Exploring Design Principles of Cellular Information Processing

by

Song Feng

Thesis Submitted to the University of Warwick

in partial fulfilment of the requirements

for admission to the degree of

Doctor of Philosophy

School of Life Sciences

November 2015

WARWICK
THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

Gratefully Acknowledges

University of Warwick and University of Exeter for 3 yrs funding.

Very very grateful to Prof. Orkun Soyer.

Prof. Peter Swain, Dr. Elisenda Feliu and Dr. Julien Ollivier.

HPC server "Zeus" in University of Exeter, OSSLab computing server.

In particular to my dearest family, my wife Xue Jiang , my parents and sister.

OSSLab group members.

Discussions with Dr. Jin Yang, Prof. Vincent Danos, Prof. Carstern Wiuf.

Support and encouragement from friends.

Inspiration from many scientists: Charles Darwin, Richard Feynman, etc.

Enlightenment from the philosopher Wang Yangming.

This thesis was typeset with LaTeX 2_ε* by the author.

*LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The style package *warwickthesis* was used.

Declarations

I hereby declare that this dissertation entitled "Exploring Design Principles of Cellular Information Processing" is an original work and has not been submitted for a degree or diploma or other qualification at any other University.

Chapter 1 introduces the main objects and the backgrounds for this study. This chapter provides information obtained from the literature as referenced to.

Chapter 2 explains the particular computational platform I constructed to evolve and simulate the signaling networks. Chapter 3 investigates the emerged mechanisms for ultrasensitive and adaptive networks as well as design of small network motif that could modulating different response dynamics based on principles discovered from evolutionary simulations. Chapter 2 and Chapter 3 are based on the following published papers respectively:

- "BioJazz: *in silico* evolution of cellular networks with unbounded complexity using rule-based modelling.", Song Feng, Julien F. Olliver, Peter S. Swain, Orkun S. Soyer, *Nucleic Acids Research*, pp. 1-15. ISSN 0305-1048 (2015)

  Author contributions: S.F., J.F.O., P.S.S. and O.S.S. designed research; S.F. performed research (J.F.O implemented encoding of rule-based models, S.F. implemented evolutionary algorithm, I/O and carried out simulations); S.F. and O.S.S. analysed data; S.F., J.F.O., P.S.S. and O.S.S. wrote the paper.

- "Proteins sequestration as a tuning point in controlling response dynamics of signalling networks.", Song Feng, Julien F. Olliver, Orkun S. Soyer, *PLoS Comput. Biol.* (Accepted)

  Author contributions: S.F. and O.S.S. designed research; S.F. performed research; S.F. and O.S.S. analysed the data; S.F., J.F.O. and O.S.S. wrote the paper.

Chapter 4 talks about a particular emergent dynamics, multistability, as byproduct of evolution on ultrasensitivity as well as the discovery of novel basic network units that enables multistationary dynamics in evolved signalling networks. At the end, it also attempts to discuss the searching of boundaries between monostationary and multistationary signalling networks in order to approach the sufficient

conditions (design principles) on multistable cellular networks. This chapter is based on unpublished data from work mainly by Song Feng and Orkun S. Soyer.

Chapter 5 is about detail investigation on one of the novel bistable motif discovered from evolved networks. Chapter 5 is based on the following working manuscript:

- "Core signalling motif displaying multistability through multi-state enzymes.", Song Feng, Meritxell Sáez, Carstern Wiuf, Elisenda Feliu, Orkun S. Soyer, Submitted to*Proc. Natl. Acad. Sci. USA*

  Author contributions: S.F., C.W., E.F. and O.S.S. designed research; S.F., M.S.C. and E.F. performed research (S.F. and C.W. performed the numerical analysis, S.F., M.S.C. and E.F. performed analytical proofs); S.F., M.S.C., C.W., E.F. and O.S.S. analysed results; S.F., M.S.C., C.W., E.F. and O.S.S. wrote the manuscript.

Chapter 6 is the conclusion for my PhD project. It concludes the results generated from *in silico* evolution and discovered new design principles. I also discussed a bit about ideas and thinking, inspired by combining my PhD study with previous literature, about evolution of biological networks.

Abstract

As a summary, this work attempts to explore and uncovered design principles of certain dynamics of cellular networks by combining evolution *in silico* with rule-based modelling approach.

Biological systems exhibit complex dynamics, due to the complex interactions in the intra- and inter- cellular biochemical reaction networks. For instance, signalling networks are composed of many enzymes and scaffolding proteins which have combinatorial interactions. These complex systems often generate response dynamics that are essential for correct decision-makings in cells. Especially, these complex interactions are results of long term of evolutionary process. With such evolutionary complexity, systems biologists aim to decipher the structure and dynamics of signalling and regulatory networks underpinning cellular responses; synthetic biologists can use this insight to alter existing networks or engineer *de novo* ones. Both tasks will benefit from an understanding of which structural and dynamic features of networks can emerge from evolutionary processes, through which intermediary steps these arise, and whether they embody general design principles. As natural evolution at the level of network dynamics is difficult to study, *in silico* evolution of network models can provide important insights.

However, current tools used for *in silico* evolution of network dynamics are limited to *ad hoc* computer simulations and models. In my PhD study, with collaborators I construct the BioJazz, an extendable, user-friendly tool for simulating the evolution of dynamic biochemical networks. Unlike previous tools for *in silico* evolution, BioJazz allows for evolution of cellular networks with theoretically unbounded complexity by combining rule-based modelling with an encoding of networks that is akin to a genome. BioJazz can be used to implement biologically realistic selective pressures, and allows exploration of the space of network architectures and dynamics that implement prescribed physiological functions. It is provided as an open-source tool to facilitate its further development and use. I use this tool to explore the possible biochemical designs for signalling networks displaying ultrasensitive and adaptive response dynamics. By running evolutionary simulations mimicking different biochemical scenarios, we find that enzyme sequestration emerges as a key biochemical mechanism for both dynamics. Detailed analysis of these evolved networks revealed that enzyme sequestration enables both ultrasensitive and adaptive

response dynamics. I verified this proposition by designing a generic model of a signalling cycle, featuring two enzymes and a sequestering (scaffold) protein. This simple system is capable of displaying both ultrasensitive and adaptive response dynamics, even more interestingly modulating the system switching between two response dynamics through perturbing the scaffold protein. These results show that enzyme sequestration can be exploited by evolution so to generate diverse response dynamics in signalling networks.

From evolutionary simulations towards ultrasensitivity, bistable dynamics emerged as an alternative solution. On one hand, inspired by such results I used the fitness function as an objective function combined with different constraints to design and optimise bistable signalling networks with completely new structure and mechanism. Studying designed bistable signalling network explicates how such bistable network can be experimentally implemented. On the other hand, from studying the evolved bistable networks allosteric enzymes catalysing futile cycles appear to be a new mechanism of bistability in signalling networks. Furthermore, one of the smallest bistable signalling motifs is derived. This motif is composed of one kinase protein with two distinct conformational states and one substrate subject to phosphorylation by the kinase and auto-dephosphorylation reactions. The sufficient and necessary condition on parameters, with which the signalling motif displays bistable response dynamics, is analytically defined. By expanding the systems with more kinases, unlimited multistability emerges with potentials of implementing complex logic gates and cell state transitions. Further exploring the discovered and natural signalling networks implies shared design patterns. Motivated by searching structural boundaries between monostationary and multistationary networks, I performed algorithmic searching of multistationary signalling networks intending to find the sufficient structural conditions for multistationarity in signalling networks.

Key words: design principles, information processing, signalling networks, ultrasensitivity, adaptation, bistability, synthetic biology, *in silico* evolution, response plasticity.

Glossary

Bipartite graph	a graph whose vertices can be divided into two disjoint sets.
Bistability	a dynamical system has two stable equilibrium states.
Cross-talk	one or more components from one signalling pathway affects others.
CRNT	chemical reaction network theory which models and studies the dynamical behaviour of chemical systems.
DSR graph	directed species-reaction graph, a signed, labelled, directed bipartite graph derived from chemical reaction networks.
Futile cycle	also known as substrate cycle, where two metabolic pathways run simultaneously in opposite directions and have no overall effect other than to dissipate energy in the form of heat.
Multistability	a property of having multiple stable equilibrium points in the vector space spanned by the states in a dynamical system.
Multistationarity	the attribute of numerous systems to possess more than one stable states.
Ultrasensitivity	an output response that is more sensitive to stimulus change than the hyperbolic Michaelis-Menten response.
Retroactivity	a phenomenon that the behavior of an upstream component is affected by the connection to a downstream component.

Chapter 1

Introduction and background

1.1 Evolutionary Systems and Synthetic Biology

Biological systems are complex. The complexity derives from combinatorial interactions between the components, the building blocks of biological systems, in multiple scales [1–5]. For instance, complex interactions between amino acids give rise to complex energy landscapes which result diverse protein functions such as allosteric regulation and catalytic activities [2, 6–9]; interactions between multi-domain proteins compute signals from fluctuating environments into reliable cellular decisions [10, 11]; gene regulatory networks of protein-DNA interactions determine the progression of cell fates [12–14]; and metabolic interactions at inter-cellular scale define the structure and dynamics of microbial communities [15–17]. All these complex interactions can not be understood without systematically investigating the dynamics of corresponding systems[11, 17–19].

Systems biology emerged as cutting edge area to deal with not only large amount of biological data but more importantly the complexity of biological systems discovered from accumulated data and knowledge [20–22]. At the centre of it, computational modelling of the biological systems is more than an aid to interpret and integrate biological data, but rather a necessity to discover and formalise

the principles governing the complex dynamics of biological systems at different scales [23–26]. In parallel, following the idea of "What I cannot create, I do not understand.", synthetic biologists dedicate to building *de novo* biological systems in hope of understanding them [15, 27–29]. In building large scale biosystems, computational predictions or theoretical guidance are tremendously helpful by reducing the searching space and directing the design [30–32]. Therefore, a computational approach is indispensable in understanding and mastering the design principles of biological systems.

Furthermore, the complexity of biological systems results from evolution [33]. The evolutionary innovations are embedded in genotype-phenotype mapping in biological systems and they emerged by long time of tinkering and optimisation under fluctuating environments [34–37]. It is reasonable to investigate the design principles in biological systems from an evolutionary perspective, not only because the studies makes no sense without evolutionary insights but also because that evolutionary studies enable us discover design principles that have not been found in natural systems yet. Even from a practical point of view, it is necessary to make accurate predictions about functional rules of proteins and the effects of modifying interactions between them so that they can improve control of natural biosystems and enable rational design of *de novo* biosystems.

In this study, I combine computational modelling and an evolutionary approach to explore design principles in one of the most important yet complicated phenomenon — the information processing in cells.

1.2 Information processing in cells

Biological cells employ complex regulatory systems to detect the states of environments they sit in, process such information into cellular decision and response accordingly in order to survive. In this regulatory systems, the signalling network is

specific for information processing so that cells utilise it to transform extracellular signals into cellular output. Many interesting and essential physiological behaviours and responses are determined by the proper functioning of signalling system in cells [38].

An astonishing fact is that the signal transduction system does not simply transmit signals rather it integrates, process and encode different external signals so that it give rise to appropriate cellular responses that guarantee the cell adapt to fluctuating environments. These cellular responses are pivotal cellular decisions determined by the temporal and spatial dynamics of signal transduction system [11]. Therefore, in order to understand the complex regulations, understanding the underlying principles and biochemical mechanisms is a necessity.

1.2.1 The structure: interconnected networks

As naturally designed, signalling networks are composed of signalling proteins such as receptors, adaptor proteins, kinases, transcription factors and second messenger like calcium and nitric oxide. Signalling proteins usually adopt conformational changes to carry out binding interactions or enzymatic reactions and consequently affect the conformation and dynamics of proteins. These signals are detected, encoded, integrated and transformed as perturbation at activity of transcription factors, through which cell responses in form of altering gene expression or as modification of molecular machines so that cell responses as, for instance, movements, neural action potential [10, 11, 38].

The conventional concept of linear signalling pathways has been replaced by the emerging viewpoint of combinatorial networks formed with interconnecting proteins with multi-domain and multi-site structure. Many signalling molecules share similar downstream or upstream signalling pathways, which brings the cross-talks between different signalling "pathways" [39]. As a result, the signal transduction systems appear as complex interacting networks between many signalling proteins

with various structures and domains.

1.2.2 The dynamics: diverse information processing functions

Different structures of biochemical reaction networks give rise to diverse functions that compute the input signals into different response dynamics [11]. There are many signalling dynamics discovered to be important and ubiquitous in biological systems, such as ultrasensitivity [40–45], adaptation [46–49], multistability [50–54], oscillation [55–58], pulsatile [59–61]. Among these signalling dynamics, ultrasensitivity, bistability and adaptation are the most fundamental dynamics in building more complex dynamical behaviour in cells [40, 62, 63].

Ultrasensitivity

The ultrasensitive or switch-like dynamics in biological systems was firstly characterised in the hemoglobin oxygen binding. The curve of oxygen-bound hemoglobin (response) to the oxygen concentration level at steady states is sigmoidal rather than proportional. Specifically, ultrasensitivity is a nonlinear information processing function where a small fraction change in the input is amplified into a large fraction of output response and form a distinct threshold. When the signal changes is at region much lower or higher than the threshold, the response has little change, while the signal changes is near threshold, the output response changes dramatically (Figure 1.1A). The ultrasensitive response curves were later found in various biological processes and playing significant roles. For instance, hemoglobin can transport more proportion of oxygen with sigmoidal binding curve than with hyperbolic ones; mating decision in yeast allows cells to filter signals to avoid inappropriate commitments with switch-like response to critical signal threshold, it makes the signalling systems robust to variations in concentrations of pheromone [43]; the switch-like dynamics in phosphorylation of isocitrate dehydrogenase can amplify the signals in metabolic regulation [64]. Such amplification of signals can be quantified by the

response coefficient (R) calculated with the input level (I) and output level (O) as well as their changes respectively (ΔI and ΔO):

$$R = \lim_{\Delta I \to 0} \frac{\Delta O/O}{\Delta I/I} = \frac{\mathrm{d}O/O}{\mathrm{d}I/I} = \frac{\mathrm{d}\ln O}{\mathrm{d}\ln I} \tag{1.1}$$

The higher the value of R, the higher the sensitivity.

Adaptation

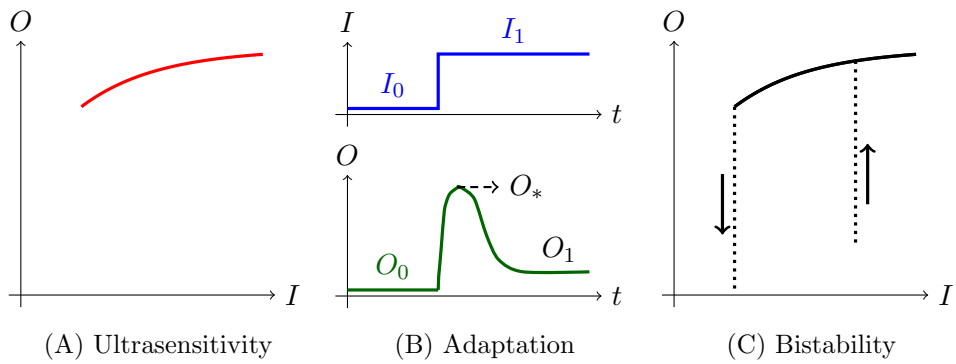
Biological systems respond to input signals and regulate cellular states not just by amplifying signals but also by adapting to them [47]. Biochemical adaptation refers to the function that many signalling systems return to their pre-stimulated state after responding to a sustained stimulus (Figure 1.1B). A mathematical description for adaptation can be quantified with two characteristic terms: adaptive sensitivity (A_{sens}) to the input perturbation and adaptive precision (A_{prec}) [65, 66], where the sensitivity is defined as the maximum of response to the sustained input stimulus and can be calculated by (Figure 1.1B):

$$A_{sens} = \left| \frac{(O_* - O_0)/O_0}{(I_1 - I_0)/I_0} \right|, \tag{1.2}$$

and the precision can be calculated by:

$$A_{prec} = \left| \frac{(O_1 - O_0)/O_0}{(I_1 - I_0)/I_0} \right|^{-1}. \tag{1.3}$$

A so called "perfect adaptation" emerges if the output response returns exactly to the pre-stimulated state ($O_1 = O_0$). Adaptation is commonly found in sensory systems and other signalling systems to either accurately detect changes in input [46, 67, 68] or maintain homeostatic condition when presented with perturbations [69, 70].



(A) Ultrasensitivity (B) Adaptation (C) Bistability

Figure 1.1: Three example response dynamics in signalling systems. Axes labelled with I represent level of input signal, the ones with O represent level of output response. (a) and (c) are showing the steady state plot of input signal and output response at steady states. (b) is showing the temporal dynamics with x-axis labelled with t representing time, I_0 and O_0 represent pre-stimulus level of input signal and output response, I_1 and O_1 represent respective levels after stimulus, O_* represents the level of output response with largest deviation from its pre-stimulus level.

Bistability

Some biological systems exhibit "all-or-none" dynamics [42, 71] and on top of it some systems also display irreversible fate induction process [72]. This particular response dynamics is due to bistability, which is very much similar to ultrasensitivity that the fast switching from one state, for example low level of response, to another state (high level of response) creates threshold at certain input signal levels. However, there are some distinctive features in bistable dynamics. The threshold is not continuous rather discrete and the switching between different levels of response is hysteretic (Figure 1.1C). This discrete threshold enable biological systems implement boolean logic [50] while the hysteresis implements biochemical memory [73, 74]. The hysteresis in bistable systems can potentially facilitates irreversible commitments in cell fate determination [75]. Also, it is implied that the bistability is the key mechanism to enable heterogeneity and bet-hedging strategy in population of cells [76–78].

Other response dynamics

Beside the response dynamics discussed above, many other response dynamics are crucial for biological systems to make appropriate decisions under certain environments. For instance, cells change their states according to the genetic program and environment in development, which requires complex transitions among many different states. Multistability, the ability of having multiple stable steady states in system dynamics, is the key player in such fate determinations. In addition, Biochemical oscillation is one the most important functions in many contexts like metabolism, signalling and development. Oscillations allow cells behaving periodically, especially when exposed periodic signals, cells use oscillators to regulate their behaviours and better adapt the environments [79–82]. Besides circadian clocks, oscillation in signalling networks can also facilitate digital activation and dynamical control of cellular behaviour [83, 84]. With more interactions involved in the network, the system can potentially generate chaotic behaviour. Such chaotic behaviours has been observed in many different cellular systems [11, 24, 85, 86].

1.2.3 Design principles: the biochemical mechanisms

By quantitative study on biological systems, some biochemical mechanisms were proposed to explain the interesting response dynamics in these systems. Such biochemical mechanisms were further formalised into design principles of corresponding response dynamics.

By studying the hemoglobin oxygen binding system, cooperative binding was uncovered as molecular mechanism for sigmoidal response curve. Several theoretical models were proposed to understand ultrasensitivity in allosteric regulated systems [87–90]. One of the most well-known is Monod-Wyman-Changeux model (MWC model, also known as the *concerted model* or *symmetry model*). The main idea of MWC model is that proteins exhibit different interconvertable states which can be regulated via interaction of protein's subunits (or domains) with other molecules

[87, 88, 91]. The ratio of different conformational states is determined by thermodynamic equilibrium. This model is a formalisation of allosteric regulation which was later found widely existing in biochemical systems and provide complex interactions between macro-molecules [88]. Comparably, another model was proposed to explain the allosteric regulation when subunits in the protein are not connected in such a way that conformational change of one induces similar change in the others [90]. In this model, all subunits are not necessarily displaying the same conformational state where substrate-binding at one subunit only slightly changes conformation of other subunits rather than propagates the conformational change to adjacent subunits. This model is called sequential model. Most allosteric effects can be explained by these two models. Both models postulate that allosteric protein exhibits in one of two distinct conformational states, tensed (T) or relaxed (R), and that these two states has different affinities and activities towards their substrates. However, the MWC model is more appropriate to explain the allosteric regulations in multi-domain proteins, since as a tightly connected entity, conformational propagation is inevitable in a folded protein. Meanwhile, treatment on the sequential model in large scale signalling networks is more complex by potentially introducing more parameters comparing to MWC model. Therefore, in this study I primarily use the MWC model to describe the allosteric regulation in signalling networks.

Besides allosteric regulation, Goldbeter and Koshland found that within an enzymatic reaction cycle, for example a phosphorylation-dephosphorylation cycle, when enzymes are saturated by the substrate, the system displays ultrasensitivity [92]. This particular mechanism is termed as zero-order sensitivity (or Goldbeter-Koshland kinetics). This mechanism is limited in the condition that enzymes has much lower concentration than the substrate so that the enzymes is near saturated.

In the following reactions:

$$\begin{aligned} \mathrm{S} + \mathrm{K} &\underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{KS} \xrightarrow{k_3} \mathrm{S_p} + \mathrm{K} \\ \mathrm{S_p} + \mathrm{P} &\underset{k_5}{\overset{k_4}{\rightleftharpoons}} \mathrm{PS_p} \xrightarrow{k_6} S + \mathrm{P} \end{aligned} \tag{1.4}$$

there are two parameters determining the saturation level of enzymes: $K_1 = \frac{k_2+k_3}{k_1[S_{tot}]}$ and $K_2 = \frac{k_5+k_6}{k_4[S_{tot}]}$, where S_{tot} is the total concentration of substrate. When K_1 and K_2 become smaller, the enzymes (K and P) becomes more saturated by substrate (S) and the system is consequently more ultrasensitive.

Build on the work of MWC model and zero-order sensitivity, more biochemical mechanisms for ultrasensitivity were discovered and studied, like signalling cascade (i.e. multiple steps of signalling cycles) [62], substrate competition[93, 94], sequestration[95–97], positive feedback[98–101]. In particular, studies showed that positive feedback either in signalling cycles or combined with signalling cascades can induce bistable response dynamics [62, 71, 72, 102, 103]. Also, mathematical proofs identified that positive feedback loops is necessary condition for generating bistability in chemical reaction networks [104, 105]. However, the positive feedback is not always manifest from the reactions displayed. As an example, double phosphorylation in mitogen-activated protein kinase (MAPK) cascade endow the system capacity to generate bistability [53, 106]. Further studies on multisite phosphorylation systems showed the complex structural conditions for multistationarity.

Adaptive response dynamics widely exists in biological systems functioning as gradient detection and homeostasis controller. The most prominent mechanism for adpative dynamics is negative feedback. Negative feedback has been well characterised and widely applied to control engineering as one of the most important engineering principles. One of the most commonly studied signalling system with adaptive response dynamics is chemotaxis in bacteria. In chemotaxis, the systems utilise negative feedback loop to achieve adaptive response. Furthermore, elabo-

rative searching all possible networks at small scale uncovered only two solutions: negative feedback loop with a buffering node and incoherent feedforward loop with a proportioner node [66].

Design principles for other more complex dynamics are also derived by studying interlinked positive and negative feedback loops as well as other complex interactions. The oscillatory response dynamics can be designed by implementing feedback loops with time delay [86]; multistability exists in the interlinked positive feedback loops and multi-domain histidine kinase systems [40, 50, 107, 108].

These mechanisms and principles are discovered by investigating recurring reaction patterns in cellular networks. The design principles requiring positive and negative feedback loops receive much appreciation with aid from computational and mathematical analysis. Especially, mathematical modelling played crucial roles in quantifying, deducing and formalising those mechanisms with verbally elaborating biochemical details. Therefore, I also take advantage of mathematical and computational modelling as the tool to study the information processing in cells.

1.3 Solutions in evolutionary landscapes

To capture the complex dynamics and uncover corresponding design principles, it is necessary to develop powerful tools such as realistic yet still executable models of large scale biological systems, and learn the lessons not only from engineered biological systems but also from those designed by evolution. Previously, various approaches, such as experimental characterisation, bioinformatic analysis, and mathematical modelling of recurring motifs in natural systems, provided many insights on the design principles [80, 86, 109–115], however the complexity of signalling systems is derived from evolutionary processes, understanding how such complex dynamics emerge from different structures and exploring potentially undiscovered design principles are not trivial [24, 116]. It is a notoriously difficult inverse problem to

characterise what design principles determine emergence of complex dynamics from evolution of signalling systems [117].

One approach would be systematic characterisation and comparison of components and their interactions in different species. This approach provides substantial information of various large scale signalling networks, it provides important clues of differences in topology and dynamics as well. For instance, comparative analysis of different prokaryotic genomics showed that network structures and response dynamics are diversified in chemotactic systems of various species [118]. The conservation and variations discovered shed light on the origin and evolution of chemotactic system. It suggests that evolutionary study of networks with defined response dynamics could be substantially helpful. The conserved features in evolved networks are much prone to be design principles for the selected response dynamics. With such knowledge and information, an alternative approach could be applying computational modelling and *in silico* evolution to forwardly understand the emergence properties of signalling networks and, more importantly, explore various design principles emerged from evolutionary processes. Previously, evolutionary *in silico* approaches has been applied to metabolic networks, gene regulatory networks and signalling networks [36, 119–121][122]. These studies provide many insights about network evolution, like robustness, evolvability, complexity and modularity, however how those specific systems dynamics emerge from evolutionary processes is much unknown. Some attempts are made by evolutionary design of gene regulatory networks with oscillatory and bistable dynamics, both known and novel design patterns for bistable gene networks emerged [123], in which the novel design was later implemented experimentally [95].

Following this line, evolution *in silico* is not limited to studying the evolution of signalling networks but also can be used to design and optimise cellular networks to achieve specific dynamics [123, 124]. Such promising applications in computer-aided design of biological systems is also one of the main objectives in synthetic

biology [32, 125, 126]. Previous studies on complex response dynamics show that some of those information processing functions such as ultrasensitivity, adaptive response and bistability are important in achieving complex functions like oscillation, homeostasis, multistability [40, 62, 127–132]. In this thesis, I primarily focus on exploring design principles of those simple response dynamics in signalling networks, namely ultrasensitivity, adaptation, bistability.

1.4 The challenges

The information processing in cells integrates multiple inputs and produce multiple outputs. In between is the networks of interacting molecules, most of the molecules are proteins which consist of multiple domains and exhibit different conformational states. The complexity of information processing emerged from such combinatorial interactions in the networks.

One of the main obstacles is combinatorial complexity due to the exceedingly high number of micro-states that grows exponentially when increasing the number of molecules and interactions in the network [116]. Such combinatorial variety was normally ignored in previous evolution *in silico* studies. Another challenge is how to encode signalling network such that *in silico* evolution of the network is close to open-ended fashion, which means expansion of signalling networks in the evolutionary processes is, at least in theory, unbounded. In order to overcome both challenges, new approaches and methods need to be adopted and/or invented.

1.5 Summary of contributions

In my PhD study, I combined the evolutionary *in silico* with rule-based modelling of cellular networks and applied this computational platform to explore evolutionary design principles of signalling networks.

- I developed a computational platform to evolve rule-based models of cellu-

lar networks. The computational platform is the first computational program that addresses both the multiple domain structure of proteins and theoretically unbounded complexity of cellular networks in evolutionary process. The encoding of rule-based models into binary string is analogous to genome sequences so that mutations are less *ad hoc*. I also implemented different fitness functions to evolve signalling networks toward desired response dynamics. Although the platform is still limited by the computational power when the networks become arbitrarily large and not completely abstraction of realistic biomolecular interactions such as spatial and geometric constraints, this is a step in the right direction towards the goal.

- By evolving signalling networks under selection pressures of ultrasensitive response and adaptive response, I discovered the protein sequestration is evolutionarily conserved in both evolved ultrasensitive and adaptive networks. Based on the discovered mechanisms, I successfully designed a single signalling cycle with sequestrating proteins that could modulate the system dynamics between ultrasensitive response and adaptive response.

- From evolved ultrasensitive networks, bistability emerged as alternative solution for thresholds that are selected for. By analysing evolved bistable networks, I discovered a genre of novel biochemical networks that displaying bistability. Following the discovery, I devised algorithmic searching procedures to search the boundary between monostationarity and multistationarity in signalling networks.

- I further studied one of the discovered bistable motif with a single allosteric regulated kinase catalysing a substrate with ability of auto-dephosphorylation. Collaborating with E. Feliu, I secured the necessary and sufficient condition on kinetic parameters that the motif display bistability. We further proved that with multiple allosteric enzymes, the system can achieve unlimited multistabil-

ity. This work expanded our current knowledge on multistability in signalling networks. Based on the necessary and sufficient condition, I did numerical study on the motif under thermodynamic constraint (i.e. detailed balancing). This work also provide insights on the constraints of detailed balancing on biochemical reaction networks.

Chapter 2

In silico evolution with unbounded complexity

2.1 Introduction

Cellular networks allow organisms to sense and process environmental information and thereby implement phenotypic behaviours that enable survival. Hence, it is of fundamental interest to understand their structure and dynamics either by experimental and modelling studies on specific examples [27, 133, 134] or by searching for recurring structural motifs in large classes of systems [66, 135–137]. Collectively, these approaches have identified key dynamical features, such as ultrasensitivity and bistability [40], and elucidated biochemical elements used for their implementation, such as feedback loops, scaffold proteins and phosphorylation cycles [43, 97, 138–141]. Despite these insights, however, we still lack an understanding of the evolutionary origins of the features of dynamical and structural networks, limiting our ability to make functional predictions based solely on the presence or absence of these features [33]. Furthermore, network elements identified from current day organisms might not constitute the only feasible solutions for achieving a specific physiological task or implementation of a specific dynamical feature. The under-

standing of the possible solution space is thus mostly lacking, but could be essential from the perspective of engineering biological systems through synthetic biology [123]. One approach for understanding the evolutionary processes leading to current day network elements and for exploring the space of possible solutions is to re-create the evolutionary dynamics of cellular networks *in silico*. This task requires computational tools that are intuitive to use, yet are sufficiently complex to capture the system dynamics of known cellular networks. Modelling of the evolution of cellular networks has so far been attempted for exploiting evolution as a design tool (e.g. [65, 123]) or for interrogating evolutionary pressures leading to particular network properties (e.g. [142–144]). It is desirable to develop further general computational tools that can achieve both aims, and that can allow unconstrained modelling of evolution, while maintaining a realistic representation of biochemistry and system dynamics. Most previous studies either focused on modelling of evolution of large networks without incorporating dynamics [36, 122, 145–147], or explicitly considered temporal dynamics of the systems that are being evolved (using for example ordinary differential equations) (e.g. [148–152]) while enforcing bounds in the size and complexity of reaction networks that they can evolve. When the modelling of dynamics is combined with unbounded system size as done in the study of evolution of gene networks through duplication [153], it was possible to better understand the evolutionary solution space for networks implementing certain dynamics. In addition, each of the different models of cellular network evolution considers specific aspects of biology that they are aimed at addressing (e.g. role of duplication in evolution of robustness), but there are still some biomolecular aspects that are yet to be incorporated in evolutionary models of cellular networks. A particular example is the allosteric and domain-based nature of proteins, which are shown to be relevant for the system dynamics in the context of signalling networks [154, 155].

In this chapter, I introduce an extendable, general tool that provides biologically realistic simulation of the evolution of dynamic biochemical networks. The

tool, called BioJazz, combines a rule-based modelling approach [156–158] with evolutionary simulation, allowing for evolution of cellular systems without any need for a priori limitations on the systems that can evolve. Thus, what is meant here by without limitations is that the structure, size and complexity of the system that is taken as an evolving entity (i.e. the modeled cellular system) is not bounded in any way (other than computational limitations). Rule-based modelling is perfectly suited for this evolutionary approach, as it is developed in the first place to overcome the combinatorial complexity arising from accounting for all possible interactions in a given biological system [158, 159]. The rule-based modelling approach and the genome-like encoding of the network also allow biologically realistic mutational events to be modeled naturally. BioJazz has the ability to change and evolve networks with respect to both topology and biochemical parameters, by starting either from a designed network *de novo* or from a partially or completely functional seed network.

I demonstrate the use of Biojazz by examining the evolution of network dynamics for two sample cases, demonstrating evolution of network architectures for ultrasensitive and adaptive response dynamics. I also use these examples to demonstrate the effects of the parameters of the simulation algorithm on the performance and evolutionary space of such signalling networks.

2.2 Materials and methods

2.2.1 Representing network interactions: rule-based model

Previous attempts to model the evolution of cellular networks relied on *ad hoc* approaches to encode network architecture and dynamics (e.g. see [65, 123, 142, 143, 160]). In this project, I make use of recently developed rule-based approaches to enable a flexible encoding of cellular networks, allowing for both realistic representation of their biochemistry and for *in silico* evolution with unbounded complexity.

Rule-based approaches are developed for addressing the combinatorial complexity arising in modelling even the biological simplest reaction systems [158, 159] and, hence, are well suited to be combined with an evolutionary approach. Although several rule-based models are now available [156, 157, 161, 162], I choose to use the Allosteric Network Compiler (ANC) [156], because it systematically incorporates the allosteric and modular nature of proteins (note that the software structure of Bio-Jazz allows other rule-based models to be incorporated in subsequent developments). ANC is a stand-alone, rule-based compiler, which turns a high-level description of allosteric proteins into the corresponding set of biochemical equations by following mass-action kinetics.

ANC has been described previously [156]. In brief, it models proteins as multi-domain entities, where each domain is an allosteric unit that can adopt two general conformational states following the Monod-Wyman-Changeux (MWC) allosteric model [87]. The two conformational states of each domain can be described as relaxed, "R", and tense, "T", and are assumed to have distinct free energies of folding as well as different binding and enzymatic characteristics. Indeed, the binding and catalytic activity of reactive sites within a domain are dependent on, and only on, the conformational state of that domain. Biochemically, domains are independent sub-units of a protein, comprising reactive sites such as catalytic or post-translational modification sites (as explained below). This choice is inspired by the structure and function of multi-domain proteins in nature. In most cases, signalling proteins are functionally modular and make use of distributed surface docking sites for recognition [163], which has been demonstrated for both natural [164, 165] and synthetic protein circuits [155, 166–169].

ANC implements allosteric regulation by modelling the effect of any binding event or post-translational modification on a given domain through modifying the R-T transition dynamics of that domain. Thus, other molecules binding to a given protein can be seen as "modifiers", which alter the distribution of the R and T

states of the domain that they bind. The transition between the R and T states is governed by the free energies of these states as well as any intermediate state between them (see Appendix B). ANC can thereby model a cellular network as a given set of proteins that comprise domains and that interact through binding and covalent modifications of reactive sites on those domains. Any of the domains can be allosteric, in which case, it would have distinct R and T states with associated allosteric rate constants, and modifications would result in altering the dynamics of the R-T transitions in the following manner:

$$k'_{RT} = k_{RT} \prod_{i=1}^{N} (\Gamma_i)^{\Phi_i} \tag{2.1}$$

$$k'_{TR} = k_{TR} \prod_{i=1}^{N} (\Gamma_i)^{\Phi_i} \tag{2.2}$$

where k_{RT} and k_{TR} are the rate constant of switching between R state and T state without any allosteric modifier, k'_{RT} and k'_{TR} are switching rate constants with modifiers accordingly, Γ_i denotes the effect of the ith modifier on the equilibrium distribution between the R and T states, and the parameter Φ_i describing the proportional effects of the ith modifier on the R-T transitions. Detailed mathematical derivation can be found in Appendix B. As explained further below, in the BioJazz implementation of ANC, the k_{RT}, k_{TR}, and Φ values for each domain and the Γ_i values for different reactive sites on a given domain are free to evolve. Note that this freedom allows us to implement easily and naturally the evolution of both individual proteins with domains that have specific internal dynamics and protein interaction networks, via the definition of binding specificities among reactive sites and Γ parameters.

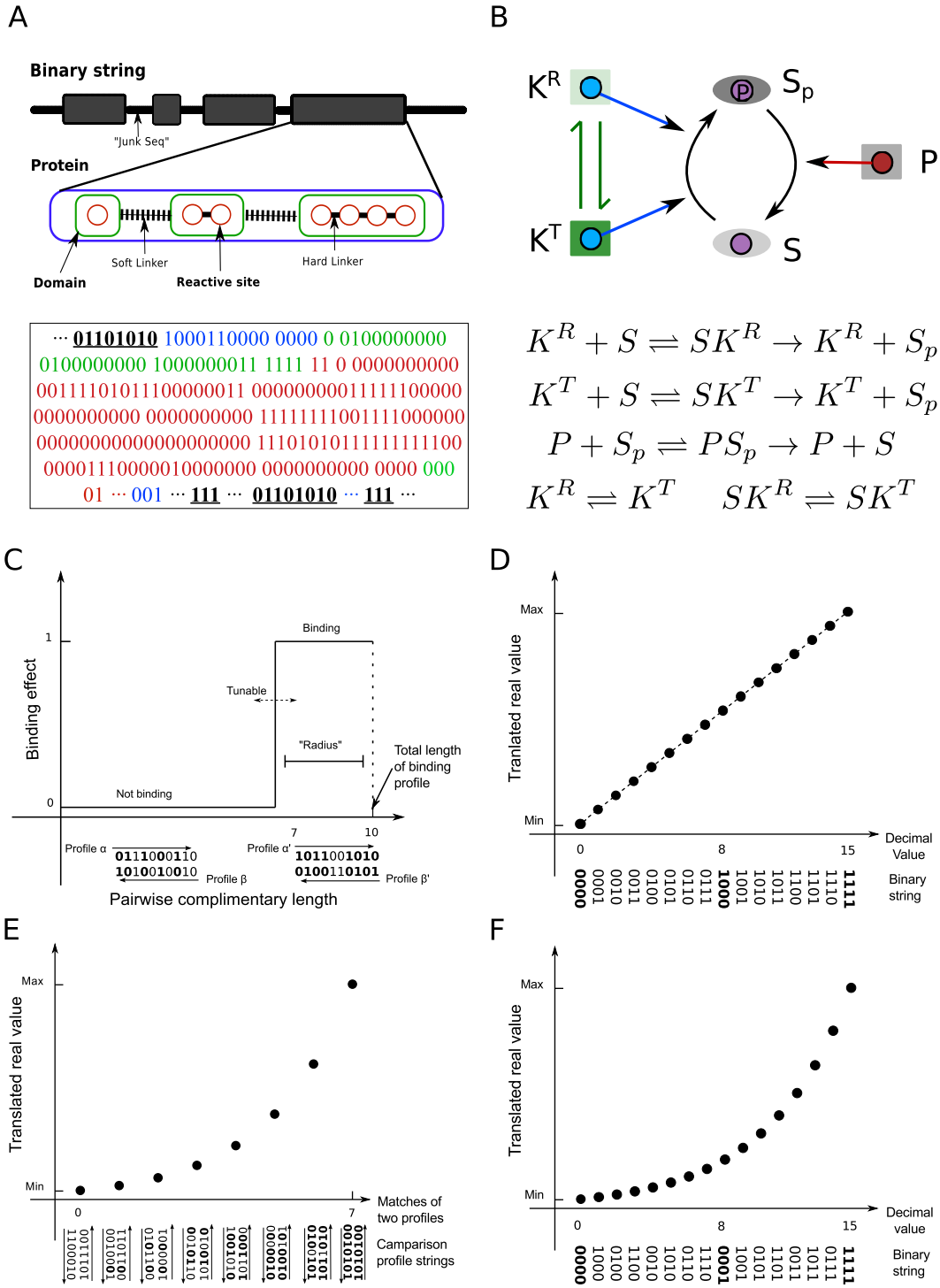


Figure 2.1: The "Genome" structure and scaling method used to encode cellular networks. **(A)** A cartoon representation of the binary string encoding the information needed to build an ANC model. The string has a hierarchical structure explained in the main text. **(B)** A cartoon representation and the resulting biochemical reactions of a sample reaction network that can be derived from a binary string (as shown in panel A). **(Caption next page...)**

Figure 2.1: **(Previous page continue) (C)** Determination of binding between two reactive sites from a binary string segment (*Methods*). The y-axis shows the binding effect; the x-axis shows complementary matches between two binding profiles. The threshold for binding, determining protein promiscuity, is user-defined. **(D, E)** Scaling of the binary string encoding of parameters into real values. The y-axis corresponds to parameter values; the x-axis shows decimal values of the binary string. For different parameters (i.e. Φ, protein concentrations, and rate constants of the conformational transition), a linear (D) or logarithmic (E) scaling is used (*Methods*). **(F)** The scaling of the binary string encoding of parameters relating to binding-mediated reactions. The y-axis corresponds to a kinetic rate value; the, x-axis shows a comparison between strings encoding for binding parameters of two reactive sites as explained in *Methods*.

2.2.2 Encoding network information: a binary string as a synthetic genome

By describing the interaction rules as well as their allosteric effects, ANC allows modelling of a reaction network of arbitrary size and complexity. To evolve cellular networks *in silico*, one needs a method to store and mutate the corresponding protein interaction rules and parameters. In BioJazz, I encode the information in an ANC model as a binary string (Figure 2.1A). Using a set of translation rules, all the information required to build an ANC model can then be extracted from a given string (Figure 2.1C-F).

The structure of the binary string is similar to a natural genome, where "non-coding" sections separate sections encoding information. This division is implemented by using "start" and "stop" strings, and allows an increase in evolutionary innovations through mutations (see below). It is also possible to start evolutionary simulations from entirely random initial points (i.e. a randomly generated binary string). The coding sections of the binary string encode the structure, dynamics and interactions of proteins as explained in detail below and in Table 1. Thus, I can parse a given binary string and translate into an ANC model (Figure 2.1B).

Protein domain structure and allosteric flag: The coding sections of the binary string contain information about the domain structure of proteins (Figure

2.1A). Each protein must contain at least one domain that contains at least one reactive site. There is no maximum limit to the number of domains and reactive sites a protein can have. As explained above, domains may be allosteric units, and hence, each domain is preceded with an allosteric flag sequence. When the allostery flag is set, the domain will undergo conformational changes and these dynamics may be affected by biochemical reactions happening at its reactive sites (note that reactions happening on other domains would not have an allosteric effect on this domain, i.e. domains are distinct and independent entities). To distinguish between domains and reactive sites on the binary string, I use soft and hard linker sequences that are inserted between domains and reactive sites respectively (Figure 2.1A). Thus, the soft linker sequences indicate start of a new domain within a protein; hard linkers indicate the different reactive sites on a given domain whose conformational dynamics is potentially modulated (provided the domain is allosteric). This structure has the additional benefit that mutations that result in joining or separating of domains can be naturally implemented (see *Mutations* section below). Reactive sites within a domain can be either a binding or catalytic site, and their nature is encoded on the binary string as shown in Table 2.1.

Table 2.1: Details of structural and encodings implemented in the binary string

Field Name	Length (L)	RegExp	Description
Binary String			
PRE JUNK	Any	$[01]^*$	Zero or more bits representing untranslated sequence preceding first protein
genes	$L\{proteins\}$	$[protein]+$	One or more genes separated by untranslated sub-sequences
POST JUNK	Any	$[01]^*$	Zero or more bits representing untranslated sequence following last protein
Protein			
START CODE	8	01111110	Fixed pattern before the string of protein indicating the starting point of a protein
Concentration	10	$[01]L$	Loglinear scaled, encodes inital concentration of protein
UNUSED	4	$[01]L$	Reserved field

Continued on next page

Table 2.1 – continued from previous page

Field Name	Length	RegExp	Description
Domains	$L\{domains\}$	$[domain]+$	One or more domains separated by a soft linker pattern '001'
STOP CODE	3	111	Terminates the protein
Domain			
Allosteric flag	1	$[01]L$	Determines the domain is allosteric regulated or not
$R \leftrightarrow T$ transition rate	10	$[01]L$	Loglinear scaled, kinetic parameter of conformation transitions in basal level
Φ	10	$[01]L$	Linear scaled into $[0, 1]$, determines changes in allosteric equilibrium under interactions
UNUSED	4	$[01]L$	Reserved field
Reactive sites	$L\{sites\}$	$[site]+$	One or more protodomains separated by a hard linker pattern '000'
Reactive site			
Type	2	$[01]L$	Reactive site type, $00 \equiv bsite$, $01 \equiv msite$, $10 \equiv csite$, $11 \equiv csite$
Substrate polarity	1	[01]	A csite to modifies (0) or unmodifies (1) the substrate
Binding profile	10	$[01]L$	Determines ligands pairs with sufficiently complementary string
k_f profile	20	$[01]L$	Loglinear scaled, determines association kinetics with Hamming distance from pairing reactive sites
k_b profile	20	$[01]L$	Loglinear scaled, determing disassociation kinetics with hamming distance from pairing reactive sites
k_p profile	10	$[01]L$	Loglinear scaled, for csite only, determines rate of post-translational modification
k_{eq} ratio	10	$[01]L$	Loglinear scaled, determines allosteric effect of msite modification, see Γ in Appendix B
k_f polarity mask	20	$[01]L$	XOR with k_f profile to determine profile of modified reactive site ($msite = 1$)
k_b polarity mask	20	$[01]L$	XOR with k_b profile to determine profile of modified reactive site ($msite = 1$)
k_f conformation mask	20	$[01]L$	XOR with k_f profile to determine new profile of T conformation
k_b conformation mask	20	$[01]L$	XOR with k_b profile to determine new profile of T conformation
k_p conformation mask	20	$[01]L$	XOR with k_p profile to determine new profile of T conformation
UNUSED	4	$[01]L$	Reserved field

Continued on next page

Table 2.1 – continued from previous page

Field Name	Length	RegExp	Description

In regular expression, the '$*$' means 'zero or more' and '$+$' means 'one or more'.

ANC intra-action fields: Intra-action fields are binary strings located at the beginning of each domain. They encode the parameters controlling the internal allosteric properties of the domain, namely the basal kinetic rates for the transitions between the R and T states (k_{RT} and k_{TR} from Equation B.3 and B.4) and the parameter Φ (which is assumed to be the same for each of the different reactive sites of the domain and, as such, encoded once per domain). The switching rates are log linearly scaled into a real value (Figure 2.1F, Table 2.1); parameter Φ is linearly scaled into the interval [0,1] (Figure 2.1D, Table 2.1).

ANC interaction fields: Interaction fields are binary strings associated with the reaction sites in each domain. They encode how a change in the state of reaction site (binding or modification) will affect the R-T transition of that domain, i.e. they encode the parameters Γ_i described above. In addition, the binary string encodes binding and rate profiles (described in the next section), as well as a site type for each reactive site. The available types are *binding*, *catalytic* or *modification* sites (Table 2.1).

Binding and rate profiles: : When the binary string is converted to an ANC model, BioJazz iterates over all pairs of reactive sites and compares their binding profiles to determine the site-specific interactions among protein domains. In each iteration, BioJazz performs an exclusive-OR (XOR) operation on the binding profiles of two given sites. The number of "1"s in the resulting string from this operation determines whether or not binding occurs based on a user-defined threshold (Figure 2.1C). Besides the binding profile, each site has also a *forward* and *backward reaction rate profile*. When two sites are found to be binding (based on their binding profiles), the XOR operation is repeated, this time using the forward and backward

rate profiles, to determine the binding coefficients (Figure 2.1E, Table 2.1). Finally, reactive sites that are catalytic encode an additional *catalytic rate profile*. If one of the sites is a catalytic site and the other a modification site, the catalytic rate profile of the former is scaled log linearly into real value and is applied as the catalytic rate constant of the corresponding Michaelis-Menten kinetics. All the translated reaction rate constants are evolvable in biologically plausible parameter ranges (Table 2.2).

Table 2.2: Parameter ranges used for the *in silico* evolution of signalling networks

Parameters	*In silico*	Measure	Reference
Concentration (μM)	$[10^{-3}, 10^{3}]$	$[0.002, 1.8]$	[139, 170–175]
Phosphorylation (s^{-1})	$[10^{-3}, 10^{3}]$	$[0.17, 8.87]$	[170–173]
Dephosphorylation (s^{-1})	$[10^{-3}, 10^{3}]$	$[0.06, 5.31]$	[170–173]
Auto-dephosphorylation (s^{-1})	N/A	$[0.00097, 0.0025]$	[170–173]
Binding membrane protein (s^{-1})	$[10^{-3}, 10^{3}]$	$[0.0036, 0.70]$	[170–173]
Unbinding membrane protein (s^{-1})	$[10^{-3}, 10^{3}]$	$[0.00016, 0.060]$	[170–173]
Protein association ($\mu\mathrm{M}^{-1} \cdot s^{-1}$)	$[10^{-3}, 10^{3}]$	$[0.10, 7.53]$	[170–173]
Protein disassociation (s^{-1})	$[10^{-2}, 10^{2}]$	$[0.015, 2.86]$	[170–173]
Basal conformational switching (s^{-1})	$[10^{-2}, 10^{2}]$	N/A	N/A
Γ	$[10^{-2}, 10^{2}]$	N/A	[156]
Φ	$[0, 1]$	N/A	[156]

Profile masks: In real proteins, the kinetic rates associated with each reaction (e.g. binding rate, catalytic rate, etc.) can be altered by the structural changes that the protein undergoes. To include such changes, the model should incorporate the possibility of alterations in the kinetic rates of each reactive site with the *R-T* state transition of the domain. I do so by implementing a *conformational mask profile*, which is applied to all rate profiles of the reactive sites and alters the outcome of the XOR operation (Table 2.1). There are therefore distinct binding rates between the R and T states. For the modification sites only, there is also a *modification mask profile* that is applied to the binding rate profiles to alter the binding rates for modified states (Table 2.1). Note that both mask profiles can evolve to have no effect on kinetic rates, i.e. a given reaction site in a given domain can have the same reaction kinetic rates under each of the R and T states by appropriate setting of the mask profiles.

2.2.3 Modelling mutations

The use of rule-based modelling and encoding of such a model in a genome-like binary string allows us to implement most biologically feasible mutations easily. Currently, the possible mutations included in BioJazz are point mutations, protein duplication, protein deletion, domain duplication, domain deletion, domain joining, domain splitting and domain shuffling. Of these, mutations involving domains were to our knowledge not considered before [65, 120, 150], but are straightforward to include in the rule-based approach. The rate of occurrence of the different mutations is controlled by user-defined parameters. Users can also restrict BioJazz to mutate a subset of the network's attributes including junk bits, linkers, binding profiles, allosteric flags, types of reactive site, etc. This flexibility is useful for example to "freeze" all or parts of a network and use BioJazz as a design tool rather than mimicking biological evolution.

Point mutation: Point mutations are implemented as flipping of specific bits in the binary string. A point mutation can alter any of the qualitative flags (explained above) or reaction parameters. Of particular note are mutations on hard and soft linkers, which can result in domain splitting and fusion respectively. The mutation algorithm parses the binary string and attempts a point mutation at each location: a bit is flipped if a randomly generated number in the interval $[0, 1]$ is smaller than a user-set probability (corresponding to a genome-wide point mutation rate).

Protein duplication/deletion: In nature, the rate of gene duplication is suggested to be a function of the size of genome [176]. Based on this observation, BioJazz implements duplication and deletion rates defined per protein. The mutation algorithm parses the binary string and attempts a duplication or deletion at each protein coding section; an entire section is duplicated or deleted if a randomly generated number in the interval $[0, 1]$ is smaller than a user-set probability. The protein duplication and deletion rates can be set independently. When a protein

coding section is duplicated, it is added to the end of the binary string. When a protein coding section is deleted, the binary string is shortened correspondingly. It is also possible that a protein is silenced by a point mutation at its "start" sequence.

Domain duplication/deletion: Bioinformatics analysis of genomes of existing organisms reveals duplication patterns of domains in proteins, where the duplication patterns show no dependence on the size of the domains involved [177]. Thus, BioJazz implements the domain duplication/deletion rate per protein. At each replication step, a randomly generated number in the interval $[0, 1]$ is generated for each protein. If this number is smaller than a user-defined probability, a random fragment of the binary string is picked. This fragment is then either deleted or copied and the new copy is inserted at the end of the originally chosen one. Note that the randomly picked segment can contain many reactive sites or none.

Domain shuffling: BioJazz implements rearrangements between two protein-encoding sections of the binary string. The mutation rate leading to rearrangements is defined per protein. For each protein coding section of the binary string, a random number drawn from uniform distribution in the interval $[0, 1]$ is compared to a user-defined probability. If the random number is smaller, a fragment containing a certain number of reactive sites is randomly chosen. Then, another subsection of a protein coding section of the binary string is randomly selected, copied and fused with the first selected fragment. Note that this approach combines sets of intact reactive sites, which can correspond to an entire domain, part of a domain, or a sequence that covers multiple domains. Besides mimicking biological domain shuffling, shuffling is expected to create novel material for subsequent evolution.

Genome Rearrangement: In biological systems, rearrangement of large genome chunks containing multiple genes also happens with certain probability. I also implemented this mutation operator in BioJazz. At each step of mutation, comparison between a random number and the rearrangement rate will determine occurrence of genome rearrangement. With rearrangement occurring, a continu-

ous segment containing multiple reactive sites that possibly cross several genes is randomly selected. Then either deletion of segment or duplication of segment is randomly chosen and executed.

Horizontal gene transfer: BioJazz also has implementation of horizontal gene transfer (HGT). Since HGT occurs between different genomes in nature, this mutational operator is only implemented when using population based selection (see below). At the mutation step of each individual, the occurrence of HGT is determined by comparing a random number to a pre-defined probability (set by the user). If the random number is smaller than this probability, a continuous segment of string containing multiple reactive sites is randomly chosen and copied from the mutating individual. Then, another genome/individual is randomly selected and the copied segment is inserted into its genome at a randomly chosen site that is between any two reactive sites.

2.2.4 Modelling evolutionary selective pressures

To simulate evolution *in silico*, I need to model selective pressures experienced by the evolving cellular networks and so link the contribution of a networks function to the overall *fitness* of the organism. Fitness is an abstract concept, representing the reproductive success of an organism and might be most tractable for microbes where it could be approximated by growth rate [178]. In BioJazz, the fitness of networks can be defined by the user, such that networks can be evolved under biologically motivated or artificial selective pressures.

