\ \left(1 - \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}}\right) (k_{\sigma;\text{rel}} *r + K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} *s)) *s \\ K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} *s) \left(\frac{K_{\sigma;m}}{K_{\sigma;\text{ic}}} *p + \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}} *s\right) + k_{\sigma;\text{rel}} \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}} *r *s \end{array} \right]$$

where

$$K_{\sigma;r} = \frac{k_{\sigma;-r}}{k_{\sigma;r}}, \quad k'_{\sigma;\text{exc}} = \frac{k_{\sigma;\text{exc}}}{k_{\sigma;-r}}, \quad k_{\sigma;\text{rel}} = \frac{k_{\sigma;\text{exc}}}{k_{\sigma;s}}.$$

Therefore, the quasi-steady-state solutions are

$$e_{\sigma}^* = \frac{(1 + k'_{\sigma;\text{exc}} *s) K_{\sigma;r} K_{\sigma;m}}{K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} *s) \left(K_{\sigma;m} \left(1 + \frac{*p}{K_{\sigma;\text{ic}}}\right) + *s\right) + (K_{\sigma;m} + k_{\sigma;\text{rel}} *s) *r} \zeta_{\sigma} \quad (5.30)$$

$$e_{\sigma \cup \{r\}}^* = \frac{K_{\sigma;m} *r}{K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} *s) \left(K_{\sigma;m} \left(1 + \frac{*p}{K_{\sigma;\text{ic}}}\right) + *s\right) + (K_{\sigma;m} + k_{\sigma;\text{rel}} *s) *r} \zeta_{\sigma} \quad (5.31)$$

$$e_{\sigma \cup \{s\}}^* = \frac{\left(1 - \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}}\right) (k_{\sigma;\text{rel}} *r + K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} *s)) *s}{K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} *s) \left(K_{\sigma;m} \left(1 + \frac{*p}{K_{\sigma;\text{ic}}}\right) + *s\right) + (K_{\sigma;m} + k_{\sigma;\text{rel}} *s) *r} \zeta_{\sigma} \quad (5.32)$$

$$e_{\sigma \cup \{p\}}^* = \frac{K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} *s) \left(\frac{K_{\sigma;m}}{K_{\sigma;\text{ic}}} *p + \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}} *s\right) + k_{\sigma;\text{rel}} \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}} *r *s}{K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} *s) \left(K_{\sigma;m} \left(1 + \frac{*p}{K_{\sigma;\text{ic}}}\right) + *s\right) + (K_{\sigma;m} + k_{\sigma;\text{rel}} *s) *r} \zeta_{\sigma} \quad (5.33)$$

where

$$\zeta_{\sigma} = e_{\sigma} + e_{\sigma \cup \{r\}} + e_{\sigma \cup \{s\}} + e_{\sigma \cup \{p\}}.$$

These solutions are relatively complex, however the numerators will simplify during the derivation of the rate of production of product. Note that the denominator of the equation for the Michaelis-Menten mechanism with product inhibition,

$$K_{\sigma;m} \left(1 + \frac{*p}{K_{\sigma;\text{ic}}}\right) + *s$$

is also found in the denominator of these solutions. Then with the assumptions

$$k_a = k_{r;a} = k_{s;a} = k_{p;a} \qquad k_{-a} = k_{r;-a} = k_{s;-a} = k_{p;-a}$$

and given the original system of ODEs, we have

$$\begin{aligned} \frac{d\bar{\zeta}}{dt} &= \frac{de}{dt} + \frac{de_r}{dt} + \frac{de_s}{dt} + \frac{de_p}{dt} \\ &= f'_{\text{adsorp}}(e, \bar{\zeta}) - k_a a^* \bar{\zeta} + k_{-a} \bar{\zeta}_a \\ \frac{d\bar{\zeta}_a}{dt} &= \frac{de_a}{dt} + \frac{de_{ra}}{dt} + \frac{de_{sa}}{dt} + \frac{de_{pa}}{dt} \\ &= k_a a^* \bar{\zeta} - k_{-a} \bar{\zeta}_a \end{aligned}$$

where

$$f'_{\text{adsorp}}(e, \bar{\zeta}) = k_{\text{on}} \Phi a e - k_{\text{off}} \left( \frac{K_s}{g} \right) \bar{\zeta}.$$

Ignoring the adsorption process, applying the rapid equilibrium assumption to the reactions between  $\bar{\zeta}$  and  $\bar{\zeta}_a$  gives

$$\bar{\zeta} = \frac{K_a}{K_a + a^*} (e_0 - e) \tag{5.34}$$

$$\bar{\zeta}_a = \frac{a^*}{K_a + a^*} (e_0 - e) \tag{5.35}$$

where  $K_a = \frac{k_{-a}}{k_a}$  and

$$e_0 = \bar{\zeta} + \bar{\zeta}_a + e = e + e_s + e_a + e_{sa} + e_r + e_{sr} + e_{ar} + e_{sar} + e.$$

We now have enough information to construct a model for the 2-site mechanism for PI4P5K. Again, the net rate of production of free PI(4,5)P<sub>2</sub> is

$$\frac{d^*p}{dt} = \sum_{\sigma \subset \{a\}} k_{\sigma;-p} e_{\sigma \cup \{p\}}^* - k_{\sigma;p} e_{\sigma}^* p$$

Substituting in values of  $e_{\sigma}^*$  and  $e_{\sigma \cup \{p\}}^*$  from equations 5.30 to 5.33 gives

$$\begin{aligned} \frac{d^*p}{dt} &= \sum_{\sigma \subset \{a\}} \frac{k_{\sigma;-p} \left( K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} s) \left( \frac{K_{\sigma;m}}{K_{\sigma;\text{ic}}} p + \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}} s \right) + k_{\sigma;\text{rel}} \frac{k'_{\sigma;\text{cat}}}{k_{\sigma;-p}} r s \right)}{K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} s) \left( K_{\sigma;m} \left( 1 + \frac{p}{K_{\sigma;\text{ic}}} \right) + s \right) + (K_{\sigma;m} + k_{\sigma;\text{rel}} s) r} \bar{\zeta}_{\sigma} \\ &\quad - \frac{k_{\sigma;p} (1 + k'_{\sigma;\text{exc}} s) K_{\sigma;r} K_{\sigma;m} p}{K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} s) \left( K_{\sigma;m} \left( 1 + \frac{p}{K_{\sigma;\text{ic}}} \right) + s \right) + (K_{\sigma;m} + k_{\sigma;\text{rel}} s) r} \bar{\zeta}_{\sigma} \end{aligned}$$

$$= \sum_{\sigma \in \{a\}} \frac{k'_{\sigma;\text{cat}} \left( (1 + k'_{\sigma;\text{exc}} *s) K_{\sigma;r} + k_{\sigma;\text{rel}} *r \right) *s}{K_{\sigma;r} (1 + k'_{\sigma;\text{exc}} *s) \left( K_{\sigma;m} \left( 1 + \frac{*p}{K_{\sigma;\text{ic}}} \right) + *s \right) + (K_{\sigma;m} + k_{\sigma;\text{rel}} *s) *r} \zeta_{\sigma}$$

Substituting in the values of  $\zeta_{\sigma}$  from equations 5.34 to 5.35, and rearranging, gives

$$\frac{d*p}{dt} = \left( \frac{k'_{\text{cat}} \left( (1 + k'_{\text{exc}} *s) K_r + k_{\text{rel}} *r \right) K_a}{K_r (1 + k'_{\text{exc}} *s) \left( K_m \left( 1 + \frac{*p}{K_{\text{ic}}} \right) + *s \right) + (K_m + k_{\text{rel}} *s) *r} + \right. \quad (5.36)$$

$$\left. \frac{k'_{a;\text{cat}} \left( (1 + k'_{a;\text{exc}} *s) K_{a;r} + k_{a;\text{rel}} *r \right) a*}{K_{a;r} (1 + k'_{a;\text{exc}} *s) \left( K_{a;m} \left( 1 + \frac{*p}{K_{a;\text{ic}}} \right) + *s \right) + (K_{a;m} + k_{a;\text{rel}} *s) *r} \right) \left( \frac{*s(e_0 - e)}{K_a + a*} \right)$$

Equation 5.36 describes the most general version of this model, with 4 variables and 15 parameters. In order to reduce the complexity of the model, I will again make the additional assumptions that the following rates are equal

$$\begin{array}{lll} K_s = K_{a;s} & K_{\text{ic}} = K_{a;\text{ic}} & K_m = K_{a;m} \\ k'_{\text{cat}} = k'_{a;\text{cat}} & k'_{\text{exc}} = k'_{a;\text{exc}} & k_{\text{rel}} = k_{a;\text{rel}} \end{array}$$

With these assumptions the model reduces to

$$\frac{d*p}{dt} = \frac{k'_{\text{cat}} \left( (1 + k'_{\text{exc}} *s) K_r + k_{\text{rel}} *r \right) *s(e_0 - e)}{K_r (1 + k'_{\text{exc}} *s) \left( K_m \left( 1 + \frac{*p}{K_{\text{ic}}} \right) + *s \right) + (K_m + k_{\text{rel}} *s) *r}. \quad (5.37)$$

This is more complex than the model for the 3-site mechanism for PI4P5K. Note that again we lose any dependence of this equation of the concentration of Arf ( $a*$ ). However now, due to the inclusion of PA within the substrate dynamics, the PA-dependence is retained.

It remains to write down an equation for the rate of change of  $e$ , the concentration of cytosolic enzyme. From the original system of ODEs we know that

$$\frac{de}{dt} = -f_{\text{adsorp}}(e, e^*) = -k_{\text{on}} \Phi a e + k_{\text{off}} e^*$$

Substituting in the value of  $e^*$  from equation 5.30 gives

$$\frac{de}{dt} = -k_{\text{on}} \Phi a e + \frac{k_{\text{off}} K_r K_m (1 + k'_{\text{exc}} *s)}{K_r (1 + k'_{\text{exc}} *s) \left( K_m \left( 1 + \frac{*p}{K_{\text{ic}}} \right) + *s \right) + (K_m + k_{\text{rel}} *s) *r} \zeta$$

Substituting in the value of  $\zeta$  from equation 5.34 gives

$$\frac{de}{dt} = -k_{\text{on}} \Phi a e + \frac{k_{\text{off}} K_a K_r K_m (1 + k'_{\text{exc}} *s) (e_0 - e)}{(K_a + a*) \left[ K_r (1 + k'_{\text{exc}} *s) \left( K_m \left( 1 + \frac{*p}{K_{\text{ic}}} \right) + *s \right) + (K_m + k_{\text{rel}} *s) *r \right]} \quad (5.38)$$

Together, equations 5.37 and 5.38 constitute what I will call the 2-site model of PI4P5K. In this model, Arf only contributes to the amount of enzyme on the membrane, and does not contribute to the actual kinetic rate of catalysis. In comparison with the 3-site model, PA is now involved in the catalytic mechanism and as such appears in the function describing the catalytic rate. I will analyse the behaviours of this model further in chapter 7.

## 5.6 Discussion

So as to develop more realistic models than the empirical models described in chapter 4 I have now constructed quasi-steady-state mechanistic models for the enzymes PLD and PI4P5K. These were based on newly proposed well-defined (complete and with all assumptions listed) molecular mechanisms describing the catalysis and regulation of the enzymes by Arf, and PI(4,5)P<sub>2</sub> or PA.

The molecular mechanisms were determined—wherever possible—based on the current understanding of the enzymes as specified by pre-existing published sources, including both from direct descriptions of molecular interactions and from arguments made via indirect evidence, for example evidence concerning binding sites. (The individual sections should be consulted for the full details and usage of these sources.) This was supplemented where necessary by sensible assumptions based on likely enzymatic activities.

Here, a ‘well-defined’ mechanism can be understood to be one that contains enough information such that it can be directly translated into a mathematical model using the law of mass-action without the immediate requirement for further mathematical assumptions. In this way, most of the primary model assumptions are clearly phrased in terms of biochemical interactions and as such could later (at least hypothetically) be investigated directly experimentally.

The application of the law of mass-action to the derived reactions schemes generates systems of ODEs. These could be directly simulated, but the requirement for many first and second order rate constants—most of which we could not easily measure—mean that this is not ideal.

Alternatively, these reaction schemes can be used to form ‘graphs on the enzyme complexes’, which describe the routes for the formation and destruction of enzyme-containing complexes. This allows the derivation of quasi-steady-state models via application of the linear framework of Gunawardena (2012) as described in section 3.3. However, in all cases, a naïve, direct application of the framework to the graphs on

the enzyme complexes generated from these mechanisms resulted in large (thousands to millions) numbers of spanning trees – which means that further, exact, derivation of mathematical models is a non-tractable problem.

An alternative approach was therefore taken, (again) using the idea of time-scale separation, now in order to separate dynamics related to substrate binding, unbinding and catalysis, from dynamics related to regulation. This required additional mathematical assumptions—the equality of certain rates—in order to generate sensible models. Further assumptions were used to simplify these to final, proposed models with fewer numbers of parameters, befitting an intent to use these as component models within a model of the Arf/PLD/PI4P5K motif. It remains to investigate whether the simplifying assumptions are valid.

This process was applied to three mechanisms:

- a 3-site mechanism for PLD, in section 5.3;
- a 3-site mechanism for PI4P5K, in section 5.5.1;
- and a 2-site mechanism for PI4P5K, in section 5.5.2.

The PI4P5K models also required the inclusion of cytosol-membrane interactions for which I have used the model of Kartal & Ebenhöf (2013) described in section 3.5.

Each model is more complicated than the equivalent empirical model discussed previously. Recall that these were already becoming too complex for complete mathematical analysis, and so the steady-state solutions for the mechanistic models of PLD and PI4P5K are very unlikely to be mathematical tractable. Most likely, investigation of these models will have to rely on numerical simulation. However, these new mechanistic models should be less dependent on arbitrary (mathematical) decisions and biases towards expected behaviours – and those assumptions that had to be taken are well-documented.

Analysis and exploration of the function and behaviours of the mechanistic models will be continued in chapter 7.

Through the development of the complete reactions schemes I have discovered that the mechanisms of regulation of PLD and PI4P5K by Arf, and PI(4,5)P<sub>2</sub> or PA, respectfully, are not particularly well understood at the scale of molecular interactions. As described in the primary literature—there remain large gaps in the understanding of both of the enzymes. This has required me to make many assumptions in the process of outlining complete, well-defined reaction schemes.

For the evidence that exists, its veracity is often uncertain. Particularly, the evidence for the 2-site model of PI4P5K—with overlapping PA- and PIP-binding sites—is based on a computational structure (Stace & Ktistakis 2006).

I have had to make unsupported assumptions in order to simplify the final models. Most importantly: that the PA (3 site model) and Arf (2/3 site models) activation of PI4P5K is solely via membrane recruitment and that there is no basal activation of PLD without the presence of both PI(4,5)P<sub>2</sub> and Arf.

Thus, I conclude that further biochemical research regarding the structure and interactions behind the catalytic and regulatory mechanisms of both PLD and PI4P5K is essential, to ensure a proper understanding of the mechanisms controlling the two enzymes.

Under the 3-site hypothesis—two independent binding sites, one for each regulator—the asymmetry in the behaviours of the two enzymes, conceals an apparently equivalent regulatory mechanism. Differences in the regulation then arise primarily via which complexes are catalytically-active and the different membrane-localisation behaviour of the two enzymes.

The complexity of the proposed reactions schemes, measured in terms of the number of enzyme complexes and reactions, is relatively large. This is the cause of the large numbers of spanning trees generated from the complete graphs on the enzyme complexes. The symmetry within these reaction schemes is then highly beneficial, as it permits a vast reduction in the complexity of the models by allowing the separation of the substrate and regulatory dynamics to be performed in a relatively simple manner.

The final models still contain including a sizeable number of parameters – particularly the 2-site model for PI4P5K. A reduction in complexity might be considered necessary for their inclusion into models of much larger systems, and it is plausible that these models could be simplified further with more stringent assumptions. However, choosing which additional assumptions to make will require a better understanding of the effects of each parameter on the behaviour of the system.

In total, I have derived suitable, sufficient mechanistic models of PLD and PI4P5K which I can use for my continuing investigation into the behaviour of the Arf/PLD/PI4P5K signalling motif. Combination of these models to explore this motif will occur in chapter 7. Additionally, these novel mechanistic models of the catalytic activity and regulation of PLD and PI4P5K, will be suitable for use—individually or together—in future models of signalling pathways and other intracellular processes.

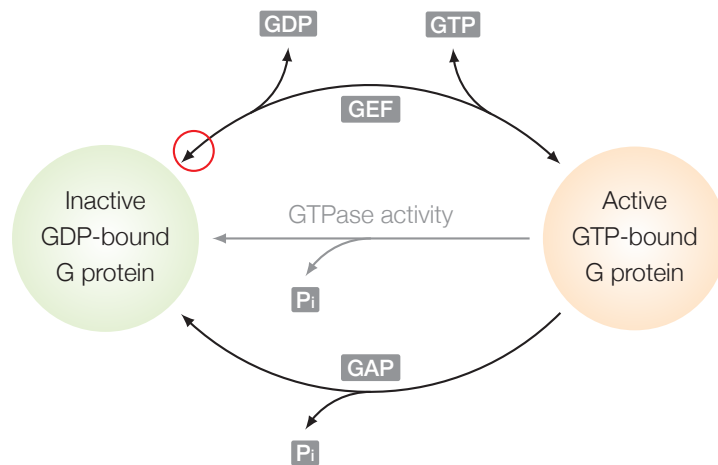
## 6 Arf family G proteins: mechanisms and models

So far I have not considered control of the activation of Arf by its regulators in any of the earlier models of the Arf/PLD/PI4P5K signalling motif. Instead, the system could be stimulated by manually changing the concentration of active Arf. Yet—as I will demonstrate in this chapter—the regulation of Arf, controlled by GEFs and GAPs, is complex. Therefore, in order to make physiologically-sound statements about the complete Arf/PLD/PI4P5K motif I consider it important to properly investigate and understand the dynamics of this regulation.

In this chapter I intend to provide the final component of a complete model of the Arf/PLD/PI4P5K motif by introducing a mechanistic model that describes the regulation of Arf by its GEFs and GAPs. With this model, the point at which the system can be manually stimulated will move upstream to the intracellular concentrations of these regulators of Arf. This will have three major advantages:

- In developing the model, I will have the opportunity to investigate the regulation of G proteins by GEFs and GAPs in a general sense.
- Incorporating this model into the complete model of the motif will give more realistic activation and inactivation profiles for Arf – which may have an important effect on the downstream signalling.
- It sets up the opportunity to investigate the effect of proposed feedback – the regulation of GEFs and GAPs by phospholipids, as discussed in section 2.7.1. While I will not attempt this in this thesis, I will discuss positive and negative feedback further in section 8.3.1.

Recall that G proteins—of which Arf is a member—are largely controlled by two mechanisms, illustrated in figure 6.1: *GTPase activity*; and *guanine nucleotide exchange*. I will briefly recap and expand on these mechanisms here. These will also



**Figure 6.1** G protein regulation. G proteins exist in one of two states depending on the bound guanine nucleotide – commonly referred to as active (GTP-bound) or inactive (GDP-bound). For monomeric G proteins, the switch between states is regulated by GTPase activity mediated by GAPs; and (reversible) guanine nucleotide exchange catalysed by GEFs. Additionally, intrinsic GTPase activity – while not present in monomeric G proteins – is present in heterotrimeric G proteins. Intrinsic activity will also be discussed in the text, as the result is catalytically equivalent, but has a mechanism that is simpler to model. The red circle highlights the fact that the GEF mechanism is completely reversible. This is often overlooked in discussions of the ‘activation/inactivation’ cycle of G proteins.



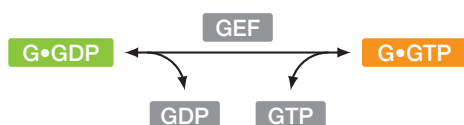
be discussed further in the following sections. Also recall that we tend to refer to GTP-bound G protein as active, and GDP-bound G protein as inactive.

GTPase activity inactivates active G proteins through the hydrolysis of the currently bound molecule of GTP to form GDP. The nucleotide remains bound in the nucleotide binding pocket of the G protein at all times throughout this process. For Arf—as a monomeric G protein—GTPase activity is catalysed by *GTPase-activating proteins (GAPs)*.



Heterotrimeric G proteins are capable of intrinsic GTPase activity, and so do not require an additional enzyme to mediate this process.

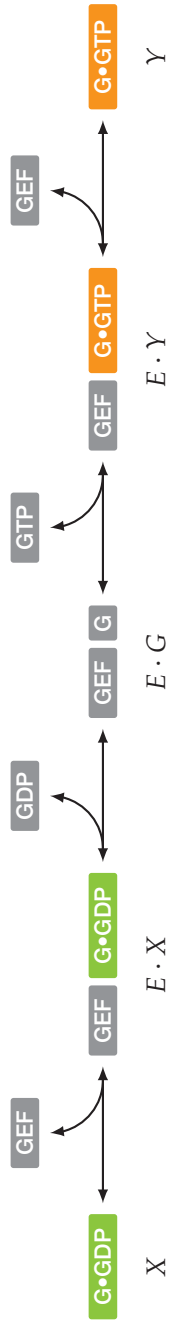
Guanine nucleotide exchange is catalysed by enzymes known as *guanine nucleotide exchange factors (GEFs)* for monomeric G proteins. These are capable of exchanging the bound GDP for GTP, and so activating the G protein.



In more detail: a GEF binds to an inactive G protein causing a conformational change which opens the G protein's guanine nucleotide binding pocket. This allows the sequential release of GDP, and the binding of GTP, into this binding pocket. Finally the GEF can disassociate from the now active G protein (Vetter & Wittinghofer 2001; Bos, Rehmann, & Wittinghofer 2007). For heterotrimeric G proteins, the role of GEFs is taken by the G protein coupled receptors (GPCRs).

While GEFs are certainly capable of mediating the activation of G proteins via this forward catalytic process, they are also known to have a completely reversible reaction mechanism (Bos, Rehmann, & Wittinghofer 2007; Goody 2014). So GEFs are also able to mediate the inactivation of G proteins by mediating this process in reverse – the release of GTP and binding of GDP.

I have noticed that the reversibility of GEFs is often overlooked when discussing the regulation of many different G proteins. In illustrations of systems including of G proteins, the arrow corresponding to GEF-mediated regulation is often drawn as unidirectional – missing the reverse arrowhead highlighted in figure 6.1. Examples of representations where this reversibility has apparently not been considered can be found as:



**Figure 6.2** The molecular mechanism for the guanine nucleotide exchange activity of GEFs. Reading left-to-right: GEF binds to an inactive G protein causing a conformational change, which allows the release of the bound GDP; this forms a nucleotide-free form of the G protein with an empty guanine nucleotide binding pocket, which remains bound to the GEF; GTP can bind into the nucleotide binding pocket, to produce active G protein; the GEF can then disassociate. Each stage of this process is completely reversible, and therefore so is the complete mechanism. Each reaction has its own rate that will depend on the G protein and GEF, and the concentrations of the participating species.

- figure 1 in Cherfils & Zeghouf (2013);
- figure 1a in Ahearn et al. (2012);
- and figure 1c in Donaldson & Jackson (2011).

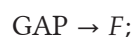
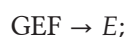
The distinction between reversible and irreversible mechanisms is an important one. In this chapter I will show that this difference leads to crucial quantitative and qualitative differences upon the stimulation of the system. Furthermore, I will conclude that the assumption of an irreversible mechanism may have led to a historic misapprehension about the role of both GEFs and GAPs in the regulation of G proteins.

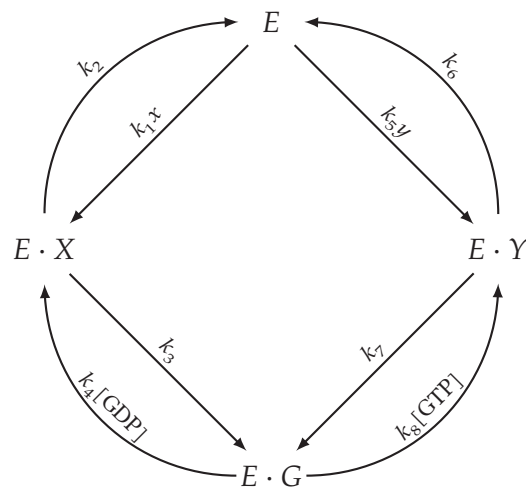
Note that an excess of GTP over GDP—as is biologically observed—means that the nucleotide-free complex is more likely to encounter GTP. However, the extent to which this drives the activation of the system is dependent on the individual kinetic rates.

Recent reports have stressed that a better understanding of the enzyme kinetics of GEFs is needed. These reports have particularly investigated GEFs in terms of initial velocity studies – therefore concerning themselves with transitory dynamics (Northup, Jian, & Randazzo 2012; Randazzo et al. 2013). I am more interested in the steady-states reached by the G protein:GEF system, as it is under this regime that stable, consistent signalling will take place. Furthermore, these steady-states can be analysed using the standard analytical techniques that I described in chapter 3. I will again stress the importance of basing these mathematical models on well-defined, and physiologically-sound, explanations of the molecular interactions. Otherwise I am not investigating the correct system, which will lead to erroneous conclusions.

In this chapter I will derive quasi-steady-state mathematical models for the regulation of a generic G protein by the action of GAPs and GEFs using the framework of Gunawardena (2012). These will allow me to make statements that should be true for G proteins in general. Following from this, I will use these models and results to derive further models and results specific to the regulation of Arf.

Throughout this chapter I will use the following notation:

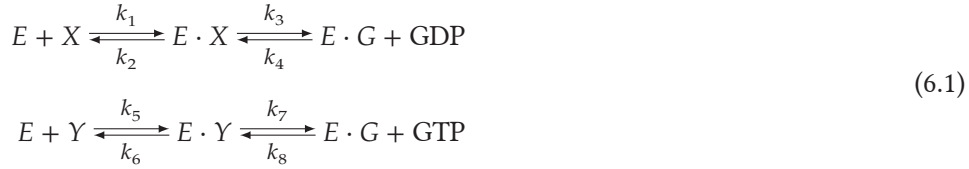




**Figure 6.3** Graph on the enzyme complexes for the GEF mechanism – a GEF ( $E$ ) acting on a G protein (GDP-bound,  $X$ ; GTP-bound,  $Y$ ; nucleotide-free,  $G$ ). Edges are first-order kinetic rates, and include partner species where applicable.

## 6.1 GEFs: mechanism and QSS model

A complete, and well-accepted (Bos, Rehmann, & Wittinghofer 2007), molecular mechanism for a generic GEF is shown in figure 6.2. Alternatively, this mechanism can be written as the reaction scheme



labelled with the first- and second-order kinetic rates ( $k_i$ ).

The concentrations of the molecular species and complexes will be denoted with the shorthand:  $x = [X]$ ;  $y = [Y]$ ;  $g = [G]$ ;  $e = [E]$ ;  $e_x = [E \cdot X]$ ;  $e_y = [E \cdot Y]$ ; and  $e_g = [E \cdot G]$ . Note that throughout the following analysis the concentrations of GDP and GTP have been assumed to be well buffered, so that they do not measurably vary.

Using the law of mass-action, the above reaction scheme implies the following ODEs.

$$\frac{dx}{dt} = -k_1xe + k_2e_x \tag{6.2}$$

$$\frac{dy}{dt} = -k_5ye + k_6e_y \tag{6.3}$$

$$\frac{de}{dt} = -(k_1x + k_5y)e + k_2e_x + k_6e_y \tag{6.4}$$

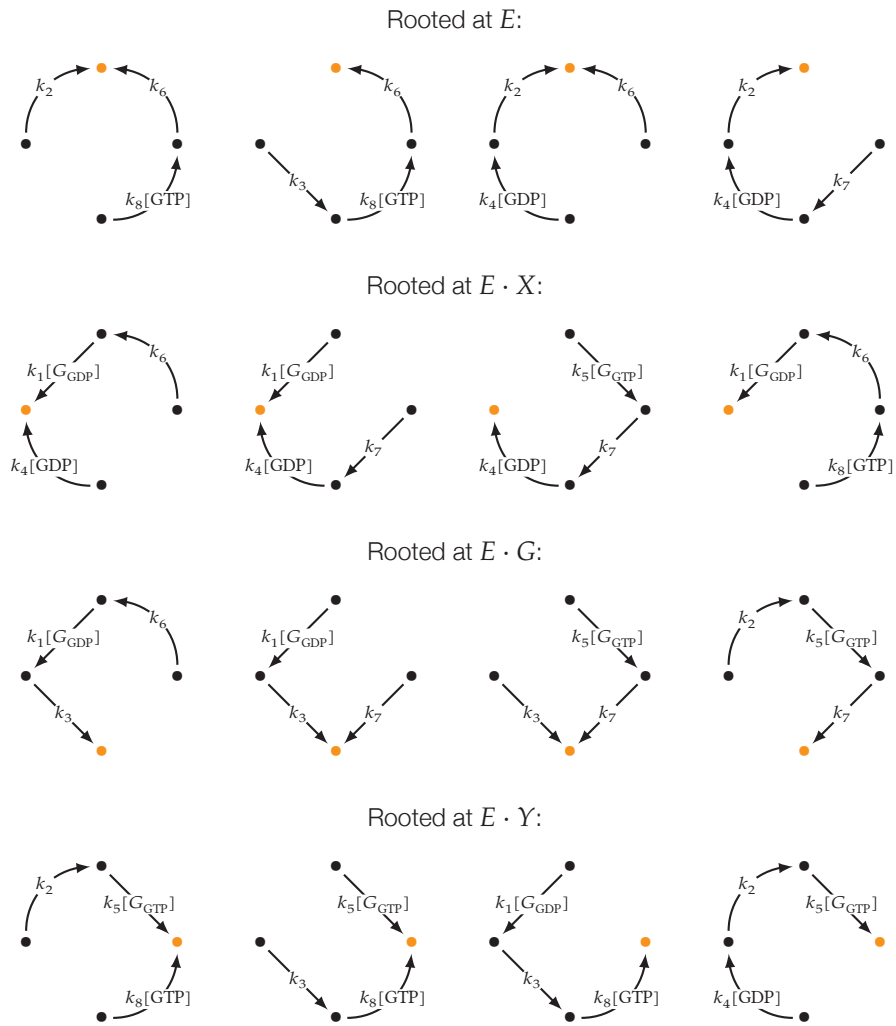
$$\frac{de_x}{dt} = -(k_2 + k_3)e_x + k_1xe + k_4[\text{GDP}]e_g \tag{6.5}$$

$$\frac{de_y}{dt} = -(k_6 + k_7)e_y + k_5ye + k_8[\text{GTP}]e_g \tag{6.6}$$

$$\frac{de_g}{dt} = -(k_4[\text{GDP}] + k_8[\text{GTP}])e_g + k_3e_x + k_7e_y \tag{6.7}$$

I will further analyse a modified form of these ODEs in section 6.3.

I will now apply the framework of Gunawardena (2012), as described in section 3.3, in order to produce a reduced QSS model of this mechanism – under the assumption that the dynamics of nucleotide exchange are—in some sense—fast. The reaction scheme can be redrawn as the graph on the enzyme complexes, shown as figure 6.3, which gives 16 directed spanning trees, shown in figure 6.4. Note that though this graph and the spanning trees appear the same as those discussed for the di-allosteric mechanism in chapter 5 they have distinct labels and so will result in a different basis element.



**Figure 6.4** The 16 spanning trees of the graph on the enzyme complexes for the GEF mechanism, shown as figure 6.3. The orange nodes denote the root of each spanning tree.

The basis element, in the order ( $e$ ,  $e_x$ ,  $e_g$ , and  $e_y$ ), is given by

$$\rho = \begin{bmatrix} K_0 \\ K_1^{ex}x + K_2^{ex}y \\ K_1^{eg}x + K_2^{eg}y \\ K_1^{ey}x + K_2^{ey}y \end{bmatrix}$$

with summary constants defined as

$$\begin{aligned} K_1^{ex} &= k_1(k_6k_8[\text{GTP}] + k_4(k_6 + k_7)[\text{GDP}]) \\ K_2^{ex} &= k_4k_5k_7[\text{GDP}] \\ K_1^{eg} &= k_1k_3(k_6 + k_7) \\ K_2^{eg} &= k_5k_7(k_2 + k_3) \\ K_1^{ey} &= k_1k_3k_8[\text{GTP}] \\ K_2^{ey} &= k_5(k_8(k_2 + k_3)[\text{GTP}] + k_2k_4[\text{GDP}]) \\ K_0 &= k_6k_8(k_2 + k_3)[\text{GTP}] + k_2k_4(k_6 + k_7)[\text{GDP}] \\ K_1 &= K_1^{ex} + K_1^{eg} + K_1^{ey} \\ K_2 &= K_2^{ex} + K_2^{eg} + K_2^{ey}. \end{aligned}$$

This then gives the steady-state concentrations of each of the GEF complexes

$$e = \left( \frac{K_0}{K_0 + K_1x + K_2y} \right) e_0 \quad (6.8)$$

$$e_x = \left( \frac{K_1^{ex}x + K_2^{ex}y}{K_0 + K_1x + K_2y} \right) e_0 \quad (6.9)$$

$$e_g = \left( \frac{K_1^{eg}x + K_2^{eg}y}{K_0 + K_1x + K_2y} \right) e_0 \quad (6.10)$$

$$e_y = \left( \frac{K_1^{ey}x + K_2^{ey}y}{K_0 + K_1x + K_2y} \right) e_0 \quad (6.11)$$

where the total mass of enzyme is given by

$$e_0 = e + e_x + e_g + e_y.$$

The rate of change of the concentration of active G protein was given in equation 6.3, this is

$$\frac{dy}{dt} = -k_5ye + k_6e_y.$$

Substituting the steady-state concentrations for  $e$ ,  $e_x$ , and  $e_y$  from equations 6.8 to 6.11 into this equation gives

$$\begin{aligned}\frac{dy}{dt} &= -k_5y \left( \frac{K_0}{K_0 + K_1x + K_2y} \right) e_0 + k_6 \left( \frac{K_1^{e_y}x + K_2^{e_y}y}{K_0 + K_1x + K_2y} \right) e_0 \\ &= \left( \frac{k_6K_1^{e_y}x + (k_6K_2^{e_y} - k_5K_0)y}{K_0 + K_1x + K_2y} \right) e_0\end{aligned}$$

and as

$$\begin{aligned}k_6K_2^{e_y} - k_5K_0 &= k_5k_6k_8(k_2 + k_3)[\text{GTP}] + k_2k_4k_5k_6[\text{GDP}] \\ &\quad - k_5(k_6k_8(k_2 + k_3)[\text{GTP}] - k_2k_4k_5(k_6 + k_7)[\text{GDP}]) \\ &= -k_2k_4k_5k_7[\text{GDP}]\end{aligned}$$

then

$$\frac{dy}{dt} = \frac{k_{\text{cat}}(x - \kappa y)e_0}{K_0 + K_1x + K_2y} \quad (6.12)$$

where

$$\begin{aligned}k_{\text{cat}} &= k_6K_1^{e_y} = k_1k_3k_6k_8[\text{GTP}] \\ \kappa &= \frac{k_2k_4k_5k_7[\text{GDP}]}{k_1k_3k_6k_8[\text{GTP}]}\end{aligned}$$

Similarly,

$$\frac{dx}{dt} = \frac{k_{\text{cat}}(\kappa y - x)e_0}{K_0 + K_1x + K_2y}. \quad (6.13)$$

Note that equation 6.12 effectively reduces to the Michaelis-Menten equation when  $y = 0$ , and is equivalent to the equation previously used for initial velocity studies of GEFs (Randazzo et al. 2013) if the concentration of GTP is extracted from the summary constants.

