

Intracellular Network
Attractor Selection
and the Problem of Cell
Fate Decision

Insights into Cancer Network Multitargeting

Nuno Rocha Nene

A thesis submitted to the University College London for the degree of Doctor of
Philosophy
CoMPLEX and UCL Cancer Institute
University College London

To my parents and my sister

Abstract

This project aims at understanding how cell fate decision emerges from the overall intracellular network connectivity and dynamics. To achieve this goal both small paradigmatic signalling-gene regulatory networks and their generalization to highdimensional space were tested. Particularly, we drew special attention to the importance of the effects of time varying parameters in the decision genetic switch with external stimulation. The most striking feature of our findings is the clear and crucial impact of the rate with which the time-dependent parameters are changed. In the presence of small asymmetries and fluctuations, slow passage through the critical region increases substantially specific attractor selection by external transient perturbations. This has strong implications for the cell fate decision problem since cell phenotype in stem cell differentiation, cell cycle progression, or apoptosis studies, has been successfully identified as attractors of a whole network expression process induced by signalling events. Moreover, asymmetry and noise naturally exist in any integrative intracellular decision network. To further clarify the importance of the rate of parameter sweeping, we also studied models from non-equilibrium systems theory. These are traditional in the study of phase transitions in statistical physics and stood as a fundamental tool to extrapolate key results to intracellular network dynamics. Specifically, we analysed the effects of a time-dependent asymmetry in the canonical supercritical pitchfork bifurcation model, both by numerical simulations and analytical solutions. We complemented the discussion of cell fate decision with a study of the effects of non-specific targets of drugs on the Epidermal Growth Factor Receptor pathway. Pathway output has long been correlated with qualitative cell phenotype. Cancer network multitargeting therapies were assessed in the context of whole network attractor phenotypes and the importance of parameter sweeping speed.

Declaration

I hereby declare that the work described in this thesis is the result of my own independent investigation, unless otherwise stated. Chapter 4 was mostly performed in collaboration with Dr.Sylvia Nagl (UCL). The work in chapters 2 and 3 were performed under the supervision of Professor Alexey Zaikin (Department of Mathematics and Institute for Women's Health, UCL).

This work has not been previously submitted for any degree and is not being concurrently submitted in candidature for any other degree.

Nuno Rocha Nene'

Table of contents

Intracellular Network Attractor Selection	1
and the Problem of Cell Fate Decision	1
Abstract	3
Declaration.....	4
Table of contents.....	5
List of Figures	11
List of Tables	16
Acknowledgements.....	17
1 Introduction	18
1.1 Thesis overview	18
1.2 Integrative intracellular network function: review of most common mechanisms	19
1.2.1 Typical mechanisms of intracellular signal transduction: the main players of the Epidermal Growth Factor Pathway.....	20
1.2.2 Cycle and cascade motifs in signalling networks.....	30
1.2.3 Modularity.....	31
1.2.4 Intracellular integration of signals and cross-talk between pathways: the mammalian mitogen activated pathway kinases (MAPK)	33
1.2.5 Transient vs. sustained activation and early gene induction: linking signals to transcriptional activity	37
1.2.6 Control of gene expression.....	39

1.3	Multitargeting therapies: network connectivity can be used to develop possible strategies.....	46
1.3.1	Directed and undirected graphs as a tool for qualitative intracellular network description.....	48
1.3.2	Global topology of networks, robustness and targeting strategies	51
1.3.3	Multiple-targeting or multiple drugs with individual single targets?	55
1.4	Dynamical systems modelling approaches in intracellular network biology.....	56
1.4.1	Continuous systems models.....	58
1.4.2	Qualitative systems models.....	62
1.4.3	Stochastic systems models.....	64
1.5	Epigenetic attractor landscapes.....	67
1.5.1	Cell phenotypes as genetic network attractors.....	68
1.5.2	Deterministic and stochastic attractor selection.....	72
1.6	Project goals.....	77

2 Attractor selection in integrative noisy signalling-transcriptional regulatory networks with time-scale separation 79

2.1	Introduction.....	79
2.2	Symmetry breaking in parametrically driven far-from equilibrium systems with time-dependent external fields.....	83
2.2.1	Critical fluctuations and long time-scales.....	88
2.2.2	Bifurcation delay and dynamical hysteresis.....	90
2.2.3	Bistable potential with time-dependent critical parameter and external asymmetry.....	91
2.2.4	Branch selectivity in the presence of additive noise: analytical results	98
2.2.5	Branch selectivity in the presence of additive noise: numerical results	119

2.3	Response of a gene regulatory decision switch to external signalling inputs.....	123
2.3.1	Model.....	127
2.3.2	Symmetry breaking and attractor selectivity by time-dependent external signals in a fast/slow stochastic cell fate decision regulatory network: numerical results	131
2.4	Summary.....	142
3	Path-dependent pattern selection effects in high-dimensional integrative signalling-gene regulatory intracellular networks .	143
3.1	Introduction	143
3.1.1	Modelling the highdimensional switch	145
3.2	Clustering of input signal combinations by the highdimensional decision switch	154
3.3	Attractor selectivity in the presence of multiplicative noise	156
3.3.1	Inter-distribution distance depends on noise	161
3.3.2	Inter-distribution distance depends on sweeping speed	165
3.4	Summary.....	168
4	Non-specific effects of small molecule kinase inhibitor therapies: insights into synergistic effects of multiple targeting	169
4.1	Introduction	169
4.2	Targeting the Systems Biology paradigmatic Epidermal Growth Factor Receptor Pathway with documented small molecule kinase inhibitor therapies.....	170
4.2.1	EGFR/MAPK pathway dynamics: focus on Schoeberl et al (2002) model.	170

4.2.2	Tyrosine Kinase Inhibitor non-specific interactions and its influence on EGFR pathway output.	177
4.3	Extending the study of non-specific small molecule kinase interactions to a larger integrative network: the Human Signalling network and the FANTOM4 regulatory network.....	204
4.3.2	Modelling the extended Human signalling network.....	207
4.4	Developing inhibition multitargeting signatures inducing the appropriate expression patterns	212
4.5	Summary.....	219
5	Summary of key results and discussion	220
5.1	Sensitivity to transient external signals in the presence of noise depends on parameter sweeping rates.....	220
5.1.1	Transient external asymmetries in supercritical pitchfork bifurcations induce higher selectivities in the vicinity of the critical region .	220
5.1.2	Combination of slow varying external signals induce symmetry breaking and increase attractor selectivity in signalling-gene regulatory decision networks	220
5.1.3	Time-scale separation decreases sensitivity to differences in external signals in the presence of fluctuations	221
5.1.4	Paradigmatic cell fate decision models for crucial mechanism identification.....	221
5.1.5	Highdimensional attractor selectivity depends strongly on structure of phase space	222
5.2	Effects of non-specificity of drugs may inform the development of relevant multitargeting therapies.....	223
5.2.1	Epidermal Growth Factor Receptor (EGFR) pathway output is affected by documented non-specific interactions	223

5.2.2	Multitargeting strategies are expected to induce the correct expression program	224
6	Conclusion.....	225
7	Future work.....	226
7.1	Decision genetic switch with external stimulation	226
7.2	Highdimensional networks with coexistence of dynamical and point attractors	227
8	Annex.....	232
8.1	Control coefficients for the Epidermal Growth Factor Receptor Network.....	232
8.2	Inhibition signature maps for small molecule kinase inhibitors ..	234
8.3	Binding affinities for small molecule kinase inhibitors.....	235
8.4	Numerical integration method	236
8.5	Selectivity expressions for the supercritical pitchfork bifurcation for specific cases of external asymmetry $g(t)$ and critical parameter $\lambda(t)$..	237
8.5.1	Maximum asymmetry before the bifurcation point	237
8.5.2	Maximum asymmetry after the bifurcation point	238
8.6	Parameter space analysis of the gene regulatory decision switch	240
8.6.1	Isolated genetic decision switch	240
8.6.2	Genetic decision switch stimulated by a combination of signalling input amplitudes.....	247
8.7	Bifurcation diagrams for specific input combinations to the highdimensional decision switch.....	251
8.7.1	Input combination I_{15}	252

8.7.2	Input combination I_{75}	253
8.7.3	Input combination I_{94}	254
8.8	Final distributions for the highdimensional switch for input combination I_{15} , I_{75} , I_{94} in the presence of fluctuations	255
8.8.1	Distribution across attractors for I_{15} , I_{75} , I_{94} with noise intensity $\sigma = 0.01$	256
8.8.2	Distribution across attractors for I_{15} , I_{75} , I_{94} with noise intensity $\sigma=0.5$	257
9	Reference List	258

List of Figures

Figure 1 Simplified representation of intercellular communication through soluble messengers and intracellular signalling response..	21
Figure 2 Simplified Epidermal Growth Factor Receptor pathway.	22
Figure 3 Typical motifs of intracellular signalling cascades.	31
Figure 4 Crosstalk between two pathways affecting downstream transcriptional activity.....	34
Figure 5 Cross-talk between mitogen activated protein kinase networks, transcription factors activated and the induction of early genes (IEG) in response to extracellular signals.....	36
Figure 6 Hardware gate abstraction of the cis-regulatory region of a gene.....	40
Figure 7 Generic representation of the process of activation (phosphorylation) of a transcription factor, cooperation between activated TF monomers and a co-activator TF to induce transcription and transport of both mRNAs and proteins.	42
Figure 8 The signalling system as a combinatorial decoder.	45
Figure 9 Representation under the graph formalism of a gene network or a network involving both genes and signalling proteins.....	49
Figure 10 Common structural motifs found in intracellular network.....	50
Figure 11 Yeast (<i>Saccharomyces cerevisiae</i>) protein interaction network.	52
Figure 12 Graphical depiction of possible strategies for targeting networks.	54
Figure 14 Hill function profile for several hill coefficients and Michaelis-Menten constants.	60
Figure 15 Simplified depiction of the possible outcomes or cell phenotypes emerging from a stem cell.....	68
Figure 16 Simplified graphical representation of highdimensional state space attractor landscapes..	70
Figure 17 Differentiation of progenitor cells into erythroid and myeloid/monocytic cell phenotypes.	72

Figure 18 Potential for a hypothetical gene regulatory decision switch (Figure 19) driven by external signals (S_1 and S_2 , see also Equation 14) leading to attractor selection.	75
Figure 19 Generic integrative signalling-transcriptional system.....	81
Figure 20 Generic representative signal shapes for incoming signals to the transcriptional network.	82
Figure 21 Attractors available for varying values of the critical parameter λ in canonical models of phase transitions.	84
Figure 22 Critical slowing down observed at a supercritical Pitchfork bifurcation or second order phase transition.	89
Figure 23 Numerical confirmation of bifurcation delay and dynamical hysteresis for a supercritical Pitchfork bifurcation.	91
Figure 24 Bifurcation curves for a one dimensional system.....	92
Figure 25 Critical parameter and asymmetry shapes in time.	93
Figure 26 Effects of a time-dependent asymmetry on the trajectories of X	95
Figure 27 Effects of time dependent asymmetry with unequal ascending and descending rates.	96
Figure 28 Comparative bifurcation diagrams for different parameters of a supercritical pitchfork bifurcation normal form.	98
Figure 29 Examples of trajectories in the presence of noise for sweeping of the critical parameter.	99
Figure 30 Potential profiles for several values of external asymmetry before, at and after the bifurcation point..	103
Figure 31 Probability evolution with time in a system with critical parameter sweeping and constant asymmetry $0 < g < 1$	104
Figure 32 Time-dependent profiles for variables λ and g	107
Figure 33 Profile for error function, $\text{erf}(z)$	111
Figure 34 Dependence of the distribution mean on the instant the maximum asymmetry is reached: before, at, and after the critical bifurcation point $\lambda=0$	114

Figure 35 Branch selection probability dependence on asymmetry and sweeping speed.	118
Figure 36 Probability P_{up} of reaching upper branch as a function of external asymmetry and sweeping speed.	120
Figure 37 Branch selectivity, theory vs. numerical results.	122
Figure 38 Phosphorylation of transcription factors necessary for initiation of transcription	125
Figure 39 Integrative signalling-gene regulatory network.....	126
Figure 40 Bifurcation diagrams and nullclines for the decision genetic switch..	134
Figure 41 Bifurcation diagram for component X with different parameterization, from P_1 to P_m	135
Figure 42 Examples of stochastic and deterministic trajectories for the decision genetic switch..	136
Figure 43 Selectivity vs. asymmetry and sweeping speed for several noise and time-scale ratios.....	138
Figure 44 Probability P_{up} of reaching the upper branch (numerical data) as a function of maximum asymmetry a for several time-scale ratios ($\tau_{TFX,Y}/\tau_{SX,Y}$) and sweeping speeds ($10/T_{S1}$).	140
Figure 45 Representation of a highdimensional genetic decision switch with 10 transcription factors (nodes 6 to 15) and 5 input signals.....	144
Figure 46 Generic representation of signal induction of immediate early gene products (IEGs) and subsequent delayed early gene products (DEGs). ...	150
Figure 47 Bifurcation diagram for each of the transcription factors for $S=S_1=S_2=S_3=S_4=S_5$	153
Figure 48 Pair-wise average distance over 100 runs (each corresponding to a different initial condition) between asymptotically stable states induced by input combinations.	155
Figure 49 Inter-trajectory distance for pairs $(I_k, I_{k'})$ inducing the same attractors.	157
Figure 50 Time-dependent evolution of the highdimensional decision switch circuit.	158

Figure 51 Initial and final attractors' positions induced by I_{15} and respective frequency.....	159
Figure 52 Typical evolution of concentrations for all the nodes TF_i	160
Figure 53 Distributions across attractors in the presence of noise intensity $\sigma=0.05$..	161
Figure 54 Distance between final distributions generated by different pairs of input combinations $(I_k, I_{k'})$ in the presence of fluctuations.	164
Figure 55 Inter-distribution distance dependence on sweeping speed.	168
Figure 56 Concentration profiles for EGFR pathway.	172
Figure 57 Diagram of the molecular species involved in Schoeberl et al model (232)..	174
Figure 58 Temporal profile of ERK-PP (total) activation with the presence of the negative feedback loop from ERK-PP to SOS.	176
Figure 59 Representation of competitive ATP and Inhibitor binding to a kinase.	178
Figure 60 Cross reactivity across the kinome for six small molecule kinase inhibitors that target the EGFR..	179
Figure 61 Profiles for inhibition constants $\varepsilon(n, \alpha)$	182
Figure 62 Geometric representation of the signal characteristics. τ -signalling time. ν -signal duration..	184
Figure 63 Inhibition signature of small molecule kinase inhibitors Gefitinib and Erlotinib.....	185
Figure 64 Vesicle formation and uncoating of clathrin vesicles. GAK is involved in the last process.	187
Figure 65. Condensed representation of control in the EGFR pathway.	188
Figure 66 ERKPP response for Gefitinib, under EGFR normal expression	190
Figure 67 EGFR and ERK-PP profiles for EGFR Inhibition with Gefitinib vs. no Inhibition.	191
Figure 68 Differences in profiles EGFR when the internalization rate and the degradation rate are inhibited..	192