The user-defined fitness function is used to evaluate the performance of a given network, encoded by a particular binary string, and to calculate a fitness score. In previous studies on the evolution of signalling and regulatory networks, the fitness function usually involved applying a stimulus to the network and evaluating its temporal or steady state response [36, 65, 123, 143, 179]. Different fitness functions relating to dynamical or structural features of the network can be easily constructed

as illustrated in the results section for ultrasensitive dynamics (additional sample files are included in the BioJazz web site) and adaptive dynamics (Figure 2.2).

Ultrasensitivity fitness function: Currently the fitness function used to score the ability of a given signalling network to generate an *ad hoc* switch-like function to mimic the ultrasensitive signal-response relationship evaluated the response to a three-step ramp-up and three-step ramp-down signal profile as shown in Figure 2.2A. For each ramp-up in the signal, the system is simulated to steady state before the next ramp is applied. The scoring function considered both the amplitude of the response to middle steps in ramp-ups and ramp-downs (amplitude score S_{amp}) and the difference of the response amplitudes between the middle steps and the other two steps (ultrasensitivity score S_{ult}). If y_{min} and y_{max} are defined as the minimum and maximum values of the response during the interval from a change in the signal to steady-state, then the response amplitude for each of the signal ramp-ups (indicated with a '+' sign) and ramp-downs (indicated with a '−' sign) is calculated as:

$$\Delta y_{i+} = y_{i+_{max}} - y_{i+_{min}} \tag{2.3}$$

$$\Delta y_{i-} = y_{i-_{max}} - y_{i-_{min}} \tag{2.4}$$

where the subscripts denote the corresponding ramps in the input signal. With these measurements, the amplitude score (S_{amp}) is given as the normalized amplitude of the change in response to the second ramp-up and ramp-down signals:

$$S_{amp} = \frac{(\Delta y_{2+} + \Delta y_{2-})/2}{y_{total}} \tag{2.5}$$

with y_{total} being the maximum possible response (i.e. the concentration difference between a fully active and fully inactive output protein), and acts as a normalisation factor ensuring S_{amp} to be between 0 and 1. In order to quantify the ultrasensitivity of the system, I use the difference between the amplitudes of the responses to the

second ramp-up/ramp-down signals, and the first choice/third choice. I first define the difference in the response to the different ramp-up and ramp-down signals as $S_{u1} = (\Delta y_{2+} + \Delta y_{2-})/(\Delta y_{1+} + \Delta y_{1-})$ and $S_{u3} = (\Delta y_{2+} + \Delta y_{2-})/(\Delta y_{3+} + \Delta y_{3-})$. Then I can derive the ultrasensitivity score (S_{ult}) as:

$$S_{ult} = \sqrt{(\frac{S_{u1}}{r_u + S_{u1}} \cdot (\frac{S_{u3}}{r_u + S_{u3}}))} \tag{2.6}$$

where, r_u is a user-defined scaling parameter that ensures the two ratios S_{u1} and S_{u3} (and thus the ultrasensitivity score) is between 0 and 1. Besides the amplitude and ultrasensitivity scores, I also define a complexity score (S_{com}). It is plausible to assume that networks are under selection to minimize their energetic burden to the cell, and this score allows us to capture network complexity. The complexity score is given by:

$$S_{com} = \frac{r_c}{r_c + C} \tag{2.7}$$

where C is the sum of the total number of rules, proteins, domains, and reactive sites in the ANC model and r_c is a user-defined scaling parameter for scaling the complexity score S_{com} between 0 and 1. Finally, the fitness function combines the three scoring functions:

$$F = (S_{amp}^{\omega_a} \cdot S_{ult}^{\omega_u} \cdot S_{com}^{\omega_c})^{\frac{1}{\omega_a + \omega_u + \omega_c}} \tag{2.8}$$

with the ω_* being user-defined parameters that control the weightings of the different scores.

Simple adaptation fitness function (Figure 2.2B): The key aspects of adaptive response dynamics are that the system shows an initial response to the input ($\Delta O_{max}^{+/-} \neq 0$) that the steady state value of the output returns to its pre-input level, irrespective of the level of the input. In other words, after a sustained change in input (e.g. a step change), the output should initially respond but ultimately

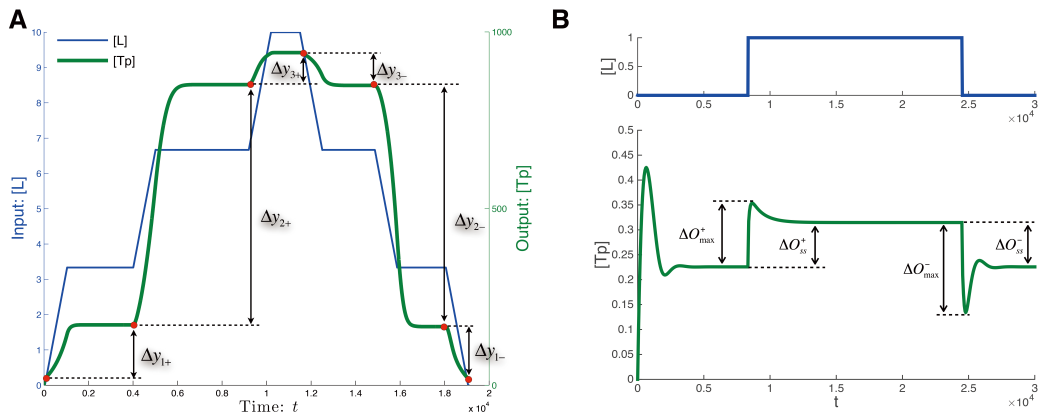


Figure 2.2: Sample fitness functions for selection for networks with ultrasensitive or adaptive response dynamics. **(A)** The input signal (blue) used in the temporal simulations of the system for ultrasensitivity. Each ramp-up and ramp-down of the signal is introduced after the system reaches steady state. The corresponding system output over time is shown in green. The differences in steady state output between different signal levels, indicated as Δy values on the plot, are used to calculate the amplitude and ultrasensitivity scores. **(B)** Illustration of the dynamics of input signal (blue) the output response (green) in simulations of the system for adaptive dynamics. The parameters in adaptive fitness function, $\Delta O^{+/}_{max}$ and $\Delta O^{+/}_{ss}$, are labelled.

settle back to its original steady state: $\Delta O^{+/-}_{ss} \approx 0$. Therefore, the adaptation fitness w can be configured as:

$$w = \sqrt{\frac{\Delta O^{+}_{max}}{C} \cdot \frac{K}{K + \Delta O^{-}_{ss}}} \cdot \sqrt{\frac{\Delta O^{-}_{max}}{C} \cdot \frac{K}{K + \Delta O^{-}_{ss}}} \tag{2.9}$$

where C is a normalization factor to scale $\Delta O^{+/-}_{max}$ and $\Delta O^{+/-}_{ss}$ in to $[0, 1]$, and K is a threshold parameter. By imposing such a selective pressure, it is possible to evolve an increased response sensitivity ($\Delta O^{+/-}_{max}$) and reduced adaptive error ($\Delta O^{+/-}_{ss}$), and so achieve networks with an adaptive response.

When the fitness function requires evaluation of the system dynamics, a temporal simulation of the network is executed by numerically integrating the set of ODEs arising from the interaction reactions in the network. To perform these simulations, BioJazz uses MATLAB® with files automatically generated from ANCs

output via the Facile tool [180]. Stochastic simulation of the ANC model is also possible by customising the fitness scoring function.

2.2.5 Modelling evolutionary dynamics

Evolutionary dynamics arising from the emergence of mutant genotypes in a population and their subsequent change in frequency can be modeled in different ways. In particular, evolution could be approximated either by a random walk in which a single beneficial (or neutral) mutant can be fixed in the population before any other mutants can arise or as occurring in a population where multiple mutants can co-exist. The former is an appropriate model for evolutionary dynamics at low mutation rate and large population size limit [181]; the latter approach can give rise to evolutionary dynamics similar to that described by the concept of quasi-species [182].Both approaches are implemented in BioJazz.

Evolution as a random walk: Under very low mutation rates and in large populations, evolutionary dynamics can be approximated by a random walk in the genotype space. Then, a single genotype dominates the population and new mutants either get fixed or are lost rapidly under natural selection and/or genetic drift [183]. The probability of fixation for such rare mutants with a given fitness effect has been approximated by Kimura [181]. This approximation can be used to model evolution under a large population and low mutation rate scenario, where the calculated probability of fixation for a mutant generated from the wildtype genotype is used to replace the wildtype or not [184]. Biojazz implements this approach by starting simulations from a given genotype and using this genotype to generate a mutant genotype. The mutation is then accepted with probability αP_{fix}, where P_{fix} (fixation probability) is calculated from the fitness of the original (w) and mutant (w') genotypes by the following equation:

$$P_{fix} = \frac{1 - e^{-2s}}{1 - e^{-4N_e s'}} \tag{2.10}$$

with s being the selection coefficient and equal to $s = \frac{w'-w}{w}$, and N_e is the effective population size (set in the range $10^5 \sim 10^8$, based on measurements for prokaryotes [185]). The coefficient α is used to tune (usually increase) the speed of simulation and is always chosen to make $\alpha P_{fix} < 1$ for all mutations [184]. A newly generated mutant will be accepted if a random number (uniformly drawn from interval $[0, 1]$) is smaller than αP_{fix}. Otherwise it is rejected. After acceptance of a given mutant, that mutant replaces the original genotype and the simulation continues. If the mutant is rejected, a new mutant is generated from the original genotype. The evolutionary simulation is continued until a user defined fitness criterion or a specified number of mutations is reached.

Population based approach: Here I consider evolution dynamics in discrete generations of an asexual population of a fixed-size [181, 183]. In a fixed-size population, selection for the next generation is implemented by sampling genotypes according to their fitness scores. Assume that there are genotypes A_1, A_2, A_3, ... with fitness w_1, w_2, w_3, ... and frequencies p_1, p_2, p_3, ... in the current population. Then the expected proportion or frequency of A_i genotypes in the next generation will be

$$p'_i = \frac{p_i w_i}{p_1 w_1 + p_2 w_2 + \cdots} = \frac{p_i w_i}{\bar{w}} \tag{2.11}$$

The p'_i is the propensity that genotype A_i is chosen for reproduction (with one progeny) in each sampling. To implement these dynamics, I start with a homogenous population. At the beginning of each generation, individuals reproduce and mutate based on mutation rates by sequentially drawing and duplicating an individual and comparing the mutation rate with a random number r_1 from $[0, 1]$. If r_1 is less than mutation rate, the reproduced individual is mutated. After reproduction the fitness scores for all mutants are recalculated. Then at the end of each generation, I apply selection. More specifically, I include all of the p'_i values in a vector and then generate another random number r_2 uniformly drawn between 0 and the length of

this vector. The individual that is selected for the next generation is determined by the index of the vector into which the random number falls. The sampling process continues until the number of individuals in the new generation reaches the defined population size.

Parallelization and choice of algorithms: Evolutionary simulations in BioJazz can be performed either on a single computer node, i.e. desktop workstation or laptop computer, or parallelly on several computer nodes in a computer cluster. If running parallelly on a cluster, the evolutionary algorithm of random walk style is implemented as setting each node running a single evolutionary program and several different simulations can be parallelly performed, while the population based approach is implemented by scoring several individuals parallelly on different nodes in a single simulation. Both approaches generate converged results from evolutionary simulations, however random walk style approach is much faster and requires less storage space, which might become a bottle neck when population based approach simulates too many generations before it converges.

2.2.6 BioJazz configuration file

BioJazz contains three key parts that are interlinked to each other: an encoding of an ANC model in the form of a binary string, evolutionary simulation of that binary string through mutations, and dynamic simulation of the ANC model and derivation of a fitness score. Many of the parameters governing the structure of these three parts and their inter linkage can be defined by the user, allowing for high customizability. These parameters are stored in a single *configuration file*.

Besides the parameters already mentioned above, the configuration file also allows setting of parameters relating to computational performance (e.g. number of nodes allocated for parallel computing, memory allocated for scoring), string encoding (e.g. fields' width and binding profiles of input and output), the evolutionary algorithm (e.g. mutation rates, population size, seed network), the dynamical

simulation of the ANC model (e.g. simulation time, numerical simulation error threshold), the scoring function (discussed below), and the output structure (e.g. frequency of output generation).

2.2.7 Post-evolutionary pruning of evolved networks and mutational analysis

It is possible that not all reactions in evolved networks are needed to achieve required function (as seen for example in previous *in silico* evolution studies [143, 186]). Thus, I incorporated ways to prune evolved final networks or apply mutations on them for further functional analyses. This can be done readily by altering the string representation of evolved networks. BioJazz stores each of the evolving networks (in the case of population based approach to modelling evolution) and the primary evolving network (in the case of random walk approach to modelling evolution) at each generation of the simulation in two separate files. The user can choose to generate these files only in a BioJazz-compatible format or in additional formats readable in ANC, Facile, and MATLAB®. Pruning and mutations can be done on these files and the resulting modified networks can be reanalysed. In the case of using BioJazz compatible files for such post-analysis, the user can make modifications on the string representation of the network and can also use existing subfunctions in the BioJazz source code. A detailed description and example of this approach is provided in BioJazz manual.

2.3 Results

To illustrate the workings of BioJazz and how it can be used to address biological questions, I consider here the evolution of signalling networks under an example selective pressure (additional selective pressures can easily be constructed by encoding an appropriate fitness function in the configuration file, as shown in *Methods*).

This demonstrates selection for ultrasensitive response dynamics as described below. Note that the fitness function used and the associated analyses are provided as an example to illustrate the applicability of BioJazz. The user has complete flexibility over the choice of fitness functions and of the parameters in a given evolutionary simulation.

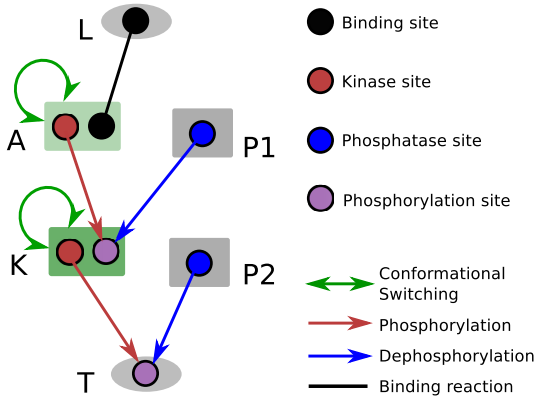


Figure 2.3: Schematic, showing the network structure used as the starting point for evolution for ultrasensitivity. The ligand (L) and the output protein (e.g. a transcription factor, T) are shaped as oval, while all other signalling proteins (e.g. a receptor/adaptor (A) protein, a kinase (K), or a phosphatase (P)) are shaped as rectangle. Black line represents binding reaction between two sites. Red arrows represent phosphorylation reactions between a kinase site (red) and a phosphorylation site (purple). Blue arrows represent dephosphorylation reactions between a phosphatase site (blue) and a phosphorylation site. The green Coloured rectangle indicates a protein domain, whose conformational switching is allosterically regulated (also indicated by a self-pointing green line with arrows at both ends).

Ultrasensitivity is observed in many biological networks, and in particular in signalling networks implementing phosphorylation cycles [40, 43, 92, 139]. An ultrasensitive response is one where a change in the input generates a non-linear change in output, specifically signal levels and response levels at steady states generate a sigmoidal curve [40, 92]. To evolve signalling networks capable of displaying ultrasensitive dynamics, I run simulations with selection under a particular fitness function.

With this fitness function, I used BioJazz to evolve ultrasensitive signalling networks. I started evolutionary simulations from a minimal seed network composed

of a receptor, a kinase, a phosphatase, and an output protein (Figure 2.3). The output protein was not allowed to duplicate or be deleted, but the rest of the network was free to evolve via all the mutations implemented in BioJazz (see *Methods*). Note that the constrained structure of the model in this case reflects a user choice rather than a limitation and allows us to demonstrate the application of BioJazz to evolve signalling networks with ultrasensitive dynamics by fixing the input and output of the evolving system. It is also possible to include the ligand as part of the evolving entity, in which case I would be able to evolve new ligands and ligand-receptor interactions, provided that an appropriate fitness function is devised. For example, to study the coevolution between ligands and the response, one can easily cluster different proteins based on the tags and prefixes of protein names implemented in the source code.

Selecting for ultrasensitivity in the signalling network using the random-walk approach (see *Methods*), I ran evolutionary simulations by assuming a high population size and low mutation rate regime (see *Methods*) and by using different complexity weightings ω_c. In particular, I ran 5 simulations in parallel each for 4 different complexity weights: $\omega_c = 0,\ 0.1,\ 1,\ 10$. Each simulation is assigned to a node in a computer cluster. I set a target fitness score of 0.8 and a maximal computation time of 120 hours per simulation. The simulations were terminated when either the target fitness score or simulation time was reached.

In all simulations, the fitness score increases over generations (Figure 2.4A) and I evolve an ultrasensitive network reaching at least a fitness score of 0.8. Analysing the evolutionary dynamics in these simulations, I find that fewer mutations were needed for simulations with ω_c set to lower values (Figure 2.4B), i.e. when the fitness penalty for complexity was low. The time required for evaluating the fitness of each mutant, however, was significantly larger with lower ω_c. These findings suggest that a weaker constraint on network complexity (i.e. smaller values of ω_c) allows the evolutionary simulations to sample a larger space of networks

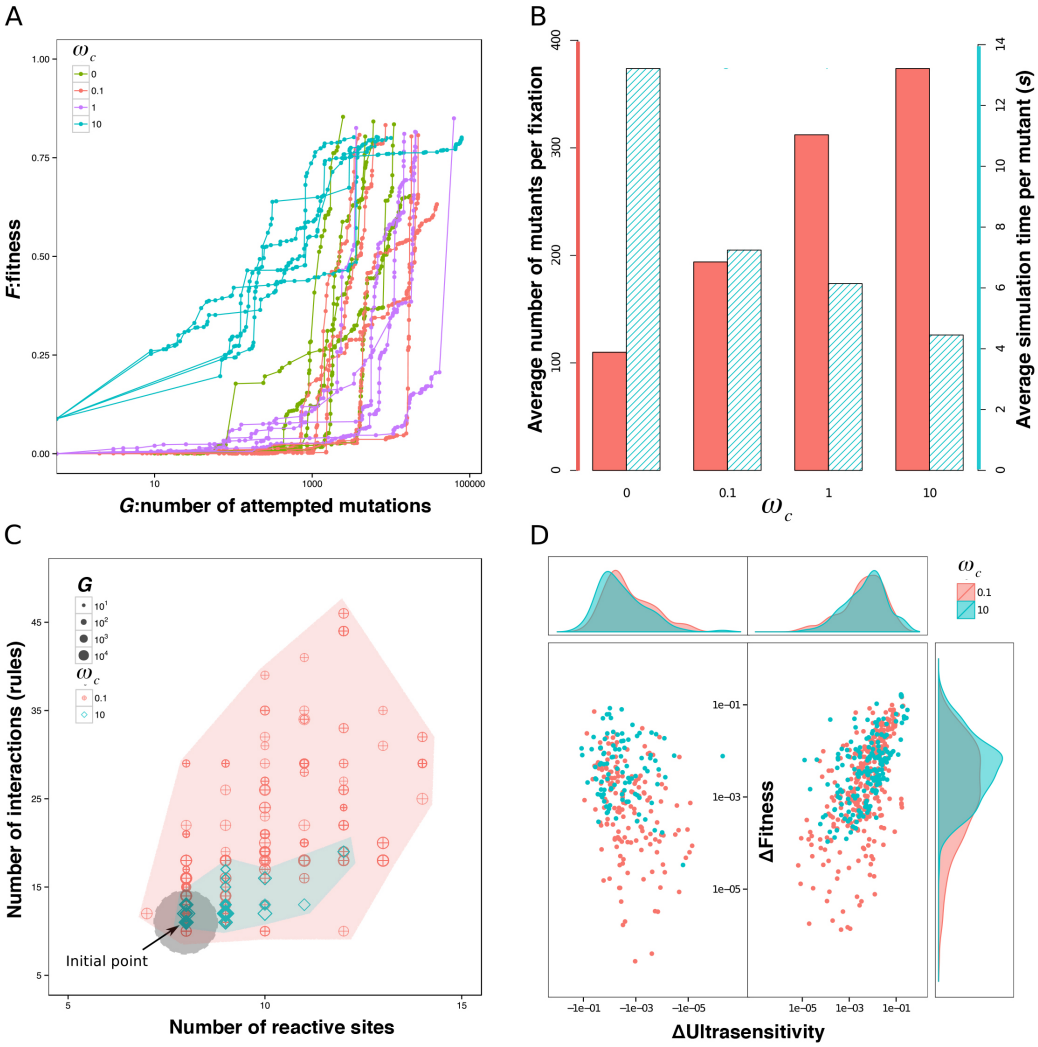


Figure 2.4: Results from sample evolutionary simulations. **(A)** The fitness score plotted against the total number of mutations sampled. Each curve depicts the results of a single evolutionary simulation, which is a biased random walk over the network space (Equation 2.10 in *Methods*). Each dot on each curve represents an accepted mutation (lines are to guide the eye). Distances along the x-axis between two dots indicate the number of mutations sampled between two accepted mutations. In all simulations, fitness increases with the number of mutations accepted, but in two simulations (one with $\omega_c = 0$ and one with $\omega_c = 0.1$) fitness fails to reach the target level of 0.8 before the maximal simulation time of 120 hours is reached. **(Caption next page...)**

Figure 2.4: **(Previous page continue) (B)** The average number of mutants sampled before a mutation is accepted increases with ω_c; the average time for evaluating the fitness of each mutant in simulations decreases with ω_c. A higher weighting of complexity score (ω_c) in the total score gives a higher penalty to mutations that generate complexity in the network structure. **(C)** The evolutionary space showing the numbers of reactive sites and of interactions for all simulations with $\omega_c = 0.1$ and 10. Each data point represents an accepted mutant network from different stages of the simulation, with the shape and colour indicating the ω_c of the simulation and the size indicating the generation number (i.e. number of mutations). Note that many of the data points from the simulations with $\omega_c = 10$ are overlapping. The initial network is at the centre of the grey area. **(D)** The distributions of mutational effects on fitness and ultrasensitivity from accepted mutations during all simulations with $\omega_c = 0.1$ and 10. Sub-graphs at the top and right are density estimates for the ultrasensitivity changes ΔS_{ult} and fitness changes ΔF of all fixation events.

and more easily find beneficial mutants. Correspondingly, the number of reactive sites and interactions in networks diverges more widely in such simulations, while network complexity is highly constrained for large ω_c (Figure 2.4C). On the other hand, a higher weighting for the complexity measure (high ω_c) can result in this measure dominating the total fitness calculation (Equation 2.8). Consequently, a larger number of mutations with detrimental or neutral effects on the ultrasensitivity and amplitude of the response may be accepted because their low scores could be absorbed by stronger effects from the complexity measure. I find that indeed this possibility is realised: the distribution of the ultrasensitivity scores of fixed mutants is slightly shifted to larger negative values in simulations with $\omega_c = 0.1$ compared to data from simulations with $\omega_c = 10$ (Figure 2.4D). A similar pattern also occurs with amplitude score.

Our implementation of evolution under a low mutation rate and high population size regime through Equation 2.10 still allows for a degree of neutral evolution. Thus, I find significant diversity in the set of ultrasensitive networks emerging at the end points of different evolutionary simulations (Figure 2.5B). This diversity confirms that different network architectures and biochemical mechanisms can generate ultrasensitivity. The evolved ultrasensitive networks I find recover known bio-

chemical mechanisms that generate ultrasensitivity. One such mechanism is enzyme saturation in a covalent modification cycle (or zero-order sensitivity, or Goldbeter-Koshland kinetics) [92]. In this mechanism, saturation of enzymes that mediate the covalent modification of a substrate generates ultrasensitivity in the modified substrate levels. In our simulations, the initial starting networks display high concentrations of kinase and phosphatase and low levels of target protein, and I analyzed the evolutionary trajectory of key kinetic parameters in a few sample simulations. In particular, I consider composite parameters K_1 and K_2, which determine the binding kinetics of the kinase and phosphatase to the output protein and should decrease with increased enzyme saturation (see the legend of Figure 2.6 for a full definition of K_1 and K_2). I find that the initial evolution of these parameters is quite erratic (Figure 2.6) until the system reaches a high level of K_2 where phosphorylation can result in low output at any signal level (network 30). Once this point is reached, evolution progresses with both K_1 and K_2 being decreased, indicating that the system spends less time in complexes of the kinase and output protein and of the phosphatase and output protein: the enzymes increasingly become saturated. Consequently, both the ultrasensitivity and the amplitude of the system response increase and reach the target fitness score in network 70. I find a similar trend in some other simulations, where decreasing K_1 and K_2 is accompanied by increasing ultrasensitivity, suggesting that these trajectories may be common in the evolution of ultrasensitive responses, at least from an initial regime of high substrate and low enzyme concentration. Nevertheless, evolved networks that are not in the regime suggest that unusual mechanisms other than zero-order sensitivity might play important roles in orchestrating the ultrasensitivie response dynamics, Chapter 3 will discuss some further studies and results on novel mechanisms of ultrasensivity from additional evolutionary simulations.

To provide a second example for the application of BioJazz, I developed a different fitness function that is designed to select for networks with adaptive re-

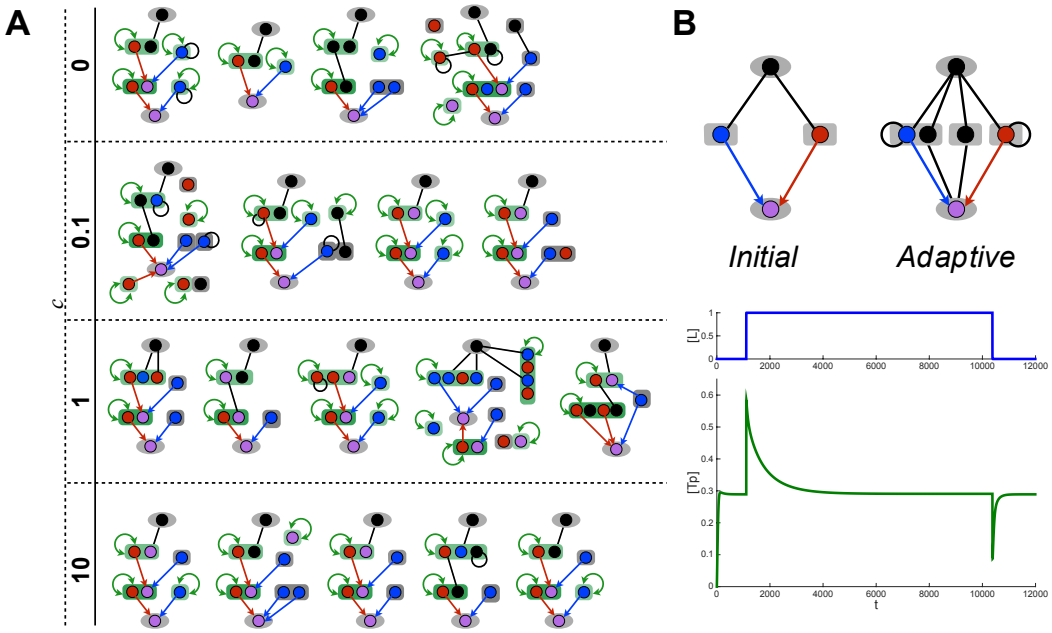


Figure 2.5: Sample evolved network structures and dynamics. **(A)** Sample network structures evolved to achieve ultrasensitivity in simulations with different weighting of the complexity score ω_c . In each network, the nodes stand for proteins and edges stand for interactions. The isolated (i.e. unconnected) nodes seen on some evolved networks represent proteins that do not interact with any other proteins (hence they can be removed without affecting the response dynamics). For explanation of labels and edge colours see legend of Figure 2.3. **(B)** An evolved network structure and its dynamics using selection for an adaptive response.

sponse dynamics (Figure 2.2B, also see *Methods*). This type of response dynamics is observed in many cellular systems and is characterised by an initial response to a persistent external stimulus that eventually returns to its pre-stimulus level. In the context of signalling networks, adaptive response dynamics are observed and studied in bacterial chemotaxis [69, 70] and the response of yeast to osmotic shock [187]. General signalling network models capable of adaptation have been presented [48] and *in silico* evolution has been successfully used to understand gene network architectures that can achieve adaptive responses [65]. Here, I have adopted the fitness function used in the latter study (Figure 2.2B) and used BioJazz to evolve signalling networks with adaptive dynamics (Figure 2.5B). I found that 9 out of 10 from the initiated simulations resulted in networks achieving high fitness solutions and adaptive response dynamics. Different from previous work on adaptive gene networks, the structures of evolved adaptive protein interaction networks do not show any obvious negative feedback [66]. Instead, I find the evolved networks commonly exploiting a buffering mechanism that could be equivalent to a feedforward mechanism [66]. In the example adaptive network shown in Figure 2.5B, the input protein can bind four binding sites in three different proteins, two of which are the kinase and phosphatase for the output protein. When a perturbation happens at the input protein concentration level, different affinities of kinase and phosphatase for binding to the input protein result in breaking the balance of phosphorylation and dephosphorylation of the output protein, inducing an initial response. Later, the binding protein in the middle (which has slower binding reaction rate constants) sequesters the input protein to re-balance the phosphorylation and dephosphorylation of the output protein. The end effect of this buffering mechanism is a response dynamics similar to that resulting from a feedforward interaction loop [66]. All other evolved adaptive networks utilised similar solutions to this example to achieve adaptive responses. However, there are two most fundamental mechanisms in achieving adaptive dynamics: negative feedback loop and incoherent feedforward loop [66].

And negative feedback loop exist widely in control engineering applications and natural biological systems to maintain stability or implement adaptive dynamics. The reason that no adaptive networks with negative feedback loops emerge is because such motif requires the network to invent additional backward inhibition (e.g. dephosphorylation of one's upstream signalling protein). This requirement installs a barrier in the evolutionary landscape which is difficult for an evolving network to overcome given a basin (i.e. where the incoherent feedforward loop mechanism is) in the evolutionary landscape. Analysing the dynamics of sample evolved networks under different levels of input perturbation I found their fitness to be sensitive to the level of the perturbation used in the fitness function. In particular, the adaptation precision (i.e. the ability to return exactly to pre-stimulus activity level after a signal) is dependent on signal level. This highlights the importance of the design of the fitness function on the types of networks that can evolve in the simulations. In Chapter 3, a new fitness function with more stringent conditions is devised for selecting networks with adaptive response dynamics in order to better understand design principles of adaptation in signalling networks.

2.4 Discussion

Here I have presented BioJazz, a tool that combines rule-based approaches and evolutionary simulation. Its key features are the implementation of biochemical interactions found in cellular networks, the simulation of dynamics arising from these interactions and their evolution with unbounded complexity through biologically plausible mutations. Previous approaches to evolutionary simulation of cellular networks have only considered a subset of these abilities. As such, I expect BioJazz to be useful both as an exploratory tool for the evolutionary systems biology community to understand evolutionary pressures leading to specific biochemical features of biological networks and as a design tool for the synthetic biology community to

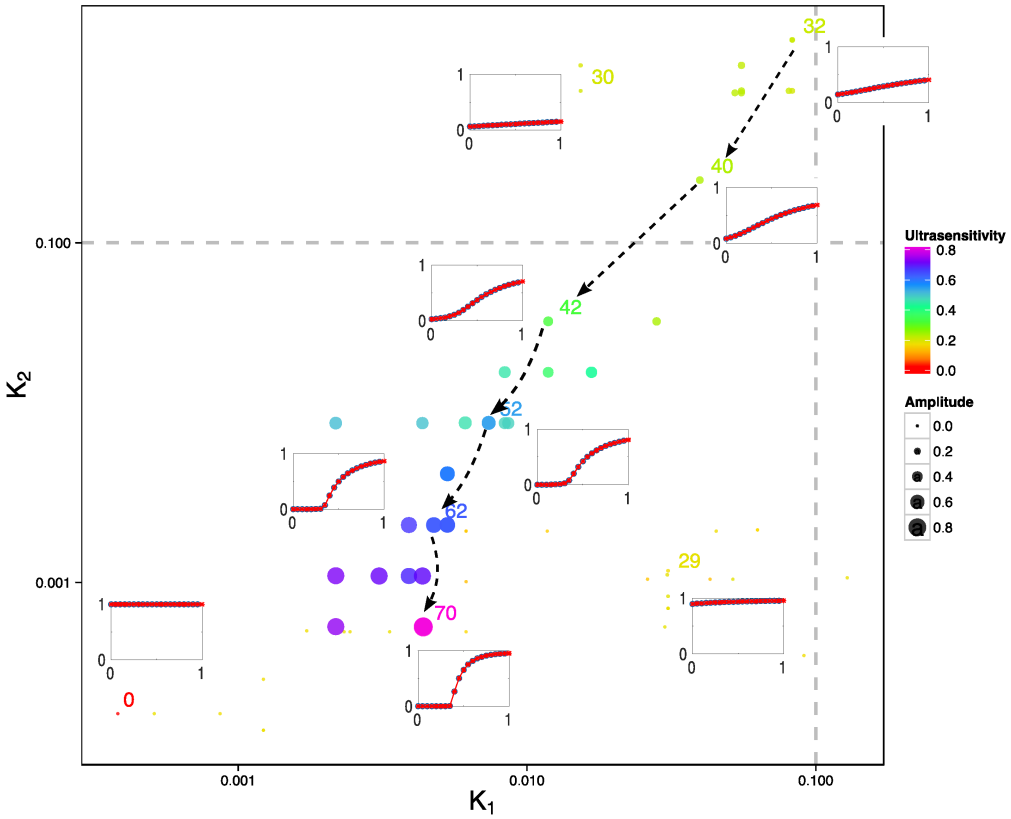


Figure 2.6: Evolution of model parameters for a sample evolutionary simulation with $\omega_c = 0$ and selecting for ultrasensitivity. Each dot represents a model from different points in the evolutionary simulation (as indicated by the generation number on each dot), while the x- and y-axis show the composite parameters, K_1 and K_2, that give the average catalytic binding efficiency of the kinases and phosphatases to the target protein respectively. The catalytic binding efficiency is defined as the Michealis-Menten constant of the enzyme (kinase or phosphatase) over the total substrate concentration, and the average is calculated as the geometric mean of individual binding efficiency of the different kinases and phosphatases and their allosteric states: $K_1 = \prod_{i=1}^{m} \sqrt[m]{(K_1^{R_i} \cdot K_1^{T_i})^{1/2}}$ and $K_2 = \prod_{j=1}^{n} \sqrt[n]{(K_2^{R_j} \cdot K_2^{T_j})^{1/2}}$. The dot size and colour indicate the response amplitude and ultrasensitivity. For selected networks the input-output response curve and the network architecture is also shown. Dashed lines with arrow heads show the trend of how ultrasensitivity increases with the evolution of decreasing values of K_1 and K_2. The initial network starts in the bottom left corner (network 0) moves to the bottom right corner (network 29) and then to the top right corner (network 32)

explore biochemically plausible implementations of different network dynamics.

As I demonstrate, BioJazz is developed in a way that allows high flexibility and user-friendliness. All parameters relating to the evolutionary simulations, as well as the fitness functions used to select networks can be specified by the user, allowing testing of different hypotheses. As a demonstration, I showed how to use BioJazz to evolve networks under different complexity constraints and to generate ultrasensitive dynamics. I found that complexity constraints can alter the efficiency of the evolutionary simulations, mainly because of their effects on the distribution of mutational effects on fitness.

Under all complexity constraints considered, I found evolutionary simulations to result in ultrasensitive networks under the appropriate fitness function. In addition, adoption of a different fitness function allowed the evolution of networks displaying adaptive dynamics. These results show that BioJazz can be used to study a range of system dynamics (i.e. ultrasensitivity, adaptation, oscillation). Networks resulting from specific simulations that implemented different selective pressures displayed specific architectures, suggesting that BioJazz can be used to study the possible repertoire of functional networks. In the case of ultrasensitivity, I found that these networks and their evolutionary dynamics highlighted known biochemical mechanisms and implied existence of unusual mechanisms as well. In particular, I found that kinetic parameters controlling binding of the enzymes and output protein evolve to favour low saturation initially for increased response amplitude and then high saturation later on for increased ultrasensitivity. BioJazz can be used to further elucidate such trends under different evolutionary scenarios. For example, the simulations I used started from high substrate and low enzyme concentrations. In the subsequent chapter, I reverse this situation and explore how ultrasensitivity can emerge under regimes where high enzyme saturation would not be possible Chapter 3. Similarly, one can use higher level selection functions, rather than *ad hoc* functions selecting for ultrasensitivity (as I have done here), to elucidate the bi-

ological origins of ultrasensitivity, for instance by imposing fluctuating signals and selecting on responses with defined thresholds and penalties. Alternatively, one can implement selection for different dynamics such as pulsatile response dynamics or oscillatory dynamics. The evolved network structures could then provide insights into which biochemical networks can implement the required dynamics and inform both systems and synthetic biology studies (as has been done before, e.g. see [36, 146, 150]). In the following chapters, it is again illustrated investigations of ultrasensitive and adaptive networks in Chapter 3 and bistable signalling networks in Chapter 4, 5 and 6.

There are notable previous works on evolutionary simulation of the structure and dynamics in cellular networks. In particular, previous studies analysed the *in silico* evolution of gene regulatory networks to understand the emergence of different dynamics [65, 123, 153, 186, 188], and their modularity and robustness [189]. The latter features were also studied in evolutionary simulations using metabolic [121] and signalling network models [120, 143], or general network models [122, 147]. As an open-source platform, BioJazz aims to further enable such studies by providing an *in silico* evolution model that explicitly considers systems dynamics and protein allostery and domain structure. The incorporation of protein allostery and domain structure is a particularly unique addition in evolutionary modelling of networks, that was not considered in any of the previous works, but whose effects on system dynamics have been studied in many previous experiments [154, 155]. In addition, the combination of rule-based modelling with *in silico* evolution is a novel attempt in modelling evolution and allows for a natural way to deal with emerging system complexity in evolutionary simulations. In particular, the rule-based modelling approach theoretically allows for simulation of arbitrarily large reaction networks as well as protein complexes. Most previous models of network evolution that considered system dynamics have used bounds on both of these features, either by imposing limitations on the number of species in evolving networks [143, 179, 190],

or by imposing limits on the protein complexes that can emerge in the evolving networks [65, 153, 188].

Although by using rule-based modelling BioJazz theoretically allows the evolution of cellular networks without restricting their complexity, there are still computational challenges with simulating large reaction networks and multi-protein complexes that give rise to the 'curse of dimensionality' [156, 158]. In particular, the ANC framework I used here generates a full set of differential equations possible in the network, prior to simulation, which creates a significant computational burden. Such technical challenges are increasingly being addressed with developing rule-based modelling frameworks. For example, the Kappa simulator KaSim [157] and BioNetGen simulator NFsim [191] allow faster simulation of reaction systems of arbitrary size. These methods are currently based on using stochastic simulations and do not consider allosteric nature of proteins as done in ANC. It should be possible to combine the best features of the developing approaches and create new rule based modelling frameworks that combine modelling of protein allostery with computationally feasible simulation routes that allow arbitrarily large networks to be simulated. Future development of BioJazz would thus explore this route of expanding the rule-based modelling aspect of its evolutionary framework towards combining best features of different methods.

Such development of the rule-based modelling component of BioJazz could allow extending its focus from encoding signalling networks to metabolic and transcriptional networks. In particular, rule-based models like Kappa and BioNetGen are able to allow modelling of degradation and synthesis reactions. This can be combined with extending the binary string patterns of BioJazz model representation to encode binding between proteins and genes, and thereby mimicking transcription factors binding on DNA. For metabolic networks, the extension would require encoding of metabolites in a form that captures basics of chemical conversion with inspirations from previous studies [121]. This would require significant further de-

velopment and interfacing rule-based models and metabolites through their corresponding representations. It is hoped that these developments will be facilitated by the open-source nature of BioJazz.

Chapter 3

Protein sequestration emerges as a tuning point

3.1 Introduction

Molecular signalling networks enable cells to generate appropriate dynamical responses to external signals including pulsed, oscillatory, ultrasensitive, and adaptive dynamics [192, 193]. Such response dynamics are also implemented in human-engineered systems, motivating the use engineering principles to understand and engineer cellular networks [194, 195]. This approach has been particularly useful in the context of gene regulatory networks, where feedback and feedforward control are successfully used to understand and even engineer specific response dynamics, such as adaptation [48, 196, 197], bistability [198–202]. While these studies demonstrate the usefulness of engineering principles, and in particular feedback control, in understanding and modulating biological systems, there is also great interest to discover and understand potential design principles that are specific to cellular networks and that are exploited by evolution to generate specific system dynamics [110, 123].

One way to identify potential evolutionary design principles is to look for features conserved across different cellular systems. For example, the high prevalence

of phosphorylation-dephosphorylation cycles in signalling networks and branching points in metabolic networks led to their identification as potential mediators of ultrasensitive dynamics [92]. Similarly, several common biochemical features of signalling networks were identified as mediators of specific response dynamics; bifunctional enyzmes mediating adaptive and pulse dynamics [113, 203], multi-site phosphorylation mediating multistability [52, 54, 106, 204], and phosphorelays mediating ultrasensitivity and multistability [50, 114, 205–207].

An alternative approach for identification of potential design principles in cellular networks is to use *in silico* evolution [65, 153]. Through the mimicking of biological evolution of cellular networks in the computer, *in silico* evolution can generate many systems with a desired response. These systems can then be analyzed to identify their key features mediating specific response dynamics. The application of this approach led to the identification and subsequent experimental implementation of sequestration as a mechanism for generating bistability and oscillation in gene regulatory networks [95, 123, 208] and also to uncovering the principle of adaptive sorting in ligand-receptor interactions, which is analogously featured in immune recognition [190]. These examples illustrate the potential utility of *in silico* evolution to allow the discovery of subtle biochemical processes, that could not be readily deduced from observations on network connectivity. In addition, the evolutionary approach allows exploring the impact of specific environmental and cellular conditions on the evolution of different design principles [121, 145, 189, 209–211]. Given that many different potential design principles can give rise to a certain dynamical response, such insights could be useful for increasing our ability to predict which designs are more likely to be found under which ecological and evolutionary setting.

Motivated by this potential, I use BioJazz (see [212] and Chapter 2) to explore the design principles of ultrasensitive and adaptive dynamics (Figure 3.1A,B) in signalling networks. I show that when possible, enzyme saturation by substrates readily evolve as a key enabling feature for ultrasensitivity. For simulations where

enzyme saturation was made difficult to evolve, I find that enzyme sequestration emerges as a key mechanism for enabling ultrasensitivity. Interestingly, this same mechanism also emerged in networks selected for adaptive dynamics, and mediated a contrasting effect on kinases and phosphatase activities. Based on these findings I design a generic model of a signalling cycle motif, featuring a scaffolding protein. I show that resulting enzyme sequestration in this motif enables it to generate both ultrasensitive and adaptive dynamics and under biologically relevant parameter regimes. Furthermore, I show that for a given set of parameters, the dynamics of such a motif can be tuned between adaptive and ultrasensitive responses through modulation of sequestrating protein concentration or affinities. These findings indicate that enzyme sequestration through scaffolding proteins provides evolution with a design principle to generate systems with plastic response dynamics and could be equally exploited in synthetic biology.

3.2 Methods

3.2.1 Evolutionary simulations

I started evolutionary simulations with three "seed" networks with different structure (Figure 3.1C). For each "seed" network, I run two groups of simulations with different total concentrations of output protein, mimicking initial conditions of enzyme saturation with substrate or not. Under each condition (and "seed" structure) I have run 10 independent evolutionary simulations.

An evolutionary algorithm implements the iterative process of mutation and selection with a predefined fitness function of ultrasensitivity. With the linear encoding of signalling networks [212], the networks, in each iterative round, undergo mutation then selection based on their new fitness scores (see next section). For simulating evolutionary dynamics, I assumed a low mutation rate high population size regime as explained in [212]. In such a regime, evolution is expected to

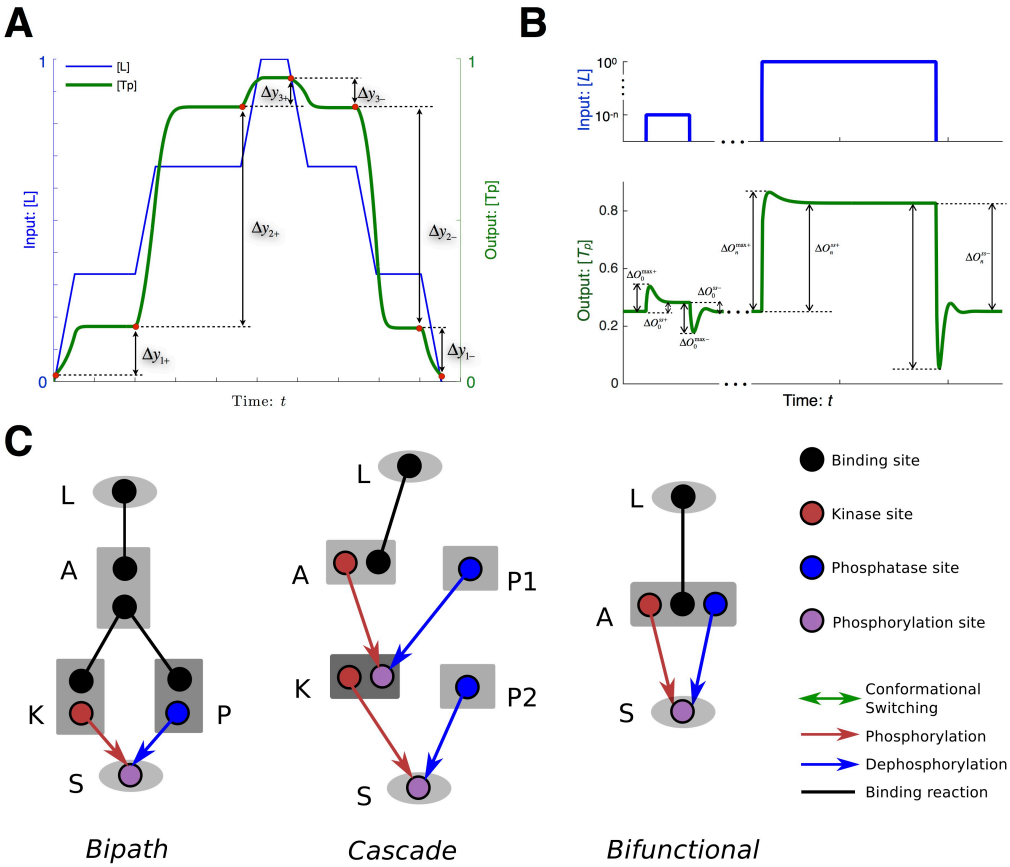


Figure 3.1: Evolutionary simulation setup. **(A)** Fitness function to score ultrasensitive networks. The input signal (blue) is used in the temporal simulations of the system for scoring ultrasensitivity. Each ramp-up and ramp-down of the signal is introduced after the system reaches steady state. The corresponding system output over time is shown in green. The differences in steady state output between different signal levels, indicated as Δy values on the plot, are used to calculate the amplitude and ultrasensitivity scores. **(B)** Fitness function to score adaptive networks. Illustration of the dynamics of input signal (blue) the output response (green) in simulations of the system for adaptive dynamics. The parameters in adaptive fitness function, $\Delta O_i^{(max+/-)}$ and $\Delta O_i^{(ss+/-)}$, are labeled. **(C)** Structure of three different starting networks. The ligand and the output protein (e.g. a transcription factor) are shaped as oval, while all other signalling proteins (e.g. receptor/adaptor proteins, kinases, or phosphatases) are shaped as rectangle. Black line represents binding reaction between two sites. Red arrows represent phosphorylation reactions between a kinase site (red) and a phosphorylation site (purple). Blue arrows represent dephosphorylation reactions between a phosphatase site (blue) and a phosphorylation site. The green colour rectangle indicates a protein domain (not shown in starting networks), whose conformational switching is allosterically regulated (also indicated by a self-pointing green line with arrows at both ends).

proceed akin to a random walk, where only fitter mutants are expected to fix and form the basis for next mutants [212]. Thus, I simulate only a single network, from which I generate mutants and replace the resident network based on the probability of fixation calculated from the fitness difference between mutants and the resident genotype as derived by Kimura [181] The additional parameters controlling evolutionary simulations, such as mutation rates and allowed size of protein complexes are summarised in a configuration file (Appendix A). All presented simulations are run with the same parameters as listed in this file.

3.2.2 Selection criteria for adaptive and ultrasensitive dynamics

For selection function of ultrasensitive dynamics, I use the same fitness function in Chapter 2 that favours large responses when presented with intermediate input signal and little responses when presented with either low or high input signals (Figure 3.1A, also see *Methods* in Chapter 2). For selection of adaptive response dynamics, I adopted the fitness function in Chapter 2 and further extended this to achieve more stringent conditions for adaptive responses. In particular, networks were evaluated for their ability to respond in transient manner to three distinct step-signals with different magnitudes (i.e. 1, 10, 100) (Figure 3.1B). Especially, the function calculates both maximum response to input perturbations, $\Delta O_i^{(max+/-)}$, and adaptive precision (i.e. different between pre- and post- input perturbations, $\Delta O_i^{(ss+/-)}$. For each square pulse signal perturbation, the score is calculated as

$$w_i = \sqrt{\frac{\Delta O_i^{(max+)}}{C} \cdot \frac{K}{K + \Delta O_i^{(ss+)}}} \cdot \sqrt{\frac{\Delta O_i^{(max-)}}{C} \cdot \frac{K}{K + \Delta O_i^{(ss-)}}} \tag{3.1}$$

where C is a normalization factor to scale $\Delta O_i^{(max+/-)}$ and $\Delta O_i^{(ss+/-)}$ into $[0, 1]$, and K is a threshold parameter (Figure 3.1B). Then the adaptive fitness is calculated as geometric mean of scores of all perturbation steps $w = \sqrt[n+1]{\prod_{i=0}^{n} w_i}$.

3.2.3 Model for signalling cycle motif featuring enzyme sequestration

In this motif, a sequestrating protein (T) can bind both the kinase (K) and the phosphatase (P), thus making these enzymes inaccessible to the substrate (S and S_p). This system gives us a generic model that can be described as a set of ordinary differential equations with 10 reaction rate constants and 9 chemical species (Appendix C). In order to explore the different response dynamics of the generic model, I sampled 11 parameter sets from a biologically feasible range (see Table 2.2 in Chapter 2, however the concentration range is further constrained into $[10^{-4}, 10]$). I used the same fitness functions as in the evolutionary simulations to characterise the response of this signalling motif to an incoming signal. The signal presence is simulated as changes in the kinase concentration level. To explore effects of enzyme saturation, I sampled the generic model at two conditions: enzyme saturated condition ($[P_{tot}] = 0.1$, $[S_{tot}] = 1$) and enzyme unsaturated condition ($[P_{tot}] = 0.1$, $[S_{tot}] = 0.1$).