From equation 6.12 we can see that

$$\frac{dy}{dt} \propto x - \kappa y = x - \frac{k_2k_4k_5k_7[\text{GDP}]}{k_1k_3k_6k_8[\text{GTP}]} \times y.$$

The constant  $\kappa$  is related to constraints on thermodynamic equilibrium around the cyclical graph shown in figure 6.3 – I will discuss this further in section 6.4.1.



## 6.2 GAPs: mechanism and QSS model

I will assume that GAPs follow a Michaelis-Menten type mechanism with product inhibition: Michaelis-Menten as GAPs act to hydrolyse GTP to GDP, which is not an atypical enzymatic activity; and product inhibition, as GDP and GTP are structurally similar.

Recall from section 3.3.2 that the Michaelis-Menten mechanism with product inhibition is described by the reaction scheme



and by the equation

$$\frac{dx}{dt} = \frac{k_{\text{ase}} y f_0}{K_m (1 + \frac{x}{K_{\text{ic}}}) + y} \quad (6.14)$$

where  $f_0$  is the total concentration of GAP;  $k_{\text{ase}}$  is the catalytic rate ( $k_{\text{cat}}$  renamed to avoid conflict with the GEF model);  $K_m$  is the Michaelis constant; and  $K_{\text{ic}}$  is the constant of inhibition.

## 6.3 Simulations of G protein regulation

Before continuing with the investigation of the QSS models of GEFs and GAPs, I want to backtrack slightly to the ODE model for the GEF mechanism, described by equations 6.2 to 6.7. Unlike the QSS model these equations account for the concentrations of the intermediate G protein-GEF complexes, and so will allow me to investigate changes in the concentrations of these in addition to the concentrations of the free active and inactive forms. I will modify the equations to incorporate different forms of GTPase activity. When simulated, this systems of ODEs will begin to help me understand the roles of both GEFs and GAPs in the regulation of G proteins.

I want to control the system by varying the total concentration of GEF, described by the parameter  $e_0$ . Recall that the conservation of mass equation for GEF is given by the equation

$$e_0 = e + e_x + e_g + e_y.$$

Therefore, by fixing  $e_0$  we lose one degree of freedom among  $e$ ,  $e_x$ ,  $e_g$ , and  $e_y$ . So I will choose to set

$$e = e_0 - e_x - e_y - e_g \quad (6.15)$$

Rate	Ran:RCC1 (Klebe et al. 1995)	Irreversible	Unit
$k_1$	$7.4 \times 10^7$	$7.4 \times 10^7$	$M^{-1} s^{-1}$
$k_2$	55	55	$s^{-1}$
$k_3$	21	21	$s^{-1}$
$k_4$	$1.1 \times 10^7$	$1.1 \times 10^7$	$M^{-1} s^{-1}$
$k_5$	$1.0 \times 10^8$	$1.0 \times 10^8$	$M^{-1} s^{-1}$
$k_6$	55	55	$s^{-1}$
$k_7$	19	0	$s^{-1}$
$k_8$	$0.6 \times 10^6$	$0.6 \times 10^6$	$M^{-1} s^{-1}$
$K_1^{ex}$	$8.466 \times 10^{16}$	$6.919 \times 10^{16}$	
$K_2^{ex}$	$2.090 \times 10^{16}$	0	
$K_1^{eg}$	$1.150 \times 10^{11}$	$8.547 \times 10^{10}$	
$K_2^{eg}$	$1.444 \times 10^{11}$	0	
$K_1^{ey}$	$9.324 \times 10^{15}$	$9.324 \times 10^{15}$	
$K_2^{ey}$	$1.061 \times 10^{17}$	$1.061 \times 10^{17}$	
$K_0$	$6.985 \times 10^{10}$	$5.836 \times 10^{10}$	
$K_1$	$9.398 \times 10^{16}$	$7.851 \times 10^{16}$	
$K_2$	$1.270 \times 10^{17}$	$1.061 \times 10^{17}$	
$k_{cat}$	$5.128 \times 10^{17}$	$5.128 \times 10^{17}$	
$\kappa$	2.242	0	
$K_s$	$2.069 \times 10^7$	$5.500 \times 10^{-7}$	

**Table 6.1** Kinetic rate parameters ( $k_1, \dots, k_8$ ) and summary rate constants for the mass-action and QSS GEF models, used for illustrative purposes. Two sets of parameters have been used: those measured for the Ran:RCC1 G protein:GEF system (Klebe et al. 1995); and those same rates except  $k_7 = 0$ , which has the effect of making the complete mechanism irreversible.

The remainder of the system is given by the equations

$$\frac{dx}{dt} = -k_1xe + k_2e_x - f_{\text{ase}}(x, y) \quad (6.16)$$

$$\frac{dy}{dt} = -k_5ye + k_6e_y + f_{\text{ase}}(x, y) \quad (6.17)$$

$$\frac{de_x}{dt} = -(k_2 + k_3)e_x + k_1xe + k_4[\text{GDP}]e_g \quad (6.18)$$

$$\frac{de_y}{dt} = -(k_6 + k_7)e_y + k_5ye + k_8[\text{GTP}]e_g \quad (6.19)$$

$$\frac{de_g}{dt} = -(k_4[\text{GDP}] + k_8[\text{GTP}])e_g + k_3e_x + k_7e_y \quad (6.20)$$

where  $f_{\text{ase}}(x, y)$  describes the GTPase activity, either:

- none,  $f_{\text{ase}}(x, y) = 0$ ;
- intrinsic, modelled by first-order (exponential) decay  $f_{\text{ase}}(x, y) = k_{\text{ase}}y$ ;
- or, GAP-mediated, modelled by equation 6.14.

These ODEs can be numerically integrated given a complete set of parameters. In this section I have chosen—for illustrative purposes—the measured parameters for the Ran:RCC1 G protein:GEF system (Klebe et al. 1995), reproduced in table 6.1. Additionally: for the GTPase activity the arbitrary parameters  $k_{\text{ase}} = 4$ ,  $K_m = 0.7$ ,  $K_{\text{ic}} = 100$ , and  $f_0 = 1$  have been used; and—for a sensible physiologically plausible ratio— $[\text{GTP}] = 10$  and  $[\text{GDP}] = 1$ , in arbitrary units.

The following simulations will be stimulated by varying the total concentration of enzyme  $e_0$ :

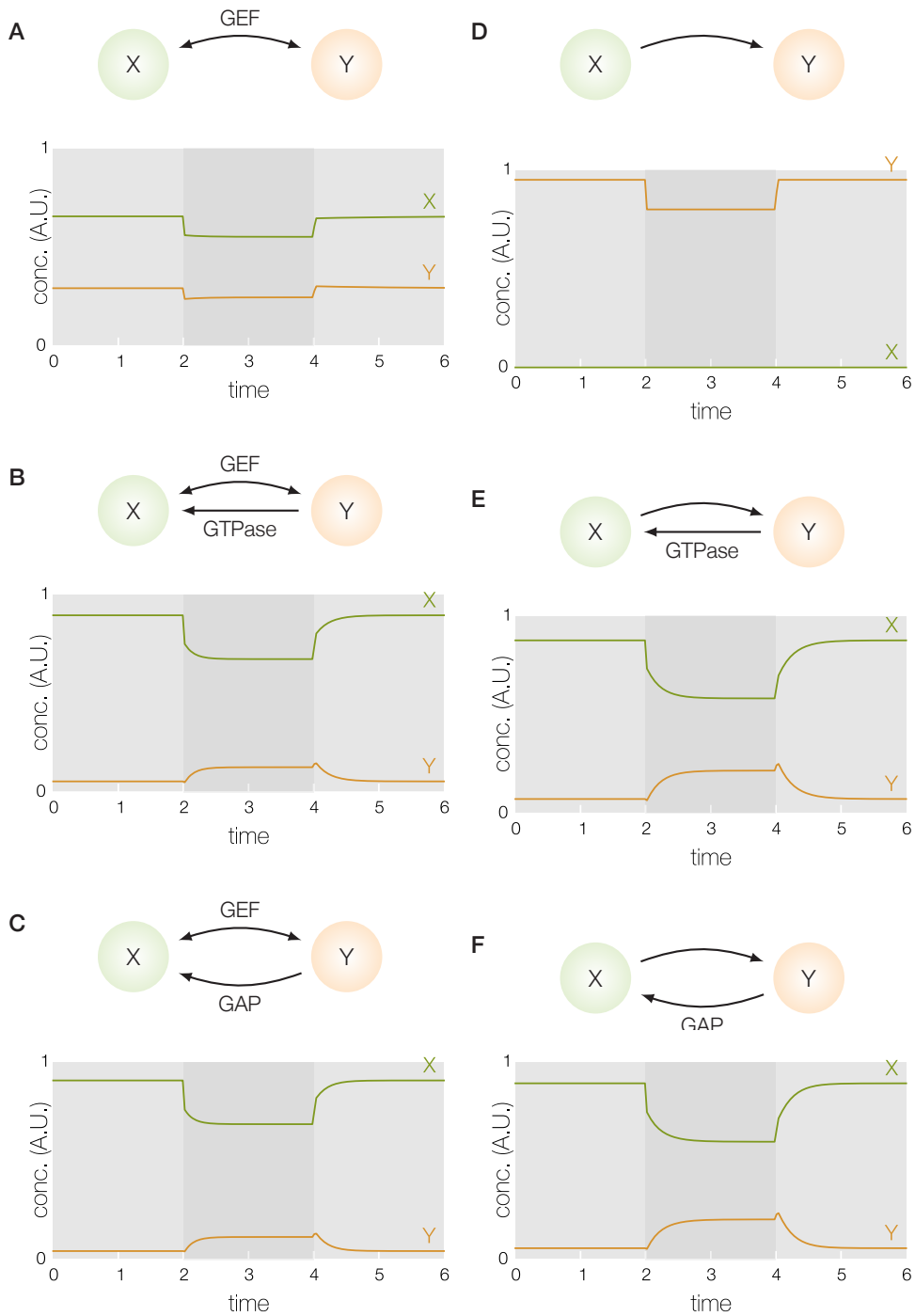
- during  $0 \leq t < 2$ ,  $e_0 = 0.05$ ;
- during  $2 \leq t < 4$ ,  $e_0 = 0.2$  (a 4-fold increase);
- and during  $4 \leq t$ , available free GEF ( $E$ ) was removed from the simulation until  $e_0 = 0.05$ .

All simulations were started from steady-state initial conditions, specific to each model. The Python function for the system of ODEs is as follows:

```

1 def gef_odes(x, t, p, gtpase=None):
2     g_gdp, g_gtp, eg_gdp, eg_gtp, eg = x
3
4     # stimulation by total GEF concentration (e0)
5     e0 = 0.2 if 2 <= t <= 4 else 0.05

```



**Figure 6.5** Simulation of the system described by equations 6.15 to 6.20, with different GTPase activity, and using either the Ran:RCC1 rates or the irreversible rates, listed in table 6.1. The darker shading denotes the time period wherein the model was stimulated by a 4-fold increase in the total concentration of GEF.

```

6
7 # set e from e0
8 e = e0 - eg_gdp - eg_gtp - eg
9 e = e if e > 0 else 0
10
11 # set appropriate function for GTPase activity
12 if gtpase == 'GTPase':
13     fase = p[kase]*g_gtp
14 elif gtpase == 'GAP':
15     fase = p[kase]*g_gtp*p[f0]/(p[km] + g_gtp)
16 else:
17     fase = 0
18
19 # ODEs
20 dg_gdp = - p[k1]*e*g_gdp + p[k2]*eg_gdp + fase
21 dg_gtp = - p[k5]*e*g_gtp + p[k6]*eg_gtp - fase
22 deg_gdp = - (p[k2] + p[k3])*eg_gdp + p[k1]*e*g_gdp \
23           + p[k4]*eg*p[gdp]
24 deg_gtp = - (p[k6] + p[k7])*eg_gtp + p[k5]*e*g_gtp \
25           + p[k8]*eg*p[gtp]
26 deg      = - eg*(p[k4]*p[gdp] + p[k8]*p[gtp]) + p[k3]*eg_gdp \
27           + p[k7]*eg_gtp
28
29 return [dg_gdp, dg_gtp, deg_gdp, deg_gtp, deg]

```

Which can be simulated using a Python script, for example:

```

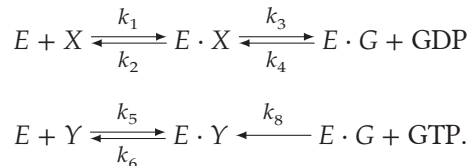
1 from sympy import *
2 import numpy as np
3 from scipy.integrate import odeint
4
5 # symbolic setup
6 k1, k2, k3, k4, k5, k6, k7, k8 = \
7     symbols('k_1 k_2 k_3 k_4 k_5 k_6 k_7 k_8', positive=True)
8 kase, km, f0 = symbols('k_ase K_m f_0', positive=True)
9 gdp, gtp = symbols('[GDP] [GTP]', positive=True)
10
11 # parameter dictionary
12 p = {k1:7.4*10**7, k2:55, k3:21, k4:1.1*10**7,
13      k5:1.0*10**8, k6:55, k7:19, k8:0.6*10**6,
14      kase:4, km:0.7, f0:1,
15      gtp:10, gdp:1}
16
17 # simulate
18 ics = [1, 0, 0, 0, 0]
19 t = np.arange(0, 6, 0.1)
20 sol = odeint(gef_odes, ics, t, (p, 'GAP'))

```

Simulations for the G protein:GEF system are shown in figure 6.5: with no GTPase activity in figure 6.5A; with intrinsic GTPase activity in figure 6.5B; and with GAP-

mediated GTPase activity in figure 6.5C.

Additionally, a irreversible model was generated from the GEF model by setting the rate of release of GTP from the active G protein:GEF complex ( $k_7$ ) equal to zero, as shown in table 6.1. This is equivalent to the reaction scheme



This would describe the situation where the enzyme can recognise both of the active and inactive forms (GDP- and GTP-bound) of the G protein, but where it is not able to catalyse the release of the bound GTP. We should expect to observe product inhibition for this model. An alternative irreversible model could be generated by setting any one or more of the reverse reaction rates ( $k_2, k_4, k_5, k_7$ ) to zero. Note that setting any of these constants to zero defines a mechanism that is thermodynamically distinct from the reversible mechanism (to keep thermodynamic equivalence we could instead assume that any of these constants is arbitrarily close to zero).

Simulations for the irreversible system are also shown in figure 6.5: with no GTPase activity in figure 6.5D; with intrinsic GTPase activity in figure 6.5E; and with GAP-mediated GTPase activity in figure 6.5F.

From these simulations—which will be discussed further in the following sections—I make the following preliminary observations:

**Figure 6.5A** The reversible model in the absence of GTPase activity shows a reduction in the concentrations of both active and inactive G protein upon stimulation by an increase in GEF. This can be explained by an equivalent increase in the concentrations of the intermediate enzyme complexes.

**Figure 6.5D** An irreversible mechanism will always convert its entire substrate into product. Product inhibition can be observed, and is a result of the location at which the mechanism was made irreversible. If, for example, the binding of active G protein to GEF ( $k_5$ ) was instead made to be zero, then there would be no response to the increase in GEF. This is the only simulation where the concentration of active G protein is greater than the concentration of inactive G protein.

**Figures 6.5B, 6.5C, 6.5E and 6.5F** Both the reversible and irreversible models display similar profiles upon increase and decrease in GEF, with only a larger

magnitude change for the irreversible model. This suggests that distinguishing these two mechanisms experimentally may be difficult. Small transitory effects can also be observed upon addition and removal of enzyme – these are small with respect to the long-term behaviours of the system, and discussion of these effects is outside the scope of the current analysis, and would require additional mathematical tools.

## 6.4 Notes on the regulation of G proteins by GEFs and GAPs

Returning to the QSS models introduced in sections 6.1 and 6.2 there are multiple mathematical results that can now be derived.

### 6.4.1 GEFs act to attain an equilibrium between active and inactive G protein

Recall that the QSS model for GEFs in equation 6.12 was derived under the quasi-steady-state assumption. If we now also assume that the active G protein concentration is at steady-state then

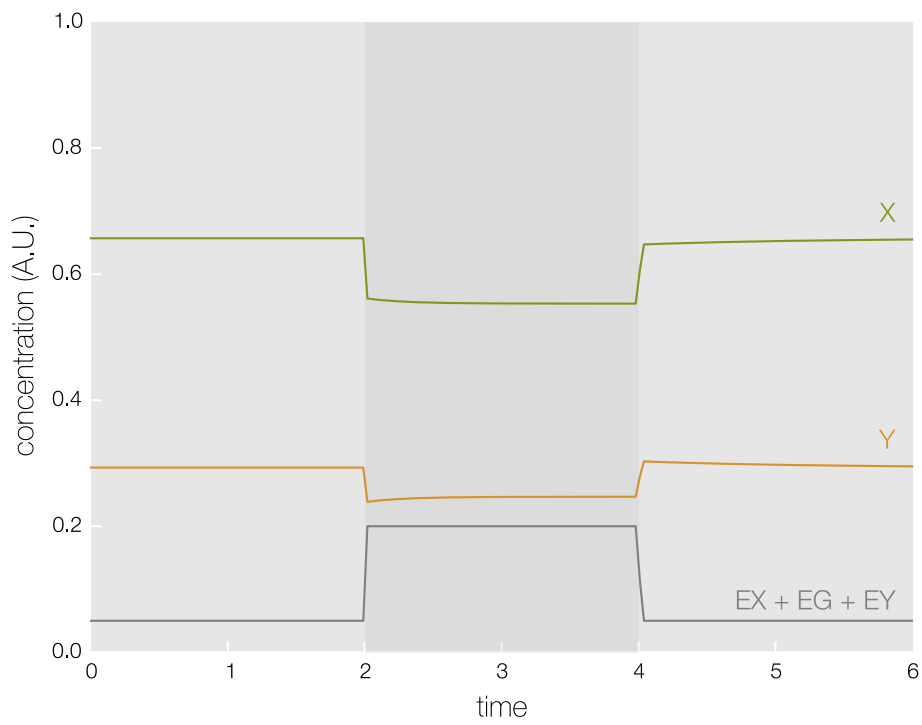
$$\begin{aligned} \frac{dy}{dt} = 0 &= \frac{k_{\text{cat}}(x - \kappa y)e_0}{K_0 + K_1x + K_2y} \\ \implies \frac{x}{y} &= \kappa. \end{aligned} \tag{6.21}$$

Alternatively this same result can be derived from the equation for  $\frac{dx}{dt}$ , equation 6.13.

Equivalently, equation 6.21 can also be derived by assuming that the cyclical graph shown in figure 6.3 is in thermodynamic equilibrium, and so the principle of microscopic reversibility will hold (Colquhoun et al. 2004). Directly, equating clockwise and anticlockwise rates, we obtain

$$k_1x k_3 k_8[GTP] k_6 = k_5y k_7 k_4[GDP] k_2 \implies \frac{x}{y} = \frac{k_2k_4k_5k_7[GDP]}{k_1k_3k_6k_8[GTP]} = \kappa.$$

Either way, this result means that the steady-state concentration of active and inactive G proteins are always in ratio, given by the summary constant  $\kappa$ . Notably, the value of  $\kappa$  is independent of the total concentration of the enzyme – it depends only on the rate parameters of the specific GEF, and the ratio of GDP to GTP.



**Figure 6.6** Enlargement of figure 6.5A with the total concentration of GEF complexes also drawn. The darker shading denotes the time period wherein the model was stimulated by a 4-fold increase in the total concentration of GEF.



For the Ran:RCC1 parameters listed in table 6.1,  $\kappa = 2.242$ . Figure 6.5A, which shows a simulation of the G protein:GEF system without GTPase activity, is enlarged in figure 6.6. In these figures, this ratio value can be observed for the steady-states for both the unstimulated and stimulated time periods in the simulation.

For the G protein:GEF system a second conservation holds, which describes the total concentration of G protein

$$g_0 = x + y + e_x + e_g + e_y.$$

In the extreme, when  $e_0 \ll g_0$  then we can say that  $e_x \approx 0$ ,  $e_g \approx 0$ , and  $e_y \approx 0$ . (Note that this assumption also applies to the derivation of the QSS model.) This means we can approximate the total concentration of G protein by

$$g_0 = x + y.$$

This means we can calculate the proportion of G protein which is active (still under the assumption that  $e_0 \ll g_0$ ),

$$\frac{y}{g_0} = \frac{y}{x + y} = \frac{y}{\kappa y + y} = \frac{y}{y(\kappa + 1)} = \frac{1}{\kappa + 1}. \quad (6.22)$$

As I will show, this is the maximum possible steady-state proportion of active G protein for systems with or without GTPase activity. For the Ran:RCC1 parameters  $\frac{1}{\kappa + 1} = 0.308$ .

#### 6.4.2 GEFs can be inhibitory

As I have previously mentioned, GEFs are often referred to as the ‘activators’ of G proteins, but their reversible reaction scheme directly implies a more complex role in the regulation of G proteins.

For instance figure 6.6 (the enlargement of figure 6.5A) shows a G protein:GEF system in the absence of GTPase activity. When stimulated by an increase in the concentration of GEF, this simulation shows a decrease in the concentration of both the inactive, and the active, forms of the G protein. This is evidence that—in at least some conditions—GEFs can actually be inhibitory towards G proteins. This is certainly contrary to their canonical role as ‘activators’ of G proteins.

The cause of this inhibitory effect is due to the formation of excess G protein:GEF complexes – intermediate complexes in the reaction scheme, that I will assume have no specific physiological role or functional interactions. The formation of these is now shown as a bulk term  $E \cdot X + E \cdot G + E \cdot Y$  in figure 6.6. Similar intermediate

complexes are present for all enzymes, but are generally ignored in reduced (QSS) models (such as the Michaelis-Menten equation). This is only acceptable when we know that the total concentration of enzyme is much less than the total concentration of substrate.

The effect we see in this simulation is in direct analogy with product inhibition. For enzymes with product inhibition, excess product leads to a reduction in the effective rate of production of further product. This is due to the enzyme being held in a state that is incapable of catalysis, where if product is bound to the catalytic site then the enzyme is not free to bind further substrate. In these simulations—which do not satisfy  $e_0 \ll g_0$ —we observe the inverse inhibition, where excess enzyme leads to a reduction in the concentration of available product. (Rather than excess product leading to a reduction in the concentration of available enzyme.)

So far this discussion has been based solely on a single set of parameters measured for a specific G protein:GEF system. It would be much more useful to be able to make statements that we are sure will hold in general for G protein:GEF systems. These statements will then certainly hold for Arf:Arf-GEF systems.

In the following mathematical analysis I will show that the observed inhibitory effect of GEFs on G proteins will hold for all G protein:GEF systems that follow the kinetics shown in figure 6.2. More precisely, I will show that there is no set of parameters which give a model that has an increase in the steady-state concentration of active G protein as the total concentration of GEF is increased. The proof of this statement follows.

As we are interested in the result at steady-state, we can use the QSS-derived equations for the concentration of the GEF complexes, originally listed in equations 6.8 to 6.11. These are

$$\begin{aligned} e &= \left( \frac{K_0}{K_0 + K_1x + K_2y} \right) e_0 \\ e_x &= \left( \frac{K_1^{ex}x + K_2^{ex}y}{K_0 + K_1x + K_2y} \right) e_0 \\ e_g &= \left( \frac{K_1^{eg}x + K_2^{eg}y}{K_0 + K_1x + K_2y} \right) e_0 \\ e_y &= \left( \frac{K_1^{ey}x + K_2^{ey}y}{K_0 + K_1x + K_2y} \right) e_0. \end{aligned}$$

We substitute these into the total concentration of G protein, given by the equation

$$\begin{aligned} g_0 &= x + y + e_x + e_g + e_y \\ &= x + y + \left( \frac{K_1x + K_2y}{K_0 + K_1x + K_2y} \right) e_0. \end{aligned}$$

It is important to note that this definition of  $g_0$  explicitly accounts for the concentrations of intermediate complexes. (For the QSS model we had to assume  $e_0 \ll g_0$ , because we were not accounting for these concentrations.) This means that in the following analysis we are free to choose  $e_0$  and  $g_0$  independently, without restriction.

From equation 6.21 we also know that at steady-state  $x = \kappa y$ , so

$$\begin{aligned} g_0 &= (\kappa + 1)y + \left( \frac{(K_1\kappa + K_2)y}{K_0 + (K_1\kappa + K_2)y} \right) e_0 \\ &= (\kappa + 1)y + \left( \frac{y}{K_s + y} \right) e_0 \end{aligned}$$

where  $K_s = \frac{K_0}{(K_1\kappa + K_2)}$ . This equation can be rearranged to give a quadratic equation in  $y$ ,

$$0 = (\kappa + 1)y^2 + 2by - K_s g_0$$

where

$$b = \frac{1}{2} (e_0 - g_0 + (\kappa + 1)K_s).$$

The coefficients of the quadratic equation are given only by combinations of  $e_0$ ,  $g_0$ , and the summary constants. This equation can be solved using the quadratic formula to give two solutions

$$y = \frac{1}{\kappa + 1} \left( -b \pm \sqrt{b^2 + (\kappa + 1)K_s g_0} \right). \quad (6.23)$$

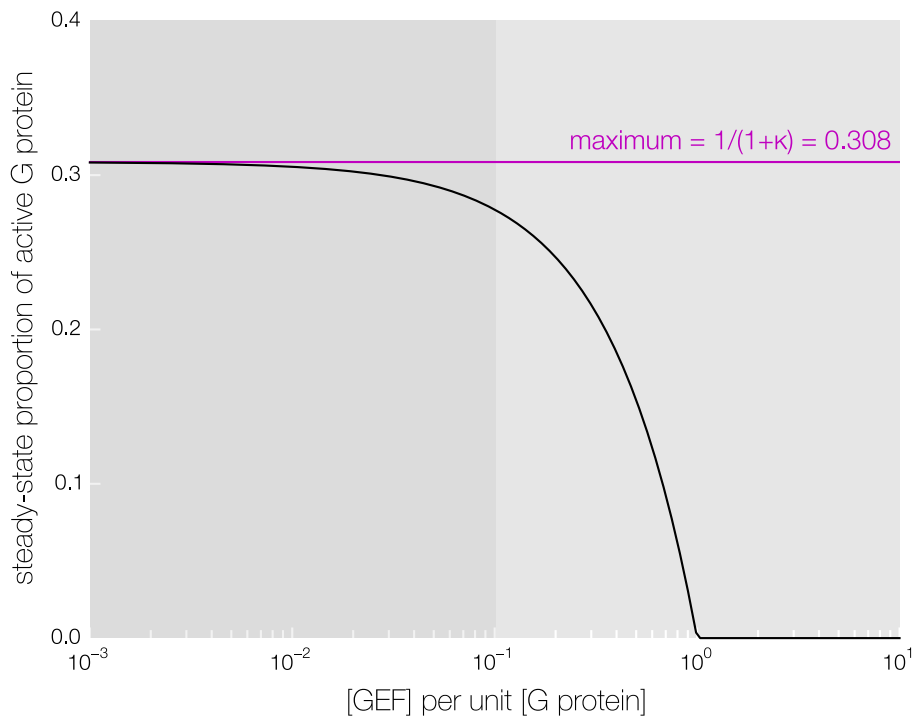
Because the concentration of active G protein must be positive, I am interested in the sign of these two solutions, which can be reduced to investigating the sign of  $-b \pm \sqrt{b^2 + q}$ , where  $q = (\kappa + 1)K_s g_0 > 0$ . Note that  $\sqrt{b^2 + q}$  is always positive. If  $b < 0$ , then

$$\begin{aligned} -b + \sqrt{b^2 + q} &> 0; \\ -b - \sqrt{b^2 + q} &< 0, \text{ by the triangle inequality.} \end{aligned}$$

Alternatively, if  $b \geq 0$ , then

$$\begin{aligned} -b + \sqrt{b^2 + q} &> 0, \text{ by the triangle inequality;} \\ -b - \sqrt{b^2 + q} &< 0. \end{aligned}$$

So the solution taking the minus sign is always negative, and the solution taking the plus sign is always positive. Therefore only the solution taking the positive sign is



**Figure 6.7** The steady-state proportion of active G protein as a function of the total concentration of GEF for the Ran:RCC1 parameters listed in table 6.1. The equation that describes this figure is given in equation 6.24. The darker shaded region indicates the region I assume is physiologically relevant in healthy conditions, where the inhibitory effect of the GEF-G protein complex formation is small. The exact form of this figure would differ for different G protein:GEF systems.

physiologically-relevant. Note that this equation gives the steady-state concentration of active G protein in terms of only the summary constants and total concentrations.

This equation can be used to derive an alternative equation for the proportion of G protein that is active (without the restriction that  $e_0 \ll g_0$ ),

$$\frac{y}{g_0} = \frac{1}{g_0(\kappa + 1)} \left( -b + \sqrt{b^2 + (\kappa + 1)K_s g_0} \right). \quad (6.24)$$

For the Ran:RCC1 parameters this equation is plotted as figure 6.7. At very low concentrations of GEF, we see that the proportion of active G protein effectively reaches the previously derived maximum value given by equation 6.22. At mid concentrations of GEF, there is inhibition of active G protein (via the formation of excess G protein:GEF complexes). At high concentrations of GEF – parity with the concentration of G protein and above, there is effectively zero free active G protein.

So far I have only discussed the steady-state dynamics of the system, however this ignores one contribution of the concentration of GEF. We also know that the concentration of GEF controls the maximum rate for the guanine nucleotide exchange activity in the system. This is directly evident from the QSS model given by equation 6.12. This suggests that there will be a tradeoff in terms of increasing the GEF concentration: low concentrations will have no inhibitory effect, but a slow total catalytic activity; high concentrations will have a inhibitory effect, but a fast total catalytic activity. Therefore, this suggests a physiologically plausible region for the concentration of GEFs. This would be where this inhibitory effect is not so pronounced that it is detrimental to the amount of active G protein, but where there is sufficient GEF present to catalyse the reaction at an appropriate rate.

Restating the initial question: how does the steady-state value of the active form of the G protein ( $y$ ) change as the total concentration of GEF ( $e_0$ ) is varied? In equation 6.23  $e_0$  is (only) found within the parameter  $b$ , so the answer can be found by investigating the derivative of equation 6.23 with respect to  $b$ ,

$$\frac{dy}{db} = \frac{1}{\kappa + 1} \left( \frac{b}{\sqrt{b^2 + (\kappa + 1)K_s g_0}} - 1 \right) < 0.$$

This is always negative as, by the triangle inequality,

$$b < \sqrt{b^2 + (\kappa + 1)K_s g_0}$$

and so

$$\frac{b}{\sqrt{b^2 + (\kappa + 1)K_s g_0}} < 1.$$

So as the amount of GEF in the system ( $e_0$ ) is increased the steady-state concentration of active G protein ( $y$ ) must decrease (and vice-versa). This proves that this inhibitory effect is present for all possible sets of parameters, and therefore all G protein:GEF systems.

### 6.4.3 GTPase activity has a crucial role in the observed activation of G protein

In the simulations in figures 6.5B and 6.5C, we have already seen that the addition of GTPase activity—both intrinsic and GAP-mediated—is sufficient to restore an apparent activation of G proteins by GEFs. In these simulations, stimulating the system by increasing the concentration of GEF does increase the steady-state concentration of active G protein. While I previously needed to prove impossibility of a particular behaviour, proving the converse—the possibility—requires only that a single example is found, for which these particular simulations suffice. However, the question of why this is sufficient remains.

We know that the role of GTPase activity is to inactivate G proteins. This irreversible process will convert the entirety of the active G protein into inactive G protein if left unchecked. So far I have demonstrated that GEFs act to create an equilibrium between active and inactive G protein. The regulation of the full G protein:GEF:GAP system is thus controlled by the competition of guanine nucleotide exchange against GTPase activity. This competition can be directly observed in figures 6.5B and 6.5C, the simulations including intrinsic and GAP-mediated GTPase activity, respectively.

In these simulations we can see that the initial steady-state concentration of active G protein appears to be suppressed in comparison to the simulation without GTPase activity, in figure 6.5A. Correspondingly, the steady-state concentration of inactive G protein is greater than in the simulation without GTPase activity. This initial regime corresponds to a low concentration of GEF, so the GTPase activity dominates – moving the system away from the GEF-mediated equilibrium values. Stimulating the system by increasing the concentration of GEF changes the relative rate of the two activities, and so the system moves towards the GEF-mediated equilibrium values.

For intrinsic GTPase activity (modelled by first-order exponential decay) I will derive some further mathematical results showing the effect of the relative rates of total GEF forward catalytic activity ( $e_0k_{\text{cat}}$ ) and GTPase activity ( $k_{\text{ase}}$ ). I have restricted this analysis to intrinsic GTPase activity, as the Michaelis-Menten model

used for the GAP-mediated GTPase activity complicates the mathematical analysis. As the GAP-mediated process is also irreversible, it is plausible that the following results (or very similar) will also hold for that process.

Assuming the QSS model for GEF, given by equation 6.12, and intrinsic GTPase activity modelled by exponential decay, we can write

$$\frac{dy}{dt} = \frac{k_{\text{cat}}(x - \kappa y)e_0}{K_0 + K_1x + K_2y} - k_{\text{ase}}y.$$

Now at steady-state,  $\frac{dy}{dt} = 0$ , and so we have

$$\begin{aligned} 0 &= k_{\text{cat}}(x - \kappa y)e_0 - k_{\text{ase}}y(K_0 + K_1x + K_2y) \\ &= (k_{\text{cat}}e_0 - k_{\text{ase}}K_1y)x - (k_{\text{ase}}K_2y + k_{\text{ase}}K_0 + k_{\text{cat}}\kappa e_0)y \\ &= (\hat{\kappa}e_0 - K_1y)x - (K_2y + K_0 + \hat{\kappa}\kappa e_0)y \end{aligned}$$

where

$$\hat{\kappa} = \frac{k_{\text{cat}}}{k_{\text{ase}}}.$$

The above equation can then be rearranged to give

$$x = \frac{K_2y^2 + (K_0 + \kappa\hat{\kappa}e_0)y}{\hat{\kappa}e_0 - K_1y}. \quad (6.25)$$

This equation gives the steady-state concentration of inactive G protein in terms of the active G protein. It is analogous to equation 6.21, which gave this same result in the absence of GTPase activity ( $x = \kappa y$ ).

The quasi-steady-state values of  $e_x$ ,  $e_g$ , and  $e_y$ , given in equations 6.8 to 6.11, and the new function for the steady-state value of  $x$  given  $y$ , given by equation 6.25, can be substituted into the equation for the total amount of G protein

$$\begin{aligned} g_0 &= x + y + e_x + e_g + e_y \\ &= x + y + \left( \frac{K_1x + K_2y}{K_0 + K_1x + K_2y} \right) e_0. \end{aligned}$$

and rearranged, to give a cubic equation in  $y$

$$0 = c_3y^3 + c_2y^2 + c_1y + c_0 \quad (6.26)$$

with the coefficients

$$c_0 = \hat{\kappa}^2 K_s g_0 e_0$$

$$\begin{aligned}
c_1 &= -2\hat{\kappa}^2 e_0 b - 2\hat{\kappa} K_s (K_1(b + g_0) - K_d K_s) \\
c_2 &= -\hat{\kappa}^2 (\kappa + 1) e_0 - 2\hat{\kappa} (K_1(b + g_0 - e_0) - 2K_d K_s) + K_1^2 K_s \\
c_3 &= 2\hat{\kappa} K_d
\end{aligned}$$

and with  $K_d = \frac{1}{2}(K_1 - K_2)$ . Cubic equations are in general solvable, and so it is possible to find analytical solutions for equation 6.26. However, the complexity of the coefficients mean that these are uninformative in providing analytical results about the system.