Figure 69 Relative order of signals for 3 consecutive α	193
Figure 70 Gefitinib ERKPP response to inhibition α , under EGFR overexpression.	194
Figure 71 Erlotinib ERKPP response to inhibition α , under normal expression of EGFR.	196
Figure 72 Erlotinib ERKPP response to inhibition α , under over expression of EGFR.	198
Figure 73 Inhibition signature for EKB-569..	199
Figure 74 EKB-569 ERKPP response to inhibition α , under EGFR normal expression..	201
Figure 75 Generic representation of the extra inhibitory interaction considered for the EKB-569 with respect to the MAPK module.....	201
Figure 76 EKB-569 ERKPP response to inhibition α , under overexpression of EGFR.	204
Figure 77. Connectivity distribution (in-degree) for the original human signalling network (64).....	205
Figure 78 Functional representation of integrated signalling-gene regulatory systems.	212
Figure 79 Simplified representation of the possible effects of therapies on the phase space of a cancer cell.	214
Figure 80 Highdimensional synthetic oscillatory genetic networks.	227
Figure 81 Inhibition signatures of 2 other small molecule kinase inhibitors targeting the EGFR.	234
Figure 82 Isolated decision genetic switch.	240
Figure 83 Bifurcation analysis of the symmetric genetic decision switch.	242
Figure 84 Bifurcation diagrams formed by extending the diagrams in Figure 83.	244
Figure 85 Bifurcation diagrams in the presence of asymmetric basal rates.	245
Figure 86 Bifurcation diagrams for asymmetric thresholds of self-activation and cross-inhibition parameters.	246

Figure 87 Bistability for the genetic decision switch with external stimulation and no cross-talk.	248
Figure 88 Multistability for the genetic decision switch with external stimulation and no cross-talk.	249
Figure 89 Effects of cross-talk on bistability.	250
Figure 90 Bifurcation diagram obtained by setting the parameters S_i following the combination of amplitudes inherent to $I_{15}(t)$	252
Figure 91 Bifurcation diagram obtained by setting the parameters S_i following the combination of amplitudes inherent to $I_{75}(t)$	253
Figure 92 Bifurcation diagram obtained by setting the parameters S_i following the combination of amplitudes inherent to $I_{94}(t)$	254
Figure 93 Distributions across attractors in the presence of noise intensity $\sigma=0.01$	256
Figure 94 Distributions across attractors in the presence of noise intensity $\sigma=0.5$	257

List of Tables

Table 1 Order of magnitude for timescales in intracellular networks in human fibroblasts.	29
Table 2 Connectivity matrix between signalling inputs and transcription factors.	151
Table 3 Boolean function table corresponding to the logical OR function with the corresponding minterms.	209
Table 4 Control coefficients for amplitude, duration and integrated response signal (area) for the processes involved in the system.	232
Table 5 Description of the processes and respective rates (125, 232) associated with each line in Table 4.	233
Table 6 Binding constants for the six EGFR inhibitors to a panel of kinases.	235

Acknowledgements

I would like to thank Professor Alexey Zaikin for all the support and stimulating in depth discussions on the applicability of stochastic calculus to intracellular-network dynamics. Professor Zaikin's excellent supervision and insight on the use of non-equilibrium phase transitions theory in biology and the problem of cell fate decision was crucial for the successful completion of the work presented ahead. I also gratefully acknowledge the possibility for all the discussions with Professor Zaikin's collaborators during crucial stages of the project; namely Dr.Ullner and Professor Volkov.

My sincere regards go to Professor Richard Begent for the encouragement and support throughout the duration of the project, for allowing me to use extensively the department's computational resources, and for the commitment to the successful PhD completion.

I also would like to thank Dr. Sylvia Nagl for the biological strategic insight, so crucial in several key points of the initial research period and for the regular discussions that challenged the validity of the assumptions behind each step of the modelling task.

For my colleagues Amos and May I would like to send also my regards and show deep gratitude for helping me with the intricate computational tasks and for the continuous encouragement. I would also like to acknowledge Ed Long and Rubi for their constructive comments during journal club sessions organized by Professor Zaikin.

I would also like to show my gratefulness to CoMPLEX for the financial support and for creating an environment where interdepartmental collaborations were always possible and fundamental for the development of the Systems Biology community at UCL.

Last, but above all, I would like to send my endless regards to my family without whom my PhD studies would not have been possible and to my Edytka whose love makes me stronger.

1 Introduction

This introductory chapter will mostly describe general properties of intracellular networks which will be relevant for the remaining chapters of this Thesis (see Thesis overview).

1.1 *Thesis overview*

Section 1.2 of the introductory chapter will be necessary to contextualize the investigations performed on the paradigmatic Epidermal Growth Factor Receptor Pathway (section 4). Section 1.3 encompasses a discussion on the rationale behind network multitargeting and contextualizes section 4. Section 1.4 reviews important modelling approaches in network systems biology and justifies the modelling techniques used in chapter 2, chapter 3 and chapter 4.

In chapter 2 we perform an extended simulation study and analytical treatment on a parametrically driven one dimensional system with both time-dependent bifurcation parameter and external field. Typical phenomena such as bifurcation delay and dynamical hysteresis will be studied (section 2.2). The effects of time-dependent characteristics of the forcing such as the sweeping speed and amplitude will be extensively analysed. Strong extrapolation of the importance of these effects to the sensitivity of genetic networks to external combinatorial signals will be formulated. Further numerical investigations on the synthetic/systems biology paradigmatic decision genetic switch (section 2.3) will help clarify and consolidate the significance of the speed of passage through the critical region for cell fate commitment, under the presence of transient asymmetries. We will also extend the findings of previous sections (section 2.2 and section 2.3) to a generalization high-dimensional of the decision genetic switch (chapter 3).

In chapter 4 the focus will be shifted to the upstream signalling processing module and the effects, on cell fate decision, of non-specific targeting common in a current class of therapies. A study of approved targeting therapies is performed

on a fully parameterized model of the Epidermal Growth Factor Receptor (EGFR) pathway (section 4). The EGFR and other members of its family have been successfully selected for targeting in therapies against cancer. This family of receptors continues to be a fundamental source of landmark discoveries on cell signalling mechanisms inherent to developmental programs, tissue homeostasis and disease. Small molecule kinase inhibitors of the EGFR have been developed and two of them, Gefitinib and Erlotinib, have already been licensed for clinical use in Non Small Cell Lung Carcinoma (NSCLC). Further investigations will be performed on modelling the dynamics of a large intracellular signalling network known as Map of Human Cancer Signalling (section 4.3.2). Insights into aspects of multitargeted therapies using high-order synergistic effects to circumvent biological robustness will be reported. Connections to the importance of whole network targeting inducing the appropriate gene regulatory network attractor selection will be highlighted.

1.2 Integrative intracellular network function: review of most common mechanisms

Cancer arises from uncontrolled cell proliferation. Several pathways have been observed with mutations that drive this abnormal scenario: EGFR, Ras, AKT/PKB and mTOR, PKC, to name a few (28). Because of this striking feature of abnormal tissue, one might expect that cancer cells reprogram severely, to their own advantage, the control circuitry regulating healthy cells. In fact, the control circuitry used by both types of cells is not drastically different. The outcome of subtle changes is on the other hand striking. Cancerous cells tweak existing controls rather than eliminate or corrupt them completely (115, 221). Several stages are thought to characterize the path to tumourigenesis and carcinogenesis as was elegantly highlighted by Weinberg and coworkers (115): evasion of apoptosis, onset of self-sufficiency in growth signals, insensitivity to anti-growth signals, capacity for tissue invasion and metastasis, unlimited replicative potential and sustained angiogenesis. The “robust yet fragile” current

approach to systems biology, proposed within the “highly optimized tolerance” (HOT) theoretical framework (47-49, 74), demands to be complemented by a “robust yet subtle” (192, 203) understanding of how cancer distributes disruption across pathway networks and how this correlates with the path to tumorigenesis. An adequate characterization of the main features of the signalling process, e.g. type of molecules involved, structure of pathways, basic modules composing intricate pathways and timescales of the inherent dynamical process, is crucial for optimizing therapeutic targeting strategies of networks with distributed disturbance and mutation (64). Recently another interesting approach for characterizing intracellular network functioning and activation in cancer has focused on pathway gene expression signatures (30, 134, 262). This is motivated by the very distributed nature of observed mutations in pathways that demands a further downstream (gene expression level) investigation of system behaviour. This comes in line with one of the problems analysed in this thesis: how does signalling affect transcription and where drugs should strike as to induce the desired expression program. In the following section we will describe the main processes involved in signal transduction, highlighting where reported cancer associated mutations occur and how this affects normal pathway functioning.

1.2.1 Typical mechanisms of intracellular signal transduction: the main players of the Epidermal Growth Factor Pathway

The remarkable structural complexity and functional capacity of multicellular organisms stems from the striking ability of cell-cell biochemical reaction coordination. Currently, several forms of communication between cells are known: cell-cell interaction via surface proteins, via gap junctions, via electrical processes and chemical messengers. Our work focuses on particular aspects of how a cell reacts internally, e.g. which transcriptional program or response is initiated or enhanced, when a chemical stimulus is induced at its surface by chemical messengers (receptor ligands) (Figure 1).

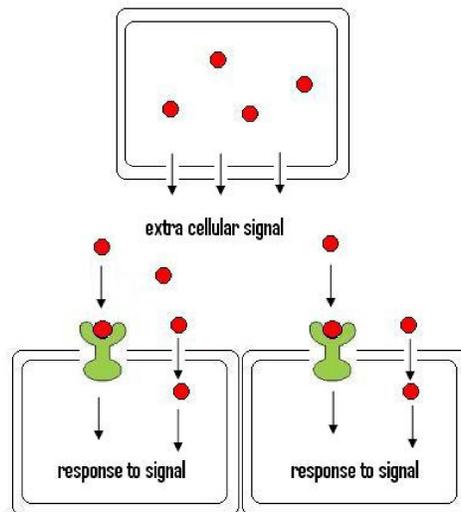


Figure 1 Simplified representation of intercellular communication through soluble messengers and intracellular signalling response. Red filled circles represent soluble messengers. Green components generically represent receptors. Receptor activation and transduction of external signals is a complex process involving other stages that will be described ahead (see also Figure 2).

Cells in higher multicellular systems operate through hundreds of kinds of signalling molecules to establish efficient communication. Among the used signals we have proteins and small peptides to dissolved gases, e.g. nitric oxide and carbon monoxide (6). A great percentage of these molecules are secreted from the cell initiating the signal and migrate, by exocytosis, to the extracellular space. Diffusion through the plasma membrane is another common process. Some others are exposed to the extracellular space but remain strongly attached to the cell's surface.

The sequence of events in the process of extracellular signal generation and response of the target cell to extracellular ligands are roughly the following (see also Figure 1):

- Formation of the signal in the original cell as a consequence of an external trigger;
- Signal transport to the target cell;
- Signal registration on the target cell;

- Signal transduction and information processing through a biochemical network;
- Termination of the signal.

The signal registration and information processing in the target cell is done through intracellular pathways forming an intricate network. The intracellular pathways can be characterized by several aspects:

- Nature of the external trigger signal;
- Mechanism of registration of the signal;
- Mechanism underlying the complex web responsible for signal transmission and termination;
- Nature of the reaction promoted in the target cell.

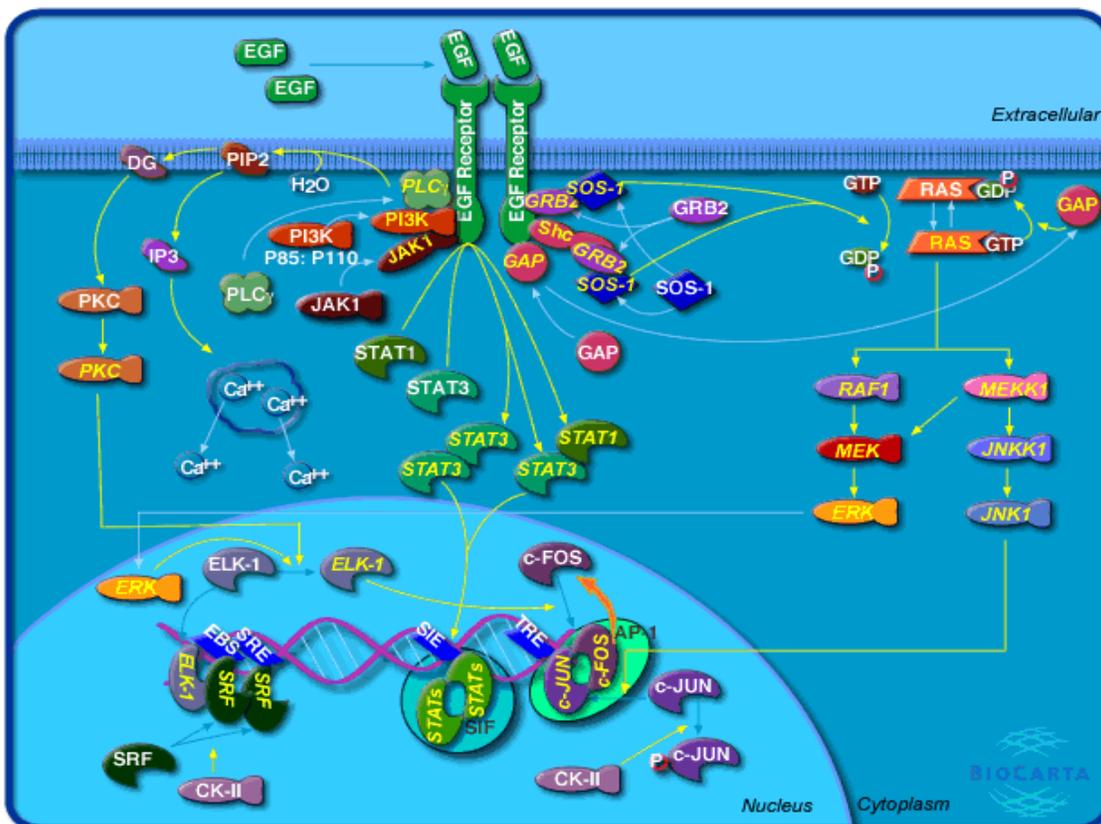


Figure 2 Simplified Epidermal Growth Factor Receptor pathway. Only main players and main processes are shown. For details see www.biocarta.com .

The reception of the extracellular signal is performed by specialized proteins termed receptors. The largest classes of receptors commonly identified are G-protein linked receptors, ion-channel linked receptors and enzyme linked receptors (6). Extracellular signalling molecules commonly activate, at very low concentrations ($[L] \leq 10^{-8}$ M), their respective target cells. The receptors that recognize these molecules typically bind them with high affinity, usually with $K_d \geq 10^8$ l/mole (6). Binding of an extracellular molecule induces activation of, for example, a transmembrane receptor and, subsequently, of a cascade of intracellular events that eventually affect the cell's behaviour. Other cases exist where we have receptors inside the target cell. In these particular scenarios the signal molecule, suitably small and hydrophobic, has to enter the cell to activate them (6).

Two important classes of receptors, G-protein-linked and enzyme-linked, after being stimulated by extracellular signals relay the information into the cell interior through a wealth of small and large intracellular signalling molecules. Figure 2 shows a paradigmatic cascade of reactions for an enzyme-linked pathway, the Epidermal Growth Factor Receptor (EGFR) network. The chain of intracellular signaling events operates on specific target proteins such as transcription factors that alter gene expression, metabolic enzymes or a cytoskeletal protein inducing cell shape changes and/or movement. Receptor tyrosine kinases (RTKs), i.e. enzyme-linked, play an essential role in several cellular functions, from regulation of embryogenesis and cell survival to motility. Their malfunction is considered to be one of the most common causes of various diseases such as: cancer, chronic inflammatory syndromes and diabetes (10, 56, 232, 276). We will focus throughout this thesis, and particular in chapter 4, on enzyme-linked pathways, particularly on the EGFR pathway.