3.3 Results

To explore design principles for generating ultrasensitive and adaptive response dynamics in signalling networks, I have evolved signalling networks under different cellular conditions and from three different starting networks composed of an input-receiving protein (L), an output protein (S), and proteins with binding, kinase and phosphatase activities and labelled as adaptor proteins (A), kinases (K) or phosphatases (P) (Figure 3.1). The initial structures were selected based on common observations from natural signalling networks. In particular, the cascade topology is based on the signalling cascades such as the Mitogen Activated Protein Kinase (MAPK) signalling networks [41, 213]; the bipath topology is based on the observations that cells utilise different signalling pathways that share specific elements

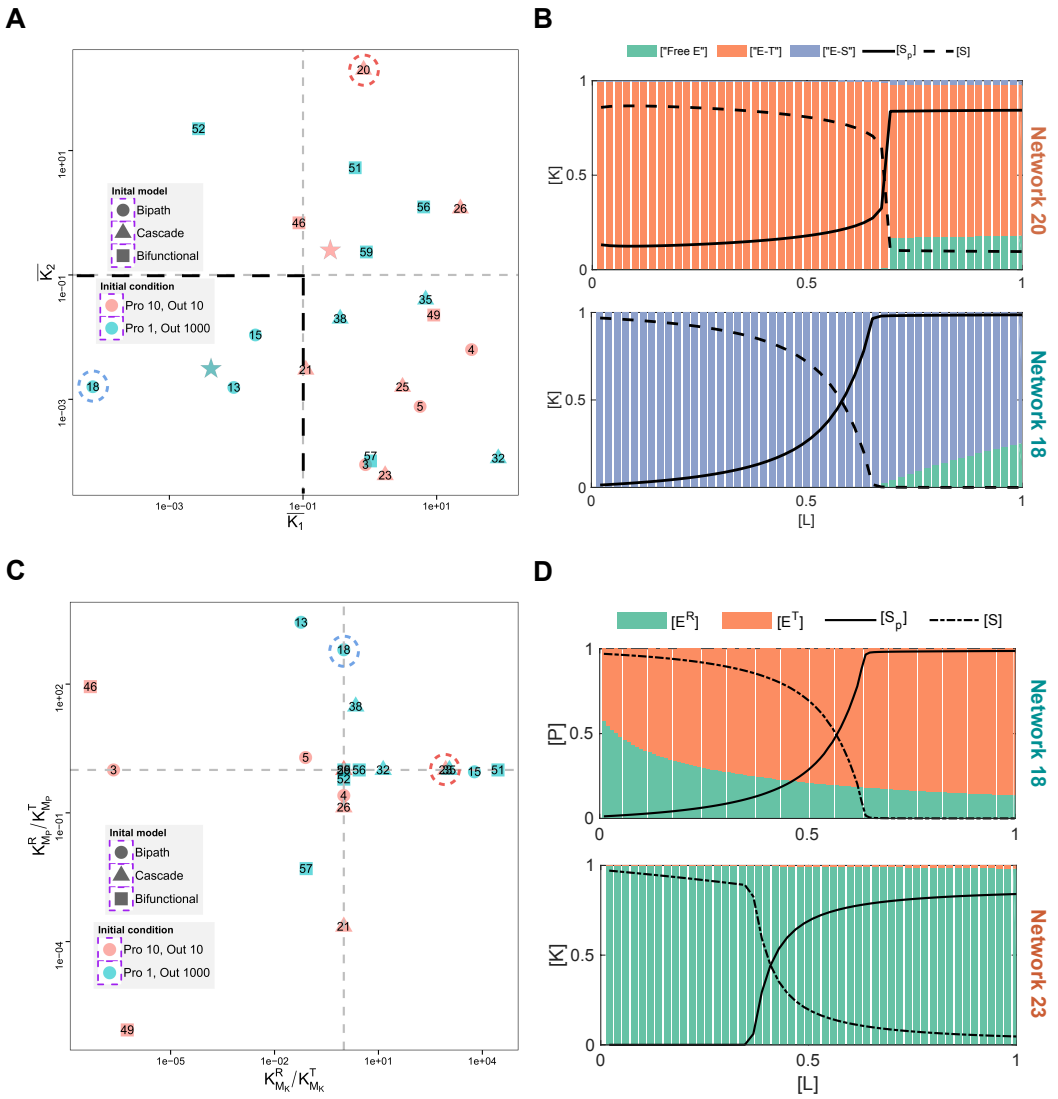


Figure 3.2: Analysis of evolved ultrasensitive networks. **(A)** Saturation parameter of all evolved ultrasensitive networks. The parameters are defined as Michealis-Menten constant of the enzyme (kinase or phosphatase) over the total substrate concentration, and the average is calculated as the geometric mean of individual binding efficiency of the different kinases and phosphatases and their allosteric states: $\bar{K_1} = \prod_{i=1}^{m} \sqrt[m]{(K_1^{R_i} \cdot K_1^{T_i})^{\frac{1}{2}}}$ and $\bar{K_2} = \prod_{i=1}^{n} \sqrt[n]{(K_2^{R_i} \cdot K_2^{T_i})^{\frac{1}{2}}}$ (see [212] and also Chapter 2). The shapes represents different starting structures in the evolutionary processes, while the colours represent two different starting conditions (i.e. blue: output protein $[S_{total}] = 1000$ and starting with all other signalling proteins, denoted as A^*, concentrations $[A^*_{total}] = 1$; red: output protein $[S_{total}] = 10$ and starting with all other signalling proteins concentrations $[A^*_{total}] = 10$). The number labelled on each data point is the unique ID used for each evolutionary simulation. The two star shaped points indicate the value of the saturation parameters at the start of evolutionary simulations. **(Caption next page...)**

Figure 3.2: **(Previous page continue)(B)** The different forms of enzymes, substrate-accessible (green), substrate-inaccessible (orange), and substrate-bounded (blue), overlaid with the dose-response dynamics for two different evolved networks (network 20 and 18 in Figure 3.3). The solid and dashed lines show the concentration of phosphorylated (i.e. response) and unphosphorylated substrate respectively. **(C)** Ratio between K_M values of different conformational states for kinase (x-axis) and phosphatase (y-axis). The colours, shapes and numbers on the dots are the same as in (A). For enzymes without allosteric regulation the ratio are set to one, so that there are no distinctive conformational differences. **(D)** The percentage of enzymes in different conformational states, relaxed "R" state (green) and tensioned "T" state (orange), overlaid with dose-response dynamics for two different evolved networks (network 18 and 23 in Figure 3.3). The solid and dashed lines show the concentration of phosphorylated (i.e. response) and unphosphorylated substrate respectively.

leading to cross-talk, as seen for example in the signalling pathways controlling yeast mating and filamentous growth responses [214–216]; the bifunctional topology is inspired by observations that many kinases can also display significant phosphatase activity, or can readily attain such activity via few mutations [217–221] Furthermore, this motif is selected as it provides a particularly minimal starting point for evolution, where I assume a generalist enzyme that contains both kinase and phosphatase activities initially and that can evolve these activities further via mutations and protein duplication. The cellular conditions were selected to mimic the presence or absence of enzyme saturation, which can mediate ultrasensitivity in signalling cycles [92] but might be lacking in natural systems [97, 222]. Thus, the evolutionary simulations allowed us to explore the role of these different features. I used specific selection criteria that operate on the response dynamics resulting from the network in presence of a signal profile (see *Methods* and Figure 3.1). I run 10 simulations for each of the conditions and for selecting ultrasensitive and adaptive dynamics, resulting in a total of 60 simulations for each dynamics.

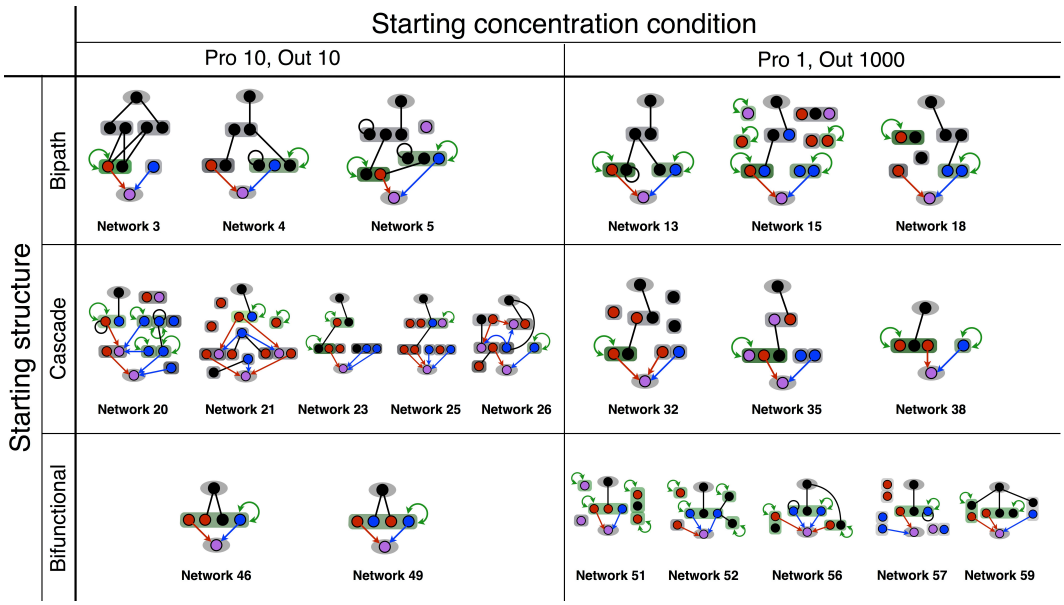


Figure 3.3: Structure of all evolved ultrasensitive networks. The categories are based on the starting concentration conditions and starting network structures. Schematic information is the same as in Figure 3.1, in these evolved networks there are several networks contain isolated proteins that evolved from duplications and mutations.

3.3.1 *In silico* evolved ultrasensitive networks display enzyme saturation, enzyme sequestration, and allosteric regulation.

It has been shown theoretically that a simple signalling motif comprising a kinase, phosphatase and their substrate can lead to an ultrasensitive input-response relation when the enzymes are fully saturated by their substrate [92]. This mechanism is termed zero-order sensitivity and can be achieved by having kinetic parameters that favour complex formation among enzymes and the substrates, and by having a large ratio of the total concentration of substrate to that of enzymes [92]. I found that when conditions allow, zero-order sensitivity readily evolves *in silico*. Of the 30 simulations, which were started with a high ratio of output protein to signalling protein concentrations, 11 have resulted in the emergence of ultrasensitivity and 8 of these successful simulations resulted in kinetic parameters where either or both kinases and phosphatases were saturated (Figure 3.2A, blue points). These results confirm that the *in silico* simulation framework can recover a known biochemical

mechanism - enzyme saturation by substrate - for achieving ultrasensitivity.

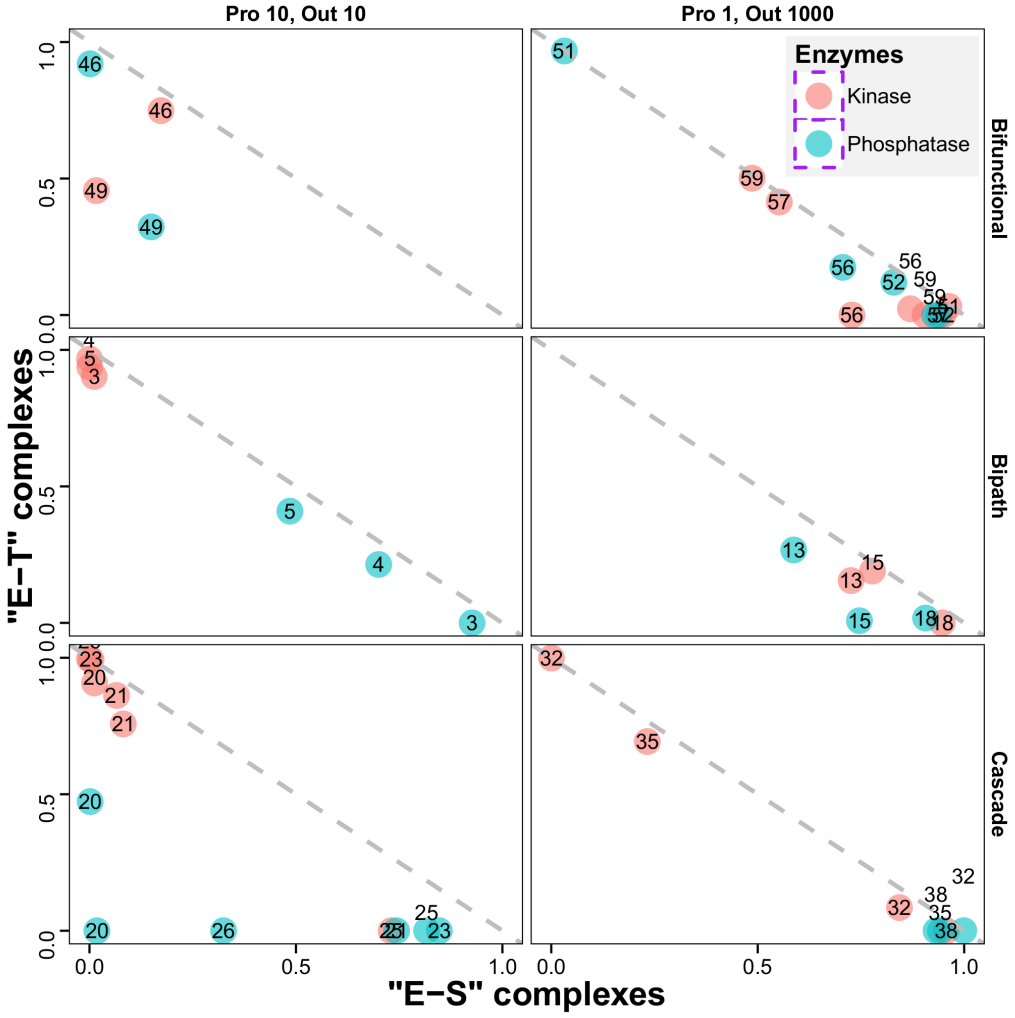


Figure 3.4: Average level of enzymes sequestrated by other proteins and average level of enzymes saturated by substrate (output protein) in all evolved ultrasensitive networks. The orange coloured dots represents kinases and the blue dots represents phosphatases. The numbers on the dots denote the network number.

While enzyme saturation mediated zero-order sensitivity is well understood theoretically, this biochemical mechanism might not be relevant for many biological systems where the ratio of substrate to enzyme concentrations is found to be low [97, 222]. To explore whether ultrasensitivity can still emerge under such conditions, I ran evolutionary simulations with equal starting concentrations for the substrate

and signalling proteins. Although concentration of signalling proteins could freely evolve in these simulations, enzyme saturation was expected to be difficult to evolve, which could lead to evolution of alternative mechanisms for ultrasensitivity. Indeed, the emerging ultrasensitive networks from these simulations (10 out of 30 simulations) did not display the kinetic parameters required for enzyme saturation (Figure 3.2A, red points). Together with three ultrasensitive networks that started with high concentrations of the substrate, but did not evolve enzyme saturation, these ultrasensitive networks clearly utilize mechanisms other than enzyme saturation.

Analysing the structure of these networks (Figure 3.3), I did not find any distinct structural features. However, I found that in many evolved networks with parameters in the non-saturating regime, there is a high prevalence of enzyme sequestration (Figure 3.2B) and also allosteric regulation of enzyme activity (Figure 3.2C) by other signalling proteins. In theory, allosteric regulation of enzyme activity by upstream proteins that are activated by signals could implement a form of ultrasensitivity [87, 88, 223] that could relax the need for enzyme saturation. I found that for at least some networks, the ratio of allosteric forms of the enzymes barely changes across the input range (Figure 3.2D), showing that allosteric regulation is not the main or sole process enabling ultrasensitivity. This suggests a more general role for enzyme sequestration, which prompted us to analyse all of the evolved networks with regard to the prevalence of the different enzyme complexes. In particular, I calculated the average proportions of ES complexes, formed by enzyme binding to substrate, and ET complexes, formed by enzyme binding to other proteins (Figure 3.4). Note that these proportions can be seen as the average level of enzyme saturation by the substrate and sequestration by other proteins in the signalling network. This analysis revealed that most of the ultrasensitive networks evolved parameters that resulted in enzymes being bound in complexes (i.e. they lie close to the line given by $[ET] = 1 - [ES]$). Moreover, contrasting the results of evolutionary simulations where enzyme saturation was made difficult to evolve vs.

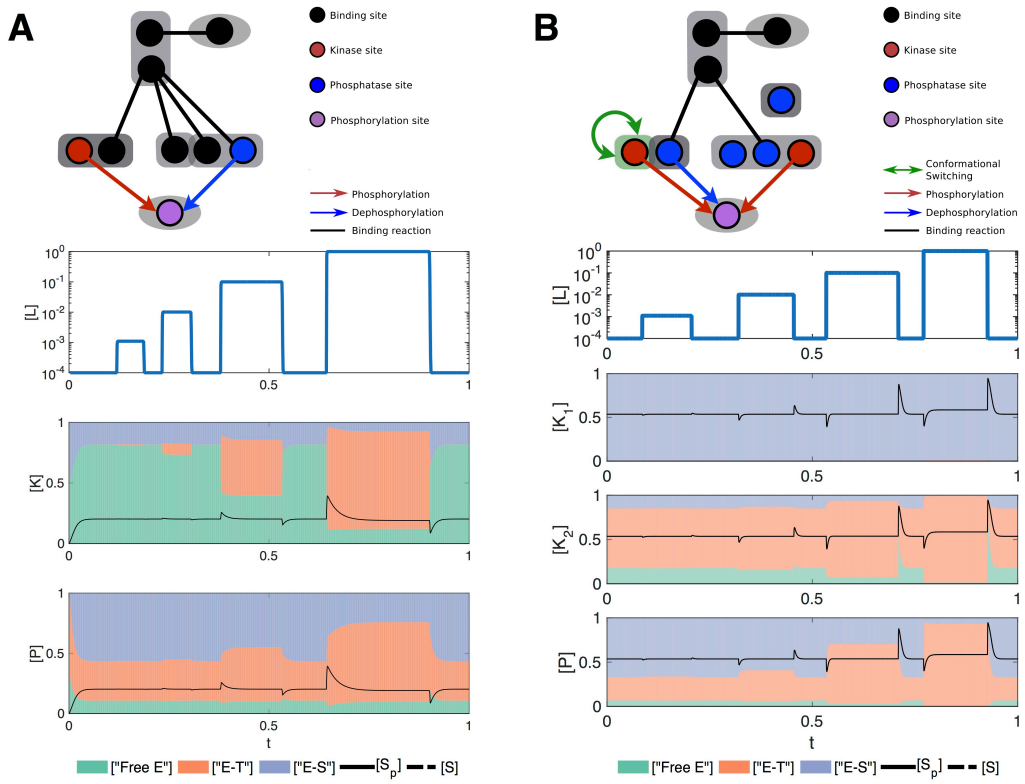


Figure 3.5: Analysis of evolved adaptive networks. **(A)** Structure and dynamics of the evolved adaptive network 1. The upper panel shows a cartoon of the network. The oval shapes represent ligand (top) and the output protein (bottom) (e.g. substrate with a phosphorylation site, S), while all other signalling proteins (e.g. receptor/adaptor proteins, kinases, or phosphatases) are shaped as rectangle. Black line represents binding reaction between two sites. Red arrows represent phosphorylation reactions between a kinase site (red) and a phosphorylation site (purple). Blue arrows represent dephosphorylation reactions between a phosphatase site (blue) and a phosphorylation site. The green coloured rectangle indicates a protein domain, whose conformational switching is allosterically regulated (also indicated by a self-pointing green line with arrows at both ends) [212]. The lower panel shows the dynamics of input signal and output response. The stacked colours represents the compositions of enzyme complexes: blue for proportion of enzyme-substrate complexes, green for free form enzymes that are accessible by the substrate, and red for complexes where enzymes are not accessible by the substrate (i.e. titrated enzymes). **(B)** Structure and dynamics of the evolved adaptive network 2. Panels are as in (A).

not, showed that the former scenario resulted in enzymes that were mostly titrated by other signalling proteins (see Figure 3.2A and Figure 3.4). These results suggest that when enzyme saturation is not readily achievable, evolution of ultrasensitivity was made possible mostly through enzyme sequestration. I analysed this proposition further with a simpler model (Figure 3.6).

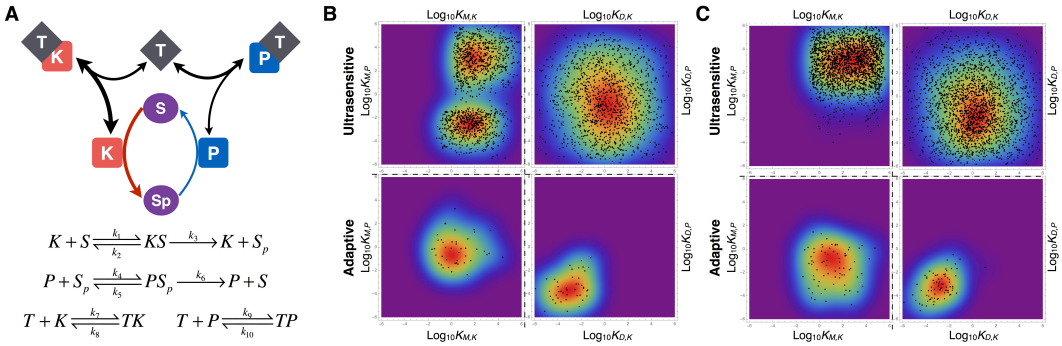


Figure 3.6: Designed signalling cycle motif and parameter space for adaptation and ultrasensitivity. **(A)** Cartoon showing the designed signalling cycle motif with a sequestering protein. The sequestrating protein T binds both the kinase K and phosphatase P, which catalyse the phosphorylation and dephosphorylation of substrate S and S_p respectively. **(B)** The values of key parameters for achieving ultrasensitive (> 0.8) and adaptive response (> 0.3), when assuming an enzyme-saturated regime ($[S_{total}] = 1$, $[P_{total}] = 0.1$). The upper and lower two panels are distribution of parameters that generate ultrasensitive and adaptive responses respectively. Panels on the left show the distribution of Michaelis-Menten constants, for kinase: $K_{M,K} = \frac{k_2+k_3}{k_1}$ (x-axis) and phosphatase $K_{M,P} = \frac{k_5+k_6}{k_4}$ (y-axis). Panels on the right show the distribution of affinities of sequestrating protein T with kinase and phosphatase: $K_{D,K} = \frac{k_8}{k_7}$ and $K_{D,P} = \frac{k_{10}}{k_9}$. Note that all four panels are plotted on the same logarithmic range. Each black dot represents a parameter set and the colours shows density of parameters. **(C)** Values of key parameters for achieving ultrasensitive (> 0.8) and adaptive response (> 0.3), when assuming an enzyme-non-saturated regime ($[S_{total}] = 0.1$ and $[P_{tot}] = 0.1$).

3.3.2 Selection for adaptive dynamics leads to networks employing differential enzyme sequestration

In order to select networks with adaptive response dynamics, I improved the adaptive fitness function by forcing the system to displaying adaptive response dynamics under input signals of several different magnitudes (*Methods*). This fitness is moti-

vated by the fact that fitness function imposing under a single input signal level will results in a pseudo-adaptation (Figure 2.5B). I found only few of the evolutionary simulations resulting in networks with adaptive dynamics (2 out of 60 simulations), potentially due to the strictness of this fitness function. Interestingly, in both of these simulations, the final evolved networks contained a protein, the role of which implements a differential sequestration of the enzymes, e.g. by sequestrating them through different number of binding sites (Figure 3.5A). The imbalanced sequestration affinity of the scaffold protein towards kinases and phosphatases enables the system to provide an initial response to a change in signal but then move back to same equilibrium points (Figure 3.5B). With every signal step, the kinase is titrated much faster compared to the phosphatase leading to an initial response that then settles back to previous levels when sequestration levels of the kinase and phosphatase equilibrate (Figure 3.5). When the scaffolding protein is fully bound, and the sequestration effect cannot operate anymore, the level of adaptation to signal is hampered (see Figure 3.5B).

3.3.3 Scaffolding protein enables ultrasensitivity and adaptive dynamics in a single signalling cycle

Interestingly, I find that scaffold proteins acting on both kinases and phosphatases as seen in evolved adaptive networks are also featured in evolved ultrasensitivity networks (see network 4 and network 13 in Figure 3.3). This suggests that such proteins could allow implementation of both adaptive and ultrasensitive dynamics. To test this idea, I developed a model of the simplest possible signalling cycle motif that features enzyme sequestration, and where incoming signals are implemented as changes in kinase concentration Figure 3.6A (also see *Methods* and Appendix C). I analysed the capacity of this model to generate ultrasensitive and adaptive responses by sampling 100,000 independent sets of kinetic parameters in a biologically feasible regime (*Methods*). I find that this generic model can achieve both adaptive and

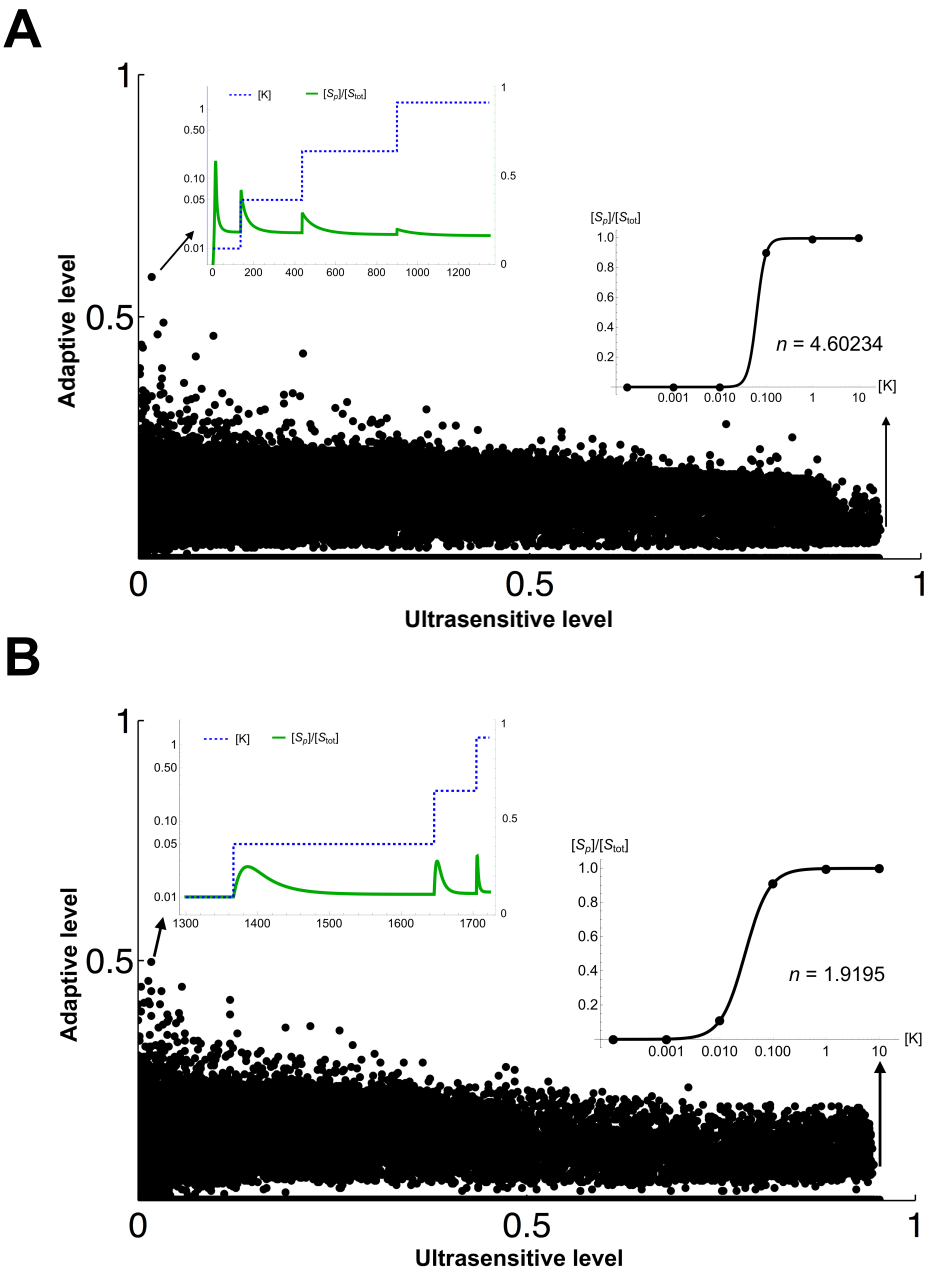


Figure 3.7: Parameters sampling of the signalling cycle in the phenotype space of ultrasensitivity and adaptation. **(A)** Sampled network parameters under substrate-saturating condition ($[S_{tot}] = 1$ and $[P_{tot}] = 0.1$). The x-axe shows the adaptive score of those sampled parameter sets, the y-axe shows the ultrasensitive score. Also both the most adaptive network (top-left corner) and the most ultrasensitive network (bottom-right corner) are shown. **(B)** Sampled network parameters under non-saturating condition ($[S_{tot}] = 0.1$ and $[P_{tot}] = 0.1$).

ultrasensitive dynamics, irrespective of imposing enzyme-saturating conditions or not (Figure 3.7).

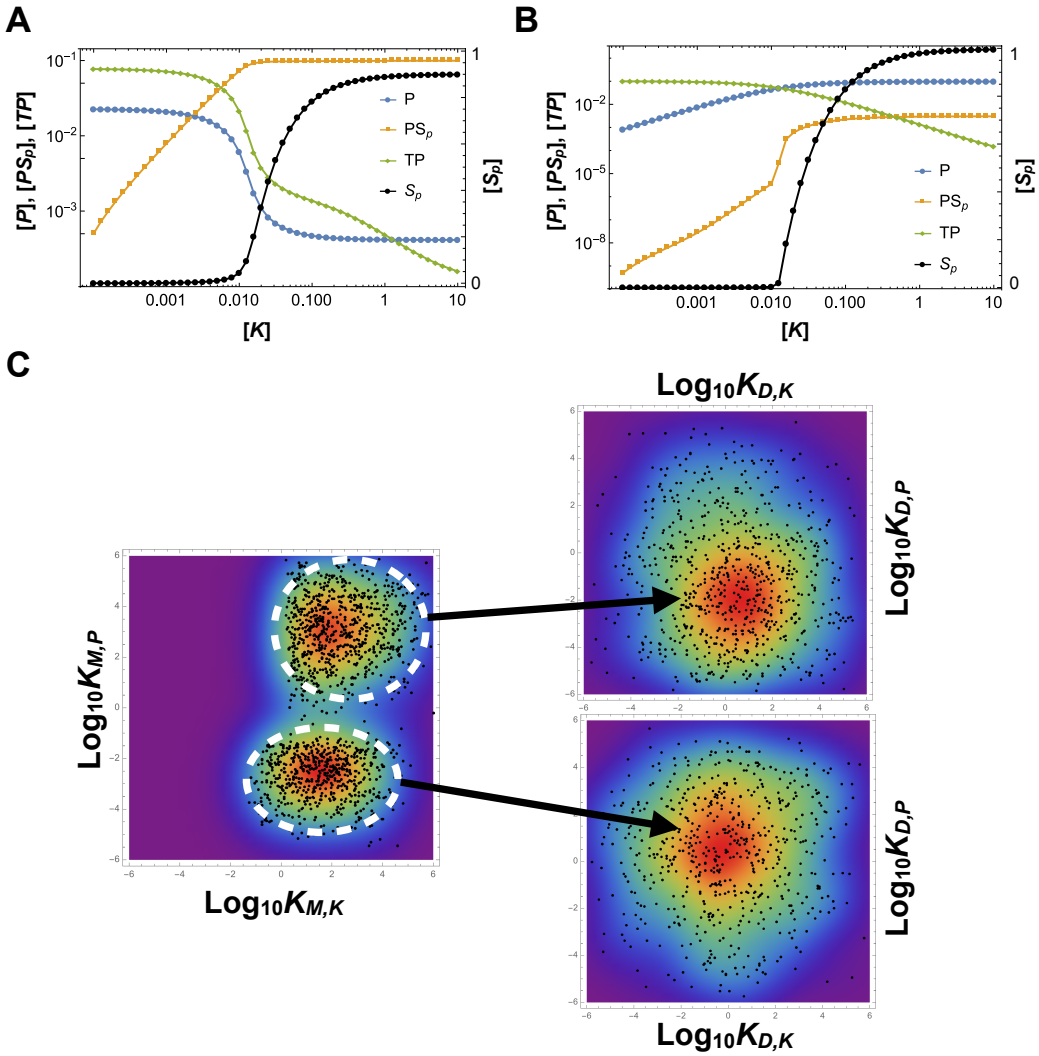


Figure 3.8: Two different parameter regimes for ultrasensitivity. **(A)** Dynamics of phosphatase in the first parameter regimes (low $K_{M,P}$ and high $K_{D,P}$). **(B)** Dynamics of phosphatase in the second parameter regimes (high $K_{M,P}$ and low $K_{D,P}$). **(C)** Separation of two different parameter regimes in parameter space of $K_{D,K}$ and $K_{D,P}$.

For the case of ultrasensitive dynamics, analysis of all "successful" parameter sets showed two distinct parameter regimes leading to ultrasensitive dynamics (Figure 3.6B). These regimes relate to enzyme saturation (i.e. large or small K_M

values); in one regime, the phosphatase has high affinity for the substrate and is fully saturated by it (small $K_{M,P}$), while the kinase has high affinity for the sequestrating protein (Figure 3.8A, C). In the second parameter regime, both the kinase and the phosphatase have large K_M values indicating a lower affinity for the substrate. Thus, the enzymes are mainly bound to the sequestrating protein (small K_D values) and are in competition for that protein (Figure 3.8B, C). In both parameter regimes, small increase of incoming signals (i.e. small increase in kinase concentration) can be "absorbed" by increased sequestration of the kinase, while higher signal levels saturate this sequestration-mediated effect, resulting in significant amounts of free kinase and resulting in a switch to high phosphorylation rates. The difference between the two parameter regimes is that in the second regime, competition among kinase and phosphatase for the sequestrating protein results in an additional feedback, where increased kinase levels enhance free phosphatase levels (through release from the sequestering protein)Figure 3.8A, B. As expected from this analysis, I find that ultrasensitivity can only be generated in the second parameter regime (i.e. large K_M values and small K_D values) when I sample parameters by forcing either enzyme to be unsaturated by the substrate (Figure 3.6C).

In the case of adaptive dynamics, I find that the parameter regime leading to highly adaptive networks corresponds to competition among kinase and phosphatase for the sequestering protein (i.e. small K_D values) (Figure 3.6B, C). In this case, incoming signals temporarily shift this competition towards the free kinase, but subsequently, the kinase binds the sequestrating protein in expense of the phosphatase. The resulting release of the phosphatase results in the balancing of the phosphorylation and desphosphorylation rates, leading to adaptive dynamics (Figure 3.6B, C).

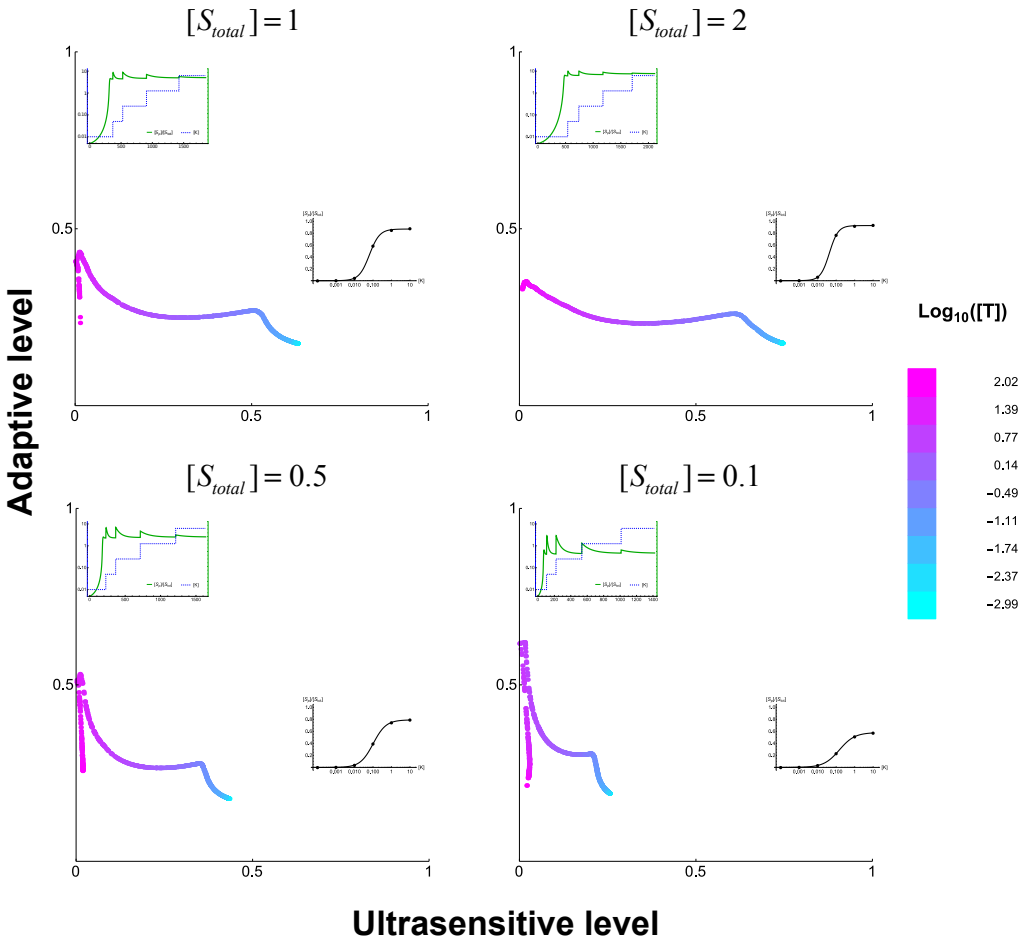


Figure 3.9: Modulation of response dynamics through tuning of scaffold protein concentration. The four panels show sampling the total concentration of sequestrating protein, $[T_{total}]$, when fixing all other parameters and with the total concentration of substrate $[S_{total}]$ as shown on the panels. The colour of each data point represents sequestrating protein concentration. In each panel, the best ultrasensitive or adaptive response dynamics that are achieved at a specific $[T_{total}]$ level are shown. Best adaptive response is shown with blue dashed line as input $[K]$ and green line as output $[Sp]$, while the best ultrasensitive response is shown with dots as steady state response level when presented with input $[K]$.

3.3.4 Scaffolding protein can act as a tuning point to generate plastic response dynamics

The intriguing similarity of the mechanisms for adaptive and ultrasensitive dynamics suggests that a single system could implement both dynamics. In particular, I note that there are parameter sets at the edges of the distinct parameter regimes leading to ultrasensitive and adaptive dynamics (Figure 3.6B). Is it possible that such parameter sets result in system where response dynamics can be modulated by the dynamics of the sequestrating protein? In order to answer the question, I sampled the concentration of sequestrating protein ($[T_{total}]$), while fixing all other parameters to check if simply varying the level of T could modulate the response dynamics. A few systems showed such modulation, where systems behave with adaptive and ultrasensitive dynamics at two distinct total concentrations of T (Figure 3.9). Interestingly, this modulation is influenced directly by the total concentration of substrate $[S_{total}]$; at high (low) substrate concentration $[S_{total}]$ modulation by $[T_{total}]$ allows an extended shift towards ultrasensitivity (adaptive) rather than adaptive (ultrasensitive) dynamics (Figure 3.9). I found that altering the affinities between the sequestrating protein T and the enzymes can also implement a similar modulation (Figure 3.10). These results show that varying concentration and/or affinities of sequestrating protein can modulate the plastic response dynamics. Furthermore, it is possible that in more complex networks (like those resulting from the evolution simulations) such response modulation is embedded within the network dynamics (i.e. scaffolding protein dynamics is allosterically regulated by other proteins or directly by the signal).

3.4 Discussion

Here, I used *in silico* evolution implementing a biologically realistic rule-based model of proteins to evolve signalling networks displaying ultrasensitive and adaptive re-

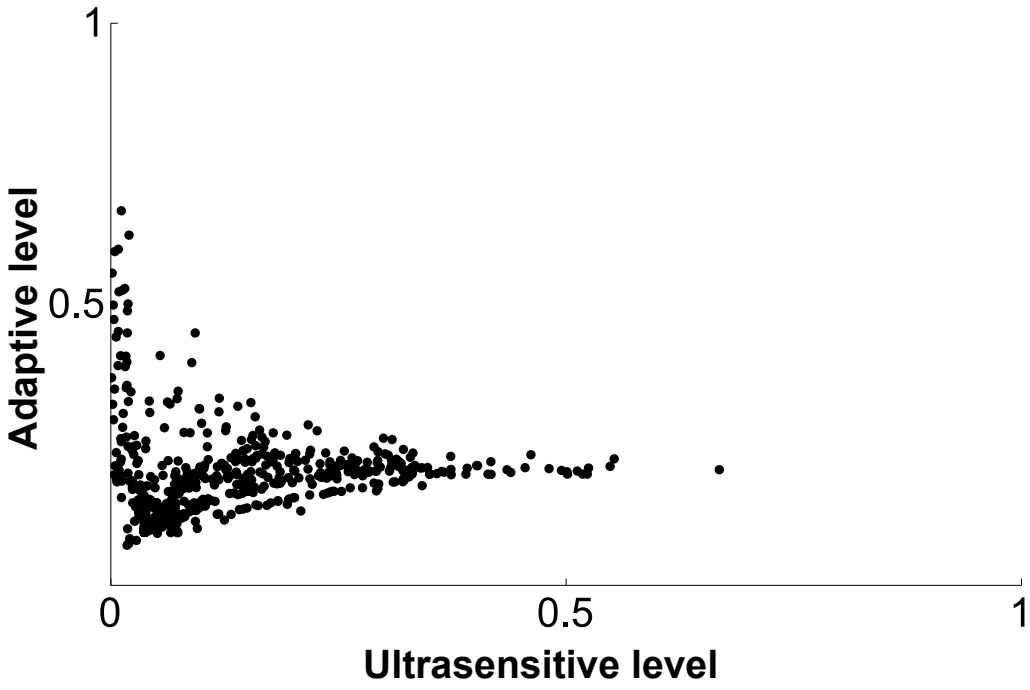


Figure 3.10: Modulation of response dynamics through altering affinity between scaffold protein and enzymes. Sampling only affinity parameters (k7, k8, k9, k10) while fixing all other parameters, x-axe represents adaptive score and y-axe represents ultrasensitive score.

sponse dynamics. Running evolutionary simulations from different starting structures and under conditions of enzyme saturation or not, I found that enzyme sequestration by scaffolding proteins is a key network feature enabling these dynamics. These results from the *in silico* evolution allowed us to design a simple network motif that can implement both adaptation and ultrasensitivity with different kinetic parameters and concentrations of the scaffolding protein. These findings highlight the role of scaffolding proteins can play in natural systems and synthetic biology applications as control point of response dynamics. However, the starting structures in this study are inspired from conserved signalling network motifs, it is possible that some of these evolved networks stay in a local optimum. A possible alternative approach is to evolve random starting networks structures by using random generated binary strings as starting network models.

In natural systems, the scaffold proteins are usually controlled by transcriptional regulation [224] that changes the concentration of scaffold proteins and/or by post-translational modifications and allosteric regulations [43] (e.g. phosphorylation, ubiquitination) that alter the affinities to their binding substrates. This suggests that evolved natural signalling systems exploit scaffolding proteins to enable diverse and/or plastic response dynamics. In particular, scaffold proteins are ubiquitously distributed in cellular signalling networks [115] and several experimental studies have shown their involvement in regulating response dynamics [43, 163, 224, 225]. Additionally, I note that the kinase and phosphatase sequestration described here is similar to bifunctional enzymes mediating robust homeostatic dynamics as identified in several biological systems [203, 226]. A possible explanation on why natural biosystems need such plastic response dynamics would be that such plastic response dynamics enable biological systems to adapt different environments. When fluctuating environment changes from one to another, the system can change their response accordingly. One related interesting questions would be under what kind of fluctuating environments are the plastic response dynamics beneficial.

From a synthetic biology perspective, our findings provide key insights on how altering scaffolding proteins can directly alter response dynamics. Synthetic manipulation of allosteric properties and/or concentrations of scaffolding proteins in the MAPK signalling pathways is already shown to result in diverse response dynamics [227–229]. It is also increasingly possible to induce or change interaction of enzymes with scaffolding proteins through alteration of common interaction domains [155], which could allow introduction of new scaffolding protein in specific systems. These experimental methods, when combined with the theoretical insights presented here can lead to scaffold proteins becoming a key engineering point for directing and manipulating signalling dynamics as noted before [168, 230, 231].

The presented study, as well as similar studies [123, 145], show that *in silico* evolution can be utilized as a useful approach to discover additional biochemical principles that are not readily discovered in experimental model systems or through analysis of conserved structural features. The ability of evolutionary simulations to provide sample systems implementing a specific functionality allows generation of hypotheses that can be subsequently tested in experiments (e.g. [95, 123]) or verified using minimalistic models, as I have done here. Thus, evolution *in silico* can provide us with insights on biochemical features that natural evolution has so successfully exploited. These features can act as evolutionary design principles that can further our ability to engineer de novo biological systems and understand the natural ones.

Chapter 4

Emergence and Design of Networks with Bistable Dynamics

4.1 Introduction

Ultrasensitivity itself is a potentially important biochemical function that allows systems respond sensitively to the modest signal changes and ignore signals that are either too low or too high. The threshold is formed by the discriminative sensitivities to different signal levels. In many biological processes, the ultrasensitivity functions as either a filter that removes the background noise or a decisive controller that switches cellular states [62]. In terms of cell state transition, bistable dynamics where the system has two distinct stable steady state can also control cell state transitions [75, 76]. When the system switches between different states, bistable response dynamics typically has hysteretic transitions which distinguish bistable dynamics from ultrasensitive but monostable systems (Figure 1.1A & 1.1C). Such hysteresis forms biochemical memory that is of great interests for synthetic biology applications. Interestingly, one more important role of ultrasensitivity is its capacity

to enable more complex dynamics such as bistability and oscillation [40][62]. In this chapter, I will discuss the emergence of bistability from previous evolutionary simulations. By dissecting the evolved bistable networks, I find the bistable units in evolved networks featuring allosteric regulation on enzymes of futile cycles, which shed light on potential design principles for bistability in signalling networks. Also inspired by the emergence of bistable dynamics, I use the fitness function to design more bistable networks where allosteric regulations are not permitted.

4.2 Methods

4.2.1 Evolutionary simulations

Previously 60 evolutionary simulations are carried in Chapter 3, where 10 independent simulations each for two different starting concentration conditions and three different starting structure conditions. Additional 60 evolutionary simulations are carried with configurations where allosteric regulations are not allowed in evolving networks (see *Methods* in Chapter 2 for detail). The additional 60 simulations started with the same conditions from previous 60 simulations, the fitness function for selecting ultrasensitive response dynamics is also the same as of previous 60 simulations.

4.2.2 Chemical reaction network toolbox

For determining the existence of multistationarity of given signalling networks, I utilised the Chemical Reaction Network Toolbox (CRNToolbox)*. Given a chemical reaction network described with mass action kinetics, CRNToolbox can determine whether multiple equilibria exist with any positive kinetic parameters. I analysed the existence of multistationarity in several different signalling networks given the chemical reactions in the networks. An example of CRNToolbox report can be found

*https://crnt.osu.edu/CRNTWin

in Appendix D. The detail usage of CRNToolbox is described in its manual.

4.3 Results

4.3.1 Bistability emerges from previous evolutionary simulations

In Chapter 3, I discussed what type of structures and biochemical mechanisms emerged under different selection pressure. One of the fitness function is ultrasensitivity which imposes an *ad hoc* threshold. The evolved systems have large responses (change of output response level) when perturbation of input signal happens near the threshold and small responses when it happens far away from the threshold. Interestingly, when I study the dose response curves of all evolved ultrasensitive networks (Figure 3.3), there are 7 networks whose dose response curves show clear hysteresis near the threshold (Figure 4.1). From low level of output to high level of output or *vice versa*, the hysteretic transitions in these networks indicate there are two distinct stable steady states in their dynamics. When input signal is in the hysteretic area, the system has two distinct levels of output response (i.e. they are bistable). However, which state the system stays in depends on the historical state where the system comes from [107, 232–235].

Examining the fitness function for selecting ultrasensitive response dynamics in Chapter 2, the fitness function actually selects for a wide threshold in the response dynamics, where signals respond most (Figure 2.2A). The hysteresis in evolved bistable networks provides sufficient threshold such that the system response mostly in the hysteretic range of input signal. Therefore, evolved bistable response dynamics is one of the possible solutions to such fitness function which was intentionally designed to selects ultrasensitive response dynamics. However, the bistable dynamics is different from ultrasensitive dynamics in the sense that the latter is monostable and without hysteresis.

The evolved bistable networks are rather complex in terms of combinatorial

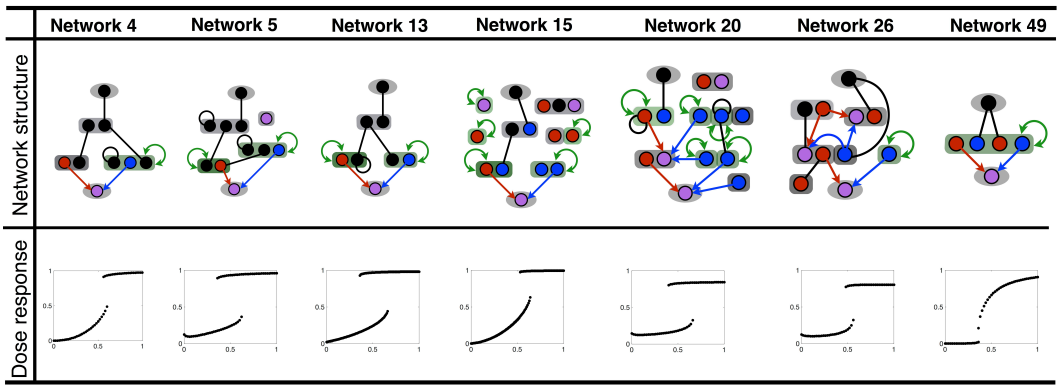


Figure 4.1: The bistable networks emerged from evolutionary simulations and their dose response curves.

interactions between signalling proteins with multiple domains (Figure 4.1). It is difficult to map their structures with the underlying mechanisms where bistability emerges. In previous studies, bistable dynamics in biological systems are commonly linked to positive feedback loops which are immediately observed from schematics of gene regulatory networks [236]. Mathematical proofs also showed that positive feedback loops is a requirement for gene networks displaying multistationarity [105]. However, the positive feedback loops required for bistable dynamics are not directly observable from all evolved bistable signalling networks. Such determinant feedback loops ought to be hidden in the complex interactions. Furthermore, investigations on MAPK signalling pathway showed that phosphorylation and desphorylation cycles of proteins with multiple phosphorylation sites can result in multistationarity even though no revealing positive feedback loops can be found in the structure of signalling cycles [106]. Examining the evolved bistable networks, there is no protein with multiple phosphorylation sites. Altogether these hint that evolved bistable networks are installed with new bistable motifs to enable hysteresis and bistability.

4.3.2 Dissecting evolved bistable networks to obtain bistable units

In order to understand the underlying bistable motifs, I started to dissect the structure of evolved bistable networks. When reducing the size and complexity of net-

works, I utilise the chemical reaction network toolbox (CRNToolbox) (see *Methods*) to examine whether simplified networks allow bistable dynamics with any positive real values of rate constants and concentrations. The CRNToolbox use chemical reaction networks theory (CRNT) [237, 238][239][240] to check several qualitative properties of chemical reaction networks with mass-action kinetics. One such property is the existence of multistationarity with parameters in positive real domain. Since all parameters are relaxed as any positive values, the toolbox checks the properties of reaction networks based on their structure only. This parameter-free approach can help us find the minimal structure basis of multistationarity in evolved networks. In each step, I simplify the evolved networks by removing a signalling protein or an interactions in the network, then use CRNToolbox to check if the network is still bistable. The network is simplified until it becomes monostable. Then the minimal network structures can be considered as candidate subnetworks enabling multistationary property in evolved networks [241]. I started from a relatively simple network (Network 15 in Figure 4.1) and continued simplifying the network results in smaller and smaller network structures that still allow bistable dynamics (Figure 4.2).

All evolved networks contain allosteric regulations that did not exist from where the evolutionary simulations started (Figure 4.1). I firstly take a route to reduce the size of network while keeping the allosteric regulation on the kinase. The derived smallest bistable subnetwork is composed of one phosphorylation-dephosporylation cycle with an allosteric enzyme where the kinase has two distinct conformational states that switch between each other. Further simplifying this bistable subnetwork by removing allosteric reactions results in the subnetwork of phosphorylation-dephosphorylation cycle which is monostable. The monostable cycle is exactly the same as the zero-order sensitivity model [92]. This supports the hypothesis that allosteric enzymes are important for bistable dynamics in signalling networks. This bistable subnetwork is thus one of the simplest motifs for generating

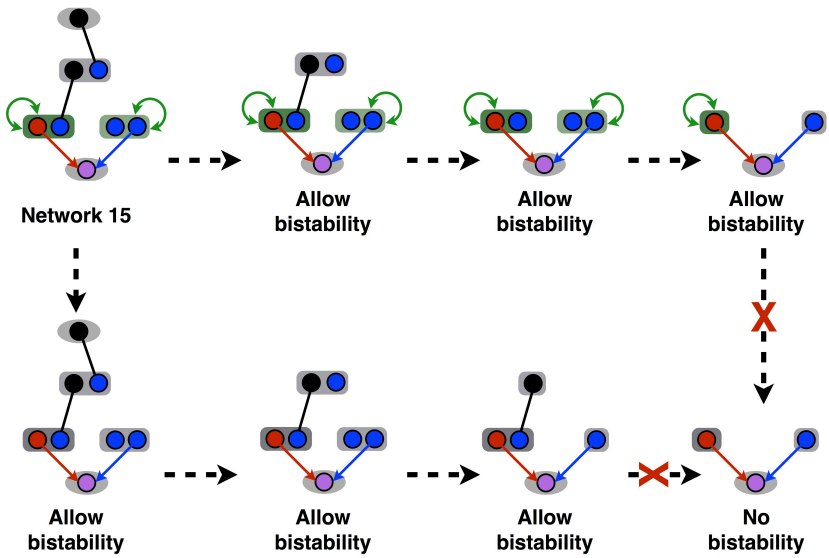


Figure 4.2: Simplification of an example bistable network emerged from evolutionary simulations (Network 15).

bistable dynamics in evolved networks.

Either the phosphorylation-dephosphorylation cycle or allosteric regulated model (MWC model) alone can generate monostable but possible ultrasensitive response dynamics [87, 92]. The uncovered bistable motif requires both present. In order to examine the necessity of allosteric regulation in evolved bistable dynamics, I took another route of simplifying the evolved network (Figure 4.2). The resulted bistable subnetwork is a phosphorylation-dephosphorylation cycle with another scaffold protein with capacity binding the kinase (Figure 4.2). Although this subnetwork does not show explicit allosteric reactions (i.e. reactions of switching between different conformational states), different catalytic capabilities in two different states of enzyme (i.e. bound and unbound with scaffold protein) again require either allosteric effect or steric effect on the enzyme's catalytic ability. These two subnetworks show the important role of allosteric enzymes combining with futile cycles in bistable signalling networks. Further analysis of allsoteric enzyme's role as a design principle of bistability appears in Chapter 5.

4.3.3 Design of bistable networks without allosteric regulations

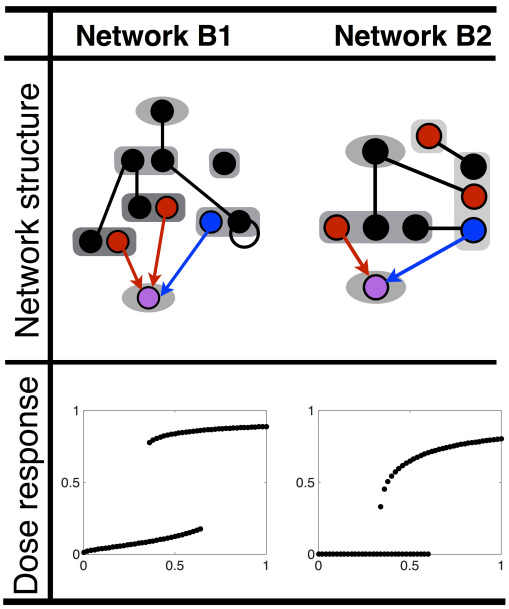


Figure 4.3: Designed bistable networks without allosteric regulations.

Since the evolutionary simulation with the discussed fitness function allows bistable response dynamics to occur, the fitness function can be used as a objective function to design bistable signalling networks. More interesting questions are whether there are other patterns for bistable dynamics in signalling networks and whether this approach can be used to design such bistable signalling networks. To prove this hypothesis, I ran another 60 simulations with the same starting conditions and fitness function as before except that no allosteric regulation are allowed to evolve or occur. In this setting, the bistable motifs with allosteric enzymes discovered from previously evolved networks can not appear in the currently evolving networks.