So instead, the equivalent equation can be re-derived under the assumption that  $e_0 \ll g_0$ . Again this means that  $e_x \approx 0$ ,  $e_g \approx 0$ , and  $e_y \approx 0$ , and so

$$\begin{aligned}
g_0 &= x + y \\
&= \frac{K_2 y^2 + (K_0 + \kappa \hat{\kappa} e_0) y}{\hat{\kappa} e_0 - K_1 y} + y
\end{aligned}$$

which can be rearranged to give a quadratic in  $y$

$$\begin{aligned}
0 &= K_2 y^2 + (K_0 + \kappa \hat{\kappa} e_0) y + (y - g_0)(\hat{\kappa} e_0 - K_1 y) \\
&= (K_2 - K_1) y^2 + 2\hat{b} y - \hat{\kappa} g_0 e_0
\end{aligned}$$

where  $\hat{b} = \frac{1}{2}(K_0 + K_1 g_0 + (\kappa + 1)\hat{\kappa} e_0)$ . The solutions to this quadratic equation are

$$y = \frac{1}{K_2 - K_1} \left( -\hat{b} \pm \sqrt{\hat{b}^2 + (K_2 - K_1)\hat{\kappa} g_0 e_0} \right).$$

I will assume for the following analysis that  $K_1 \neq K_2$ .

The concentration of active G protein must be positive, so I am interested in the sign of these two solutions. Let  $\delta = (K_2 - K_1)^{-1}$ ,  $q = \hat{\kappa} g_0 e_0 > 0$ , and note that  $\hat{b}$  is always positive. If  $\delta > 0$  or  $K_2 > K_1$ , then

$$\begin{aligned}
\delta \left( -\hat{b} + \sqrt{\hat{b}^2 + q/\delta} \right) &> 0 \text{ by the triangle inequality as } q/\delta > 0; \\
\delta \left( -\hat{b} - \sqrt{\hat{b}^2 + q/\delta} \right) &< 0.
\end{aligned}$$

Alternatively, if  $\delta < 0$  or  $K_2 < K_1$ , then

$$\begin{aligned}
\delta \left( -\hat{b} + \sqrt{\hat{b}^2 + q/\delta} \right) &> 0 \text{ by the triangle inequality as } q/\delta < 0; \\
\delta \left( -\hat{b} - \sqrt{\hat{b}^2 + q/\delta} \right) &> g_0 > 0.
\end{aligned}$$

In this case then there appear to be two positive solutions, however the second of these is always greater than the total concentration of G protein. This can be shown by a proof by contradiction – assume the converse

$$g_0 \geq \delta \left( -\hat{b} - \sqrt{\hat{b}^2 + q/\delta} \right)$$



then, by the triangle inequality,

$$g_0 > -2\delta\hat{b}$$

We know  $\delta^{-1} = K_2 - K_1 < 0$  and so substituting in the values of  $\hat{b}$  and  $\delta$  gives

$$\begin{aligned} g_0(K_2 - K_1) &< -(K_0 + K_1g_0 + (\kappa + 1)\hat{\kappa}e_0) \\ g_0K_2 &< -(K_0 + (\kappa + 1)\hat{\kappa}e_0) \\ \Rightarrow K_2 &< -\frac{K_0 + (\kappa + 1)\hat{\kappa}e_0}{g_0} < 0 \end{aligned}$$

but this is a contradiction as  $K_2 \geq 0$ . The following code in the Mathematica programming language—which evaluates to false—also suffices to prove this statement:

```
1 (* k = kappa, h = kappa hat *)
2 constraints = {K0>0, K1>0, K2>0, k>0, h>0, e0>0, g0>0};
3 b = (K0 + K1 g0 + (k + 1) h e0)/2;
4 expr = (-b - Sqrt[b^2 + h g0 e0 (K2 - K1)])/(K2 - K1);
5 Reduce[Union[{expr < g0, K2 < K1}, constraints]] (* false *)
```

I also note that  $\hat{b}^2 + e_0/g_0 > 0$ , and so all solutions are real. The following code in the Mathematica programming language—which again evaluates to false—suffices to prove this statement:

```
1 (* k = kappa, h = kappa hat *)
2 constraints = {K0>0, K1>0, K2>0, k>0, h>0, e0>0, g0>0};
3 b = (K0 + K1 g0 + (k + 1) h e0)/2;
4 expr = b^2 + h g0 e0 (K2 - K1);
5 Reduce[Union[{expr < 0, K2 < K1}, constraints]] (* false *)
```

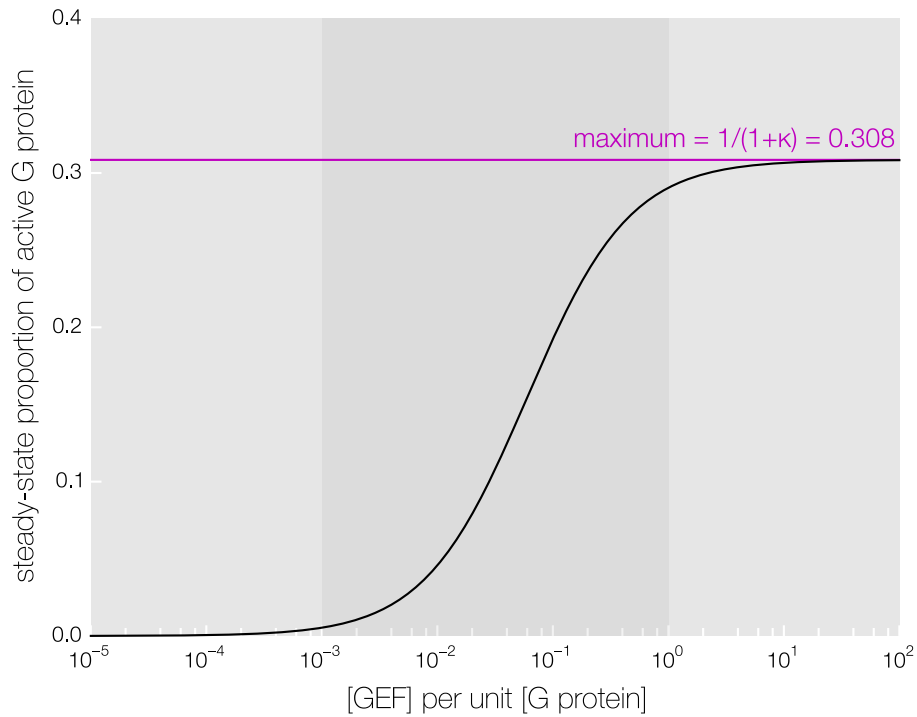
Therefore, there is only one physiologically relevant solution, which is

$$y = \frac{1}{K_2 - K_1} \left( -\hat{b} + \sqrt{\hat{b}^2 + (K_2 - K_1)\hat{\kappa}g_0e_0} \right).$$

Again, this equation can be used to derive an alternative equation for the proportion of G protein that is active,

$$\frac{y}{g_0} = \frac{1}{g_0(K_2 - K_1)} \left( -\hat{b} + \sqrt{\hat{b}^2 + (K_2 - K_1)\hat{\kappa}g_0e_0} \right). \quad (6.27)$$

The relative strength of the total forward catalytic activity of the GEF ( $e_0k_{\text{cat}}$ ) and the GTPase activity ( $k_{\text{ase}}$ ) is given by  $e_0\hat{\kappa}$ . So I am interested in what happens as this value changes. Though we are restricted by the constraint  $e_0 \ll g_0$ , we remain unrestricted (mathematically) in the choice of both  $k_{\text{cat}}$  and  $k_{\text{ase}}$ .



**Figure 6.8** The steady-state proportion of active G protein as a function of the total concentration of GEF divided by the total rate of GTPase activity for the Ran:RCC1 parameters listed in table 6.1. The equation that describes this figure is given in equation 6.27. Note that although this figure follows the constraint  $e_0 \ll g_0$ , we can choose GTPase activity small enough to compensate. Therefore it may be misleading to directly compare this figure with figure 6.7. The darker shaded region indicates the region I assume is physiologically relevant in healthy conditions, where the system is responsive to changes in the concentration of GEF. The exact form of this figure would differ for different G protein:GEF systems.

For the Ran:RCC1 parameters equation 6.27 is plotted in figure 6.8. Note that we have already set the value of  $k_{\text{cat}}$  with the choice of these parameters, so the x-axis in this figure corresponds to the value of  $\frac{e_0}{k_{\text{ase}}}$ , which we are free to change under the assumption that  $e_0$  remains small. In which case increasing  $e_0$  or decreasing  $k_{\text{ase}}$  gives an equivalent response.

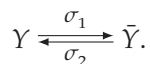
When there is high relative GTPase activity there is effectively zero active G protein. When there is low relative GTPase activity we again reach the previously derived maximum value given by equation 6.22. In between, the response is sigmoidal and is such that increasing the concentration of GEF (relative to the GTPase activity) is now capable of increasing the steady-state concentration of active G protein.

This is further proof that GTPase activity is required for GEFs to be able to regulate G proteins. Here, we can now directly see how the competition between the two activities affects the steady-state concentration of the active form of the G protein. GTPase activity appears to act to suppress the system away from the maximum equilibrium value. The strength of this suppression depends only on the relative total rate of guanine nucleotide exchange to GTPase activity. As the relative GTPase activity decreases—for example, through an increase in concentration of GEF—then the proportion of active G protein returns towards the maximum equilibrium value. Therefore, in order to properly respond to an activating or inhibitory signal (increasing or decreasing the GEF or GAP concentration) the system must lie somewhere on the sigmoidal region.

#### 6.4.4 Sequestration is not sufficient to restore the activation of G proteins by GEFs

Alternatively, other mechanisms may be sufficient to explain an observed activation of G proteins by GEFs. However—as I will demonstrate—simple reversible sequestration of active G protein is insufficient to explain this effect.

Sequestration of active G protein (into a new form  $\bar{Y}$ ,  $\bar{y} = [\bar{Y}]$ ) could be caused by: binding to a partner species; changing form; or, moving location. The simplest version of this process can be modelled by the inclusion of an extra reversible reaction step to the reaction scheme shown in equation 6.1,



This implies an extra ODE

$$\frac{d\bar{y}}{dt} = \sigma_1 y - \sigma_2 \bar{y}.$$

Which at steady-state,  $\frac{d\bar{y}}{dt} = 0$ , implies

$$\bar{y} = \bar{\kappa}y, \text{ where } \bar{\kappa} = \frac{\sigma_1}{\sigma_2}.$$

Assuming no GTPase activity, we still know that  $x = \kappa y$ , which can be derived from  $\frac{dx}{dt} = 0$ .

Now the conservation of mass equation for the G protein is given by

$$\bar{g}_0 = x + y + e_x + e_g + e_y + \bar{y}$$

into which we can substitute the quasi-steady-state values of  $e_x$ ,  $e_g$ , and  $e_y$ , given in equations 6.8 to 6.11, and the steady-state values of  $x$  and  $\bar{y}$  to obtain a quadratic equation,

$$0 = (\bar{\kappa} + \kappa + 1)y^2 + 2\bar{b}y - K_s\bar{g}_0$$

where  $\bar{b} = \frac{1}{2}(e_0 - \bar{g}_0 + (\bar{\kappa} + \kappa + 1)K_s)$ .

By identical reasoning to that in section 6.4.2, there is one positive solution

$$y = \frac{1}{\bar{\kappa} + \kappa + 1} \left( -\bar{b} + \sqrt{\bar{b}^2 + (\bar{\kappa} + \kappa + 1)K_s\bar{g}_0} \right)$$

and again

$$\frac{dy}{d\bar{b}} < 0.$$

Now under the assumption that both the active and the sequestered forms of the G protein are functionally active, then we are interested in how  $y + \bar{y}$  changes with respect to  $\bar{b}$ ,

$$\frac{d}{d\bar{b}}(y + \bar{y}) = \frac{d}{d\bar{b}}(y + \bar{\kappa}y) = (1 + \bar{\kappa}) \frac{dy}{d\bar{b}} < 0.$$

This result shows that even with this sequestration scheme, the total amount of active G protein is inversely correlated with the total amount of GEF in the system. Therefore, this simple sequestering mechanism is insufficient to explain the role of GEFs as activators of G proteins. However, it is still plausible that a more complicated sequestering model may be sufficient.

#### 6.4.5 Regulation of G protein regulation by GEFs and GAPs is by a balance/imbalance mechanism

Taking the above analysis together, it becomes difficult to defend the canonical description of GEFs as the ‘activators’ of G proteins. This is because I have shown,

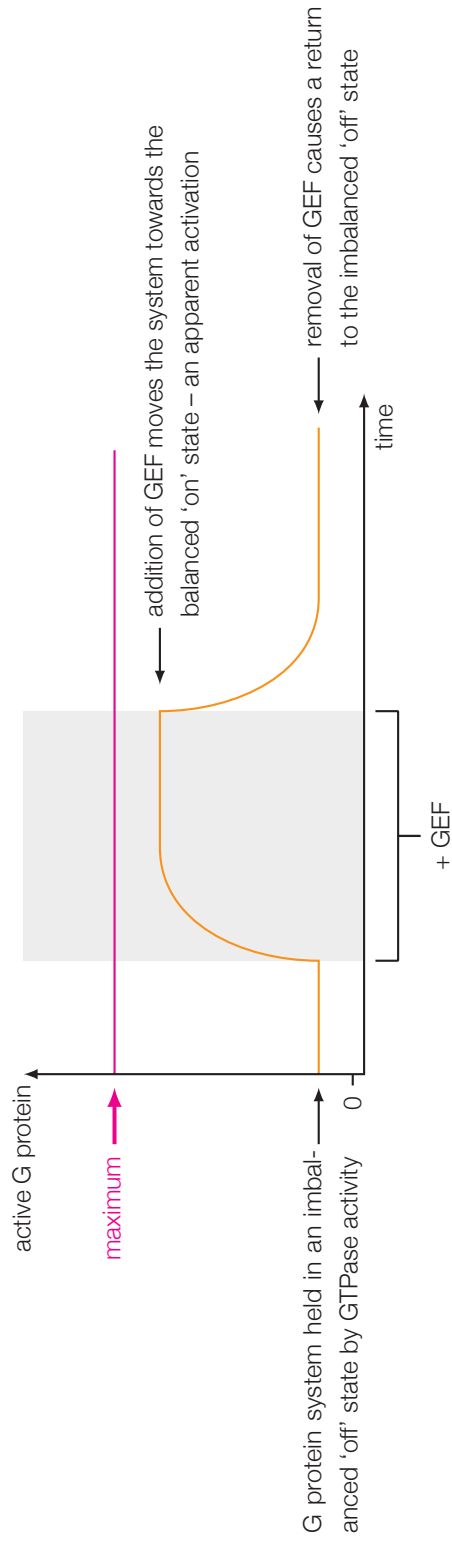
in section 6.4.1, that in the absence of GTPase activity, they seem more likely to act as inhibitors towards the concentration of active G protein. I have also shown, in section 6.4.1, how the true role of GEFs appears to be to drive the system towards an equilibrium ratio of active to inactive forms of the G protein, given by the constant  $\kappa$ . So GEFs could be said to act as a *balancing force* in the system.

In contrast, GTPase activity drives the system away from this equilibrium, towards the state where none of the G protein is in the active form. So GTPase activity could be said to act as an *imbaling force* in the system.

In this sense the regulation of G proteins by GEFs and GAPs could be best described as a *balance/imbalance* mechanism. The relative strengths of the guanine nucleotide exchange and the GTPase activity will control the extent to which the system is either in the balanced or imbalanced state, and so control the activation of the system. This is illustrated in figure 6.9, which describes the form of the dynamics seen in figures 6.5B and 6.5C. The important point to note from this illustration is that the G protein can be activated by increasing the concentration GEFs, but only by starting from an initially suppressed state. This activated state is then maintained only so long as there is sufficient GEF present. Both the suppressed and activated states are below the maximum (equilibrium) concentration of active G protein, which we know is reached only when  $e_0 \ll g_0$  and  $k_{ase} = 0$ . So GTPase activity could be described as crucial in allowing the differential activation of G proteins by GEFs. Therefore I suggest that G protein regulation must include both a functioning GEF and a functioning GAP to display the complete and correct range of regulatory behaviours.

This negative regulation of G proteins by GTPase activity is already known to be crucial in the healthy control of intracellular signalling. For example, it is common for mutations in Ras implicated in cancer to cause insensitivity to its GAPs (Stephen et al. 2014). This leads to the G protein becoming constitutively active, where the system is insensitive to the suppression normally supplied by GTPase activity. Similarly, the vast majority of GTP analogues used to investigate G protein signalling experimentally are non-hydrolysable, and so generate a form of active Arf which is immune to GTPase activity.

In totality, the regulation of G proteins by GEFs will be some sum of the two processes illustrated as figures 6.7 and 6.8. That is, the total activation will be some sum of the inhibitory effect caused by the formation of excess intermediate complexes and the activation caused by pushing the system towards equilibrium. This will primarily depend on the total amount of G protein, and the strength of the



**Figure 6.9** The proposed balance/imbalance mechanism by which GEFs and GAPs have distinct and crucial roles, and collaborate to regulate G proteins. GEFs act as the balancing force, attempting to move the system towards a maximum activation. GAPs act as an imbalancing force, attempting to move the system towards an imbalanced state favouring the inactive form of the G protein.

GTPase activity.

Finally, I predict that experiments that attempt to regulate G proteins by over-expression of GEFs are likely to produce unexpected behaviour – as in some circumstances this may cause the seemingly paradoxical inhibition of the G protein rather than the activation. The activation of G proteins through indirect methods should therefore be preferentially attempted, for example through reduction of GTPase activity.

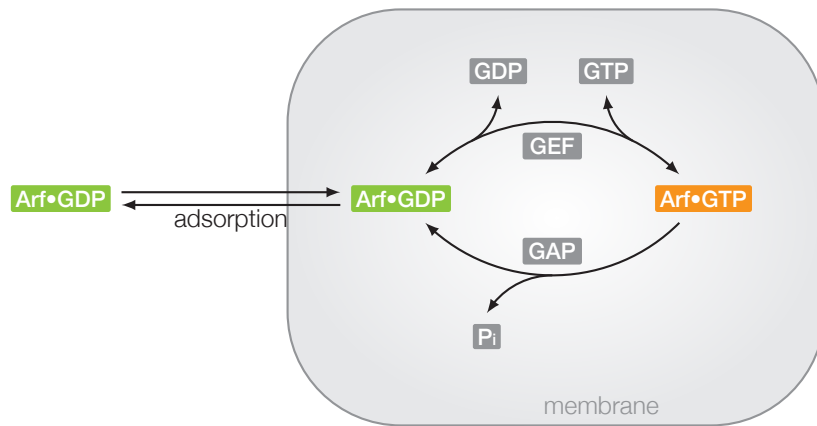
The mathematical and theoretical underpinning to these result means that it should hold for all G proteins (heterotrimeric or small) as long as the mechanism of nucleotide exchange is comparable to that in figure 6.2. Dynamics for more complicated mechanisms—for instance systems with implicit GEF:G protein:GAP complexes, which have previously been observed (Berstein et al. 1992)—would require additional analysis.

## 6.5 Regulation of Arf: specific mechanisms

The above analysis focused on developing results that applied to a generic G protein regulated by a generic GEF and GTPase activity, either intrinsic to the G protein or by a generic GAP. Therefore these results should hold for any G protein:GEF:GAP system which matches this regulation.

The Arf family of G proteins are regulated by whole classes of GEFs and GAPs (Donaldson & Jackson 2011). I will assume that these behave exactly as described above: Arf GEFs with a reversible reaction scheme; and Arf GAPs with Michaelis-Menten with product inhibition reaction scheme. However one difference in regards to the system as described so far is that while inactive Arfs can be cytosolic, the regulation of Arf by GEFs is restricted to membranes. This is because Arf GEFs are themselves localised to membranes (Kolanus 2007; Casanova 2007; Donaldson & Jackson 2011); and as active Arf remains membrane bound, so GAP activity must also be on the membrane.

Recall that active Arf is known to be membrane-localised through a myristoylation on its N-terminus. In inactive Arf this myristoylation is hidden, and so the process of activation causes the insertion of this domain into the lipid membrane. Inactivation of Arf via GAPs—and as I will continue to assume—via GEFs mediating guanine nucleotide exchange in reverse will cause the retraction of the myristoylated N-terminus from the membrane – and so inactive Arf loses its membrane-tether. I will need to include this differential localisation of Arf into the final model. The loss of



**Figure 6.10** Regulation of Arf by GEFs, GAPs, and an adsorption process. Guanine nucleotide exchange and GTPase activity for Arf occurs solely on the membrane, but inactive Arf has no specific membrane-localisation. Therefore inactive Arf must approach the membrane before activation can occur.



membrane localisation of inactive Arf is—in one sense—theoretically equivalent to sequestration of the inactive form of Arf.

The complete mechanism of the regulation of Arf by GEFs/GAPs will need to describe at least the following steps:

- Inactive, cytosolic Arf approaches the membrane.
- Already membrane-localised GEF catalyses the guanine nucleotide exchange of the G protein. During this process the myristoylated N-terminus of the G protein is inserted into the membrane.
- A membrane-localised GAP catalyses the hydrolysis of the GTP bound to the active G protein, thus inactivating the G protein. During this process the myristoylated N-terminus of the G protein is retracted from the membrane.
- Alternatively, GEFs retain the ability to inactivate the G protein by mediating guanine nucleotide exchange in reverse.
- Inactive G protein is free to diffuse away from the membrane.

These processes are illustrated as figure 6.10.

## 6.6 Regulation of Arf: QSS model

Recall that equation 6.12 gives a general equation for the regulation of active G protein by a GEF,

$$f_{\text{GEF}}(x, y) = \frac{k_{\text{cat}}(x - \kappa y)e_0}{K_0 + K_1x + K_2y}.$$

Also that equation 6.14 gives a general equation for the regulation of active G protein by a GAP,

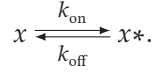
$$f_{\text{GAP}}(x, y) = -\frac{k_{\text{ase}}yf_0}{K_m(1 + \frac{x}{K_{\text{ic}}}) + y}.$$

Here  $x$  is the concentration of inactive G protein;  $y$  is the concentration of active G protein;  $e_0$  is the total concentration of GEF;  $f_0$  is the total concentration of GAP; and the other symbols are summary constants.

Now, we need to indicate that  $x$  and  $y$  are extrinsic components of the membrane, and so we relabel  $x \rightarrow x^*$  and  $y \rightarrow y^*$ . So we are interested in

$$f_{\text{GEF}}(x^*, y^*) \text{ and } f_{\text{GAP}}(x^*, y^*).$$

We also need reactions that describe rate of binding and unbinding of the inactive Arf to the membrane,



In the previous discussion of surface-active enzymes, in section 3.5, I considered enzymes which were adsorbed on the membrane. Here it is not the enzyme, but the regulated species that is adsorbed. Equation 3.17 can easily be modified accordingly to give the relevant equation,

$$f_{\text{adsorption}}(x, x^*) = k_{\text{on}}\Phi[A]x - k_{\text{off}}x^*$$

where  $[A]$  is the surface area per unit volume; and  $\Phi$  is the available area function for which a suitable form needs to be determined.

Figure 6.10 shows the relationship between the GEF, GAP, and adsorption processes. This directly implies the equations that model these processes,

$$\begin{aligned} \frac{dx}{dt} &= -f_{\text{adsorption}}(x, x^*) \\ \frac{dx^*}{dt} &= -f_{\text{GEF}}(x^*, y^*) + f_{\text{GAP}}(x^*, y^*) + f_{\text{adsorption}}(x, x^*) \\ \frac{dy^*}{dt} &= f_{\text{GEF}}(x^*, y^*) - f_{\text{GAP}}(x^*, y^*). \end{aligned}$$

This system of equations follows the assumptions that  $e_0 \ll g_0$  and  $f_0 \ll g_0$ , where

$$g_0 = x + x^* + y^*.$$

I can use this conservation of mass equation to reduce the three equations ( $\frac{dx}{dt}$ ,  $\frac{dx^*}{dt}$ ,  $\frac{dy^*}{dt}$ ) to two. I am free to choose which equation I remove: I want to keep the equation for  $y^*$ , the concentration of active Arf; I will also choose to keep the equation for  $x$ , the concentration of cytosolic Arf. So I will let  $x^* = g_0 - x - y^*$ , and so

$$\begin{aligned} \frac{dx}{dt} &= -f_{\text{adsorption}}(x, g_0 - x - y^*) \\ \frac{dy^*}{dt} &= f_{\text{GEF}}(g_0 - x - y^*, y^*) - f_{\text{GAP}}(g_0 - x - y^*, y^*). \end{aligned}$$

Written out in full, we have

$$\begin{aligned} \frac{dx}{dt} &= -k_{\text{on}}\Phi[A]x + k_{\text{off}}(g_0 - x - y^*) \\ \frac{dy^*}{dt} &= \frac{k_{\text{cat}}(g_0 - x - y^* - \kappa y^*)e_0}{K_0 + K_1(g_0 - x - y^*) + K_2 y^*} - \frac{k_{\text{ase}} y^* f_0}{K_m \left(1 + \frac{(g_0 - x - y^*)}{K_{\text{ic}}}\right) + y^*} \end{aligned}$$

which can be rearranged to give

$$\frac{dx}{dt} = k_{\text{off}} (g_0 - (1 + k_A \Phi[A]) x - y^*) \quad (6.28)$$

$$\frac{dy^*}{dt} = \frac{k_{\text{cat}}(g_0 - x - (\kappa + 1)y^*)e_0}{K_0 + K_1(g_0 - x) + K_2' y^*} - \frac{k_{\text{ase}} y^* f_0}{K_0'' - K_1'' x + K_2'' y^*} \quad (6.29)$$

where

$$\begin{aligned} k_A &= \frac{k_{\text{on}}}{k_{\text{off}}}; & K_2' &= K_2 - K_1; \\ K_0'' &= K_m \left(1 + \frac{g_0}{K_{\text{ic}}}\right); & K_1'' &= \frac{K_m}{K_{\text{ic}}}; & K_2'' &= 1 - \frac{K_m}{K_{\text{ic}}}. \end{aligned}$$

Equations 6.28 and 6.29 give what I will refer to as the final model for the regulation of Arf by an Arf-GEF and an Arf-GAP. It consists of two equations which model the regulation of Arf as was illustrated in figure 6.10.

Alternatively, it is possible to reduce this model to a single equation using the rapid equilibrium assumption (see section 3.2.4). Assuming

$$f_{\text{adsorption}}(x, x^*) = 0 \implies x^* = k_A \Phi[A] x.$$

Substituting into the conservation of mass equation gives

$$g_0 = (1 + k_A \Phi[A]) x + y^*$$

which implies

$$x = \frac{1 + k_A \Phi[A]}{g_0 - y^*}.$$

This can be substituted into equation 6.29 to give

$$\frac{dy^*}{dt} = \frac{a_0 + a_1 y^* + a_2 y^{*2}}{a_3 + a_4 y^* + a_5 y^{*2}} + \frac{b_1 y^* + b_2 y^{*2}}{b_3 + b_4 y^* + b_5 y^{*2}} \quad (6.30)$$

where

$$\begin{aligned} a_0 &= k_{\text{cat}} e_0 (g_0^2 - X); & a_1 &= -k_{\text{cat}} e_0 g_0 (2 + \kappa); & a_2 &= k_{\text{cat}} e_0 (1 + \kappa); \\ a_3 &= g_0 K_0' - K_1' X; & a_4 &= g_0 K_2' - K_0'; & a_5 &= -K_2'; \\ & & b_1 &= -k_{\text{ase}} f_0 g_0; & b_2 &= k_{\text{ase}} f_0; \\ b_3 &= g_0 K_0'' - K_1'' X; & b_4 &= g_0 K_2'' - K_0''; & b_5 &= -K_2''; \end{aligned}$$

with  $X = (1 + k_A \Phi[A])$ .

## 6.7 Discussion

I have now derived a quasi-steady-state model of nucleotide exchange on G proteins catalysed by GEFs based on a well-described, well-accepted molecular reaction scheme. This is a new derivation, using the linear framework of Gunawardena (2012) described in section 3.3, of the model previously described by Randazzo et al. (2013). Note that this derivation gives quasi-steady-state concentrations of the intermediate enzyme complexes – information that was crucial to prove most of the included results.

I have also determined sensible reaction schemes for GTPase activity, both intrinsic (modelled by exponential decay), and mediated by GAPs (modelled by Michaelis-Menten with product inhibition).

Together, the models of nucleotide exchange and GTPase activity form a complete model of the regulation of a generic G protein. From this, a specific model for Arf was generated by the inclusion of differential localisation of its active and inactive forms, modelled using the approach of Kartal & Ebenhöh (2013) described in section 3.5. This final regulatory model will be investigated further—independently and together with the mechanistic models for PLD and PI4P5K—in chapter 7.

Both full mass-action-derived systems of ODEs, and the quasi-steady-state solutions found by the application of the linear framework were used to derive mathematical results concerning the possible behaviours of the system. I have shown that there are certain universal properties of GEFs that arise from their reversible mechanism and which are independent of any particular kinetic rates.

There is a fundamental flaw in the language commonly used to describe the regulation of G proteins by nucleotide exchange and GTPase activity. Rather than an ‘activation/inactivation’ cycle, this regulation is better described as a system controlled through ‘regulated balance/imbalance’. Here, GEFs act to balance the concentrations of active and inactive G proteins, and GTPase activity acts to imbalance the concentrations in favour of the inactive form. The distinction from the previous description is that description of GEFs as ‘activators’ is only true when the system is already imbalanced.

I have shown mathematically that GEFs, in the absence of GTPase activity, cannot increase the concentration of the active G protein above a theoretical maximum value defined by the ratio of the forwards and backwards kinetic rates. At steady-state, stimulation by increasing the GEF concentration can therefore not *positively* regulate the system. Addition of GTPase activity then restores the ability for GEFs to activate the system, by causing an imbalanced state.

The requirement for GTPase activity as a functional component in the activation of G proteins may have previously been under-appreciated due to the almost exclusive use of experimental systems where the GDP form of the G protein is the unique starting condition and where uptake of GTP is monitored. Overall, many previous studies that ignored the reversibility of GEFs will make valid conclusions under some—but crucially not all—physiologically-plausible conditions. However, this description is better placed to explore extreme scenarios, such as systems where GTPase activity is diminished, for example constitutively active transforming mutations in Ras common in cancers (Stephen et al. 2014). Note also that the irreversible model (figures 6.5B and 6.5C) and the reversible model (figures 6.5E and 6.5F) in the presence of GTPase activity have similar profiles, and so it may be difficult to experimentally distinguish these schemes.

These results also suggest that attempting to regulate G proteins by the over-expression of GEFs is likely to produce unexpected behaviour as in some circumstances this may cause inhibition of the G protein rather than activation. Activation of G proteins should therefore be preferentially attempted through reduction of the relevant GTPase activity. However, this suggests a novel approach for the inhibition of G proteins where it is desired, and suggests that GEF up-regulation could be a novel, naïvely paradoxical, therapeutic target.

The mathematical underpinning to these results mean that they should hold for all G proteins:GEF systems as long as the mechanism remains identical to figure 6.2. Conclusions based on alternative mechanisms, for instance systems with an implicit G protein alpha subunit:GEF:GAP complex (Berstein et al. 1992), would require further analysis.

I once more urge caution against description of GEFs as ‘enzymes that activate G proteins’, and against continued representations that show this mechanism as irreversible. This analysis demonstrates that GEFs should not be described as enzymes that simply convert a substrate into product, but as enzymes that act to attain an equilibrium – a balance – of active and inactive G protein. The two key roles of GTPase activity are then to: drive the system away from this equilibrium – to create an imbalance; and, to confer a direction for the apparent activation/inactivation cycle.



## 7 The Arf/PLD/PI4P5K motif: mechanistic models

I will now bring together the mechanistic models developed in chapters 5 and 6 to construct a complete mechanistic model of the entire Arf/PLD/PI4P5K motif including regulation of Arf by a GEF and a GAP.

In order to combine the component models into a complete model of the Arf/PLD/PI4P5K motif I require sensible relative concentrations of the involved species, derived in section 7.1; and I need to discuss the adsorption processes and associated constants, derived in section 7.2. Also I have yet to discuss the individual behaviours and characteristics of any of the component models and so I will briefly reintroduce and discuss these in sections 7.3 to 7.5. Finally, I will explain the construction of (multiple versions of) the complete model and demonstrate that it is capable of physiologically plausible behaviours in section 7.6.

Each of the models (component and complete) is suitable for inclusion into larger models of signalling pathways and other intracellular processes – it only remains to state a suitable set of parameters. Here I will derive a set of parameters that will be sufficient and suitable to allow preliminary computational simulation and provide a working basis for later numerical (and otherwise) investigation.

Information about the kinetic parameters of these enzymes is sparse, and so I will supplement the data that does exist with plausible values derived from alternative sources. This will include kinetic data from the BRENDA database (Schomburg et al. 2013); numbers from the BioNumbers database (Milo et al. 2010) with BNID; structural details from the PDB database (Berman et al. 2000) with PDB ID; and proportional abundances of proteins from the PaxDB database (Wang et al. 2012). Still, there will remain certain parameters for which there is no known data and for which I can determine no sensible method for their derivation via suitable proxy data. In these cases, in order to complete the parameterisation, I must choose arbitrary (but sensible) values – I will clearly note wherever I am forced to do so.

Name (Specific isozyme)	mol%	PPM*	Conc.	Source
Arf6	—	39.5	197 nM	PaxDB database (Wang et al. 2012) <sup>a</sup>
PLD2	—	0.22	1.10 nM	PaxDB database (Wang et al. 2012) <sup>a</sup>
PIP5K1C	—	0.71	3.55 nM	PaxDB database (Wang et al. 2012) <sup>a</sup>
Arf-GAP (AGAP1)	—	0.01	0.05 nM	PaxDB database (Wang et al. 2012) <sup>a</sup>
Arf-GEF (Cytohesin 1)	—	0.53	2.65 nM	PaxDB database (Wang et al. 2012) <sup>a</sup>
PC	43.20	—	14.26 mM	Pankov et al. (2006) <sup>b</sup>
PA	1.45	—	480 nM	Pankov et al. (2006) <sup>b</sup>
PI4P	(0.76)	—	250 nM	(see note) <sup>b,c</sup>
PI(4,5)P <sub>2</sub>	(0.07)	—	25 nM	(see note) <sup>b,c</sup>
GTP	—	—	305 μM	Traut (1994)
GDP	—	—	36 μM	Traut (1994)

**Table 7.1** Relative abundances of the species included in the models of PLD, PI4P5K, and the regulation of Arf.

\* Parts Per Million.

<sup>a</sup> As given by the 'Human, PaxDB integrated dataset'; scaled by 5 mM total concentration of protein (see section 7.1).

<sup>b</sup> Scaled by 33 μM total lipid (see section 7.2).

<sup>c</sup> PI 7.62 mol% as given by Pankov et al. (2006); assuming phosphorylated forms each an order of magnitude less, supported by statements in Di Paolo & De Camilli (2006).



Throughout this chapter I will use the following notation:

total concentration of PLD  $\rightarrow \text{PL}_t$   
total concentration of PI4P5K  $\rightarrow \text{PK}_t$   
concentration of cytosolic PI4P5K  $\rightarrow [\text{PK}_c]$   
total concentration of Arf  $\rightarrow \text{A}_t$   
concentration of cytosolic Arf·GDP  $\rightarrow [\text{A}_c]$   
concentration of membrane-associated Arf·GTP  $\rightarrow [\text{A}]$   
total concentration of (active) GEF  $\rightarrow \text{GEF}_t$   
total concentration of (active) GAP  $\rightarrow \text{GAP}_t$ .