Figure 2, taken as a guide for intracellular molecules and processes participating in signal transduction, shows also small signalling molecules called small intracellular mediators, or second messengers (e.g. cyclic AMP and Ca^{2+}), that also perform important functions. Second messengers are generated in large numbers upon receptor activation and communicate the signal to other parts of

the cell mainly by rapid diffusion. Throughout this thesis we will not address this type of signal transduction mechanism. We will centre most of our investigations on aspects of intracellular protein cascades mediated by protein-protein (chapter 4) interactions or protein-gene interactions (section 2.3 and chapter 3).

The large intracellular signalling molecules represented in Figure 2, e.g. Ras and Raf, are intracellular signalling proteins. Many of these transmit the signal further into the cell by activating “the first” signalling protein of a cascade culminating in the nucleus or through other processes such as generation of small intracellular mediators (see below).

The Epidermal Growth Factor Receptor Pathway

The Epidermal Growth Factor Receptor (EGFR or ERBB1) signalling pathway is one of the fundamental pathways regulating growth, survival, proliferation, and differentiation in mammalian cells (57, 173) and has been a paradigmatic network for pathway biology. It is the best studied RTK pathway and in synergy with other members of the ERBB family plays a critical role in carcinogenesis (10, 56, 81). It is also involved in phenomena associated with cell plasticity, such as the epidermal to mesenchymal transition (EMT) (106, 177), crucial in embryonic development but also extremely important in late stage events in cancer such as metastasis, one of the hallmarks of cancer (115).

The family of ERBB proteins is made of 4 receptors, ERBB1-4 (HER1-4), and additionally 13 polypeptide extracellular ligands. All of these have a conserved EGF domain. An interesting characteristic of the ERBB family network lies in the fact that two of its members, ERBB2 and ERBB3, do not have autonomy in their signalling functionality. The first is deficient in its capacity to interact with a growth ligand and the second has defective kinase activity. This lack of autonomy doesn't weight on their ability to heterodimerize to form complexes with other ERBB family members that generate potent intracellular signals.

Due to its importance in cell behaviour the EGFR pathway has been the focus of several computational models. The first mechanistic description of the

temporal dynamics of the signalling responses was with parameters extracted from experiments in liver cells stimulated by EGF (151). These models have looked into a range of aspects of EGFR dependent signalling such as:

- transient versus sustained signal response of the GTPases Ras and Rap1 to a number of growth factors (38, 183, 226);
- the non-linearity in amplitude response of MAPK activation reaction due to variation in EGFR numbers (151, 230);
- autocrine positive feedback loops (235);
- crosstalk between MAPK and Akt pathways (119);
- synergistic EGFR signalling from the plasma membrane and membrane bound internal compartments (endosomes) (51, 220, 230).

1.2.1.1 Cascade of events in the Epidermal Growth Factor pathway

EGF binding, receptor dimerization and cross-activation

Binding of the EGF mitogen ligand contributes to EGFR receptor dimerization and autophosphorylation of intracellular domains (see Figure 2). Following this reaction, the phosphorylation of multiple receptor tyrosine residues enables the transmission of a biochemical signal to several cytoplasmic proteins (adaptors and enzymes) that are mobilized to the receptor. The cellular response to these stimuli is due to an elaborate biological circuitry. For any individual receptor pathway there isn't a unique single protein or gene that determines specificity in physiological responses. Nevertheless, one can identify key players in the receptor pathway with the intent of simplifying the modelling task. The major downstream targets of the EGFR signalling are the PI3K (phosphatidylinositol 3-kinase)^a and Raf kinases (via Grb2, SOS and Ras, Figure 2).

^a Phosphorylates phosphatidylinositol bisphosphate (PIP2) to generate PIP3(intensely charged lipid). PIP3 recruits proteins with PH domains like PKB (6)

Ras activation and downstream targets

Ras is recruited by several receptors. It belongs to the large Ras superfamily of monomeric GTPases^b and is fundamentally linked to other proteins in the cell. The Ras superfamily also contains two other families: the Rho (ras homolog gene) family which participates in signal transduction and cascading to the actin cytoskeleton, and the Rab family which regulates transport vesicles traffic. Ras crucially participates in transmitting signals from the cell surface to other parts of the cell. It is frequently necessary for receptor tyrosine kinase signal relay to the nucleus to induce, through differential gene expression, cell proliferation or differentiation.

Two major pathways lead to activation of Ras-GTP: Shc-dependent and Shc-independent (Figure 2). A number of activated RTKs do not exhibit the specific necessary phosphotyrosines for Grb-2 docking^c (6). These recruit another adaptor protein called Src homology and collagen domain protein (Shc). This particular protein has the capacity to equally bind to activated receptors and Grb-2, thus coupling RTKs to Sos by an alternative more indirect route (6). The complex RTK-Grb-2-Sos, or alternatively RTK-Shc-Grb-2-Sos, allows for nearby Ras molecules to be activated by Sos through an exchange of bound GDP for GTP. Very frequently the capacity of these receptors to initiate a cascade of events depends on relay chains of protein-protein interactions. The cytoplasmic tyrosine kinase Src, for example, also binds these receptors and phosphorylates other signalling proteins on tyrosines. Similar circuitry and proteins activating the GTPase Ras are thought to operate in all animals (6).

^b GTPases are enzymes that hydrolyze GTP (guanosine triphosphate). The name is usually restricted to the family of proteins that bind GTP which induces a new conformation capable of activating target proteins. In order to return to their original form they need to hydrolyze the bound GTP (6). The process of activation and deactivation of GTPases is performed with the help of guanine exchange nucleotide factor (GEF) and the GTPase-activating protein (GAP) (6).

^c The Grb-2 protein in mammalian cells binds through an SH2 domain to specific phosphotyrosines on activated RTKs and through SH3 domains to proline-rich motifs on a GEF called Son of sevenless (Sos) (6).

Mitogen Activated Protein Kinase (MAPK) network activation and end targets

In order to induce any cell phenotype such as proliferation or differentiation, the short-lived signalling events described above have to initiate a cascade of longer signals to reach the nucleus and change the pattern of gene expression. Activated Ras is fundamental in eliciting this conversion by starting a sequence of downstream serine/threonine phosphorylations. These are sustained for longer periods than tyrosine phosphorylations. A number of serine/threonine kinases are involved in these long phosphorylation cascades. Nevertheless, three types are central to the cascade: MAPKKK (or Raf), MAPKK (or MEK) and MAPK (6). In molecular biology the MAPK pathway is considered to be a paradigm for signal transduction. Multiple manifestations of the pathway (see Figure 5) have been found in all eukaryotic cells and have been studied in organisms ranging from yeast to humans. The system of three kinases is activated by sequential phosphorylation in response to a multitude of stimuli such as: cytokines, growth factors (e.g. EGF, see Figure 2), neurotransmitters, cellular, etc. The MAPK pathway utilizes a generic signalling design characteristic of biological transduction: a cycle made by a kinase that phosphorylates a target protein and a phosphatase responsible for dephosphorylation. The characteristic dynamics will be discussed in following sections.

An unusual feature of MAP-kinase cascade participating proteins such as Raf and Mek (Figure 2), is that for its activation to be complete and the protein to be able to relay signals to downstream components, it requires both threonine and tyrosine residues to be phosphorylated. These particular residues are separated in the protein structure by a single amino acid (6).

The MAPK pathway has been found to be constitutively activate by mutation in many human tumours (126, 221). Hyperactive mutant Ras has also seen has one of the forms for enhanced or aberrant signalling through the MAPK cascade (6).

Activation of the MAPK pathway by Ras ultimately (through the output kinase ERK which migrates to the nucleus) causes the phosphorylation of several cytoplasmic and nuclear targets such as transcription factors (c-Fos to

create AP-1 and ELK-1) that are involved in cell cycle progression (173). Eukaryotic cells contain possibly more than 12 different MAPKKs, seven MAPKKs and eight MAPKs (see also Figure 5 for links between receptors and end targets). These can be associated with four different MAPK modules that have evolved by gene duplication. Another important set of targets of EGF that are represented in Figure 2 are STAT-1 (latent gene regulatory protein, Signal Transducer and Activators of Transcription) and STAT-3. Activation of these transcription factors by the set of JAK kinases in response to EGF stimulation adds to the set of preconditions inducing cell proliferation (6).

Phosphatidylinositol 3-kinase network activation (PI3K)

The EGFR also targets the PI3K which is involved in activating pro-survival kinases such as PKB/Akt and pro-growth/survival kinases such as PKC. PKB/Akt is activated by PIP3 when recruited to the plasma membrane. It then returns to the cytoplasm and phosphorylates a variety of serine/threonine target proteins. One of these targets, known as BAD, is a protein that usually is adjuvant to programmed cell death (apoptosis). Phosphorylated BAD is inactive (it then is able to associate with 14-3-3).

Ca²⁺ level response

The EGFR also has an impact in cytosolic Ca²⁺ levels (see Figure 2) through the action of Phospholipase C- γ (PLC- γ , activated by tyrosine phosphorylation). This enzyme hydrolyses PIP2, generating IP3 (inositol triphosphate). Through this pathway, receptor tyrosine kinases are able to increase cytosolic Ca²⁺ levels (IP3 acts on the endoplasmic reticulum and the latter releases calcium) (6).

1.2.1.2 Timescales for intracellular network functioning

Signal transduction biochemical circuits transmit information on the state of the cell and its surroundings to other regulatory networks. They are essential in the complex cellular decision making that induces diverse cell phenotypes such

as cell division, apoptosis and differentiation(122). Gene regulatory networks (also known as transcription factor networks (7)) play a fundamental part in this phenotype selection and are influenced by signal transduction networks. Moreover, they exert feedback on the pathway. This could be crucial for signal spreading in the cell in normal cases or in the presence of deleterious mutations. On a more qualitative level, attempts to deal with the pathway signalling complexity like comprehensive signalling pathway diagrams (157, 158) have been decisive in clarifying molecules in the cell machinery to be modelled. Yet, cellular interactions are highly dynamical and non-linear and cell processes occur on different time-scales (Table 1). Therefore, a more resourceful approach is necessary to deal with intracellular network complexity involving dynamical behaviour. We will review and use modelling approaches, throughout this thesis, that respond accurately to these issues and take into account the wealth of time-scales in cellular processes.

Cellular signalling process (Human fibroblast)	Time
Kinase/phosphatase reactions	$\sim 10^{-3}$ sec
Protein conformational changes	$\sim 10^{-3}$ sec
Cell-scale protein diffusion (passive)	~ 100 sec
Cell-scale protein diffusion (active)	$< 10^0$ sec
Diffusion of small molecule across cell	~ 0.1 sec
Time to transcribe a gene	~ 30 min (including mRNA processing)
Time to translate a protein	~ 30 min (including mRNA nuclear export)
Typical mRNA lifetime	~ 10 min to over 10h
Cell generation time	20h non-dividing
Cell migration	$10^0 - 10^2$
Receptor internalization	10^2

Table 1 Order of magnitude for timescales in intracellular networks in human fibroblasts. Based on (7, 208).

1.2.2 Cycle and cascade motifs in signalling networks

After a signal is translated into the cell, a cascade of downstream events takes place as was described above for the representative EGFR pathway (see Figure 2). The elements of this cascade usually function as molecular switches: upon receiving a signal they suffer a transition, or switch, from an inactive to an active state, until another process, exerted by an additional protein or simply due to instability, switches them off. Examples of proteins functioning according to this paradigm are kinases and GTPases. Switching off proteins that have been activated is as important as the process of switching them on. The inactive protein level of a pathway has to be maintained in order for a message to be passed again. A common fundamental motif found in intracellular networks is made of proteins having interconvertible forms (see Figure 3). The change in protein state is performed by two opposing enzymes. For phosphoproteins the enzymes are the kinases that add a phosphate group and the phosphatases that remove it (Figure 3). The great majority of proteins underlying phosphorylation signalling cascades are serine/threonine kinases phosphorylating proteins on serine residues and less frequently on threonine residues. Others are tyrosine kinases, phosphorylating proteins on tyrosine residues. Genome sequencing has revealed that protein kinases are encoded by about 2% of the genome, and it is thought that around 518 distinct types of protein kinases operate in a typical mammalian cell (182, 216). Kinases are commonly mutated in cancers and constitute prominent therapeutic targets (28).

Another form of the active/inactive cycle motif is seen in activation of small G-proteins (e.g. Ras, Rho, Rab, Ran or Arf), which are also crucial for the continued transmission of the signal. In this case the guanine exchange nucleotide factor (GEF) and the GTPase-activating protein (GAP) execute the changes (Figure 3) (6). Take the case of the EGFR pathway (section 1.2.1, Figure 2); the GAP enzyme increases the rate of hydrolysis of bound GTP by Ras, which inactivates Ras (GTPase).

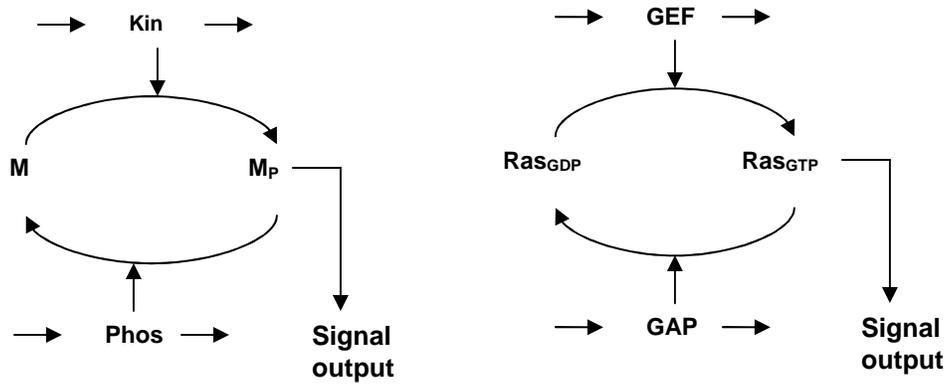


Figure 3 Typical motifs of intracellular signalling cascades. Left: one site phosphorylation cycle involving kinases and phosphatases. Right: Ras activation/inactivation involving GAP and GEF enzymes. Kin: kinase. Phos: phosphatase. M: generic protein. M_P: generic protein phosphorylated.

These types of cycles are the building blocks of all signalling networks studied until recently and seem to pervade intracellular network structure in both prokaryotes and eukaryotes (6). The cycles represented in Figure 3 are often interlinked forming layers of cycles known as cascades. A crucial aspect of these types of networks is that they typically establish cross-talk with other pathway cascades forming an intricate regulatory web. A key question being asked in this area is what is the purpose of all this complexity? Tools from electric circuit design have been useful in understanding, in the light of man-made system's concepts, the properties of these biological networks (255).