These evolutionary simulations have resulted in only 3 networks that become "ultrasensitive" (fitness score > 0.8). However, from those 3 networks, two of them have hysteresis in their dose response curve thus are bistable (Figure 4.3). This clearly shows that there are mechanisms other than allostery that are endowing the

evolved networks with bistable dynamics. Using the similar deducing approach, I dissect one of the two bistable networks (Network B2) and derived a simple bistable motif (Figure 4.4) featuring a futile cycle with both enzymes sequestrating each other. This bistable motif is similar to the signalling cycle with both enzyme sequestrated by a common scaffold protein (Chapter 3), however that motif is not bistable for any parameters that are positive (Figure 4.4).

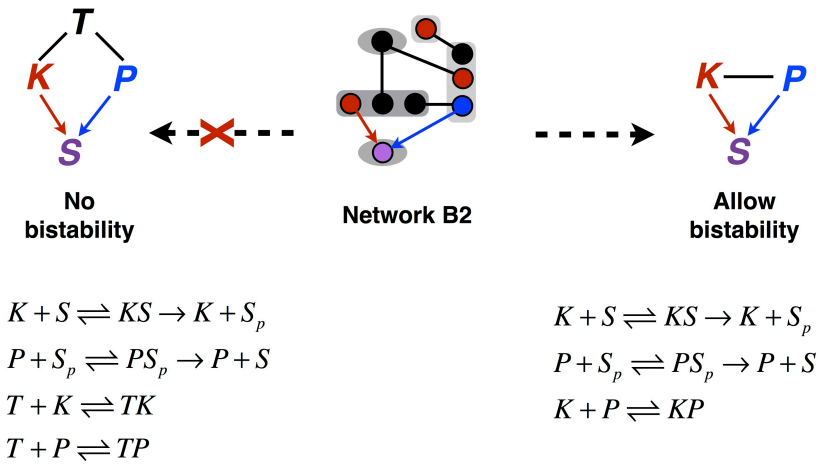


Figure 4.4: Simplification of designed bistable network (Network B2)

4.4 Discussion

For many different complex dynamics in biological systems, it is very difficult to depict their genotype-phenotype mapping. One particular reason is that the mapping between genotypes and phenotypes are rather complex than one to one mapping. Study in excitable gene regulatory circuits, two different architectures of the circuits both emerged as solutions for pulse dynamics however with diffferent noise dependency and tunability [201, 242, 243]. Such complex genotype-phenotype mapping also appears in previous studies on the evolution of bistable switches in gene regulatory networks where three different mechanisms evolved as the solutions for

bistable switches in gene networks, one of which is not reported in natural biological systems [123] but latter implemented experimentally [95]. The emergence of bistable dynamics in this study, where I evolved signalling networks with the *ad hoc* fitness function with intention of selecting ultrasensitive response dynamics, is another indication that solution to certain response dynamics in cellular networks is not simplex, rather that many different solutions may emerge.

The bistable dynamics has its biological significance. From the perspective of systems dynamics, the systems will rest in one state depending on their historical position because of hysteretic transitions. In biological systems, bistable systems generate heterogeneous responses in a population of cells. Especially, such heterogeneity in microbes is considered as a bet-hedging strategy that enable survival of the species in fluctuating environments [244]. For multicellular organisms, the bistable dynamics determines the cell fates in differentiation and development [75, 233, 245]. Therefore, it is possible to design and engineer bistable dynamics in biological systems. In particular, this Chapter showed that the *ad hoc* fitness function can potentially be used as an objective function to design and optimise bistable signalling networks through evolutionary simulation.

Furthermore, understanding the design principles of bistable dynamics can be of great interest and benefit. Not only such design principles can help us understand natural biological systems, but also allow us to design and build novel bistable biological systems. Those subnetworks derived from evolved bistable networks provide specific design pattern for bistable dynamics in signalling networks (Chapter 5). Particularly, the futile cycle with allosteric enzymes is prevalent in biological systems, and the detailed analysis of this systems reveals interesting design principles, which is provided in Chapter 5. However, the structural conditions that distinguish multistationary signalling networks from monostationary ones are rather subtle and unapparent. Comparison between monostationary signalling networks and multistationary ones can potentially uncover the definitive structural

patterns of multistationary signalling networks. Such structural patterns are the necessary and sufficient structural conditions on the capacity allowing multistationarity of a general mass-action chemical reaction network. This motivated me to construct an algorithmic approach to enumerate reaction networks on the purpose of searching the boundary between monostationarity and multistationarity in signalling networks. This work is introduced in Chapter 6.

Chapter 5

Core signalling motif displaying multistability through multi-state enzymes

5.1 Introduction

Cells sense environmental stimuli and use these to initiate appropriate physiological responses. Understanding such cellular information processing in healthy and diseased states [246–248], and engineering it through synthetic biology [111, 249–251], requires better insights into the relation between different interaction motifs found in signalling networks and their potential roles in the ensuing system dynamics [77]. To this end, a key interaction motif found predominantly in eukaryotic signalling systems is that of a futile signalling cycle, where a substrate protein is phosphorylated by a kinase and dephosphorylated by a phosphatase. When these enzymes are saturated by their substrate, this motif can display ultrasensitive response dynamics, enabling threshold responses to graded input signals [92]. It can also be shown theoretically, that the futile cycle motif in its simple form cannot enable bistability (see below). Experimental studies of cellular systems embedding the futile signalling

cycle for several physiological responses, including cell fate determination and cell division [23, 43, 72, 193], found ultrasensitive-responses and in some cases bistability [75, 199, 236, 252–256]. While the presence of bistability has been indicated to be functionally significant, for example in the generation of phenotypic variability [76, 257–259], its molecular implementations have not been fully elucidated.

To achieve bistability in a futile signalling cycle motif, the originally studied structure of this motif needs to be extended with additional features. Theoretical studies have shown that bistability can be achieved if there are feedback interactions between the substrate and its acting enyzmes (i.e. the kinase or phosphatase) [11, 48, 241, 260, 261], or if the substrate has multiple phosphorylation sites [52, 54, 106, 262]. The latter proposition is particularly interesting as the presence of multiple phosphorylation sites on signalling proteins is a common phenomenon [263, 264]. Kinases, phosphatases, as well as their substrates readily exhibit two or more conformational states that are associated with different levels of phosphorylation and that result in different catalytic activity levels [227, 265, 266]. In the signalling pathways regulating the cell cycle for example, it has been hypothesised that signalling proteins with multiple phosphorylation sites act as multi-state enzymes that can embed complex signal-processing [265–268]. It is also shown that the different activity levels of signalling proteins can be regulated through allosteric interactions with ligands or other proteins, such as so-called scaffolding proteins [216, 224, 225, 227]. Scaffolding proteins, which are ubiquitous in signalling systems [224, 225], can also have multiple phosphorylation and binding sites themselves and, as such, are key regulators in signalling pathways [43, 163, 269–271]. Despite these experimental findings and observations on specific signalling proteins and pathways, it has been difficult to elucidate any particular features, or design principles, that can provide a clear understanding between the nature of signal processing that a system implements and the presence of multi-phosphorylation-site-featuring, multi-state enzymes. This difficulty arises partially from the fact that modelling of signalling pathways with

multi-state enzymes becomes increasingly complex, with a combinatorial explosion of possible interactions in the system.

In Chapter 4, it has been demonstrated that only a futile cycle cannot display bistability (also see), however this result changes and bistability becomes possible if we consider the allosteric nature of kinases and phosphatases (Figure 4.2). In this chapter, I perform a systematic, mathematical analysis of the effects of having multi-state kinases on the response dynamics and the number of steady states in this simple and core futile signalling cycle motif. I first show that when this motif is analysed with the assumption of single-state enzymes, the resulting system cannot display bistability for any positive kinetic parameter values. This situation changes and bistability becomes possible only with the introduction of a two-state kinase, leading to one of the smallest signalling systems that is bistable. Using this minimalist system as a tractable core motif, I am able to derive mathematical conditions on the kinetic parameters and/or the total concentrations of substrate and kinase that are necessary and sufficient for the existence of three steady states. This allows an intuitive insight that bistability in this minimalist system arises from the competition between the different kinase states for the substrate. Extending from this intuition, I show that increasing the number of kinase states in the system leads to a linear increase in the number of steady states. I show that both multi-state enzymes and the discussed core motif are prevalent in many signalling pathways and that the identified parameter ranges for bistability are biologically plausible. These results provide an intuitive view on multi-state enzymes leading to bistability and multistability through competition for their substrates. As such, the multi-state nature of enzymes can be exploited to better understand natural signalling pathways and to engineer novel ones.

5.2 Methods

5.2.1 Model for a futile signalling cycle with two-state kinase

The core futile signalling cycle I consider here has been considered before in seminal works and consists of a covalent modification, i.e. de/phosphorylation, of a substrate by a kinase and a phosphatase [47, 92]. In Chapter 4 I derived one of the smallest signalling motif which is similar to the futile signalling cycle but with an allosteric kinase. Here, I take the bistable signalling motif with an allosteric kinase. For the case of the allosteric kinase, I consider two distinct states (K_r and K_t) catalysing a substrate (S) into product (S_p). To simplify the system, I do not model the phosphatase directly, but rather consider the reverse reaction as an auto-dephosphorylation reaction. The corresponding reactions including kinase transformations between different states and considering catalytic reaction cycle is given by:

$$
\begin{aligned}
K_r + S &\underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} K_rS \xrightarrow{\kappa_3} K_r + S_p \\
K_t + S &\underset{\kappa_5}{\overset{\kappa_4}{\rightleftharpoons}} K_tS \xrightarrow{\kappa_6} K_t + S_p \\
S_p &\xrightarrow{\kappa_7} S \\
K_r &\underset{\kappa_9}{\overset{\kappa_8}{\rightleftharpoons}} K_t \\
K_rS &\underset{\kappa_{11}}{\overset{\kappa_{10}}{\rightleftharpoons}} K_tS,
\end{aligned}
$$

where, the parameters κ_1, κ_2, $\cdots$, κ_{11} represent the kinetic reaction rates. The system is composed of 6 species, of which two are complexes. Based on the reaction network, I constructed a mathematical model containing a set of 6 ordinary

differential equations:

$$
\begin{aligned}
\frac{\mathrm{d}[K_r]}{\mathrm{d}t} &= -\kappa_1[K_r][S] + \kappa_2[K_rS] + \kappa_3[K_rS] - \kappa_8[K_r] + \kappa_9[K_t] \\
\frac{\mathrm{d}[K_t]}{\mathrm{d}t} &= -\kappa_4[K_t][S] + \kappa_5[K_tS] + \kappa_6[K_tS] + \kappa_8[K_r] - \kappa_9[K_t] \\
\frac{\mathrm{d}[K_rS]}{\mathrm{d}t} &= \kappa_1[K_r][S] - \kappa_2[K_rS] - \kappa_3[K_rS] - \kappa_{10}[K_rS] + \kappa_{11}[K_tS] \\
\frac{\mathrm{d}[K_tS]}{\mathrm{d}t} &= \kappa_4[K_t][S] - \kappa_5[K_tS] - \kappa_6[K_tS] + \kappa_{10}[K_rS] - \kappa_{11}[K_tS] \\
\frac{\mathrm{d}[S]}{\mathrm{d}t} &= -\kappa_1[K_r][S] + \kappa_2[K_rS] - \kappa_4[K_t][S] + \kappa_5[K_tS] + \kappa_7[S_p] \\
\frac{\mathrm{d}[S_p]}{\mathrm{d}t} &= -\kappa_7[S_p] + \kappa_3[K_rS] + \kappa_6[K_tS],
\end{aligned}
\tag{5.1}
$$

And the system need to follow these conservation equations:

$$
\begin{aligned}
[S_{\mathrm{tot}}] &= [S] + [S_p] + [K_rS] + [K_tS] \\
[K_{\mathrm{tot}}] &= [K_r] + [K_t] + [K_rS] + [K_tS]
\end{aligned}
$$

where I introduce two concentration invariants, namely $[K_{\mathrm{tot}}]$ and $[S_{\mathrm{tot}}]$, representing the total concentration of the kinase and the substrate respectively. This equates to the biological assumption that total concentration of these signalling proteins are constant over the relevant time scales of signalling (i.e. the model does not consider dynamics arising from gene regulation and expression).

5.2.2 Analytical solutions

The mathematical analysis on model of minimal bistable signalling motif results in the necessary condition on parameters (i.e. 2 total concentration values and 11 kinetic rate constants) under which the motif exhibits bistable dynamics (see Appendix E):

$$
\alpha_1 K_{\mathrm{tot}} + \alpha_2 < S_{\mathrm{tot}} < \alpha_3 K_{\mathrm{tot}} + \alpha_4, \tag{5.2}
$$

where,

$$\alpha_1 = \frac{\kappa_1\kappa_4[(\kappa_6+\kappa_7)\kappa_{10}+(\kappa_3+\kappa_7)\kappa_{11}]}{\kappa_1\kappa_4\kappa_7(\kappa_{10}+\kappa_{11})},$$
$$\alpha_2 = \frac{(\kappa_2+\kappa_3)\kappa_4\kappa_7(\kappa_8+\kappa_{11})+(\kappa_5+\kappa_6)\kappa_1\kappa_7(\kappa_9+\kappa_{10})+\kappa_7(\kappa_1\kappa_9+\kappa_4\kappa_8)(\kappa_{10}+\kappa_{11})}{\kappa_1\kappa_4\kappa_7(\kappa_{10}+\kappa_{11})}$$
$$\alpha_3 = \frac{[(\kappa_1\kappa_9+\kappa_4\kappa_8)((\kappa_6+\kappa_7)\kappa_{10}+(\kappa_3+\kappa_7)\kappa_{11})(\kappa_2+\kappa_3)\kappa_4\kappa_8(\kappa_6+\kappa_7)+(\kappa_5+\kappa_6)\kappa_1\kappa_9(\kappa_3+\kappa_7)]}{\kappa_7[(\kappa_1\kappa_9+\kappa_4\kappa_8)(\kappa_{10}+\kappa_{11})+(\kappa_2+\kappa_3)\kappa_4(\kappa_8+\kappa_{11})+(\kappa_5+\kappa_6)\kappa_1(\kappa_9+\kappa_{10})]}$$
$$\alpha_4 = \frac{[(\kappa_2+\kappa_3)\kappa_{11}+(\kappa_5+\kappa_6)\kappa_{10}+(\kappa_2+\kappa_3)(\kappa_5+\kappa_6)]\kappa_7(\kappa_8+\kappa_9)}{\kappa_7[(\kappa_1\kappa_9+\kappa_4\kappa_8)(\kappa_{10}+\kappa_{11})+(\kappa_2+\kappa_3)\kappa_4(\kappa_8+\kappa_{11})+(\kappa_5+\kappa_6)\kappa_1(\kappa_9+\kappa_{10})]}.$$

For each fixed value of K_{tot}, the solution to the system of inequalities (5.2) is either empty or an interval. Since $\alpha_i > 0$ for $i = 1, 2, 3, 4$, $\alpha_1 K_{\text{tot}} + \alpha_2$ and $\alpha_3 K_{\text{tot}} + \alpha_4$ are increasing straight lines in K_{tot} with positive intercept. The region is described by a sector intersected with the positive orthant of $\mathbb{R}^2$. If the two lines are parallel, the valid region is the region between the two lines intersected with the positive orthant.

The necessary and sufficient condition under which the system exhibits bistable dynamics with parameters in positive real domain is as following (see Appendix E):

$$(\kappa_3-\kappa_6)(\eta_r\kappa_9\kappa_{10}-\eta_t\kappa_8\kappa_{11}) > ((\kappa_6+\kappa_7)\kappa_{10}+(\kappa_3+\kappa_7)\kappa_{11})(\eta_r\kappa_{10}+\eta_t\kappa_{11}) \tag{5.3}$$

where,

$$\eta_r = \frac{\kappa_1}{\kappa_2+\kappa_3} \qquad \eta_t = \frac{\kappa_4}{\kappa_5+\kappa_6}$$

are the inverses of the Michaelis-Menten constants of the kinases K_r and K_t respectively.

5.2.3 Parameter sampling

The parameter sampling is performed by drawing random number r from uniform distribution in interval $[\ln 10^{-3}, \ln 10^3]$, then scale the random number through $\kappa = e^r$. In this approach, I confine the sampled parameters in biologically relevant

ranges (Table 2.2). For sampling on interdependent parameters in detailed balancing conditions (i.e. $\kappa_1 \cdot \kappa_5 \cdot \kappa_9 \cdot \kappa_{10} = \kappa_2 \cdot \kappa_4 \cdot \kappa_8 \cdot \kappa_{11}$), I first draw a random number γ from gamma distribution with probability density function as $p(x) = x^{\alpha-1} \frac{e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}$ where Γ is a gamma function, $\alpha = 2$ and $\beta = 7$. Then draw two sets uniformly distributed random numbers, each set has four random numbers (e.g. r_1, r_5, r_9, r_{10} and r_2, r_4, r_8, r_{11}) such that $r_1 + r_5 + r_9 + r_{10} = r_2 + r_4 + r_8 + r_{11} = 1$. Then accordingly, the kinetic rate constants can be scaled by $\kappa_i = e^{-r_i \cdot \gamma}$, where $i = 1, 2, 4, 5, 8, 9, 10, 11$.

5.3 Results

5.3.1 The futile signalling cycle with a two-state kinase is a bistable motif

A key interaction motif found in eukaryotic signalling networks is the so-called futile signalling cycle, where a protein substrate is covalently phosphorylated and dephosphorylated by a kinase and phosphatase (Figure 5.1A). In Chapter 4, I discovered a novel bistable subnetwork by simplifying emerged bistable networks in evolutionary simulations. This subnetwork is composed of a futile signalling cycle with a kinase and a phosphatase where the kinase has two distinct conformational states switching between each other. As shown before, the futile signalling cycle when considered with a single phosphorylation site on the substrate and a single-state kinase and phosphatase cannot display bistability for any parameters with positive values [262, 272] (Figure 4.2 and 5.1A) as can be proven by the deficiency one theorem [237, 238, 273, 274], but shows ultrasensitivity under saturating [92]. When I extended this system with a two-state kinase, this key result changed and bistability was possible. I introduced the two-state kinase such that each state can bind the substrate and catalyse its phosphorylation, and where transitions between the two states are possible irrespective of substrate binding (Figure 5.1B). The two-state kinase, as I introduced in this simple model, switches between two conformational states with a

constant rate. The two states show differential catalytic activity towards the substrate (see Figure 5.1B and *Methods*). While this is the simplest model to introduce the idea of multi-state enzymes into the core futile cycle motif, it is readily possible to assume more complex models. In particular, the conformational change between kinase states can be modelled as an allosteric regulation [87, 88, 90, 275, 276], whereby it is linked to binding of the kinases by a ligand or other proteins, or as arising from covalent phosphorylation events as commonly observed in signalling proteins [11, 264, 277]. I consider such models below, but note that they do not alter the key conclusions of this study on bistability and multi-stability.

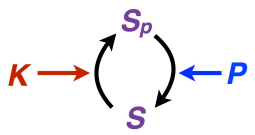
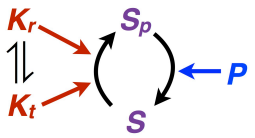
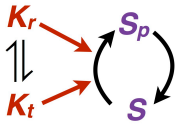
	Structure	Reactions	Allow bistability
A		$K+S \rightleftharpoons KS \rightarrow K+S_p$ $P+S_p \rightleftharpoons PS_p \rightarrow P+S$	NO
B		$K_r+S \rightleftharpoons K_rS \rightarrow K_r+S_p$ $K_t+S \rightleftharpoons K_tS \rightarrow K_t+S_p$ $P+S_p \rightleftharpoons PS_p \rightarrow P+S$ $K_r \rightleftharpoons K_t \quad K_rS \rightleftharpoons K_tS$	Yes
C		$K_r+S \rightleftharpoons K_rS \rightarrow K_r+S_p$ $K_t+S \rightleftharpoons K_tS \rightarrow K_t+S_p$ $S_p \rightarrow S$ $K_r \rightleftharpoons K_t \quad K_rS \rightleftharpoons K_tS$	Yes

Figure 5.1: Different signalling futile cycles, corresponding chemical reactions and their capacity for bistable dynamics.

I find that the core motif with two-state enzymes can be further simplified without compromising bistability by removing the phosphatase and letting the dephosphorylation of the substrate happen through auto-hydrolysis at a constant rate (see Figure 5.1C and *Methods*). In this way, I obtain a minimalist core signalling system driven by a two-state kinase, which displays bistability. The system contains

only six species, making it one of the smallest signalling motifs that are bistable.

5.3.2 Conditions for bistability in the core motif are satisfied in a biologically plausible range

The simplicity of this core motif allowed me to analytically study the solutions to the steady state equations (see Appendix E). In particular, I was able to derive a set of inequalities in the kinetic parameters and total concentrations of the substrate and kinase that provide a set of *necessary and sufficient conditions* for the existence of three steady states in the system (Equation 5.3, see also Appendix E for the derivation of this equation). From these conditions I derive the following necessary condition for bistability (the indexing of the rate constants is given in Equation 5.1 in *Methods*:

$$(\kappa_3 - \kappa_6)\left(\eta_r \kappa_9 \kappa_{10} - \eta_t \kappa_8 \kappa_{11}\right) > 0 \tag{5.4}$$

where,

$$\eta_r = \frac{\kappa_1}{\kappa_2 + \kappa_3} \qquad \eta_t = \frac{\kappa_4}{\kappa_5 + \kappa_6}$$

are the inverses of the Michaelis constants of the kinases K_r (the kinase at the relaxed state) and K_t (the kinase at the tense state) respectively. Analysis of this equation reveals key features of the system that are necessary for bistability. I find that the switching reactions between the two states of the kinase, as well as between the kinase-substrate complexes are crucial for bistability. That is, both κ_8 and κ_9 cannot be zero, and both κ_{10} and κ_{11} cannot be zero. Thus, the structure of the reaction system composing of a futile signalling cycle driven by a two-state kinase is crucial for enabling bistability.

Equation 5.4 provides two key features for bistability. Firstly, the two interconnected futile cycles between S and S_p, defined by the two kinase states, need to operate at different catalytic rates (i.e. $\kappa_3 \neq \kappa_6$). Secondly, the switching between these cycles through the four forms of the kinase (i.e. K_r, K_t, K_rS,

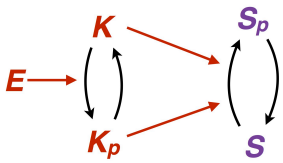
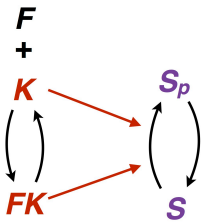
Structure	Reactions
A 	$K + S \rightleftharpoons KS \rightarrow K + S_p$ $K_p + S \rightleftharpoons K_pS \rightarrow K_p + S_p$ $E + K \rightleftharpoons EK \rightarrow E + K_p$ $E + KS \rightleftharpoons EKS \rightarrow E + K_pS$ $S_p \rightarrow S \quad K_p \rightarrow K \quad K_pS \rightarrow KS$
B 	$K + S \rightleftharpoons KS \rightarrow K + S_p$ $FK + S \rightleftharpoons FKS \rightarrow FK + S_p$ $F + K \rightarrow FK$ $FKS \rightarrow KS + F$ $S_p \rightarrow S$

Figure 5.2: Expanded signalling networks without detailed balance. (A) Extended network obtained by adding an enzyme catalysing one of the transitions between the two states of the kinase K in the core motif. (B) Extended network obtained by adding a protein to the core motif such that steric effects from the binding of the added protein with the enzyme makes the transitions between different states of the kinase K irreversible. Both extensions maintain the capacity for bistability.

K_tS) needs to occur at different rates, and in a way opposing the difference in the catalytic rates. Specifically, if the futile cycle for the relaxed state of the kinase (i.e. K_r and K_rS) has the highest catalytic activity (i.e. $\kappa_3 > \kappa_6$), then $\eta_r\kappa_9\kappa_{10}$ needs to be larger than $\eta_t\kappa_8\kappa_{11}$. As a consequence, the clockwise interchanging cycle, $K_r \to K_rS \to K_tS \to K_t \to K_r$, corresponding to the product of the rate constants $\kappa_1\kappa_{10}(\kappa_5+\kappa_6)\kappa_9$, needs to dominate over the anti-clockwise cycle, $K_r \to K_t \to K_tS \to K_rS \to K_r$, corresponding to the product $\kappa_4\kappa_{11}(\kappa_2 + \kappa_3)\kappa_8$. Symmetrically, if K_t has higher catalytic activity than K_r (i.e. $\kappa_3 < \kappa_6$) then the anti-clockwise cycle needs to dominate.

A further constraint on the rates governing the transitions among the four forms of the kinase might arise from thermodynamics. Particularly, these transitions form a local state cycle, which must follow the principle of detailed balance if we assume no additional energy input into the system [278–281]. This results in a thermodynamic constraint on the reaction kinetics such that the product of the rate constants in the clockwise direction must equal the product of the reverse rate constants (i.e. $\kappa_1\kappa_5\kappa_9\kappa_{10} = \kappa_2\kappa_4\kappa_8\kappa_{11}$). It must also be noted, however, that this constraint would be relaxed if the conformational switching between the enzyme states were directed by energy input (e.g. phosphorylation-dephosphorylation reactions, (Figure 5.2A) or steric effects with enzyme binding with other proteins or enzymes (Figure 5.2B).

Table 5.1: Number of bistable parameter sets found by sampling parameters of the core motif. Sampling is performed under two conditions, relaxed form and under the thermodynamic constraint. The total number of sampled parameter sets is 10^5.

With thermodynamic constraint	Without thermodynamic constraint
2787($\sim$ 2.8%)	14492($\sim$ 14%)

To determine whether these conditions on kinetic rates can be simultaneously satisfied in cellular signalling networks, I tabulated kinetic parameters from the literature (see Table 2.2 and references therein). I then sampled 105 parameter

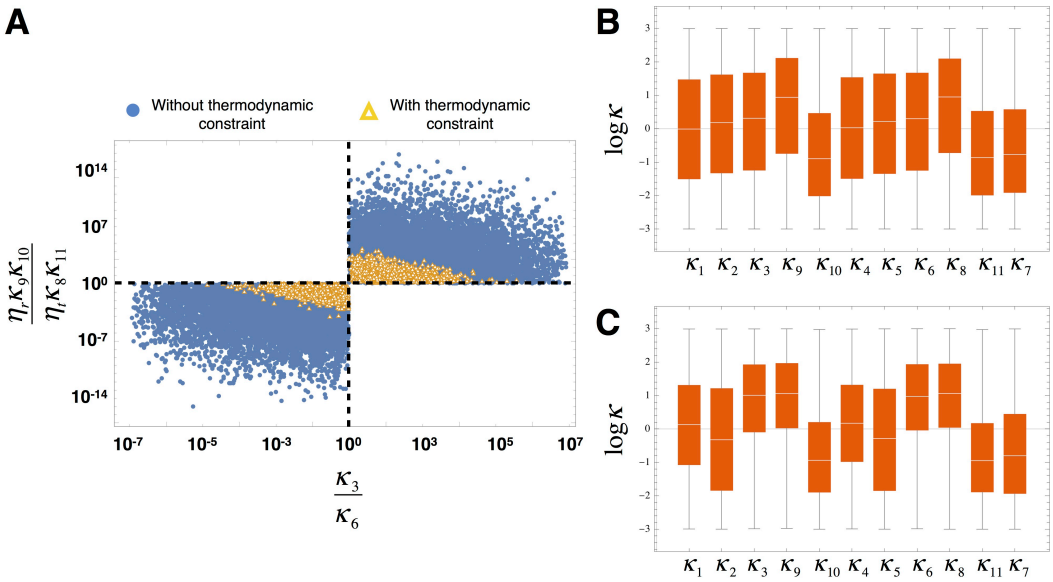


Figure 5.3: Parameter sets that allow for bistability, sampled in a biologically feasible range. (A) Sampled parameter sets plotted in the space of $\frac{\kappa_3}{\kappa_6}$ *vs.* $\frac{\eta_r \kappa_9 \kappa_{10}}{\eta_t \kappa_8 \kappa_{11}}$. The blue dots (resp. yellow triangles) correspond to the parameter sets sampled without (resp. with) the thermodynamic constraint. In accordance with the sufficient and necessary condition (see *Methods*), all sampled parameters that allow for bistability fall into the two regions that meet at $(1, 1)$. (B, C) Boxplots of the rate constants sampled without (B) and with (C) the thermodynamic constraint, shown on log10-scale. The conditioning on bistability changes the distribution of the rate constants. In the inequality for bistability (Equation 5.3 in the main text) the groups of rate constants κ_1, κ_2, κ_3, κ_9, κ_{10} and κ_4, κ_5, κ_6, κ_8, κ_{11} appear symmetrically in the inequality in the sense that if the two groups of parameters are swapped, the inequality is fulfilled if and only if it was so before swapping. Hence the rate constants κ_1 and κ_4 follow the same distribution, κ_2 and κ_5 do as well, and so on. This symmetry is visible in the boxplots. The range of each parameter generally shrinks under the thermodynamic constraint compared to without the constraint.

sets around these known kinetic parameters and checked whether the necessary and sufficient conditions for bistability were satisfied (see *Methods*). This analysis showed that the futile signalling cycle displays bistability in a biologically plausible parameter regime, even when thermodynamic constraints are taken into account (Figure 5.3 and Table 5.1).

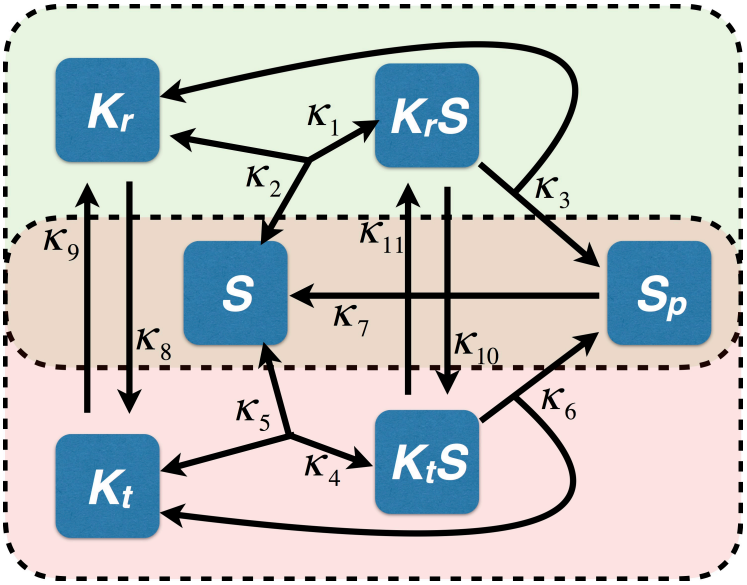


Figure 5.4: Schematic of minimal signalling motif displaying bistability. Cartoon representation of the two interconnected reaction cycles constituting the core bistable system. The arrows represent reactions in the system and are labelled with the kinetic parameters from Equation 5.1. Two rectangles (dashed line) with different background colour show the two futile cycles with K_r (green) and K_t (red) competing for the substrate (in the intersected region of the two rectangles).

5.3.3 Bistability can be seen as arising from competition between the kinase states for the substrate

It is interesting to note that the mathematical conditions derived in Equation 5.4 impose a specific structure onto the core motif, which can be seen as two connected reaction cycles that are driven by the two states of the kinase competing for the same substrate (Figure 5.4). Equation 5.4 shows that the flows of these two competing reaction cycles need to have a specific relationship for bistability to emerge. To

better understand these ensuing reaction fluxes, I have analysed the steady states of the system for increasing total kinase concentration, as a proxy for an increasing signal (Figure 5.5, see also Appendix E). For a fixed set of parameters in the bistable regime such that $\kappa_3 > \kappa_6$, $\kappa_9 > \kappa_8$, $\kappa_{10} > \kappa_{11}$, and $\eta_r\kappa_9\kappa_{10} > \eta_t\kappa_8\kappa_{11}$ (see Table 5.2), I find that in the low signal regime, where the total level of kinase is low, there is a large flux from K_rS into K_tS, resulting in the accumulation of K_tS. Thus in this low signal regime, the slow futile cycle driven by K_t (which has the lower catalytic activity) dominates (i.e. $[K_r] + [K_rS] < [K_t] + [K_tS]$) and the system is at low state (i.e. small $[S_p]$) (Figure 5.5, red dots). In the high signal regime, the fast futile cycle driven by K_r dominates (i.e. $[K_r] + [K_rS] > [K_t] + [K_tS]$) and the system is at the high state (i.e. large $[S_p]$). The substrate is largely converted to the phosphorylated form, which results into the accumulation of K_r (Figure 5.5, green dots). Whether the K_r mediated or K_t mediated cycle dominates is primarily determined by the condition $\eta_r\kappa_9\kappa_{10} > \eta_t\kappa_8\kappa_{11}$, which relates to the inverse of Michealis-Menten constants associated with each kinase forms and the transition rates between these forms in a free and substrate-bound state.

Table 5.2: Example parameter sets that enable bistable dynamics in the core signalling motif. The table shows the parameter sets used for the generation of the bifurcation plot in Figure 5.5.

Parameter	Unit	Value	Reaction
κ_1	$\mu M^{-1}s^{-1}$	86.78	$K_r + S \rightarrow K_rS$
κ_2	s^{-1}	3.583	$K_rS \rightarrow K_r + S$
κ_3	s^{-1}	92.84	$K_rS \rightarrow K_r + S_p$
κ_4	$\mu M^{-1}s^{-1}$	1.200	$K_t + S \rightarrow K_tS$
κ_5	s^{-1}	0.02626	$K_tS \rightarrow K_t + S$
κ_6	s^{-1}	0.2644	$K_tS \rightarrow K_t + S_p$
κ_7	s^{-1}	2.357	$S \rightarrow S_p$
κ_8	s^{-1}	0.01310	$K_r \rightarrow K_t$
κ_9	s^{-1}	0.7842	$K_t \rightarrow K_r$
κ_{10}	s^{-1}	1.041	$K_rS \rightarrow K_tS$
κ_{11}	s^{-1}	0.008057	$K_tS \rightarrow K_rS$
$[S_{\text{tot}}]$	μM	9.994	—
$[K_{\text{tot}}]$	μM	$0 \sim 3$	—

This analysis derived from the necessary parameter conditions leads to an intuitive view, in which the bistability in the system is understood as a result of the two futile cycles driven by the two forms of the kinase competing for the substrate. Furthermore, the competing kinase forms need to have opposite dominance in terms of being able to bind the substrate and their catalytic activity, such that the form dominating catalytically ($\kappa_3 > \kappa_6$) needs to be weaker in terms of substrate binding kinetics (i.e. assuming $\kappa_9 = \kappa_{10} = \kappa_8 = \kappa_{11}$, we need to have $\eta_r > \eta_t$).

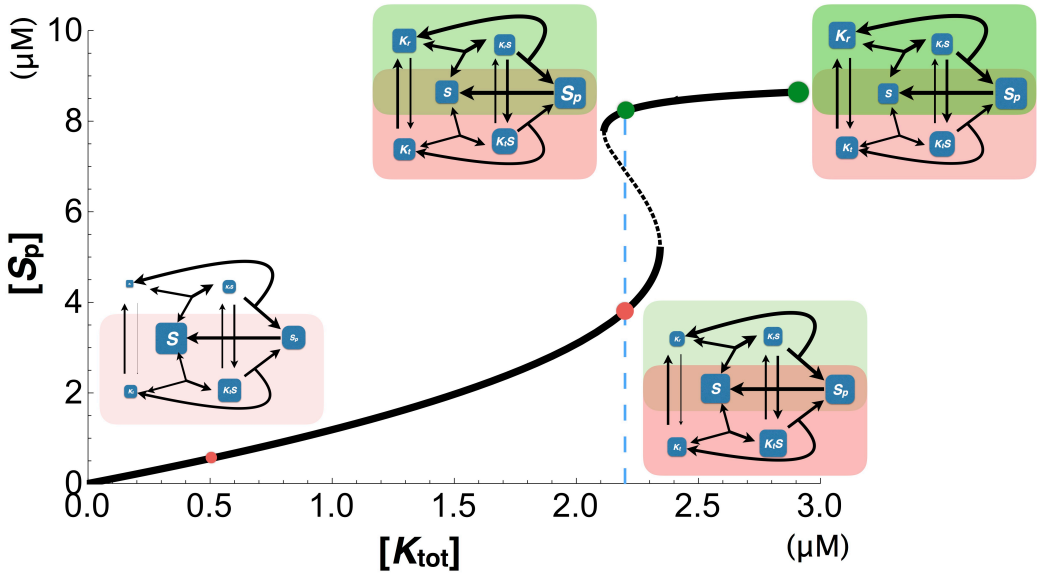


Figure 5.5: Bifurcation plot of core bistable signalling motif. The solid line corresponds to the stable steady state levels of $[S_p]$ with increasing signal given by the total concentration of kinase $[K_{\text{tot}}]$. The dashed line corresponds to the unstable steady states. The parameter values used to generate the bifurcation plot are listed in Table 5.2. The four little cartoons, drawn as inset, are showing the allocation of all species concentration and corresponding reaction fluxes at the different levels of $[K_{\text{tot}}]$, as indicated by the coloured dots. Within each cartoon, the size of each blue box stands for the relative amount of species (logarithmically scaled), while the thickness of the arrows stands for the relative levels of the reaction fluxes (logarithmically scaled) calculated with mass-action kinetics, namely $\kappa_1[K_r][S]$, $\kappa_2[K_rS]$, $\kappa_3[K_rS]$, $\kappa_4[K_t][S]$, $\kappa_5[K_tS]$, $\kappa_6[K_tS]$, $\kappa_7[S_p]$, $\kappa_8[K_r]$, $\kappa_9[K_t]$, $\kappa_{10}[K_rS]$, $\kappa_{11}[K_tS]$.

5.3.4 Increasing the number of kinase states in the signalling cycle leads to unbounded multistationarity

Recognising that bistability in the core motif is linked to the competition between the two futile cycles, it is intriguing to consider whether adding more competing cycles increases the number of steady states. To expand from the simplest motif towards more complicated systems, one way of increasing competing cycles is to increase the number of two-state kinases, while the other is to increase the number of states of a single kinase. I find that both expansions of the minimal system result in an increase of the number of steady states.

Firstly, I considered the case of multiple kinases with two states (Figure 5.6A). In this case, multiple two-state kinases in a futile cycle lead to multistationarity (Figure 5.6A). With the number of kinases n increasing, the number of steady states linearly scales with n. We prove that the system can admit at most $2n + 1$ steady states and further that n of them are unstable (see Appendix E). The other $n + 1$ steady states are presumably stable. Secondly, multistability can be achieved by one kinase with multiple states (Figure 5.6B). When the kinase has 3 distinct states, the system can have 3 steady states at most, but a four-state kinase results in the possibility of 5 steady states at most (Figure 5.6B, see Appendix E). The general scenario with an n-state kinase is too complex mathematically, and does not admit the approach used to analyse systems with multiple two-state kinases. However, we conjecture that the number of positive steady states grows linearly with n as well, such that the system admits at most $n + 1$ positive steady states if n is even and n positive steady states if n is odd.

5.3.5 Multistability enables complex state transitions

The above results confirm that a single futile signalling cycle with a two-state kinase can generate bistable dynamics and that such a system can be expanded by increasing the number of kinase states to achieve unlimited multistationarity. In this

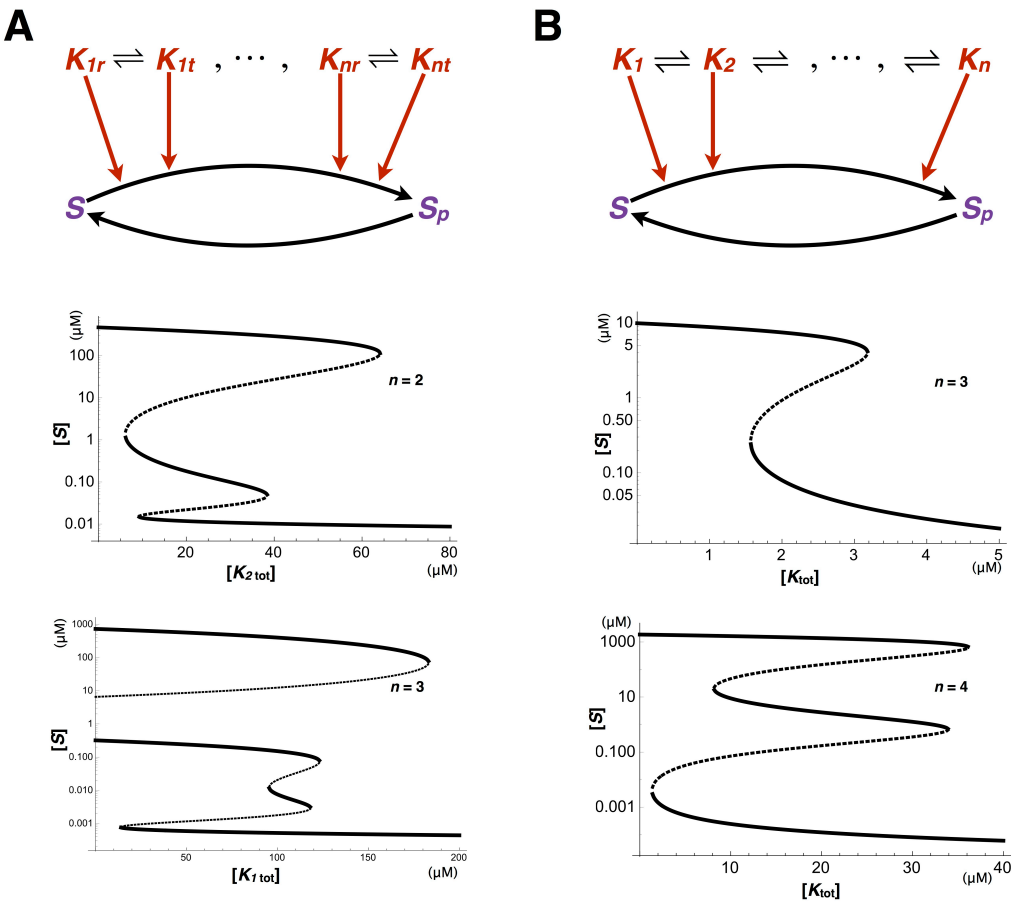


Figure 5.6: Implementation of multistability by expanding the minimal bistable motif. (A) Multistability generated from signalling cycle with multiple two-state kinases. Top-left: A schematic of multiple kinases. Middle-left: Bifurcation plot for a system with two allosteric kinases, the x-axis shows the signal level $[K_{2\text{tot}}]$ (total concentration of the second kinase K_2), the y-axix shows the level of $[S]$ (unphosphorylated substrate S). Bottom-left: Bifurcation plot of a system with three kinases. (B) Multistability generated from signalling cycle with multi-state kinase. Top-right: schematic of a multi-state kinase catalysing a futile signalling cycle. Middle-right: Bifurcation plot of system with a three-state kinase. Bottom right: Bifurcation plot of a system with four-state kinase. The x- and y-axis are as above. In all bifurcation plots, solid lines correspond to stable steady states, while dashed lines correspond to unstable steady states. All axes use the unit of concentration μM.

scenario, each additional kinase state drives potentially the generation of a pair of steady states, one stable and one unstable, due to the competition for the substrate. Thus, it should be possible to use the total concentrations (or kinetic parameters) of the different kinases to change the signal thresholds to switch between steady states and implement logic gates in this way. More specifically, in the system with multiple two-state kinases, varying the total concentration of a kinase can dictate the system transitions among the different steady states resulting from multistability.

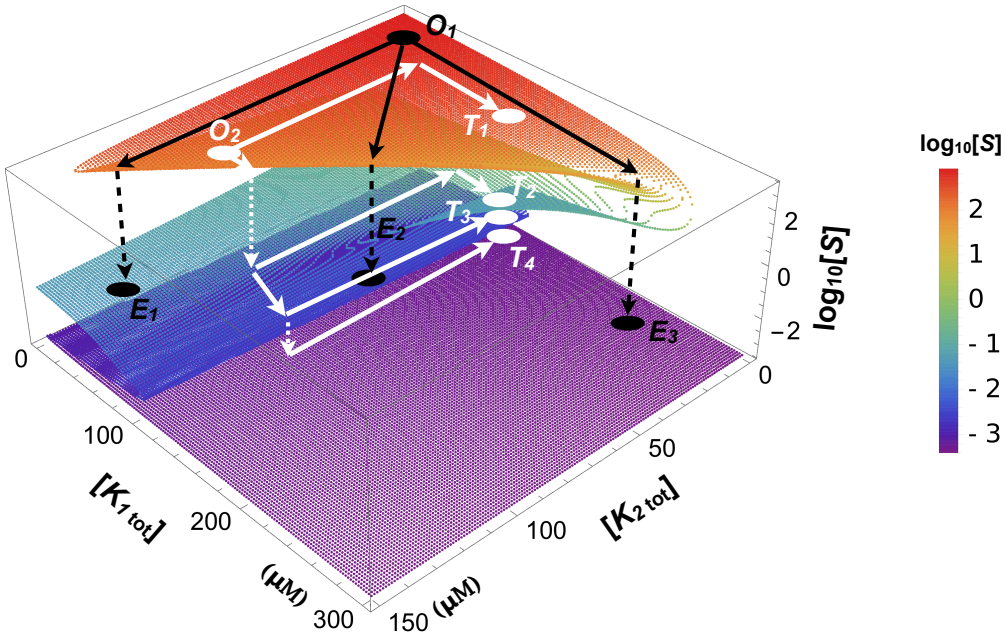


Figure 5.7: Multistability installs complex state transitions. The steady state level of the unphosphorylated substrate, $[S]$, for different levels of the two kinases, $[K_{1,\text{tot}}]$ and $[K_{2,\text{tot}}]$. The colour-coding shows the level of unphosphorylated substrate, $[S]$, for each amount of kinase. The black and white dots represent specific states of the system. The black and white arrows show the hypothetical trajectories described when the kinase levels are perturbed in various combinatorial ways, as discussed in the main text.

Here, I show that by combinatorial perturbations of different kinases, a system with three two-state kinases can perform complex state transitions (Figure 5.7). The varying parameters are the total concentrations of the first two kinases, namely $[K_{1\text{tot}}]$ and $[K_{2\text{tot}}]$. I assume that the system starts off at a given state (O_1 in

Figure 5.7) with low total concentration of both kinases. By increasing the total concentration of either kinase (K_1, K_2) or both, the system can be made to switch to three different end-states of $[S]$ (Figure 5.7, points E_3, E_1 or E_2). It is also possible to bring the system into different states by perturbing the total concentrations of both kinases by a fixed amount each, but following different sequential moves (Figure 5.7, from O_2 to T_1, T_2, T_3 and T_4). In these examples, the final system output is a function of the combinatorial activity patterns of both kinases. In contrast, different perturbations would result in the same output state in a monostable system. Therefore, multistability can encode the specific order of changes in the environmental signals (i.e. different kinase activities) into different system outputs at steady state. The result is a potential increase in the systems capacity to store information, e.g. relating to fluctuating or complex environments.

5.3.6 Real biological systems display complex interactions leading to multi-state enzymes and potential for multistability

As discussed in the introduction, futile signalling cycles are ubiquitous motifs in natural signalling networks, where they feature multi-state enzymes. To demonstrate this point, I explore two example cases of natural signalling cycles. The first example comprises the signalling networks controlling the cell cycle, in particular networks involving cyclin-dependent kinases (Cdks). It is argued that the activity of Cdks is a key mechanism for ensuring appropriate switching dynamics for the cell cycle [265–267]. The activity level of Cdk1 is regulated by four different mechanisms: (1) activating phosphorylation by Cdk-activating kinases (CAKs), where phosphorylation by a CAK of a threonine residue increases the kinase activity of Cdk1 [282]; (2) inhibitory phosphorylation by Wee1, where phosphorylation of a tyrosine residue by Wee1 reduces kinase activity of Cdk1 [283]; (3) cyclin binding, where cyclins binding cooperatively to Cdk1 and their substrates promote Cdk1 kinase activity [284]; and (4) Cdk-inhibitor (CKI) binding, where CKIs bind to Cdk1

and block their active sites [283] (Figure 5.8A). Such combinatorial interactions (i.e. regulations) thus correspond to different Cdk1s "states" (i.e. phosphorylated at different positions, bound/unbound, *etc*) that can display different activity levels and that compete for the same downstream substrates. Moreover, several homologous Cdks are shown to compete for the same substrates [268, 283], Similarly, Wee1 has differentially phosphorylated forms that have different activity towards Cdk1 [285, 286], and ubiquitination of Wee1 leading to its degradation also affects the phosphorylation of Cdk1 by Wee1 [287]. The second example I focus on comprises the MAPK signalling cascades [106]. For instance the MAPK signalling networks controlling yeast mating response and filamentous growth response share the signalling proteins Ste11 and Ste7, both of which have two phosphorylation sites and can bind to a scaffolding protein Ste5 (Figure 5.8B) [216]. The possible combinatorial interactions and the different phospho-states of these proteins, as well as their downstream interaction partners such as Fus3 and Kss1 provide a system with multiple kinase states

The picture emerging from the Cdk as well as the MAPK pathways is one with multiple steady states and several enzymes in competition for the same substrates. This picture fits in the simplified motifs as analysed above (and shown in Figure 5.6 and 5.7), making it theoretically possible for these pathways to display bistability and multistability. Towards experimental verification of such possibility, it would be a good starting point to measure *in vitro* the catalytic and binding rates of different enzyme forms found in these systems.

5.4 Discussion

The key finding of this study is that the presence of a multi-state kinase in the common futile signalling cycle motif allows this functional interaction system to display bistability. Thus, a phosphorylable substrate with a two-state kinase forms one of

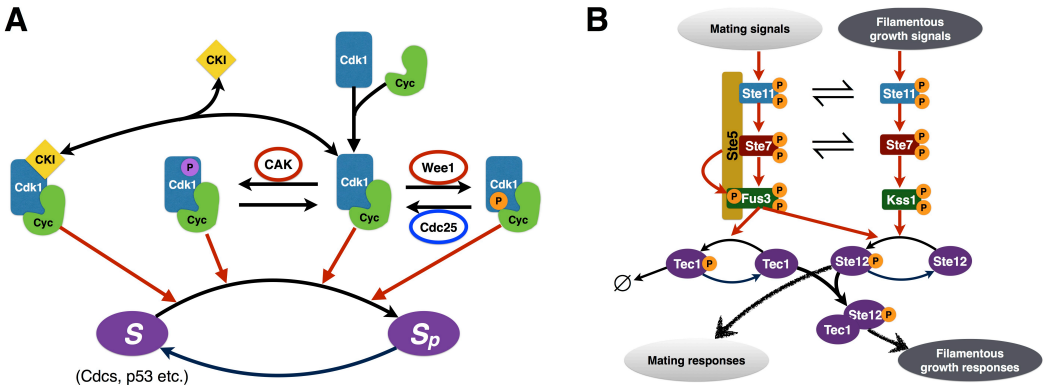


Figure 5.8: The bistable signalling motif in biological systems. (A) Different forms of regulation of Cdk1s catalytic activity give rise to different states of Cdk1. The multiple states of Cdk1 are involved in catalysing many downstream substrates, including Cdc and p53. Such catalytic reactions show precisely the structural pattern in Figure 5.6B. (B) The two MAPK cascades in yeast mating response and filamentous growth response. The two cascades share Ste11 and Ste7. Ste11, Ste7 and Kss1 have two phosphorylation sites while Fus3 has three phosphorylation sites, one of which is phosphorylated by Ste5. This schematic shows that in both cascades all three layers of signalling enzymes, MAP3K (i.e. Ste11), MAP2K (i.e. Ste7) and MAPKs (i.e. Fus3 and Kss1) exhibit different states that compete for their substrates. Thus, the cross-talk between the two cascades and the presence of the scaffold protein increases the number of states of the enzymes, resulting in a system similar to that considered in Figure 5.6A.

the smallest bistable signalling motifs. The emergence of bistability in this simple system relates closely to the two states of the kinase forming two futile cycles that are competing for the substrate. I define conditions on the kinetic parameters of these two competing cycles that are necessary and sufficient for three steady states. I show that these conditions are met under biologically feasible parameter regimes. Finally, I find that increasing either the number of two-state kinases acting on the same substrate or the number of distinct states that a single kinase can exhibit increases the number of steady states in an unbounded manner. Particularly, the unlimited multistability with by increasted enzyme states suggest that this theoretical analysis can potentially help us design cellular signalling systems with various biochemical memories for recording environmental information. The implementation of multistability experimentally requires tuning the kinetic parameters which are potentially in very narrow spaces. The possible ways to overcome such constraint are either using evolutionary experiments and direct evolution method to optimise the kinetic parameters or designing larger networks to relax the constraints imposed on it. For instance, it is possible to implement system with 5 steady states with 3 or more two-state enzymes and embed such motif into a larger signalling network to expand the kinetic parameter space.

The core bistable signalling motif featuring multi-state enzymes is prevalent in biological systems. Presence of multiple conformational states with differential activity is a common feature of many enzymes [88], and particularly in signalling networks, where many kinases and phosphatases display multiple states that display different levels of activity and that are regulated through covalent modification or interaction with scaffold proteins [224, 288]. As I have shown above, using Cdks and MAPK pathways as examples, there are several natural cases where such interactions create or embed the described core bistable motifs or extensions of it. Our findings thus provide mathematical proof that these natural systems can theoretically allow bistability and potentially unbounded multistability. Transitions between the steady

states can underpin the capacity of cells to map environmental states to internal gene expression and physiology, increasing their ability to adapt to different or fluctuating environments. The validation and further interrogation of these possibilities must come from experimental studies. In particular, synthetic biology approaches can be used to implement the core bistable motif described here using existing multistate proteins and kinases from nature and analysing their dynamics in a controlled manner. These approaches are already being employed to study MAPK and two-component signalling systems [231, 288–290], and can be further extended using the presented results as guiding principles for experiments.