Broadly—square brackets are retained for concentrations of proteins that will vary throughout a simulation and dropped for concentrations that will be fixed for the duration of a simulation. The exception to this rule is  $\text{GAP}_t$  which can be manually varied in order to perturb the complete system. I will also retain square brackets for all concentrations of lipids ( $[\text{PC}]$ ,  $[\text{PA}]$ ,  $[\text{PIP}]$ ,  $[\text{PIP}_2]$ ).

Parameters have been renamed so that these are distinct across the component models. Notation specific to membrane binding has been dropped for brevity.

In each model I will assume that concentrations of lipid substrates ( $[\text{PC}]$  and  $[\text{PIP}]$ ) are constant (well-buffered by unspecified processes), and that products are removed by an exponential decay (mediated by an unspecified sink). Note that in experimental systems with fixed total concentrations of lipid these assumptions may have to be re-evaluated.

Although I have previously included product inhibition in the models for PI4P5K and GAPs, I have been unable to find suitable numbers for the associated inhibition constants. So for the purposes of the following discussion I have chosen to assume that product inhibition is not observed; equivalently  $K_{ic} = \infty$ .

## 7.1 Estimates of protein & lipid concentrations

I require estimates of baseline concentrations for each of the involved proteins and lipids, for use as initial conditions and in the calculation of the decay constants (see below).

Relative abundances of the proteins were obtained from the ‘Human, PaxDB integrated dataset’ from the PaxDB database (Wang et al. 2012). These numbers

specify the parts per million (PPM) for each protein, so to convert these into concentrations we require an estimate of the total intracellular concentration of proteins. An estimate of  $3 \times 10^6$  proteins per  $\mu\text{m}^3$  ( $3 \times 10^{24}$  proteins per  $\text{m}^3$ ) is given by Milo (2013). Using Avagadro's constant ( $6.022 \times 10^{23} \text{mol}^{-1}$ ) this gives an estimated total protein concentration of

$$\frac{3 \times 10^{24}}{6.022 \times 10^{23}} \approx 5 \text{ mol m}^{-3} = 5 \text{ mM}.$$

This estimated total concentration was used to scale the PPM numbers to produce the concentrations listed in table 7.1.

Relative abundances (mol%) of lipids were obtained from Pankov et al. (2006). To produce the concentrations listed in table 7.1 the total concentration of lipids was assumed to be 3300 nM, see section 7.2.2.

## 7.2 Notes on adsorption processes

The models of PI4P5K and the regulation of Arf include adsorption processes. In order to simulate these models I need to choose the available area function,  $\Phi(\theta)$  (either Langmuir's model, section 3.5.3; or RSA, section 3.5.4), and specify a form for the fractional surface coverage,  $\theta$ . I will also require sensible estimates for the maximum surface concentration of adsorbates,  $n_0$ ; and the surface area per unit volume,  $[\alpha]$ .

### 7.2.1 A form for the fractional surface coverage

The coverage,  $\theta$ , describes the fraction of the membrane that is already covered by adsorbate, and so will be a function of the total concentration of membrane-associated molecules. Here I will identify assumptions that will specify a simple form for  $\theta$ .

First, I will assume that the concentrations of the constitutively membrane-associated species (GEF, GAP, PLD) can be ignored. Equivalently, there is additionally surface area sufficient for these that I will otherwise not discuss. Therefore,  $\theta$  will be a function of only the concentrations of membrane-bound Arf and PI4P5K.

Next, I need to know the relative areas occluded by a molecule of Arf and a molecule of PI4P5K. I will assume that the surface area occluded by either can be approximated by a disc with a given radius (e.g. caused by the projection of an approximately spherical protein onto the membrane).

It is possible to use the 3V tool (Voss & Gerstein 2010) on a known structure of Arf (Pasqualato et al. 2001) to estimate its radius as 13.75 Å. While there is no known structure for PI4P5K, structures for PI5P4Ks (PDB IDs: 2GK9, DOI:10.2210/pdb2gk9/pdb; 2YBX, DOI:10.2210/pdb2ybx/pdb) give effective radii of 14.98 Å and 16.54 Å. Alternatively, we know that Arfs have a molecular weight of around 21 kDa (Liu, Kahn, & Prestegard 2009) and PI4P5Ks have a molecular weight of around 68 kDa (Ishihara et al. 1996). Given these values, Erickson (2009) gives a formula which suggests a radius of 18.2 Å for Arf, and 26.9 Å for PI4P5K. Therefore, the radius of PI4P5K is likely to be 1.1–1.5 times as large as that of Arf; and so the area occluded 1.2–2.2 times as large. However, in order to specify a particularly simple form for  $\theta$ , I will assume that both Arf and PI4P5K occlude identical areas.

Given the above assumptions, we have

$$\theta = \frac{([\text{PK}_t] - [\text{PK}_c]) + (A_t - [A_c])}{[\alpha]n_0}. \quad (7.1)$$

## 7.2.2 Estimating membrane surface area

I need a plausible value for the total concentrations of lipids. This will give the factor used to scale the percentage compositions listed in table 7.1. During this derivation, I will need to choose a sensible value for the surface area per unit volume,  $[\alpha]$ .

I will assume that all lipids contribute identical surface area to the membrane. Two estimates for the surface area per lipid molecule are:

- 0.5 nm<sup>2</sup> from Brugger et al. (2006), BNID 106993
- 0.65 nm<sup>2</sup> from Nagle & Tristram-Nagle (2000), BNID 102781

So I will assume a (surface) concentration of

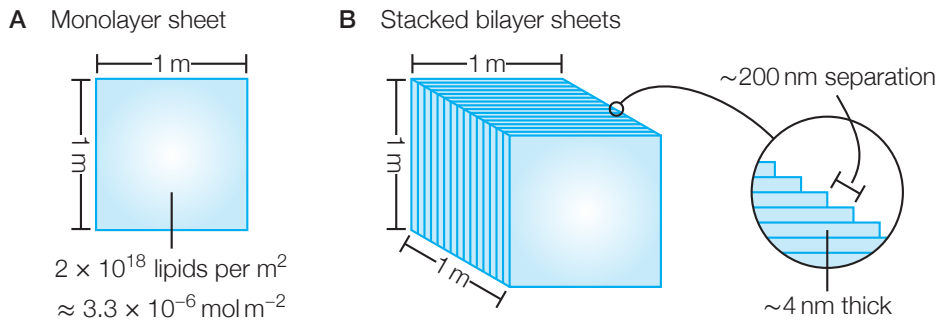
$$2 \text{ lipids per nm}^2 = 2 \times 10^{18} \text{ lipids per m}^2.$$

Using Avogadro's constant, this gives an estimated surface concentration of

$$\frac{2 \times 10^{18}}{6.022 \times 10^{23}} \approx 3.3 \times 10^{-6} \text{ mol m}^{-2}.$$

This is illustrated in figure 7.1A.

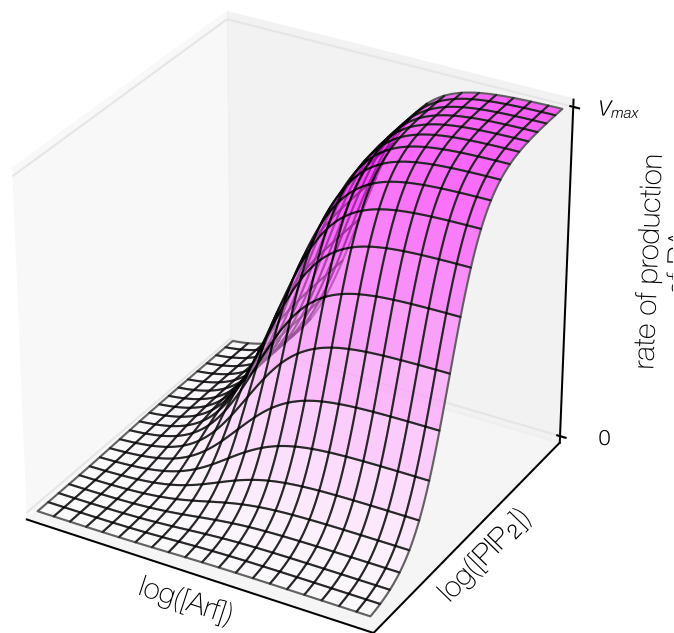
To convert this surface concentration to a volume concentration I need to specify a number for the surface area per unit volume,  $[\alpha]$ . In practice (in biological situations and in experimental setups) the value of  $[\alpha]$  will depend on multiple factors. For



**Figure 7.1** Illustration of lipid surface and volume concentrations.

**A** A  $1 \text{ m}^2$  membrane monolayer sheet is estimated to contain  $2 \times 10^{18}$  lipids, which equates to a surface concentration of approximately  $3.3 \times 10^{-6} \text{ mol m}^{-2}$ .

**B** A  $1 \text{ m}^3$  volume filled with 5000 equally spaced membrane bilayer sheets implies  $[\alpha] = 10^3 \text{ m}^2 \text{ per m}^3$ , which equates to a volume concentration of lipids of approximately  $33 \mu\text{M}$ . As the thickness of a bilayer is only approximately 4 nm (Rawicz et al. 2000) (BNID 105298) then each sheet is  $\sim 200 \text{ nm}$  away from its neighbours.



**Figure 7.2** Composition of Hill functions. Illustration of the effect on the rate of production of PA by the concentrations of Arf and PI(4,5)P<sub>2</sub>.

my purposes I want to ensure that the concentration of lipids is at least an order of magnitude higher than the protein concentration. So I will choose

$$[\alpha] = 10^4 \text{ m}^2 \text{ per m}^3 = 10^4 \text{ m}^{-1}.$$

This gives a total volume concentration of lipids of

$$3.3 \times 10^{-2} \text{ mol m}^{-3} = 33 \mu\text{M}.$$

If we assumed that the volume was filled with homogeneous sheets of lipid bilayers, then this would mean that the sheets are (on average) spaced approximately 200 nm apart. This is illustrated in figure 7.1B. It is therefore plausible that we could choose a still higher value for  $[\alpha]$  and retain a physically sensible system.

This derivation assumes that all lipids contribute to the cytosol-facing side of the membrane. This would be certainly true if the lipids naturally form micelles or monolayers. Alternatively, in practice (again, in biological situations and in experimental setups) we may need to halve the actual (experimental) concentration of lipids for use in the model.

### 7.2.3 Estimating maximal surface concentration of adsorbate

Finally, I need a sensible value for the maximum surface concentration of the adsorbate,  $n_0$ .

As in section 7.2.1 I will assume that molecules of Arf and PI4P5K occlude disc-shaped areas of the membrane of identical size. I will assume these discs have a radius of 18 Å,

$$\text{occluded area} = \pi \times 18^2 \approx 10^3 \text{ \AA}^2 = 10^{-17} \text{ m}^2.$$

The maximum density of circles on a surface is known to be (approximately) 0.9069 (Chang & Wang 2010). So once more using Avogadro's constant, an estimate of the maximum density of Arf and PI4P5K on the membrane surface is

$$n_0 = 0.9069 \times \frac{10^{17}}{6.022 \times 10^{23}} \approx 1.5 \times 10^{-7} \text{ mol m}^{-2} = 0.15 \text{ nM m}.$$

## 7.3 PLD model

Equation 5.18 described the final version of the 3-site mechanistic model for PLD. With new notation and including exponential decay, this becomes

$$\frac{d[\text{PA}]}{dt} = \frac{\lambda_{\text{cat}} [\text{PC}] [\text{A}] [\text{PIP}_2] \text{PL}_t}{(L_s + [\text{PC}])(L_a + [\text{A}])(L_r + [\text{PIP}_2])} - \lambda_{\text{decay}}[\text{PA}]. \quad (7.2)$$

Parameter	Value	Source	
$\lambda_{\text{cat}}$	$0.27 \text{ s}^{-1}$	Henage, Exton, & Brown (2006)	<sup>a</sup>
$L_s (K_m)$	400 nm	Vinggaard & Hansen (1995)	
$L_a$	10 nm	Henage, Exton, & Brown (2006)	<sup>b</sup>
$L_r$	3 nm	Henage, Exton, & Brown (2006)	<sup>c</sup>
$\lambda_{\text{decay}}$	$3.6 \times 10^{-4} \text{ nM}^{-1} \text{ s}^{-1}$	(see note)	<sup>d</sup>

**Table 7.2** Provisional parameters for the PLD model.

<sup>a</sup> As cited in the BRENDA database (Schomburg et al. 2013).

<sup>b</sup> The citation states that Arf1 activation of PLD is not saturated at  $10 \mu\text{M}$ ; plausibly from Figure 1C this is the EC50 (the concentration that gives half of the maximal response).

<sup>c</sup> The citation states that PLD achieves maximal velocity at  $\sim 5 \text{ mol}\%$  PI(4,5)P<sub>2</sub>; total lipid mass is  $117.6 \mu\text{M}$ ; and assuming EC50 is 2.5 mol% of total lipid mass =  $2.94 \mu\text{M}$ .

<sup>d</sup> Calculated so that steady-state concentrations agree with table 7.1, see section 7.3.1.

Parameter	Value	Source	
$\mu_{\text{cat}}$	$0.026 \text{ s}^{-1}$	Bazenet & Anderson (1992)	<sup>a</sup>
$M_s (K_m)$	1200 nm	Bazenet & Anderson (1992)	
$M_a$	19.7 nm	(arbitrary)	<sup>b</sup>
$M_r$	480 nm	(arbitrary)	<sup>c</sup>
$\mu_{\text{decay}}$	$2.1 \times 10^{-4} \text{ nM}^{-1} \text{ s}^{-1}$	(see note)	<sup>d</sup>
$\mu_{\text{exc}}$	$0.5 \text{ nM}^{-1}$	(arbitrary)	<sup>e</sup>
$\mu_{\text{rel}}$	10	(arbitrary)	<sup>e</sup>
$\mu_{\text{off}}$	$10^{-1} \text{ s}^{-1}$	(arbitrary)	<sup>f</sup>
$\mu_A$	$10^{-5} \text{ nM}$	(arbitrary)	<sup>f</sup>

**Table 7.3** Provisional parameters for the 3-site and 2-site PI4P5K models.

<sup>a</sup> Converted from specific activity of  $0.023 \mu\text{mol min}^{-1} \text{ mg}^{-1}$ , as cited in the BRENDA database (Schomburg et al. 2013); based on a molecular weight of 68 kDa (Ishihara et al. 1996).

<sup>b</sup> Based on 10% of the concentration listed in table 7.1.

<sup>c</sup> Based on 100% of the concentration listed in table 7.1.

<sup>d</sup> Calculated so that steady-state concentrations agree with table 7.1, see section 7.4.3.

<sup>e</sup> See section 7.4.6.

<sup>f</sup> The parameter  $\mu_A$  describes the ratio of the on rate to the off rate,  $\mu_{\text{off}}$ , for PI4P5K membrane-adsorption. Therefore,  $\mu_{\text{on}} = 10^5 \text{ nM s}^{-1}$ . (These parameters control both the rate of membrane-adsorption and the steady-state concentration of PI4P5K on the membrane.)

Alternatively, this model can be written as

$$\frac{d[\text{PA}]}{dt} = \lambda_{\text{cat}} \times \frac{[\text{PC}]}{L_s + [\text{PC}]} \times \frac{[\text{A}]}{L_a + [\text{A}]} \times \frac{[\text{PIP}_2]}{L_r + [\text{PIP}_2]} \times \text{PL}_t - \lambda_{\text{decay}}[\text{PA}].$$

which demonstrates that it is the composition of three Hill functions (see section 4.3): one for each of the substrate and the two regulators. Increasing any of the concentrations of PC, Arf, or PI(4,5)P<sub>2</sub> will increase the rate of production of PA up to the maximum,  $V_{\text{max}} = \lambda_{\text{cat}} \text{PL}_t$ . This is illustrated for Arf and PI(4,5)P<sub>2</sub> (with fixed PC) in figure 7.2.

### 7.3.1 Parameters for the PLD model

Provisional parameters allowing simulation of this system are listed in table 7.2.

The value for  $\lambda_{\text{decay}}$  was calculated so that the concentrations agreed with the concentrations listed in table 7.1 (assuming 10% Arf in the active form), using the following Python script:

```

1 from mechanistic import *
2 from scipy.integrate import odeint
3 from scipy.optimize import newton
4
5 def ss_ldecay(est, val): # log10(ldecay) estimate
6     # setup parameters, initial conditions, time
7     p = {kdecay:10**est,kcat:0.27,Ks:400,Ka:10,Kr:3}
8     ic = conclist({'PLDt':1.1,'PC':14260,'PIP2':25,'ARF':19.7})
9     t = np.linspace(0, 1e9, 100)
10    # simulate, get final values
11    final_vals = odeint(PLD_model, ic, t, (p,))[-1]
12
13    return condict(final_vals)['PA'] - val # divergence
14
15 # use Newton-Raphson method to estimate log10(kdecay)
16 #   initial estimate = 10^-5
17 #   desired [PA] = 480nM
18 ldecay_val = 10**newton(ss_ldecay, -5, args=(480,))
19 print("%.1e" % ldecay_val)

```

## 7.4 PI4P5K models

Recall that I have developed two contradictory mechanistic models for PI4P5K: one based on a 3-site assumption and the other based on a 2-site assumption.

### 7.4.1 PI4P5K 3-site model

Equations 5.28 and 5.29 described the final version of the 3-site mechanistic model for PI4P5K. With new notation, removing product inhibition, and including exponential decay, these become

$$\frac{d[\text{PIP}_2]}{dt} = \frac{\mu_{\text{cat}}[\text{PIP}] (\text{PK}_t - [\text{PK}_c])}{M_s + [\text{PIP}]} - \mu_{\text{decay}}[\text{PIP}_2] \quad (7.3)$$

$$\frac{d[\text{PK}_c]}{dt} = \mu_{\text{off}} \left( \frac{M_s M_a M_r (\text{PK}_t - [\text{PK}_c])}{(M_s + [\text{PIP}]) (M_a + [\text{A}]) (M_r + [\text{PA}])} - \mu_A \Phi(\theta)[\alpha][\text{PK}_c] \right). \quad (7.4)$$

Thus two equations: the net rate production of PI(4,5)P<sub>2</sub>; and the net rate of de-adsorption of enzyme. Note that the equation for the rate of production of PI(4,5)P<sub>2</sub> does not directly rely on the concentrations of Arf or PA, but that the rate of de-adsorption does.

### 7.4.2 PI4P5K 2-site model

Equations 5.37 and 5.38 described the final version of the 3-site mechanistic model for PI4P5K. With new notation, removing product inhibition, and including exponential decay, these become

$$\frac{d[\text{PIP}_2]}{dt} = \frac{\mu_{\text{cat}}[\text{PIP}] (M_r (1 + \mu_{\text{exc}}[\text{PIP}]) + \mu_{\text{rel}}[\text{PA}])}{\chi} - \mu_{\text{decay}}[\text{PIP}_2] \quad (7.5)$$

$$\frac{d[\text{PK}_c]}{dt} = \mu_{\text{off}} \left( \frac{M_s M_a M_r (1 + \mu_{\text{exc}}[\text{PIP}])}{\chi (M_a + [\text{A}])} - \mu_A \Phi(\theta)[\alpha][\text{PK}_c] \right). \quad (7.6)$$

where

$$\chi = \frac{M_r (1 + \mu_{\text{exc}}[\text{PIP}]) (M_s + [\text{PIP}]) + (M_s + \mu_{\text{rel}}[\text{PIP}]) [\text{PA}]}{\text{PK}_t - [\text{PK}_c]}.$$

Again, two equations: the net rate of production of PI(4,5)P<sub>2</sub>; and the net rate of de-adsorption of enzyme. Note that the equation for the rate of production of PI(4,5)P<sub>2</sub> does not rely on the concentrations of Arf but does depend on the concentration of PA; the rate of de-adsorption depends on both. This demonstrates the fact that PA is involved within the substrate dynamics in this model.

### 7.4.3 Parameters for the PI4P5K models

Provisional estimates for the parameters for both PI4P5K models are listed in table 7.3. Parameters related to the adsorption process were discussed in section 7.2. A majority of values were unknown; as I can see no obvious method for their (indirect) determination plausible (arbitrary) values have been assumed.



The value for  $\mu_{\text{decay}}$  was calculated so that the concentrations agreed with the concentrations listed in table 7.1 (assuming 10% Arf in the active form), using the following Python script:

```

1 from mechanistic import *
2 from scipy.integrate import odeint
3 from scipy.optimize import newton
4
5 def ss_mdecay(est, val): # log10(mdecay) estimate
6     # setup parameters, initial conditions, time
7     p = {kdecay:10**est, kcat:0.026, Ks:1200,
8         koff:1e-1, kA:1e-5, Ka:19.7, Kr:480}
9     ic = conclist({'PIPKt':3.55, 'PIP':250, 'PA':480, 'ARF':19.7})
10    t = np.linspace(0, 1e6, 100)
11    # simulate, get final values
12    final_vals = odeint(PIPK_model3, ic, t, \
13                       (p, 0.15, 1e4, phi_langmuir))[-1]
14
15    return condict(final_vals)['PIP2'] - val # divergence
16
17 # use Newton-Raphson method to estimate log10(kdecay)
18 #   initial estimate = 10^-2
19 #   desired [PIP2] = 25nM
20 mdecay_val = 10**newton(ss_mdecay, -3, args=(25,))
21 print("%.1e" % mdecay_val)

```

The parameters unique to the 2-site model  $\mu_{\text{rel}}$  and  $\mu_{\text{exc}}$  are discussed in section 7.4.6.

#### 7.4.4 Equality of models when $[PA] = 0$

The models for PI4P5K both reduce to the same simplified model whenever  $[PA] = 0$ ,

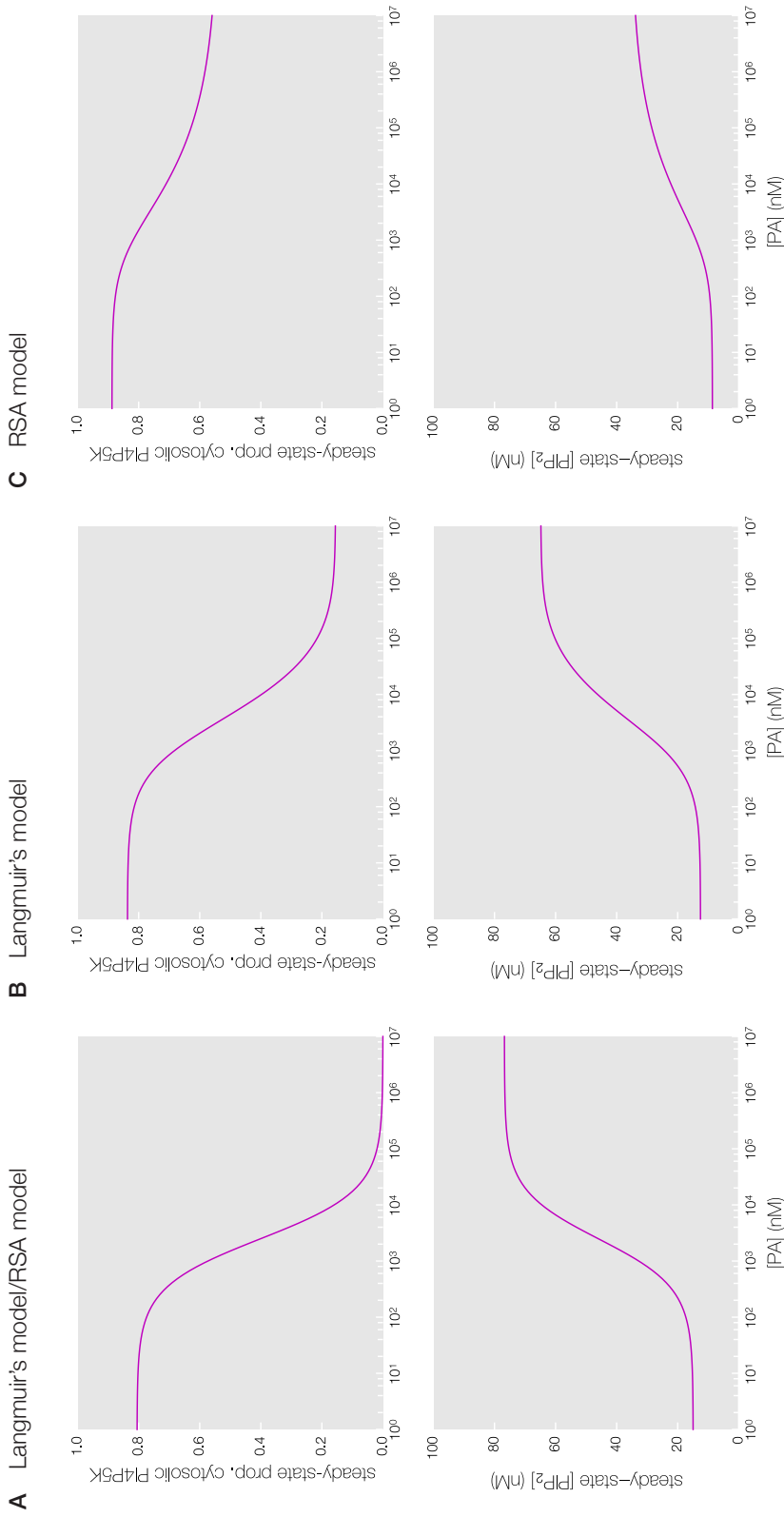
$$\frac{d[\text{PIP}_2]}{dt} = \frac{\mu_{\text{cat}}[\text{PIP}](\text{PK}_t - [\text{PK}_c])}{M_s + [\text{PIP}]} - \mu_{\text{decay}}[\text{PIP}_2]$$

$$\frac{d[\text{PK}_c]}{dt} = \mu_{\text{off}} \left( \frac{M_s M_a (\text{PK}_t - [\text{PK}_c])}{(M_s + [\text{PIP}])(M_a + [A])} - \mu_A \Phi(\theta)[\alpha][\text{PK}_c] \right).$$

This is not surprising as it removes the only distinction between the 2-site and 3-site assumptions for PI4P5K—the mechanism of regulation by PA.

#### 7.4.5 Steady-state behaviour of the 3-site model

Figure 7.3 shows the effect of varying  $[PA]$  on the steady-state proportion of cytosolic PI4P5K and PI(4,5)P<sub>2</sub> for the 3-site model with both adsorption models. With the value



**Figure 7.3** The effect of the concentration of PA on the steady-state proportion of cytosolic PI4P4K (concentration divided by the total concentration of PI4P5K) and the steady-state concentration of PI(4,5)P<sub>2</sub> for the 3-site model of PI4P5K. Note that the membrane-associated PI4P5K is 1 — the cytosolic PI4P5K. Parameters are as given in table 7.3, with other variables as in table 7.1 except [A] = 19.7.  
**A** With large maximum binding concentration ( $r_0 = 0.15$ , see section 7.2.3) both adsorption models give identical behaviours at steady-state.  
**B, C** With (artificially) small maximum binding concentration ( $r_0 = 0.003$ ) the models have distinct behaviours.

of  $n_0$  I estimated in section 7.2.3 (which is large in respect to the concentration of PI4P5K) Langmuir's model and the RSA model have identical responses to increases in the concentration of PA, see figure 7.3A. Reducing the value of  $n_0$  leads to a clearer distinction between the behaviours of the system under the different adsorption models, see figures 7.3B and 7.3C. Specifically, we can observe a reduction in PI4P5K binding under the RSA model due to jamming, see section 3.5.4.

For each of these systems, the response to changes in the concentration of PA is not very tightly regulated; the activation profile from minimum to maximum spans 2/3 orders of magnitude. Whether this is due to the parameterisation or inherent to the model structure is difficult to say, however this does suggest that PA activation of PI4P5K is graded and gradual, rather than switch-like.

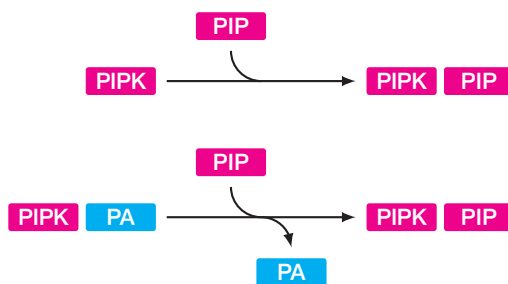
#### 7.4.6 PI4P5K as a 'competitive activator'

The 2-site model of PI4P5K has two additional parameters:

- $\mu_{\text{exc}}$ , the ratio of the exchange rate (PA  $\rightarrow$  PIP) against the off-rate for PA;
- $\mu_{\text{rel}}$ , the ratio of the exchange rate (PA  $\rightarrow$  PIP) against the on-rate for PIP.

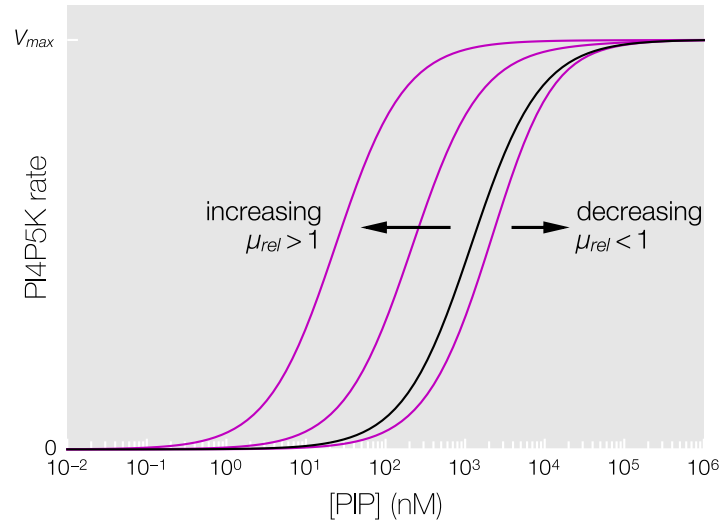
I have provided the first description of these parameters, and so it will not be possible to find existing values for these. Therefore I will need to suggest sensible values for both, based only on their qualitative effect on the dynamics of the regulation.

Recall that in the 2-site model there are two routes to producing the enzyme substrate complex:

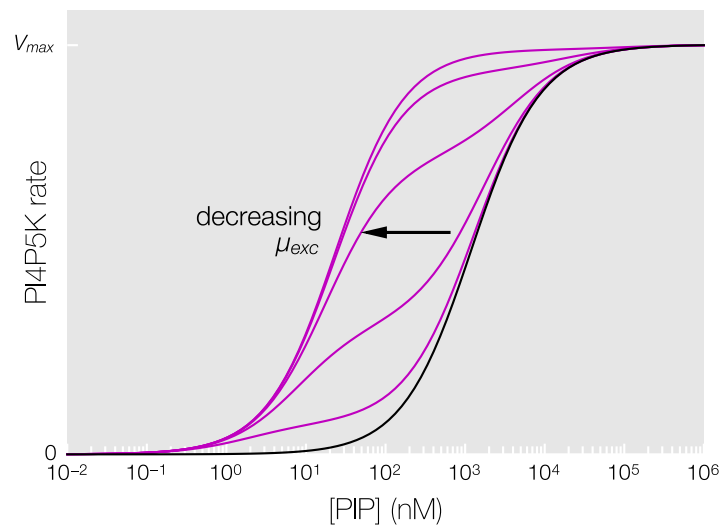


That is, a *direct* route and a *indirect* route (via a PA bound complex), respectively. By definition, if  $\mu_{\text{rel}} = 1$  then the rate of production of enzyme-substrate complex will be identical through both routes. If  $\mu_{\text{rel}} < 1$  then the rate for the indirect route will be slower, and so we would observe a form of competitive inhibition (a larger effective  $K_m$  with the same  $V_{\text{max}}$ ). If  $\mu_{\text{rel}} > 1$  then the rate for the indirect route

**A**  $\mu_{exc} = 10^5$ ; black line  $\mu_{rel} = 1$



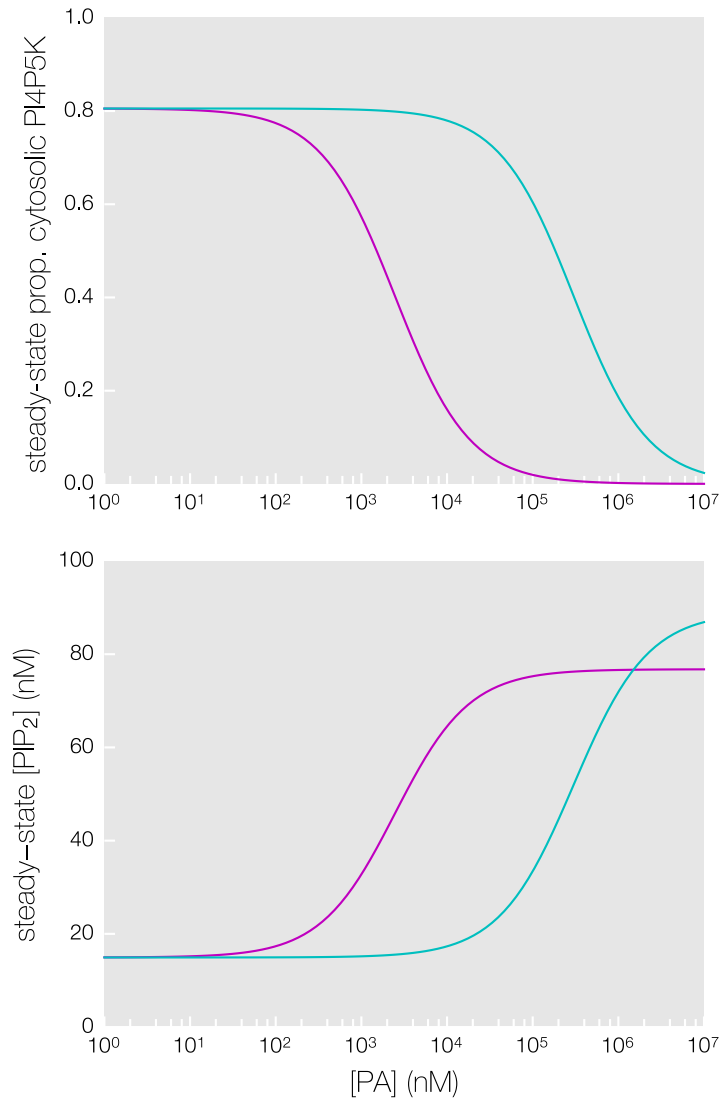
**B**  $\mu_{rel} = 100$ ; black line  $\mu_{exc} \gg 0$



**Figure 7.4** The effect of varying the parameters  $\mu_{rel}$  and  $\mu_{exc}$  on the rate of the 2-site PI4P5K model. PA concentration fixed, and substrate concentration varied. Other parameters are as given in table 7.3, and other variables as in table 7.1 except  $[A] = 19.7$ .

**A** If  $\mu_{rel} < 1$  then PA acts as a competitive inhibitor. Thus when  $\mu_{rel} > 1$  PA could be said to act as a competitive activator.

**B** Decreasing the value of  $\mu_{exc}$  increases the strength of the activation of PA in a nonlinear manner.



**Figure 7.5** The effect of the concentration of PA on the steady-state proportion of cytosolic PI4P4K (concentration divided by the total PI4P5K concentration) and the steady-state concentration of PI(4,5)P<sub>2</sub> for the 3-site model (magenta line; as in figure 7.3A) and the 2-site model (blue line) with  $k_{\text{exc}} = 0.5$  and  $k_{\text{rel}} = 1.2$ . Other parameters are as given in table 7.3, with other variables as in table 7.1 except  $[A] = 19.7$ .