1.2.3 Modularity

Hartwell and coworkers (117) defended that the recognition of functional modules is fundamental for understanding biological complexity. Modules are defined as *"...discrete entities, whose relatively autonomous functions are separable (through spatial localization or chemical specificity) from those of other modules, and designed (or evolved) so that interconnecting modules allow higher level functions to be built."* (70). Highlighted examples that have been seen as conforming to this view in signalling are the MAPK cascades. Lauffenburger (167) further elaborated on the possibility to understand biology in

a “*hierarchical or nested manner*”, comparable to engineering design: components and their dynamical regimes are studied in isolation and incorporated into larger networks. Modular structures are thought to promote evolvability, by nesting specific cellular functions in distinguishable modules allowing robustness to core functions but variations in inter-module connections. To find these appropriate functional modules or understand why a subnetwork may be interpreted as being a module some approaches have been developed; some based on structural properties of the graph identifying a module as a tightly connected sub-graph or community, others at a more functional level defending that negative feedback loops may provide special characteristics to the elements of the network encompassed by it (70, 223). This facilitates intuitive understanding of network function and may inform on targeting strategies. The existing negative feedback between ERK-PP and the input to the MAPK cascade has been seen as a crucial point in isolating the cascade into a module (228). One conclusion taken from identifying the MAPK cascade (Raf/MEK/ERK) as a feedback amplifier module is that we should never target nodes inside the module because these perturbation will be overcome by the natural dynamical function and feedback of the network (228). Conclusions from a simulation study of the EGFR pathway (see section 4.2.2) will test this assumption. The usefulness of using these modules as the ultimate information for embedded network function is reduced when the intracellular networks to be modelled are extremely intricate and involve multiple feedback loops and cross-talk. Also, due to the proved existence of redundancy in intracellular networks (64), cellular systems biology may be harder to understand from a purely engineering modular or compositional approach. Intracellular networks have not only distributed robustness but also control (125, 175). In this thesis we will follow an approach which is closer to understanding intracellular network function not as composed of functional or structural modules but as a whole network information processing unit with extensive distributed control.

Below we focus on the paradigmatic MAPK pathways and the importance of cross-talk in cell fate decision.

1.2.4 Intracellular integration of signals and cross-talk between pathways: the mammalian mitogen activated pathway kinases (MAPK)

Signaling pathways have traditionally been interpreted as almost independent or isolated sets of reactions or modules, exchanging limited information. A number of studies have been devoted to understanding crosstalk within the theoretical framework of pathway specificity (20, 21, 159). Mechanisms such as cross mutual inhibition (20, 159, 185) (Figure 4) and kinetic insulation (21) have been thought as being crucial for avoidance of excessive signalling crosstalk.

On a more structural level another definition of crosstalk has been elaborated. It stands on the notion of cross-interactions relative frequency, i.e., *“number of cross-interactions found normalized to maximal number of possible cross-interactions”* (31, 32).

If one considers the wealth of signaling molecules and respective intricate web of interactions as “one big signalling system” (41), then pathway cross-communication is the standard and not the exception. As can be seen in Figure 5 cross-talk between MAPK pathways is quite frequent. This perspective shifts the research question of avoidance of crosstalk to understanding what its purpose is. In the context of therapeutic intervention the effort also shifts from how to develop drugs that circumvent the intricate crosstalk to using it to attain the desired outcome, e.g. cell phenotype changing from proliferation to apoptosis. A proposal based on a more functional analysis of signalling properties was reported in (25). Looking at the signalling system as a combinatorial decoder allows one to understand Input-output mapping.

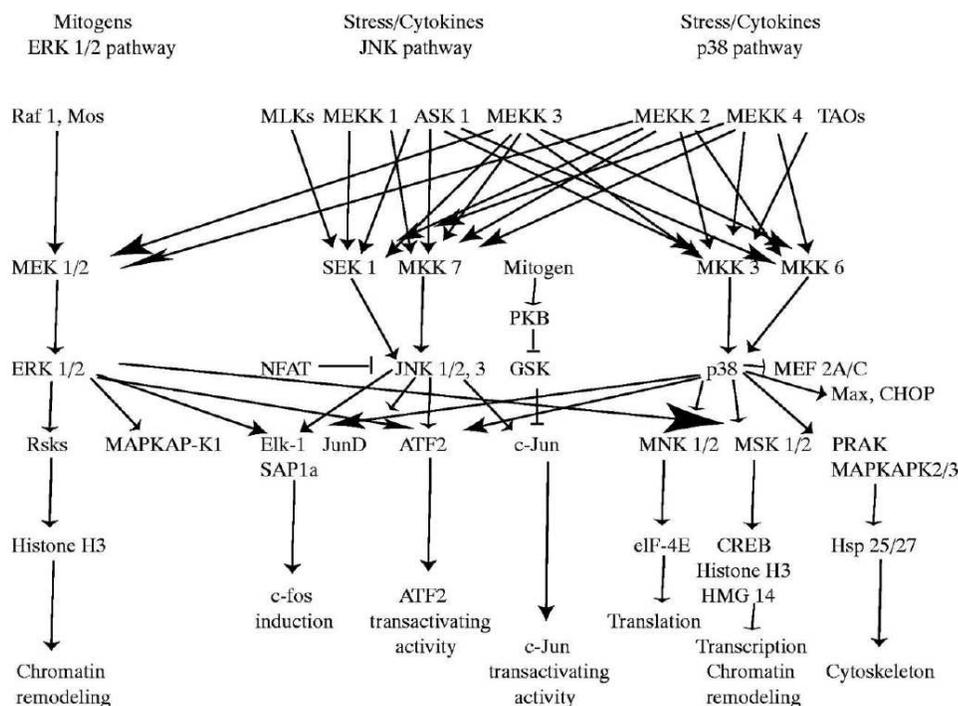


Figure 5 Cross-talk between mitogen activated protein kinase networks, transcription factors activated and the induction of early genes (IEG) in response to extracellular signals. For details about each of the protein see (59) and <http://www.ncbi.nlm.nih.gov/omim>.

The MAPK signalling pathways play a central role in gene expression regulation in eukaryotic cells. MAPKs activation of transcription factors are seen as one of the major mechanisms for inducing gene expression. Numerous mammalian transcription factors and their co-regulators have been identified as MAPK cascade targets (275). Four main groups of MAPKs have been reported in mammalian cells: p38, extracellular signal regulated kinase (ERK), extracellular regulated kinase-5 (ERK5, also known as BMK1) and c-Jun N terminal kinase (JNK). Each group contains a number of gene products and additional isoforms that are generated by alternative splicing, e.g. *Jnk* genes generate ten distinct isoforms (111). The structure of the mammalian MAPK pathways is highly conserved in yeast. It is therefore probable that the mechanisms governing this core cascade functioning is conserved across

eukaryotes (92). A representation of the cross-talk within the mammalian MAPK network is presented in Figure 5 along with the respective transcription factor targets and cellular responses. Recent studies have demonstrated that cross-talk is a fundamental aspect of cellular decision making (96) in situations where simultaneous external signals operate on the cell. The combinatorial and sequential complexity of the network has to elicit the right cellular response or phenotype, e.g. apoptosis, proliferation, differentiation.

In the following section we highlight the importance of signalling output node activation and phenotype commitment.

1.2.5 Transient vs. sustained activation and early gene induction: linking signals to transcriptional activity

The classical view of discrete linear networks of components connecting receptors to transcription factors has been substituted by an extremely intricate network where there's no clear view of specificity in signal-response events as was highlighted before. Specificity in cellular dynamics is performed both by spatial dynamics and temporal profiles of downstream signalling components. Take for instance the example of the MAPK cascade activation through the growth factor EGF or the nerve growth factor (NGF) (184, 195):

- EGF-induced transient MAPK activation stimulates proliferation
- NGF-induced sustained activation results in differentiation

Nevertheless, the control of the MAPK cascade is extremely complex. MAPK cascades are able to generate bistable dynamics and oscillations (26, 27, 149). These behaviours depend on subcellular localization and recruitment to scaffolds (116, 264). The activation profile has been the main indicator, in certain cell lines, of cell phenotype. Here we discuss the connection between activation profile of output signalling proteins and transcription factor activation. It will be crucial for

understanding how these signals regulate externally the transcriptional machinery and induce the correct phenotype.

MAPKs phosphorylation of target proteins, such as transcription factors, is secured through their docking sites, which can act as sensor units with the capacity to detect MAPK signal duration as well as strength (193-195). MAPK substrates contain binding sites for the MAPKs and Ser/Thr-Pro phosphoacceptor motifs. Several docking sites, crucial for efficient phosphorylation and enhancing specificity by selectively binding to MAPK subtypes, have been uncovered (193-195). Two that are particularly well characterized are the D-domains and DEF domains. The DEF domain is the docking site for ERK (266). It has been a common experimental fact that specific characteristics of the output signal of the MAPK cascade, ERK, correlates with mutually exclusive cell fate decisions. Take PC12 pheochromocytoma cells where ERK transient activation is associated with cell proliferation and sustained activation promotes cell differentiation (184). In other cell lines the profile of ERK activation has also been correlated with cell fate decision: sustained ERK activation in fibroblasts is a necessary condition for cell cycle re-entry and proliferation (184, 193, 274).

Let us take the example of the immediate early gene product c-Fos^d which expression is promoted by ERK activation. c-Fos is unstable if not phosphorylated at its C-terminus taking place. If ERK is activated transiently the c-Fos protein is not phosphorylated and thus is unstable and degraded. On the other hand if ERK is activated in a sustained fashion, c-Fos is phosphorylated (primed), the DEF domain becomes available to ERK and further phosphorylation leads to activation and initiation of a subsequent transcriptional program (194). Other DEF domain containing IEG products also function as ERK activity sensors, e.g. c-Myc (194, 195). DEF domain containing immediate early (IEG) products can also function as gatekeepers by only allowing DEF domain dependent phosphorylation when ERK activation profile attains a certain threshold. Furthermore, ERK may be retained in the nucleus in an active form

^d Part of the AP-1 transcription factor complexes.

and shielded from phosphatases by DEF domain binding (193). DEF domains also bind the protein p38 α in addition to ERK (91). Other ligands such as TNF may induce MAPK activity that also determines cell fate decision and may compete with the induction from other ligands such as EGF. JNK activation to TNF, a stress signal, determines distinct cell fates (91, 259).

We see from the above information that the time-dependent activation profile of the output nodes of signalling networks influences clearly the transcriptional programs induced.

In the following sections we review the most common mechanisms of initiation of transcription and integration of external signals into gene expression programs.

1.2.6 Control of gene expression

The process by which genetic sequences are translated into amino acid sequences in proteins or into nucleotide sequences of RNA is termed gene expression. It entails in eukaryotes the subsequent steps (6):

- Production of a primary transcript, known as pre-mRNA; this step is also known as transcription;
- Change of the initial pre-mRNA into a mature mRNA. This includes processing, splicing and subsequent transport or diffusion from inside to outside the nucleus;
- Synthesis of proteins into the ribosome.

The regulated transcription of genes requires the reorganization and modification of the chromatin which is affected by DNA methylation of cytidine residues. After this essential step, initiation of transcription involves selection of the target gene and formation of an intricate initiation complex. The main components of this complex are:

- The RNA polymerase;
- General and specific transcription factors;
- Specific cofactors that assist in the coordination of chromatin changes and RNA synthesis.

The class of proteins called transcription factors are produced in order to allow a signalling pathway to convey information with several genes. Via transcription factors a gene also communicates with others genes although in a slower time scale (see Table 1). Transcription factors are fully capable of forming complexes and are susceptible to degradation, all of which have a fundamental role in the regulation of gene activity.

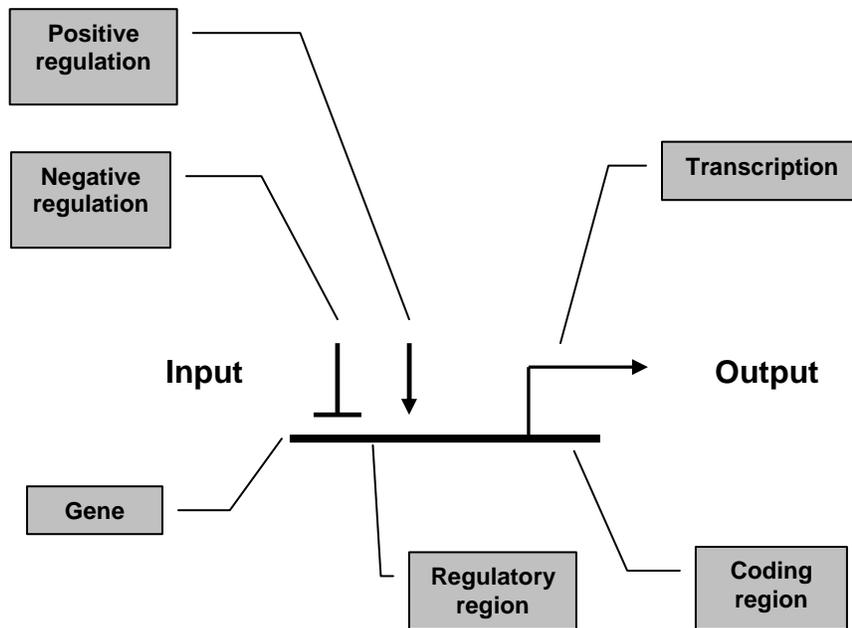


Figure 6 Hardware gate abstraction of the cis-regulatory region of a gene.

A simplified depiction of the cis-regulatory region of a gene as a hardware gate may be seen in Figure 6. It is a stretch of DNA with two regions which may not be continuous:

- a regulatory region which serves as input; this contains essential protein binding sites for transcription factors;
- A coding region which serves as the output; this region may code for one or several proteins.

Although one may model this system using a gate abstraction there are cases where this may be too coarse. Nevertheless, this modelling approach is very common (33, 34). The coding or output region represented in Figure 6 follows

roughly the following genetic code: triplets of nucleotides (codons) are associated with one of 20 amino acids, with the addition of start and stop triplets.

The input region obeys principles that are much more complex. The subtleties corresponding to this region are still poorly understood. The process of transcription involves transcription factors (with specific 3D shapes) binding, with variable strength, to nucleotide sequences in the input region. We have, therefore, an “analog” character inherent to this process. Various shapes recognize to a different degree another type of information (digital) in the form of a string. Again, this string is an abstraction from a 3D chemical entity, also analog. Nonetheless, it is a custom modelling practice to use the abstract hardware gate approach and simply measure the effectiveness of gene to gene connections by tracking gene product concentrations. The modeller has to understand, in any case, the limitations of such an abstraction.

The external inputs affecting a gene or gene regulatory network, i.e. a system where several transcription factors regulate each other's expression, are signals that carry information from the environment and inside the cell. An example of a cascade of signals was already described before (see Figure 2). Each signal is a small molecule or molecular partner that affects activity of one of the target transcription factors (Figure 7). For a view of the main transcription factor targets of the mammalian MAPK cascade module see Figure 5.