An intuitive interpretation of our results is that competition of different futile cycles for the same substrate is a key prerequisite for bistability. This intuitive view can also be applied to understand previously described bistable and multistable signalling motifs. For instance, a substrate with multiple phosphorylation sites that are acted upon by the same kinase is shown to implement bistability and multistability [52–54, 106, 262]. This system is almost a symmetric version of the system I consider here, as it features futile cycles involving differently phosphorylated substrates competing for the same enzyme. Another example of a bistable system is where a futile cycle can take place in two different compartments, with both substrates and enzymes shuttling between the two compartments. This again fits our intuitive view, where the separation of enzymes and substrates in different compartments creates a set of futile cycles that are competing for both substrates and enzymes [51].

These examples indicate that competing futile cycles could provide a general condition for determining bistability. In order to validate this idea, further exploration of different motifs and the structural conditions on multistationarity is required. One possible approach would be to enumerate a large set of small signalling networks and compare structural differences between monostationary and multistationary networks. Specific structural patterns might emerge, which can be

validated by further mathematical analyses. These mathematically derived conditions can then be utilised to better understand natural signalling systems and design bistable signalling networks and biochemical memory through synthetic biology. Motivated by this hypothesis, in Chapter 6 I constructed an algorithmic approach to enumerate chemical reaction networks of given size. By comparing structural differences of monostationary and multistationary networks, I expect to approach such structural determination of multistationarity in these signalling networks.

Chapter 6

Design Principles of Multistability in Signalling Networks

6.1 Introduction

From Chapter 4 and 5, I derived several network motifs that give rise to bistable dynamics. One of these motifs can be expanded by introducing more multi-state enzymes such that the system has capacity for unbounded multistability. While the analysis of the found individual motifs provided us with important insights into molecular basis of bistability, these were not enough to develop mathematically sufficient conditions for bistability that can differentiate between mono- and multistable systems. Achieving such mathematical conditions that are solely based on network topology would provide highly valuable, as they would allow us to discern biological networks capable of multistability from information on protein-protein interactions without the need to measure kinetic rates.

Previously, mathematical analysis on the structures of general chemical reaction networks produced fruitful results, particularly several important theorems

related to several structural properties of reaction networks as included in chemical reaction network theory [237, 238] which concentrates on reaction networks with general mass action kinetics [291, 292]. Mathematical analysis of general networks led to positive feedback loop being one of the topological requirements for multistationarity [105]. Further study by linking the monostationarity of reaction network with its structural properties revealed that two intersecting positive feedback loops are the necessary condition such that the network has capacity for multistationarity [239, 240]. Together, these theorems composed the compass for design of reaction networks with multiple steady states. However, these conditions are only necessary ones, I can only reject networks that can not give rise to multistationarity based on them, but not directly construct networks guaranteed with capacity for bistable dynamics. With sufficient conditions for multistationarity, one can design bistable networks instantly with confidence by following the conditions. Such sufficient conditions are exactly design principles required to both understand and design multistable networks.

Based on the bistable motifs derived from previous chapters, the common structural features of those bistable networks are two reaction cycles competing at the substrate. I hypothesised that interconnecting loops in a graph and competition among such loops could be a sufficient condition for multistabilty. Is such competition a sufficient condition for multistationary signalling network? If not, what other conditions are there discriminating multistationary reaction networks from monostationary ones? Motivated by the hypothesis and these questions, I constructed an algorithmic approach to enumerate reaction networks with limited sizes, towards identifying certain topological features that can distinguish between mono- and multi-stable networks, or even act as sufficiency conditions.

6.2 Methods

6.2.1 Reaction networks

A general form of reaction network is composed of a set of species $\{X_1, \ldots, X_n\}$ with concentrations $\{x_1, \ldots, x_n\}$ respectively and a set of reactions:

$$r_j : \sum_{i=1}^{n} \alpha_{ij} X_i \to \sum_{i=1}^{n} \beta_{ij} X_i, \quad j = 1, \ldots, m \tag{6.1}$$

where α_{ij}, β_{ij} are stoichiometric coefficients with nonnegative integer values. The stoichiometric matrix of the network can be defined as $A = (a_{ij})$, where $a_{ij} = \beta_{ij} - \alpha_{ij}$. The rates vector of reactions in the network is $v = (v_1, \ldots, v_m)$ with

$$v_j(x) = \kappa_j x_1^{\alpha_{1j}} \cdot \ldots \cdot x_n^{\alpha_{nj}}, \ x \in \Omega_v. \tag{6.2}$$

where $\mathbb{R}^n_{>0} \subseteq \Omega_v \subseteq \mathbb{R}^n_{\geq 0}$. The general form of different equation describing such reaction network is:

$$\dot{x} = Av(x), \ x \in \Omega_v. \tag{6.3}$$

Since the structure properties of the networks is only determined by the stoichiometric matrix A, I enumerate the reaction networks by constructing and dealing with the stoichiometric matrices of all possible networks.

6.2.2 DSR graph

The DSR graph is defined as a labelled bipartite directed graph with node set $\{X_1, \ldots, X_n, r_1, \ldots, r_m\}$ such that:

- There is an edge from X_i to r_j with label z_{ij} if $z_{ij} \neq 0$.
- There is an edge from X_i to r_j with label z_{ij} if $z_{ij} \neq 0$.

where z_{ij} is defined as:

$$z_{ij} = \begin{cases} 1 & \text{if } v_j(x) \text{ increases } x_i, \\ -1 & \text{if } v_j(x) \text{ decreases } x_i, \\ 0 & \text{if } v_j(x) \text{ is constant in } x_i. \end{cases}$$

In the DSR graph, a circuit in a graph $\mathscr{G}$ is a sequence of distinct nodes $i_1, \ldots, i_q$ such that there is a directed edge from i_k to i_{k+1} for all $kq1$ and one from i_q to i_1. A circuit with positive label is a positive feedback loop.

6.2.3 Enumeration of small reaction networks

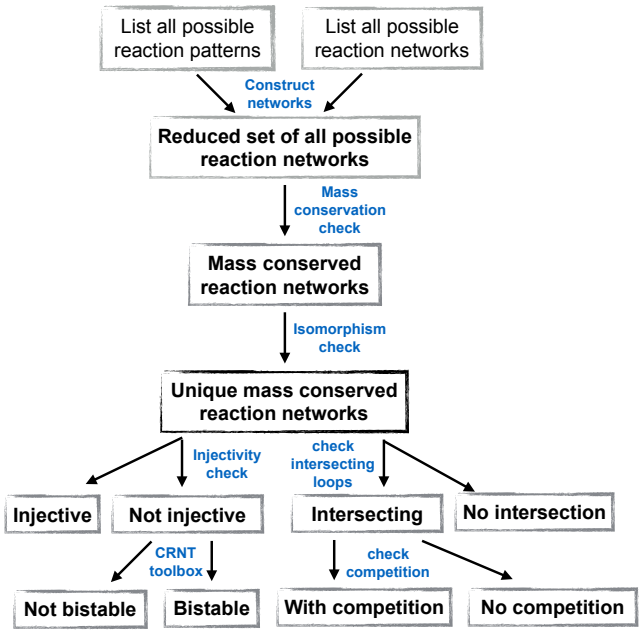


Figure 6.1: The schematic chart illustrating enumeration procedures. In the box are constructed or categorised sets of reaction networks.

Firstly the enumeration process constructs all possible reaction networks with given size. I have then developed algorithmic approaches to select from this full set of networks those that are biologically plausible; this involved eliminating isomorphic networks, as well as networks that do not fit with mass conservation (Figure 6.1).

In particular, elimination of isomorphic networks is implemented by comparing the full of permutations of newly constructed reaction network with each member in the set of unique reaction networks; if no member in the set of unique networks is the same as any permutation of the new network, then add the new network into the set of unique networks; if there are any member in the set of unique networks is the same as any permutation of the new network, then reject the new network as redundant. The process continues to comparing all constructed networks. This ensures all enumerated networks are unique reaction networks.

With the set of biologically plausible networks, I use the injectivity theory to determine whether a network is monostationary [293], then check whether there are two positive loops in its DSR graph that intersects each other [294]. Competition between loops is checked by searching the loop where two species both interact with the another species. These steps classify the biologically plausible networks into 8 categories. Comparison between monostationary and multistationary networks can be performed further from this point.

6.3 Results

6.3.1 Classifying reaction networks with 5 species and 5 reactions

Mathematical proofs provided insights of topological requirements [104, 105, 295][237, 238][239, 240, 293, 296, 297] for multistationarity of chemical reaction networks. These results defined structural necessary condition for bistable dynamics in chemical reaction networks. Although necessary conditions are powerful in preclusion of monostationary networks, it is also appealing to find any general sufficient structural conditions for multistationarity in signalling networks. Such sufficient conditions then can be used as design principles for bistable networks. Nevertheless, the conditions for stability in chemical reaction networks are potentially more subtle and odd. Here, I devised an algorithmic approach to enumerate all possible reaction networks

with fixed number of reactions and species (i.e. 5 species with 5 reactions, 6 species with 6 reactions). The logic behind this approach is that enumerating all possible reaction networks in small scale and characterising their stability help us find the sufficient condition for multistationarity in small chemical reaction networks. Then comparing structural differences between monostationary and multistationary networks can potentially provide certain clues of the topological conditions for multistationarity.

Based on previous necessary conditions, I search the potential necessary condition(s) for multistability in cell signalling network, a special case of chemical reaction networks. By comparing the multistable networks and monostable networks I observed, multiple futile cycles with proteins transiting among multiple states are commonly found in multistable networks rather than monostable ones. The gaps between this necessary condition and potential sufficient condition(s) are subtle and somewhat odd. To reach the possible sufficient condition(s), I enumerated all possible reaction networks with relative low dimensions .

I am mostly interested in biochemical reactions, especially signalling networks with protein interactions, therefore I mainly study reaction networks composed of enzymatic catalysis (e.g. $E+S \rightleftharpoons C \rightarrow E+P$), conformational change (e.g. $S \rightarrow P$) and binding/disassociation (e.g. $A + B \rightarrow C$, $D \rightarrow E + F$). In order to reduce the complexity of enumeration, I excluded the reaction networks with homo-dimerisation and corresponding disassociation reactions (i.e. $2M \rightarrow N$, $P \rightarrow 2Q$).

In the first attempt of enumeration, all possible reaction networks with 5 species and 5 reactions are constructed and examined with several checks. Detail algorithms are described in *Methods*. With 5 species and 5 reactions, I have in total 80 reaction patterns with allowed reactions among 5 species. And the number of all combinatorial reaction networks with 5 reactions is $\binom{80}{5} = 24,040,016$, after partially exclude the isomorphic reaction networks, I reduced the number of reaction networks for enumeration down to $43 \cdot \binom{80}{3} = 3,532,880$, which is computationally

feasible. Then all reaction networks are constructed and only 8933 of these networks are inline with mass conservation required from biologically plausible networks. In those 8933 reaction networks, I further excluded isomorphic networks. This gives a set of 6171 unique mass conserved reaction networks with 5 species and 5 reactions. Among those 6171 unique mass conserved reaction networks, only 68 of them allow multi-stationary dynamics with some positive rate constants and species concentrations. The rest of reaction networks cannot have multiple equilibria, regardless of rate constants and species concentrations (Figure 6.2).

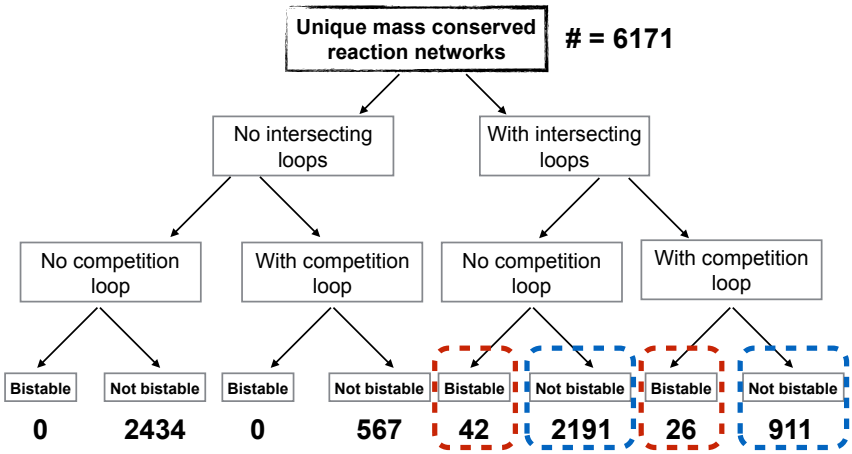


Figure 6.2: The schematic chart illustrating enumeration results. In the box are different categorised sets of reaction networks through multistationarity check and positive feedback loop checks. Detail of algorithmic checks can be found in *Methods*.

6.3.2 Comparison between different categories

From the enumeration and classification of reaction networks of give size, the category of networks with intersecting loops and also competition loop has 26 bistable networks and 911 monostable networks. This clearly shows that the condition in our hypothesis that competing research cycles is not sufficient enough to guarantee the bistability in reaction networks (Figure 6.2). The results also support that networks without intersecting loops are all monostationary (Figure 6.2).

6.4 Discussion

Understanding the design principles of bistable dynamics can be of great interest and benefit. Not only such design principles can help us understand natural biological systems, but also allow us to design and build novel bistable biological systems. Particularly the two bistable motifs derived from evolved bistable networks provide specific design pattern for bistable dynamics in signalling networks (Chapter 5). Following the interest on multistationarity in signalling networks, algorithmic approach of enumerating reaction networks is constructed on the purpose of searching the potential sufficient conditions for multistationarity in small chemical reaction networks. Although conditions in our hypothesis is not sufficient enough for guaranteeing the bistability in reaction networks, the enumerated networks can potentially provide insights on the topological conditions of multistationarity in chemical reaction networks, especially signalling networks. The comparisons between monostable networks with competition loops bistable networks without competition loops can be quite useful. Further comparisons of monostationary and multistationary networks resulted from enumeration is planned in future projects, thus not included in this PhD study.

However, from the enumeration results, it is clear that my hypothesis about competing loops is not sufficient for explaining the bistability emerged. However, there are many bistable networks fulfil this conditions. This suggest that the sufficient and necessary condition for bistability in signalling networks is very strange and somehow subtle. It shows that such problem is a very challenging problem. Nevertheless, from the enumerated networks, I might be able to derive further patterns and hypothesis to complement our current (failed) hypothesis, eventually toward the determining the sufficient and necessary condition(s).

Chapter 7

Conclusion

7.1 Evolution *in silico*: as taught and as practiced

In my PhD study, I mainly took the approach of evolution *in silico* to explore the potential design principles of signalling networks, the information processing systems in biology. The earliest computational simulations of evolutionary process can be traced back to the time when computer was invented [298]. Later, evolution *in silico* was widely recognised as a powerful optimisation method, subsequent algorithms were devised and applied in areas like optimisation and artificial intelligence. In parallel, evolution *in silico* is also utilised to study the evolution itself of biological systems. This approach of studies about biological systems has been applied in many scales. At molecular level, *in silico* evolution of proteins was adopted to understand the landscape of protein structures and dynamics, which is crucial to understand the function and evolution of proteins [299][184]. At cellular level, *in silico* evolution of regulatory networks provide insights about modularity and complexity of network evolution [119, 122] as well as designing networks with desired functions [123]. Evolutionary landscapes and genotype-phenotype maps were characterised in morphogenesis and development of teeth through evolution *in silico* at cellular level [209, 300]. The application of this approach to study evolution of biological systems

can both potentially test the hypothesis about evolution itself such as evolvability, robustness, plasticity and also provide insights on the evolutionary design principles of biological systems. Especially the later part can be used in engineering and synthetic biology.

In order to study evolution of signalling systems in cells, I combined the evolution *in silico* approach with rule-based modelling and devised a computational tool — BioJazz [212]. Adopting the rule-based modelling is critical in this study, the rule-based approach relaxed the constraints on complexity of networks in evolution of signalling networks. Also, directly manipulating rules that are used to describe interactions between proteins makes evolving the signalling network much easier than conventional approach. More importantly, the rule-based approach encloses multi-domain structure in the model of signalling networks. This is particularly useful for uncovering design principles with more biochemical details so that such principles are more applicable in designing signalling networks with desired functions.

Then I applied BioJazz to study the evolution of ultrasenstivity and adaptation with synthesised fitness function that can sufficiently evolve networks with ultrasensitive and adaptive response dynamics. By analysing the evolved networks, I discovered two interesting design principles, the first design principle is that protein sequestration (e.g. through scaffold proteins) can generate both ultrasensitive and adaptive response, even more interestingly modulate the signalling network switching between those two distinct response dynamics; the second principle is that enzymatic futile cycles with allosteric enzymes can give rise to bistable dynamics. I also provided potential application in synthetic biology for both design principles.

This PhD study of cellular information processing indicates that evolution *in silico* can help to understand the genotype-phenotype mapping of cellular systems and also explore the potentially undiscovered design principles and solutions. However, the return is possibly not always as expected. The detailedness of design

principles discovered from evolution *in silico* is largely dependent on the level of abstraction in its computational model of biological systems. For instance, spatial diffusion is a important factor affecting the signalling dynamics and information processing in cells, it is impossible for BioJazz to discover design principles with compartmentalisation or limit diffusion under homogeneity assumption. Therefore, better computational modelling approach can potentially converge with and be adopted into evolution *in silico* approach.

To sum up, *in silico* evolution approach enables actively searching design principles, rather than studying the recurring biological systems case by case, as such it provides a big leap from studying known to exploring unknown.

7.2 Design principles: the contexts and the applicability

The two design principles discovered in this PhD study are embedded in certain contexts and certain models. In the first design principles, protein sequestration has been found promoting or diminishing the ultrasensitivity levels in different networks. To sum up, the functions of protein sequestration in signalling cycles are based on zero-order sensitivity, if the sequestration happens at the enzymes, the saturation of enzyme by substrate will be enhanced thus sequestration promotes ultrasensitivity. However, if sequestration happens at the substrate such that the enzymes are less saturated, the sequestration diminishes ultrasensitivity [222]. The modulation of signalling cycles between ultrasensitivity and adaptation also appears in other network motifs [210]. This indicates that different mechanisms can potentially generate similar dynamics and behaviours under different contexts.

In the design principle for bistable signalling networks, I discussed the condition of detailed balance on the smallest bistable signalling motif (Chapter 5). The detailed balancing casts thermodynamics constraint on state transitions of kinase

in the smallest bistable motif (Chapter 5). As a result, detailed balancing [278, 280] reduced the parameter space for bistable dynamics because of its constraint on reaction rate constants. Therefore, under such context bistable dynamics is more difficult to be implemented than relaxed conditions. It is possible to make bistable dynamics more easily implemented either by embedding the bistable motif into a larger systems or relaxing the condition of detailed balancing, for instance, extending the allosteric switching induced by covalent modification (e.g. phosphorylation and dephosphorylation reactions) will relieve the constraints on its rate constants. This is a good example that the applicability of design principles depends on the context of the network and physical properties of its reactions. Such applicability can be extended by properly relaxing the conditions of network contexts.

The design principle for bistable signalling network derived from evolved networks can still be rather specific and in a narrow scope. In order to approach a more general design principle, I took the chemical reaction network theory (CRNT) to algorithmically search the possible boundaries between monostationarity and multistationarity. The CRNT is well grounded by mathematical proofs and only constrained by mass-action kinetics. Such perspective provide more general design principles for us to understand what can possibly work and what cannot [112]. The work towards a more defined necessary and sufficient condition may emerge from the area of CRNT.

7.3 Evolutionary innovations: what can we learn for engineering?

My PhD study not only generated interesting discoveries and insights about design principles of signalling networks, but also these results stimulated some potential new questions and clues about the evolution of signalling networks. Here I took a specific point of view to discuss about the thinking that arose from analysing the

evolved networks and their design principles. I will discuss the evolutionary innovations of cellular networks from the perspective of engineering and synthetic biology. Specifically, the retroactivity, cross-talks, futile cycles and noise are somehow detrimental in conventional engineering, however their functional roles under different scenarios might provide insights on design principles and potential applications in synthetic biology.

7.3.1 Retroactivity: functional roles in different contexts

Retroactivity was proposed in synthetic biology and under the background of engineering biological systems by integrating modules which perform well-defined functions (semi)independently into more complex cellular networks. However, biological systems in nature only display certain degree of modularity. These modules are interconnected with combinatorial interactions that may affect the dynamics and functions of such "modules", such effects are termed as "retroactivity". One simple example is the sequestration in gene regulatory networks, where translated proteins are sequestrated by downstream transcriptional components. The dynamics of the protein expression module is affected where sequestrated protein is the output [301]. Also, in a bifunctional enzyme catalysed signalling cycle, the sequestration of substrate by downstream targets dramatically decreases the sensitivity [260]. Therefore, following conventional engineering principles when multiple modules are integrated, retroactivity is a repellent side effect being diminished. Such retroactivity can be attenuated through implementing insulation and/or amplification [301] with certain energy cost [302].

However, it is possible that in certain context network structures the retroactivity has functional roles. The sequestration of signalling protein by both kinase and phosphatase in evolved adaptive networks showed the necessity of retroactivity on the sequestrated signalling protein (Figure 3.5) so that it can convert the linear input signal into downstream signalling cycle that transforms the signal into adaptive

response. Retroactivity between different signalling cycles where enzymes competing with the same substrate such retroactivity on contrary enhances the sensitivity of both signalling cycles [303]. These evidences strongly suggest the retroactivity has potential functional roles in implementing complex regulations and response dynamics. Our understanding of both biological systems and engineering principles may benefit from searching and studying from such evolutionary design.

7.3.2 Cross-talks: is multistability a potential rescuer?

The cross-talk is another nuisance in engineering that a signal transmitted in one circuit creates undesired effects in another circuit. In engineering, such undesired effects is likely to be avoided as much as possible so that the dynamics is predictable and reliable. One of the conventional ways is to implementing an insulator between circuits to keep the modules more independent. The cross-talks are pervasive in biological systems especially in information processing systems. Most of the signalling pathways are interconnected due to cross-talks [11]. One example is the three mitogen-activated protein kinase (MAPK) signalling cascades where several signalling proteins are shared among these cascades [304].

The direct mystery is how such information processing systems maintain specificity, given that different pathways detects different signals through various receptors and functions differently in cell behaviours. Several principles or mechanisms for specificity of MAPK signalling are proposed including scaffold protein as insulator [305], temporal specificity through transcriptional control [214] and kinetic insulation [306]. Here, the uncovered design principle for multistability in signalling networks can potentially be used as a mechanism of specificity maintenance in signalling pathways with cross-talks. Previous study on two component systems (TCS) in bacteria showed that TCS with multi-domain histidine kinase (HK) can give rise to multistability. With additional HKs the system can perform logic gates through cross-talks between different phosphorelays [50]. Similarly, in the phosphorylation-

dephosphorylation cycles with allosteric enzymes, additional allosteric enzymes provide multistability. It is very appealing to examine whether such logic gates can be implemented in the MAPK cascades, particularly whether an exclusive "OR" logic gate can be implemented through the multistability. Future works on validating such hypothesis might again extend our knowledge on functional roles of cross-talks in signalling networks and engineering specificity with cross-talks.

7.3.3 Modularity or complexity: plasticity in response dynamics

As discussed above, both retroactivity and cross-talks are prevalent in biological systems, while they act as nuisance in conventional engineering principles. Such contradictions suggest that the evolutionary designed biological systems can potentially provide new perspectives and principles for engineering such perspectives and principles may be applicable to other engineering areas. For biological systems, the contradictions indicate the gaps between modular biology and "systems" biology. Again, it encourages us to study the biological systems under the light of evolution. As the biological systems are results from evolution in fluctuating environments, their regulation systems were never selected by a single function rather by multiple objectives. Evolution under such multiple objectives inevitably brings retroactivity and cross-talks between modules. The hypothetical solutions provided by evolution is probably the plasticity in cellular networks that networks can perform multiple functions through minimal regulations and costs [35, 167, 206, 246, 307–311]. Validation and formalisation of such hypothesis requires further research inputs.

Appendix A

Manual of BioJazz

A.1 Introduction

Biological systems employ sophisticated mechanisms to sense and process information then achieve proper phenotypic behaviours so that it enables their survival in environments. The essential part of the regulation involves large-scale biochemical reaction networks that accurately compute the input signal into output response, though the computational capabilities results from interactions between proteins with merely two types of reactions: non-covalent binding reaction and post-translational modification. Observations from experiments reveal evolutionary innovations from complex signalling networks, such as allosteric regulation, cross-talk, regulatory motifs, facilitating computability of the cell [**Rowland:2014bk**, 88, 122, 193, 312, 313].

To fully understand the complexity of signalling network and its evolution, one need to utilize computational models rather than intuitively trying to capture its dynamics. Besides, it is necessary to study the evolution of complex signalling networks in order to uncover principles of nature design as well as to reverse engineer it or design novel functions beyond nature. Many researches have been carried out about evolutionary simulation of metabolic networks, or gene regulatory net-

works [121, 123]. Here, we introduce a tool for evolutionary simulating dynamic biochemical networks, aiming to explore the design principles of signalling network in cells.

BioJazz is a tool for evolving and designing biochemical reaction networks using genetic algorithm (GA). Typically, a BioJazz user wishes to evolve or design a small network or motif that accomplishes a specific function, such as a switch or an oscillator module. The network comprises a set of proteins whose attributes are encoded in a network's "genome". The "genome" is a binary string which contains all the information necessary to determine how many proteins are present in the network, their structure, which proteins interact and the biochemical parameters of their interaction.

BioJazz implements a genetic algorithm through a process of replication, mutation, and selection, attempts to incrementally improve how well those "genomes" perform a user-specified function. By encoding the network in a fashion that mimics the way nature does, BioJazz can use a larger variety of mutational operators than do traditional GAs (which use point mutations and crossover), such as gene duplications, gene deletions, and domain shuffling. Thus, BioJazz has the ability to change and evolve networks with respect to both topology and biochemical parameters, by starting from a designed network "*de novo*", or a partially or completely functional seed network. While the genetic algorithm itself is not very tasking, scoring each individual of a population of genomes may require a lot of processing power. Therefore, BioJazz has an integrated capability to use workstation clusters to speed the computation.

Much of BioJazz's ability to design realistic networks comes from the accompanying Allosteric Network Compiler (ANC) [156]. ANC is a stand-alone, rule-based compiler which has the ability to turn a high-level description of allosteric proteins into the corresponding set of biochemical equations. The proteins can exhibit many of the behaviours observed in nature, such as co-localization, allosteric transitions,

binding and catalytic reactions. The rule-based approach implemented in ANC fits in with allosteric biochemical networks. It utilizes thermodynamically grounded methodology to abstract protein structures and allosteric regulation.

Rule-based model not only solve the combinatorial explosion occurred in modelling signaling networks, but more importantly, it also makes network restructuring possible due to clustering reaction patterns by interaction rules and parameterisation of allosteric regulation with two key parameters, "Γ" and "Φ" [156], based on thermodynamic changes of protein conformation when under binding and post-translational modifications. BioJazz is likely the first tool to couple a rule-based compiler with an evolutionary algorithm.

To evolve the protein-protein interaction networks, one need to store and mutate the network of which protein structures, reaction rules and corresponding parameters are the most important. BioJazz encodes all information with binary string, that can be "transcribed" into interaction networks without losing any information. Moreover, in order to study the evolution of complex interaction networks, we need to embed the mutations of networks, both structure and kinetic parameters, into a realistic matter rather than choosing arbitrarily alter network structure and kinetic parameters. Therefore, binary string encoding provides an advantage on storing and mutating biochemical networks as an analogue of the real biological systems.

BioJazz is also highly configurable. For example, the user can specify evolutionary parameters such as mutation rates. Also, the user may restrict BioJazz to changing a subset of the network's attributes. This is useful to "freeze" the network topology, with the effect that only the network's biochemical parameters and not its structure are allowed to evolve.

The main features of BioJazz are:

- evolves both network topology and connection weights

- designs a network "*de novo*", or starting from user-specified seed network
- uses workstation clusters to speed up the design
- produces a human-readable model of network
- highly configurable

A.2 Installation and usage

A.2.1 Download

The code is hosted in Github.com and distributed with GPLv3 licence. The BioJazz code can be downloaded from **http://oss-lab.github.io/biojazz/**, **http://osslab.lifesci.warwic**
or cloned with **git clone https://github.com/OSS-Lab/biojazz.git**.

A.2.2 Installation

BioJazz requires the *ANC* and Facile tools. You can tell BioJazz where to get them by setting the *ANC_HOME* and *FACILE_HOME* environment variables to point to the appropriate directories. It is recommended to add the following lines to your " /.bashrc" file:

```
export ANC_HOME     = ~/workspace/anc
export FACILE_HOME = ~/workspace/facile
alias  anc='$ANC_HOME/anc.pl'
alias  facile='$FACILE_HOME/facile.pl'

export BIOJAZZ_HOME = ~/workspace/biojazz
alias  biojazz='$BIOJAZZ_HOME/biojazz.pl'
```

BioJazz requires Matlab to be installed on all nodes used for computation, and assumes Matlab can be started with the command "matlab". Here is an example of configuration in " /.bashrc" file (on Mac OS X):

```
export MATLAB_HOME = /Applications/MATLAB_R2011b.app/bin
alias  matlab='$MATLAB_HOME/matlab'
export PATH = $MATLAB_HOME:\$PATH

DYLD_LIBRARY_PATH = /Applications/MATLAB_R2011b.app/bin/maci64:/
    Applications/MATLAB_R2011b.app/sys/os/maci64:/Applications/
    MATLAB_R2011b.app/runtime/maci64:$DYLD_LIBRARY_PATH
export DYLD_LIBRARY_PATH
```

Note that if you decide to use a cluster of workstations, these installation instructions apply to all workstations used.

CPAN modules

CPAN is an internet database of Perl modules. BioJazz/ANC/Facile uses several of them and they must be installed prior to use. You will need system administrator priviledge to install these modules (or see for instructions on how to install them in your home directory). You or your system administrator will typically need to run the following commands on each system used:

```
cpan -i Class::Std
cpan -i Class::Std::Storable
cpan -i String::CRC32
cpan -i Expect
cpan -i Carp
cpan -i WeakRef
cpan -i IPC::Shareable
cpan -i Linux::Pid
cpan -i Text::CSV
```

You should use *sudo* as prefix if available, if you don't have an admin privilege here is a solution* that lets you install perl modules in your user directory. Then you can test your installation by running Facile, ANC and BioJazz without any arguments:

*http://twiki.org/cgi-bin/view/TWiki/HowToInstallCpanModules#Install_CPAN_modules_into_your_l

```
$FACILE_HOME/facile.pl
$ANC_HOME/anc.pl
$BIOJAZZ_HOME/biojazz.pl
```

An error will be reported if any of the required modules are still missing. Simply run CPAN again to install the missing module. If you would like ANC to generate diagrams of the reaction network and species, you will also need the "dot" application and the following CPAN module:

```
cpan -i GraphViz
```

A.2.3 Usage

Workspace creation

Depending on your specific application, BioJazz will require some customized configuration and scoring functions. Also, during a single design runs, BioJazz will generate large number of files. For this reason, the user must create a properly configured workspace which will contain the appropriate configuration files, scoring functions, and design files. To facilitate this, BioJazz can create the workspace for you and populate it with the required directories and with template files to get you started. To do this, run the following command:

```
biojazz --command='create_workspace("bjazz")'
```

This will create the directory bjazz and various sub-directories including config and custom. Your configuration files go in the config directory, while your custom scoring functions go in the custom directory. At this point, the user should familiarize him/herself with some the template files that are provided, and try to run BioJazz.

The example file will try to design a network which contains a signalling cascades, and demonstrates how to use some functions available to the user.

```
cd bjazz
```

```
less config/ultrasensitive.cfg   # ultrasensitive configuration file
less config/Ultrasensitive.pm    # ultrasensitive application-specific
    scoring function
```

Running BioJazz

After installing the required Perl modules, it is time to run BioJazz. The *cluster_type* and *cluster_size* arguments override the specification contained in the configuration file, and will launch both slave nodes of the cluster on your machine.

```
biojazz --config=config/template.cfg --tag=first_try --cluster_type="
   LOCAL" --cluster_size=2
```

This will evolve the network for only a couple generations. The tag argument is very important. In BioJazz, each design attempt is associated with a specific, user-specified tag. BioJazz will create a directory in your workspace containing all the results and other files generated during the optimization. This allows the user to attempt several optimizations simultaneously without fear of accidental loss of files. The name of the design's working directory is *work_dir/tag*. The *work_dir* parameter is specified in your configuration file (and has a value of template in this example). The results of the above run are contained in the directory *ultrasensitive/first_try*.

```
[user@host bjazz]\$ ls -la ultrasensitive/first_try/
total 168
drwx------ 5 user group  4096 2013-06-03 14:53 .
drwx------ 3 user group  4096 2013-06-03 14:51 ..
drwx------ 2 user group  4096 2013-06-03 14:53 matlab
drwx------ 2 user group  4096 2013-06-03 14:53 obj
drwx------ 1 user group  4096 2013-06-03 14:53 report
drwx------ 1 user group  4096 2013-06-03 14:53 stat
drwx------ 2 user group  4096 2013-06-03 14:51 source_2013
   -06-03-14:51:58
```

The *obj* directory contains all the genomes generated in a machine-readable form. The *matlab* contains the models generated by ANC, and the Matlab scripts

generated by Facile. The *stat* contains the output information of each genome in each generation in *.csv* files. The *source** directory is a snapshot of the source code used for that run such as your configuration and custom scoring files. Now you can try modifying the configuration file to use other available workstations and run BioJazz again.

Workspace directory structure

```
bjazz                                # workspace home
  config                             # configuration files
  custom                             # application-specific modules and functions (incl. scoring function)
  test/custom                        # recommended location for test results of custom modules
  test/modules                       # BioJazz module test results
  ultrasensitive                     # application-specific directory
    first_try                        # results directory for run with TAG=08jun01
      matlab                         # ANC genome models, eqn files, and matlab files
      obj                            # genome objects in binary form
      report                         # post evolution analysis
      stat                           # information about individual genome in each generation
```

Initial Generation

The initial generation can be either generated randomly or loaded from disk, as specified by the *initial_generation* parameter of the configuration file. In the random case, the user can also specify the number of individuals to create (parameter *inum_genomes*) and the genome length (parameter – currently fixed at 5000). Loading from disk is useful to resume work on a partially completed design starting from the last generation created, or to load hand-crafted seed designs. The following

shows some examples for each case:

```
initial_genome = random                                    # random generation
initial_genome = load test/modules/Ultrasensitive.obj      # load a hand-crafted network
initial_genome = load ultrasensitive/test/obj/G427_I*.obj  # load all individuals of generation 427 of previous run
```

Regardless of how the initial generation is created, each network is stored under the following name in the working directory of the design:

```
$DESIGN_WORK/obj/G***_I%%.obj
```

Where *** is the generation number and %% is the individual number.

A.2.4 Scoring

The principal user input consists of a scoring function, which evaluates a particular genome against the desired functionality, and returns a score reflecting the network's performance. This score is compared against the score of other networks to determine whether the network survives to the next generation and replicates. Generally speaking, this involves applying a stimulus to the network and evaluating it's response. Simulation of the network is accomplished by integrating a set of ordinary differential equations (ODEs) in "Matlab". The required Matlab files are automatically generated from ANC's output using a tool called Facile.[180]

Scoring part is composed of three main parts: stimulus class, scoring class and specific scoring subclass. The stimulus class is used to generate a stimulus waveform to apply on a specific node/species (usually the ligand) in the reaction network. ANC constructs biochemical equations for stimuli consisting of either a time-varing *source* or *sink* or both, which expressed as:

$$\text{null} \longrightarrow \text{X}; \qquad \text{source} = f(t)$$

$$\text{X} \longrightarrow \text{null}; \qquad \text{sink} = g(t)$$

The scoring class uses MATLAB® to simulate the network and return the simulation results. The customized subclass constructs the network input and output and uses the simulated results to scoring the network based on certain input-output response pattern, such as ultransensitivity, oscillation, linear, hyperbolic *etc.*

A.3 Example *config* file

Here is an example of configuration file:

```
#####################################################
# BioJazz configuration
#####################################################

#-----------------------------------------
# CPU AND CLUSTER SETTINGS
#-----------------------------------------
cluster_type = LOCAL
cluster_size = 1
nice = 15
vmem = 200000000

#-----------------------------------------
# WORKSPACE AND CUSTOM SCORING MODULES
#-----------------------------------------
scoring_class = Ultrasensitive
work_dir = ultrasensitive
local_dir = ultrasensitive/localdir
```

```
19
20 initial_genome = random
21 #initial_genome = load test/custom/Ultrasensitive.obj
22
23
24 #----------------------------------------
25 # GENOME PARAMS
26 #----------------------------------------
27
28 # Scaling: all concentrations in uM, all 2nd-order rates in uM^-1 s^-1
29
30 # Genome class
31 radius = 3      # should be reasonable. Binomial[Width,radius..0]/2^
     width
32 kf_max = 1e3     # uM^-1 s^-1
33 kf_min = 1e-3
34 kb_max = 1e3
35 kb_min = 1e-3
36 kp_max = 1e3
37 kp_min = 1e-3
38
39 # Gene class
40 regulated_concentration_width = 10
41 gene_unused_width = 4
42 regulated_concentration_max = 1e3    # 1mM
43 regulated_concentration_min = 1e-3   # 1nM  ~ 1 molecule in prokaryote
44
45 # Domain class
46 RT_transition_rate_width = 10
47 TR_transition_rate_width = 10
48 RT_phi_width = 10
49 domain_unused_width = 4
50 RT_transition_rate_max = 1e2
51 RT_transition_rate_min = 1e-2
```

```
TR_transition_rate_max = 1e2
TR_transition_rate_min = 1e-2
RT_phi_max = 1.0
RT_phi_min = 0.0

# ProtoDomain class
binding_profile_width = 10
kf_profile_width = 20
kb_profile_width = 20
kp_profile_width = 10
steric_factor_profile_width = 20
Keq_profile_width = 10
protodomain_unused_width = 4
Keq_ratio_max = 1e2
Keq_ratio_min = 1e-2

#----------------------------------------
# EVOLUTION PARAMS
#----------------------------------------
num_generations = 10000
target_score = 0.8
first_generation = 0    # define the number of the first generation, either 0 or 1
continue_sim = 0
continue_init = 0
remove_old_files = 1
score_initial_generation = 1
rescore_elite = 0
report_on_fly = 1
report_selection = 0        # because of fossil_epoch you may lose information if collect information later!!

# selection method: kimura selection
selection_method = kimura_selection
```

```
84 effective_population_size = 1e8 # for kimura_selection only
85 amplifier_alpha = 1e3   # for kimura_selection only, speed up the
       evolution, range: The lowe bound is 1.16 the upper bound is 2*
       effective_population_size
86 max_mutate_attempts = 100000 # default -1 or not defined as unlimited,
       should be an integer;
87
88 # selection method: population-based selection
89 #selection_method = population_based_selection
90 #fossil_epoch = 10    # for genome storage and records of genomes in
       certain generations, comment if using kimura selectio method (must)
       or record every generation
91 #inum_genomes = 50    # for kimura selection method, doesn't matter
       because it's set 1 as default.
92 #evolve_population = 1000    # for population-based selection method
       only
93 #mutation_rate = 0.05  # For population-based model
94
95 # mutation settings
96 mutation_rate_params = 0.0
97 mutation_rate_global = 0.01
98 gene_duplication_rate = 0.005
99 gene_deletion_rate = 0.005
100 domain_duplication_rate = 0.005
101 domain_deletion_rate = 0.005
102 recombination_rate = 0.01
103 hgt_rate = 0.01  # currently not implemented yet
104
105 #----------------------------------------
106 # ANALYSIS PARAMS (POST-EVOLUTION)
107 #----------------------------------------
108 #report_on_collection = 1    # for population based method usually set
       as 1 !!(Currently not implemented)
109 restore_genome = 0
```

```
analysis_dir = analysis

#####################################################
# User-defined, application-specific configuration
#####################################################

#-----------------------------------------
# ANC PARAMS
#-----------------------------------------
max_external_iterations = -1
max_internal_iterations = -1
max_complex_size = 3  #MATLAB has maximal length of names, if using MATLAB as simulator, this value should always be less than 9. Either -1(unlimited) or 6 should be resonable, please ref the Plos ONE paper from Vincent Danos group.
max_species = 512
max_csite_bound_to_msite_number = 1 # originally set as 1, but if consider more complex situation, we should put this unlimited, which means in complex multiple csite-msite bindings could happen.
default_max_count = 2          # this prevents polymerization (see ANC manual)
default_steric_factor = 1000      # in micro-mol/L
export_graphviz = nothing
#export_graphviz = network,collapse_states,collapse_complexes
#export_graphviz = network,collapse_states,collapse_complexes,primary,scalar,ungrouped,canonical # possibly there are more information could be output

#-----------------------------------------
# FACILE/MATLAB SETTINGS
#-----------------------------------------
solver = ode23s
#solver = stoch

```

```
sampling_interval = 1.0
SS_timescale = 500.0

# MATLAB odeset params
InitialStep = 1e-8
AbsTol = 1e-9
RelTol = 1e-3
MaxStep = 500.0

#----------------------------------------
# SIMULATION/SCORING PARAMS
#----------------------------------------
plot_input = 1
plot_output = 1
plot_species = 0
plot_phase = 1
plot_min = -1

round_values_flag = 0

steady_state_threshold = 1000   # IC settling time
steady_state_score_threshold = 0.5

delta_threshold = 0.01          # relative measure of amplitude used to filter out integration noise
amplitude_threshold = 0.01      # absolute measure of amplitude
ultrasensitivity_threshold = 5  # ratio of 2nd step over 1st step
complexity_threshold = 250
expression_threshold = 500

w_n = 0.0
w_c = 0.0   # complexity score weight
w_e = 0.0
w_s = 1.0
```

```
w_a = 1.0
w_u = 1.0
w_u1 = 1.0
w_u3 = 1.0

LG_range = 10           # uM (about 6 molecules in 1e-18L vol ???)
LG_delay = ~
LG_strength = 4.0       # in Hz
LG_ramp_time = 3000
LG_steps = 3

LG_timeout = 20000

#stimulus = staircase_equation
#stimulus = ramp_equation
stimulus = ss_ramp_equation

#hill_n = 8
hill_n = 40
hill_k = 5

TG_init = 1000  # uM
cell_volume = 1e-18             # 1e-18L --> sub-cellular volume

# to make sure the input and output have relatively large distance and also have relative large distance from themselves
# and also make sure their binding partner to have relatively large distance in this case the intermediate binding profile could be 0010110100 have both 5 distanct to all four binding profiles
# it depends the problem, whether want far distances between initial profiles or shorter distances
lg_binding_profile = 0100111010
tg_binding_profile = 0111000110

```

```
#-----------------------------------------
# SPREADSHEET EXPORT/ANALYSIS
#-----------------------------------------
genome_attribute_names = \
                         score, \
                         ultrasensitivity_score, \
                         expression_score, \
                         amplitude_score, \
                         complexity_score, \
                         steady_state_score, \
                         complexity,\
                         num_anc_species,\
                         num_rules,\
                         num_genes,\
                         num_pruned_genes,\
                         num_domains,\
                         num_protodomains,\
                         num_allosteric_domains,\
                         num_allosteric_protodomains,\
                         num_binding_protodomains,\
                         num_phosphorylation_protodomains,\
                         num_catalytic_protodomains,\
                         num_kinase_protodomains,\
                         num_phosphatase_protodomains,\
                         num_adjacent_kinases,\
                         num_adjacent_phosphatases,\
                         num_receptive_protodomains,\
                         tg_K1,\
                         tg_K2,\
                         tg_K1_concentration,\
                         tg_K2_concentration,\
```

Appendix B

Thermodynamic framework in ANC and BioJazz

B.1 Thermodynamic framework for modelling allosteric regulation

The Arrhenius equation gives the kinetic rate of the R-T transition as: $k_{RT} = k_{R\dagger} = Ae^{-\Delta G_{R\dagger}/kT}$, with "†" denoting the transition state, A denoting the Arrhenius constant, and kT being the product of Boltzmanns constant and temperature [314]. Similarly, $k_{TR} = k_{T\dagger} = Ae^{-\Delta G_{T\dagger}/kT}$. The equilibrium distribution of the R and T states will be governed by the equilibrium constant K_{RT}, which is given by k_{RT}/k_{TR}, where $K_{RT} = Ae^{(-\Delta G_{R\dagger}+\Delta G_{T\dagger}/kT)} = Ae^{-\Delta G_{RT}/kT}$.

In ANC, modifiers are assumed to contribute independently to the free energy of each conformational state, R and T, allowing us to formulate the free energy difference between these two states ($\Delta G'_{RT}$) in a given domain with N modifiers as:

$$\Delta G'_{RT} = \Delta G_{RT} + \sum_{i=1}^{N} (\Delta G_T^{(i)} - \Delta G_R^{(i)}) \tag{B.1}$$

where $\Delta G_T^{(i)}$ and $\Delta G_R^{(i)}$ give the effect of the ith modifier free energies of the R and T

states. While $\Delta G'_{RT}$ could be evaluated via Equation B.1, this requires assignment of the $\Delta G_T^{(i)}$ and $\Delta G_R^{(i)}$ values. Instead of doing this, we can exponentiate Equation B.1 and thus equivalently define the effect of each modifier on the overall equilibrium distribution between the R and T states. To do so, we defined the relation of the equilibrium constant of the domain without any modifiers (K_{RT}) to that with modifiers (K'_{RT}) as:

$$\frac{k'_{RT}}{k'_{TR}} = K'_{RT} = K_{RT} \prod_{i=1}^{N} \Gamma_i \tag{B.2}$$

where $\Gamma_i = e^{-(\Delta G_T^{(i)} - \Delta G_R^{(i)})/kT}$ denotes the effect of the ith modifier on the equilibrium distribution between the R and T states. The Γ_i relate to the altered kinetic rate constants in the presence of the ith modifier in the following manner:

$$k'_{RT} = k_{RT} \prod_{i=1}^{N} (\Gamma_i)^{\Phi_i} \tag{B.3}$$

$$k'_{TR} = k_{TR} \prod_{i=1}^{N} (\Gamma_i)^{\Phi_i} \tag{B.4}$$

with the parameter Φ_i describing the proportional effects of the ith modifier on the R-T transitions. To simplify the implementation of this approach, all modifiers acting on different reactive sites of a domain are assumed to employ the same Φ value (i.e. $\Phi_i = \Phi_j (i \neq j)$ for all reactive sites in one domain) [156].

Appendix C

Mathematical model of sequestration motif

C.1 Mathematical model of sequestration motif

This particular motif describes one phosphorylation-dephosphorylation cycle of substrate protein (S and S_p), which can potentially be generalised into any futile cycles, with both kinase (K) and phosphatase (P) that are sequestrated by a scaffold protein (T). The corresponding chemical reactions are:

$$K + S \rightleftharpoons KS \rightarrow K + S_p$$

$$P + S_p \rightleftharpoons PS_p \rightarrow P + S$$

$$T + K \rightleftharpoons TK$$

$$T + P \rightleftharpoons TP$$

The above reactions show a simple system that composed of one scaffold protein, one kinase, one phosphatase and one substrate. Here we try to describe this simple

system with differential equation following the mass action kinetics:

$$
\begin{aligned}
\frac{\mathrm{d}[K]}{\mathrm{d}t} &= -\kappa_1[K][S] + \kappa_2[KS] + \kappa_3[KS] - \kappa_7[T][K] + \kappa_8[TK] \\
\frac{\mathrm{d}[K]}{\mathrm{d}t} &= -\kappa_4[K][S] + \kappa_5[PS_p] + \kappa_6[PS_p] - \kappa_9[T][P] + \kappa_{10}[TP] \\
\frac{\mathrm{d}[S]}{\mathrm{d}t} &= -\kappa_1[K][S] + \kappa_2[KS] + \kappa_6[PS_p] \\
\frac{\mathrm{d}[S_p]}{\mathrm{d}t} &= -\kappa_4[P][S_p] + \kappa_3[KS] + \kappa_5[PS_p] \\
\frac{\mathrm{d}[KS]}{\mathrm{d}t} &= \kappa_1[K][S] - \kappa_2[KS] - \kappa_3[KS] \\
\frac{\mathrm{d}[PS_p]}{\mathrm{d}t} &= \kappa_4[P][S_p] - \kappa_5[PS_p] - \kappa_6[PS_p] \\
\frac{\mathrm{d}[T]}{\mathrm{d}t} &= -\kappa_7[T][K] + \kappa_8[TK] - \kappa_9[T][P] + \kappa_{10}[TP] \\
\frac{\mathrm{d}[TK]}{\mathrm{d}t} &= \kappa_7[T][K] - \kappa_8[TK] \\
\frac{\mathrm{d}[TP]}{\mathrm{d}t} &= \kappa_9[T][P] - \kappa_{10}[TP].
\end{aligned}
$$

And the system need to follow these conservation equations:

$$
\begin{aligned}
[S_{tot}] &= [S] + [S_p] + [KS] + [PS_p] \\
[K_{tot}] &= [K] + [KS] + [TK] \\
[P_{tot}] &= [P] + [PS_p] + [TP] \\
[T_{tot}] &= [T] + [TK] + [TP].
\end{aligned}
$$

Appendix D

CRNToolbox analysis

D.1 CRNToolbox analysis of the simplest bistable motif

CRNToolbox is a powerful tool to analyse the dynamical behaviours of chemical reaction networks based on their structural properties, including deficiency and injectivity [237–240]. The program can be downloaded from (**https://crnt.osu.edu/CRNTWin**).

Here, I take the simplest bistable motif discovered in Chapter 4 and 5 to illustrate how to use CRNToolbox to analysis (bio)chemical reaction networks. The reaction network of bistable motif is as follows:

$$
\begin{aligned}
K_r + S &\rightleftharpoons K_rS \rightarrow K_r + S_p \\
K_t + S &\rightleftharpoons K_tS \rightarrow K_t + S_p \\
S_p &\rightarrow S \\
K_r &\rightleftharpoons K_t \\
K_rS &\rightleftharpoons K_tS.
\end{aligned}
$$

The reaction network is composed of 11 reactions and 6 species. First, we need to type the reaction networks into CRNToolbox, then get the basic analysis about deficiency of the reaction network:

```
                    ==========================================
                    BASIC REPORT: simplest bistable motif
                    ==========================================
Reaction network:

                              Kr + S <-> KrS
                                 KrS <-> KtS
                                 KrS -> Kr + Sp
                              Kt + S <-> KtS
                                 KtS -> Kt + Sp
                                  Sp -> S
                                  Kr <-> Kt

                           Graphical Properties
                           ====================
Number of complexes = 10
Number of linkage classes = 3:

 Linkage class no. 1: {Kr + S, KrS, Kt + S, KtS, Kr + Sp, Kt + Sp}
 Linkage class no. 2: {Sp, S}
 Linkage class no. 3: {Kr, Kt}

Number of TERMINAL strong linkage classes = 4:

 Strong linkage class no. 1: {Kr, Kt}
 Strong linkage class no. 2: {Kr + Sp}
 Strong linkage class no. 3: {Kt + Sp}
 Strong linkage class no. 4: {S}

Number of NON-TERMINAL strong linkage classes = 2:

 Strong linkage class no. 5: {Kr + S, KrS, Kt + S, KtS}
 Strong linkage class no. 6: {Sp}

```

```
The network is neither reversible nor weakly reversible.

The network is conservative. (There exists a positive vector
   orthogonal to all reaction vectors.)

                              Rank Information
                              ================

Rank of entire network = 4

                            Deficiency Information
                            ======================

Deficiency of entire network = 3

Deficiency of linkage class no. 1 = 1
Deficiency of linkage class no. 2 = 0
Deficiency of linkage class no. 3 = 0

                                  Analysis
                                  ========
    This is a deficiency three network. It is an excellent candidate
   for application of HIGHER DEFICIENCY THEORY (tailored mostly to
   networks with deficiencies greater than one).

    Whether results will be obtained, will depend on whether or not
   the reaction network has certain additional structural attributes
   that help reduce the problem to a study of systems of linear
   inequalities.

    If a network is "good", higher deficiency theory will determine,
   either affirmatively or negatively, whether there are positive rate
   constant values such that the corresponding mass action differential
    equations admit multiple (positive) steady states. If the answer is
    affirmative, higher deficiency theory will generate a sample set of
```

```
 rate constants and a pair of distinct steady states that are
consistent with those rate constants.

 If a network is "bad", some additional nonlinear analysis might be
 required, and the program might not be able to ascertain the
network's capacity for multiple  positive steady states. If definite
 conclusions can be reached they they will be reported. Otherwise
the program will tell you that it cannot reach a conclusion.

 Higher deficiency theory will also determine, either affirmatively
 or negatively, whether there can exist a set of rate constants such
 that the corresponding mass action differential equations admit a
positive steady state having a zero eigenvalue (corresponding to an
eigenvector in the stoichiometric subspace).  When the answer is
affirmative, the theory will produce such a set of rate constants, a
 positive steady state, and an eigenvector (in the stoichiometric
subspace) corresponding to an eigenvalue of zero. Results of this
kind are contained after running the Zero Eigenvalue Report.

 For information about still other reports (including those that
provide information when the kinetics is not mass action) see the
CRNToolbox Guide pdf file that accompanied this program.

        Introductory References for Chemical Reaction Network Theory

 The following provides a general introduction to parts of Chemical
Reaction Network Theory that are centered on the network's
deficiency:

Feinberg, M., Chemical reaction network structure and the stability of
 complex isothermal reactors. I. The deficiency zero and deficiency
one theorems, Chem. Eng. Science, 42, 2229-2268 (1987).

 The following is a typewritten set of lectures on reaction
```

```
networks that are aimed at mathematicians:

Feinberg, M. Lectures on Chemical Reaction Networks, Written versions
   of lectures given at the Mathematics Research Center, University of
   Wisconsin, Autumn, 1979, available at: http://www.crnt.osu.edu/
   LecturesOnReactionNetworks

      An introduction to more recent work can be found here:

Craciun, G., Y. Z. Tang, and M. Feinberg. 2006. Understanding
   bistability in complex enzyme-driven reaction networks. Proc. Natl
   Acad Sci USA 103:8697-8702 (2006).
```

The reaction network of bistable motif has deficiency higher than 1. As suggested, I did the higher deficiency analysis to determine whether the network allows multistationarity with some positive parameters.

```
                    ==========================================
                    HIGHER DEFICIENCY REPORT: simplest bistable motif
                    ==========================================
                                    Analysis
                                    ========
      Taken with mass action kinetics, the network DOES have the
   capacity for multiple steady states. That is, there are rate
   constants that give rise to two or more positive (stoichiometrically
    compatible) steady states -- you'll see an example below. There
   MIGHT also exist rate constants for which there is a steady state
   having an eigenvector (in the stoichiometric subspace) corresponding
    to an eigenvalue of zero. (To try to construct rate constants that
   give a degenerate steady state, use the Zero Eigenvalue Report.)

A mass action system example is also given below:

                        Example No. 1: Multiple Steady States

```

```
  The following mass action system gives rise to multiple steady states:

                        Kr + S ---22161.264-> KrS
                           KrS ---6.979721--> Kr + S
                           KrS ---39.54158--> KtS
                           KrS ---12.030025-> Kr + Sp
                        Kt + S ---31729.032-> KtS
                           KtS ---2.4214323-> KrS
                           KtS ---23.071536-> Kt + S
                           KtS -------1-----> Kt + Sp
                            Sp ---3.0100083-> S
                            Kr ---63.181325-> Kt
                            Kt ---109.9872--> Kr

  The steady states shown below are both consistent with the mass action system indicated.

 Steady State No. 1          Species  Steady State No. 2

     4.6744 E-3                  Kr          1.4042 E-2
     1.7012 E-2                  S           7.6443 E-3
     4.2314 E-2                 KrS          5.1682 E-2
     0.26723489                  Sp          0.29534023
     1.1395 E-2                  Kt          2.0763 E-2
     0.29534023                 KtS          0.26723489

                    Eigenvalues for Steady State No. 1

                              -3484.6708
                              2.1929672
                              -179.33244
                              -1901.9663
```

```
                        Steady State No. 1 is unstable.

                      Eigenvalues for Steady State No. 2

                                  -4111.8342
                                  -3.6463844
                                  -192.90991
                                  -1255.3861

                 Steady State No. 2 is asymptotically stable.

                                  References

1. Feinberg, M., Chemical reaction network structure and the stability
    of complex isothermal reactors. I. The deficiency zero and
   deficiency one theorems, Chem. Eng. Science, 42, 2229-2268 (1987).

2. Ellison, P. and Feinberg, M. How catalytic mechanisms reveal
   themselves in multiple steady state data. I. Basic principles, The
   Journal of Molecular Catalysis A: Chemical, 154, 155 - 167, 2000.

3. Ellison, P. PhD. Thesis. Rochester, NY: Department of Chemical
   Engineering, University of Rochester; 1998. The advanced deficiency
   algorithm and its applications to mechanism discrimination.

4. Ji, H. PhD. Thesis. Columbus, OH: Department of Mathematics, The
   Ohio
State University; 2011. Uniqueness of equilibria for complex chemical
   reaction
networks.
```

The analysis shows that the reaction network indeed admits multiple steady states. The toolbox also provides an instance of parameter set that enables multi-stationarity.