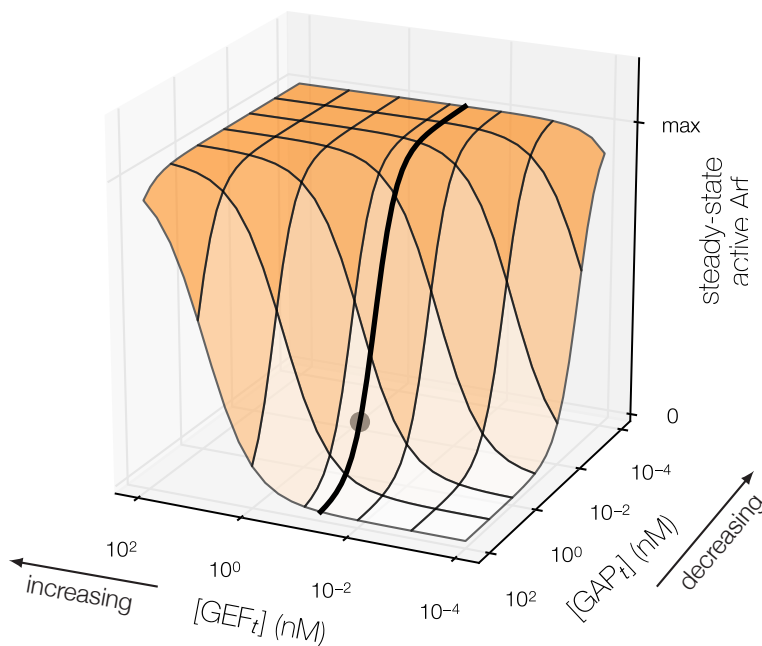
Parameter	Value	Source	
$k_{\text{cat}}$	$1.56 \times 10^4 \text{ nmol}^{-1} \text{ s}^{-4}$	Klebe et al. (1995)	<sup>a</sup>
$\kappa$	2.65	Klebe et al. (1995)	<sup>a</sup>
$K_0$	$2.38 \times 10^6 \text{ s}^{-3}$	Klebe et al. (1995)	<sup>a</sup>
$K_1$	$3.31 \times 10^3 \text{ nmol}^{-1} \text{ s}^{-3}$	Klebe et al. (1995)	<sup>a</sup>
$K_2$	$1.15 \times 10^3 \text{ nmol}^{-1} \text{ s}^{-3}$	Klebe et al. (1995)	<sup>a</sup>
$k_{\text{ase}}$	$8 \text{ s}^{-1}$	Ahmadian et al. (1997)	<sup>b</sup>
$K_M$	$5 \times 10^3 \text{ nM}$	Ahmadian et al. (1997)	<sup>b</sup>
$k_{\text{off}}$	$10^{-6} \text{ s}^{-1}$	(arbitrary)	
$k_A$	$10^{-5} \text{ nM}^{-1}$	(arbitrary)	<sup>c</sup>

**Table 7.4** Provisional parameters for the regulation of Arf model.

<sup>a</sup> Derived from kinetic parameters describing the Ran:RCC1 system, with concentrations of GTP and GDP as in table 7.1.

<sup>b</sup> Parameters for the the H-Ras:GAP-334 system.

<sup>c</sup> See section 7.5.1.



**Figure 7.6** The effect of varying the active concentrations of GEF and GAP on the steady-state concentration of active Arf. The thick line corresponds to the concentration of GEF listed in table 7.1; the dot corresponds to also the concentration of GAP listed in that table.

will be faster, and in direct analogy to the inhibition we could term this *competitive activation* (a smaller effective  $K_m$  with the same  $V_{\max}$ ), see figure 7.4A.

By definition, if  $\mu_{\text{exc}}$  is large, then the off-rate for PA will be much larger than the exchange rate. This gives an unstable enzyme-PA complex which decays into the component molecules (PA and PI4P5K) before the exchange process can occur. Therefore, we should expect  $\mu_{\text{exc}}$  to be (relatively) small, see figure 7.4B. In this figure, we can also observe a complicated response to changes in the concentration of PA for intermediate values of  $\mu_{\text{exc}}$  which is likely caused by substrate/regulator competition.

Together this means that—as we expect PA to be an activator of PI4P5K—we must have  $\mu_{\text{rel}} > 1$  and  $\mu_{\text{exc}}$  (relatively) small. With these constraints, we can observe an activation profile that is distinct from the activation profile of the 3-site model, see figure 7.5: an increase in the maximum steady-state concentration of PI(4,5)P<sub>2</sub>, but with an increase in the PA concentration for which the maximum is achieved.

## 7.5 Arf regulation model

Equations 6.28 and 6.29 describe a model for the dynamics of Arf regulation by GEFs, GAPs, and including the membrane-adsorption. With new notation and removing product inhibition from the GAP model, these become

$$\frac{d[A]}{dt} = \frac{k_{\text{cat}}(A_t - [A_c] - (\kappa + 1)[A])\text{GEF}_t}{K_0 + K_1(A_t - [A_c]) + K_2[A]} - \frac{k_{\text{ase}}[A]\text{GAP}_t}{K_m + [A]}. \quad (7.7)$$

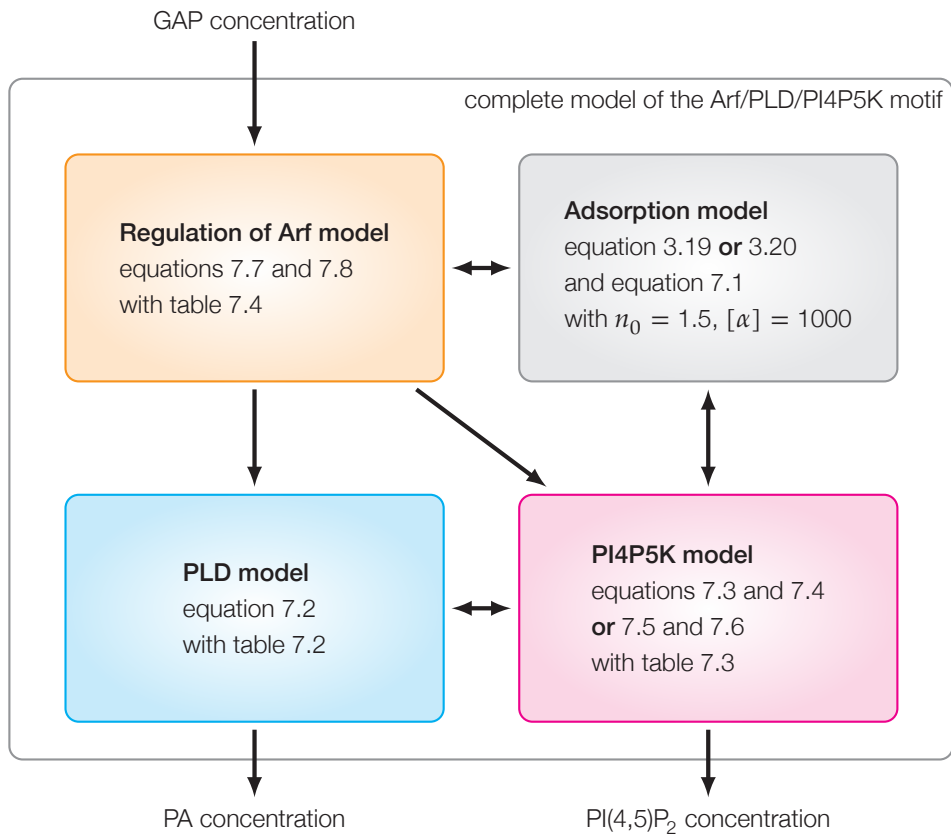
$$\frac{d[A_c]}{dt} = k_{\text{off}}(A_t - (1 + k_A\Phi(\theta)[\alpha])[A_c] - [A]) \quad (7.8)$$

Thus two equations: the rate of change of membrane-associated GTP-bound (active) Arf; and the rate of change of cytosolic GDP-bound (inactive) Arf. Note that the rate of change of membrane-associated GDP-bound (inactive) Arf,  $[A_m]$ , can be found from

$$\frac{d[A_m]}{dt} = -\left(\frac{d[A]}{dt} + \frac{d[A_c]}{dt}\right).$$

### 7.5.1 Parameters for the Arf regulation model

In chapter 6 I analysed the model of a generic G protein:GEF system (without membrane-adsorption) using the measured rates for the Ran:RCC1 system (Klebe et al. 1995), which I included in table 6.1. I know of no similar measurements for the kinetic rates of an Arf:Arf GEF system, so as Ran is also a monomeric G protein of



**Figure 7.7** Diagram showing the components of the complete model of the Arf/PLD/PI4P5K motif. Two choices remain: the adsorption model, Langmuir's model or RSA; and the PI4P5K model, the 3-site model or the 2-site model. Primary perturbation of the model should be through the concentration of GAP; primary read-outs are the concentrations of PA and PI(4,5)P<sub>2</sub>.



the Ras superfamily I will continue use these rates as (preliminary) values for Arf with a (yet unspecified) GEF. These rate parameters give the summary constants for the GEF component of the Arf regulation model.

Similarly, I have been unable to find kinetic measurements for any Arf:Arf-GAP system and so I will use values reported for the H-Ras:GAP-334 system (Ahmadian et al. 1997).

I will continue to assume that the GTP-bound myristoylated Arf is unable to leave the membrane surface. The association constant for myristoylated protein (derived from the model system of a small myristoylated peptide) has been reported as  $10^4 \text{ M}^{-1} = 10^{-5} \text{ nM}^{-1}$  (McLaughlin & Aderem 1995) (BNID 105722). I will use this number as a preliminary estimate for the association constant for the GDP-bound form of Arf (though of course in this case the myristoylation is ordinarily hidden).

In total, parameters allowing simulation of this model are listed in table 7.4. Parameters related to the adsorption process were discussed in section 7.2.

### 7.5.2 Perturbing the complete model via the GAP concentration

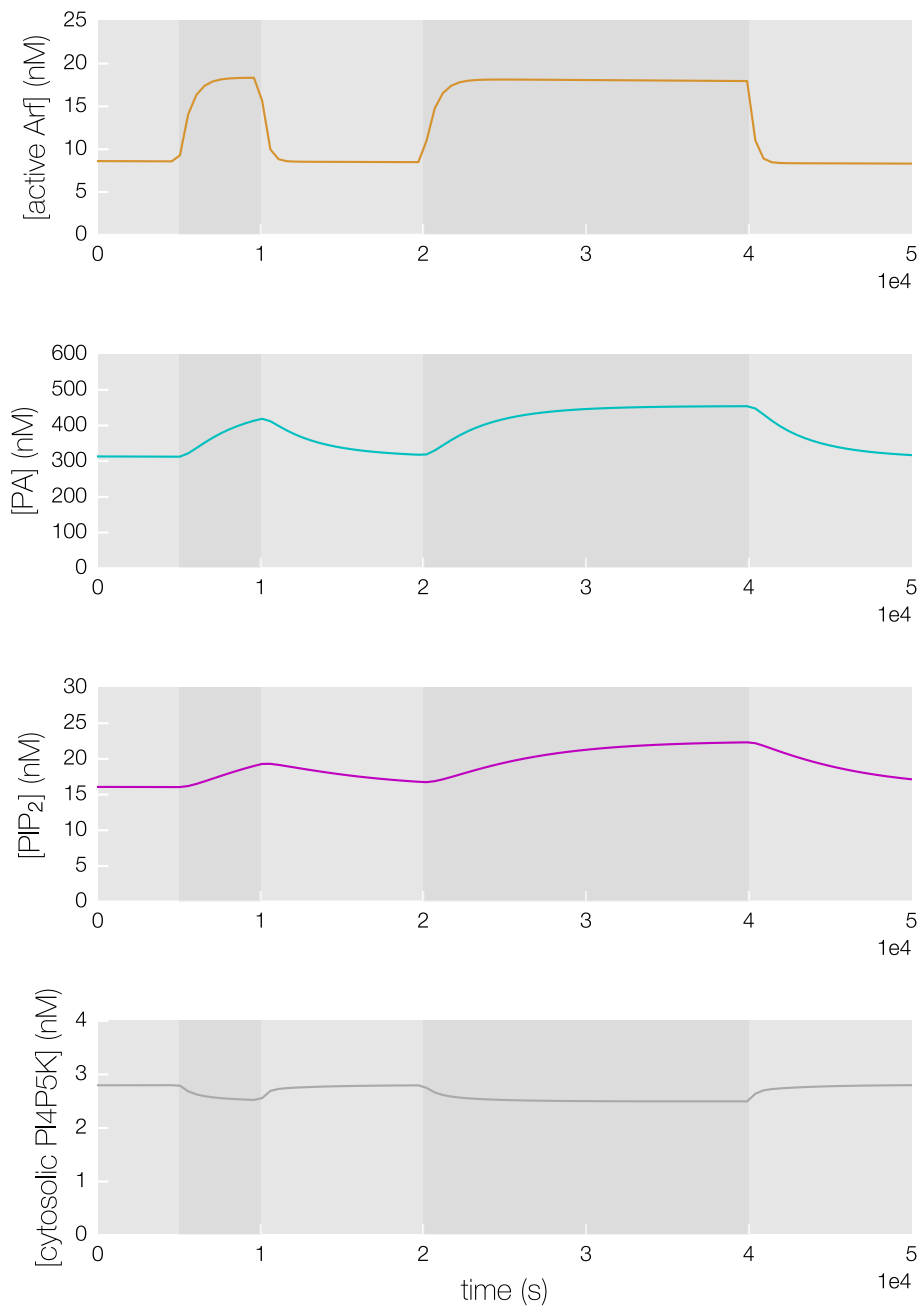
The effect of varying the concentrations of the regulators (GEF and GAP) on the steady-state concentration of active Arf is shown in figure 7.6. We observe the expected outcome: increasing the concentration of GEF or decreasing the concentration of GAP leads to an increase in the amount of active Arf.

As previously discussed in chapter 6, control of the activation of a G protein is more reliably performed via controlling the concentration of active GAP. Therefore, in the following simulations I will fix the concentration of GEF and vary the concentration of GAP. This will restrict the steady-state activation profile of Arf onto the thick line drawn in figure 7.6.

## 7.6 Arf/PLD/PI4P5K motif: complete models

We now have every component needed to construct a complete mechanistic model of the Arf/PLD/PI4P5K motif. Or rather—given the choice of PI4P5K model and adsorption model—we have the components to propose and construct a *family* of mechanistic models consisting of multiple versions of the model within the same structure. Figure 7.7 illustrates how the component models are connected in order to generate these complete models.

The component and complete models have been implemented in the Python programming language, included here in appendix C.



**Figure 7.8** Simulation of the complete mechanistic model of the Arf/PLD/PI4P5K motif, see figure 7.7 assuming the 3-site model of PI4P5K and adsorption follows Langmuir's model. The basal concentration of GAP was taken as 50% of the value in table 7.1. During the stimulation of the system (the darker shaded regions) this concentration was reduced to 10% of this value.

### 7.6.1 Preliminary analysis of the complete models

It only remains to demonstrate that simulating the complete model with the parameters that have been derived in this chapter gives physiologically plausible behaviours.

A simulation of the complete model with the 3-site PI4P5K model and Langmuir's model is shown in figure 7.8. Here we can see that perturbing the system by decreasing the concentration of active GAP produces the expected behaviour: an increase in active Arf, which in turn gives an increase in the concentrations of PA and PI(4,5)P<sub>2</sub>. Increasing the concentration of GAP restores the original state of the system.

Qualitatively similar behaviour was observed for the complete models with the 2-site PI4P5K model and/or the RSA model. These simulations have been included as figure C.1.

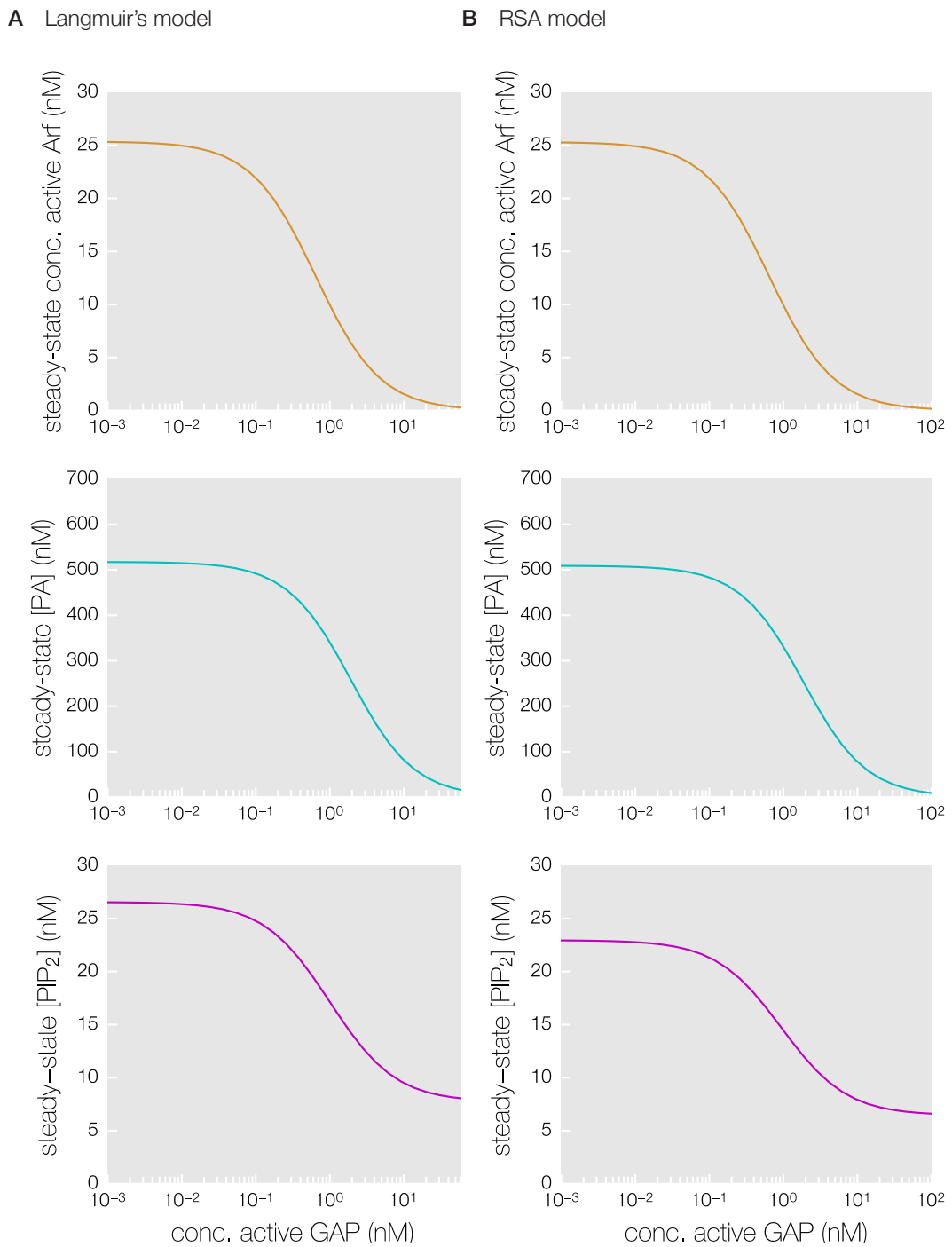
Alternatively, the behaviour of the system can be investigated through its steady-state dynamics. The effect of the concentration of GAP on the steady-state concentrations of active Arf, PA, and PI(4,5)P<sub>2</sub> are shown in figure 7.9 for the complete model with the 3-site model of PI4P5K. Here we observe a sigmoidal response, a nonlinear decrease in the activation state of the system (and so reduction in the PA and PI(4,5)P<sub>2</sub> concentrations) on increase of the GAP concentration. Again, the complete models with the 2-site PI4P5K models gives very qualitatively similar behaviours; these have been included as figure C.2.

These figures are sufficient to show that each of the versions of the complete model proposed for the Arf/PLD/PI4P5K motif are capable of behaviours that are consistent with the known biology.

## 7.7 Discussion

I have now described the first complete mechanistic model of the Arf/PLD/PI4P5K signalling motif, the construction of which is summarised by figure 7.7. These complete models are built from the component models of PLD and PI4P5K described in chapter 5, and the component model for the regulation of Arf described in chapter 6. Four versions of the mechanistic model are generated through two choices: the model for PI4P5K (2-site or 3-site); and the choice of adsorption model (Langmuir's model or the RSA model).

The models presented here are large and complex, and so unsuitable for direct mathematical analysis using the methods previously discussed. Instead I have determined a full set of (preliminary) parameters which is sufficient to allow the simulation of each of the component and complete models.



**Figure 7.9** Steady-state concentrations of active Arf, PA, and PI(4,5)P<sub>2</sub> as a function of the concentration of active GAP, assuming the 3-site model of PI4P5K.

As a first step towards validation, each version of the proposed model has been shown to be capable of displaying physiologically plausible behaviours – see figures 7.8, 7.9, C.1 and C.2. In every case we observe a system where a decrease in the concentration of GAP (the inactivator of the system) generates a corresponding increase in the concentration of the concentrations of PA and PI(4,5)P<sub>2</sub>. In the presented simulations, these four versions have near identical steady-state behaviours except for relatively small changes in magnitude. This means that, currently, each of the four versions are equally plausible for further investigation and, potentially, suitable for inclusion in larger models of signalling pathways.

Only one (or possibly none) of these models will accurately represent the actual underlying biochemistry of the Arf/PLD/PI4P5K motif. If suitable (experimental) kinetic data can be gathered or identified, Bayesian model selection could be used to distinguish which model is best supported by the data (Liepe et al. 2014; Toni et al. 2009). Bayesian model selection may be particularly applicable for the determination of the best model for the activation of PI4P5K by PA as I have now demonstrated that the 2-site and 3-site models can display distinct activation profiles, see figure 7.5.

In respect to the choice of adsorption process—we already know that the RSA model is more realistic (and more complex) than Langmuir’s model (recall section 3.5). However, the adsorption model does not appear to significantly alter the behaviour of PI4P5K adsorption when the maximal surface concentration ( $n_0$ ) is large. For the simulations of the complete models—while we can observe an effect of the choice of adsorption model on the magnitude of the response in the simulations shown (see figures 7.9 and C.2), the qualitative behaviour remains similar. However, while numerical simulation remains the primary method of analysis, there is little disadvantage in assuming the RSA model.

Equally important for the adsorption process is the choice of function for the fractional surface coverage ( $\theta$ ). Here, in order to make this function as simple as possible, I have assumed that a molecule of Arf and PI4P5K occlude identical area on the surface of the membrane. As per the analysis in section 7.2.1 it is likely that this assumption is not true, however a more realistic form for this function would require a much better understanding of the interaction of both Arf and PI4P5K with membranes, as well as a structure for PI4P5K to be determined.

Additionally, the current analysis ignores the effect of other membrane-associated proteins, including the possibility of crowding on the membrane surface. It also ignores any macro-scale effects of the organisation of lipids.

Together, these considerations suggest that the adsorption model should be care-

fully reconsidered during application of, or further investigation into, any of the proposed models.

Finally, it should be stressed that the arbitrary nature of many of the individual parameters suggested here (in tables 7.1 to 7.4) means that the complete set of parameters is currently only suitable for preliminary analysis and so unsuitable for predictive use. Future effort needs to be directed towards determining more physiologically-sound sets (or ranges) of parameters. It is also—so far—unclear as to the sensitivity of the system to most of these parameters, and whether any of these must be particularly tightly regulated in order for the system to display healthy behaviour. Analysis of the model in this way (e.g. via sensitivity analysis) is likely to lead to a better understanding of the underlying biology of the system.

To summarise, while there remain important considerations before they are ready for predictive work, I have constructed a novel mechanistic model of the Arf/PLD/PI4P5K signalling motif and have demonstrated, using a set of preliminary parameters, that each version of this model is capable of displaying physiologically-realistic behaviours.

## 8 Highlights & perspectives

For the full implications of many of the results presented within this thesis, I direct the reader to the previous discussion sections:

**Section 4.5** concerning initial, empirical models of the Arf/PLD/PI4P5K signalling motif.

**Section 5.6** concerning definition of plausible and complete molecular reaction schemes for the enzymes PLD and PI4P5K; and construction of quasi-steady-state mechanistic models.

**Section 6.7** concerning the derivation of the quasi-steady-state models for the regulation of G proteins by GEFs and GAPs; an argument towards the re-definition of this regulation in terms of a balance/imbalance mechanism; and modification of the regulatory models to be applicable to Arf family G proteins.

**Section 7.7** concerning parameterisation and further consideration of the PLD, PI4P5K, and regulation of Arf models, and construction of (multiple versions of) a complete mechanistic model of the Arf/PLD/PI4P5K signalling motif.

In this chapter: I will summarise the major results; briefly describe plausible experiments that could further the understanding of the Arf/PLD/PI4P5K motif; lay out possibilities for future research connected with the themes presented; and draw some final conclusions

### 8.1 Summary & highlights

The following subsections recap the major results, and discussion thereof, presented within this thesis.

### 8.1.1 Semi-automation of model construction using the Python module `enzymegraph`

To simplify and semi-automate the derivation of quasi-steady-state models of enzymes using the linear framework of Gunawardena (2012) in section 3.4 I introduced a new module for the Python programming language called `enzymegraph`. The source code is available in appendix A and online. It has been released under the MIT license, giving express permission for the dissemination and modification of the module by future researchers.

The module contains a direct implementation of the linear framework of Gunawardena (2012). It uses a modified form of the algorithm from Gabow & Myers (1978) to enumerate the rooted, directed spanning trees of a given graph. It utilises the Python module `sympy` to allow symbolic edge labels and algebraic manipulation.

An `enzymegraph` object can be created with an input graph—the graph on the enzyme complexes for an enzyme—in one of several forms. This object can be queried to return:

- the directed spanning trees of the graph (as `enzymegraph` objects);
- the basis element in terms of the edge labels;
- and, the quasi-steady-state concentrations of the enzyme complexes in terms of the edge labels.

Additionally, the `enzymegraph` object can output (the precursors to) the spanning-tree figures shown throughout this thesis (e.g. figure 5.7).

For an example script that uses the module see section 3.4.2. The `enzymegraph` module was used in the derivation of the models in chapters 5 and 6.

### 8.1.2 Construction of both empirical and mechanistic models of the Arf/PLD/PI4P5K signalling motif

Recall that in section 1.3 I defined two approaches for the construction of mathematical model of biochemical systems: empirical and mechanistic. Each has specific advantages and disadvantages.

Empirical models require less in-depth understanding of specific molecular interactions, and allow us to choose their mathematical description to model the observed qualitative behaviour of the system. Because we have this choice, we can often



choose forms that will give models that are simpler than the equivalent mechanistic model.

Mechanistic models require a well-defined molecular reaction scheme to be available, but can be less biased towards the modeller's expectations. Because of this they may more accurately predict emergent behaviours for a system, especially in extreme scenarios.

Here, I have constructed models of the Arf/PLD/PI4P5K signalling motif under both paradigms. Empirical models in chapter 4, and mechanistic models in chapters 5 to 7. Further discussion of these continues below.

### 8.1.3 Comparison with theoretical systems

Throughout the previous discussion I have frequently noted the differences in the catalytic and regulatory mechanisms for PLD and PI4P5K, this leads to an asymmetry in the Arf/PLD/PI4P5K motif that may not initially be apparent in illustrations (e.g. figure 2.8).

Symmetric *theoretical* motifs in analogy to the Arf/PLD/PI4P5K motif were formed by mirroring the behaviour of each enzyme. This approach highlighted behaviours that were shared with the real motif; implying that the Arf/PLD/PI4P5K motif has characteristics and advantages of both theoretical motifs.

This approach was used for the empirical models analysed in chapter 4. It would also be possible to replicate the approach for the complete mechanistic models in chapter 7 by again constructing models of the theoretical motifs using two copies of either PLD or PI4P5K mechanistic model.

### 8.1.4 Empirical models of the Arf/PLD/PI4P5K signalling motif

In chapter 4 I have described three families of empirical models:

**Pseudo-mass-action models** (sections 4.1 to 4.2) based on a previous model of the Arf/PLD/PI4P5K signalling motif (Stanley 2011). These models were capable of displaying unbounded growth, and as such were deemed unrealistic.

**Hill models** (section 4.3) which included Hill functions in order to bound the behaviour of the systems.

**Conservation of mass models** (section 4.4) which included conservation of mass on one or both of the enzymes in order to bound the behaviour of the systems.

Much of the mathematical tractability of the pseudo-mass-action model was lost when the complexity was increased through inclusion of Hill functions or conservation of mass. However, further qualitative results were obtained by taking additional assumptions on rate constants, and considering numerical results.

### 8.1.5 Empirical models: hypothesised mechanism for the Arf/PLD/PI4P5K signalling motif

The results derived from the Hill models and the results derived from the conservation of mass models both suggested a mechanism of signal propagation for the Arf/PLD/PI4P5K motif. In summary, this analysis suggested that the system is able to produce two distinct output signals (the concentration of PA and  $\text{PI}(4,5)\text{P}_2$ ) from a single initial input signal (the concentration of active Arf). Based on the results for non-extreme parameterisations, I hypothesise that:

- PI4P5K acts to produce  $\text{PI}(4,5)\text{P}_2$  so as to replicate the original signal with little to no modulation.
- PLD acts to produce PA only when there is a sufficient concentration of Arf (above a threshold value). This also means that PLD could act to reduce low levels of noise occurring in the upstream signal.

So when there is no active Arf there is no activation of either enzyme; at low concentrations of Arf only PI4P5K is activated; and at high concentrations of Arf both enzymes are activated.

The division of one signal into two signals means that downstream processes could act in response to either one or both of these signals. This provides evidence that the ‘simple’ Arf/PLD/PI4P5K motif is capable of complex behaviours.

### 8.1.6 Mechanistic models of PLD & PI4P5K

In chapter 5 sensible and complete (well-defined) molecular reaction schemes were proposed for PLD and PI4P5K based on a mixture of evidence from published literature, and inference and extrapolation (see section 8.1.7 below). These reaction schemes were used to derive three mechanistic models: the 3-site model for PLD; the 3-site model for PI4P5K; and the 2-site model for PI4P5K. PI4P5K was modelled in both instances as a surface-active enzyme using the approach of Kartal & Ebenhöf (2013) described in section 3.5.

The 3-site hypothesis describes an enzyme that has independent binding sites for substrate and two regulators. In this case, equivalent mechanisms for binding to the allosteric regulations implies a hidden equivalent regulatory mechanism for the two enzymes.

The 2-site hypothesis was based on conjecture by Stace et al. (2008) that the binding sites for PI4P and PA on PI4P5K overlap. Whether the 2-site model is a more accurate representation is unclear.

### 8.1.7 A lack of well-defined biochemistry

In chapter 5 the determination of molecular reaction schemes for PLD and PI4P5K was impeded due to a lack of information regarding the specifics of the molecular interactions of their catalytic and regulatory mechanisms. The existing body of research is insufficient for immediately and uniquely determining suitable molecular reaction schemes. This highlights a major requirement for further investigation into these biochemical processes.

Amongst other factors, more evidence is needed to determine or reject: a specific ordering of interactions between the enzymes and each regulator; the existence of product inhibition; the independence of binding sites; the importance of multiple PA binding sites on PI4P5K; the mechanism of activation of PI4P5K by Arf (membrane recruitment and/or allosteric activation).

Here, sufficient (well-defined) descriptions of the molecular interactions of PLD and PI4P5K were required in order to progress to (candidate) mathematical models. Therefore, existing evidence had to be supplemented by additional biochemical assumptions, each of which was noted and described in the text.

### 8.1.8 Complexity of mechanistic ( $n$ -site) models

A measure of the complexity of the catalytic and regulatory mechanisms of an enzyme is given by the number of spanning trees of its graph on the enzyme complexes. The proposed 2-site and 3-site mechanisms for PLD and PI4P5K lead to graphs with spanning trees numbering in the thousands or millions. These spanning trees lead directly to the basis element and the quasi-steady-state solutions, which will therefore be very long and complicated functions of the parameters, which would be highly impractical to manipulate. This large number of spanning trees is caused by the large number of vertices and (reversible) edges between these vertices. Therefore, one reason to attempt further experimental investigation of the two enzymes would

be to determine whether there is any restriction on the order of binding of the regulators.

So in order to construct models for each of the proposed mechanisms I could not simply (naïvely) apply the linear framework of Gunawardena (2012), but instead I had to take a modified approach. Specifically, I chose to separate the dynamics related to regulation and dynamics related to catalytic activity. This used the assumption that the catalytic dynamics act much faster than the regulatory dynamics, and so I could use an argument in respects to time-scale separation (see section 3.2.3). This vastly reduced the complexity of the system, and allowed the construction of the mechanistic models.

### 8.1.9 Membrane-recruitment of PI4P5K by Arf

Both of the 2-site and 3-site mechanistic models of PI4P5K are described by two equations: one for the rate of production of PI(4,5)P<sub>2</sub> by membrane-associated PI4P5K; and another for the amount of cytosolic PI4P5K.

Note that for the 3-site model both Arf and PA are allosteric regulators; for the 2-site model only Arf is an allosteric regulator. In the final version of both models I have assumed no allosteric activation of PI4P5K (catalytic rates are independent of whether the allosteric regulators are bound). In this case, the equation for the rate of production of product loses all dependence on the concentrations of the allosteric regulators. Then, the concentrations of the allosteric regulators only determine the amount of enzyme on the membrane. This model describes the hypothesis that membrane-recruitment is the sole mechanism of PI4P5K activation by Arf.

### 8.1.10 Mechanistic model of Arf regulation by GEFs and GAPs

Based on a model for the regulation of a generic G protein by guanine nucleotide exchange catalysed by a GEF, and GTPase activity catalysed by a GAP, a mechanistic model of Arf regulation was derived in section 6.6. This includes the differential localisation of the active and inactive forms of Arf modelled as a cytosol-membrane transition using the approach of Kartal & Ebenhöf (2013) discussed in section 3.5.

### 8.1.11 Regulation of G proteins by GEFs and GTPase activity is via a balance/imbalance mechanism

During the construction of the model of the regulation of Arf in chapter 6, I first constructed and analysed models of the regulation of generic G proteins by guanine nucleotide exchange catalysed by GEFs and GTPase activity. The results of this analysis have implications for much of the current understanding of G proteins.

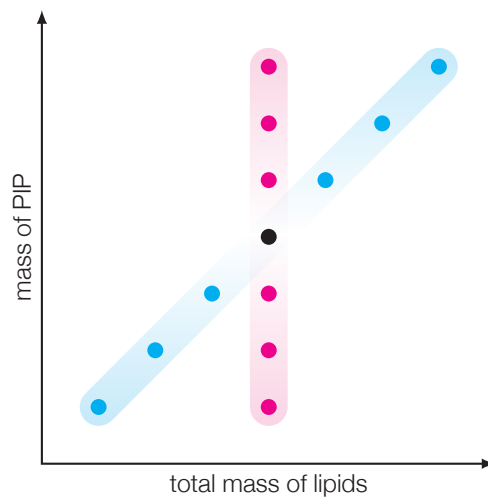
I have determined that the fully reversible mechanism of GEFs (and GPCRs) means that they act to produce a balance of active and inactive forms of the substrate G protein. Only when the system is imbalanced towards an inactive state by the presence of GTPase activity does stimulation of guanine nucleotide exchange result in an observable activation event. In situations where the system is already in balance then—in a manner equivalent with product inhibition—on stimulation of guanine nucleotide exchange the amount of available active G protein will decrease. This places GTPase activity as the key regulator of G protein signalling as it allows a full, controlled, spectrum of signals to propagate downstream of the G protein. In this way, I conclude that G protein:GEF:GAP systems are best described as controlled via a balance/imbalance mechanism.