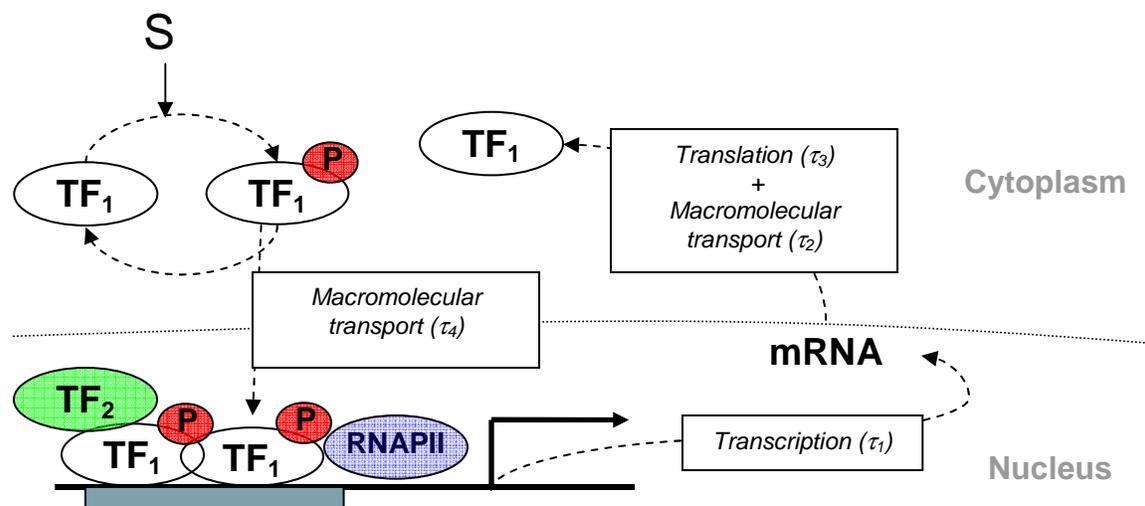


Figure 7 Generic representation of the process of activation (phosphorylation) of a transcription factor, cooperation between activated TF monomers and a co-activator TF to induce transcription and transport of both mRNAs and proteins. TF_1 -transcription factor. TF_2 - RNA polymerase II necessary for initiation of transcription. Each τ_i represents delays associated with transcription, translation or molecular transport. See also Table 1 for approximate values of the characteristic times each of these reactions take.

The signal can also be as simple as a sugar molecule that enters the cell and directly binds the transcription factor. The signal usually initiates a physical transformation (often via phosphorylation, for a review see reference (39)) in the shape of the transcription factor protein, causing it to assume an active molecular state (6). Activated transcription factors subsequently bind to DNA sequences, the responsive elements, and control the transcription of particular nearby genes. The initiation of transcription may be dependent on the interaction of several transcription factors and cofactors, together with the RNA polymerase II complex (Figure 7), leading to a transcription initiation process with complicated combinatorial logic (33, 34, 40). Transcription factors frequently interact with DNA in homodimer form or as heterodimers. The potential complexity of genetic regulation is further advanced by the fact that certain transcription factors are able to activate their own transcription (e.g. Jun) (248). As will be seen in further chapters this is a fundamental characteristic for simultaneous existence of multiple stable states in genetic regulatory circuits (see section 8.6). In other

cases, transcription factors repress their own transcription. This is performed either directly, e.g. Fos, or by inducing repressors such as the inducible Ca^{2+} /cAMP-responsive early repressor (ICER) protein. The latter is increasingly transcribed if phosphorylated dimers of CREB (cAMP response element binding) bind to a nearby Ca^{2+} /cAMP-responsive element (CRE). ICER suppresses its own transcription.

The required presence of several components together with the activation profile of each of them performed by an upstream signalling protein may lead to changes in the overall network. Transcription factors are embedded in transcription factor or gene regulatory networks responsible for the transcriptional program ending in the appropriate cell fate (7).

Transcription networks show strong separation of timescales as was previously seen in Table 1: the input signals usually change the transcription factor activities on a sub-second timescale as mentioned earlier. Binding of the active transcription factor to its DNA sites often reaches equilibrium in seconds. Transcription and translation of the target genes takes minutes and the accumulation of the protein product minutes to hours (see Table 1). This process plus transport of mRNAs and proteins in the cytoplasm is usually modelled in the literature through delays (236, 237) (see also Figure 7).

Signal-transduction networks made of interacting proteins typically operate much faster than transcription networks (Table 1). Thus, they can be considered, to a degree, to be approximately at steady state on the slow timescales of transcription networks. This is the customary approach to modelling intracellular network dynamics. Nevertheless, composite motifs (including fast and slow time scales) may be of crucial importance in cell functioning (7). Also, the traditional division between slow and fast time scales for simulating and solving hard numerical problems may not be appropriate if we take into account variations in the parameters, which may give rise to bifurcations in the dynamical system (85-87, 87). This completely changes the fast time scale network behaviour and perturbs differently the slow network dynamics, possibly inducing different transcriptional programs.

We have reviewed so far the main aspects of signal transduction (sections 1.2.1 and 1.2.4) and the mechanisms by which the transduced signals affect transcription factor activity (sections 1.2.5 and 1.2.6). The molecular aspects of the pathway described before (see Figure 2) will be fundamental to follow the perturbation studies performed in section 4.

The following picture (see Figure 8) represents the signalling apparatus as an analog combinatorial decoder (25), where a combination of inputs (S_i) with different profiles (impulse like, sustained, etc) are processed by the signalling block involving several pathway modules with crosstalk. This is actually in line with data collected from cross-talk studies in signalling networks (181) (see also Figure 5). Therefore, the working hypothesis of signalling systems functioning as a whole information processing network is indeed a viable one (122). Downstream of the signalling module different activation profiles, e.g. transient/sustained, of the elements (TF_i) operate on an intricate gene regulatory network to induce the correct transcription pattern and consequently cell phenotype.

We believe that this whole-network approach, and not a collection of independent pathways, is the most suitable to clarify how cell fate decision arises in intracellular networks. Moreover, current efforts in pathway biology have focused on signalling module response studies (96), with cross-talk being a fundamental aspect. The question lies with understanding how do the combinations of inputs to a transcription network, in the case of Figure 8 TF_i -PP activation profiles, select appropriate gene expression patterns in a concerted fashion. Below we will extend this rationale to multitargeting therapy design. This class of therapies operate on signalling pathways that affect transcription network activity. Hence, selecting targets is in effect optimizing the combination of TF_i -PP activation profiles inducing the correct cell response. The concept of cell phenotype being associated with a gene expression pattern will be reviewed in section 1.5 and linked with the whole-network approach to cellular dynamics defended in this thesis. The effect of different TF_i -PP signal shapes in cell

phenotype or gene expression pattern selection will be explored at length in section 2.3 and section 3.

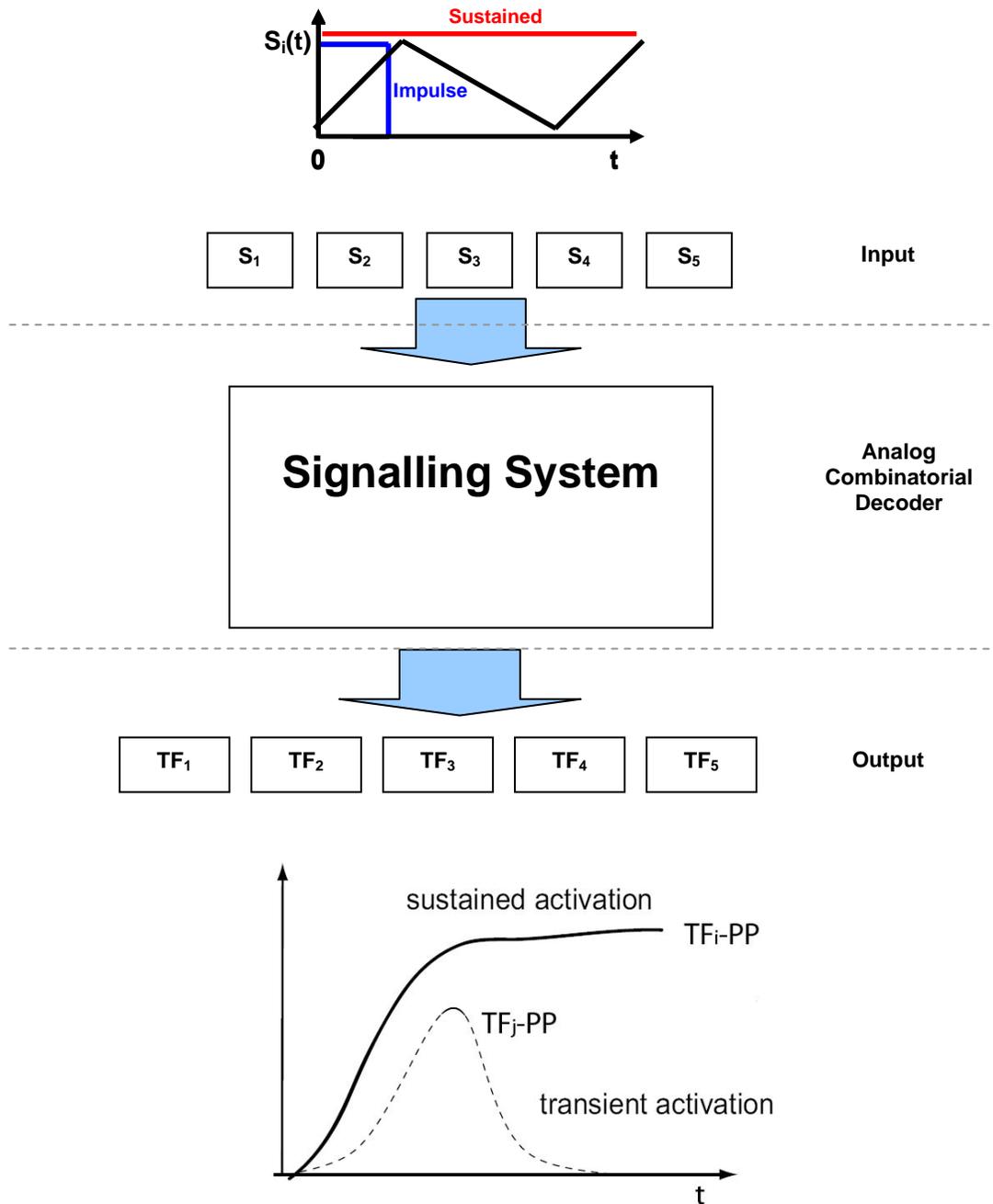


Figure 8 The signalling system as a combinatorial decoder. Upper line: generic signal shapes normally used in experimental perturbation studies (140). Bottom line: generic shapes for activated (doubly phosphorylated) forms of transcriptions factors. The double phosphorylation represents in a simplistic form the mechanisms described in section 1.2.5. S_i - external signals. TF_i -transcription factors activated in response to external signals. Inspired in (25).

1.3 Multitargeting therapies: network connectivity can be used to develop possible strategies.

Targeted therapy is an expression associated with a new generation of drugs, e.g. Tyrosine Kinase Inhibitors (TKIs) (280), designed to disrupt or modulate specific molecular targets which are believed to have a fundamental role, when in a mutated form, in tumour growth and progression. This new approach is in contrast with a more empirical one practiced in the last few decades, cytotoxic chemotherapeutics, which is not based on thorough molecular characterization of protein targets. At the moment, clinical trials are in progress for a range of drugs targeting signal transduction components. Among these are second-messenger generators, receptor tyrosine kinases but also kinases involved in cell-cycle regulation (204, 229). Several of these drugs will be discussed further ahead (section 1.3.3). One of the main advantages over conventional chemotherapeutic agents that has been reported in the literature is their capacity to obstruct specific deregulated pathways with diminished effect on cell functioning (219). In section 4 we will, on the other hand, prove that the application of a number of these drugs may not be as efficient as required, due to non-specificity in binding (79).

Drugs devised against individual molecular targets embedded in a convoluted protein-protein-gene interaction network with redundant and multifunctional cellular mechanisms (43, 44, 124) cannot usually fight against multigenic diseases, e.g. cancer, or diseases disrupting multiple tissues or cell types, e.g. diabetes or immunoinflammatory diseases (see graph representing the number of perturbations necessary for relevant reduction in output response in Figure 13). Although the single target molecule has led to growth in the industry, successful development of first-in-class drugs is extremely difficult (224). A number of novel therapies have been developed that correct some of the short comings of single target therapies by recurring to a network control approach (82, 147) in signalling, transcription or metabolic systems (192, 203). This aims to identify possible targets for the application of multiple drugs either simultaneous or in a timed manner. The systematic search of combinations of drugs in vitro can avoid the traps of intracellular evolved systems. Yet, it demands large-scale searches in an

enormous space of candidate target combinations using cell-based experiments (282). Approaches to multitargeting based on stochastic search algorithms are currently being studied that use both topology information and pathway details (46, 273).

Targeting multiple nodes simultaneously on a dynamical model of the signalling cascade of the Epidermal Growth Factor Receptor has been performed by Araujo and coworkers with positive conclusions (14). It was observed that attenuation of the output signal is enhanced with multiple targets which reinforces the idea that understanding of the dynamics of pathway functioning and the inherent distributed control is crucial for therapeutics. One interesting and solid concept that has been devised and brought from systems engineering and that could widely help in efficient targeting is robustness (153-155). Biological systems show fail-safe capacity through redundancy and diversity which allows continued operability. Pathway cross-talk (section 1.2.4) is also a mechanism for robust behaviour but may diminish drug efficacy.

The inefficiency of certain drugs relates sometimes to the fact that the inherent robustness of the patient's systems or of pathogens balances any desired changes caused by drugs. Yet, systems may exhibit a trade-off between vulnerability and robustness. The extreme vulnerability verified in some biological systems when faced with perturbations (e.g. mutations on the p53 protein crucial for induction of apoptosis (221)) for which have not been optimized illustrates their robust yet fragile nature, an idea proposed under the concept of Highly Optimized Tolerance (HOT) (47, 49). Side effects provoked by drugs are probably an interference with certain features that stand out as a fragility of the system. Kitano and coworkers have long defended that "*understanding the relationships between robustness, disease, drug efficacy and side effects is the first step towards the design of drugs that can target robust systems to achieve the desired therapeutic goals*" (155), with which we agree. Biological networks are expected to demonstrate robustness in response to node removal, "fail-off" failure, but vulnerability to sustained malfunction of nodes, "fail-on" failure. The most disruptive impact on biological systems comes from invading agents,

overexpression and amplification of genes and uncontrolled activation of their regulatory loops, such as the case of Ras (see section 1.2.1.1), so common in cancers.

Cancer can be compared to a “*parasitic disease*” (155). Under this perspective tumour cells establish a symbiotic relationship with the host by controlling its robustness mechanisms. In this way they protect themselves and divide uncontrollably. Also, they are able to establish robust mechanisms for themselves. Regarding therapeutic interventions in tumours, their robust behaviour is manifested through multiple drug resistance, micro-environment remodelling, intra-tumoural genetic heterogeneity (155).

Several questions arise in the face of the astonishing biological complexity of intracellular circuits. How should drugs be designed to incorporate the concepts relating to robustness, both structural and dynamical, and overcome the problems observed in their application? Regarding intracellular networks where should drug molecules strike? When should they strike and in which dosage?

The complex machinery of a cell may be characterized as a network where interactions represented by links determine in a very coarse grained fashion possible functions (section 1.3.1). The availability of information on networks at genome scale, ranging from metabolic reactions to gene regulatory and protein interaction networks, has created an interest in these large scale networks’ topological properties or wiring (4, 5, 36, 136, 137, 187, 233).

Below we review main aspect of network representation that will clarify the mathematical modelling throughout the thesis and the determination of adequate multitargeting strategies in general (see also section 4.4).