Appendix E

Proof of Multistability in Allosteric Motif

E.1 A model for an allosteric kinase

E.1.1 Model description

We consider a reaction network consisting of an allosteric kinase for one substrate. We let K be the kinase that exists in two conformations: K_r (relaxed state) and K_t (tensed state). Each of the conformations acts as a kinase for a common substrate S. We let S_p denote the phosphorylated form of the substrate. We assume that the intermediate kinase-substrate complexes, K_rS and K_tS, also undergo conformational change.

These considerations give rise to a reaction network with the following reactions:

- Phosphorylation of S:

$$K_r + S \underset{\kappa_2}{\overset{\kappa_1}{\rightleftharpoons}} K_rS \xrightarrow{\kappa_3} K_r + S_p \qquad K_t + S \underset{\kappa_5}{\overset{\kappa_4}{\rightleftharpoons}} K_tS \xrightarrow{\kappa_6} K_t + S_p$$

- Dephosphorylation of S_p:

$$S_p \xrightarrow{\kappa_7} S$$

- Conformational change:

$$K_r \underset{\kappa_9}{\overset{\kappa_8}{\rightleftharpoons}} K_t \qquad\qquad K_rS \underset{\kappa_{11}}{\overset{\kappa_{10}}{\rightleftharpoons}} K_tS.$$

We denote the concentration of the 6 species of the network as follows:

$$x_1 := [K_r] \quad x_2 := [K_t] \quad x_3 := [K_rS] \quad x_4 := [K_tS] \quad x_5 := [S] \quad x_6 := [S_p].$$

Under the law of mass action, the dynamics of the concentrations is modeled over time by the following system of ordinary differential equations:

$$\begin{aligned}
\dot{x}_1 &= -\kappa_1 x_1 x_5 + (\kappa_2 + \kappa_3)x_3 - \kappa_8 x_1 + \kappa_9 x_2 \\
\dot{x}_2 &= -\kappa_4 x_2 x_5 + (\kappa_5 + \kappa_6)x_4 + \kappa_8 x_1 - \kappa_9 x_2 \\
\dot{x}_3 &= \kappa_1 x_1 x_5 - (\kappa_2 + \kappa_3)x_3 - \kappa_{10} x_3 + \kappa_{11} x_4 \\
\dot{x}_4 &= \kappa_4 x_2 x_5 - (\kappa_5 + \kappa_6)x_4 + \kappa_{10} x_3 - \kappa_{11} x_4 \\
\dot{x}_5 &= -\kappa_1 x_1 x_5 - \kappa_4 x_2 x_5 + \kappa_2 x_3 + \kappa_5 x_4 + \kappa_7 x_6 \\
\dot{x}_6 &= \kappa_3 x_3 + \kappa_6 x_4 - \kappa_7 x_6,
\end{aligned}$$

where $\dot{x}$ denotes the derivative of x with respect to time t and reference to time t is omitted, that is, $x_* = x_*(t)$ and $\dot{x}_* = \dot{x}_*(t)$.

Since

$$\dot{x}_1 + \dot{x}_2 + \dot{x}_3 + \dot{x}_4 = 0 \quad \text{and} \quad \dot{x}_3 + \dot{x}_4 + \dot{x}_5 + \dot{x}_6 = 0,$$

the sums $x_1 + x_2 + x_3 + x_4$ and $x_3 + x_4 + x_5 + x_6$ are constant over time. This leads

to the following two *conservation laws*:

$$x_1 + x_2 + x_3 + x_4 = K_{\text{tot}}, \qquad x_3 + x_4 + x_5 + x_6 = S_{\text{tot}}. \tag{E.1}$$

Here $K_{\text{tot}}, S_{\text{tot}} > 0$ are positive *total amounts*.

E.1.2 Summary of results

The results for the model with one allosteric kinase can be summarised in the following way. In subsection E.1.3 we show that the steady states of the system can be given in terms of the concentration x_5 of the substrate S only. That is, knowing the value of x_5 at steady state allows us to calculate the value of the remaining concentrations from x_5 alone. Further, we show that the system can have up to three positive steady states by choosing the reaction rate constants and total amounts appropriately.

In subsections E.1.4-E.1.6 we study necessary and sufficient conditions for multistationarity to occur. In subsection E.1.4 necessary conditions for multistationarity on the reaction rate constants and the total amounts are given. Specifically, a necessary condition for multistationarity is

$$\alpha_1 K_{\text{tot}} + \alpha_2 < S_{\text{tot}} < \alpha_3 K_{\text{tot}} + \alpha_4,$$

where $\alpha_1, \ldots, \alpha_4$ depend on the reaction rate constants.

In subsection E.1.5 we focus on conditions that are both necessary and sufficient for multistationarity. We show that if the following inequality on the reaction rate constants is fulfilled, then the system exhibits multistationarity by choosing appropriate total amounts:

$$(\kappa_3 - \kappa_6)(\eta_r \kappa_9 \kappa_{10} - \eta_t \kappa_8 \kappa_{11}) > ((\kappa_6 + \kappa_7)\kappa_{10} + (\kappa_3 + \kappa_7)\kappa_{11})(\eta_r \kappa_{10} + \eta_t \kappa_{11})$$

where

$$\eta_r = \frac{\kappa_1}{\kappa_2 + \kappa_3}, \qquad \text{and} \qquad \eta_t = \frac{\kappa_4}{\kappa_5 + \kappa_6}.$$

If the inequality is not fulfilled, then there cannot be multistationarity for any choice of total amounts. Moreover, by inspecting the inequality, necessary conditions for multistationarity might be induced. For example, one of the following two constraints is necessary for multistationarity to occur:

(a) $\kappa_3 > \kappa_6$ and $\eta_r \kappa_9 \kappa_{10} > \eta_t \kappa_8 \kappa_{11}$.

(b) $\kappa_3 < \kappa_6$ and $\eta_r \kappa_9 \kappa_{10} < \eta_t \kappa_8 \kappa_{11}$.

If a set of rate constants fulfil the necessary and sufficient conditions for multistationarity, then the next question is to find total amounts for which it occurs. The linear inequalities in S_{tot} and K_{tot} given above restrict the possible values considerably. However, it is also possible to give necessary and sufficient conditions involving all parameters, that is, the reaction rate constants *and* the total amounts. These conditions are easy to check for a specific choice of parameters but are little illuminating in themselves.

In subsection E.1.6 we discuss how to explicitly find parameter sets for which multistationarity arises, using the conditions discussed above, and illustrate it with one example.

In subsection E.1.7 we show the steady states of the system *cannot* be given in terms of the concentration x_6 of the modified substrate S_p only, since when multistationarity occurs x_5 cannot be expressed as a function of x_6. (The substrates S and S_p do *not* appear in a symmetric way in the reactions). Further, we describe in detail the species concentrations at steady state as functions of x_5.

In subsection E.1.8, we consider bifurcation plots in the multistationary setting. We study the effect of changing the total amounts of the substrate and the kinase on the number of steady states. We encounter here again the necessary and sufficient conditions from subsection E.1.5.

E.1.3 Positive steady states

Parameterization of steady states

The positive steady states of the system are the solutions to the equations $\dot{x}_1, \ldots, \dot{x}_6 = 0$, constrained by the conservation laws (E.1). Due to the conservation laws, the equations $\dot{x}_1 = 0$ and $\dot{x}_5 = 0$ can be disregarded.

Consider first the system of equations given by $\dot{x}_2 = \dot{x}_3 = \dot{x}_4 = \dot{x}_6 = 0$ and the first conservation law in (E.1). That is, consider the system of equations:

$$
\begin{aligned}
0 &= -\kappa_4 x_2 x_5 + (\kappa_5 + \kappa_6) x_4 + \kappa_8 x_1 - \kappa_9 x_2 \\
0 &= \kappa_1 x_1 x_5 - (\kappa_2 + \kappa_3) x_3 - \kappa_{10} x_3 + \kappa_{11} x_4 \\
0 &= \kappa_4 x_2 x_5 - (\kappa_5 + \kappa_6) x_4 + \kappa_{10} x_3 - \kappa_{11} x_4 && \text{(E.2)} \\
0 &= \kappa_3 x_3 + \kappa_6 x_4 - \kappa_7 x_6, && \text{(E.3)} \\
K_{\text{tot}} &= x_1 + x_2 + x_3 + x_4.
\end{aligned}
$$

This system is linear in x_1, x_2, x_3, x_4, x_6 with coefficients involving the reaction rate constants and x_5. We obtain the following algebraic expressions for x_1, x_2, x_3, x_4, x_6

at steady state, which depend on the value of x_5 at steady state:

$$x_1 = \frac{K_{\text{tot}}}{q(x)}\Big((\kappa_2+\kappa_3)\kappa_4\kappa_{11}x_5 + \kappa_9((\kappa_2+\kappa_3)(\kappa_5+\kappa_6) + (\kappa_2+\kappa_3)\kappa_{11} + (\kappa_5+\kappa_6)\kappa_{10})\Big) \tag{E.4}$$

$$x_2 = \frac{K_{\text{tot}}}{q(x)}\Big((\kappa_5+\kappa_6)\kappa_1\kappa_{10}x_5 + \kappa_8((\kappa_2+\kappa_3)(\kappa_5+\kappa_6) + (\kappa_2+\kappa_3)\kappa_{11} + (\kappa_5+\kappa_6)\kappa_{10})\Big) \tag{E.5}$$

$$x_3 = \frac{K_{\text{tot}}x_5}{q(x)}\Big(\kappa_1\kappa_4\kappa_{11}x_5 + \kappa_1\kappa_9(\kappa_5+\kappa_6+\kappa_{11}) + \kappa_4\kappa_8\kappa_{11}\Big) \tag{E.6}$$

$$x_4 = \frac{K_{\text{tot}}x_5}{q(x)}\Big(\kappa_1\kappa_4\kappa_{10}x_5 + \kappa_4\kappa_8(\kappa_2+\kappa_3+\kappa_{10}) + \kappa_1\kappa_9\kappa_{10}\Big) \tag{E.7}$$

$$x_6 = \frac{K_{\text{tot}}x_5}{\kappa_7 q(x)}\Big(\kappa_1\kappa_4\left(\kappa_3\kappa_{11}+\kappa_6\kappa_{10}\right)x_5 + \kappa_1\kappa_3\kappa_9(\kappa_5+\kappa_{11}) + \kappa_4\kappa_8(\kappa_2\kappa_6+\kappa_3\kappa_{11}) + \kappa_6(\kappa_3+\kappa_{10})(\kappa_1\kappa_9+\kappa_4\kappa_8)\Big) \tag{E.8}$$

$$\begin{aligned} q(x) :=& \kappa_1\kappa_4(\kappa_{10}+\kappa_{11})x_5^2 \\ &+ ((\kappa_2+\kappa_3)\kappa_4(\kappa_8+\kappa_{11}) + (\kappa_5+\kappa_6)\kappa_1(\kappa_9+\kappa_{10}) + (\kappa_{10}+\kappa_{11})(\kappa_1\kappa_9+\kappa_4\kappa_8))x_5 \\ &+ (\kappa_8+\kappa_9)((\kappa_2+\kappa_3)(\kappa_5+\kappa_6+\kappa_{11}) + \kappa_{10}(\kappa_5+\kappa_6)). \end{aligned}$$

The expressions for x_1, x_2, x_3, x_4, x_6 are positive provided x_5 is positive.

The steady state polynomial

All concentrations are expressed as functions of x_5. After replacing x_3, x_4, x_6 in the second conservation law in (E.1) by their expressions in (E.6),(E.7),(E.8), we obtain that the value of x_5 at a positive steady state satisfies the equation:

$$\begin{aligned} 0 = (x_5 - S_{\text{tot}}) + \frac{K_{\text{tot}}x_5}{\kappa_7 q(x)}\Big(&\big(\kappa_1\kappa_4\kappa_{11}x_5 + \kappa_1\kappa_9(\kappa_5+\kappa_6+\kappa_{11}) + \kappa_4\kappa_8\kappa_{11}\big)\kappa_7 \\ &+ \big(\kappa_1\kappa_4\kappa_{10}x_5 + \kappa_4\kappa_8(\kappa_2+\kappa_3+\kappa_{10}) + \kappa_1\kappa_9\kappa_{10}\big)\kappa_7 \\ &+ \Big(\kappa_1\kappa_3\kappa_9(\kappa_5+\kappa_{11}) + \kappa_4\kappa_8(\kappa_2\kappa_6+\kappa_3\kappa_{11}) + \kappa_6(\kappa_3+\kappa_{10})(\kappa_1\kappa_9+\kappa_4\kappa_8)\Big)\Big). \end{aligned}$$

By clearing the denominator $\kappa_7 q(x)$, the positive solutions to the above equation agree with the positive solutions to the polynomial given by the numerator. This polynomial is the following polynomial in x_5:

$$
\begin{aligned}
p(x_5) =& \kappa_1\kappa_4\kappa_7(\kappa_{10}+\kappa_{11})x_5^3 \qquad \text{(E.9)}\\
&+ \big(K_{\text{tot}}\kappa_1\kappa_4((\kappa_6+\kappa_7)\kappa_{10}+(\kappa_3+\kappa_7)\kappa_{11}) - S_{\text{tot}}\kappa_1\kappa_4\kappa_7(\kappa_{10}+\kappa_{11})\\
&+ (\kappa_2+\kappa_3)\kappa_4\kappa_7(\kappa_8+\kappa_{11}) + (\kappa_5+\kappa_6)\kappa_1\kappa_7(\kappa_9+\kappa_{10})\\
&+ \kappa_7(\kappa_1\kappa_9+\kappa_4\kappa_8)(\kappa_{10}+\kappa_{11})\big)x_5^2\\
&+ \big((\kappa_1\kappa_9+\kappa_4\kappa_8)(K_{\text{tot}}((\kappa_6+\kappa_7)\kappa_{10}+(\kappa_3+\kappa_7)\kappa_{11}) - S_{\text{tot}}\kappa_7(\kappa_{10}+\kappa_{11}))\\
&+ (\kappa_2+\kappa_3)\kappa_4(K_{\text{tot}}\kappa_8(\kappa_6+\kappa_7) - S_{\text{tot}}\kappa_7(\kappa_8+\kappa_{11}))\\
&+ (\kappa_5+\kappa_6)\kappa_1(K_{\text{tot}}\kappa_9(\kappa_3+\kappa_7) - S_{\text{tot}}\kappa_7(\kappa_9+\kappa_{10}))\\
&+ ((\kappa_2+\kappa_3)\kappa_{11}+(\kappa_5+\kappa_6)\kappa_{10}+(\kappa_2+\kappa_3)(\kappa_5+\kappa_6))\kappa_7(\kappa_8+\kappa_9)\big)x_5\\
&- S_{\text{tot}}\kappa_7(\kappa_8+\kappa_9)((\kappa_2+\kappa_3)(\kappa_5+\kappa_6)+(\kappa_2+\kappa_3)\kappa_{11}+(\kappa_5+\kappa_6)\kappa_{10}).
\end{aligned}
$$

The polynomial $p(x_5)$ has degree 3. Any positive root of this polynomial gives rise to a positive steady state using the expressions (E.4)-(E.8) and, similarly, the value of x_5 for any positive steady state of the system is a root of the polynomial. That is, positive steady states of the network fulfilling the conservation laws (E.1) are in one-to-one correspondence with the positive roots of this polynomial.

We note that this polynomial has at least one positive root since $p(0) < 0$ and $p(+\infty) > 0$. In subsection E.2.4 we show that the reaction rate constants and the total amounts can be chosen such that $p(x_5)$ has indeed three positive roots. Therefore, there exist reaction rate constants and total amounts such that the system has three positive steady states.

The result is first shown by setting the reaction rate constants $\kappa_2 = \kappa_5 = \kappa_9 = \kappa_{10}$ to zero. This corresponds to making some reversible reactions irreversible. Subsequently, we apply a result by Joshi and Shiu [315] to conclude that existence of

three positive steady states can be lifted to the network with all rates being positive.

E.1.4 Necessary conditions for bistability

Following Descartes' rule of signs, a necessary condition for $p(x_5)$ to have 3 positive roots is that the coefficients of the polynomial have alternating signs. Since the leading coefficient is positive and the independent term is negative, a necessary condition is that the coefficient of degree 2 is negative and the coefficient of degree 1 is positive, that is:

$$K_{\text{tot}}\kappa_1\kappa_4[(\kappa_6+\kappa_7)\kappa_{10}+(\kappa_3+\kappa_7)\kappa_{11}]+(\kappa_2+\kappa_3)\kappa_4\kappa_7(\kappa_8+\kappa_{11})$$
$$+(\kappa_5+\kappa_6)\kappa_1\kappa_7(\kappa_9+\kappa_{10})+\kappa_7(\kappa_1\kappa_9+\kappa_4\kappa_8)(\kappa_{10}+\kappa_{11})<S_{\text{tot}}\kappa_1\kappa_4\kappa_7(\kappa_{10}+\kappa_{11})$$

and

$$K_{\text{tot}}[(\kappa_1\kappa_9+\kappa_4\kappa_8)((\kappa_6+\kappa_7)\kappa_{10}+(\kappa_3+\kappa_7)\kappa_{11})(\kappa_2+\kappa_3)\kappa_4\kappa_8(\kappa_6+\kappa_7)$$
$$+(\kappa_5+\kappa_6)\kappa_1\kappa_9(\kappa_3+\kappa_7)]+[(\kappa_2+\kappa_3)\kappa_{11}+(\kappa_5+\kappa_6)\kappa_{10}+(\kappa_2+\kappa_3)(\kappa_5+\kappa_6)]\kappa_7(\kappa_8+\kappa_9)$$
$$>S_{\text{tot}}\kappa_7[(\kappa_1\kappa_9+\kappa_4\kappa_8)(\kappa_{10}+\kappa_{11})+(\kappa_2+\kappa_3)\kappa_4(\kappa_8+\kappa_{11})+(\kappa_5+\kappa_6)\kappa_1(\kappa_9+\kappa_{10})].$$

In contrast to the condition that will be derived in the next subsection, these two conditions involve the total amounts. These conditions can be rewritten as

$$\alpha_1K_{\text{tot}}+\alpha_2<S_{\text{tot}}<\alpha_3K_{\text{tot}}+\alpha_4, \tag{E.10}$$

where:

$$\alpha_1 = \frac{\kappa_1\kappa_4[(\kappa_6+\kappa_7)\kappa_{10}+(\kappa_3+\kappa_7)\kappa_{11}]}{\kappa_1\kappa_4\kappa_7(\kappa_{10}+\kappa_{11})},$$
$$\alpha_2 = \frac{(\kappa_2+\kappa_3)\kappa_4\kappa_7(\kappa_8+\kappa_{11})+(\kappa_5+\kappa_6)\kappa_1\kappa_7(\kappa_9+\kappa_{10})+\kappa_7(\kappa_1\kappa_9+\kappa_4\kappa_8)(\kappa_{10}+\kappa_{11})}{\kappa_1\kappa_4\kappa_7(\kappa_{10}+\kappa_{11})}$$
$$\alpha_3 = \frac{[(\kappa_1\kappa_9+\kappa_4\kappa_8)((\kappa_6+\kappa_7)\kappa_{10}+(\kappa_3+\kappa_7)\kappa_{11})(\kappa_2+\kappa_3)\kappa_4\kappa_8(\kappa_6+\kappa_7)+(\kappa_5+\kappa_6)\kappa_1\kappa_9(\kappa_3+\kappa_7)]}{\kappa_7[(\kappa_1\kappa_9+\kappa_4\kappa_8)(\kappa_{10}+\kappa_{11})+(\kappa_2+\kappa_3)\kappa_4(\kappa_8+\kappa_{11})+(\kappa_5+\kappa_6)\kappa_1(\kappa_9+\kappa_{10})]}$$
$$\alpha_4 = \frac{[(\kappa_2+\kappa_3)\kappa_{11}+(\kappa_5+\kappa_6)\kappa_{10}+(\kappa_2+\kappa_3)(\kappa_5+\kappa_6)]\kappa_7(\kappa_8+\kappa_9)}{\kappa_7[(\kappa_1\kappa_9+\kappa_4\kappa_8)(\kappa_{10}+\kappa_{11})+(\kappa_2+\kappa_3)\kappa_4(\kappa_8+\kappa_{11})+(\kappa_5+\kappa_6)\kappa_1(\kappa_9+\kappa_{10})]}.$$

For each fixed value of K_{tot}, the solution to the system of inequalities (E.10) is either empty or an interval. Since $\alpha_i > 0$ for $i = 1, 2, 3, 4$, $\alpha_1 K_{\text{tot}} + \alpha_2$ and $\alpha_3 K_{\text{tot}} + \alpha_4$ are increasing straight lines in K_{tot} with positive intercept. Therefore the region is described by a sector intersected with the positive orthant of $\mathbb{R}^2$. If the two lines are parallel then the valid region is the region between the two lines intersected with the positive orthant.

An example of how such a sector might look like is given in Example 2 below.

E.1.5 Necessary and sufficient conditions for multistationarity

Conditions involving only reaction rate constants

In order to find sufficient conditions for multistationarity, we apply the strategy introduced in [316]. In that paper, sufficient conditions for multistationarity, based on the reaction rate constants only, were found for a two-site phosphorylation cycle in which both the kinase and the phosphatase follow a sequential and distributive mechanism. The strategy is based on Brouwer Degree Theory.

The steps of the procedure are as follows:

(1) Compute the determinant of the Jacobian matrix associated with the function

given by the two conservation laws and the expressions for $\dot{x_2}, \dot{x_3}, \dot{x_4}$ and $\dot{x_6}$:

$$\begin{aligned} f(x) =& (-\kappa_4 x_2 x_5 + (\kappa_5 + \kappa_6)x_4 + \kappa_8 x_1 - \kappa_9 x_2, \kappa_1 x_1 x_5 - (\kappa_2 + \kappa_3)x_3 - \kappa_{10} x_3 + \kappa_{11} x_4, \\ & \kappa_4 x_2 x_5 - (\kappa_5 + \kappa_6)x_4 + \kappa_{10} x_3 - \kappa_{11} x_4, \kappa_3 x_3 + \kappa_6 x_4 - \kappa_7 x_6, \\ & x_1 + x_2 + x_3 + x_4, x_3 + x_4 + x_5 + x_6). \end{aligned}$$

Let $\det(J_\kappa(x))$ denote this determinant.

(2) Find a parameterisation of the positive steady states in terms of x_1 and x_5. That is, consider the steady state equations $\dot{x}_2 = \dot{x}_3 = \dot{x}_4 = \dot{x}_6 = 0$ and solve them for x_2, x_3, x_4, x_6 in terms of x_1, x_5.

(3) Substitute the values of x_2, x_3, x_4, x_6 found in the previous step into the determinant of the Jacobian. The resulting expression is a quotient of polynomials in x_1, x_5, where all coefficients of the polynomial in the denominator are positive.

Let $b_\kappa(x_1, x_5)$ be the numerator of $\det(J_\kappa(x))$ after the substitution in step (3). Brouwer Degree Theory gives us that multistationarity occurs if and only if the polynomial $b_\kappa(x_1, x_5)$ is positive for some positive values of x_1, x_5 [316].

Viewed as a polynomial in x_1, x_5, all coefficients of $b_\kappa(x_1, x_5)$ are polynomials in κ. All coefficients have negative sign, independently of the values of κ_i, except for one coefficient which is:

$$\begin{aligned} \alpha(\kappa) = & (\kappa_3 - \kappa_6)\left(-\kappa_4 (\kappa_2 + \kappa_3)\kappa_8\kappa_{11} + \kappa_1 (\kappa_5 + \kappa_6)\kappa_9\kappa_{10}\right) \\ & - \left((\kappa_6 + \kappa_7)\kappa_{10} + (\kappa_3 + \kappa_7)\kappa_{11}\right)\left(\kappa_1 (\kappa_5 + \kappa_6)\kappa_{10} + \kappa_4\kappa_{11}(\kappa_2 + \kappa_3)\right). \end{aligned} \tag{E.11}$$

Clearly, if this coefficient is negative, then all coefficients are negative and multistationarity cannot occur. Assume now that $\alpha(\kappa)$ is positive. We want to show that in this case the polynomial $b_\kappa(x_1, x_5)$ is positive for some values of x_1, x_5. The coefficient $\alpha(\kappa)$ is the coefficient of the monomial $x_1 x_5^2$. The other monomials of the

polynomial are $1, x_1, x_5, x_1x_5, x_5^2, x_5^3$. If we can choose x_1, x_5 such that the monomial $x_1x_5^2$ dominates the other monomials, then $b_\kappa(x_1, x_5)$ becomes positive. For this, let $x_5 = T$ and $x_1 = T^2$. Then $b_\kappa(T^2, T)$ is a polynomial in T of degree 4 with leading positive coefficient $\alpha(\kappa)$. By letting T be arbitrarily large, $b_\kappa(T^2, T)$ becomes eventually positive.

This shows that $b_\kappa(x_1, x_5)$ is positive for some values of x_1, x_5, if and only if the coefficient $\alpha(\kappa)$ is positive and hence

multistationarity occurs if and only if $\alpha(\kappa)$ is positive.

After rearranging the terms of the coefficient $\alpha(\kappa)$, we obtain the following necessary and sufficient condition for multistationarity:

$$(\kappa_3 - \kappa_6)(\eta_r\kappa_9\kappa_{10} - \eta_t\kappa_8\kappa_{11}) > ((\kappa_6 + \kappa_7)\kappa_{10} + (\kappa_3 + \kappa_7)\kappa_{11})(\eta_r\kappa_{10} + \eta_t\kappa_{11}) \tag{E.12}$$

where

$$\eta_r = \frac{\kappa_1}{\kappa_2 + \kappa_3} \qquad \eta_t = \frac{\kappa_4}{\kappa_5 + \kappa_6}$$

are the inverses of the Michaelis-Menten constants of the kinases K_r and K_t respectively.

By inspecting the inequality, we can find some necessary conditions for multistationarity. For example:

- Either κ_9 and κ_{10} need to be nonzero or κ_8 and κ_{11} need to be nonzero. That is, allosteric changes must occur both for the kinase and the kinase-substrate complexes.

- Since the left-hand side of the inequality must be positive for the inequality to hold, we deduce that one of the following conditions is necessary for multistationarity:

  (a) $\kappa_3 > \kappa_6$ and $\eta_r\kappa_9\kappa_{10} > \eta_t\kappa_8\kappa_{11}$.

(b) $\kappa_3 < \kappa_6$ and $\eta_r\kappa_9\kappa_{10} < \eta_t\kappa_8\kappa_{11}$.

Conditions involving reaction rate constants and total amounts

Here we provide necessary and sufficient conditions on *all* parameters (reaction rate constants and total amounts) of the system for multistationarity to occur. To obtain the conditions, we apply Sturm's Theorem:

Theorem 1 (Sturm)**.** *Let $p(x)$ be a real polynomial. Define recursively the Sturm sequence by*

$$p_0(x) = p(x),\ p_1(x) = p'(x), \quad \textit{and} \quad p_{i+1}(x) = -\mathrm{rem}(p_{i-1}, p_i),$$

for $i \geq 1$, where $\mathrm{rem}(p_{i-1}, p_i)$ denotes the reminder of p_{i-1} divided by p_i. The sequence stops when $p_{i+1} = 0$. Let p_m be the last nonzero polynomial.

For $c \in \mathbb{R}$, let $\sigma(c)$ be the number of sign changes in the sequence $p_0(c), \ldots, p_m(c)$. Let $a < b$ and assume that neither a nor b are multiple roots of $p(x)$. Then $\sigma(a) - \sigma(b)$ is the number of distinct roots of $p(x)$ in the interval $(a, b]$.

We are interested in the positive roots of the polynomial $p(x) = p(x_5)$ in (E.9). That is we should take $a = 0$ and b so large that all positive roots are in $(a, b] = (0, b]$. If b is large then the signs of $p_0(b), \ldots, p_m(b)$ are determined by the leading coefficients of the polynomials $p_0, \ldots, p_m$. Because b is an arbitrarily large number, we write $b = +\infty$ and the sequence is written as $p_0(+\infty), \ldots, p_m(+\infty)$. Observe that $a = 0$ is not a root of $p(x)$ and hence the hypothesis of Sturm's theorem applies.

According to the theorem, $\sigma(0) - \sigma(+\infty)$ equals the number of distinct positive roots of $p(x)$. In our case, we have $m = 3$, that is, $p_4(x) = 0$ (see below), and hence $0 \leq \sigma(c) \leq 3$ for $c \geq 0$. Therefore, the number of distinct roots will be 3, that is, there will be three positive steady states, if and only if $\sigma(0) = 3$ and $\sigma(+\infty) = 0$.

We computed in Maple the Sturm sequence $p_0(x), \ldots, p_3(x)$ ($p_4(x) = 0$). For a generic polynomial of degree 3, $p_0(x) = a_0x^3 + a_1x^2 + a_2x + a_3$, the sequence is:

$$
\begin{aligned}
p_0(x) &= a_0x^3 + a_1x^2 + a_2x + a_3 \\
p_1(x) &= 3a_0x^2 + 2a_1x + a_2 \\
p_2(x) &= -\frac{6a_0a_2x - 2a_1^2x + 9a_0a_3 - a_1a_2}{9a_0} \\
p_3(x) &= -\frac{9a_0(27a_0^2a_3^2 - 18\,a_0a_1a_2a_3 + 4a_0a_2^3 + 4a_1^3a_3 - a_1^2a_2^2)}{4(3a_0a_2 - a_1^2)^2}.
\end{aligned}
$$

In our case, the coefficients are:

$$
\begin{aligned}
a_0 =& \kappa_1\kappa_4\kappa_7(\kappa_{10} + \kappa_{11}) \\
a_1 =& \big(K_{\text{tot}}\kappa_1\kappa_4((\kappa_6 + \kappa_7)\kappa_{10} + (\kappa_3 + \kappa_7)\kappa_{11}) - S_{\text{tot}}\kappa_1\kappa_4\kappa_7(\kappa_{10} + \kappa_{11}) \\
&+ (\kappa_2 + \kappa_3)\kappa_4\kappa_7(\kappa_8 + \kappa_{11}) + (\kappa_5 + \kappa_6)\kappa_1\kappa_7(\kappa_9 + \kappa_{10}) + \kappa_7(\kappa_1\kappa_9 + \kappa_4\kappa_8)(\kappa_{10} + \kappa_{11})\big) \\
a_2 =& \big((\kappa_1\kappa_9 + \kappa_4\kappa_8)(K_{\text{tot}}((\kappa_6 + \kappa_7)\kappa_{10} + (\kappa_3 + \kappa_7)\kappa_{11}) - S_{\text{tot}}\kappa_7(\kappa_{10} + \kappa_{11})) \\
&+ (\kappa_2 + \kappa_3)\kappa_4(K_{\text{tot}}\kappa_8(\kappa_6 + \kappa_7) - S_{\text{tot}}\kappa_7(\kappa_8 + \kappa_{11})) \\
&+ (\kappa_5 + \kappa_6)\kappa_1(K_{\text{tot}}\kappa_9(\kappa_3 + \kappa_7) - S_{\text{tot}}\kappa_7(\kappa_9 + \kappa_{10})) \\
&+ ((\kappa_2 + \kappa_3)\kappa_{11} + (\kappa_5 + \kappa_6)\kappa_{10} + (\kappa_2 + \kappa_3)(\kappa_5 + \kappa_6))\kappa_7(\kappa_8 + \kappa_9)\big) \\
a_3 =& - S_{\text{tot}}\kappa_7(\kappa_8 + \kappa_9)((\kappa_2 + \kappa_3)(\kappa_5 + \kappa_6) + (\kappa_2 + \kappa_3)\kappa_{11} + (\kappa_5 + \kappa_6)\kappa_{10}).
\end{aligned}
\tag{E.13}
$$

Since $p_0(0) = a_3 < 0$, for $\sigma(0) = 3$ we need

$$p_1(0) > 0,\ p_2(0) < 0 \quad \text{and} \quad p_3(0) > 0.$$

On the other hand,

$$p_0(+\infty) = a_0 > 0 \quad \text{and} \quad p_1(+\infty) = 3a_0 > 0.$$

Therefore, for $\sigma(+\infty) = 0$ we require $p_2(+\infty), p_3(+\infty) > 0$.

The polynomial $p_3(x)$ has degree zero, and hence $p_3(0) = p_3(+\infty)$. Therefore, we are left with 4 conditions on the parameters that fully characterise the region of the parameter space with three steady states, namely

$$p_1(0) > 0, \quad p_3(0) > 0, \quad p_2(+\infty) > 0 \quad \text{and} \quad p_2(0) < 0.$$

Using that $a_0 > 0$ and $a_3 < 0$, these conditions simplify to the following conditions, where $a_0, \ldots, a_3$ need to be substituted by their respective expressions in (E.13):

$$\begin{aligned} a_2 > 0 \qquad & (p_1(0) > 0) \\ -9a_0a_3 + a_1a_2 < 0 \qquad & (p_2(0) < 0) \\ 27a_0^2a_3^2 - 18a_0a_1a_2a_3 + 4a_0a_2^3 + 4a_1^3a_3 - a_1^2a_2^2 < 0 \qquad & (p_3(0) > 0) \\ -6a_0a_2 + 2a_1^2 > 0 \qquad & (p_2(+\infty) > 0). \end{aligned} \tag{E.14}$$

That is, the system has three positive steady states if and only if the 4 inequalities above are satisfied using (E.13).

E.1.6 Necessary and sufficient conditions in practice

In order to find explicit values of the parameters such that the system exhibits multistationarity, the procedure is the following:

1. First, use the necessary and sufficient condition given in (E.12) to find appropriate values for the reaction rate constants.

2. Second, substitute these values of the reaction rate constants into (E.14). This yields a system of 4 inequalities in K_{tot} and S_{tot}. The positive values of $(K_{\text{tot}}, S_{\text{tot}})$ fulfilling the inequalities correspond to parameter sets for which there are three positive steady states. By the results above, there are always values of $(K_{\text{tot}}, S_{\text{tot}})$ for which this is the case.

After the first step we might use the necessary conditions for multistationarity from subsection E.1.4, that is, using the inequalities in (E.10) instead of the conditions in (E.14). This gives (simpler) regions of the parameter space of total amounts containing all pairs $(K_{\text{tot}}, S_{\text{tot}})$ for which there is multistationarity. Remember though that not all pairs $(K_{\text{tot}}, S_{\text{tot}})$ satisfying the inequalities yield multistationarity as the conditions are only necessary and not sufficient.

Example 2. Consider the set of parameters

$$\begin{aligned} &\kappa_1 = 5, \quad \kappa_2 = 0.1, \quad \kappa_3 = 1, \quad \kappa_4 = 2, \quad \kappa_5 = 0.1, \quad \kappa_6 = 2, \\ &\kappa_7 = 0.01, \quad \kappa_8 = 0.8, \quad \kappa_9 = 0.1, \quad \kappa_{10} = 0.01, \quad \kappa_{11} = 0.1 \end{aligned} \tag{E.15}$$

for which (E.12) is satisfied.

The system of inequalities (E.10) is

$$110.1K_{\text{tot}} + 3.06 < S_{\text{tot}} < 144.16K_{\text{tot}} + 0.65,$$

and the pairs $(K_{\text{tot}}, S_{\text{tot}})$ fulfilling the inequalities are highlighted in light blue in Figure 1(a,b).

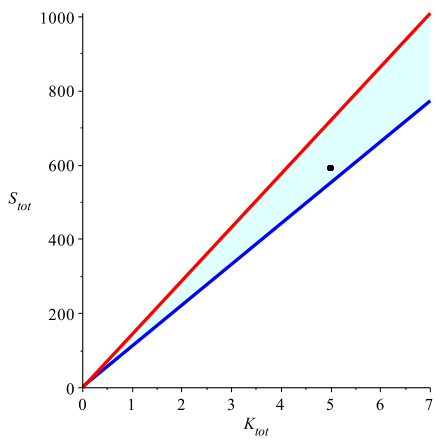


Figure 1a.

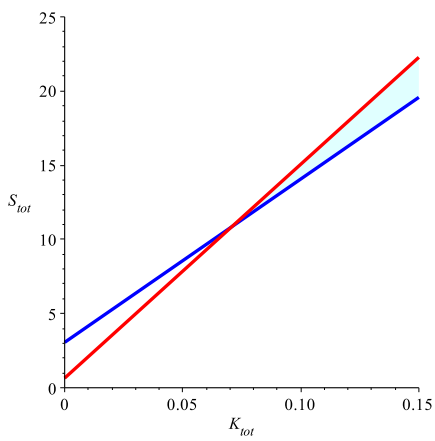


Figure 1b.

The plot in Figure 1b illustrates that for very small values of K_{tot} there

is no value of S_{tot} for which the inequalities are satisfied. The dot in Figure 1a corresponds to $S_{\text{tot}} = 591$ and $K_{\text{tot}} = 5$, for which there is multistationarity.

We use Sturm's conditions (E.14) to find a precise characterization of the pair of total amounts for which multistationarity occurs. The conditions translate into the following set of inequalities:

$$
\begin{aligned}
0 <& 4.852K_{\text{tot}} - 0.03366S_{\text{tot}} + 0.02197 \\
0 >& 5.876K_{\text{tot}}^2 - 0.09414K_{\text{tot}}S_{\text{tot}} + 0.0003703S_{\text{tot}}^2 \\
&+ 0.0008003S_{\text{tot}} + 0.1899K_{\text{tot}} + 0.0007395 \\
0 >& -2.013 \cdot 10^{-8}S_{\text{tot}}^4 - 34.531K_{\text{tot}}^4 + 0.9503S_{\text{tot}}K_{\text{tot}}^3 - 0.008961S_{\text{tot}}^2K_{\text{tot}}^2 \\
&+ 0.00003108K_{\text{tot}}S_{\text{tot}}^3 - 1.232 \cdot 10^{-7}S_{\text{tot}}^3 + 2.795K_{\text{tot}}^3 - 0.04015S_{\text{tot}}K_{\text{tot}}^2 \\
&+ 0.0001532K_{\text{tot}}S_{\text{tot}}^2 + 0.02351K_{\text{tot}}^2 + 1.786 \cdot 10^{-4}K_{\text{tot}}S_{\text{tot}} - 2.689 \cdot 10^{-7}S_{\text{tot}}^2 \\
&+ 2.823 \cdot 10^{-5}K_{\text{tot}} - 2.460 \cdot 10^{-7}S_{\text{tot}} - 8.029 \cdot 10^{-8} \\
0 <& 2.933K_{\text{tot}}^2 - 0.05328K_{\text{tot}}S_{\text{tot}} + 0.000242S_{\text{tot}}^2 - 0.1572K_{\text{tot}} \\
&+ 0.0007405S_{\text{tot}} + 0.0008160.
\end{aligned}
$$

One set of total amounts fulfilling the above system of inequalities is $S_{\text{tot}} = 591$ and $K_{\text{tot}} = 5$ (the point plotted in the figure above).

In the following two plots (Figure 2(a,b)) the yellow region is the region of common solutions to the first, second and fourth inequalities, and the blue region the solution to the third inequality. The intersection of the two regions is the solution set to the four inequalities. It is the small blue region inside the yellow region. Figure 2b is a magnification of Figure 2a, in which also the point $(5, 591)$ is shown.

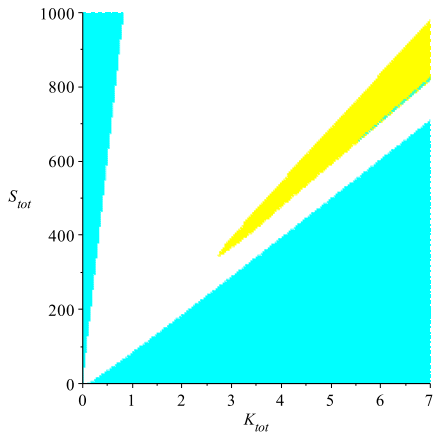


Figure 2a.

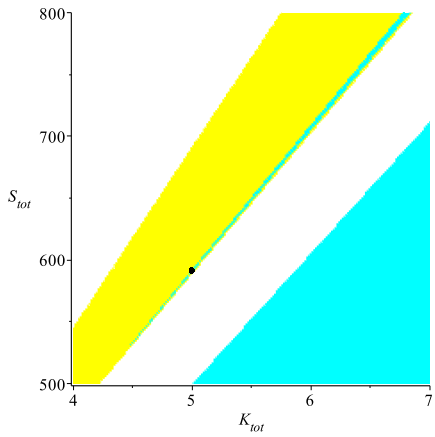


Figure 2b.

Note that the region for which multistationarity exists is much smaller than the region given in Figure 1(a,b). In practice, it is not straightforward to solve the Sturm's inequalies for S_{tot} and K_{tot}.

E.1.7 Describing the steady states

In this section we describe the intersection of the solution set to the steady state equations, with the linear space defined by $x_1 + x_2 + x_3 + x_4 = K_{\text{tot}}$, using the parametrization (E.4)-(E.8). To illustrate the results of this section we choose a set of parameters for which multistationarity occurs:

$$\begin{aligned} &\kappa_1 = 5, \quad \kappa_2 = 0.1, \quad \kappa_3 = 1, \quad \kappa_4 = 2, \quad \kappa_5 = 0.1, \quad \kappa_6 = 2, \\ &\kappa_7 = 0.01, \quad \kappa_8 = 0.8, \quad \kappa_9 = 0.1, \quad \kappa_{10} = 0.01, \quad \kappa_{11} = 0.1. \end{aligned} \tag{E.16}$$

The concentration $[S_p]$ as a function of $[S]$

First, we discuss how the concentration of S_p, x_6, changes according to the concentration of S, x_5, at steady state using the parametrization (E.8).

Consider the expression (E.8) as a function of x_5:

$$x_6 = \varphi_6(x_5).$$

We are interested in the steady states for a fixed value of $S_{\text{tot}} = x_3 + x_4 + x_5 + x_6$. Therefore, we also consider the rational function of x_6 obtained by substitution of (E.6) ($x_3 = \varphi_3(x_5)$) and (E.7) ($x_4 = \varphi_4(x_5)$) into the conservation law for S_{tot}: $x_6 = S_{\text{tot}} - \varphi_3(x_5) - \varphi_4(x_5) - \varphi_6(x_5)$. This expression is a rational function in x_5 whose numerator has degree three and the denominator has degree two. We plot the two functions using the parameters in (E.16), $K_{\text{tot}} = 5$ and $S_{\text{tot}} = 591$.

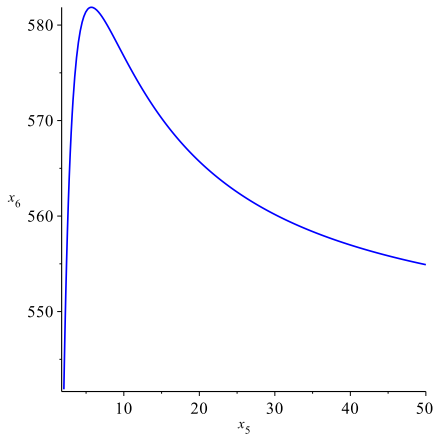


Figure 3a.

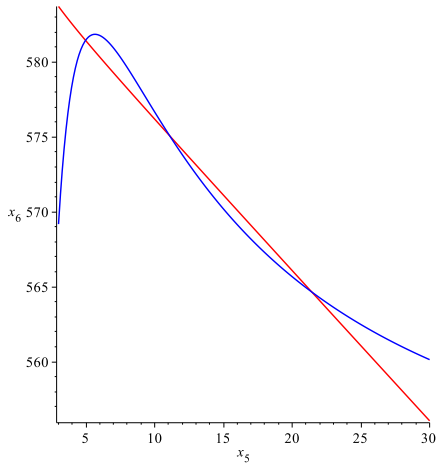


Figure 3b.

Figure 3a shows the graph of the function $\varphi_6(x_5)$. Figure 3b shows the graph of the function $\varphi_6(x_5)$ together with the function $x_6 = S_{\text{tot}} - \varphi_3(x_5) - \varphi_4(x_5) - \varphi_6(x_5)$. The intersection points of the two graphs in Figure 3b are the pairs (x_5, x_6) for the three steady states in this stoichiometric compatibility class.

We observe that $\varphi_6(x_5)$ increases for small values of x_5, until it reaches a maximum and then decreases towards a limit value for large x_5,

$$\lim_{x_5 \to +\infty} \varphi_6(x_5) = \frac{K_{\text{tot}} \left(\kappa_3 \kappa_{11} + \kappa_6 \kappa_{10}\right)}{\kappa_7 \left(\kappa_{10} + \kappa_{11}\right)}. \tag{E.17}$$

Next we show that this shape is necessary for multistationarity. The deriva-

tive of $\varphi_6(x_5)$ with respect to x_5 is:

$$\varphi_6'(x_5) = \frac{K_{\text{tot}}\ \big(\lambda_1(\kappa)x_5^2 + \lambda_2(\kappa)x_5 + \lambda_3(\kappa)\big)}{\kappa_7^2 q(x_5)^2},$$

where $\lambda_2(\kappa), \lambda_3(\kappa)$ are positive polynomials in the reaction rate constants, and

$$\lambda_1(\kappa) = -\alpha(\kappa) - \overline{\lambda}_1(\kappa),$$

with $\alpha(\kappa)$ as in (E.11) and $\overline{\lambda}_1(\kappa)$ a positive polynomial in the reaction rate constants.

By the results of subsection E.1.5, multistationarity occurs if and only if $\alpha(\kappa) > 0$. In this case $\lambda_1(\kappa)$ is negative. Therefore the numerator of $\varphi_6'(x_5)$ is a second degree polynomial with negative leading coefficient and the rest of the coefficients are positive. Since $\lambda_1(\kappa)$ is the leading coefficient of this polynomial, this implies that the polynomial is negative for large values of x_5. For small values of x_5, the polynomial is positive. It follows that there exists a unique positive value of x_5, $\widehat{x}_5$ for which $\varphi_6'(\widehat{x}_5) = 0$. The derivative is positive for $x_5 < \widehat{x}_5$ and negative for $x_5 > \widehat{x}_5$.

As a consequence, the function $\varphi_6(x_5)$ has the shape as Figure 3a. That is, $\varphi_6(x_5)$ increases up to a value $\widehat{x}_5$ and decreases towards (E.17) for large x_5.

If $\lambda_1(\kappa)$ is positive, then $\alpha(\kappa)$ is negative and hence multistationarity cannot occur. In that case, $\varphi_6'(x_5) > 0$ for all $x_5 > 0$ and hence $\varphi_6(x_5)$ is an increasing function that approaches the limit (E.17) from below. Note that we cannot express x_5 as a function of x_6 because the function is not injective when multistationarity occurs. Therefore, we cannot use x_6 to parameterize the set of steady states.

The concentrations $[K_r]$ and $[K_t]$ as functions of $[S]$

We consider the rational functions $x_1 = \varphi_1(x_5)$ in (E.4) and $x_2 = \varphi_2(x_5)$ in (E.5) using the parameters in (E.16) and $K_{\text{tot}} = 5$. The plot of these functions are shown in Figure 4(a,b).

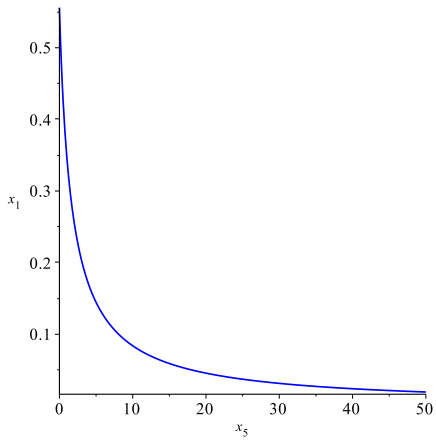


Figure 4a.

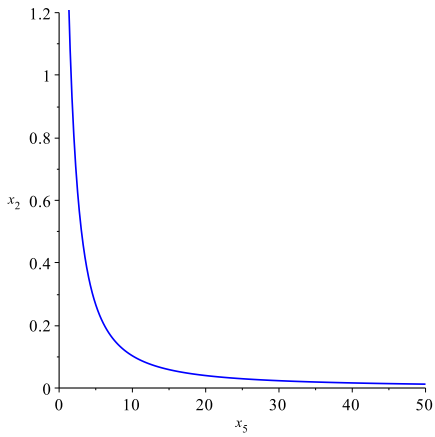


Figure 4b.

From (E.4) and (E.5), it follows that the numerator of both $\varphi_1(x_5)$ and $\varphi_2(x_5)$ has degree one and the denominator degree 2. Hence, $\varphi_1(x_5)$ and $\varphi_2(x_5)$ tend to zero as x_5 tends to infinity. The derivatives of $\varphi_1(x_5)$ and $\varphi_2(x_5)$ are of the form

$$\begin{aligned}\varphi_1'(x_5) =& \frac{K_{\text{tot}}\ \left(a_1(\kappa)x_5^2 + a_2(\kappa)x_5 + a_3(\kappa)\right)}{q(x_5)^2}\\ \varphi_2'(x_5) =& \frac{K_{\text{tot}}\ \left(b_1(\kappa)x_5^2 + b_2(\kappa)x_5 + b_3(\kappa)\right)}{q(x_5)^2}\end{aligned}$$

where $a_1(\kappa), a_2(\kappa), b_1(\kappa), b_2(\kappa)$ are negative polynomials in the reaction rate constants, and

$$\begin{aligned}a_3(\kappa) &= -\overline{a}_3(\kappa)\kappa_4\kappa_8(\kappa_3+\kappa_2)(\kappa_9-\kappa_{11}) + \kappa_9((\kappa_1\kappa_9+\kappa_4\kappa_8)(\kappa_{10}+\kappa_{11})\\ &\quad + \kappa_1(\kappa_5+\kappa_6)(\kappa_{10}+\kappa_9)),\\ b_3(\kappa) &= -\overline{b}_3(\kappa)(\kappa_1\kappa_9(\kappa_8-\kappa_{10})(\kappa_5+\kappa_6) + \kappa_8(\kappa_4\left(\kappa_8+\kappa_{11}\right)\left(\kappa_2+\kappa_3\right)\\ &\quad + \left(\kappa_{10}+\kappa_{11}\right)\left(\kappa_1\kappa_9+\kappa_4\kappa_8\right))),\end{aligned}$$

with $\overline{a}_3(\kappa)$ and $\overline{b}_3(\kappa)$ being positive polynomials in the reaction rate constants.

For small values of x_5, the functions $\varphi_1(x_5)$ and $\varphi_2(x_5)$ can be increasing or

decreasing depending on the values of the reaction rate constants (that is, the sign of $a_3(\kappa)$ and $b_3(\kappa)$, respectively). Since $a_1(\kappa), b_1(\kappa) < 0$, for large values of x_5, the two functions decrease and tend to zero as x_5 tends to infinity.