These results have implications in many systems where the perturbation of G proteins has been previously studied. Over-expression of the ‘activator’ could possibly lead to unexpected results, these systems should be preferentially controlled through regulation in the GTPase activity. This does, however, suggest a novel route for the negative control of G proteins which have become GTPase activity-insensitive, as is the case for mutations in the G protein Ras which are common in cancers (Stephen et al. 2014).

### 8.1.12 Mechanistic models of the Arf/PLD/PI4P5K signalling motif

The individual mechanistic models for PLD, PI4P5K, and the regulation of Arf were combined to form a complete mechanistic model of the Arf/PLD/PI4P5K signalling motif in chapter 7. As I had previously derived two distinct models of PI4P5K (2-site or 3-site) and suggested two different adsorption models (Langmuir’s model or the RSA model), this led to four distinct versions of the model being proposed. While it remains uncertain which version most accurately describes the actual biochemical system, each of these was shown to display plausibly physiologically realistic behaviour when simulated using a series of preliminary parameters.

These models constitute the first known mechanistic (derived from individual



**Figure 8.1** Two approaches for the production of lipid vesicles with varying concentrations of PIP. Either (diagonal, blue line) the PIP:total lipid mass ratio can be fixed and the total lipid mass varied. Or (vertical, magenta line) the total lipid mass can be fixed and the PIP:total lipid mass ratio varied.

molecular interactions) models of the Arf/PLD/PI4P5K signalling motif, and as such are now suitable for inclusion into the future models of larger signalling processes.

## 8.2 Experimental approaches to validation

I have identified the following experimental approaches for the further investigation into the Arf/PLD/PI4P5K signalling motif and validation of some of the theoretical results presented within this thesis. This section is not exhaustive; further proposals for sensible experiments would result from additional investigation into each of the models.

In addition to the following suggestions—and, as mentioned previously—there remains a major requirement for further fundamental biochemical research into the structure, function, and regulation of the enzymes PI4P5K and PLD.

### 8.2.1 Biochemical assays

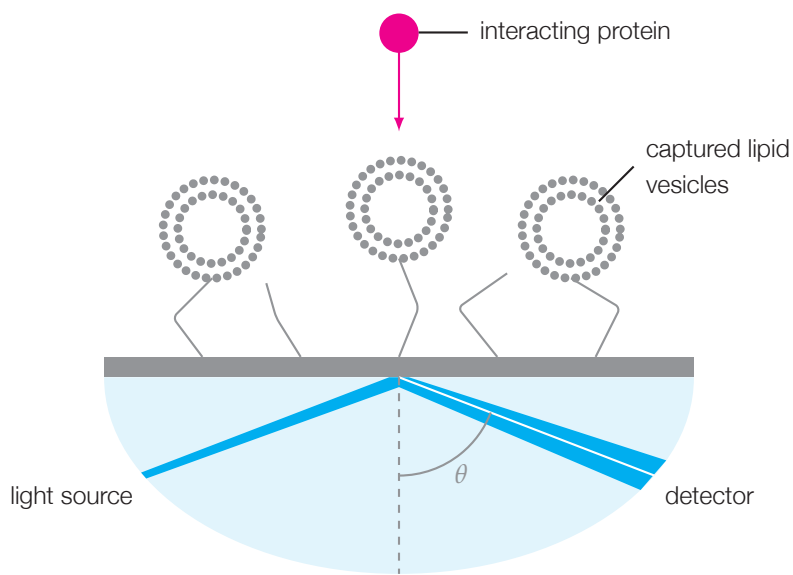
In chapter 5 I developed two contradictory models for the catalysis and regulation of PI4P5K. Only one of these (or neither) will accurately represent the biochemistry. With sufficient data it becomes possible to use Bayesian approaches to determine the model that is most plausible (given the data) (Toni et al. 2009).

Suitable data would be in the form of dose-response curves obtained (for example) from *in vitro* kinase assays, by varying the concentrations of substrate and regulators. Measurement can be made by a radioactive signal: ATP acts as the phosphate donor for the phosphorylation of PIP; introducing [ $\gamma$ - $^{32}\text{P}$ ]ATP (ATP with a radioactive phosphate group) means we can radioactively tag the resulting PI(4,5)P<sub>2</sub>.

Vesicles can be constructed from PI4P, PA, and (inert, in this context) PC. To these vesicles, Arf will associate via incubation with Gpp(NH)p (a non-hydrolysable analogue of GTP). The proportion of Arf that binds to the vesicles can be quantified by measuring the amount of Arf remaining in solution (e.g. using a dot blot assay).

A reaction is completed by addition of PI4P5K and [ $\gamma$ - $^{32}\text{P}$ ]ATP. Thin-layer chromatography (TLC) can be used to separate the radioactive phospholipids from the bulk solution, the radioactive signal imaged on a phosphorimager, and quantified. The amount of PI4P5K remaining in the cytosol could also be quantified at this stage.

Note that the concentration of the substrate PI4P can be varied in two ways, illustrated in figure 8.1:



**Figure 8.2** Outline of surface plasmon resonance using sensor chip with captured lipid vesicles. Molecules interacting with a structure on a sensor chip causes the refractive index of the surface of the chip to change.



- By fixing the PIP:total lipid mass ratio and varying the total lipid mass. This requires only a single preparation of vesicles.
- By fixing the total lipid mass, and varying the PIP:total lipid mass ratio.

These methods will lead to different dose-response curves for the activity of PI4P5K: the latter will have an (approximately) fixed surface area per unit volume, whereas for the former this will be variable. Data from both approaches is valuable, as this difference only determines whether  $\alpha$ —the variable denoting surface area per unit volume—should be fixed or variable.

The full or component models presented here could also be usefully compared with time-series data, if suitable data can be gathered. This would let us directly compare with model simulations (such as figure 7.8) so as to perform Bayesian parameter estimation and model selection (Toni et al. 2009; Liepe et al. 2010).

### 8.2.2 G protein assays using fluorescent GTP

Data for the parameterisation of the models for the regulation of Arf (and other G proteins) given a specific GEF and GAP is often obtained using measurement of the binding of fluorescent analogues of GTP and GDP: mantGTP and mantGDP (mantGXP) (Rojas et al. 2003). On binding to the G protein, mantGXP undergoes an increase in fluorescence intensity. So using a fluorometer it is possible to record a time-course of the effects of controlling the G protein, GEF, GAP, and nucleotide concentrations. Using this approach it is possible to quantify the transient and long-term dynamics of interactions in a G protein:GEF:GAP system.

Specifically, in relation to the work presented in this thesis, these nucleotides could be used to investigate the strength of the inhibitory effect of over-expression of GEFs (see section 6.4.2). This would further validate our understanding of this regulatory system, and allow us to determine whether this effect is likely at physiological concentrations of the species.

However, mantGXPs have been shown to affect the kinetics of the system in an unpredictable manner (Mazhab-Jafari et al. 2010). Alternatively, other approaches for the measurement of the kinetics of G protein:GEF:GAP systems are possible, such as real-time NMR (Marshall et al. 2009).

### 8.2.3 Surface-adsorption assays using surface plasmon resonance

Membrane-adsorption was shown to be an important regulatory step in the activation of Arf and PI4P5K. These effects could be quantified using *surface plasmon resonance* (Beseničar et al. 2006). Molecules interacting with a structure on the surface of a prepared sensor chip cause its refractive index to vary; this is measured by a decrease in the intensity of reflected light hitting a detector at a critical angle. In this manner, binding processes can be detected.

For the purpose of lipid-binding assays, it is possible to coat a chip with either a lipid monolayers or with a structure which can bind lipid vesicles, see figure 8.2.

## 8.3 Proposed extensions

I have identified the following areas as suitable for potential further investigation regarding the theoretical understanding of the Arf/PLD/PI4P5K signalling motif.

### 8.3.1 Feedback of phospholipids on GAPs and GEFs

As discussed in section 2.7.1 there exists evidence for feedback of phospholipids onto Arf-GEFs and Arf-GAPs. It is plausible that this regulation occurs via membrane-recruitment effects, as Arf is known to be activated and inactivated on the membrane surface. Modelling these regulations will require a thorough understanding of the underlying molecular interactions.

Ultimately, inclusion of these interactions will increase the complexity of the system and so, perhaps, lead to distinct, emergent behaviours. Therefore, predicting the effect of these feedbacks without mathematical or computational analysis would be difficult and somewhat inappropriate. However, I would conjecture that positive feedback on the activation of Arf activation will broadly act to reinforce a signal, and negative feedback to diminish a signal.

### 8.3.2 Inclusion of ATP/ADP dynamics in kinase models

The mechanistic reaction schemes developed for PI4P5K have so far ignored the crucial steps of ATP-binding, and ADP-unbinding. (This can be compared to the reaction scheme used for the GEF model which did include GDP and GTP binding and unbinding as individual kinetic steps.) We know that these events must occur at some stage of the catalysis for the conversion of each molecule of PI4P to PI(4,5)P<sub>2</sub>, as ATP is the provider of the additional phosphate group.

Commonly, ATP/ADP dynamics are ignored for models of kinases as we assume that ATP is well-buffered in the cytosol (otherwise the cell will be experiencing other far more deleterious effects). However, there is some evidence that these dynamics are important, and that crucially fluctuations in the concentrations of ATP may have consequences to signalling (Nirody & Rangamani 2014).

### 8.3.3 $n$ -site models

It would be possible to directly extend the construction of the 2-site and 3-site models to  $n$ -site models, using similar time-scale separation arguments to simplify the analysis whenever necessary.

This would be one approach to modelling the multiple, independent and allosteric binding sites that have been hypothesised for PA on PI4P5K (Stace et al. 2008).

Additionally, multi-site models of phospholipid binding have some analogy with multi-site phosphorylation models, which have been shown to be capable of complex behaviours (Manrai & Gunawardena 2008; Thomson & Gunawardena 2009).

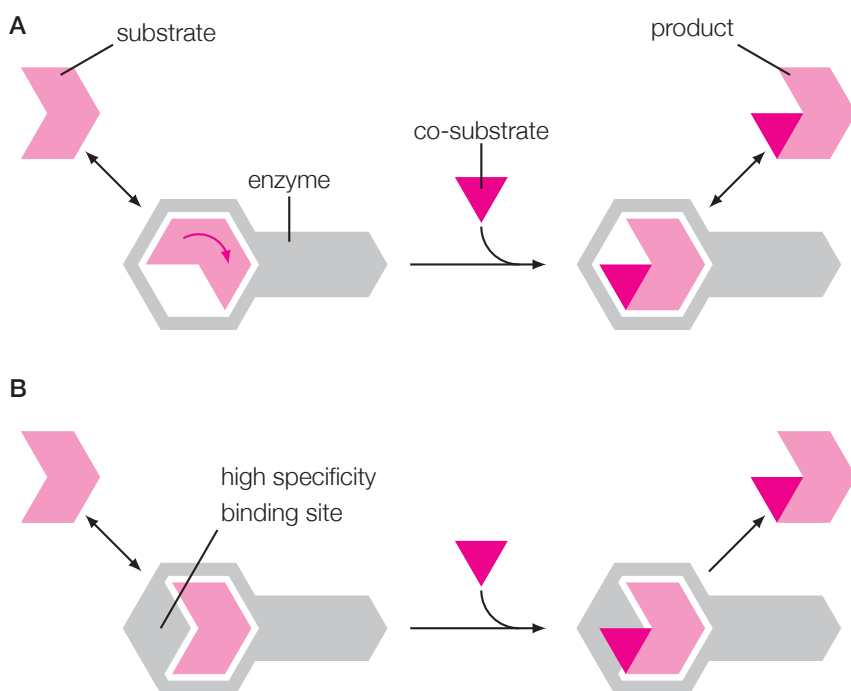
## 8.4 Perspectives

The results presented in this thesis have led me towards a series of ideas (or perspectives) for areas of potential future research that could potentially lead to greater understanding of intracellular signalling processes. These ideas are presented here in sketch form only.

### 8.4.1 Theoretical & synthetic biology

I have used theoretical motifs alongside the real Arf/PLD/PI4P5K in an attempt to better understand the effect on signal propagation of the combination of regulatory mechanism of the enzymes and cross-talk of their products. These theoretical motifs were designed as to be similar in scale and form to the original system. Here, analysis of the models of these motifs has helped in the understanding of the original system. I propose that this approach is useful and generally transferable to other systems.

These theoretical systems could be considered to describe unhealthy, diseased, or mutated states arising from the original system. Therefore their analysis could be considered interesting in a therapeutic sense.



**Figure 8.3** Demonstrating a possible trade-off between specificity to substrate orientation and product-inhibition, geometrically.

**A** A mechanism displaying product-inhibition, but where the enzyme has a low specificity to substrate orientation. Thus a high on-rate of substrate.

**B** A mechanism not displaying product-inhibition, but where the enzyme has a high specificity to substrate orientation. Thus a low on-rate of substrate.

Also, this approach has links to the field of synthetic biology – in which we could hope to actually construct these motifs and so investigate their behaviours experimentally.

#### 8.4.2 Evolutionary constraints on enzyme kinetics, product inhibition

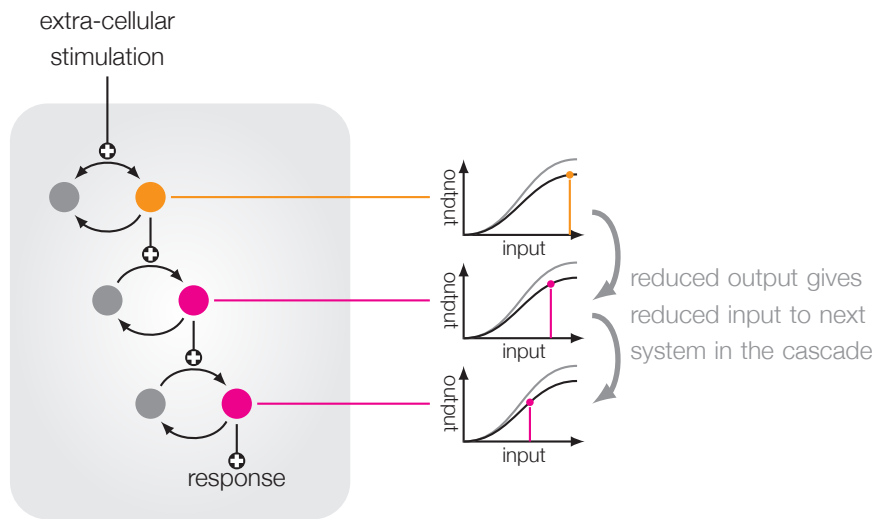
Some enzymes display product inhibition, while others do not. Here evidence suggested that PI4P5K does have product inhibition, whereas the evidence for PLD was less certain (ultimately I assumed it did not). This leads to the question: is product inhibition due to a high structural similarity between substrate and product; or is product inhibition a mechanism which has some role in signalling, for example limiting the maximum concentration of the product?

I hypothesise the following. Suppose that an enzyme would evolve to maximise a rate of production of product (in the absence/presence of appropriate regulators). Then if the product and substrate are structurally similar, product inhibition may imply a limitation to the maximum rate. One solution would be for the enzyme to evolve a more specific binding site, one that would bind the substrate, but not the product. However, by argument to geometry—see figure 8.3—this might further restrict the optimal substrate orientations for binding, so reduce the on-rate, and so limit the maximum catalytic rate. An optimum balance between specificity to substrate orientation versus product inhibition would then depend on the roles and requirements for a specific enzyme.

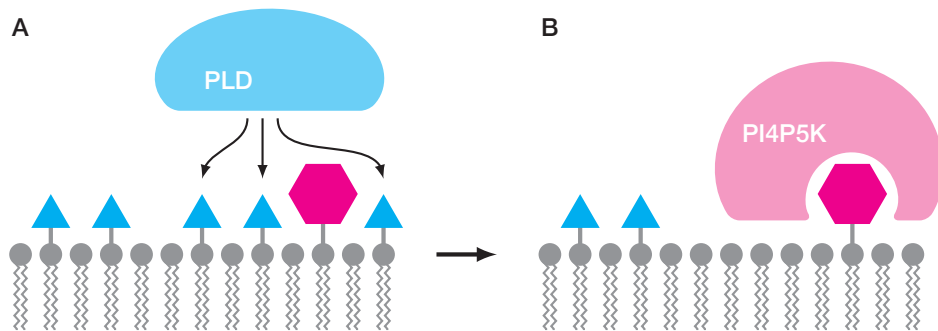
Complicating matters is that (models of) systems with product inhibition can have fundamentally distinct behaviour to (models of) systems without product inhibition, for example see Ortega et al. (2002). Particularly, product inhibition has been shown to have an important role in transferring information about the concentration of the end products of a (metabolic) system upstream, and as such assisting the system in reaching steady-state (Fell & Sauro 1985; Cornish-Bowden & Cárdenas 2001).

#### 8.4.3 Perturbing intracellular signalling through nucleotides

During analysis of the model for G protein regulation, I demonstrated that GEFs act to obtain a balance between active and inactive forms of the G protein; furthermore the specific balance obtained strongly depends upon the GTP/GDP ratio. So fluctuations in the concentration of GTP and/or GDP will affect the maximum possible ratio



**Figure 8.4** Proposed effect of altering nucleotide levels for a cascade of G protein and kinase signalling. A reduction in ATP and GTP levels would diminish the maximum concentration (subgraphs; grey line, reduced to black line) of activate G protein or phosphorylated protein in individual systems. If multiple G protein or kinase systems form a tightly tuned signalling cascade, then it is plausible that a reduction in the activation of the first system will further reduce the activation of the next system (subgraphs; dots, vertical lines). In this way, the reduction in signal could further compound through the pathway (an amplification of downregulations) and lead to a severely diminished response to the initial stimulation.



**Figure 8.5** Proposed mechanism for which the choline-headgroup of PC is an inhibitors of PI4P5K activity by restricting access of the kinase domain of PI4P5K to the PIP.

**A** Activated PLD cleaves the choline headgroup from PC (blue triangles), for a region of the membrane.

**B** This allows PI4P5K to locate and bind to its substrate, PIP (magenta hexagon).

of active to inactive form of the G protein, and so affect the maximum strength of activation.

We also know that ATP concentrations will affect the maximum rate of kinases (see section 8.3.2). Guanine and adenosine nucleotides are interrelated by nucleoside-diphosphate kinases (NDKs) which catalyse the reaction



Throughout the cell there exist signalling cascades that make use of coupled G proteins and kinases, such as the Ras-Raf-MEK-ERK pathway (Kolch 2000). In such situations, it is plausible that altered nucleotide levels could cause a reduction in signal strength, compounded through the signalling pathway. This is illustrated in figure 8.4.

The strength of any inhibitory effect would depend on the stimulation required for activation of the downstream elements, and the nonlinearity of this response. Additional modelling should inform us of the likelihood or importance of this effect.

It is known that modifying the GTP concentration can control the cellular behaviours including apoptosis and differentiation (Meshkini, Yazdanparast, & Nouri 2011).

#### 8.4.4 Topology of membranes

Lipid membranes are not flat, homogeneous surfaces. Size, shape, and charge differences for individual lipids will cause regions of intracellular membranes to have distinct properties. Furthermore, proteins (such as PLD and PI4P5K) are able to bind to and change the composition of the membranes. Some of these temporal changes may have fundamental importance in the processes inherent to the Arf/PLD/PI4P5K signalling motif.

It is known that the phospholipids PC, PA, and the PIs have distinct three-dimensional shapes, and are so able to affect the curvature and topology of the membranes they are part of (McMahon & Gallop 2005). Changing the topology may also change the surface area. Different phospholipids have different charges, and so attract differently charged regions on proteins. It is plausible that the proteins considered here are affected by these effects and so further investigation into the static and dynamic physical membrane-protein effects may lead to crucial insights into the regulation of the system.

Indeed, observe from figure 2.4 that phosphatidylcholine has a much larger head-group than phosphatidic acid. Therefore, I hypothesize that a plausible mechanism for

the membrane recruitment of PI4P5K (and perhaps part of the apparent activation by PA) is the removal of choline groups from the membrane surface by the action of PLD. In this way, perhaps a region of the membrane is cleared of choline head-groups allowing PI4P5K to get closer to its substrate. This is illustrated in figure 8.5.

## 8.5 Conclusion

This thesis has documented my progress towards a better theoretical understanding of the mechanisms and behaviours of the Arf/PLD/PI4P5K signalling motif, formed by the activation of the lipid modifying enzymes PLD and PI4P5K by Arf, and the cross-regulation of their products. This system is of fundamental importance in lipid signalling: it controls the concentrations of lipids which are important in many intracellular processes; and it is ubiquitously expressed across eukaryotic cells in distinct triple of isoforms in multiple locations inside every cell.

I have presented multiple new mathematical models of the component parts of the motif, and have combined these into mathematical models of the entire motif, using both empirical and mechanistic approaches. Each of these models is formed of one or more ordinary differential equations.

Initial empirical models suggested a non-trivial behaviour for the motif: an ability to produce two distinct signals from a signal input signal. Moderate parameters suggested that PLD acts in a switch-like manner with an insensitivity to low-level noise; only above a critical stimulation does PLD act to propagate a signal. PI4P5K, in contrast, appears to act to propagate a signal largely as-is, likely resulting from an assumption of a less restrictive mechanism of activation. The arbitrary construction of these models and a limit to their mathematical tractability meant that effort was focused on the construction of mechanistic models.

Individual mechanistic models were developed for PLD and PI4P5K in chapter 5, and the regulation of Arf in chapter 6. The framework of Gunawardena (2012) was used to construct suitable quasi-steady-state (à la Michaelis-Menten) models for each enzyme. Where appropriate, additionally simplifying assumptions were taken to produce reasonably compact final models. Cytosol-membrane interactions were included in the models of PI4P5K and the regulation of Arf using the approach described by Kartal & Ebenhöh (2013).

The component mechanistic models were finalised with the definition of a full set of (preliminary parameters) and combined to form a new model of the Arf/PLD/PI4P5K signalling motif. Initial simulations of (each version of) this model show



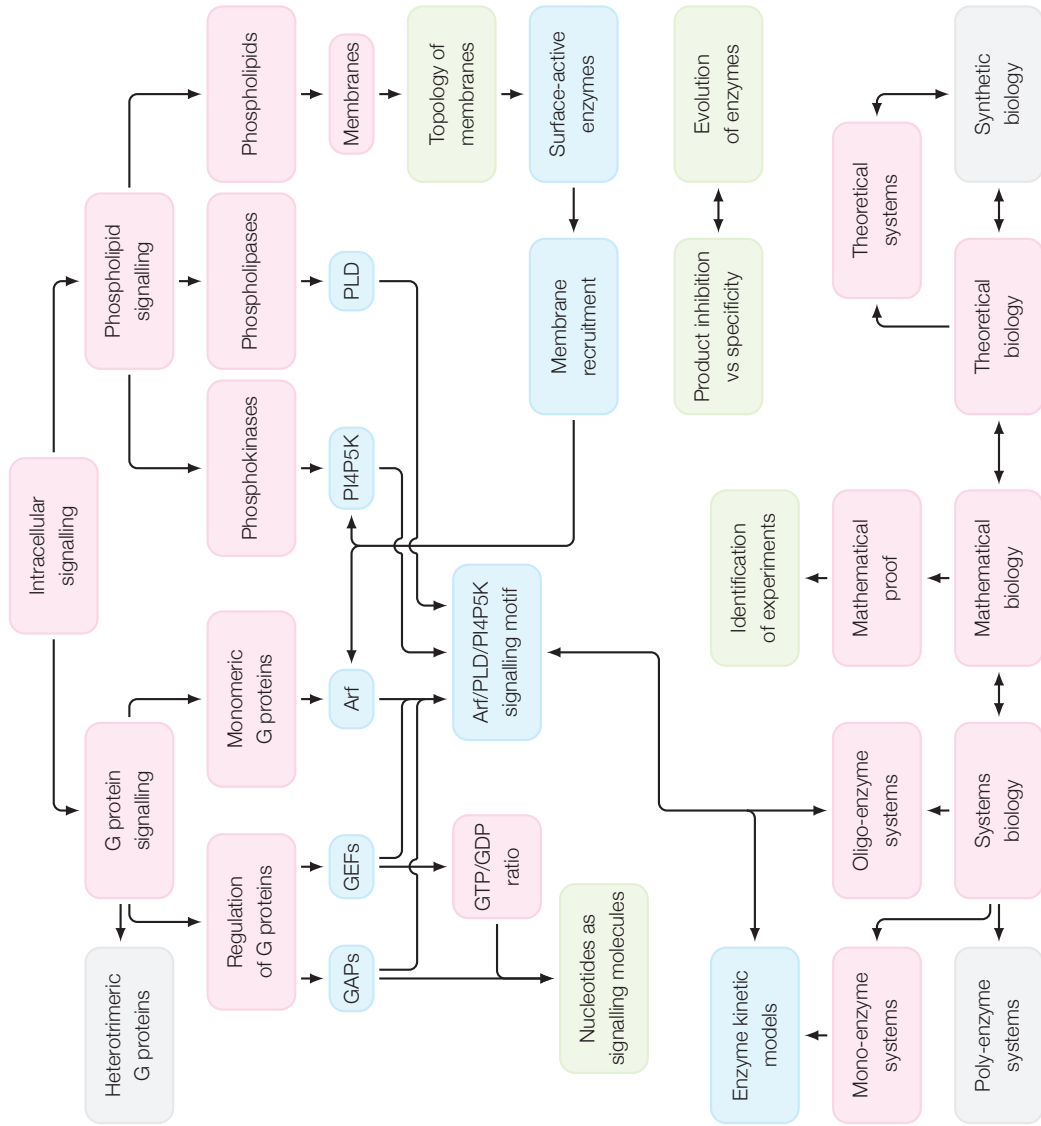
that it is capable of displaying physiologically-realistic behaviour consistent with the known biology.

The final, complete model of the Arf/PLD/PI4P5K signalling motif is suitable for further individual analysis (under assumptions related to both healthy and diseased states) and inclusion into other, larger models. For comparison with this model, this final chapter has indicated potential approaches for gathering suitable experimental data.

Finally—in addition to the derivation of these mathematical models—there are three other major conclusions for the work presented within this thesis:

- I have identified a lack of current understanding of the molecular mechanisms behind the regulatory and catalytic actions of the enzymes PLD and PI4P5K. More experimental and theoretical evidence is needed to correctly verify (or reject) the models proposed here or those constructed in any subsequent attempts.
- G protein:GEF:GAP systems do not behave, as commonly described, as simple activation/inactivation cycles but are better described as regulation through a balance/imbalance mechanism. This result has fundamental implications for the study of these wide-spread and crucial signalling components.
- ‘Simple’, oligo-enzyme biochemical systems such as the Arf/PLD/PI4P5K signalling motif can display complex behaviours. So care must be taken for the inclusion of such motifs into models of larger processes and signalling pathways, that the small components are adequately and suitably modelled. The Arf/PLD/PI4P5K system in particular has a greater complexity than is initially apparent: physical interactions with the membrane; differential localisation (surface adsorption); regulatory interactions on uncertain, perhaps complex, binding sites; and close integration with other signalling systems (as a lipid-modifying process). These properties of the system have each been considered within this thesis to varying degrees, somewhat dependent on the limits of the current level of knowledge, and the availability and complexity of suitable mathematical tools. However, to properly understand the behaviours of the system each of these properties (and many other) must be fully understood at the scale of the molecular processes.

I have summarised the key themes of this thesis in figure 8.6.



**Figure 8.6** Key themes and connections between themes within this thesis.

## A Python module enzymegraph: source code

This appendix contains the entire source code for version 1.0.2 of the Python module enzymegraph, written for use in this thesis. See section 3.4 for further description of the implementation and usage. As of the date of submission, the source code is also available in the repository:

<https://github.com/robjstan/python-enzymegraph>.

This source code has been released under the MIT license:

The MIT License (MIT)

Copyright (c) 2014 Rob J Stanley

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

```

1 from sympy import *
2 import copy
3
4 # -----
5 class enzymegraph:
6     # initialise from edges and (optionally) vertices
7     def __init__(self, edges, vertices = None):
8         # import edges
9         if isinstance(edges, dict) and \
10            enzymegraph.validate_edge_format([e for e in edges]):
11            self.edges = edges
12        elif isinstance(edges, list) and \
13            enzymegraph.validate_edge_format(edges):
14            self.edges = {edge:1 for edge in edges}
15        else:
16            raise Exception("Unexpected edge list format.")
17
18        # import vertices
19        if vertices:
20            self.vertices = vertices
21        else:
22            self.vertices = list(set([vertex for edge in edges \
23                                   for vertex in edge]))
24
25        # -----
26        # initialise from matrix and (optionally) vertices
27        @classmethod
28        def from_matrix(cls, matrix, vertices=None):
29            if not vertices:
30                vertices = [x for x in range(len(matrix))]
31
32            edges={vertices[i], vertices[j]):mij \
33                   for i, mi in enumerate(matrix) \
34                   for j, mij in enumerate(mi) \
35                   if mij}
36
37            return cls(edges, vertices)
38
39
40 # -----
41 # validate a list of edges as [(a1, b1), ..., (an, bn)]
42 # throw error if not
43 @classmethod
44 def validate_edge_format(cls, edges):
45     try:

```

```

46         return all([len(edge) == 2 for edge in edges])
47     except:
48         raise Exception("Unexpected edge format.")
49
50     # -----
51     # check subgraph is valid
52     # throw error if not
53     def validate_subgraph(self, subgraph):
54         vertex_check = all([stv in self.vertices \
55             for stv in subgraph.vertices])
56         edge_check = all([ste in self.edges for ste in subgraph.edges])
57
58         if not (vertex_check and edge_check):
59             raise Exception("Not a subtree of this graph!")
60
61     # -----
62 # -----
63     # return the (current) size of the graph (number of vertices)
64     def size(self):
65         return len(self.vertices)
66
67     # -----
68     # list all inwards edges from subgraph
69     def inward_edges(self, subgraph):
70         self.validate_subgraph(subgraph)
71
72         return {edge:k for edge, k in self.edges.items() \
73             if (edge[1] in subgraph.vertices) \
74             and (edge[0] not in subgraph.vertices)}
75
76     # -----
77     # list all outward edges from subgraph
78     def outward_edges(self, subgraph):
79         self.validate_subgraph(subgraph)
80
81         return {edge:k for edge, k in self.edges.items() \
82             if (edge[0] in subgraph.vertices) \
83             and (edge[1] not in subgraph.vertices)}
84
85     # -----
86     def forwards_edges(self, vertex):
87         return self.outward_edges(enzymegraph([], [vertex]))
88
89     # -----
90     def backwards_edges(self, vertex):

```

```

91     return self.inward_edges(enzymegraph([], [vertex]))
92
93 # -----
94 # append edge to graph, with label
95 def append_edge(self, edge, label = 1):
96     # check the edge is in a valid format
97     enzymegraph.validate_edge_format([edge])
98
99     # add edge
100    self.edges[edge] = label
101
102    # add vertices
103    self.append_vertex(edge[0])
104    self.append_vertex(edge[1])
105
106    # -----
107    # append vertex to the graph
108    def append_vertex(self, vertex):
109        # only if it is not already present
110        if vertex not in self.vertices:
111            (self.vertices).append(vertex)
112
113    # -----
114    # remove edge from the graph
115    def remove_edge(self, edge):
116        try:
117            del self.edges[edge]
118        except:
119            pass
120
121    # -----
122    # remove vertex from the graph
123    def remove_vertex(self, vertex):
124        try:
125            # make a new subgraph
126            graph_vertex = enzymegraph([], vertices=[vertex])
127            # remove all edges in and out
128            [self.remove_edge(edge) for edge in \
129             self.inward_edges(graph_vertex)]
130            [self.remove_edge(edge) for edge in \
131             self.outward_edges(graph_vertex)]
132        except:
133            pass
134
135

```

```

136 # -----
137 # output all spanning trees (as a graph) rooted at a given vertex
138 # modified from:
139 # Gabow, H. N., & Myers, ODE. W. (1978).
140 # Finding All Spanning Trees of Directed and Undirected Graphs.
141 # SIAM Journal on Computing, 7(3), -280287. doi:10.1137/0207024
142 def spanning_trees(self, root=None):
143     # subroutine grow
144     def grow(graph, tree):
145         # return if we've reached the bottom
146         if self.size() == tree.size():
147             yield tree
148
149         else:
150             # make a copy of the current graph
151             new_graph = copy.deepcopy(graph)
152
153             # for each inwards edge
154             for edge in graph.inward_edges(tree):
155                 # make a copy of the tree, and add this new edge
156                 new_tree = copy.deepcopy(tree)
157                 new_tree.append_edge(edge, label=graph.edges[edge])
158
159                 # grow the graph, return the result
160                 for subtree in grow(new_graph, new_tree):
161                     yield subtree
162
163                 # remove the edge just used
164                 # otherwise other subtrees will use this edge!
165                 new_graph.remove_edge(edge)
166
167     roots = [root] if root else self.vertices
168     for root in roots:
169         for spanning_tree in grow(self, enzymegraph([], \
170                                 vertices=[root])):
171             yield spanning_tree
172
173 # -----
174 # get basis element from spanning trees
175 def basis_element(self):
176     return {root:sum([prod(t.edges.values()) \
177                     for t in self.spanning_trees(root)]) \
178            for root in self.vertices}
179
180 # -----

```

```

181 # get normed basis element
182 def qssa_replacements(self, total_value):
183     rho = self.basis_element()
184     return {k:rho[k]/sum(rho.values())*total_value for k in rho}
185
186 # -----
187 # get ODE model
188 def ode_model(self):
189     return {v: \
190         sum([k[0]*v for k, v in self.backwards_edges(v).items()]) \
191         -sum([k[0]*v for k, v in self.forwards_edges(v).items()]) \
192         for v in self.vertices}
193
194 # -----
195 # get ODE model as function (suitable for integrating with scipy)
196 def ode_function(self, param_dict, variables, extra_odes={}):
197     # get odes
198     odes = self.ode_model()
199     for k, v in extra_odes.items():
200         # try adding to existing key, or add new key
201         try:
202             odes[k] += v
203         except:
204             odes[k] = v
205
206     # get sympy function
207     f_sympy = lambdify(variables, [odes[v].subs(param_dict) \
208         for v in variables])
209     # need to map this to a form that scipy expects
210     return (lambda x, t: f_sympy(*x))
211
212
213 # -----
214 default_text_formatter = lambda text: '$'+latex(text)+'$'
215
216 # get graph as tikz
217 def graph_as_tikz(self, vertex_pos, edge_styles={},
218                 relabel=True, head=True,
219                 vertex_text_formatter=default_text_formatter,
220                 edge_text_formatter=default_text_formatter):
221     before = '\\begin{tikzpicture}'
222     after = '\\n\\end{tikzpicture}'
223
224     # sort vertices (for consistent tikz output)
225     self.vertices = sorted(self.vertices, key=default_sort_key)

```



```

226
227     # dictionary of vertex information
228     vertex_dict = { \
229         vertex : {'label'      : vertex_text_formatter(vertex),
230                  'v_id'       : 'v%02d' % i if relabel else vertex,
231                  'position'   : vertex_pos[vertex],} \
232         for i, vertex in enumerate(self.vertices)}
233
234     # dictionary of edge information
235     edge_dict = { \
236         edge : {'start_v_id' : vertex_dict[edge[0]]['v_id'],
237                'end_v_id'   : vertex_dict[edge[1]]['v_id'],
238                'line_style' : edge_styles[edge] \
239                    if edge in edge_styles else '--',
240                'label'      : edge_text_formatter(self.edges[edge]),} \
241         for edge in self.edges}
242
243     # return tikz commands
244     return (self._tikz_head() if head else '') \
245         + before \
246         + self._tikz_nodes(vertex_dict) \
247         + self._tikz_paths(edge_dict) \
248         + after
249
250     # -----
251     # get spanning trees as tikz
252     def spanning_trees_as_tikz(self, vertex_pos, edge_styles={},
253                               relabel=True, head=True,
254                               vertex_text_formatter=default_text_formatter,
255                               edge_text_formatter=default_text_formatter):
256         tikz = [st.graph_as_tikz(vertex_pos, edge_styles, relabel,
257                                 head=False, vertex_text_formatter=vertex_text_formatter, \
258                                 edge_text_formatter=edge_text_formatter) \
259                for st in self.spanning_trees()]
260
261         return (self._tikz_head() if head else '') + "\n%\n".join(tikz)
262
263     # -----
264     # tikz head
265     def _tikz_head(self):
266         return '\\tikzset{edge label/.style=' \
267             + '{font=\\tiny, fill=white, inner sep=1}}\n\n'
268
269     # -----
270     # format graph nodes as tikz

```

```

271     def _tikz_nodes(self, vertex_dict):
272         before = '\n\t'
273         after = '\n'
274         template = '\\node %(v_id)s at %(position)s {%(label)s};'
275         nodes = [template % vertex for vertex in vertex_dict.values()]
276
277         return before + '\n\t'.join(sorted(nodes)) + after
278
279     # -----
280     # format graph edges as tikz
281     def _tikz_paths(self, edge_dict):
282         before = '\t\\begin{scope}[every node/.style=edge label]\n\t\t'
283         after = '\n\t\\end{scope}'
284         template = '\\draw[->] %(start_v_id)s %(line_style)s ' \
285                 + 'node {%(label)s} %(end_v_id)s;'
286         edges = [template % vertex for vertex in edge_dict.values()]
287
288         return before + '\n\t\t'.join(sorted(edges)) + after
289
290     # -----
291     # helper function to generate (arbitrary parameters)
292     def parameter_generator(i=0, label='k'):
293         while True:
294             yield symbols('%s_%d' % (label, i), positive=True)
295             i += 1

```

## B Derivation of the PMA model

This material was originally included in Stanley (2011). The PMA model of the Arf/PLD/PI4P5K motif is described by the system of ODEs

$$\begin{aligned}\frac{dx}{dt} &= a_0 + a_1zy - a_r x \\ \frac{dy}{dt} &= b_0 + b_1x + b_2z - b_r y.\end{aligned}$$

This system has a single steady-state, at

$$\begin{aligned}x_{ss} &= \frac{a_0 b_r + a_1 z(b_0 + b_2 z)}{a_r b_r - a_1 b_1 z} \\ y_{ss} &= \frac{a_0 b_1 + a_r(b_0 + b_2 z)}{a_r b_r - a_1 b_1 z}.\end{aligned}$$

This steady-state will be positive in both  $x$  and  $y$  when  $a_r b_r > a_1 b_1 z$ , negative whenever in both  $x$  and  $y$  when  $a_r b_r < a_1 b_1 z$ , and undefined when  $a_r b_r = a_1 b_1 z$ .