1.3.1 Directed and undirected graphs as a tool for qualitative intracellular network description.

One of the most straightforward ways to create a model of an intracellular network is to view it as a graph, directed or undirected (Figure 11). A graph can be defined as a tuple $\langle V, E \rangle$. Here V stands for the set of vertices and E the set of edges. A directed edge is a tuple $\langle i, j \rangle$ of vertices, with i denoting the head and

j standing for the tail of the chosen edge. In this description, each vertex corresponds to a gene or protein and each edge to an interaction. In an undirected graph the representation of the nature of the interaction between two vertices usually is not directional. Physical interactions involving protein-protein interactions without explicit activation or inhibition can be represented by undirected graphs.

Another representation known as a hypergraph can be used to represent for instance cooperation of two proteins in the regulation of the expression of a gene. The formal expression of a hypergraph is now $\langle i, J, S \rangle$, where J constitutes the list of genes regulating i and S the respective list of signs showing the nature of the regulatory influence (Figure 9).

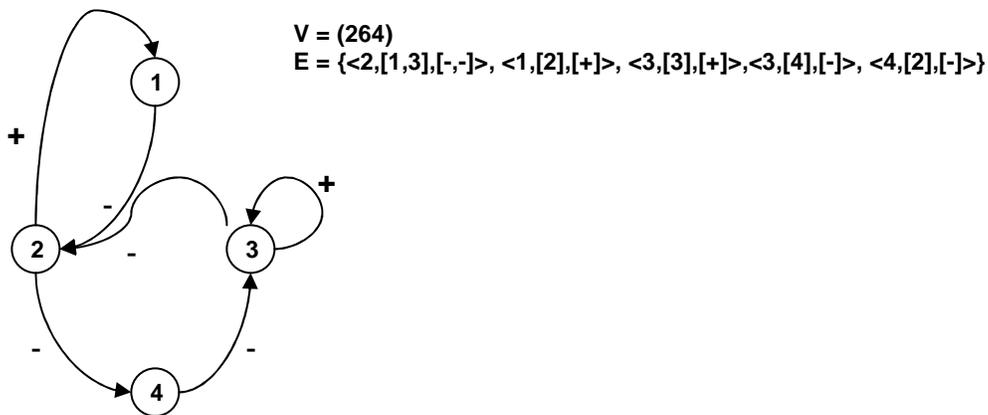


Figure 9 Representation under the graph formalism of a gene network or a network involving both genes and signalling proteins.

Current databases, such as KEGG, that provide information about regulatory interactions can be viewed as annotated graph representations.

Resorting to graphs a number of biologically relevant predictions can be made. Searching for paths between two genes may be a way of finding missing regulatory interactions in the database or providing evidence about redundancy in the network. Furthermore, the existence of feedback cycles may inform on the possibility of homeostasis or multistationarity (253). Network connectivity translated into quantities such as the average and the distribution of the number

of regulatory proteins per gene shows us in reasonably simple terms the complexity inherent to a specific network.

Furthermore, the existence of modules of regulatory activity may be identified through tightly connected subgraphs. These may be associated with isolated functional activity and further elucidate the level of complexity of an intracellular network. Nevertheless, although biology is largely accepted as being modular there's no unique formal method for devising modules on intricate networks (117, 167, 223, 227, 231, 271).

Of particular interest is also the recent study of network properties such as “network motifs” (187). These are locally defined structures or directed subgraphs which stand out as patterns of interaction between a reduced number of nodes. In bacterial transcription factor networks typical motifs such as the feed-forward loop (233) (Figure 10 A) are significantly over represented. For each directed edge a sign can be associated ($s=\pm 1$), separating the FFL motif into two classes

according to the number of negative edges: incoherent (odd number) and coherent (even number) (7). In signalling networks one may also represent interactions through directed graphs. Mostly, these interactions are associated with direct activation, e.g. phosphorylation. Nevertheless, signalling networks may also involve events associated with protein complex formation without direct activation, e.g. binding of kinase to scaffold proteins or adaptor proteins binding to a tyrosine receptor. Common motifs in signalling networks are the BIFAN

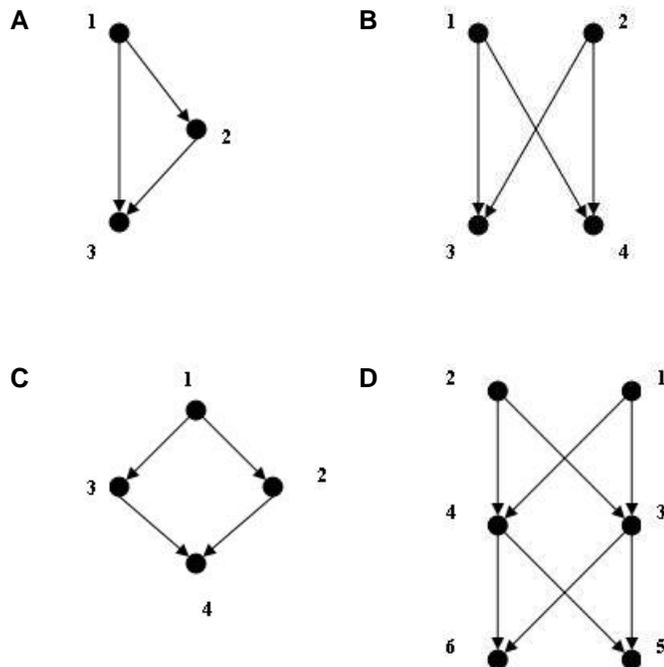


Figure 10 Common structural motifs found in intracellular network. A) Feed-forward loop motif. B) Motif known as bifan. C) Diamond motif. D) Multilayer perceptron motif. See also (7).

(Figure 10 B) and the Diamond (Figure 10 C) with possible and also found generalizations such as the multilayer perceptron (Figure 10 D) (7) known in the area of artificial neural networks for its pattern classification capabilities (35). The strong presence of the multilayer perceptron motif in signalling networks reinforces our interpretation of the function of this particular module as an analog combinatorial decoder (see Figure 8).

The use of graphs as a modelling technique also provides us with a way of comparing regulatory or signalling networks from different species and understand which parts have been conserved during the evolutionary process (65). The graph models may be constructed by extensively consulting the information stored in databases or by reverse engineering approaches to gene expression data. Several clustering algorithms have been devised that group genes having comparable temporal expression patterns, under the rationale that they regulate each other or are coregulated by another (8, 123).

This methodological approach is interesting for obtaining structural aspects of the system under study in this project. Yet, this does not restrict completely the possible dynamical behaviours. One can only speculate about the outcome of the interactions. Parameters regulating interaction activities are crucial at this point (255). If certain parameters are not available, abstractions based on Boolean logic or continuous extension to Boolean logic may be resourceful (to be discussed ahead).

1.3.2 Global topology of networks, robustness and targeting strategies

Systematic network topology information has been made accessible for metabolic networks in various microorganisms as well as for genome-wide protein–protein interaction networks such as the yeast *Saccharomyces cerevisiae* (Figure 11). Network data is also available for *Escherichia coli*. In

several existing databases, e.g., KEGG^e, pSTIING^f, Reactome, all of this information is collected, curated and available for the research community.

Interpretation of genome-wide complex networks has been based mostly on topological information. Highly connected hubs in these networks have been proposed to be more likely essential genes. Power-laws (Equation 1, probability of finding a node with d links), which coexist with hubs, have been shown to generically lead to networks more tolerant to random failure of individual nodes. Yet, they show vulnerability to specific targeted attacks on the hubs (5).

Equation 1

$$P(d) \propto d^{-\gamma}$$

This topology is usually known as scale-free due to the invariance of $P(d)$ under a rescaling of d (Equation 1). Scale-free topologies have been found to lie in the interval $2 < \gamma < 3$ (4).

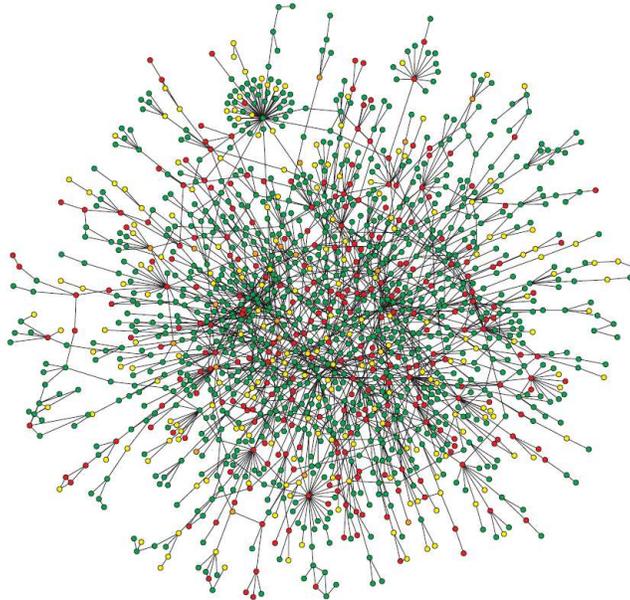


Figure 11 Yeast (*Saccharomyces cerevisiae*) protein interaction network. Only the largest cluster containing approximately 78% of all proteins is shown. Nodes are colour coded according to the phenotypic effect arising from removing the respective protein: red reflects lethality, green stands for non-lethality, orange for slow growth, and yellow for unknown effect. Taken from (136).

^e KEGG- Kyoto Encyclopedia of Genes and Genomes.

^f pSTIING-Protein, Signalling, Transcriptional Interactions & Inflammation Networks Gateway,

In cancer, intracellular network proteins such as Ras (involved in cell proliferation) or p53 (involved in apoptosis) are often mutated (221). These constitute important hubs for cell signalling. They have the highest degree (number of links). Attacking mutated Ras or proteins upstream that activate it (e.g. EGFR) has been a successful strategy in cancer therapies (114). Globally, another measure that can be helpful in understanding cross-talk between pathways is betweenness (36). This is used as a standard measure quantifying node centrality in intracellular networks and can be defined as the number of geodesics going through a particular node of interest. Targeting a node with high betweenness may be seen as targeting bridges or links between pathways (Figure 12). Another measure commonly used is edge betweenness which is defined according to (36) as the “number of shortest paths between pairs of nodes that run through that specific edge”. Targeting an edge following this measure could also be viewed as targeting a link with crucial importance to pathway cross-talk.

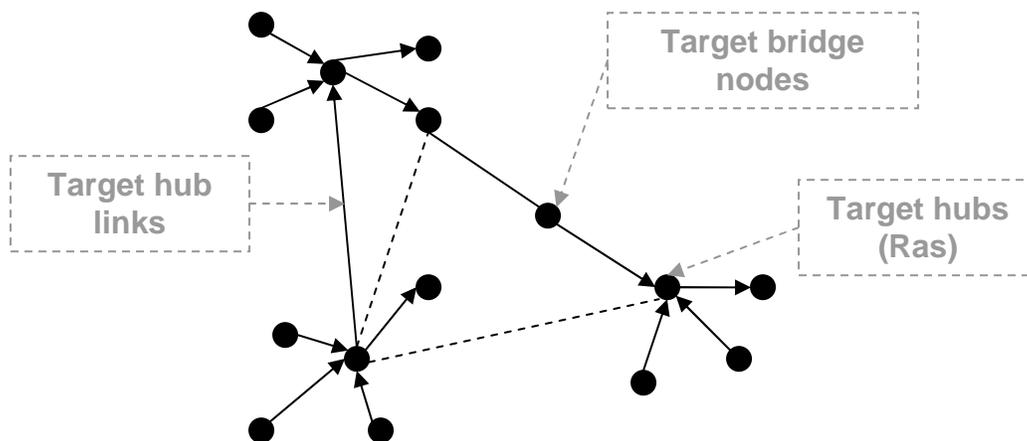


Figure 12 Graphical depiction of possible strategies for targeting networks. Black dots represent proteins. Links between proteins represent in a simplistic fashion complex reactions.

It has been verified that although power-law networks constructed through random node attachment algorithms show robustness to randomly selected

nodes (5), the network becomes progressively more susceptible to disruption as the number of nodes removed or weakly modulated increases (3, 62). This has motivated therapeutic approaches looking for synergistic effects of drug combinations with reduced toxicity (169-172, 282). These multitarget combinatorial approaches explore cross-talk and redundancy in intracellular networks (see Figure 13).

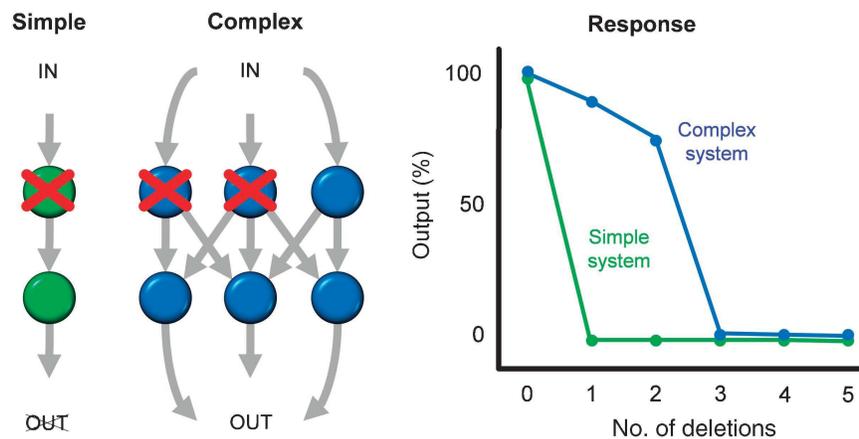


Figure 13 Simplified representation of how complex networks have a high-threshold to higher-order perturbations. From simple linear pathways to interlocked pathways the response to deletions changes. Complex networks are only sufficiently disrupted reaching what is known as ‘combination order of fragility’ (COF) by a higher number of deletions (169). This fact justifies the use of multitargeting therapies in intracellular network phenotype selection.

As can be seen in Figure 13 the representation of the complex systems of interlocking pathways is a multilayer perceptron, a generalization of the BIFAN and Diamond structures (Figure 10). This multiple-in-multiple-out structure, commonly referred to as signalling bow-tie (63), generates a classification of the combination of inputs based on their level (dosage) and timing (122), recurring to the multiple interlinked chemical processes, transforming these into a combination of activation concentration profiles of the output nodes crucial for cell fate decision (see section 1.2.5 and Figure 8). Regarding targeting strategies, the question arises of how to modulate the biochemical processes as to induce the correct combination of concentration profiles of output nodes (Figure 8) and how to interpret this combination (see section 1.5). Also, recalling the results

presented above in Figure 13 and the fact that scale-free networks are more susceptible to weakly modulated nodes (3, 62), we ask if lower affinity molecules with broader specificity or therapies relying on administration of multiple drugs can be more efficient than high-affinity, highly specific drugs (see also section 1.3.3).