The concentrations $[K_rS]$ and $[K_tS]$ as functions of $[S]$

We consider the rational functions $x_3 = \varphi_3(x_5)$ in (E.6) and $x_4 = \varphi_4(x_5)$ in (E.7) using the parameters in (E.16) and $K_{\text{tot}} = 5$. The plot of these functions are in Figure 5(a,b).

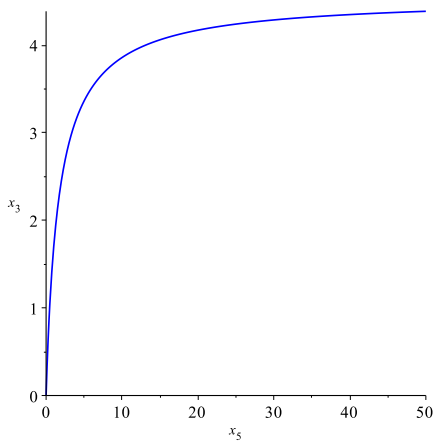


Figure 5a.

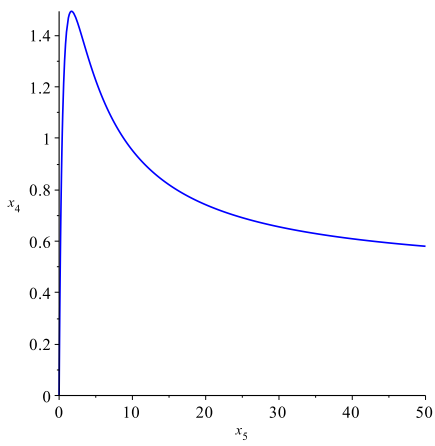


Figure 5b.

From (E.6) and (E.7), it follows that the numerator and the denominator of both $\varphi_3(x_5)$ and $\varphi_4(x_5)$ have degree two. Hence:

$$\lim_{x_5 \to +\infty} \varphi_3(x_5) = \frac{K_{\text{tot}}\kappa_{11}}{\kappa_{10} + \kappa_{11}}, \qquad \lim_{x_5 \to +\infty} \varphi_4(x_5) = \frac{K_{\text{tot}}\kappa_{10}}{\kappa_{10} + \kappa_{11}}. \tag{E.18}$$

The derivatives of $\varphi_3(x_5)$ and $\varphi_4(x_5)$ are

$$\begin{aligned}
\varphi_3'(x_5) =& \frac{K_{\text{tot}}\ \big(c_1(\kappa)x_5^2 + c_2(\kappa)x_5 + c_3(\kappa)\big)}{q(x_5)^2} \\
\varphi_4'(x_5) =& \frac{K_{\text{tot}}\ \big(d_1(\kappa)x_5^2 + d_2(\kappa)x_5 + d_3(\kappa)\big)}{q(x_5)^2}
\end{aligned}$$

where $c_2(\kappa), c_3(\kappa)$ are positive polynomials in the reaction rate constants, and

$$c_1(\kappa) = \overline{c}_1(\kappa)(\kappa_1\kappa_{10}\,(\kappa_{11} - \kappa_9)\,(\kappa_5 + \kappa_6) + \kappa_4\kappa_{11}\,(\kappa_8 + \kappa_{11})\,(\kappa_2 + \kappa_3)),$$
$$d_1(\kappa) = \overline{d}_1(\kappa)(\kappa_4\kappa_{11}\,(\kappa_{10} - \kappa_8)\,(\kappa_2 + \kappa_3) + \kappa_1\kappa_{10}\,(\kappa_{10} + \kappa_9)\,(\kappa_5 + \kappa_6)),$$

with $\overline{c}_1(\kappa)$ and $\overline{d}_1(\kappa)$ positive polynomials in the reaction rate constants. The only coefficients in the numerators of $\varphi_3(x_5)$ and $\varphi_4(x_5)$ that have undetermined sign are thus $c_1(\kappa)$ and $d_1(\kappa)$. It follows that for small values of x_5, both derivatives take positive values and the functions are increasing. For large x_5, the derivatives are positive or negative, depending on the signs of $c_1(\kappa)$ and $d_1(\kappa)$. Thus each of the functions $\varphi_3(x_5)$ and $\varphi_4(x_5)$ can either be increasing towards the limit (E.18) (as in Figure 5a) or be increasing towards a maximum value and then be decreasing towards the limit in (E.18) (as in Figure 5b).

These results show that for a steady state with a large concentration of S, the concentrations of K_rS and K_tS are close to a limit value, and the concentrations of K_r and K_t are close to zero. This confirms mathematically that saturation occurs: large amounts of substrate imply that the kinase is essentially only in bound form.

E.1.8 Bifurcation plots with K_{tot} and S_{tot}

In this subsection we investigate how the number of steady states depends on the total amounts of kinase and substrate.

At steady state, $p(x_5) = 0$. The polynomial $p(x_5)$ is linear in K_{tot} and in S_{tot}. Hence, we can use the equation $p(x_5) = 0$ to isolate K_{tot} (resp. S_{tot}) and get an expression of K_{tot} (resp. S_{tot}) as a function of x_5, the reaction rate constants and S_{tot} (resp. K_{tot}).

In this way we get two functions

$$S_{\text{tot}} = \psi_S(x_5), \qquad K_{\text{tot}} = \psi_K(x_5),$$

which hold at steady state and which we analyse below.

Changing S_{tot}

We fix first the total amount of kinase K_{tot} and analyse $\psi_S(x_5)$. The function $\psi_S(x_5)$ is a rational function whose numerator has degree three and the denominator has degree two. The coefficients of these polynomials are positive polynomials in the reaction rate constants and K_{tot}. Hence, $\psi_S(x_5)$ tends to infinity as x_5 goes to infinity.

Fixing $K_{\text{tot}} = 5$ and the values of the parameters as in (E.16), we plot the value of S_{tot} against x_5 at steady state. Figure 6b is the graph of $\psi_S(x_5)$, while Figure 6a is obtained by interchanging the axes of the plot in Figure 6b.

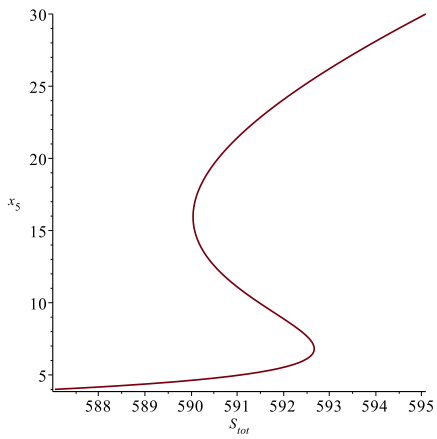


Figure 6a.

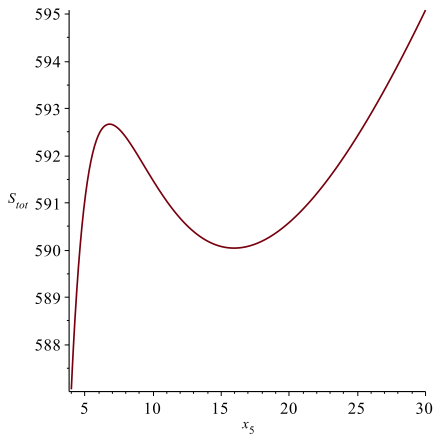


Figure 6b.

From Figure 6a, we conclude that there is a range of S_{tot} values for which multistationarity occurs: each value of S_{tot} corresponds to 3 values of x_5, which in turn give rise to three positive steady states. For low values of S_{tot} the concentration of x_5 is low and for higher values of S_{tot} the concentration of x_5 is high.

To understand Figure 6a, we study the function $\psi_S(x_5)$ in Figure 6b, since the bifurcation plot is simply obtained by interchanging the axes. We do this because we do not have an analytical expression of the type $x_5 = \Phi(S_{\text{tot}})$.

We do a similar plot of S_{tot} against x_6 (both are functions of x_5).

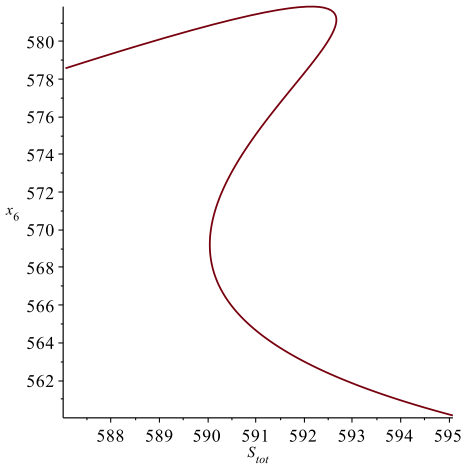


Figure 7a.

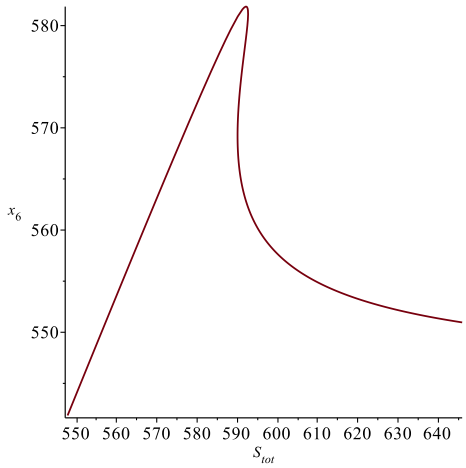


Figure 7b.

In Figure 7a, the values of S_{tot} for which multistationarity occurs can be identified. In Figure 7b, we used a larger range of x_5 values in the plot. We find that x_6 increases with S_{tot} towards a maximum value. Then, there is a transition phase where multistationarity occurs and finally x_6 decreases and there is one steady state.

When multistationarity occurs, the shape of S_{tot} as a function of x_5 is always as illustrated above in Figure 6b. We show this below. The same cannot be done for S_{tot} as a function of x_6, because we cannot get an expression of x_5 as a function of x_6.

The derivative of $\psi_S(x_5)$ with respect to x_5 is the rational function

$$\psi_S'(x_5) = \frac{\mu_1(\kappa)x_5^4 + \mu_2(\kappa)x_5^3 + \mu_3(\kappa, K_{\text{tot}})x_5^2 + \mu_4(\kappa, K_{\text{tot}})x_5 + \mu_5(\kappa, K_{\text{tot}})}{q_2(\kappa, x_5)},$$

where $\mu_1(\kappa), \mu_2(\kappa)$ are positive polynomials in the reaction rate constants, $\mu_4(\kappa, K_{\text{tot}})$, $\mu_5(\kappa, K_{\text{tot}})$ are positive polynomials in the reaction rate constants and K_{tot} (depending linearly on K_{tot}), $q_2(\kappa, x_5)$ is a degree 4 polynomial in x_5 whose coefficients are positive polynomials in the reaction rate constants and

$$\mu_3(\kappa, K_{\text{tot}}) = \mu_3'(\kappa) - K_{\text{tot}}\alpha(\kappa),$$

with $\mu_3'(\kappa)$ a positive polynomial in the reaction rate constants and $\alpha(\kappa)$ as in (E.11).

It follows that $\psi_S'(x_5)$ is positive for small and large values of x_5, and hence for these two cases, the function $\psi_S(x_5)$ is increasing. If multistationarity occurs, then there must be values of x_5 for which the corresponding values of $S_{\text{tot}} = \psi_S(x_5)$ agree. As a consequence $\psi_S(x_5)$ must decrease in some interval (it cannot be an increasing function). This can only occur if $\mu_3(\kappa, K_{\text{tot}})$ is negative.

Note that the sequence of coefficients of the polynomial in the numerator of $\psi_S'(x_5)$ has at most two changes of sign (which occur when $\mu_3(\kappa, K_{\text{tot}}) < 0$). Descartes rule of signs tells us that $\psi_S'(x_5) = 0$ has at most two solutions. Combined with the discussion on the increasing/decreasing behavior of $\psi_S(x_5)$, we deduce that there is exactly one local maximum and one local minimum when the system has three steady states. We conclude that the graph of $\psi_S(x_5)$ must be as in Figure 6b when multistationarity occurs, that is, it has an S-shape.

In fact, since $\mu_1(\kappa), \mu_2(\kappa)$ do not depend on K_{tot} and $\mu_3(\kappa, K_{\text{tot}})$, $\mu_4(\kappa, K_{\text{tot}})$, $\mu_5(\kappa, K_{\text{tot}})$ depend linearly on K_{tot}, we deduce that if $\alpha(\kappa) > 0$ there are always values of K_{tot} such that $\psi_S'(x_5)$ is negative for some values of x_5 (for K_{tot} large enough such that $-K_{\text{tot}}\alpha(\kappa)x_5^2$ dominates). Therefore, we see again that $\alpha(\kappa) > 0$ is a necessary and sufficient condition for multistationarity.

Changing K_{tot}

We consider the value of S_{tot} fixed and analyse $K_{\text{tot}} = \psi_K(x_5)$. The function $\psi_K(x_5)$ is a rational function whose numerator has degree three and the denominator has degree two. Note that x_5 cannot increase beyond the bound given by the fixed total amount S_{tot}.

Figure 8 shows the function $\psi_K(x_5)$ with $S_{\text{tot}} = 591$ and the reaction rate constants as in (E.16), with the axes interchanged:

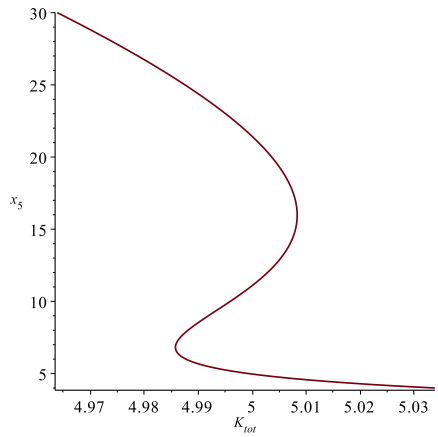


Figure 8.

To investigate the behavior of $\psi_K(x_5)$, we plot the function $\psi_K(x_5)$ for two different domains of x_5, see Figure 9(a,b).

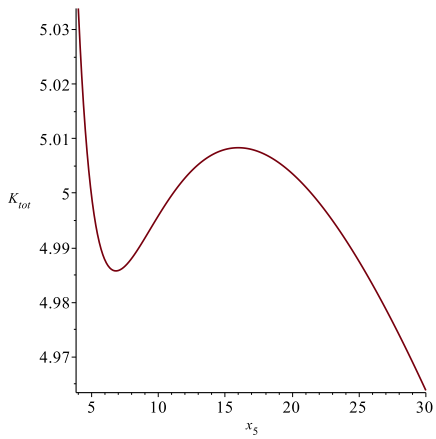


Figure 9a.

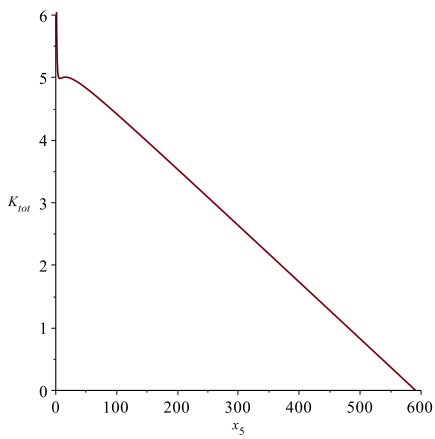


Figure 9b.

We also plot the value of K_{tot} against the value of x_6 (both are functions of x_5), see Figure 10(a,b).

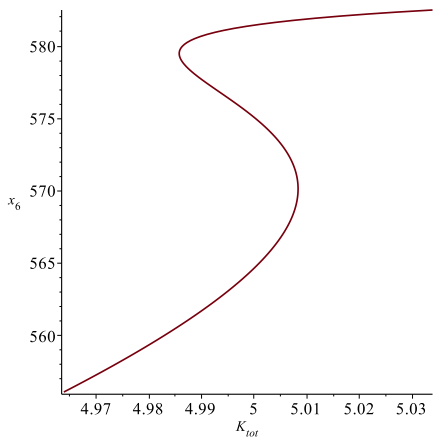


Figure 10a.

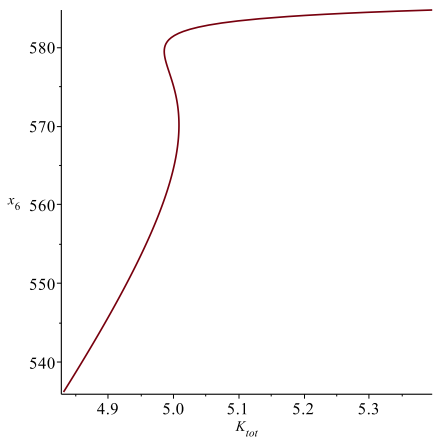


Figure 10b.

For low values of K_{tot}, the concentration of x_6 increases, until it reaches a transitional phase, after which it tends towards an upper bound, the total amount of substrate S_{tot}.

The derivative of $\psi_K(x_5)$ with respect to x_5 is a rational function

$$\psi_K'(x_5) = \frac{\gamma_1(\kappa)x_5^4 + \gamma_2(\kappa)x_5^3 + \gamma_3(\kappa, S_{\text{tot}})x_5^2 + \gamma_4(\kappa, S_{\text{tot}})x_5 + \gamma_5(\kappa, S_{\text{tot}})}{q_3(\kappa, x_5)},$$

where $\gamma_1(\kappa), \gamma_2(\kappa)$ are negative polynomials in the reaction rate constants, $\gamma_4(\kappa, S_{\text{tot}})$, $\gamma_5(\kappa, S_{\text{tot}})$ are negative polynomials in the reaction rate constants and S_{tot} (they are linear in S_{tot}), $q_3(\kappa, x_5)$ is a degree 4 polynomial in x_5 whose coefficients are positive polynomials in the reaction rate constants, and

$$\gamma_3(\kappa, S_{\text{tot}}) = S_{\text{tot}}\kappa_7\alpha(\kappa) + \gamma_3'(\kappa),$$

where $\gamma_3'(\kappa)$ is a polynomial in the reaction rate constants with positive and negative terms, and $\alpha(\kappa)$ is as in (E.11).

For small and large values of x_5, $\psi_K'(x_5)$ is negative and thus $\psi_K(x_5)$ decreases (because $\gamma_1(\kappa), \gamma_5(\kappa, S_{\text{tot}}) < 0$). If multistationarity occurs for some values of K_{tot}, $\psi_K(x_5)$ must increase in some interval, where necessarily $\gamma_3(\kappa, S_{\text{tot}}) > 0$. When multistationarity occurs, we argue as above for $\psi_S(x_5)$ to conclude that the

function has exactly one local maximum and one local minimum. In this case the graph of $\psi_K(x_5)$ has the same S-shape as in the example graph in Figure 9a.

Since $\gamma_1(\kappa)$, $\gamma_2(\kappa)$ do not depend on S_{tot} and $\gamma_3(\kappa, S_{\text{tot}})$, $\gamma_4(\kappa, S_{\text{tot}})$, $\gamma_5(\kappa, S_{\text{tot}})$ depend linearly on S_{tot}, we deduce that if $\alpha(\kappa) > 0$, then there are always values of S_{tot} that make $\psi'_K(x_5)$ negative for certain values of x_5 (for S_{tot} large enough such that $S_{\text{tot}}\kappa_7\alpha(\kappa)x_5^2$ dominates). Therefore, we see once again that $\alpha(\kappa) > 0$ is a necessary and sufficient condition for multistationarity.

E.2 The core model for n allosteric kinase competing for the same substrate

In this section we consider a simplified model of the model in subsection E.1.1, in which $\kappa_2 = \kappa_5 = 0$ and $\kappa_9 = \kappa_{10} = 0$. Furthermore, we consider the case where there are n allosteric kinases for the same substrate.

This simplified model is still multistationary as we will show below. Furthermore, the result of Joshi and Shiu [315] on multistationarity of reaction networks applies: If a reduced model has the same stoichiometric subspace as the full model, and the reduced model has N (non-degenerate) steady states, then this is also the case for the full model. We will apply this to the full model in subsection E.1.1.

E.2.1 Model description

We study the reduced system consisting of n allosteric kinases competing for the same substrate.

Let K^i, for $i = 1, \ldots, n$, denote the n allosteric kinases. We use subindices r, t to denote the relaxed or tensed state (respectively) of each of them. The set of reactions given in the previous subsection are reproduced for the n allosteric kinases, after making some reversible reactions irreversible and renaming the reaction rate constants accordingly. That is, for $i = 1, \ldots, n$, the reactions under consideration

are as follows:

$$K_r^i + S \xrightarrow{\kappa_{i,1}} K_r^i S \xrightarrow{\kappa_{i,2}} K_r^i + S_p \qquad\qquad K_r^i \xrightarrow{\kappa_{i,5}} K_t^i$$

$$K_t^i + S \xrightarrow{\kappa_{i,3}} K_t^i S \xrightarrow{\kappa_{i,4}} K_t^i + S_p \qquad\qquad K_t^i S \xrightarrow{\kappa_{i,6}} K_r^i S.$$

In addition, there is a dephosphorylation reaction

$$S_p \xrightarrow{\kappa_7} S.$$

We denote the concentration of the species as follows:

$$x_{i,1} := [K_r^i] \quad x_{i,2} := [K_t^i] \quad x_{i,3} := [K_r^i S] \quad x_{i,4} := [K_t^i S] \quad x_5 := [S] \quad x_6 := [S_p],$$

for $i = 1, \ldots, n$. We proceed as in the previous section and model the dynamics of the concentrations over time under the law of mass action by the following system of ordinary differential equations:

$$\begin{aligned}
\dot{x}_{i,1} &= -\kappa_{i,1} x_{i,1} x_5 + \kappa_{i,2} x_{i,3} - \kappa_{i,5} x_{i,1} \\
\dot{x}_{i,2} &= -\kappa_{i,3} x_{i,2} x_5 + \kappa_{i,4} x_{i,4} + \kappa_{i,5} x_{i,1} \\
\dot{x}_{i,3} &= \kappa_{i,1} x_{i,1} x_5 - \kappa_{i,2} x_{i,3} + \kappa_{i,6} x_{i,4} \\
\dot{x}_{i,4} &= \kappa_{i,3} x_{i,2} x_5 - \kappa_{i,4} x_{i,4} - \kappa_{i,6} x_{i,4} \\
\dot{x}_5 &= -\sum_{j=1}^{n} \left(\kappa_{j,1} x_{j,1} x_5 + \kappa_{j,3} x_{j,2} x_5\right) + \kappa_7 x_6 \\
\dot{x}_6 &= \sum_{j=1}^{n} \left(\kappa_{j,2} x_{j,3} + \kappa_{j,4} x_{j,4}\right) - \kappa_7 x_6
\end{aligned}$$

for $i = 1, \ldots, n$. The system has $n + 1$ conservation laws. Namely, for $i = 1, \ldots, n$, we have

$$x_{i,1} + x_{i,2} + x_{i,3} + x_{i,4} = K_{\text{tot}}^i \tag{E.19}$$

for some $K^i_{\text{tot}} > 0$, and for $S_{\text{tot}} > 0$,

$$x_5 + x_6 + \sum_{i=1}^{n}(x_{i,3} + x_{i,4}) = S_{\text{tot}}. \tag{E.20}$$

E.2.2 Summary of results

The results for the model with n allosteric kinases competing for the same substrate can be summarised in the following way. In subsection E.2.3 we show that the steady states of the system can be given in terms of the concentration x_5 of the substrate S only. That is, knowing the value of x_5 at steady state allows us to calculate the value of the remaining concentrations from x_5 alone.

Further, we show that there are reaction rate constants such that the system has exactly $2m + 1$, $m = 0, \ldots, n$, positive steady states. In particular this is true for $m = n$ in which case there are $2n+1$ positive steady state. In fact, we show that $2n + 1$ is the maximal possible number of steady states, positive as well as steady states for which at least one concentration is zero. In subsection E.2.5 we consider the stability of the steady states and show that if there are $2n + 1$ positive steady states then at least n of them are unstable.

E.2.3 Positive steady states

The positive steady states of the system are the solutions to the equations $\dot{x}_{i,1} = \dot{x}_{i,2} = \dot{x}_{i,3} = \dot{x}_{i,4} = 0$, for $i = 1, \ldots, n$, together with $\dot{x}_5 = \dot{x}_6 = 0$, constrained by the conservation laws (E.19) and (E.20). We reason as in the previous section and disregard the steady state equations $\dot{x}_5 = 0$ and $\dot{x}_{i,1} = 0$, for $i = 1, \ldots, n$. Using the equations $\dot{x}_{i,2} = \dot{x}_{i,3} = \dot{x}_{i,4} = 0$ and (E.19), we obtain algebraic expressions for $x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4}$ at steady state, depending on the value of x_5 at steady state,

analogous to the expressions (E.4)-(E.7):

$$x_{i,1} = \frac{K^i_{\text{tot}}\kappa_{i,2}\kappa_{i,3}\kappa_{i,6}x_5}{q_i(x)} \tag{E.21}$$

$$x_{i,2} = \frac{K^i_{\text{tot}}(\kappa_{i,4}+\kappa_{i,6})\kappa_{i,2}\kappa_{i,5}}{q_i(x)} \tag{E.22}$$

$$x_{i,3} = \frac{K^i_{\text{tot}}\kappa_{i,3}\kappa_{i,6}(\kappa_{i,1}x_5+\kappa_{i,5})x_5}{q_i(x)} \tag{E.23}$$

$$x_{i,4} = \frac{K^i_{\text{tot}}\kappa_{i,2}\kappa_{i,3}\kappa_{i,5}x_5}{q_i(x)} \tag{E.24}$$

$$q_i(x) = \kappa_{i,1}\kappa_{i,3}\kappa_{i,6}x_5^2 + \kappa_{i,3}(\kappa_{i,2}\kappa_{i,5}+\kappa_{i,2}\kappa_{i,6}+\kappa_{i,5}\kappa_{i,6})x_5 + \kappa_{i,2}\kappa_{i,5}(\kappa_{i,4}+\kappa_{i,6}).$$

These expressions are positive provided x_5 is positive. From the equation $\dot{x}_6 = 0$ we obtain

$$x_6 = \frac{\sum_{j=1}^n (k_{j,2}x_{j,3} + k_{j,4}x_{j,4})}{k_7} \tag{E.25}$$

which, using expressions (E.23) and (E.24), is positive provided $x_5 > 0$.

All concentrations are expressed as functions of x_5. We replace x_6 and subsequently x_3, x_4 in (E.20) by their expressions in (E.23)-(E.25) to obtain

$$(x_5 - S_{\text{tot}}) + \sum_{i=1}^n \frac{K^i_{\text{tot}}x_5\kappa_{i,3}((1+\kappa_{i,2}/\kappa_7)\,\kappa_{i,6}(\kappa_{i,1}x_5+\kappa_{i,5}) + (1+\kappa_{i,4}/\kappa_7)\,\kappa_{i,2}\kappa_{i,5})}{q_i(x)} = 0. \tag{E.26}$$

By clearing denominators, that is, by multiplying this equation by $\prod_{i=1}^n q_i(x)$, we obtain a polynomial $p(x_5)$ of degree $2n+1$ in x_5. As argued in the previous section, any positive root of the polynomial corresponds to a positive steady state. We note again that $p(x_5)$ has at least one positive root since $p(0) < 0$ and $p(+\infty) > 0$.

E.2.4 Existence of $2n+1$ positive steady states.

We have shown that the positive steady states of the system with n allosteric kinases competing for the same substrate are determined by the positive solutions to a polynomial $p(x_5)$ of degree $2n+1$. By the fundamental theorem of algebra, a

polynomial of degree $2n+1$ has $2n+1$ complex roots counted with multiplicity. Therefore, such a polynomial can *at most* have $2n+1$ distinct positive real roots.

We show in this section that there exist choices of reaction rate constants κ_i and total amounts $K^i_{\text{tot}}, S_{\text{tot}}$ such that the polynomial has exactly $2n+1$ distinct positive real roots. As a consequence, this proves that the system with n allosteric kinases competing for the same substrate admits $2n+1$ positive steady states for some choice of reaction rate constants and total amounts. As argued at the beginning of the section, this result holds for the general system where some reactions are made reversible.

The proof of this statement consists of a series of simplifications and constructions analogous to those in [50].

First of all observe that the steady states of the system are invariant by multiplication of all reaction rate constants by some scalar $\lambda > 0$. Therefore, we can assume that $\kappa_7 = 1$. For simplicity we write x for x_5. We let

$$\alpha_{i,1} = (\kappa_{i,2}+1)\, K^i_{\text{tot}} \kappa_{i,1}\kappa_{i,3}\kappa_{i,6} \tag{E.27}$$

$$\alpha_{i,2} = K^i_{\text{tot}} \kappa_{i,3}\kappa_{i,5}((\kappa_{i,2}+1)\,\kappa_{i,6} + (\kappa_{i,4}+1)\,\kappa_{i,2}) \tag{E.28}$$

$$\alpha_{i,3} = \kappa_{i,1}\kappa_{i,3}\kappa_{i,6} \tag{E.29}$$

$$\alpha_{i,4} = \kappa_{i,3}\left(\kappa_{i,2}\kappa_{i,5} + \kappa_{i,2}\kappa_{i,6} + \kappa_{i,5}\kappa_{i,6}\right) \tag{E.30}$$

$$\alpha_{i,5} = \kappa_{i,2}\kappa_{i,5}\left(\kappa_{i,4} + \kappa_{i,6}\right), \tag{E.31}$$

such that we write

$$\frac{K^i_{\text{tot}} x_5 \kappa_{i,3}((1+\kappa_{i,2}/\kappa_7)\,\kappa_{i,6}(\kappa_{i,1}x_5 + \kappa_{i,5}) + (1+\kappa_{i,4}/\kappa_7)\,\kappa_{i,2}\kappa_{i,5})}{q_i(x)} = \frac{\alpha_{i,1}x^2 + \alpha_{i,2}x}{\alpha_{i,3}x^2 + \alpha_{i,4}x + \alpha_{i,5}}.$$

Lemma 1. *For any positive values $\alpha_{i,1}, \ldots, \alpha_{i,5} > 0$, there exist $\kappa_{i,1}, \ldots, \kappa_{i,6} > 0$ and $K^i_{\text{tot}} > 0$ such that* (E.27)-(E.31) *are fulfilled.*

Proof. To simplify the notation, we prove that for all $\alpha_1, \ldots, \alpha_5 > 0$ there exist

$\kappa_1, \ldots, \kappa_6, K_{\text{tot}} > 0$ such that

$$\alpha_1 = (\kappa_2 + 1) K_{\text{tot}} \kappa_1 \kappa_3 \kappa_6 \tag{E.32}$$

$$\alpha_2 = K_{\text{tot}} \kappa_3 \kappa_5 ((\kappa_2 + 1) \kappa_6 + (\kappa_4 + 1) \kappa_2) \tag{E.33}$$

$$\alpha_3 = \kappa_1 \kappa_3 \kappa_6 \tag{E.34}$$

$$\alpha_4 = \kappa_3 (\kappa_2 \kappa_5 + \kappa_2 \kappa_6 + \kappa_5 \kappa_6) \tag{E.35}$$

$$\alpha_5 = \kappa_2 \kappa_5 (\kappa_4 + \kappa_6) . \tag{E.36}$$

Using the expressions for α_1, α_3, α_4 and α_5 we solve for $\kappa_1, \kappa_3, \kappa_5, K_{\text{tot}}$ and obtain

$$K_{\text{tot}} = \frac{\alpha_1}{(\kappa_2 + 1) \alpha_3} \qquad \kappa_3 = \frac{\alpha_4 \kappa_2 (\kappa_4 + \kappa_6)}{\kappa_2^2 \kappa_4 \kappa_6 + \kappa_2^2 \kappa_6^2 + \kappa_2 \alpha_5 + \kappa_6 \alpha_5}$$

$$\kappa_1 = \frac{\alpha_3 \left(\kappa_2^2 \kappa_4 \kappa_6 + \kappa_2^2 \kappa_6^2 + \kappa_2 \alpha_5 + \kappa_6 \alpha_5\right)}{\alpha_4 \kappa_2 \kappa_6 (\kappa_4 + \kappa_6)} \qquad \kappa_5 = \frac{\alpha_5}{\kappa_2 (\kappa_4 + \kappa_6)}.$$

Finally, using the equation for α_2 after plugging the previous expressions, we obtain

$$\alpha_2 = \frac{\alpha_1 \alpha_4 \alpha_5 \kappa_6}{\alpha_3 \left(\kappa_2^2 \kappa_4 \kappa_6 + \kappa_2^2 \kappa_6^2 + \kappa_2 \alpha_5 + \kappa_6 \alpha_5\right)} + \frac{(\kappa_4 + 1) \alpha_1 \kappa_2 \alpha_4 \alpha_5}{(\kappa_2 + 1) \alpha_3 \left(\kappa_2^2 \kappa_4 \kappa_6 + \kappa_2^2 \kappa_6^2 + \kappa_2 \alpha_5 + \kappa_6 \alpha_5\right)},$$

which is equivalent to the polynomial equation

$$\begin{aligned} 0 = &\left(-\kappa_2^3 \alpha_2 \alpha_3 - \kappa_2^2 \alpha_2 \alpha_3\right) \kappa_6^2 + \\ &\left(-\kappa_2^3 \kappa_4 \alpha_2 \alpha_3 - \kappa_2^2 \kappa_4 \alpha_2 \alpha_3 + \kappa_2 \alpha_1 \alpha_4 \alpha_5 - \kappa_2 \alpha_2 \alpha_3 \alpha_5 + \alpha_1 \alpha_4 \alpha_5 - \alpha_2 \alpha_3 \alpha_5\right) \kappa_6 \\ &- \kappa_2^2 \alpha_2 \alpha_3 \alpha_5 + \kappa_2 \kappa_4 \alpha_1 \alpha_4 \alpha_5 + \kappa_2 \alpha_1 \alpha_4 \alpha_5 - \kappa_2 \alpha_2 \alpha_3 \alpha_5. \end{aligned}$$

We obtained a polynomial of degree 2 in κ_6 with negative leader term. If the independent term is positive, then the polynomial has one positive root. Hence, we want to show that there exist values of κ_2, κ_4 such that

$$-\kappa_2 \alpha_5 (\kappa_2 \alpha_2 \alpha_3 - \kappa_4 \alpha_1 \alpha_4 - \alpha_1 \alpha_4 + \alpha_2 \alpha_3) > 0$$

or equivalently, that

$$\kappa_2\alpha_2\alpha_3 - \kappa_4\alpha_1\alpha_4 - \alpha_1\alpha_4 + \alpha_2\alpha_3 < 0.$$

For a fixed value of κ_2, this expression is a decreasing linear function on κ_4. Therefore, we can find a positive value of κ_4 such that it is negative.

We conclude that for any $\kappa_2 > 0$, we can find values of $\kappa_1, \kappa_3, \kappa_4, \kappa_5, \kappa_6 > 0$ and $K_{\text{tot}} > 0$ satisfying (E.32)-(E.36). □

As a consequence of Lemma 1, there exist values of the reaction rate constants and total amounts such that (E.26) holds if we can find $\alpha_{i,1}, \ldots, \alpha_{i,5} > 0$ such that

$$0 = x - S_{\text{tot}} + \sum_{i=1}^{n} \frac{\alpha_{i,1}x^2 + \alpha_{i,2}x}{\alpha_{i,3}x^2 + \alpha_{i,4}x + \alpha_{i,5}}. \tag{E.37}$$

With this notation, we want to determine the positive real roots of the polynomial obtained by clearing denominators in (E.37):

$$p(x) = (x - S_{\text{tot}}) \prod_{i=1}^{n} (\alpha_{i,3}x^2 + \alpha_{i,4}x + \alpha_{i,5}) + \sum_{i=1}^{n} \Big((\alpha_{i,1}x^2 + \alpha_{i,2}x) \prod_{j \neq i} (\alpha_{j,3}x^2 + \alpha_{j,4}x + \alpha_{j,5}) \Big). \tag{E.38}$$

The coefficient of degree $2n+1$ of $p(x)$ is $\prod_{i=1}^{n} \alpha_{i,3}$ and the independent term of $p(x)$ is $-S_{\text{tot}} \prod_{i=1}^{n} \alpha_{i,5}$. We set $\alpha_{i,4} = 0$ and $\alpha_{i,1} = 0$. Setting these two constants to zero, for $i = 1, \ldots, n$, does not change the degree of the polynomial. By the continuity of the isolated roots of a polynomial as functions of the coefficients of the polynomial, if we can find $\alpha_{i,2}, \alpha_{i,3}, \alpha_{i,5} > 0$ such that with $\alpha_{i,4} = \alpha_{i,1} = 0$, the polynomial $p(x)$ has $2n+1$ distinct positive real roots, then for $\alpha_{i,4}, \alpha_{i,1}$ small enough, the polynomial $p(x)$ still has $2n+1$ distinct positive real roots.

We further let $\alpha_{i,3} = 1$ for all $i = 1, \ldots, n$, and $S_{\text{tot}} = 1$. To ease the notation,

we write $a_i = \alpha_{i,2}$ and $b_i = \alpha_{i,5}$, such that the polynomial of interest becomes

$$p(x) = (x-1)\prod_{i=1}^{n}(x^2+b_i) + \sum_{i=1}^{n}\Big(a_i x \prod_{j\neq i}(x^2+b_j)\Big). \tag{E.39}$$

Lemma 2. *There exist $a_i, b_i > 0$, for $i = 1, \dots, n$, such that $p(x)$ has $2n+1$ positive roots.*

Proof. The statement is a consequence of Lemmas 2, 3 and 4 in the Supplementary Information of [50]. □

We are ready to prove the main result on the number of positive steady states.

Theorem 3. *For any $n \geq 1$, there exists a choice of reaction rate constants $\kappa_6 > 0$, $\kappa_{i,1}, \kappa_{i,2}, \kappa_{i,3}, \kappa_{i,4} > 0$ and total amounts $S_{\text{tot}}, K^i_{\text{tot}} > 0$, for $i = 1, \dots, n$, such that the system with n allosteric kinases competing for the same substrate has $2n+1$ distinct positive steady states.*

Proof. Let $b_i, a_i > 0$, for $i = 1, \dots, n$, as in Lemma 2, such that $p(x)$ in (E.39) has $2n+1$ distinct positive real roots. We set $\alpha_{i,2} = a_i, \alpha_{i,5} = b_i, \alpha_{i,3} = 1$, for $i = 1, \dots, n$, $S_{\text{tot}} = 1$ and let $\alpha_{i,1}, \alpha_{i,4} > 0$ be small enough such that the polynomial $p(x)$ in (E.38) has $2n+1$ distinct positive real roots. We set $k_6 = 1$. By construction, any choice $\kappa_{i,1}, \kappa_{i,2}, \kappa_{i,3}, \kappa_{i,4} > 0$ and $K^i_{\text{tot}} > 0$ such that (E.27)-(E.31) are fulfilled provides a set of parameters with $2n+1$ distinct positive steady states. Such a choice exists by Lemma 1. □

Remark 4. Consider the polynomial $p(x)$ in (E.39) and assume that $a_k = 0$ for a certain k. We get the polynomial

$$\begin{aligned}\tilde{p}(x) &= (x-1)\prod_{i=1}^{n}(x^2+b_i) + \sum_{i=1, i\neq k}^{n}\Big(a_i x \prod_{j\neq i}(x^2+b_j)\Big)\\ &= (x^2+b_k)\cdot\Big((x-1)\prod_{i=1, i\neq k}^{n}(x^2+b_i) + \sum_{i=1, i\neq k}^{n}\Big(a_i x \prod_{j\neq i,k}(x^2+b_j)\Big)\Big).\end{aligned}$$

Assuming $b_k > 0$, the polynomial $\tilde{p}(x)$ has a factor of degree two with non-real roots and a factor with the same form of the original p with degree $2(n-1)+1 = 2n-1$. We will show below that the later factor admits $2n-1$ positive roots for some choice of a_i, b_i. Therefore, we conclude that $\tilde{p}(x)$ admits $2n-1$ positive roots for some choice of parameters as well. By the continuity of the roots of a polynomial (to have $a_k > 0$), this implies that we can find reaction rate constants and total amounts such that the system has $2n-1$ positive steady states.

We can repeat the argument by letting m of the parameters among $a_1, \ldots, a_n$ be equal to zero, and conclude that we can find reaction rate constants and total amounts such that the system has $2(n-m)+1$ positive steady states.

E.2.5 n unstable steady states

In this subsection we show that, considering the $2n+1$ steady states ordered increasingly by their value $x = x_5$, then the steady states number $2, 4, \ldots, 2n$ are unstable relative to the stoichiometric compatibility class they belong to, that is, relative to the invariant subspaces described by the conservation laws (E.19) and (E.20).

Since the system with n allosteric kinases competing for the same substrate has $4n+2$ variables and $n+1$ conservation laws, the Jacobian of f in $\dot{x} = f(x)$ always has $n+1$ zero eigenvalues. The remaining $3n+1$ eigenvalues (which could include zero) have corresponding eigenvectors in the stoichiometric subspace and dictate the dynamics around the steady state and within the stoichiometric compatibility class. If the steady state is locally stable relative to the stoichiometric compatibility class, then the product of these $3n+1$ eigenvalues has sign $(-1)^{3n+1}$. Therefore, if the sign of the product of these eigenvalues is $(-1)^{3n}$, then the steady state is necessarily locally unstable relative to the stoichiometric compatibility class. We argue in the proof of the next theorem that this is the case for the steady states in even position $2, 4, \ldots, 2n$.

Theorem 5. *The $2, 4, \ldots, 2n$-th steady states are unstable relative to the stoichio-*

metric compatibility class.

Proof. We order the variables of the system as $x_{1,1}$, $x_{1,2}$, $x_{1,3}$, $x_{1,4}$, $\ldots$, $x_{n,1}$, $x_{n,2}$, $x_{n,3}$, $x_{n,4}$, x_5, x_6. It follows from [296, Prop. 5.3] that the product of the $3n+1$ eigenvalues of the Jacobian with eigenvectors in the stoichiometric space agrees with the determinant of the Jacobian of the function $g\colon \mathbb{R}^{4n+2} \to \mathbb{R}^{4n+2}$ where

$$
\begin{aligned}
g_{4(i-1)+1}(x) &= x_{i,1} + x_{i,2} + x_{i,3} + x_{i,4} - K^i_{\text{tot}} \\
g_{4(i-1)+2}(x) &= -\kappa_{i,3} x_{i,2} x_5 + \kappa_{i,4} x_{i,4} + \kappa_{i,5} x_{i,1} \\
g_{4(i-1)+3}(x) &= \kappa_{i,1} x_5 x_{i,1} - \kappa_{i,2} x_{i,3} + \kappa_{i,6} x_{i,4} \\
g_{4(i-1)+4}(x) &= \kappa_{i,3} x_{i,2} x_5 - \kappa_{i,4} x_{i,4} - \kappa_{i,6} x_{i,4}
\end{aligned}
$$

for $i = 1, \ldots, n$ and

$$
\begin{aligned}
g_{4n+1}(x) &= x_5 + x_6 + \sum_{j=1}^{n} (x_{j,3} + x_{j,4}) - S_{\text{tot}}, \\
g_{4n+2}(x) &= \sum_{j=1}^{n} (\kappa_{j,2} x_{j,3} + \kappa_{j,4} x_{j,4}) - \kappa_7 x_6.
\end{aligned}
$$

We now apply the method described in [54] (see the Electronic Supplementary Material of the paper), to determine the sign of the determinant of the Jacobian of g from iterative eliminations. One can check that the expressions in (E.21)-(E.24) are obtained from iteratively eliminating $x_{i,1}, \ldots, x_{i,4}$ from the equations $g_{4(i-1)+1}(x) = \cdots = g_{4(i-1)+4}(x) = 0$ respectively, which correspond to the conservation law with total amount K^i_{tot} together with $\dot{x}_{i,2} = \dot{x}_{i,3} = \dot{x}_{i,4} = 0$. Note that in each step we eliminate the first variable from the first function and for each group the first one is increasing and the other three are decreasing in the eliminated variable. Finally, we eliminate x_6 from $g_{4n+2}(x)$. Note that in that step we are eliminating the second variable from the second (remaining) function and that function is decreasing in x_6.

Let $p(x_5)$ be the polynomial obtained after clearing denominators in (E.26). Then, by [54], the sign of the determinant of the Jacobian of g at a steady state agrees with the sign of the derivative of $p(x_5)$, $p'(x_5)$, times $(-1)^{3n+1}$. Therefore, if $p'(x_5)$ is negative, then the corresponding steady state is locally unstable. Since $p(0)$ is negative, the first real root of $p(x_5)$ has positive derivative, and then the signs alternate. Therefore, the steady states corresponding to the $2, 4, \ldots, 2n$-th roots are locally unstable relatively to the stoichiometric compatibility class. □

E.3 Allosteric kinases with several states

In this section we consider the case in which the allosteric kinase is specific to one substrate, but the kinase might have more than 2 states. Let K_i, $i = 1, \ldots, n$ denote the n states of the kinase. The general model is in this case:

$$K_i + S \rightleftharpoons K_iS \longrightarrow K_i + S_p \qquad S_p \longrightarrow S \qquad K_i \rightleftharpoons K_j \qquad K_iS \rightleftharpoons K_jS$$

for all $i, j = 1, \ldots, n$, $i \neq j$.

In addition we will consider a simplified model. The simplified model is easier to analyse mathematically and we will demonstrate that for $n = 3, 4$ there exist reaction rate constants such that multistationarity occurs. A result by Joshi and Shiu [315] and Feliu and Wiuf [317] then allows us to conclude that there exist reaction rate constants for the original full model such that multistationarity also occurs in the full model.

The simplified model we consider is:

$$\begin{array}{llll} K_i + S \xrightarrow{\kappa_i} Y & \text{for } i = 1, \ldots, n & Y \xrightarrow{\eta_1} K_1 + S_p & \\ K_i + S \xrightarrow{\mu_{i,j}} K_j + S_p & \text{for } i \geq j \geq 2 & K_n \xrightarrow{\eta_n} K_{n-1} \xrightarrow{\eta_{n-1}} \ldots \xrightarrow{\eta_2} K_1 & \\ S_p \xrightarrow{\kappa_{n+1}} S. & & & \end{array}$$

When $n = 3$, the simplified model cannot have more that 3 positive steady states, because we can reduce the steady state equations with the conservation laws to a polynomial of degree 4 with positive independent term and negative leading term. By the Descartes' rule of sign, the polynomial can at most have 3 positive real roots.

For $n = 4$, we find the following instance of parameters that give 5 positive steady states.

$$\begin{aligned}
\kappa_1 &= 0.0369, & \kappa_2 &= 0.000172, & \kappa_3 &= 1.4 \cdot 10^{-9}, & \kappa_4 &= 0.00011, \\
\kappa_5 &= 1069.496, & \mu_{2,2} &= 0.0003, & \mu_{3,2} &= 0.000008426, & \mu_{4,2} &= 0.00016, \\
K_{\text{tot}} &= 102, & \mu_{3,3} &= 0.000085, & \mu_{4,3} &= 1999.97, & \mu_{4,4} &= 25165410, \\
S_{\text{tot}} &= 120, & \eta_1 &= 0.0107, & \eta_2 &= 0.1, & \eta_3 &= 0.001, \\
\eta_4 &= 0.1.
\end{aligned}$$

The general scenario is too complex mathematically and cannot be analysed as we did for the model in section E.2. We conjecture though that the number of positive steady states grows with n as well, such that the system admits at least $m + 1$ positive steady states where $m = n$ if n is even and $m = n - 1$ if n is odd.

Bibliography

[1] Joseph Walpole, Jason A Papin, and Shayn M Peirce. "Multiscale Computational Models of Complex Biological Systems". In: *Annual review of biomedical engineering* 15.1 (July 2013), pp. 137–154.

[2] F H Arnold. "Combinatorial and computational challenges for biocatalyst design." In: *Nature* 409.6817 (Jan. 2001), pp. 253–257.

[3] Eric J Deeds et al. "Combinatorial complexity and compositional drift in protein interaction networks." In: *PloS one* 7.3 (2012), e32032.

[4] Daniel M Cornforth et al. "Combinatorial quorum sensing allows bacteria to resolve their social and physical environment." In: *Proceedings of the National Academy of Sciences of the United States of America* 111.11 (Mar. 2014), pp. 4280–4284.

[5] Nicolas E Buchler, Ulrich Gerland, and Terence Hwa. "On schemes of combinatorial transcription logic." In: *Proceedings of the National Academy of Sciences USA* 100.9 (Apr. 2003), pp. 5136–5141.

[6] Gregory D Friedland and Tanja Kortemme. "Designing ensembles in conformational and sequence space to characterize and engineer proteins." In: *Current opinion in structural biology* 20.3 (June 2010), pp. 377–384.

[7] Wenfei Li, Wei Wang, and Shoji Takada. "Energy landscape views for interplays among folding, binding, and allostery of calmodulin domains." In: *Pro-*

ceedings of the National Academy of Sciences of the United States of America 111.29 (July 2014), pp. 10550–10555.

[8] Wenfei Li, Peter G Wolynes, and Shoji Takada. "Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.9 (Mar. 2011), pp. 3504–3509.

[9] Wenfei Li et al. "Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.44 (Oct. 2012), pp. 17789–17794.

[10] Nancy E Hynes et al. "Signalling change: signal transduction through the decades." In: *Nature reviews. Molecular cell biology* 14.6 (June 2013), pp. 393–398.

[11] Boris N Kholodenko. "Cell-signalling dynamics in time and space." In: *Nature reviews. Molecular cell biology* 7.3 (Mar. 2006), pp. 165–176.

[12] Gábor Balázsi, Alexander van Oudenaarden, and James J Collins. "Cellular decision making and biological noise: from microbes to mammals." In: *Cell* 144.6 (Mar. 2011), pp. 910–925.

[13] Jordi Garcia-Ojalvo and Alfonso Martinez Arias. "Towards a statistical mechanics of cell fate decisions." In: *Current opinion in genetics & development* 22.6 (Dec. 2012), pp. 619–626.

[14] Thomas M Norman et al. "Memory and modularity in cell-fate decision making." In: *Nature* 503.7477 (Nov. 2013), pp. 481–486.

[15] Tobias Grosskopf and Orkun S Soyer. "Synthetic microbial communities." In: *Current opinion in microbiology* 18 (Apr. 2014), pp. 72–77.

[16] Michael T Mee et al. "Syntrophic exchange in synthetic microbial communities." In: *Proceedings of the National Academy of Sciences of the United States of America* 111.20 (May 2014), E2149–56.

[17] Allan Konopka, Stephen Lindemann, and Jim Fredrickson. "Dynamics in microbial communities: unraveling mechanisms to identify principles." In: *The ISME journal* (Dec. 2014).

[18] Roger P Alexander et al. "Understanding modularity in molecular networks requires dynamics." In: *Science signaling* 2.81 (2009), pe44–pe44.

[19] Frederic Rousseau and Joost Schymkowitz. "A systems biology perspective on protein structural dynamics and signal transduction." In: *Current opinion in structural biology* 15.1 (Feb. 2005), pp. 23–30.

[20] Hans V Westerhoff et al. "Systems biology towards life in silico: mathematics of the control of living cells." In: *Journal of mathematical biology* 58.1-2 (Jan. 2009), pp. 7–34.

[21] Jeremy Gunawardena. "Systems biology. Biological systems theory." In: *Science (New York, N.Y.)* 328.5978 (Apr. 2010), pp. 581–582.

[22] Uri Alon. *An Introduction to Systems Biology.* Design Principles of Biological Circuits. CRC Press, July 2006.

[23] J J Tyson, K Chen, and B Novak. "Network dynamics and cell physiology." In: *Nature reviews. Molecular cell biology* 2.12 (Dec. 2001), pp. 908–916.

[24] G Weng, U S Bhalla, and R Iyengar. "Complexity in biological signaling systems." In: *Science (New York, N.Y.)* 284.5411 (Apr. 1999), pp. 92–96.

[25] Hiroaki Kitano. "Systems biology: a brief overview." In: *Science (New York, N.Y.)* 295.5560 (Mar. 2002), pp. 1662–1664.

[26] Hiroaki Kitano. "Computational systems biology." In: *Nature* 420.6912 (Nov. 2002), pp. 206–210.

[27] Michael Elowitz and Wendell A Lim. "Build life to understand it". In: *Nature* 468.7326 (Dec. 2010), pp. 889–890.

[28] Ernesto Andrianantoandro et al. "Synthetic biology: new engineering rules for an emerging discipline." In: *Molecular systems biology* 2.1 (2006), p. 2006.0028.

[29] Shankar Mukherji and Alexander van Oudenaarden. "Synthetic biology: understanding biological design from synthetic circuits". In: *Nature reviews. Genetics* (Dec. 2009).

[30] Jennifer A N Brophy and Christopher A Voigt. "Principles of genetic circuit design." In: *Nature methods* 11.5 (May 2014), pp. 508–520.

[31] C A Voigt et al. "Computational method to reduce the search space for directed protein evolution." In: *Proceedings of the National Academy of Sciences USA* 98.7 (Mar. 2001), pp. 3778–3783.

[32] Kevin Clancy and Christopher A Voigt. "Programming cells: towards an automated 'Genetic Compiler'." In: *Current Opinion in Biotechnology* 21.4 (Aug. 2010), pp. 572–581.

[33] Orkun S Soyer and Maureen A O'Malley. "Evolutionary systems biology: what it is and why it matters." In: *BioEssays : news and reviews in molecular, cellular and developmental biology* 35.8 (Aug. 2013), pp. 696–705.

[34] Christopher V Rao, John R Kirby, and Adam P Arkin. "Design and diversity in bacterial chemotaxis: a comparative study in Escherichia coli and Bacillus subtilis." In: *PLoS biology* 2.2 (Feb. 2004), E49.

[35] João F Matias Rodrigues and Andreas Wagner. "Evolutionary plasticity and innovations in complex metabolic reaction networks." In: *PLoS Computational Biology (PLOSCB) 6(11)* 5.12 (Dec. 2009), e1000613.

[36] Anton Crombach and Paulien Hogeweg. "Evolution of Evolvability in Gene Regulatory Networks." In: *PLoS Computational Biology (PLOSCB) 4(7)* 4.7 (2008), e1000112.

[37] U Alon. "Biological networks: the tinkerer as an engineer." In: *Science (New York, N.Y.)* 301.5641 (Sept. 2003), pp. 1866–1867.