The change of variables

$$\begin{aligned}u &= x - x_{ss} \\ v &= y - y_{ss}\end{aligned}$$

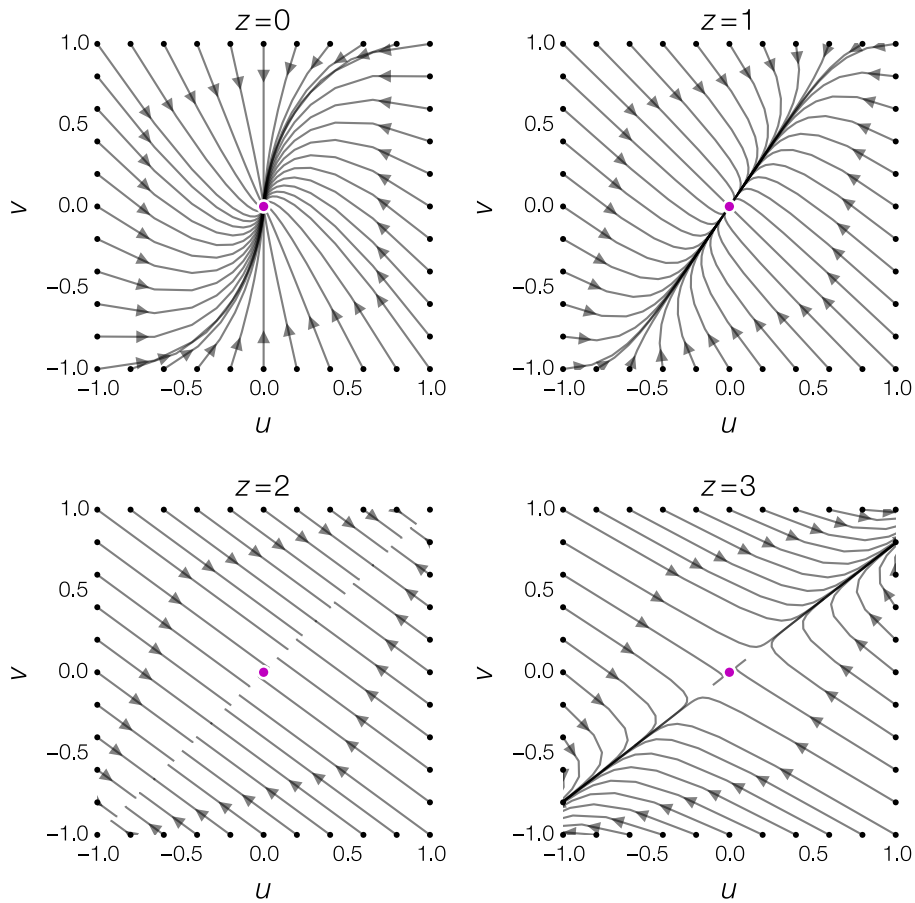
creates a new homogeneous system  $(u, v)$ ,

$$\frac{du}{dt} = a_1 z v - a_r u \tag{B.1}$$

$$\frac{dv}{dt} = b_1 u - b_r v. \tag{B.2}$$

where the steady-state is now at the origin. The system can be written in matrix form as

$$\begin{pmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{pmatrix} = \mathbf{A} \begin{pmatrix} u \\ v \end{pmatrix}, \text{ where } \mathbf{A} = \begin{pmatrix} -a_r & a_1 z \\ b_1 & -b_r \end{pmatrix}.$$



**Figure B.1** Phase portraits of equation B.2 with  $a_1 = 0.2$ ,  $a_r = 0.4$ ,  $b_1 = b_r = 0.3$ .  $z < 2$  gives a stable steady-state;  $z = 2$  gives a bifurcation and a stable line of equilibrium points;  $z > 2$  gives a saddle node. Black circles are the initial conditions for the connected trajectory. Drawn using the implicit solution derived in section 4.1.1.

and so the eigenvalues of  $\mathbf{A}$  are

$$\lambda_{\pm} = \frac{1}{2} \left[ -(a_r + b_r) \pm \sqrt{(a_r - b_r)^2 + 4a_1 b_1 z} \right].$$

One eigenvalue ( $\lambda_-$ ) is always negative as it is the sum of two negative components.

The other eigenvalue ( $\lambda_+$ ) is negative if and only if

$$\begin{aligned} a_r + b_r &> \sqrt{(a_r - b_r)^2 + 4a_1 b_1 z} \\ (a_r + b_r)^2 &> (a_r - b_r)^2 + 4a_1 b_1 z \\ (a_r + b_r)^2 - (a_r - b_r)^2 &> 4a_1 b_1 z \\ 4a_r b_r &> 4a_1 b_1 z \\ a_r b_r &> a_1 b_1 z \end{aligned}$$

So for  $a_r b_r > a_1 b_1 z$  there is a positive, stable steady-state; for  $a_r b_r < a_1 b_1 z$  there is a negative, unstable steady-state; and so when  $a_r b_r = a_1 b_1 z$  there is a bifurcation. Phase planes demonstrating different behaviours of this system for different values of  $z$  are shown in figure B.1.



## C Complete model: source code and additional figures

This appendix contains the source code for Python implementations of the models listed in chapter 7.

```

1 from sympy import *
2 import numpy as np
3
4 # -----
5 # symbolic setup
6 kcat, Ks, Ka, Kr = symbols('k_cat K_s K_a K_r')
7 kdecay          = symbols('k_decay')
8 koff, kA        = symbols('k_off k_A')
9 kexc, krel      = symbols('k_exc k_rel')
10 kappa, K0, K1, K2 = symbols('kappa K_0 K_1 K_2')
11 kase, Km        = symbols('k_ase K_m')
12
13 # -----
14 # ADSORPTION MODELS
15 # calculating theta
16 def theta(x, n0, alpha):
17     ARFt, ARFc, ARF, PC, PA, PIP, PIP2, GEft, GAPt, PLDt, PIPKt, PIPKc = x
18     return ( PIPKt - PIPKc + ARFt - ARFc )/(n0*alpha)
19
20 # membrane-adsorption model, Langmuir
21 def phi_langmuir(x, n0, alpha):
22     return 1 - theta(x, n0, alpha)
23
24 # membrane-adsorption model, RSA
25 def phi_RSA(x, n0, alpha):
26     thc = 0.547; a1 = 6.107; a2 = 9.059; a3 = 8.669; a4 = -28.56
27     th = theta(x, n0, alpha)
28     if th > thc: return 0 # approximation only valid for th <= thc
29     else: return (thc - th)**3*(a1 + a2*th + a3*th**2 + a4*th**3)

```

## 272 COMPLETE MODEL: SOURCE CODE AND ADDITIONAL FIGURES

```

30
31 # -----
32 # PLD MODEL
33 def PLD_model(x, t, p):
34     ARFt, ARFc, ARF, PC, PA, PIP, PIP2, GEft, GAPt, PLDt, PIPKt, PIPKc = x
35
36     # compute net catalytic rate
37     cat = ( p[kcat] * ARF * PIP2 * PC * PLDt ) \
38           / ( ( p[Ks] + PC)*(p[Ka] + ARF)*(p[Kr] + PIP2) )
39
40     return conclist({'PA': cat - p[kdecay]*PA})
41
42 # -----
43 # PI4P5K MODEL (3-site)
44 def PIPK_model3(x, t, p, n0, alpha, phi=phi_langmuir):
45     ARFt, ARFc, ARF, PC, PA, PIP, PIP2, GEft, GAPt, PLDt, PIPKt, PIPKc = x
46
47     # compute net catalytic rate
48     cat = ( p[kcat]*PIP*(PIPKt - PIPKc) ) / ( p[Ks] + PIP)
49
50     # compute net PIPK adsorption
51     dPIPKc = p[koff]*p[Ks]*p[Ka]*p[Kr]*(PIPKt - PIPKc) \
52            / ( ( p[Ks] + PIP)*(p[Ka] + ARF)*(p[Kr] + PA) ) \
53            - p[koff]*p[kA]*phi(x, n0, alpha)*alpha*PIPKc
54
55     return conclist({'PIP2': cat - p[kdecay]*PIP2, 'PIPKc': dPIPKc})
56
57 # -----
58 # PI4P5K MODEL (2-site)
59 def PIPK_model2(x, t, p, n0, alpha, phi=phi_langmuir):
60     ARFt, ARFc, ARF, PC, PA, PIP, PIP2, GEft, GAPt, PLDt, PIPKt, PIPKc = x
61
62     chi = ( p[Kr]*(1 + p[kexc]*PIP)*(p[Ks] + PIP) \
63            + (p[Ks] + p[krel]*PIP)*PA ) / (PIPKt - PIPKc)
64
65     # compute net catalytic rate
66     cat = p[kcat]*PIP*( p[Kr]*( 1 + p[kexc]*PIP ) + p[krel]*PA ) / chi
67
68     # compute net PIPK adsorption
69     dPIPKc = p[koff]*p[Ks]*p[Ka]*p[Kr]*(1 + p[kexc]*PIP) \
70            / ( chi*(p[Ka] + ARF) ) \
71            - p[koff]*p[kA]*phi(x, n0, alpha)*alpha*PIPKc
72
73     return conclist({'PIP2': cat - p[kdecay]*PIP2, 'PIPKc': dPIPKc})
74

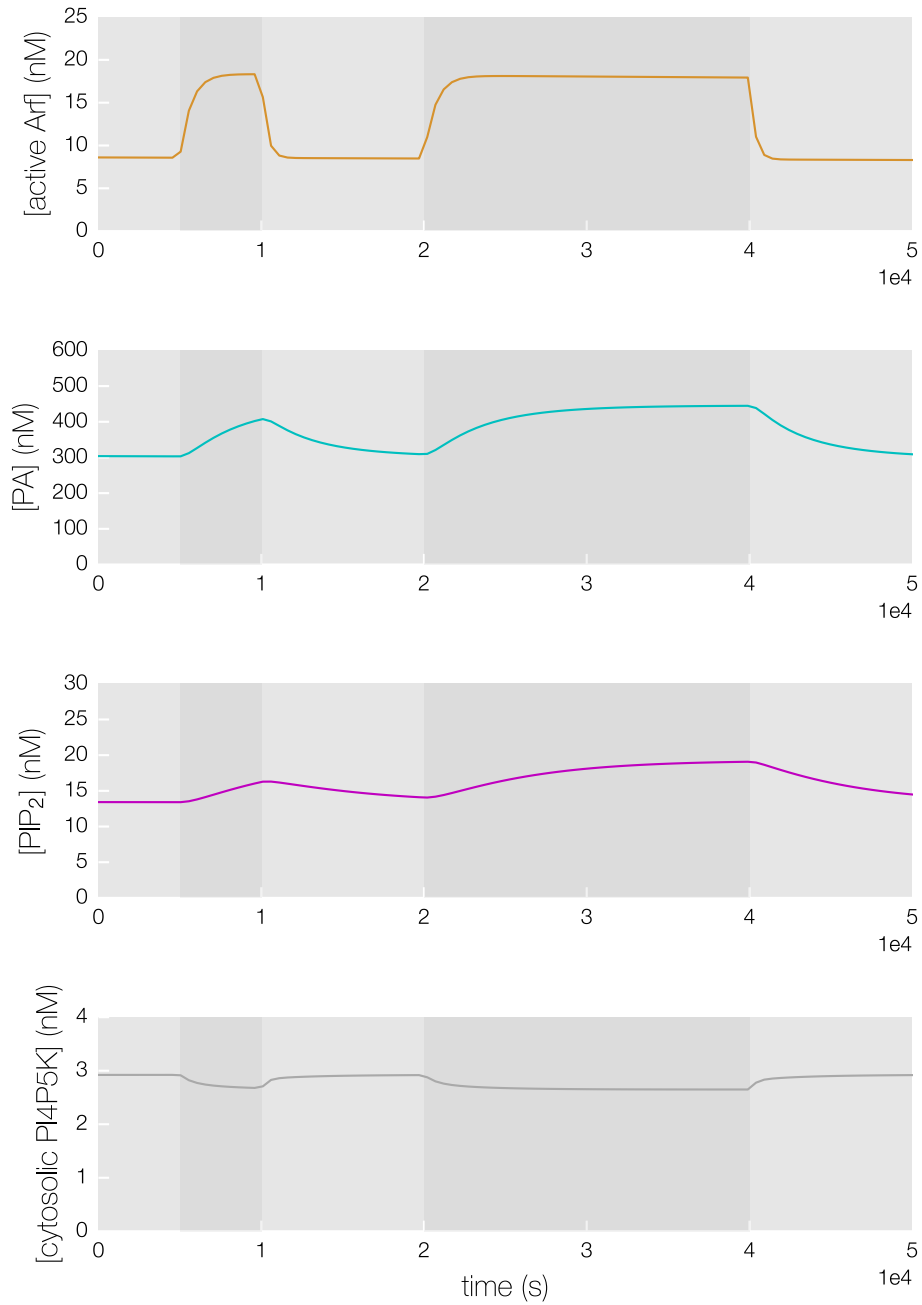
```



```

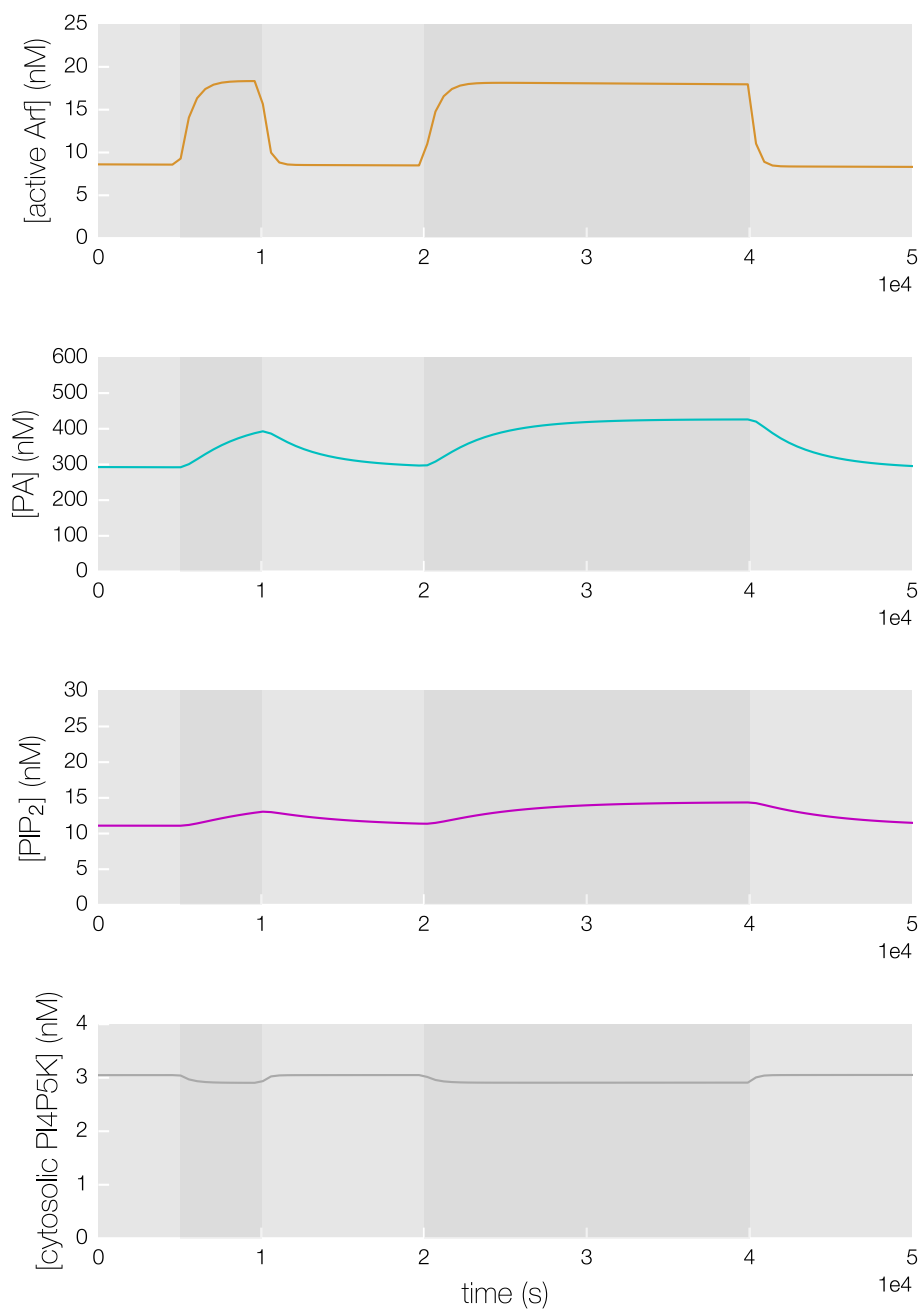
75 # -----
76 # ARF REGULATION MODEL
77 def Arf_model(x, t, p, n0, alpha, phi=phi_langmuir):
78     ARFt, ARFc, ARF, PC, PA, PIP, PIP2, GEft, GAPt, PLDt, PIPkt, PIPKc = x
79
80     # compute net catalytic rates
81     gef = p[kcat]*(ARFt - ARFc - (p[kappa] + 1)*ARF)*GEft \
82         / (p[K0] + p[K1]*(ARFt - ARFc) + p[K2]*ARF)
83     gap = p[kase]*ARF*GAPt / (p[Km] + ARF)
84
85     # compute net ARF adsorption
86     dARFc = p[koff]*(ARFt - (1 + p[kassoc]*phi(x, n0, alpha)*alpha)*ARFc \
87         - ARF)
88
89     return conclist({'ARF': gef - gap, 'ARFc': dARFc})
90
91 # -----
92 # COMPLETE MODEL
93 def motif_model(x, t, p, n0, alpha, phi=phi_langmuir,
94                 PIPK_model=PIPK_model2, fun_perturb=lambda x,t: x):
95
96     # perturb the model, depending on the supplied function
97     x = fun_perturb(x, t)
98
99     return np.array(Arf_model(x, t, p['Arf'], n0, alpha, phi)) \
100         + np.array(PLD_model(x, t, p['PLD'])) \
101         + np.array(PIPK_model(x, t, p['PIPK'], n0, alpha, phi))
102
103 # -----
104 # HELPER FUNCTIONS
105 # variable names, in order
106 vars = ['ARFt', 'ARFc', 'ARF', 'PC', 'PA', 'PIP', 'PIP2', 'GEft', \
107         'GAPt', 'PLDt', 'PIPkt', 'PIPKc']
108
109 # return list of concentrations given dictionary
110 def conclist(d, default=0):
111     return [d[var] if var in d else default for var in vars]
112
113 # return dictionary of concentrations given list
114 def concdict(l, default=0):
115     return {k:v for k,v in zip(vars,l)}

```

**A** 3-site model of PI4P5K; RSA model

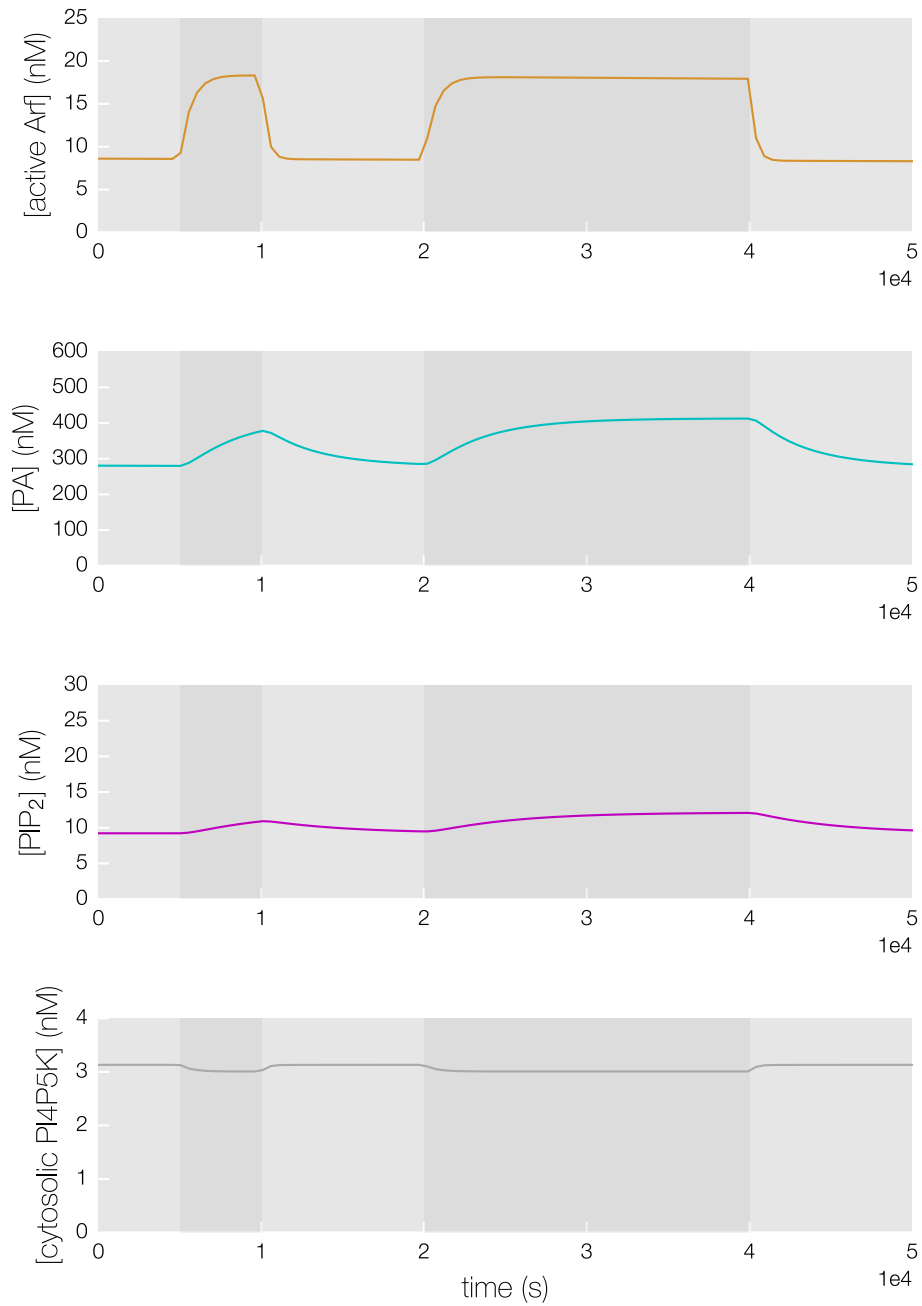
**Figure C.1** Simulations of the complete mechanistic model of the Arf/PLD/PI4P5K motif, see figure 7.7. As figure 7.8.

**B** 2-site model of PI4P5K; Langmuir's model



**Figure C.1** Simulations of the complete mechanistic model of the Arf/PLD/PI4P5K motif, see figure 7.7. As figure 7.8.

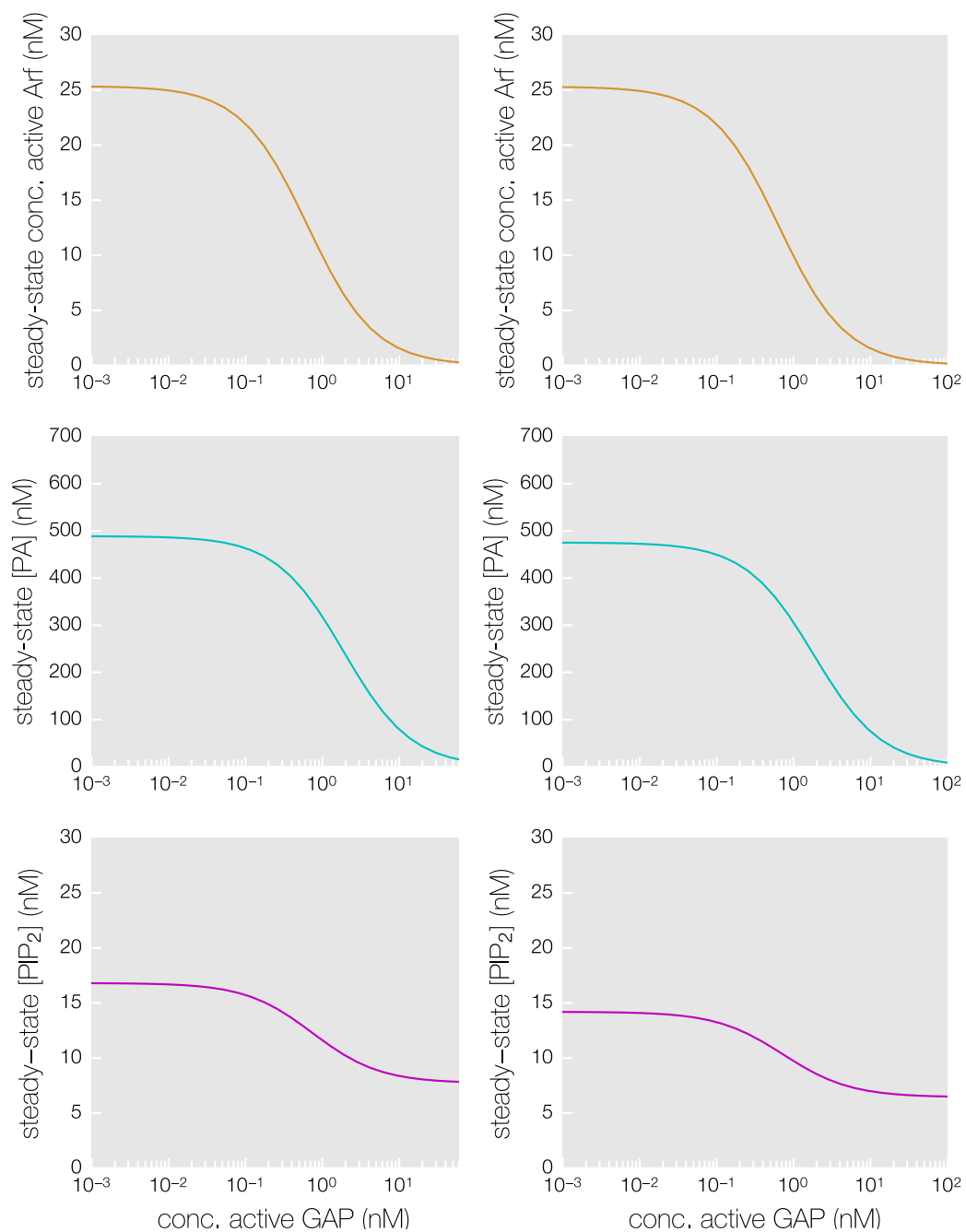
## C 2-site model of PI4P5K; RSA model



**Figure C.1** Simulations of the complete mechanistic model of the Arf/PLD/PI4P5K motif, see figure 7.7. As figure 7.8.

A Langmuir's model

B RSA model



**Figure C.2** Steady-state concentrations of active Arf, PA, and PI(4,5)P<sub>2</sub> as a function of the concentration of active GAP, assuming the 2-site model of PI4P5K. As figure 7.9.



## Bibliography

- Ahearn, I. M., Haigis, K., Bar-Sagi, D., & Philips, M. R. (2012). 'Regulating the regulator: post-translational modification of RAS.' *Nat. Rev. Mol. Cell Biol.* **13**(1), 39–51.
- Ahmadian, M. R., Hoffmann, U., Goody, R. S., & Wittinghofer, A. (1997). 'Individual rate constants for the interaction of Ras proteins with GTPase-activating proteins determined by fluorescence spectroscopy.' *Biochemistry* **36**(15), 4535–4541.
- Alt, F. L. (1972). 'Archaeology of computers: reminiscences, 1945-1947'. *Communications of the ACM* **15**(7), 693–694.
- Antonny, B., Béraud-Dufour, S., Chardin, P., & Chabre, M. (1997). 'N-terminal hydrophobic residues of the G-protein ADP-ribosylation factor-1 insert into membrane phospholipids upon GDP to GTP exchange.' *Biochemistry* **36**(15), 4675–4684.
- Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). 'The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.' *Bioinformatics* **22**(2), 195–201.
- Bazenet, C. E. & Anderson, R. A. (1992). 'Phosphatidylinositol-4-phosphate 5-kinases from human erythrocytes.' *Meth. Enzymol.* **209**, 189–202.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). 'The Protein Data Bank.' *Nucleic Acids Research* **28**(1), 235–242.
- Berridge, M. J. (2014). *Cell Signalling Biology*.
- Berstein, G., Blank, J. L., Jhon, D.-Y., Exton, J. H., Rhee, S. G., & Ross, E. M. (1992). 'Phospholipase C- $\beta$ 1 is a GTPase-activating protein for Gq/11, its physiologic regulator'. *Cell* **70**(3), 411–418.
- Beseničar, M., Maček, P., Lakey, J. H., & Anderluh, G. (2006). 'Surface plasmon resonance in protein-membrane interactions.' *Chem. Phys. Lipids* **141**(1-2), 169–178.

- Bhalla, U. S. & Iyengar, R. (1999). 'Emergent properties of networks of biological signaling pathways.' *Science* **283**(5400), 381–387.
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., & Schwede, T. (2014). 'SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information.' *Nucleic Acids Research* **42**(Web Server issue), W252–8.
- Bos, J. L., Rehmann, H., & Wittinghofer, A. (2007). 'GEFs and GAPs: critical elements in the control of small G proteins.' *Cell* **129**(5), 865–877.
- Botts, J. & Morales, M. (1953). 'Analytical description of the effects of modifiers and of enzyme multivalency upon the steady state catalyzed reaction rate.' *Transactions of the Faraday Society* **49**, 696–707.
- Briggs, G. E. & Haldane, J. B. (1925). 'A Note on the Kinetics of Enzyme Action.' *Biochem. J.* **19**(2), 338–339.
- Brown, F. D., Thompson, N., Saqib, K. M., Clark, J. M., Powner, D., Thompson, N. T., Solari, R., & Wakelam, M. J. (1998). 'Phospholipase D1 localises to secretory granules and lysosomes and is plasma-membrane translocated on cellular stimulation.' *Curr. Biol.* **8**(14), 835–838.
- Brugger, B., Glass, B., Haberkant, P., Leibrecht, I., Wieland, F. T., & Krausslich, H. G. (2006). 'The HIV lipidome: A raft with an unusual composition'. *Proc. Natl. Acad. Sci. U.S.A.* **103**(8), 2641–2646.
- Bui, Q. T., Golinelli-Cohen, M.-P., & Jackson, C. L. (2009). 'Large Arf1 guanine nucleotide exchange factors: evolution, domain structure, and roles in membrane trafficking and human disease.' *Mol. Genet. Genomics* **282**(4), 329–350.
- Burkhardt, U., Stegner, D., Hattingen, E., Beyer, S., Nieswandt, B., & Klein, J. (2014). 'Impaired brain development and reduced cognitive function in phospholipase D-deficient mice.' *Neurosci. Lett.* **572**, 48–52.
- Cai, J., Abramovici, H., Gee, S. H., & Topham, M. K. (2009). 'Diacylglycerol kinases as sources of phosphatidic acid.' *Biochim. Biophys. Acta* **1791**(9), 942–948.
- Casanova, J. E. (2007). 'Regulation of Arf activation: the Sec7 family of guanine nucleotide exchange factors.' *Traffic* **8**(11), 1476–1485.
- Cauvin, C. & Echard, A. (2014). 'Phosphoinositides: Lipids with informative heads and mastermind functions in cell division.' *Biochim. Biophys. Acta* **1851**(6), 832–843.
- Chang, H.-C. & Wang, L.-C. (2010). 'A Simple Proof of Thue's Theorem on Circle Packing'. arXiv: 1009.4322.