1.3.3 Multiple-targeting or multiple drugs with individual single targets?

Clinical experience has shown that drug combinations can be more efficient than single agents (282). In many cancer treatments combinations have become the main strategic approach (66). The development of combinatorial targeting strategies may take two broad avenues: combination of multiple agents with high specificity or development of an agent with multiple targets (66). The latter might address concerns related to efficiency of drugs with a limited spectrum of action. Development of a single agent with multiple targets might overcome molecular heterogeneity and have higher probability of succeeding if tumour markers are not systematically collected or established. Nevertheless, the mechanisms of its action are difficult to understand and the development of a drug with optimal potency against several targets is a difficult endeavour. Another problematic aspect stems from the additional targets (non-specific interactions) which may or may not be relevant in a particular tumour molecular profile, thus increasing the probability for toxicity. For example, non-specific kinase inhibitors have been identified through extensive screening for activity against particular kinase(s) of interest by Fabian and coworkers (79) (see also section 4.2)⁹. Development of kinase inhibitors with a specific kinase-inhibitory profile showing optimal potency for each of the chosen multiple targets is still not possible and non-intended off-targeting may erase or mask the intended drug effect.

By contrast, the strategy resorting to combinations of more specific targeted agents might be more appropriate for personalized regimes with the molecular

⁹ The most recent study of non-specificity of small molecule kinase inhibitors was performed by Karaman and coworkers (144)

profile of the patient's tumour analysed systematically. The toxicity of this type of combination may also be more circumspect given that off-target effects are reduced (66). We have, therefore, a possibly conflicting set of approaches. In chapter 4, we explore the effects of non-specific interactions of small molecule kinase inhibitors, which, in effect, follow the approach of development of an agent with multiple less specific targets. We will not compare, nevertheless, multiple highly specific drug targeting and low affinity multitargeting. Either way, a multitargeting approach has been both theoretically (3, 62) and experimentally (169) justified even if in certain scenarios non-specificity may increase toxicity.

Although network structural properties have been seen as a precious resource for development of therapies, biological systems are highly dependent on time-dependent processes (see for example the discussion of section 1.2.5). In order to represent in a more faithful way the inherent complexity of biosystems, we need to model them by applying tools that allow representation of dynamic processes. Below we highlight several modelling approaches that pervade systems biology, and that will be fundamental in answering specific dynamical aspects of cell fate decision throughout this thesis.

1.4 Dynamical systems modelling approaches in intracellular network biology

Biological systems are complex systems where a large number of heterogeneous entities interact and give rise to system-level behaviour and processes. They commonly possess certain characteristics that make them difficult to study and predict such as: hierarchy, heterogeneity, polymorphism, context dependency, evolution, reprogrammability, emergence and non-linearity. Computational approaches to modelling such overwhelming complexity have been brought from computer science, physics, mathematics and engineering, creating a new perspective on how biology should be addressed. This new interdisciplinary endeavour is known in the literature as Systems Biology and its goal *“is to analyze the behaviour and interrelationships of entire functional*

biological systems' (152). Extremely powerful technologies for making comprehensive measurements on DNA sequence, gene expression profiles, protein-protein interactions, etc, give the necessary information on biological systems for the understanding at the systems level.

Computational techniques have contributed immensely to understanding structure and dynamics of biological systems. Two main avenues have been explored: knowledge discovery or data mining, and simulation based analysis. Data mining aims at extracting hidden patterns from experimental data such as gene expression profiles. One example is the inference of Boolean networks from gene expression data resorting to computational learning theory and non-linear signal processing methods (234). The simulation based analysis completes the in-vivo or in-vitro experiments by hypothesis testing and validation (152). Database processing, modelling, simulation, and analysis are all current major efforts of the scientific community. Tools and concepts brought from control theory like feedback regulation, or from concurrent systems analysis like process algebras, or even from statistical physics for the analysis of large ensembles of entities, have provided new avenues for describing and understanding systems at the cellular level and also at the tissue level.

Each model is a representation of a system, an abstraction which involves simplification and aggregation of details. What do certain approaches offer in modelling biological network systems when tested against others?

Our project will mostly use the ODE formalism, with extensive bifurcation analysis (section 8.6), and in fundamental sections for attractor selection and cell fate decision will resort to SDE theory (section 2.3). Noise will be central in understanding mechanisms behind cell-fate decision. We will also describe a modelling methodology in a final section, where, by using qualitative data collected from the literature and putting this in a Boolean formalism, we transform a rule based system into an ODE model (section 4.3.2.1). The following sections highlight main aspects of modelling approaches in systems biology and justify the use of the modelling techniques in sections 2, 3 and 4.

1.4.1 Continuous systems models

One of most important goals of computational cell biology is to achieve mechanistic understanding of cell behaviour. This means we aim to elucidate what determines each physiological state and what forces or influences transitions between them. The profusion of techniques in molecular biology has enabled the characterization of physiological states in terms of molecular concentrations. Therefore, the dynamical behaviour of the cell can to a certain extent be translated in terms of evolution of concentrations.

The continuous branch of modelling and simulation is held together by the differential equation paradigm for model representation and numerical integration (an example is the simple Heun method which will be used in numerical integration throughout the project, see section 8.4). Qualitative dynamical properties of the system can be studied by various types of ordinary differential equations (ODE) and many others such as delay differential equations and partial differential equations. In our study we consider time t to be the only independent variable. The ODE formalism seeks to model the concentration of RNAs, proteins, and other molecules by variables depending on time. Regulatory interactions are expressed in functional and differential relations involving the system's variables of interest. They take the following form (Equation 2):

Equation 2
$$\frac{dX_i(t)}{dt} = f_i(\mathbf{X}(t)), X_i(t_0) = X_{0i} \quad 1 \leq i \leq n.$$

The previous set of equations is a mapping from the state space of \mathbf{X} (the initial state \mathbf{X}_0) to the function space of \mathbf{X} :

Equation 3
$$F : X \rightarrow S_{X,t>t_0}$$

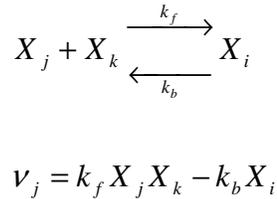
Each function f_i expresses usually a non-linear relation between quantities in the system. In most models of biochemical reactions the right hand side of Equation 2 takes the following form (see Equation 4).

Equation 4

$$f_i = \sum_{j=1}^r n_{ij} v_j,$$

where $i=1,\dots,n$, and n equals the number of biochemical species with concentration X_i . Additionally, r is the number of reactions with rate v_j and the quantities n_{ij} stand for stoichiometric coefficients. According to the information available for the system's kinetics and the detail desired, the reaction rates included in the model may follow sophisticated kinetic laws (60, 120). Usually, mass action kinetics is used (see Equation 5).

Equation 5



The parameters k_f and k_b are rate constants. Under specific approximations (60) the dynamics may be represented by Michaelis–Menten kinetics ($m=1$) that are a specific case of Hill-type dynamics:

Equation 6

$$v_j = V_{\max} H^+(X_j, K_j, m)$$

$$H^+(X_j, K_j, m) = \frac{X_j^m}{X_j^m + K_j^m}$$

K_j stands for the threshold and m for the steepness of the function. V_{\max} is the maximal rate. The higher the steepness the closer the Hill function gets to a step function ($m=\infty$) (Equation 7, Figure 14, observe transition from blue to red and green line).

Equation 7

$$s^+(X_j, K_{ij}) = \begin{cases} 1, & X_j > K_{ij} \\ 0, & X_j < K_{ij} \end{cases} \text{ and } s^-(X_j, K_{ij}) = 1 - s^+(X_j, K_{ij})$$

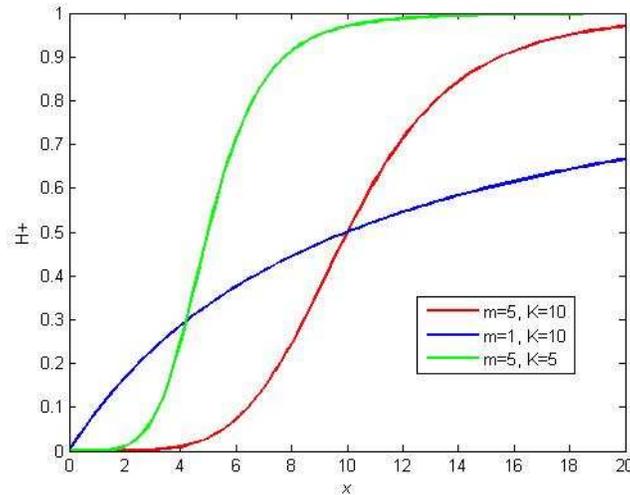


Figure 14 Hill function profile for several hill coefficients and Michaelis-Menten constants. See Equation 6.

Continuous models are widely spread in the systems biology community and are seen as a good reflection what is measured in cell biology samples over time (60, 120): concentrations of the DNA, enzymes, or metabolites. One can also use discrete time delays in this type of formulation to express for instance lagging processes arising from transcription, translation, and diffusion to the place where a protein is active (Equation 8, Figure 7). This additional feature is more faithful to the real biological system and hence an upgrade of the ODE formalism (148) to a DDE formalism.

Equation 8

$$\frac{dX_i}{dt} = f_i(X_1(t - \tau_{i1}), \dots, X_n(t - \tau_{in})),$$
$$X_j[t_0 - \tau_{ij}] = g_{0j} : [t_0 - \tau_{ij}, t_0] \rightarrow X_i, 1 \leq i \leq n.$$

The set of equations represented by Equation 8 is a mapping from the initial function space to the function space of X_i for a duration equivalent to the respective time delay (in Equation 9 the time delay is assumed to be the maximum of the time delays involved in Equation 8):

Equation 9

$$F : S_{X, t_0 - \tau < t < t_0} \rightarrow S_{X, t > t_0}$$

The increase in the variety of the initial conditions generally also results in a higher number of solutions too. Oscillations have been reported in signalling-transcriptional systems modelled through explicit time delays (188). Another way to use time delays with this type of differential description and particularly in the case of distributed time delays is incorporating integrals (218). Nevertheless, the experimental capacity available may not be enough to determine, to a satisfactory degree of accuracy, the kinetic parameters involved in our theoretical description. In these particular situations several other approaches may be justified where time is not represented in a continuous fashion but the system is stepped according to its inherent rules. Time stepped methods such as Petri Nets brought from concurrent systems theory have been applied to qualitatively describe biochemical reaction networks. Many extensions to the simple Petri Net model have been developed for various modelling and simulation purposes (212). Petri Nets and the π -calculus (215) are formal approaches commonly used in discrete event systems modelling where parametric information is scarce. An extension to the π -calculus which stems from the necessity to represent real biological scenarios arising from low numbers is the stochastic π -calculus (215). Other process algebras based projects like the Performance Evaluation Process

Algebra (PEPA) (45) workbench have also pushed forward the application of stochastic process algebras.

1.4.2 Qualitative systems models

Often the information provided by cell biology is qualitative in nature: interactions between components are described only as activation/inhibition. Is it possible to formalize the qualitative reasoning that has been applied by biologists to small data sets and generalize it to larger data sets? Qualitative models (90) enable us to describe the system in a less thorough way but provide us with valuable insight into the existence of stable states or limit cycles.

In Boolean networks, initially applied to genetic networks in (146), the state of the gene is approximately represented by a Boolean variable: active (on) usually associated with 1 or inactive (off) usually represented by 0. Gene products are modelled as either being present or absent, respectively. In this type of representation the functional links between nodes can be encapsulated by Boolean functions determining the state of a particular gene from the state of those genes connected to it. Dynamical behaviour is oversimplified but existent. The transitions in the network's state space are normally deterministic, with one output state for a given combination of input node states. Another characteristic is the synchronous nature of the transitions: outputs of all the network nodes are simultaneously updated. Due to the state space being finite, the number of states visited in a trajectory of the system is also finite. Eventually, the system will reach a stable steady state or a cycle state. The other states not belonging to an attractor are referred to as transient states. The set of attractor state plus the transients converging onto the attractor point make the basin of attraction. Calculating the attractors and basins of attraction is feasible by hand for small networks but constitutes an arduous task when network size goes up (146).

An original application of the Boolean network formalism has been performed in the study of global properties of large genetic networks (146). Properties like the average number of regulators affecting a gene and the type of

boolean functions, e.g. canalizing functions, representing the functional aspects operating on the promoter regions were studied. The attractors generated plus the transient trajectories and basins of attraction in state space were also investigated and local properties correlated with global network dynamics. The central feature of this simple type of Boolean network is functional, as opposed to cumulative change, which characterizes the ODE formalism.

The Boolean network formalism is a rather efficient though simplified way of studying large regulatory networks. Given that the states allowed are 0 or 1 the intermediate expression levels are neglected. Moreover, the transition between states for all the components of the network is synchronous. This assumption is in clear contrast with biological evidence. Several extensions of the original Boolean formalism have been proposed, especially to deal with extremely important aspects of biological systems: asynchrony between processes and therefore explicit continuity of time. Dynamics of asynchronous random Boolean networks have been analysed where asynchrony was generated by stochastic processes including: Poisson processes, random walks, birth and death processes, and Brownian motion (71). There is a crucial limit to the traditional Boolean networks, which is the restricted solution space arising from discrete time structured networks. Adaptations of a class of Boolean networks that evolve in continuous time has proved to have rich behaviour including aperiodic solutions (205). These are a natural extension of the formalism known as Boolean Delay Equations (BDE) developed and applied in the field of climate modelling and seismology (69, 97). We have, therefore, a special class of systems allowing us to work with a restricted state space but with a temporal functional space with a much higher resolution. Comparing the continuous time Boolean formalism with that represented by piecewise linear differential equations (67, 68) one sees that the latter differs from the former by allowing intermediate levels for the concentrations but still include an update or interaction function working based on Boolean logic.

Generalizations of the Boolean formalism to multilevel logic such as the kinetic logic have been a field of active research. This extension allows variables

to have several levels of activity (>2) and transitions to be asynchronous (250-253). It was developed initially to model genetic networks. It is nevertheless an analytical technique not for simulation purposes. Transitions between states of each variable are made to be switch like. It relies on the fact that at a threshold the effect of the regulator or the upstream protein rapidly levels up and hence also approximates the usual sigmoidal function to a step function. This fully asynchronous description has been proved to fit extremely well with the differential description of systems (251-253, 253). Again we emphasize that it is an analytical technique. This logical method has been implemented and demonstrated to be effective when applied to small genetic regulatory systems; successful examples are the λ phage infection in *E.coli* (249), the dorso-ventral pattern formation and gap gene control in fly *Drosophila* (225) and the network underlying flower morphogenesis in *Arabidopsis thaliana* (186). One interesting aspect of the kinetic logic modelling approach is that it enables the use of the logical method for induction of models, a problem referred in this literature as “reverse logic” (251-253, 253).

1.4.3 Stochastic systems models.

If stochastic processes are considered one needs to express the uncertainty inherent to fluctuating biological processes. A number of methods are dedicated to the problems of stochastic modelling. The small numbers of molecules of some of the components in a biological regulatory system (e.g. number of molecules of a specific transcription factor in the nucleus is in the tens order), invalidates the continuity assumption underlying the differential description. Moreover, the use of the deterministic differential operator may not be as appropriate as expected in the traditional view of a cell (e.g. mass-action kinetics). There exist considerable fluctuations in the timing of cellular events (e.g. time delay in transcription) that may have as a consequence unpredictable outcomes even given the same initial conditions for regulatory systems.

To overcome some of these problems one may choose to model the system as being constituted by a discrete number of \mathbf{n} particles and having a stochastic nature. Discrete amounts of certain molecules are the state variables and the evolution of the system is described by a joint probability distribution $p(\mathbf{n},t)$ obeying Equation 10 (94).