[38] D L Nelson and M M Cox. *Lehninger Principles of Biochemistry.* W.H. Freeman, 2013.

[39] Joseph Schlessinger. "Common and distinct elements in cellular signaling via EGF and FGF receptors." In: *Science (New York, N.Y.)* 306.5701 (Nov. 2004), pp. 1506–1507.

[40] Qiang Zhang, Sudin Bhattacharya, and Melvin E Andersen. "Ultrasensitive response motifs: basic amplifiers in molecular signalling networks." In: *Open biology* 3.4 (Apr. 2013), pp. 130031–130031.

[41] James E Ferrell and Sang Hoon Ha. "Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity." In: *Trends in biochemical sciences* 39.10 (Oct. 2014), pp. 496–503.

[42] A Novick and M Weiner. "Enzyme induction as an all-or-none phenomenon". In: *Proceedings of the National Academy of Sciences of the United States of America* 43.7 (July 1957), pp. 553–566.

[43] Mohan K Malleshaiah et al. "The scaffold protein Ste5 directly controls a switch-like mating decision in yeast." In: *Nature* 465.7294 (May 2010), pp. 101–105.

[44] Gustavo J Melen et al. "Threshold responses to morphogen gradients by zero-order ultrasensitivity." In: *Molecular systems biology* 1.1 (2005), 2005.0028–E11.

[45] Hong Qian. "Cooperativity in cellular biochemical processes: noise-enhanced sensitivity, fluctuating enzyme, bistability with nonlinear feedback, and other mechanisms for sigmoidal responses." In: *Annual review of biophysics* 41.1 (2012), pp. 179–204.

[46] Bernardo A Mello and Yuhai Tu. "Perfect and near-perfect adaptation in a model of bacterial chemotaxis." In: *Biophysical journal* 84.5 (May 2003), pp. 2943–2956.

[47] D E Koshland, A Goldbeter, and J B Stock. "Amplification and adaptation in regulatory and sensory systems." In: *Science (New York, N.Y.)* 217.4556 (July 1982), pp. 220–225.

[48] Marcelo Behar et al. "Mathematical and computational analysis of adaptation via feedback inhibition in signal transduction pathways." In: *Biophysical journal* 93.3 (Aug. 2007), pp. 806–821.

[49] Attila Csikász-Nagy and Orkun S Soyer. "Adaptive dynamics with a single two-state protein." In: *Journal of the Royal Society, Interface / the Royal Society* 5 Suppl 1 (Aug. 2008), S41–7.

[50] Varun B Kothamachu et al. "Unlimited multistability and Boolean logic in microbial signalling." In: *Journal of the Royal Society, Interface / the Royal Society* 12.108 (July 2015), pp. 20150234–20150234.

[51] Heather A Harrington et al. "Cellular compartments cause multistability and allow cells to process more information." In: *Biophysical journal* 104.8 (Apr. 2013), pp. 1824–1831.

[52] Matthew Thomson and Jeremy Gunawardena. "Unlimited multistability in multisite phosphorylation systems." In: *Nature* 460.7252 (July 2009), pp. 274–277.

[53] Fernando Ortega et al. “Bistability from double phosphorylation in signal transduction. Kinetic and structural requirements.” In: *The FEBS journal* 273.17 (Sept. 2006), pp. 3915–3926.

[54] Elisenda Feliu and Carsten Wiuf. “Enzyme-sharing as a cause of multi-stationarity in signalling systems.” In: *Journal of the Royal Society, Interface / the Royal Society* 9.71 (June 2012), pp. 1224–1232.

[55] Naama Geva-Zatorsky et al. “Oscillations and variability in the p53 system.” In: *Molecular systems biology* 2.1 (2006), p. 2006.0033.

[56] Fernando Siso-Nadal et al. “Cross-talk between signaling pathways can generate robust oscillations in calcium and cAMP.” In: *PloS one* 4.10 (2009), e7189.

[57] Michael J Rust et al. “Ordered phosphorylation governs oscillation of a three-protein circadian clock.” In: *Science (New York, N.Y.)* 318.5851 (Nov. 2007), pp. 809–812.

[58] Alexander Hoffmann et al. “The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation.” In: *Science (New York, N.Y.)* 298.5596 (Nov. 2002), pp. 1241–1245.

[59] Louise Ashall et al. “Pulsatile stimulation determines timing and specificity of NF-kappaB-dependent transcription.” In: *Science (New York, N.Y.)* 324.5924 (Apr. 2009), pp. 242–246.

[60] Y Li and A Goldbeter. “Frequency encoding of pulsatile signals of cAMP based on receptor desensitization in Dictyostelium cells.” In: *Journal of theoretical biology* 146.3 (Oct. 1990), pp. 355–367.

[61] Alexander Loewer et al. “Basal dynamics of p53 reveal transcriptionally attenuated pulses in cycling cells.” In: *Cell* 142.1 (July 2010), pp. 89–100.

[62] James E Ferrell and Sang Hoon Ha. "Ultrasensitivity part III: cascades, bistable switches, and oscillators." In: *Trends in biochemical sciences* 39.12 (Dec. 2014), pp. 612–618.

[63] Paul François, Nicolas Despierre, and Eric D Siggia. "Adaptive temperature compensation in circadian oscillations." In: *PLoS Computational Biology (PLOSCB) 6(11)* 8.7 (2012), e1002585.

[64] D C LaPorte and D E Koshland. "Phosphorylation of isocitrate dehydrogenase as a demonstration of enhanced sensitivity in covalent regulation." In: *Nature* 305.5932 (Sept. 1983), pp. 286–290.

[65] Paul François and Eric D Siggia. "A case study of evolutionary computation of biochemical adaptation." In: *Physical biology* 5.2 (2008), p. 026009.

[66] Wenzhe Ma et al. "Defining network topologies that can achieve biochemical adaptation." In: *Cell* 138.4 (Aug. 2009), pp. 760–773.

[67] William Pontius, Michael W Sneddon, and Thierry Emonet. "Adaptation dynamics in densely clustered chemoreceptors." In: *PLoS Computational Biology (PLOSCB) 6(11)* 9.9 (2013), e1003230.

[68] Diana Clausznitzer et al. "Chemotactic Response and Adaptation Dynamics in ¡i¿Escherichia coli¡/i¿." In: *PLoS Computational Biology (PLOSCB) 6(5)* 6.5 (2010), e1000784.

[69] N Barkai and S Leibler. "Robustness in simple biochemical networks." In: *Nature* 387.6636 (June 1997), pp. 913–917.

[70] U Alon et al. "Robustness in bacterial chemotaxis." In: *Nature* 397.6715 (Jan. 1999), pp. 168–171.

[71] James E Ferrell. "Feedback regulation of opposing enzymes generates robust, all-or-none bistable responses." In: *Current biology : CB* 18.6 (Mar. 2008), R244–5.

[72] J E Ferrell and E M Machleder. "The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes." In: *Science (New York, N.Y.)* 280.5365 (May 1998), pp. 895–898.

[73] J E Lisman. "A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase." In: *Proceedings of the National Academy of Sciences USA* 82.9 (May 1985), pp. 3055–3057.

[74] Ian B Dodd et al. "Theoretical analysis of epigenetic cell memory by nucleosome modification." In: *Cell* 129.4 (May 2007), pp. 813–822.

[75] James E Ferrell. "Bistability, bifurcations, and Waddington's epigenetic landscape." In: *Current biology : CB* 22.11 (June 2012), R458–66.

[76] Jan-Willem Veening, Wiep Klaas Smits, and Oscar P Kuipers. "Bistability, epigenetics, and bet-hedging in bacteria." In: *Annual review of microbiology* 62.1 (2008), pp. 193–210.

[77] Theodore J Perkins and Peter S Swain. "Strategies for cellular decision-making." In: *Molecular systems biology* 5.1 (2009), p. 326.

[78] Bruno M C Martins and James C W Locke. "Microbial individuality: how single-cell heterogeneity enables population level strategies." In: *Current opinion in microbiology* 24 (Apr. 2015), pp. 104–112.

[79] Joseph S Markson et al. "Circadian control of global gene expression by the cyanobacterial master regulator RpaA." In: *Cell* 155.6 (Dec. 2013), pp. 1396–1408.

[80] D A Rand et al. "Design principles underlying circadian clocks." In: *Journal of the Royal Society, Interface / the Royal Society* 1.1 (Nov. 2004), pp. 119–130.

[81] Michael J Rust, Susan S Golden, and Erin K O'Shea. "Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator." In: *Science (New York, N.Y.)* 331.6014 (Jan. 2011), pp. 220–223.

[82] Vikram Vijayan, Rick Zuzow, and Erin K O'Shea. "Oscillations in supercoiling drive circadian gene expression in cyanobacteria." In: *Proceedings of the National Academy of Sciences of the United States of America* 106.52 (Dec. 2009), pp. 22564–22568.

[83] Ryan A Kellogg and Savaş Tay. "Noise facilitates transcriptional control under dynamic inputs." In: *Cell* 160.3 (Jan. 2015), pp. 381–392.

[84] Savaş Tay et al. "Cells Respond Digitally to Variation in Signal Intensity via Stochastic Activation of NF-κB". In: *Biophysical journal* 98.3 (Jan. 2010), 429a–430a.

[85] O Decroly and A Goldbeter. "Birhythmicity, chaos, and other patterns of temporal self-organization in a multiply regulated biochemical system." In: *Proceedings of the National Academy of Sciences USA* 79.22 (Nov. 1982), pp. 6917–6921.

[86] Bela Novak and John J Tyson. "Design principles of biochemical oscillators." In: *Nature reviews. Molecular cell biology* 9.12 (Dec. 2008), pp. 981–991.

[87] J MONOD, J WYMAN, and J P CHANGEUX. "**On the nature of allosteric transitions: a plausible model**". In: *Journal of molecular biology* 12 (May 1965), pp. 88–118.

[88] Jean-Pierre Changeux and Stuart J Edelstein. "Allosteric mechanisms of signal transduction." In: *Science (New York, N.Y.)* 308.5727 (June 2005), pp. 1424–1428.

[89] L Pauling. "The oxygen equilibrium of hemoglobin and its structural interpretation." In: *Proceedings of the National Academy of Sciences USA* 21.4 (Apr. 1935), pp. 186–191.

[90] D E Koshland, G Némethy, and D Filmer. "Comparison of experimental binding data and theoretical models in proteins containing subunits." In: *Biochemistry* 5.1 (Jan. 1966), pp. 365–385.

[91] Sarah Marzen, Hernan G Garcia, and Rob Phillips. "Statistical mechanics of Monod-Wyman-Changeux (MWC) models." In: *Journal of molecular biology* 425.9 (May 2013), pp. 1433–1460.

[92] A Goldbeter and D E Koshland. "An amplified sensitivity arising from covalent modification in biological systems." In: *Proceedings of the National Academy of Sciences of the United States of America* 78.11 (Nov. 1981), pp. 6840–6844.

[93] Sun Young Kim and James E Ferrell. "Substrate competition as a source of ultrasensitivity in the inactivation of Wee1." In: *Cell* 128.6 (Mar. 2007), pp. 1133–1145.

[94] Ronny Straube. "Analysis of substrate competition in regulatory network motifs: Stimulus-response curves, thresholds and ultrasensitivity." In: *Journal of theoretical biology* 380 (May 2015), pp. 74–82.

[95] Nicolas E Buchler and Frederick R Cross. "Protein sequestration generates a flexible ultrasensitive response in a genetic network." In: *Molecular systems biology* 5 (2009), p. 272.

[96] Xinfeng Liu, Lee Bardwell, and Qing Nie. "A combination of multisite phosphorylation and substrate sequestration produces switchlike responses." In: *Biophysical journal* 98.8 (Apr. 2010), pp. 1396–1407.

[97] Bruno M C Martins and Peter S Swain. "Ultrasensitivity in phosphorylation-dephosphorylation cycles with little substrate." In: *PLoS Computational Biology (PLOSCB) 6(11)* 9.8 (2013), e1003175.

[98] Dennis D Wykoff et al. "Positive feedback regulates switching of phosphate transporters in S. cerevisiae." In: *Molecular cell* 27.6 (Sept. 2007), pp. 1005–1013.

[99] Hisaaki Shinohara et al. "Positive feedback within a kinase signaling complex functions as a switch mechanism for NF-κB activation." In: *Science (New York, N.Y.)* 344.6185 (May 2014), pp. 760–764.

[100] Liam J Holt, Andrew N Krutchinsky, and David O Morgan. "Positive feedback sharpens the anaphase switch." In: *Nature* 454.7202 (July 2008), pp. 353–357.

[101] Onn Brandman and Tobias Meyer. "Feedback loops shape cellular signals in space and time." In: *Science (New York, N.Y.)* 322.5900 (Oct. 2008), pp. 390–395.

[102] James E Ferrell and Wen Xiong. "Bistability in cell signaling: How to make continuous processes discontinuous, and reversible processes irreversible." In: *Chaos (Woodbury, N.Y.)* 11.1 (Mar. 2001), pp. 227–236.

[103] James E Ferrell et al. "Simple, realistic models of complex biological processes: positive feedback and bistability in a cell fate switch and a cell cycle oscillator." In: *FEBS letters* 583.24 (Dec. 2009), pp. 3999–4005.

[104] Gheorghe Craciun, Yangzhong Tang, and Martin Feinberg. "Understanding bistability in complex enzyme-driven reaction networks." In: *Proceedings of the National Academy of Sciences USA* 103.23 (June 2006), pp. 8697–8702.

[105] Christophe Soulé. "Graphic Requirements for Multistationarity". In: *Complexus* 1.3 (2003), pp. 123–133.

[106] Nick I Markevich, Jan B Hoek, and Boris N Kholodenko. "Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades." In: *The Journal of cell biology* 164.3 (Feb. 2004), pp. 353–359.

[107] David Angeli, James E Ferrell, and Eduardo D Sontag. "Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems." In: *Proceedings of the National Academy of Sciences USA* 101.7 (Feb. 2004), pp. 1822–1827.

[108] Maier S Avendaño, Chad Leidy, and Juan M Pedraza. "Tuning the range and stability of multiple phenotypic states with coupled positive-negative feedback loops." In: *Nature communications* 4 (2013), p. 2605.

[109] Leland H Hartwell et al. "From molecular to modular cell biology". In: *Nature* 402.supp (Dec. 1999), pp. C47–C52.

[110] Wendell A Lim, Connie M Lee, and Chao Tang. "Design principles of regulatory networks: searching for the molecular algorithms of the cell." In: *Molecular cell* 49.2 (Jan. 2013), pp. 202–212.

[111] Markus Kollmann et al. "Design principles of a bacterial signalling network." In: *Nature* 438.7067 (Nov. 2005), pp. 504–507.

[112] Guy Shinar and Martin Feinberg. "Design principles for robust biochemical reaction networks: what works, what cannot work, and what might almost work." In: *Mathematical biosciences* 231.1 (May 2011), pp. 39–48.

[113] Yuval Hart et al. "Design principles of cell circuits with paradoxical components." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.21 (May 2012), pp. 8346–8351.

[114] Abhinav Tiwari et al. "Bistable responses in bacterial genetic networks: designs and dynamical consequences." In: *Mathematical biosciences* 231.1 (May 2011), pp. 76–89.

[115] Jianfei Hu et al. "Systematic Prediction of Scaffold Proteins Reveals New Design Principles in Scaffold-Mediated Signal Transduction". In: *PLoS Computational Biology (PLOSCB) 6(11)* 11.9 (Sept. 2015), e1004508.

[116] William S Hlavacek and James R Faeder. "The complexity of cell signaling and the need for a new mechanics." In: *Science signaling* 2.81 (2009), pe46–pe46.

[117] Heinz W Engl et al. "Inverse problems in systems biology". In: *Inverse Problems* 25.12 (Dec. 2009), p. 123014.

[118] Kristin Wuichet and Igor B Zhulin. "Origins and diversification of a complex signal transduction system in prokaryotes." In: *Science signaling* 3.128 (2010), ra50–ra50.

[119] Orkun S Soyer and Sebastian Bonhoeffer. "Evolution of complexity in signaling pathways." In: *Proceedings of the National Academy of Sciences USA* 103.44 (Oct. 2006), pp. 16337–16342.

[120] Orkun S Soyer and Thomas Pfeiffer. "Evolution under fluctuating environments explains observed robustness in metabolic networks." In: *PLoS Computational Biology (PLOSCB) 6(11)* 6.8 (2010), e1000907.

[121] Thomas Pfeiffer, Orkun S Soyer, and Sebastian Bonhoeffer. "The evolution of connectivity in metabolic networks." In: *PLoS biology* 3.7 (July 2005), e228.

[122] N Kashtan and U Alon. "Spontaneous evolution of modularity and network motifs". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.39 (Sept. 2005), pp. 13773–13778.

[123] Paul François and Vincent Hakim. "Design of genetic networks with specified functions by evolution in silico." In: *Proceedings of the National Academy of Sciences of the United States of America* 101.2 (Jan. 2004), pp. 580–585.

[124] Javier Carrera, Santiago F Elena, and Alfonso Jaramillo. "Computational design of genomic transcriptional networks with adaptation to varying environments." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.38 (Sept. 2012), pp. 15277–15282.

[125] Daniel J Mandell and Tanja Kortemme. "Computer-aided design of functional protein interactions." In: *Nature chemical biology* 5.11 (Nov. 2009), pp. 797–807.

[126] James T MacDonald et al. "Computational design approaches and tools for synthetic biology." In: *Integrative biology : quantitative biosciences from nano to macro* 3.2 (Feb. 2011), pp. 97–108.

[127] Peipei Zhou et al. "Mechanisms generating bistability and oscillations in microRNA-mediated motifs." In: *Physical review. E, Statistical, nonlinear, and soft matter physics* 85.4 Pt 1 (Apr. 2012), p. 041916.

[128] Daniele Soroldoni et al. "Genetic oscillations. A Doppler effect in embryonic pattern formation." In: *Science (New York, N.Y.)* 345.6193 (July 2014), pp. 222–225.

[129] Liang Qiao et al. "Bistability and oscillations in the Huang-Ferrell model of MAPK signaling." In: *PLoS Computational Biology (PLOSCB) 6(11)* 3.9 (Sept. 2007), pp. 1819–1826.

[130] Tathagata Dasgupta et al. "A fundamental trade-off in covalent switching and its circumvention by enzyme bifunctionality in glucose homeostasis." In: *The Journal of biological chemistry* 289.19 (May 2014), pp. 13010–13025.

[131] Yuval Hart et al. "Paradoxical signaling by a secreted molecule leads to homeostasis of cell levels." In: *Cell* 158.5 (Aug. 2014), pp. 1022–1032.

[132] Amit Tzur et al. "Cell growth and size homeostasis in proliferating animal cells." In: *Science (New York, N.Y.)* 325.5937 (July 2009), pp. 167–171.

[133] Bree B Aldridge et al. "Physicochemical modelling of cell signalling pathways." In: *Nature cell biology* 8.11 (Nov. 2006), pp. 1195–1203.

[134] Wendell A Lim. "Designing customized cell signalling circuits." In: *Nature reviews. Molecular cell biology* 11.6 (June 2010), pp. 393–403.

[135] Ron Milo et al. “Superfamilies of evolved and designed networks.” In: *Science (New York, N.Y.)* 303.5663 (Mar. 2004), pp. 1538–1542.

[136] Uri Alon. “Network motifs: theory and experimental approaches.” In: *Nature reviews. Genetics* 8.6 (June 2007), pp. 450–461.

[137] Ralf Steuer et al. “Robust signal processing in living cells.” In: *PLoS Computational Biology (PLOSCB) 6(11)* 7.11 (Nov. 2011), e1002218.

[138] Joe H Levine et al. “Pulsed feedback defers cellular differentiation.” In: *PLoS biology* 10.1 (Jan. 2012), e1001252.

[139] C Y Huang and J E Ferrell. “Ultrasensitivity in the mitogen-activated protein kinase cascade.” In: *Proceedings of the National Academy of Sciences USA* 93.19 (Sept. 1996), pp. 10078–10083.

[140] Caleb J Bashor et al. “Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics.” In: *Science (New York, N.Y.)* 319.5869 (Mar. 2008), pp. 1539–1543.

[141] Peter Lenz and Peter S Swain. “An entropic mechanism to generate highly cooperative and specific binding from protein phosphorylations.” In: *Current biology : CB* 16.21 (Nov. 2006), pp. 2150–2155.

[142] Kirill Evlampiev and Hervé Isambert. “Conservation and topology of protein interaction networks under duplication-divergence evolution.” In: *Proceedings of the National Academy of Sciences of the United States of America* 105.29 (July 2008), pp. 9863–9868.

[143] Orkun S Soyer, Thomas Pfeiffer, and Sebastian Bonhoeffer. “Simulating the evolution of signal transduction pathways.” In: *Journal of theoretical biology* 241.2 (July 2006), pp. 223–232.

[144] Nadav Kashtan et al. "Extinctions in heterogeneous environments and the evolution of modularity." In: *Evolution; international journal of organic evolution* 63.8 (Aug. 2009), pp. 1964–1975.

[145] Otto X Cordero and Paulien Hogeweg. "Feed-forward loop circuits as a side effect of genome evolution." In: *Molecular biology and evolution* 23.10 (Oct. 2006), pp. 1931–1936.

[146] Dov J Stekel and Dafyd J Jenkins. "Evolution of resource and energy management in biologically realistic gene regulatory network models." In: *Advances in experimental medicine and biology* 751.Chapter 14 (2012), pp. 301–328.

[147] Evan A Variano, Jonathan H McCoy, and Hod Lipson. "Networks, Dynamics, and Modularity". In: *Physical review letters* 92.18 (May 2004), p. 188701.

[148] Anastasia Deckard and Herbert M Sauro. "Preliminary studies on the in silico evolution of biochemical networks." In: *Chembiochem : a European journal of chemical biology* 5.10 (Oct. 2004), pp. 1423–1431.

[149] Dafyd J Jenkins and Dov J Stekel. "A new model for investigating the evolution of transcription control networks." In: *Artificial life* 15.3 (2009), pp. 259–291.

[150] Orkun S Soyer. "Emergence and maintenance of functional modules in signaling pathways." In: *BMC evolutionary biology* 7.1 (2007), p. 205.

[151] Marcel Salathé and Orkun S Soyer. "Parasites lead to evolution of robustness against gene loss in host signaling networks." In: *Molecular systems biology* 4 (2008), p. 202.

[152] Richard A Goldstein and Orkun S Soyer. "Evolution of Taxis Responses in Virtual Bacteria: Non-Adaptive Dynamics." In: *PLoS Computational Biology (PLOSCB) 6(11)* 4.5 (2008), e1000084.

[153] Paul François and Eric D Siggia. "Predicting embryonic patterning using mutual entropy fitness and in silico evolution." In: *Development (Cambridge, England)* 137.14 (July 2010), pp. 2385–2395.

[154] Caleb J Bashor et al. "Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems." In: *Annual review of biophysics* 39.1 (2010), pp. 515–537.

[155] Sergio G Peisajovich et al. "Rapid diversification of cell signaling phenotypes by modular domain recombination." In: *Science (New York, N.Y.)* 328.5976 (Apr. 2010), pp. 368–372.

[156] Julien F Ollivier, Vahid Shahrezaei, and Peter S Swain. "Scalable Rule-Based Modelling of Allosteric Proteins and Biochemical Networks." In: *PLoS Computational Biology (PLOSCB) 6(11)* 6.11 (2010), e1000975.

[157] Vincent Danos et al. "Rule-Based Modelling of Cellular Signalling." In: *CONCUR* 4703.Chapter 3 (2007), pp. 17–41.

[158] William S Hlavacek et al. "Rules for modeling signal-transduction systems." In: *Science's STKE : signal transduction knowledge environment* 2006.344 (July 2006), re6–re6.

[159] Lily A Chylek et al. "Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems." In: *Wiley interdisciplinary reviews. Systems biology and medicine* 6.1 (Jan. 2014), pp. 13–36.

[160] Nadav Kashtan et al. "An analytically solvable model for rapid evolution of modular structure." In: *PLoS Computational Biology (PLOSCB) 6(11)* 5.4 (Apr. 2009), e1000355.

[161] Justin S Hogg et al. "Exact hybrid particle/population simulation of rule-based models of biochemical systems." In: *PLoS Computational Biology (PLOSCB) 6(11)* 10.4 (Apr. 2014), e1003544.

[162] J Yang, X Meng, and W S Hlavacek. "Rule-based modelling and simulation of biochemical systems with molecular finite automata." In: *IET systems biology* 4.6 (Nov. 2010), pp. 453–466.

[163] Roby P Bhattacharyya et al. "The Ste5 scaffold allosterically modulates signaling output of the yeast mating pathway." In: *Science (New York, N.Y.)* 311.5762 (Feb. 2006), pp. 822–826.

[164] Jing Jin et al. "Eukaryotic protein domains as functional units of cellular evolution." In: *Science signaling* 2.98 (2009), ra76–ra76.

[165] Najeeb Halabi et al. "Protein sectors: evolutionary units of three-dimensional structure." In: *Cell* 138.4 (Aug. 2009), pp. 774–786.

[166] Brian J Yeh et al. "Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors." In: *Nature* 447.7144 (May 2007), pp. 596–600.

[167] John E Dueber et al. "Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry." In: *Current opinion in structural biology* 14.6 (Dec. 2004), pp. 690–699.

[168] Sang-Hyun Park, Ali Zarrinpar, and Wendell A Lim. "Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms." In: *Science (New York, N.Y.)* 299.5609 (Feb. 2003), pp. 1061–1064.

[169] John E Dueber et al. "Reprogramming control of an allosteric signaling switch through modular recombination." In: *Science (New York, N.Y.)* 301.5641 (Sept. 2003), pp. 1904–1908.

[170] Ty M Thomson et al. "Scaffold number in yeast signaling system sets tradeoff between system output and dynamic range." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.50 (Dec. 2011), pp. 20265–20270.

[171] Birgit Schoeberl et al. "Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors." In: *Nature biotechnology* 20.4 (Apr. 2002), pp. 370–375.

[172] Brian D Slaughter, Joel W Schwartz, and Rong Li. "Mapping dynamic protein interactions in MAP kinase signaling using live-cell fluorescence fluctuation spectroscopy and imaging." In: *Proceedings of the National Academy of Sciences of the United States of America* 104.51 (Dec. 2007), pp. 20320–20325.

[173] Celine I Maeder et al. "Spatial regulation of Fus3 MAP kinase activity through a reaction-diffusion mechanism in yeast pheromone signalling." In: *Nature cell biology* 9.11 (Nov. 2007), pp. 1319–1326.

[174] Aki Fujioka et al. "Dynamics of the Ras/ERK MAPK cascade as monitored by fluorescent probes." In: *The Journal of biological chemistry* 281.13 (Mar. 2006), pp. 8917–8926.

[175] Mariko Hatakeyama et al. "A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling." In: *The Biochemical journal* 373.Pt 2 (July 2003), pp. 451–463.

[176] Jianzhi Zhang. "Evolution by gene duplication: an update". In: *Trends in Ecology & Evolution* 18.6 (June 2003), pp. 292–298.

[177] Asa K Björklund, Diana Ekman, and Arne Elofsson. "Expansion of protein domain repeats." In: *PLoS Computational Biology (PLOSCB) 6(11)* 2.8 (Aug. 2006), e114.

[178] H Allen Orr. "Fitness and its role in evolutionary genetics". In: *Nature reviews. Genetics* 10.8 (Aug. 2009), pp. 531–539.

[179] Orkun S Soyer, Marcel Salathé, and Sebastian Bonhoeffer. "Signal transduction networks: topology, response and biochemical processes." In: *Journal of theoretical biology* 238.2 (Jan. 2006), pp. 416–425.

[180] Fernando Siso-Nadal, Julien F Ollivier, and Peter S Swain. "Facile: a command-line network compiler for systems biology." In: *BMC systems biology* 1.1 (2007), p. 36.

[181] James F Crow and Motoo Kimura. *An Introduction to Population Genetics Theory*. Blackburn Press, Jan. 2009.

[182] Manfred Eigen, John McCaskill, and Peter Schuster. "Molecular quasi-species". In: *The Journal of Physical Chemistry* 92.24 (Dec. 1988), pp. 6881–6891.

[183] Josef Hofbauer and Karl Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press, 2009.

[184] David D Pollock, Grant Thiltgen, and Richard A Goldstein. "Amino acid coevolution induces an evolutionary Stokes shift." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.21 (May 2012), E1352–9.

[185] Michael Lynch and John S Conery. "The origins of genome complexity." In: *Science (New York, N.Y.)* 302.5649 (Nov. 2003), pp. 1401–1404.

[186] Paul François. "Evolving phenotypic networks in silico." In: *Seminars in cell & developmental biology* 35 (Nov. 2014), pp. 90–97.

[187] Edda Klipp et al. "Integrative model of the response of yeast to osmotic shock." In: *Nature biotechnology* 23.8 (Aug. 2005), pp. 975–982.

[188] Paul François, Vincent Hakim, and Eric D Siggia. "Deriving structure from evolution: metazoan segmentation." In: *Molecular systems biology* 3 (2007), p. 154.

[189] Kirsten H ten Tusscher and Paulien Hogeweg. "Evolution of Networks for Body Plan Patterning; Interplay of Modularity, Robustness and Evolvability." In: *PLoS Computational Biology (PLOSCB) 7(10)* 7.10 (2011), e1002208.

[190] Jean-Benoît Lalanne and Paul François. "Principles of adaptive sorting revealed by in silico evolution." In: *Physical review letters* 110.21 (May 2013), p. 218102.

[191] Michael W Sneddon, James R Faeder, and Thierry Emonet. "Efficient modeling, simulation and coarse-graining of biological complexity with NFsim." In: *Nature methods* 8.2 (Feb. 2011), pp. 177–183.

[192] John J Tyson, Katherine C Chen, and Bela Novak. "Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell." In: *Current opinion in cell biology* 15.2 (Apr. 2003), pp. 221–231.

[193] John J Tyson and Bela Novak. "Functional motifs in biochemical reaction networks." In: *Annual review of physical chemistry* 61.1 (2010), pp. 219–240.

[194] Adrian L Slusarczyk, Allen Lin, and Ron Weiss. "Foundations for the design and implementation of synthetic genetic circuits." In: *Nature reviews. Genetics* 13.6 (June 2012), pp. 406–420.

[195] Priscilla E M Purnick and Ron Weiss. "The second wave of synthetic biology: from modules to systems." In: *Nature reviews. Molecular cell biology* 10.6 (June 2009), pp. 410–422.

[196] T M Yi et al. "Robust perfect adaptation in bacterial chemotaxis through integral feedback control." In: *Proceedings of the National Academy of Sciences USA* 97.9 (Apr. 2000), pp. 4649–4653.

[197] Dale Muzzey et al. "A systems-level analysis of perfect adaptation in yeast osmoregulation." In: *Cell* 138.1 (July 2009), pp. 160–171.

[198] T S Gardner, C R Cantor, and J J Collins. "Construction of a genetic toggle switch in Escherichia coli." In: *Nature* 403.6767 (Jan. 2000), pp. 339–342.

[199] Ertugrul M Ozbudak et al. "Multistability in the lactose utilization network of Escherichia coli." In: *Nature* 427.6976 (Feb. 2004), pp. 737–740.

[200] János Z Kelemen et al. "Spatial epigenetic control of mono- and bistable gene expression." In: *PLoS biology* 8.3 (Mar. 2010), e1000332.

[201] Gürol M Süel et al. "An excitable gene regulatory circuit induces transient cellular differentiation." In: *Nature* 440.7083 (Mar. 2006), pp. 545–550.

[202] Vijay Chickarmane, Boris N Kholodenko, and Herbert M Sauro. "Oscillatory dynamics arising from competitive inhibition and multisite phosphorylation." In: *Journal of theoretical biology* 244.1 (Jan. 2007), pp. 68–76.

[203] Yuval Hart and Uri Alon. "The utility of paradoxical components in biological circuits." In: *Molecular cell* 49.2 (Jan. 2013), pp. 213–221.

[204] Ling Yang et al. "Multisite phosphorylation and network dynamics of cyclin-dependent kinase signaling in the eukaryotic cell cycle." In: *Biophysical journal* 86.6 (June 2004), pp. 3432–3443.

[205] Munia Amin et al. "Phosphate sink containing two-component signaling systems as tunable threshold devices." In: *PLoS Computational Biology (PLOSCB) 6(11)* 10.10 (Oct. 2014), e1003890.

[206] Varun B Kothamachu et al. "Phosphorelays provide tunable signal processing capabilities for the cell." In: *PLoS Computational Biology (PLOSCB) 6(11)* 9.11 (2013), e1003322.

[207] Jatin Narula et al. "Ultrasensitivity of the Bacillus subtilis sporulation decision." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.50 (Dec. 2012), E3513–22.

[208] Nicolas E Buchler and Matthieu Louis. "Molecular titration and ultrasensitivity in regulatory networks." In: *Journal of molecular biology* 384.5 (Dec. 2008), pp. 1106–1119.

[209] Isaac Salazar-Ciudad and Miquel Marín-Riera. "Adaptive dynamics under development-based genotype-phenotype maps." In: *Nature* 497.7449 (May 2013), pp. 361–364.

[210] Orkun S Soyer, Hiroyuki Kuwahara, and Attila Csikász-Nagy. "Regulating the total level of a signaling protein can vary its dynamics in a range from switch like ultrasensitivity to adaptive responses." In: *The FEBS journal* 276.12 (June 2009), pp. 3290–3298.

[211] Hiroyuki Kuwahara and Orkun S Soyer. "Bistability in feedback circuits as a byproduct of evolution of evolvability." In: *Molecular systems biology* 8 (2012), p. 564.

[212] Song Feng et al. "BioJazz: in silico evolution of cellular networks with unbounded complexity using rule-based modeling." In: *Nucleic acids research* (June 2015), gkv595.

[213] James E Ferrell and Sang Hoon Ha. "Ultrasensitivity part II: multisite phosphorylation, stoichiometric inhibitors, and positive feedback." In: *Trends in biochemical sciences* 39.11 (Nov. 2014), pp. 556–569.

[214] Leon O Murphy and John Blenis. "MAPK signal specificity: the right place at the right time." In: *Trends in biochemical sciences* 31.5 (May 2006), pp. 268–275.

[215] Christina Kiel and Luis Serrano. "Challenges ahead in signal transduction: MAPK as an example." In: *Current Opinion in Biotechnology* 23.3 (June 2012), pp. 305–314.

[216] Franziska Witzel, Louise Maddison, and Nils Blüthgen. "How scaffolds shape MAPK signaling: what we know and opportunities for systems approaches". In: *Frontiers in physiology* 3 (2012).

[217] A M Stock, V L Robinson, and P N Goudreau. "Two-component signal transduction." In: *Annual review of biochemistry* 69.1 (2000), pp. 183–215.

[218] Rui Alves and Michael A Savageau. "Comparative analysis of prototype two-component systems with either bifunctional or monofunctional sensors: differences in molecular structure and physiological function." In: *Molecular microbiology* 48.1 (Apr. 2003), pp. 25–51.

[219] D C LaPorte and D E Koshland. "A protein with kinase and phosphatase activities involved in regulation of tricarboxylic acid cycle." In: *Nature* 300.5891 (Dec. 1982), pp. 458–460.

[220] Fernando Ortega et al. "Product dependence and bifunctionality compromise the ultrasensitivity of signal transduction cascades." In: *Proceedings of the National Academy of Sciences USA* 99.3 (Feb. 2002), pp. 1170–1175.

[221] Jeffrey M Skerker et al. "Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis." In: *PLoS biology* 3.10 (Oct. 2005), e334.

[222] Nils Blüthgen et al. "Effects of sequestration on signal transduction cascades." In: *The FEBS journal* 273.5 (Mar. 2006), pp. 895–906.

[223] Bruno M C Martins and Peter S Swain. "Trade-Offs and Constraints in Allosteric Sensing." In: *PLoS Computational Biology (PLOSCB) 6(11)* 7.11 (2011), e1002261.

[224] Matthew C Good, Jesse G Zalatan, and Wendell A Lim. "Scaffold proteins: hubs for controlling the flow of cellular information." In: *Science (New York, N.Y.)* 332.6030 (May 2011), pp. 680–686.

[225] András Zeke et al. “Scaffolds: interaction platforms for cellular signalling circuits.” In: *Trends in cell biology* 19.8 (Aug. 2009), pp. 364–374.

[226] Yuval Hart et al. “Robust control of nitrogen assimilation by a bifunctional enzyme in E. coli.” In: *Molecular cell* 41.1 (Jan. 2011), pp. 117–127.

[227] Jesse G Zalatan et al. “Conformational control of the Ste5 scaffold protein insulates against MAP kinase misactivation.” In: *Science (New York, N.Y.)* 337.6099 (Sept. 2012), pp. 1218–1222.

[228] Nan Hao et al. “Regulation of cell signaling dynamics by the protein kinase-scaffold Ste5.” In: *Molecular cell* 30.5 (June 2008), pp. 649–656.

[229] Stephen A Chapman and Anand R Asthagiri. “Quantitative effect of scaffold abundance on signal propagation.” In: *Molecular systems biology* 5.1 (2009), p. 313.

[230] Patrick Guye et al. “Rapid, modular and reliable construction of complex mammalian gene circuits.” In: *Nucleic acids research* 41.16 (Sept. 2013), e156–e156.

[231] Weston R Whitaker et al. “Engineering robust control of two-component system phosphotransfer using modular scaffolds.” In: *Proceedings of the National Academy of Sciences of the United States of America* 109.44 (Oct. 2012), pp. 18090–18095.

[232] Wei Sha et al. “Hysteresis drives cell-cycle transitions in Xenopus laevis egg extracts.” In: *Proceedings of the National Academy of Sciences USA* 100.3 (Feb. 2003), pp. 975–980.

[233] Joseph R Pomerening, Eduardo D Sontag, and James E Ferrell. “Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2.” In: *Nature cell biology* 5.4 (Apr. 2003), pp. 346–351.

[234] Onn Brandman et al. “Interlinked fast and slow positive feedback loops drive reliable cell decisions.” In: *Science (New York, N.Y.)* 310.5747 (Oct. 2005), pp. 496–498.

[235] Jayajit Das et al. “Digital signaling and hysteresis characterize ras activation in lymphoid cells.” In: *Cell* 136.2 (Jan. 2009), pp. 337–351.

[236] Wen Xiong and James E Ferrell Jr. “A positive-feedback-based bistable ‘memory module′ that governs a cell fate decision”. In: *Nature* 426.6965 (Nov. 2003), pp. 460–465.

[237] Martin Feinberg. “Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency zero and deficiency one theorems”. In: *Chemical Engineering Science* 42.10 (Jan. 1987), pp. 2229–2268.

[238] Martin Feinberg. “Chemical reaction network structure and the stability of complex isothermal reactors—II. Multiple steady states for networks of deficiency one”. In: *Chemical Engineering Science* 43.1 (Jan. 1988), pp. 1–25.

[239] Gheorghe Craciun and Martin Feinberg. “Multiple Equilibria in Complex Chemical Reaction Networks: I. The Injectivity Property.” In: *SIAM Journal of Applied Mathematics ()* 65.5 (2005), pp. 1526–1546.

[240] Gheorghe Craciun and Martin Feinberg. “Multiple Equilibria in Complex Chemical Reaction Networks: II. The Species-Reaction Graph.” In: *SIAM Journal of Applied Mathematics ()* 66.4 (2006), pp. 1321–1338.

[241] Carsten Conradi et al. “Subnetwork analysis reveals dynamic features of complex (bio)chemical networks.” In: *Proceedings of the National Academy of Sciences of the United States of America* 104.49 (Dec. 2007), pp. 19175–19180.

[242] Gürol M Süel et al. “Tunability and noise dependence in differentiation dynamics.” In: *Science (New York, N.Y.)* 315.5819 (Mar. 2007), pp. 1716–1719.

[243] Tolga Çağatay et al. “Architecture-dependent noise discriminates functionally analogous differentiation circuits.” In: *Cell* 139.3 (Oct. 2009), pp. 512–522.

[244] Murat Acar, Jerome T Mettetal, and Alexander van Oudenaarden. “Stochastic switching as a survival strategy in fluctuating environments.” In: *Nature genetics* 40.4 (Apr. 2008), pp. 471–475.

[245] Guang Yao et al. “Origin of bistability underlying mammalian cell cycle entry.” In: *Molecular systems biology* 7.1 (Apr. 2011), pp. 485–485.

[246] Walter Kolch et al. “The dynamic control of signal transduction networks in cancer cells.” In: *Nature reviews. Cancer* 15.9 (Sept. 2015), pp. 515–527.

[247] Paul Nurse. “Life, logic and information”. In: *Nature* 454.7203 (July 2008), pp. 424–426.

[248] R N Germain. “The art of the probable: system control in the adaptive immune system.” In: *Science (New York, N.Y.)* 293.5528 (July 2001), pp. 240–245.

[249] Anselm Levskaya et al. “Spatiotemporal control of cell signalling using a light-switchable protein interaction.” In: *Nature* 461.7266 (Oct. 2009), pp. 997–1001.

[250] Anselm Levskaya et al. “Synthetic biology: engineering Escherichia coli to see light.” In: *Nature* 438.7067 (Nov. 2005), pp. 441–442.

[251] Jason S Park et al. “Synthetic control of mammalian-cell motility by engineering chemotaxis to an orthogonal bioinert chemical signal.” In: *Proceedings of the National Academy of Sciences of the United States of America* 111.16 (Apr. 2014), pp. 5896–5901.

[252] Jeremy B Chang and James E Ferrell. “Mitotic trigger waves and the spatial coordination of the Xenopus cell cycle.” In: *Nature* 500.7464 (Aug. 2013), pp. 603–607.

[253] Murat Acar, Attila Becskei, and Alexander van Oudenaarden. “Enhancement of cellular memory by reducing stochastic transitions.” In: *Nature* 435.7039 (May 2005), pp. 228–232.

[254] Lei Dai, Kirill S Korolev, and Jeff Gore. “Relation between stability and resilience determines the performance of early warning signals under different environmental drivers.” In: *Proceedings of the National Academy of Sciences of the United States of America* 112.32 (Aug. 2015), pp. 10056–10061.

[255] Oliver Kotte et al. “Phenotypic bistability in Escherichia coli’s central carbon metabolism.” In: *Molecular systems biology* 10.7 (2014), pp. 736–736.

[256] J Barrett Deris et al. “The innate growth bistability and fitness landscapes of antibiotic-resistant bacteria.” In: *Science (New York, N.Y.)* 342.6162 (Nov. 2013), pp. 1237435–1237435.

[257] J Hasty et al. “Noise-based switches and amplifiers for gene expression.” In: *Proceedings of the National Academy of Sciences USA* 97.5 (Feb. 2000), pp. 2075–2080.

[258] M Thattai and A van Oudenaarden. “Intrinsic noise in gene regulatory networks.” In: *Proceedings of the National Academy of Sciences USA* 98.15 (July 2001), pp. 8614–8619.

[259] Thomas M Norman et al. “Stochastic Switching of Cell Fate in Microbes.” In: *Annual review of microbiology* 69.1 (Aug. 2015), p. 150902154308007.

[260] Alejandra C Ventura et al. “Signaling properties of a covalent modification cycle are altered by a downstream target.” In: *Proceedings of the National Academy of Sciences of the United States of America* 107.22 (June 2010), pp. 10032–10037.

[261] Alejandra C Ventura, Jacques-A Sepulchre, and Sofía D Merajver. "A Hidden Feedback in Signaling Cascades Is Revealed." In: *PLoS Computational Biology (PLOSCB) 6(11)* 4.3 (2008), e1000041.

[262] Liming Wang and Eduardo D Sontag. "On the number of steady states in a multiple futile cycle." In: *Journal of mathematical biology* 57.1 (July 2008), pp. 29–52.

[263] Mardo Kõivomägi et al. "Cascades of multisite phosphorylation control Sic1 destruction at the onset of S phase." In: *Nature* 480.7375 (Dec. 2011), pp. 128–131.

[264] Jesper V Olsen et al. "Global, in vivo, and site-specific phosphorylation dynamics in signaling networks." In: *Cell* 127.3 (Nov. 2006), pp. 635–648.

[265] Mardo Kõivomägi et al. "Dynamics of Cdk1 substrate specificity during the cell cycle." In: *Molecular cell* 42.5 (June 2011), pp. 610–623.

[266] Mardo Kõivomägi et al. "Multisite phosphorylation networks as signal processors for Cdk1." In: *Nature structural & molecular biology* 20.12 (Dec. 2013), pp. 1415–1424.

[267] Mart Loog and David O Morgan. "Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates." In: *Nature* 434.7029 (Mar. 2005), pp. 104–108.

[268] Denise A McGrath et al. "Cks confers specificity to phosphorylation-dependent CDK signaling pathways." In: *Nature structural & molecular biology* 20.12 (Dec. 2013), pp. 1407–1414.

[269] Matthew Good et al. "The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation." In: *Cell* 136.6 (Mar. 2009), pp. 1085–1097.

[270] Hao Song and Lingchong You. "Evolving sensitivity." In: *ACS chemical biology* 1.11 (Dec. 2006), pp. 681–682.

[271] Song Feng, Julien F Ollivier, and Orkun S Soyer. "Enzyme sequestration as a tuning point in controlling response dynamics of signalling networks". In: *PLoS Computational Biology, under review* (2016).

[272] Elisenda Feliu et al. "An algebraic approach to signaling cascades with N layers." In: *Bulletin of mathematical biology* 74.1 (Jan. 2012), pp. 45–72.

[273] Balázs Boros. "On the existence of the positive steady states of weakly reversible deficiency-one mass action systems." In: *Mathematical biosciences* 245.2 (Oct. 2013), pp. 157–170.

[274] Martin Feinberg. "Multiple steady states for chemical reaction networks of deficiency one". In: *Archive for Rational Mechanics and Analysis* 132.4 (1995), pp. 371–406.

[275] Hesam N Motlagh et al. "The ensemble nature of allostery." In: *Nature* 508.7496 (Apr. 2014), pp. 331–339.

[276] Vincent J Hilser, James O Wrabl, and Hesam N Motlagh. "Structural and Energetic Basis of Allostery". In: *Annual review of biophysics* 41.1 (June 2012), pp. 585–609.

[277] Janine Mok et al. "Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs." In: *Science signaling* 3.109 (2010), ra12–ra12.

[278] G N Lewis. "A new principle of equilibrium." In: *Proceedings of the National Academy of Sciences of the United States of America* 11.3 (Mar. 1925), pp. 179–183.

[279] D Walz and S R Caplan. "Consequences of detailed balance in a model for sensory adaptation based on ligand-induced receptor modification." In: *Proceed-*

ings of the National Academy of Sciences USA 84.17 (Sept. 1987), pp. 6152–6156.

[280] R H Fowler and E A Milne. "A note on the principle of detailed balancing." In: *Proceedings of the National Academy of Sciences of the United States of America* 11.7 (July 1925), pp. 400–402.

[281] O G Berg, J Paulsson, and M Ehrenberg. "Fluctuations in repressor control: thermodynamic constraints on stochastic focusing." In: *Biophysical journal* 79.6 (Dec. 2000), pp. 2944–2953.

[282] David O Morgan. "Principles of CDK regulation". In: *Nature* 374.6518 (Mar. 1995), pp. 131–134.

[283] David Owen Morgan. *The Cell Cycle*. Principles of Control. New Science Press, 2007.

[284] Samyabrata Bhaduri et al. "A docking interface in the cyclin Cln2 promotes multi-site phosphorylation of substrates and timely cell-cycle entry." In: *Current biology : CB* 25.3 (Feb. 2015), pp. 316–325.

[285] Z Tang, T R Coleman, and W G Dunphy. "Two distinct mechanisms for negative regulation of the Wee1 protein kinase." In: *The EMBO Journal* 12.9 (Sept. 1993), pp. 3427–3436.

[286] T R Coleman, Z Tang, and W G Dunphy. "Negative regulation of the wee1 protein kinase by direct action of the nim1/cdr1 mitotic inducer." In: *Cell* 72.6 (Mar. 1993), pp. 919–929.

[287] N Watanabe, M Broome, and T Hunter. "Regulation of the human WEE1Hu CDK tyrosine 15-kinase during the cell cycle." In: *The EMBO Journal* 14.9 (May 1995), pp. 1878–1891.

[288] Carlo Chan et al. “Protein scaffolds can enhance the bistability of multisite phosphorylation systems.” In: *PLoS Computational Biology (PLOSCB) 6(11)* 8.6 (2012), e1002551.

[289] John E Dueber, Ethan A Mirsky, and Wendell A Lim. “Engineering synthetic signaling proteins with ultrasensitive input/output control.” In: *Nature biotechnology* 25.6 (June 2007), pp. 660–662.

[290] John E Dueber et al. “Synthetic protein scaffolds provide modular control over metabolic flux.” In: *Nature biotechnology* 27.8 (Aug. 2009), pp. 753–759.

[291] F Horn and R Jackson. “General mass action kinetics”. In: *Archive for Rational Mechanics and Analysis* 47.2 (1972), pp. 81–116.

[292] Martin Feinberg. “Complex balancing in general kinetic systems”. In: *Archive for Rational Mechanics and Analysis* 49.3 (1972), pp. 187–194.

[293] Elisenda Feliu and Carsten Wiuf. “A computational method to preclude multistationarity in networks of interacting species.” In: *Bioinformatics (Oxford, England)* 29.18 (Sept. 2013), pp. 2327–2334.

[294] Elisenda Feliu and Carsten Wiuf. “Finding the positive feedback loops underlying multi-stationarity.” In: *BMC systems biology* 9.1 (2015), p. 22.

[295] M Kaufman, C Soulé, and R Thomas. “A new necessary condition on interaction graphs for multistationarity.” In: *Journal of theoretical biology* 248.4 (Oct. 2007), pp. 675–685.

[296] Carsten Wiuf and Elisenda Feliu. “Power-Law Kinetics and Determinant Criteria for the Preclusion of Multistationarity in Networks of Interacting Species.” In: *SIAM J. Applied Dynamical Systems ()* 12.4 (2013), pp. 1685–1721.

[297] Stefan Müller et al. “Sign Conditions for Injectivity of Generalized Polynomial Maps with Applications to Chemical Reaction Networks and Real

Algebraic Geometry". In: *Foundations of Computational Mathematics* (Jan. 2015), pp. 1–29.

[298] David B Fogel. *Evolutionary Computation.* The Fossil Record. Wiley-IEEE Press, May 1998.

[299] Daniel M Weinreich et al. "Darwinian evolution can follow only very few mutational paths to fitter proteins." In: *Science (New York, N.Y.)* 312.5770 (Apr. 2006), pp. 111–114.

[300] Isaac Salazar-Ciudad and Jukka Jernvall. "A computational model of teeth and the developmental origins of morphological variation." In: *Nature* 464.7288 (Mar. 2010), pp. 583–586.

[301] Domitilla Del Vecchio, Alexander J Ninfa, and Eduardo D Sontag. "Modular cell biology: retroactivity and insulation." In: *Molecular systems biology* 4.1 (2008), p. 161.

[302] John P Barton and Eduardo D Sontag. "The energy costs of insulators in biochemical networks." In: *Biophysical journal* 104.6 (Mar. 2013), pp. 1380–1390.

[303] Michael A Rowland, Walter Fontana, and Eric J Deeds. "Crosstalk and competition in signaling networks." In: *Biophysical journal* 103.11 (Dec. 2012), pp. 2389–2398.

[304] Haruo Saito. "Regulation of cross-talk in yeast MAPK signaling pathways." In: 13.6 (Dec. 2010), pp. 677–683.

[305] Monica A Schwartz and Hiten D Madhani. "Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae." In: *Annual review of genetics* 38.1 (2004), pp. 725–748.

[306] Marcelo Behar, Henrik G Dohlman, and Timothy C Elston. "Kinetic insulation as an effective mechanism for achieving pathway specificity in intracellu-

lar signaling networks." In: *Proceedings of the National Academy of Sciences USA* 104.41 (Oct. 2007), pp. 16146–16151.

[307] Avraham E Mayo et al. "Plasticity of the cis-regulatory input function of a gene." In: *PLoS biology* 4.4 (Apr. 2006), e45.

[308] Richard Harrison et al. "Plasticity of genetic interactions in metabolic networks of yeast." In: *Proceedings of the National Academy of Sciences USA* 104.7 (Feb. 2007), pp. 2307–2312.

[309] Sam P Brown, Daniel M Cornforth, and Nicole Mideo. "Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control." In: *Trends in microbiology* 20.7 (July 2012), pp. 336–342.

[310] Ryan Suderman and Eric J Deeds. "Machines vs. Ensembles: Effective MAPK Signaling through Heterogeneous Sets of Protein Complexes." In: *PLoS Computational Biology (PLOSCB) 6(11)* 9.10 (2013), e1003278.

[311] Adam Filipczyk et al. "Network plasticity of pluripotency transcription factors in embryonic stem cells." In: *Nature cell biology* 17.10 (Sept. 2015), pp. 1235–1246.

[312] D Bray. "Protein molecules as computational elements in living cells." In: *Nature* 376.6538 (July 1995), pp. 307–312.

[313] Roby P Bhattacharyya et al. "Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits." In: *Annual review of biochemistry* 75.1 (2006), pp. 655–680.

[314] Peter Atkins and Julio de Paula. *Atkins' Physical Chemistry.* Oxford University Press, Mar. 2014.

[315] Badal Joshi and Anne Shiu. "Atoms of multistationarity in chemical reaction networks". In: *Journal of Mathematical Chemistry* 51.1 (Aug. 2012), pp. 153–178.

[316] Carsten Conradi and Maya Mincheva. “Catalytic constants enable the emergence of bistability in dual phosphorylation.” In: *Journal of the Royal Society, Interface / the Royal Society* 11.95 (June 2014), pp. 20140158–20140158.

[317] Elisenda Feliu and Carsten Wiuf. “Simplifying biochemical models with intermediate species.” In: *Journal of the Royal Society, Interface / the Royal Society* 10.87 (Oct. 2013), pp. 20130484–20130484.