- Chen, J.-S. & Exton, J. H. (2004). 'Regulation of phospholipase D2 activity by protein kinase C alpha.' *J. Biol. Chem.* **279**(21), 22076–22083.
- Cherfils, J. & Zeghouf, M. (2013). 'Regulation of small GTPases by GEFs, GAPs, and GDIs.' *Physiol. Rev.* **93**(1), 269–309.
- Cochet, C. & Chambaz, E. M. (1986). 'Catalytic properties of a purified phosphatidylinositol-4-phosphate kinase from rat brain.' *Biochem. J.* **237**(1), 25–31.
- Cockcroft, S. (2009). 'Phosphatidic acid regulation of phosphatidylinositol 4-phosphate 5-kinases.' *Biochim. Biophys. Acta* **1791**(9), 905–912.
- Colley, W. C., Sung, T. C., Roll, R., Jenco, J., Hammond, S. M., Altshuler, Y., Bar-Sagi, D., Morris, A. J., & Frohman, M. A. (1997). 'Phospholipase D2, a distinct phospholipase D isoform with novel regulatory properties that provokes cytoskeletal reorganization.' *Curr. Biol.* **7**(3), 191–201.
- Colquhoun, D., Dowsland, K. A., Beato, M., & Plested, A. J. R. (2004). 'How to impose microscopic reversibility in complex reaction mechanisms.' *Biophys. J.* **86**(6), 3510–3518.
- Cornish-Bowden, A. & Cárdenas, M. L. (2001). 'Information transfer in metabolic pathways. Effects of irreversible steps in computer models.' *Eur. J. Biochem.* **268**(24), 6616–6624.
- Cornish-Bowden, A. (2012). *Fundamentals of Enzyme Kinetics*. Wiley-VCH.
- Di Paolo, G. & De Camilli, P. (2006). 'Phosphoinositides in cell regulation and membrane dynamics.' *Nature* **443**(7112), 651–657.
- Divecha, N., Roefs, M., Halstead, J. R., D'Andrea, S., Fernandez-Borga, M., Oomen, L., Saqib, K. M., Wakelam, M. J., & D'Santos, C. (2000). 'Interaction of the type Ialpha PIPkinase with phospholipase D: a role for the local generation of phosphatidylinositol 4, 5-bisphosphate in the regulation of PLD2 activity.' *EMBO J.* **19**(20), 5440–5449.
- Donaldson, J. G. & Jackson, C. L. (2011). 'ARF family G proteins and their regulators: roles in membrane transport, development and disease.' *Nat. Rev. Mol. Cell Biol.* **12**(6), 362–375.
- D'Souza-Schorey, C. & Chavrier, P. (2006). 'ARF proteins: roles in membrane traffic and beyond.' *Nat. Rev. Mol. Cell Biol.* **7**(5), 347–358.
- Du, G. (2003). 'Regulation of phospholipase D1 subcellular cycling through coordination of multiple membrane association motifs.' *J. Cell Biol.* **162**(2), 305–315.

- Duex, J. E., Nau, J. J., Kauffman, E. J., & Weisman, L. S. (2006). 'Phosphoinositide 5-phosphatase Fig 4p is required for both acute rise and subsequent fall in stress-induced phosphatidylinositol 3,5-bisphosphate levels.' *Eukaryotic Cell* 5(4), 723–731.
- Elvers, M., Grenegård, M., Khoshjabinzadeh, H., Münzer, P., Borst, O., Tian, H., Di Paolo, G., Lang, F., Gawaz, M., Lindahl, T. L., & Fälker, K. (2012). 'A novel role for phospholipase D as an endogenous negative regulator of platelet sensitivity.' *Cell. Signal.* 24(9), 1743–1752.
- Erickson, H. P. (2009). 'Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy'. *Biological Procedures Online* 11(1), 32–51.
- Feder, J. (1980). 'Random Sequential Adsorption'. *J. Theor. Biol.* 87(2), 237–254.
- Fell, D. A. & Sauro, H. M. (1985). 'Metabolic control and its analysis. Additional relationships between elasticities and control coefficients.' *Eur. J. Biochem.* 148(3), 555–561.
- Foskett, J. K., White, C., Cheung, K.-H., & Mak, D.-O. D. (2007). 'Inositol trisphosphate receptor Ca<sup>2+</sup> release channels.' *Physiol. Rev.* 87(2), 593–658.
- Freyberg, Z. (2002). 'Phospholipase D2 Is Localized to the Rims of the Golgi Apparatus in Mammalian Cells'. *Mol. Biol. Cell* 13(11), 3930–3942.
- Freyberg, Z., Sweeney, D., Siddhanta, A., Bourgoin, S., Frohman, M., & Shields, D. (2001). 'Intracellular localization of phospholipase D1 in mammalian cells.' *Mol. Biol. Cell* 12(4), 943–955.
- Funakoshi, Y., Hasegawa, H., & Kanaho, Y. (2011). 'Regulation of PIP5K activity by Arf6 and its physiological significance.' *J. Cell. Physiol.* 226(4), 888–895.
- Gabow, H. N. & Myers, E. W. (1978). 'Finding All Spanning Trees of Directed and Undirected Graphs'. *SIAM J. Comput.* 7(3), 280–287.
- Godi, A., Pertile, P., Meyers, R., Marra, P., Di Tullio, G., Iurisci, C., Luini, A., Corda, D., & De Matteis, M. A. (1999). 'ARF mediates recruitment of PtdIns-4-OH kinase-beta and stimulates synthesis of PtdIns(4,5)P<sub>2</sub> on the Golgi complex.' *Nat. Cell Biol.* 1(5), 280–287.
- Gomez-Cambronero, J. (2014). 'Phospholipase D in cell signaling: from a myriad of cell functions to cancer growth and metastasis.' *J. Biol. Chem.* 289(33), 22557–22566.
- Goody, R. S. (2014). 'How not to do kinetics: examples involving GTPases and guanine nucleotide exchange factors.' *FEBS J.* 281(2), 593–600.

- Guex, N., Peitsch, M. C., & Schwede, T. (2009). 'Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective.' *Electrophoresis* **30 Suppl 1**(S1), S162–73.
- Gunawardena, J. (2012). 'A linear framework for time-scale separation in nonlinear biochemical systems.' *PLoS ONE* **7**(5), e36321.
- Gunawardena, J. (2014). 'Time-scale separation - Michaelis and Menten's old idea, still bearing fruit.' *FEBS J.* **281**(2), 473–488.
- Hammes, G. G. (1978). 'Control of the Catalytic Activity of Enzymes by the Near and Remote Environment of a Polyatomic Framework'. In: *Advances in Chemical Physics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 177–227.
- Hammond, S. M., Altshuller, Y. M., Sung, T. C., Rudge, S. A., Rose, K., Engebrecht, J., Morris, A. J., & Frohman, M. A. (1995). 'Human ADP-ribosylation factor-activated phosphatidylcholine-specific phospholipase D defines a new and highly conserved gene family.' *Journal of Biological Chemistry* **270**(50), 29640–29643.
- Hammond, S. M., Jenco, J. M., Nakashima, S., Cadwallader, K., Gu, Q., Cook, S., Nozawa, Y., Prestwich, G. D., Frohman, M. A., & Morris, A. J. (1997). 'Characterization of two alternately spliced forms of phospholipase D1. Activation of the purified enzymes by phosphatidylinositol 4,5-bisphosphate, ADP-ribosylation factor, and Rho family monomeric GTP-binding proteins and protein kinase C-alpha.' *Journal of Biological Chemistry* **272**(6), 3860–3868.
- Hanahan, D. J. & Chaikoff, I. L. (1948). 'On the nature of the phosphorus-containing lipides of cabbage leaves and their relation to a phospholipide-splitting enzyme contained in these leaves.' *J. Biol. Chem.* **172**(1), 191–198.
- Henage, L. G., Exton, J. H., & Brown, H. A. (2006). 'Kinetic analysis of a mammalian phospholipase D: allosteric modulation by monomeric GTPases, protein kinase C, and polyphosphoinositides.' *Journal of Biological Chemistry* **281**(6), 3408–3417.
- Hirano, S., Sekine, K., Handa, T., & Nakano, M. (2012). 'Continuous monitoring of phospholipid vesicle hydrolysis by phospholipase D (PLD) reveals differences in hydrolysis by PLDs from 2 *Streptomyces* species.' *Colloids Surf B Biointerfaces* **94**, 1–6.
- Hiroshima, M. & Exton, J. H. (2005). 'Localization and regulation of phospholipase D2 by ARF6.' *J. Cell. Biochem.* **95**(1), 149–164.
- Hodgkin, M. N., Masson, M. R., Powner, D., Saqib, K. M., Ponting, C. P., & Wakelam, M. J. (2000). 'Phospholipase D regulation and localisation is dependent upon a phosphatidylinositol 4,5-bisphosphate-specific PH domain.' *Curr. Biol.* **10**(1), 43–46.

- Honda, A., Nogami, M., Yokozeki, T., Yamazaki, M., Nakamura, H., Watanabe, H., Kawamoto, K., Nakayama, K., Morris, A. J., Frohman, M. A., & Kanaho, Y. (1999). 'Phosphatidylinositol 4-phosphate 5-kinase alpha is a downstream effector of the small G protein ARF6 in membrane ruffle formation.' *Cell* **99**(5), 521–532.
- Hunter, J. D. (2007). 'Matplotlib: A 2D graphics environment'. *Computing in Science & Engineering* **9**(3), 90–95.
- Imamura, S. & Horiuti, Y. (1979). 'Purification of *Streptomyces chromofuscus* phospholipase D by hydrophobic affinity chromatography on palmitoyl cellulose.' *Journal of Biochemistry* **85**(1), 79–95.
- Ishihara, H., Shibasaki, Y., Kizuki, N., Katagiri, H., Yazaki, Y., Asano, T., & Oka, Y. (1996). 'Cloning of cDNAs encoding two isoforms of 68-kDa type I phosphatidylinositol-4-phosphate 5-kinase.' *J. Biol. Chem.* **271**(39), 23611–23614.
- Ishihara, H., Shibasaki, Y., Kizuki, N., Wada, T., Yazaki, Y., Asano, T., & Oka, Y. (1998). 'Type I phosphatidylinositol-4-phosphate 5-kinases. Cloning of the third isoform and deletion/substitution analysis of members of this novel lipid kinase family.' *J. Biol. Chem.* **273**(15), 8741–8748.
- Itoh, T. (2000). 'Autophosphorylation of Type I Phosphatidylinositol Phosphate Kinase Regulates Its Lipid Kinase Activity'. *Journal of Biological Chemistry* **275**(25), 19389–19394.
- Itoh, T. & Takenawa, T. (2002). 'Phosphoinositide-binding domains: Functional units for temporal and spatial regulation of intracellular signalling.' *Cell. Signal.* **14**(9), 733–743.
- Jarquín-Pardo, M., Fitzpatrick, A., Galiano, F. J., First, E. A., & Davis, J. N. (2007). 'Phosphatidic acid regulates the affinity of the murine phosphatidylinositol 4-phosphate 5-kinase-beta for phosphatidylinositol-4-phosphate.' *J. Cell. Biochem.* **100**(1), 112–128.
- Jean, S. & Kiger, A. A. (2012). 'Coordination between RAB GTPase and phosphoinositide regulation and functions.' *Nat. Rev. Mol. Cell Biol.* **13**(7), 463–470.
- Jenkins, G. H., Fiset, P. L., & Anderson, R. A. (1994). 'Type I phosphatidylinositol 4-phosphate 5-kinase isoforms are specifically stimulated by phosphatidic acid.' *J. Biol. Chem.* **269**(15), 11547–11554.
- Jeon, H., Kwak, D., Noh, J., Lee, M. N., Lee, C. S., Suh, P.-G., & Ryu, S. H. (2011). 'Phospholipase D2 induces stress fiber formation through mediating nucleotide exchange for RhoA.' *Cell. Signal.* **23**(8), 1320–1326.

- Jiang, H., Luo, J. Q., Urano, T., Frankel, P., Lu, Z., Foster, D. A., & Feig, L. A. (1995). 'Involvement of Ral GTPase in v-Src-induced phospholipase D activation.' *Nature* 378(6555), 409–412.
- Johnson, K. A. & Goody, R. S. (2011). 'The original Michaelis constant: Translation of the 1913 Michaelis-Menten Paper'. *Biochemistry* 50(39), 8264–8269.
- Jones, D. H., Morris, J. B., Morgan, C. P., Kondo, H., Irvine, R. F., & Cockcroft, S. (2000). 'Type I phosphatidylinositol 4-phosphate 5-kinase directly interacts with ADP-ribosylation factor 1 and is responsible for phosphatidylinositol 4,5-bisphosphate synthesis in the golgi compartment.' *J. Biol. Chem.* 275(18), 13962–13966.
- Kahn, R. A. & Gilman, A. G. (1984). 'Purification of a protein cofactor required for ADP-ribosylation of the stimulatory regulatory component of adenylate cyclase by cholera toxin.' *J. Biol. Chem.* 259(10), 6228–6234.
- Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Assad-Garcia, N., Glass, J. I., & Covert, M. W. (2012). 'A whole-cell computational model predicts phenotype from genotype.' *Cell* 150(2), 389–401.
- Kartal, O. & Ebenhöf, O. (2013). 'A generic rate law for surface-active enzymes.' *FEBS Letters* 587(17), 2882–2890.
- Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., & Schwede, T. (2009). 'The SWISS-MODEL Repository and associated resources.' *Nucleic Acids Research* 37(Database issue), D387–92.
- Klebe, C., Prinz, H., Wittinghofer, A., & Goody, R. S. (1995). 'The kinetic mechanism of Ran–nucleotide exchange catalyzed by RCC1.' *Biochemistry* 34(39), 12543–12552.
- Klein, J. (2005). 'Functions and pathophysiological roles of phospholipase D in the brain'. *J Neurochem* 94(6), 1473–1487.
- Kolanus, W. (2007). 'Guanine nucleotide exchange factors of the cytohesin family and their roles in signal transduction.' *Immunol. Rev.* 218(1), 102–113.
- Kolch, W. (2000). 'Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions.' *Biochem. J.* 351 Pt 2, 289–305.
- Kong, A. M., Speed, C. J., O'Malley, C. J., Layton, M. J., Meehan, T., Loveland, K. L., Cheema, S., Ooms, L. M., & Mitchell, C. A. (2000). 'Cloning and characterization of a 72-kDa inositol-polyphosphate 5-phosphatase localized to the Golgi network.' *Journal of Biological Chemistry* 275(31), 24052–24064.
- Kooijman, E. E. & Burger, K. N. J. (2009). 'Biophysics and function of phosphatidic acid: a molecular perspective.' *Biochim. Biophys. Acta* 1791(9), 881–888.

- Kunz, J., Wilson, M. P., Kisseleva, M., Hurley, J. H., Majerus, P. W., & Anderson, R. A. (2000). 'The activation loop of phosphatidylinositol phosphate kinases determines signaling specificity.' *Molecular Cell* 5(1), 1–11.
- Langmuir, I. (1918). 'The adsorption of gases on plane surfaces of glass, mica and platinum.' *J. Am. Chem. Soc.* 40(9), 1361–1403.
- Liang, J. O., Sung, T. C., Morris, A. J., Frohman, M. A., & Kornfeld, S. (1997). 'Different domains of mammalian ADP-ribosylation factor 1 mediate interaction with selected target proteins.' *J. Biol. Chem.* 272(52), 33001–33008.
- Liepe, J., Barnes, C., Cule, E., Erguler, K., Kirk, P., Toni, T., & Stumpf, M. P. H. (2010). 'ABC-SysBio—approximate Bayesian computation in Python with GPU support'. *Bioinformatics* 26(14), 1797–1799.
- Liepe, J., Kirk, P., Filippi, S., Toni, T., Barnes, C. P., & Stumpf, M. P. H. (2014). 'A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation.' *Nat Protoc* 9(2), 439–456.
- Ling, L. E., Schulz, J. T., & Cantley, L. C. (1989). 'Characterization and purification of membrane-associated phosphatidylinositol-4-phosphate kinase from human red blood cells.' *J. Biol. Chem.* 264(9), 5080–5088.
- Liscovitch, M., Chalifa, V., Pertile, P., Chen, C. S., & Cantley, L. C. (1994). 'Novel function of phosphatidylinositol 4,5-bisphosphate as a cofactor for brain membrane phospholipase D.' *J. Biol. Chem.* 269(34), 21403–21406.
- Liscovitch, M., Czarny, M., Fiucci, G., & Tang, X. (2000). 'Phospholipase D: molecular and cell biology of a novel gene family.' *Biochem. J.* 345 Pt 3, 401–415.
- Liu, Y., Kahn, R. A., & Prestegard, J. H. (2009). 'Structure and membrane interaction of myristoylated ARF1.' *Structure* 17(1), 79–87.
- Loijens, J. C. & Anderson, R. A. (1996). 'Type I phosphatidylinositol-4-phosphate 5-kinases are distinct members of this novel lipid kinase family.' *J. Biol. Chem.* 271(51), 32937–32943.
- Loirand, G., Sauzeau, V., & Pacaud, P. (2013). 'Small G proteins in the cardiovascular system: physiological and pathological aspects.' *Physiol. Rev.* 93(4), 1659–1720.
- Lopez, I., Arnold, R. S., & Lambeth, J. D. (1998). 'Cloning and initial characterization of a human phospholipase D2 (hPLD2). ADP-ribosylation factor regulates hPLD2.' *J. Biol. Chem.* 273(21), 12846–12852.
- Mahankali, M., Henkels, K. M., Alter, G., & Gomez-Cambronero, J. (2012). 'Identification of the catalytic site of phospholipase D2 (PLD2) newly described guanine nucleotide exchange factor activity.' *J. Biol. Chem.* 287(49), 41417–41431.

- Mahankali, M., Alter, G., & Gomez-Cambronero, J. (2014). 'Mechanism of Enzymatic Reaction and Protein-Protein Interactions of PLD from a 3D Structural Model.' *Cell. Signal.* 27(1), 69–81.
- Mahankali, M., Peng, H.-J., Henkels, K. M., Dinauer, M. C., & Gomez-Cambronero, J. (2011). 'Phospholipase D2 (PLD2) is a guanine nucleotide exchange factor (GEF) for the GTPase Rac2.' *Proceedings of the National Academy of Sciences* 108(49), 19617–19622.
- Malcolm, K. C., Elliott, C. M., & Exten, J. H. (1996). 'Evidence for Rho-mediated agonist stimulation of phospholipase D in Rat1 fibroblasts: Effects of Clostridium botulinum C3 exoenzyme.' *Journal of Biological Chemistry* 271(22), 13135–13139.
- Manciu, M. & Ruckenstein, E. (2004). 'Estimation of the available surface and the jamming coverage in the Random Sequential Adsorption of a binary mixture of disks.' *Colloids and Surfaces a-Physicochemical and Engineering Aspects* 232(1), 1–10.
- Manrai, A. K. & Gunawardena, J. (2008). 'The geometry of multisite phosphorylation.' *Biophys. J.* 95(12), 5533–5543.
- Marshall, C. B., Ho, J., Buerger, C., Plevin, M. J., Li, G.-Y., Li, Z., Ikura, M., & Stambolic, V. (2009). 'Characterization of the intrinsic and TSC2-GAP-regulated GTPase activity of Rheb by real-time NMR.' *Sci Signal* 2(55), ra3–ra3.
- Martin, T. F. (2001). 'PI(4,5)P(2) regulation of surface membrane traffic.' *Current Opinion in Cell Biology* 13(4), 493–499.
- Mazhab-Jafari, M. T., Marshall, C. B., Smith, M., Gasmi-Seabrook, G. M. C., Stambolic, V., Rottapel, R., Neel, B. G., & Ikura, M. (2010). 'Real-time NMR study of three small GTPases reveals that fluorescent 2'(3')-O-(N-methylanthraniloyl)-tagged nucleotides alter hydrolysis and exchange kinetics.' *J. Biol. Chem.* 285(8), 5132–5136.
- McLaughlin, S. & Aderem, A. (1995). 'The myristoyl-electrostatic switch: a modulator of reversible protein-membrane interactions.' *Trends Biochem. Sci.* 20(7), 272–276.
- McMahon, H. T. & Gallop, J. L. (2005). 'Membrane curvature and mechanisms of dynamic cell membrane remodelling.' *Nature* 438(7068), 590–596.
- Meshkini, A., Yazdanparast, R., & Nouri, K. (2011). 'Intracellular GTP level determines cell's fate toward differentiation and apoptosis.' *Toxicol. Appl. Pharmacol.* 253(3), 188–196.
- Michaelis, L. & Menten, M. L. (1913). 'Die kinetik der invertinwirkung.' *Biochem. Z* 49(333-369), 352.

- Milo, R. (2013). 'What is the total number of protein molecules per cell volume? A call to rethink some published values.' *Bioessays* 35(12), 1050–1055.
- Milo, R., Jorgensen, P., Moran, U., Weber, G., & Springer, M. (2010). 'BioNumbers—the database of key numbers in molecular and cell biology'. *Nucleic Acids Research* 38(Database issue), D750–D753.
- Morgan, C. P., Sengelov, H., Whatmore, J., Borregaard, N., & Cockcroft, S. (1997). 'ADP-ribosylation-factor-regulated phospholipase D activity localizes to secretory vesicles and mobilizes to the plasma membrane following N-formylmethionyl-leucyl-phenylalanine stimulation of human neutrophils'. *Biochemical Journal* 325(3), 581–585.
- Moritz, A., De Graan, P. N., Gispen, W. H., & Wirtz, K. W. (1992). 'Phosphatidic acid is a specific activator of phosphatidylinositol-4-phosphate kinase.' *J. Biol. Chem.* 267(11), 7207–7210.
- Nagle, J. F. & Tristram-Nagle, S. (2000). 'Structure of lipid bilayers.' *Biochim. Biophys. Acta* 1469(3), 159–195.
- Narayan, K. & Lemmon, M. A. (2006). 'Determining selectivity of phosphoinositide-binding domains.' *Methods* 39(2), 122–133.
- Nirody, J. & Rangamani, P. (2014). 'ATP concentration regulates enzyme kinetics'. *arXiv*. arXiv: 1406.2074v1 [q-bio.MN].
- Northup, J. K., Jian, X., & Randazzo, P. A. (2012). 'Nucleotide exchange factors: Kinetic analyses and the rationale for studying kinetics of GEFs.' *Cellular Logistics* 2(3), 140–146.
- Ocana, A., Vera-Badillo, F., Al-Mubarak, M., Templeton, A. J., Corrales-Sanchez, V., Diez-Gonzalez, L., Cuenca-Lopez, M. D., Seruga, B., Pandiella, A., & Amir, E. (2014). 'Activation of the PI3K/mTOR/AKT pathway and survival in solid tumors: systematic review and meta-analysis.' *PLoS ONE* 9(4), e95219.
- O'Hayre, M., Vázquez-Prado, J., Kufareva, I., Stawiski, E. W., Handel, T. M., Seshagiri, S., & Gutkind, J. S. (2013). 'The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer'. *Nat. Rev. Cancer* 13(6), 412–424.
- Okamura, S. & Yamashita, S. (1994). 'Purification and characterization of phosphatidylcholine phospholipase D from pig lung.' *Journal of Biological Chemistry* 269(49), 31207–31213.
- Oldham, W. M. & Hamm, H. E. (2008). 'Heterotrimeric G protein activation by G-protein-coupled receptors.' *Nat. Rev. Mol. Cell Biol.* 9(1), 60–71.



- Ortega, F., Acerenza, L., Westerhoff, H. V., Mas, F., & Cascante, M. (2002). 'Product dependence and bifunctionality compromise the ultrasensitivity of signal transduction cascades.' *Proc. Natl. Acad. Sci. U.S.A.* **99**(3), 1170–1175.
- Oude Weernink, P. A., Schmidt, M., & Jakobs, K. H. (2004). 'Regulation and cellular roles of phosphoinositide 5-kinases.' *Eur. J. Pharmacol.* **500**(1-3), 87–99.
- Pankov, R., Markovska, T., Antonov, P., Ivanova, L., & Momchilova, A. (2006). 'The plasma membrane lipid composition affects fusion between cells and model membranes.' *Chem. Biol. Interact.* **164**(3), 167–173.
- Paris, S., Béraud-Dufour, S., Robineau, S., Bigay, J., Antonny, B., Chabre, M., & Chardin, P. (1997). 'Role of protein-phospholipid interactions in the activation of ARF1 by the guanine nucleotide exchange factor Arno.' *J. Biol. Chem.* **272**(35), 22221–22226.
- Park, S. J., Itoh, T., & Takenawa, T. (2001). 'Phosphatidylinositol 4-phosphate 5-kinase type I is regulated through phosphorylation response by extracellular stimuli.' *Journal of Biological Chemistry* **276**(7), 4781–4787.
- Pasqualato, S., Ménétrey, J., Franco, M., & Cherfils, J. (2001). 'The structural GDP/GTP cycle of human Arf6.' *EMBO Rep* **2**(3), 234–238.
- Peng, X. & Frohman, M. A. (2012). 'Mammalian phospholipase D physiological and pathological roles.' *Acta Physiol (Oxf)* **204**(2), 219–226.
- Perez-Mansilla, B., Ha, V. L., Justin, N., Wilkins, A. J., Carpenter, C. L., & Thomas, G. M. H. (2006). 'The differential regulation of phosphatidylinositol 4-phosphate 5-kinases and phospholipase D1 by ADP-ribosylation factors 1 and 6.' *Biochim. Biophys. Acta* **1761**(12), 1429–1442.
- Pertile, P., Liscovitch, M., Chalifa, V., & Cantley, L. C. (1995). 'Phosphatidylinositol 4,5-bisphosphate synthesis is required for activation of phospholipase D in U937 cells.' *J. Biol. Chem.* **270**(10), 5130–5135.
- Pettitt, T. R., Martin, A., Horton, T., Liou, C., Lord, J. M., & Wakelam, M. J. (1997). 'Diacylglycerol and phosphatidate generated by phospholipases C and D, respectively, have distinct fatty acid compositions and functions. Phospholipase D-derived diacylglycerol does not activate protein kinase C in porcine aortic endothelial cells.' *J. Biol. Chem.* **272**(28), 17354–17359.
- Rameh, L. E., Toliás, K. F., Duckworth, B. C., & Cantley, L. C. (1997). 'A new pathway for synthesis of phosphatidylinositol-4,5-bisphosphate.' *Nature* **390**(6656), 192–196.
- Randazzo, P. A. (1997). 'Functional interaction of ADP-ribosylation factor 1 with phosphatidylinositol 4,5-bisphosphate.' *J. Biol. Chem.* **272**(12), 7688–7692.

- Randazzo, P. A., Jian, X., Chen, P.-W., Zhai, P., Soubias, O., & Northup, J. K. (2013). 'Quantitative analysis of guanine nucleotide exchange factors (GEFs) as enzymes'. *Cellular Logistics* 3(1), e27609.
- Randazzo, P. A. & Kahn, R. A. (1994). 'GTP hydrolysis by ADP-ribosylation factor is dependent on both an ADP-ribosylation factor GTPase-activating protein and acid phospholipids.' *Journal of Biological Chemistry* 269(14), 10758–10763.
- Rao, V. D., Misra, S., Boronenkov, I. V., Anderson, R. A., & Hurley, J. H. (1998). 'Structure of type IIbeta phosphatidylinositol phosphate kinase: a protein kinase fold flattened for interfacial phosphorylation.' *Cell* 94(6), 829–839.
- Rawicz, W., Olbrich, K. C., McIntosh, T., Needham, D., & Evans, E. (2000). 'Effect of chain length and unsaturation on elasticity of lipid bilayers.' *Biophys. J.* 79(1), 328–339.
- Riebeling, C., Morris, A. J., & Shields, D. (2009). 'Phospholipase D in the Golgi apparatus.' *Biochim. Biophys. Acta* 1791(9), 876–880.
- Rojas, A. M., Fuentes, G., Rausell, A., & Valencia, A. (2012). 'The Ras protein superfamily: evolutionary tree and role of conserved amino acids.' *J. Cell Biol.* 196(2), 189–201.
- Rojas, R. J., Kimple, R. J., Rossman, K. L., Siderovski, D. P., & Sondek, J. (2003). 'Established and emerging fluorescence-based assays for G-protein function: Ras-superfamily GTPases.' *Comb. Chem. High Throughput Screen.* 6(4), 409–418.
- Rudge, S. A. & Wakelam, M. J. O. (2009). 'Inter-regulatory dynamics of phospholipase D and the actin cytoskeleton.' *Biochim. Biophys. Acta* 1791(9), 856–861.
- Rusinova, R., Hobart, E. A., Koeppe, R. E., & Andersen, O. S. (2013). 'Phosphoinositides alter lipid bilayer properties.' *J. Gen. Physiol.* 141(6), 673–690.
- Sarri, E., Pardo, R., Fensome-Green, A., & Cockcroft, S. (2003). 'Endogenous phospholipase D2 localizes to the plasma membrane of RBL-2H3 mast cells and can be distinguished from ADP ribosylation factor-stimulated phospholipase D1 activity by its specific sensitivity to oleic acid.' *Biochem. J.* 369(Pt 2), 319–329.
- Sasaki, T., Takasuga, S., Sasaki, J., Kofuji, S., Eguchi, S., Yamazaki, M., & Suzuki, A. (2009). 'Mammalian phosphoinositide kinases and phosphatases.' *Prog Lipid Res* 48(6), 307–343.
- Schleifer, L. S., Kahn, R. A., Hanski, E., Northup, J. K., Sternweis, P. C., & Gilman, A. G. (1982). 'Requirements for cholera toxin-dependent ADP-ribosylation of the purified regulatory component of adenylate cyclase.' *Journal of Biological Chemistry* 257(1), 20–23.

- Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., & Schomburg, D. (2013). 'BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA.' *Nucleic Acids Research* 41(Database issue), D764–72.
- Sciorra, V. A. (2002). 'Dual role for phosphoinositides in regulation of yeast and mammalian phospholipase D enzymes'. *J. Cell Biol.* 159(6), 1039–1049.
- Sciorra, V. A., Rudge, S. A., Prestwich, G. D., Frohman, M. A., Engebrecht, J., & Morris, A. J. (1999). 'Identification of a phosphoinositide binding motif that mediates activation of mammalian and yeast phospholipase D isoenzymes.' *EMBO J.* 18(21), 5911–5921.
- Seixas, E., Barros, M., Seabra, M. C., & Barral, D. C. (2013). 'Rab and Arf proteins in genetic diseases.' *Traffic* 14(8), 871–885.
- Selvy, P. E., Lavieri, R. R., Lindsley, C. W., & Brown, H. A. (2011). 'Phospholipase D: enzymology, functionality, and chemical modulation.' *Chem. Rev.* 111(10), 6064–6119.
- Simon, M. I., Strathmann, M. P., & Gautam, N. (1991). 'Diversity of G proteins in signal transduction.' *Science* 252(5007), 802–808.
- Skippen, A., Jones, D. H., Morgan, C. P., Li, M., & Cockcroft, S. (2002). 'Mechanism of ADP ribosylation factor-stimulated phosphatidylinositol 4,5-bisphosphate synthesis in HL60 cells.' *J. Biol. Chem.* 277(8), 5823–5831.
- Stace, C. L. & Ktistakis, N. T. (2006). 'Phosphatidic acid- and phosphatidylserine-binding proteins.' *Biochim. Biophys. Acta* 1761(8), 913–926.
- Stace, C., Manifava, M., Delon, C., Coadwell, J., Cockcroft, S., & Ktistakis, N. T. (2008). 'PA binding of phosphatidylinositol 4-phosphate 5-kinase.' *Adv. Enzyme Regul.* 48(1), 55–72.
- Stanley, R. J. (2011). 'Signalling through G-proteins & phospholipids: Modelling the properties of an unusual self-reinforcing network motif'. MRes thesis.
- Stephen, A. G., Esposito, D., Bagni, R. K., & McCormick, F. (2014). 'Dragging ras back in the ring.' *Cancer Cell* 25(3), 272–281.
- Strogatz, S. H. (2014). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Studies in Nonlinearity. Westview Press.
- Su, W., Chen, Q., & Frohman, M. A. (2009). 'Targeting phospholipase D with small-molecule inhibitors as a potential therapeutic approach for cancer metastasis.' *Future Oncol* 5(9), 1477–1486.

- Sung, T. C., Altshuler, Y. M., Morris, A. J., & Frohman, M. A. (1999a). 'Molecular analysis of mammalian phospholipase D2.' *J. Biol. Chem.* 274(1), 494–502.
- Sung, T. C., Zhang, Y., Morris, A. J., & Frohman, M. A. (1999b). 'Structural analysis of human phospholipase D1.' *J. Biol. Chem.* 274(6), 3659–3666.
- Terui, T., Kahn, R. A., & Randazzo, P. A. (1994). 'Effects of acid phospholipids on nucleotide exchange properties of ADP-ribosylation factor 1. Evidence for specific interaction with phosphatidylinositol 4,5-bisphosphate.' *J. Biol. Chem.* 269(45), 28130–28135.
- Thomson, M. & Gunawardena, J. (2009). 'Unlimited multistability in multisite phosphorylation systems'. *Nature* 460(7252), 274–277.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. H. (2009). 'Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems'. *Journal of The Royal Society Interface* 6(31), 187–202.
- Traut, T. W. (1994). 'Physiological concentrations of purines and pyrimidines'. *Mol. Cell. Biochem.* 140(1), 1–22.
- Valencia, A., Kjeldgaard, M., Pai, E. F., & Sander, C. (1991). 'GTPase domains of ras p21 oncogene protein and elongation factor Tu: analysis of three-dimensional structures, sequence families, and functional sites.' *Proc. Natl. Acad. Sci. U.S.A.* 88(12), 5443–5447.
- Van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). 'The NumPy Array: A Structure for Efficient Numerical Computation'. *Computing in Science & Engineering* 13(2), 22–30.
- Van Dongen, C. J., Zwiers, H., & Gispen, W. H. (1984). 'Purification and partial characterization of the phosphatidylinositol 4-phosphate kinase from rat brain.' *Biochem. J.* 223(1), 197–203.
- Van Hecke, M. (2010). 'Jamming of soft particles: geometry, mechanics, scaling and isostaticity'. *J. Phys.: Condens. Matter* 22(3), 033101.
- Van Meer, G. & de Kroon, A. I. P. M. (2011). 'Lipid map of the mammalian cell.' *J. Cell. Sci.* 124(Pt 1), 5–8.
- Van Meer, G., Voelker, D. R., & Feigenson, G. W. (2008). 'Membrane lipids: where they are and how they behave.' *Nat. Rev. Mol. Cell Biol.* 9(2), 112–124.
- Vetter, I. R. & Wittinghofer, A. (2001). 'The guanine nucleotide-binding switch in three dimensions.' *Science* 294(5545), 1299–1304.
- Vigil, D., Cherfils, J., Rossman, K. L., & Der, C. J. (2010). 'Ras superfamily GEFs and GAPs: validated and tractable targets for cancer therapy?' *Nat. Rev. Cancer* 10(12), 842–857.

- Vinggaard, A. M. & Hansen, H. S. (1995). 'Characterization and partial purification of phospholipase D from human placenta'. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **1258**(2), 169–176.
- Voss, N. R. & Gerstein, M. (2010). '3V: cavity, channel and cleft volume calculator and extractor.' *Nucleic Acids Research* **38**(Web Server issue), W555–62.
- Wang, M., Weiss, M., Simonovic, M., Haertinger, G., Schrimpf, S. P., Hengartner, M. O., & von Mering, C. (2012). 'PaxDb, a database of protein abundance averages across all three domains of life.' *Mol. Cell Proteomics* **11**(8), 492–500.
- Wang, X., Devaiah, S. P., Zhang, W., & Welti, R. (2006). 'Signaling functions of phosphatidic acid.' *Prog Lipid Res* **45**(3), 250–278.
- Wiener, N. (1965). *Cybernetics Or Control and Communication in the Animal and the Machine*. MIT Press.
- Yin, H. L. & Janmey, P. A. (2003). 'Phosphoinositide regulation of the actin cytoskeleton.' *Annu. Rev. Physiol.* **65**(1), 761–789.
- Young, A., Lyons, J., Miller, A. L., Phan, V. T., Alarcón, I. R., & McCormick, F. (2009). 'Ras signaling and therapies.' *Adv. Cancer Res.* **102**, 1–17.
- Zhang, Y. & Du, G. (2009). 'Phosphatidic acid signaling regulation of Ras superfamily of small guanosine triphosphatases.' *Biochim. Biophys. Acta* **1791**(9), 850–855.