Equation 10

$$p(\mathbf{n},t + \Delta t) = p(\mathbf{n},t) \left(1 - \sum_{j=1}^m \alpha_j \Delta t \right) + \sum_{j=1}^m \beta_j \Delta t$$

In Equation 10 parameter m represents the number of reactions that occur in the system, the product $\alpha_j \Delta t$ stands for the probability that a certain reaction j is going to take place in the interval with duration Δt given that the system is in the state \mathbf{n} at time t . The term $\beta_j \Delta t$ stands for the probability of reaction j bringing the system in state \mathbf{n} from another state during $[t, t + \Delta t]$. Subtracting each side of Equation 10 by $p(\mathbf{n},t)$, and subsequently dividing by Δt and taking the limit $\Delta t \rightarrow 0$, we arrive at an expression commonly known as the master equation (94) (see Equation 11).

Equation 11

$$\frac{\partial}{\partial t} p(\mathbf{n},t) = \sum_{j=1}^m (\beta_j - \alpha_j p(\mathbf{n},t))$$

Equation 11 describes how the probability distribution of the system in state \mathbf{n} changes with time. It is a relatively intuitive description of the stochastic processes underlying the dynamics of the regulatory system. Nevertheless, analytical solutions are extremely difficult to find and even numerical simulation is an arduous task given the $\mathbf{n} + 1$ independent variables. The master equation can be transformed into a stochastic differential equation known as Langevin equation consisting of an ODE with a noise term (see Equation 12) (94).

Equation 12
$$\frac{dX_i(t)}{dt} = f_i(\mathbf{X}(t)) + \xi_i(t), X_i(t_0) = X_{0i} \quad 1 \leq i \leq n.$$

Equation 13
$$\begin{aligned} \langle \xi_i(t) \rangle &= 0 \\ \langle \xi_i(t) \xi_j(t') \rangle &= 2\sigma^2 \delta_{ij} \delta(t-t') \end{aligned}$$

Where ξ stands for the uncorrelated noise or Gaussian noise (Equation 13). The parameter σ (standard deviation) controls the magnitude of noise in the system. The noise term must be specified so that it mimics the effects of thermal fluctuations and so models successfully intrinsic noise (243). The Langevin theory for stochastic systems is strictly valid near a steady-state and when numbers of molecules are sufficiently large. The approximation may not always be valid for genetic regulatory systems (98). We will nevertheless use the Langevin approach to modelling stochastic effects in intra-cellular dynamics (section 2.3). It constitutes a fairly simple modelling approach and serves our modelling purpose (see project goals in section 1.6).

Due to the fact that many equations are necessary to model systems involving more than a few reactions or species when the master equation approach is used, Monte-Carlo algorithms such as those developed by Gillespie (98-103) are more commonly used to appropriately model the effects of noise and randomness in intracellular network dynamics. Stochastic simulation algorithms give us information on individual behaviours while the original master equation provides information on averages and variances of each discrete variable by operating on $p(\mathbf{n}, t)$. Improvements to this type of formalism have been developed and used also in signal transduction pathways (189, 190).

The stochastic simulation approach reflects closer the molecular reality of gene regulation but it requires the detailed knowledge about reaction involved, which sometimes is not available, and estimation of probability densities crucial to the process. One has to trace each reaction involved in the simulation. It is, therefore, computationally time consuming. Moreover, sometimes stochastic

effects level out with time and a more deterministic approach constitutes a good approximation (272).

Formalized rules for protein-protein interaction or protein-gene interaction may be extracted from the literature where most of the interactions are characterized qualitatively. Techniques such as those in STOCHSIM taking advantage of those rules (168) and use them during simulation to generate discrete reaction events. Each protein has domains which are “flagged” if a conformational change occurs. The likelihood of the reaction involved in the conformational change is determined by the Monte Carlo method by discretization of the master equation into small time steps. These kinds of approaches utilize both qualitative and quantitative information about the system. Nevertheless, for large intracellular networks they become time consuming.

1.5 Epigenetic attractor landscapes

The traditional approach to understanding cell regulation involves breaking the information into signalling pathways, linking activated cell membrane receptors to gene transcription, with the intent of gaining better insight into the function of the underlying control mechanisms. Although the tendency is to attribute specific functions to particular signalling molecules or pathways the biological reality is far from being completely localized and determined by a specific pathway. It is scattered across numerous pathways exhibiting crosstalk (see section 1.2.4). A “globalist view” (129, 131) of the dynamics of the network is extremely relevant for a characterization of the non-linear effects of network causal interactions. The coherent nature of whole-cell behaviour during, for example, the switching from phenotype to phenotype, may be a reflection of higher order dynamics of global intracellular networks (129, 131). The concept of epigenetic attractor landscapes illustrates well the globalist, pattern oriented nature of cell fate decision.

1.5.1 Cell phenotypes as genetic network attractors

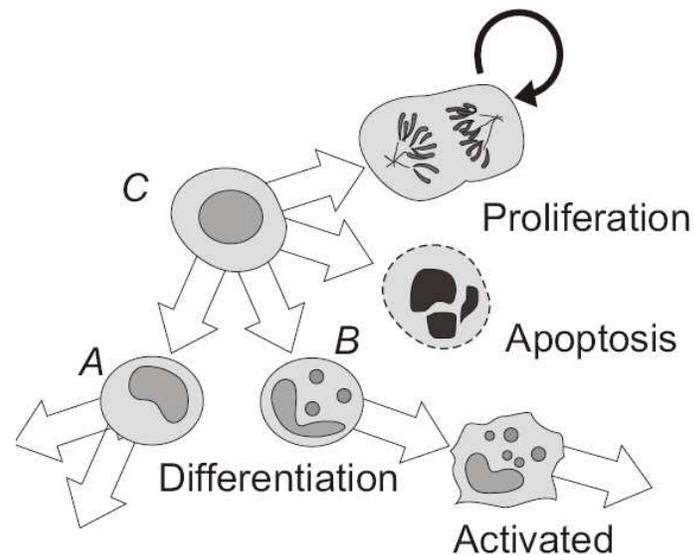


Figure 15 Simplified depiction of the possible outcomes or cell phenotypes emerging from a stem cell. To each cell phenotype a specific gene expression pattern is associated (132). The positive feed-back loop represents the reiteration of the proliferation program cell division (130).

Although each of the 250 distinct kinds of cells in the human body have extremely specialized function they carry an identical copy of the DNA. This constitutes an intriguing problem for biology and applied mathematical and computational sciences. Understanding the functioning of organisms at the molecular level requires necessarily crucial information on which genes are expressed: when, where and to what extent. Gene expression regulation is achieved through transcription factor networks (section 1.2.6). The genome has a vital role in cellular processes control such as the response to environmental stress and differentiation in developmental programs. To understand how genes are implicated in the control of intracellular and intercellular processes, a systemic approach is indispensable to link sequences coding for proteins and regulatory systems behind genetic expression profiles. Take as a paradigmatic example the process of cell differentiation. One might describe it through a process of selection of distinct existing states of expression in genetic regulatory networks. Each type of cell (see Figure 15) corresponds, under this approach, to

a distinct “attractor” state of the intracellular network dynamics (131). Cell differentiation involves, therefore, the transition of the cell state from one attractor to another (Figure 16).

There have been attempts at understanding how in large networks of master regulator transcription factors the switch-like behaviour or differentiation arises (54, 55, 127, 132). Experimental observations have also shown that distinct cell differentiated phenotypes exhibit a different expression pattern (133) or that signals that induce the same phenotype have equivalent gene expression patterns (131). This constitutes a strong case for globalist approaches to intracellular network functioning and a strong indication that therapy design focusing on single targets may be inefficient (169).

Understanding functionality requires considering dynamics on top of network structure, as was stated before. Consider a singular network node i , gene or protein. Its state may be characterized by a concentration value $X_i(t)$ of the protein coded for by the gene the node represents, or of an activated form of the protein the node represents. The node’s state arises as a response to the activity of incoming nodes. Together with interaction rules obeying a certain logic (40) or continuous function, the incoming nodes determine how the level of a certain node “ i ” is perturbed by its interaction partners. The dynamics of the system is, therefore, the combined evolution of the concentration levels $X_i(t)$ of the set nodes of the network, and is amenable to a representation in high-dimensional state by a state vector $\mathbf{X}(t) = [X_1(t), X_2(t), \dots, X_N(t)]$ (N node network). Although one cannot draw an N dimensional state space if $N > 3$, a graphical representation of how a network would evolve may be reasonably intuitive, if we think of a system as following an epigenetic landscape (see Figure 16).

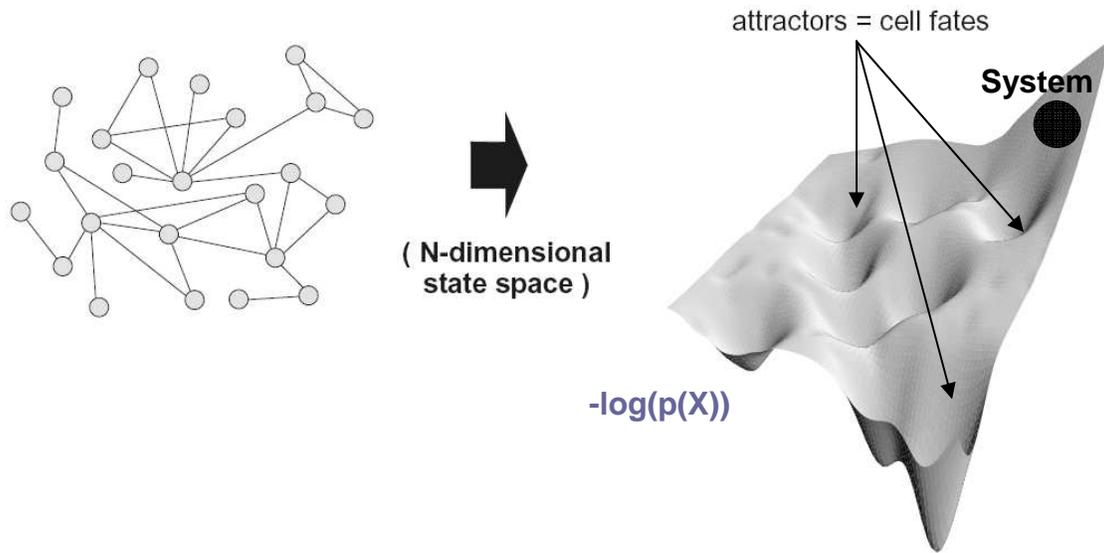


Figure 16 Simplified graphical representation of highdimensional state space attractor landscapes. Left: graph of an N node network. Right: intuitive representation of an epigenetic landscape. The pits represent the minima of an energy function associated with the state vector $X(t)$ or just the most probable states (highest frequency) according to numerical simulations. Notice that the N -dimensional state space is represented by two independent variables only. These can be seen as a projection of high-dimensional state space onto 2 essential variables, determined for example by principal component analysis. The cell is assumed to stabilize in one of these states (attractor=cell fate) according to specific stimuli: extracellular signals initiated by ligand binding and transduced by pathways. The landscape can be thought to be similar to Waddington's 'epigenetic landscape' (261). Here, $p(X)$ is probability of state X and the gray surface corresponds to $-\log(p(X))$, with s being a particular stable combination $[X_1, \dots, X_N]$.

The interactions of the network impose restrictions on the evolution of the state vector $X(t)$ in the state space. This limitation of the high-dimensional trajectories is where the overwhelming complexity of large networks subsides and a reduced set of possible network behaviours arises such as multiple highdimensional attractors.

To overcome the lack parametric information on genes and their precise connectivity, Kauffman opted to study a simple class of dynamical system, random Boolean networks, as models for testing the hypothesis of cell phenotypes being associated with high-dimensional attractors (146). The idea was to generate randomly connected networks and study the effects of properties such as the average connectivity and different classes of Boolean functions. Kauffman was able to show that in the presence of sparse connectivity, the

global dynamics of a highdimensional network demonstrates mostly a set of point attractors and less frequently disordered regimes. The attractor states (Figure 16) can be assigned to various differentiated cell types in a multicellular organism as was previously mentioned. Another possible association for the attractors in Boolean networks is that they stand for distinct cellular phenotypes such as proliferation (cell cycle), apoptosis (programmed cell death), together with differentiation (127, 128). This interpretation can provide new insights into structural aspects and dynamics behind cellular homeostasis and cancer progression. In the case of a certain mutation affecting a structural property that weights in a negative way on the probability of the network entering the apoptosis attractor, the cell is unable to begin this emergency procedure and will exhibit uncontrolled growth. Another possible analogy for tumourigenesis is the existence of a large basin of attraction for the proliferation attractor (132). This may result in excessive stem cell population growth, which is also a hallmark of cancer (115).

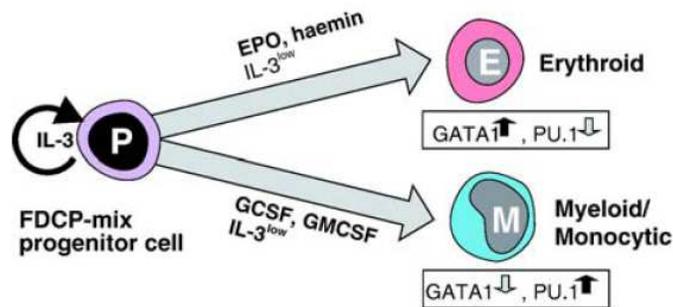
A central concept in dynamical systems is that of structural stability, which can be defined as the persistent behaviour of a system under small perturbations. This is definitely a property of real genetic networks or any intracellular network since homeostasis as to be maintained in the presence of external perturbations and stimuli, in metabolism and developmental programs. Boolean networks capture this important phenomenon as the system usually converges to the available attractors when some genes, represented by nodes, have their state perturbed (146). The observed dynamics of a Boolean network with attractor states naturally models the fundamental properties of cell fate dynamics: mutual exclusivity between cell phenotypes^h, robustness and all transitions between cell fates arising from the combinatorial complexity of external and internal signals. Nevertheless, the deterministic parallel updating scheme associated with the traditional Boolean network formalism is not realistic. The significance of the Boolean paradigm results is further limited if we take in consideration the omnipresence of noise in biological systems; certain attractors

^h For example a cell either proliferates or enters apoptosis.

in Boolean networks have been proved to be an artifact of the updating scheme (156) and thus not robust to noise. We will in upcoming sections use, therefore, the ODE and SDE formalisms (see section 1.4.1 and 1.4.3) to avoid the pitfalls of Boolean dynamics.

1.5.2 Deterministic and stochastic attractor selection

A



B

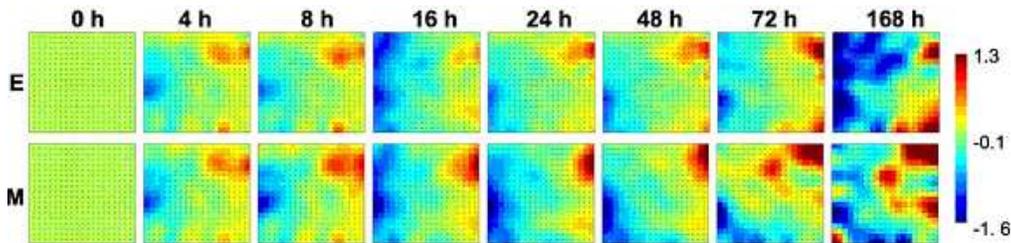


Figure 17 Differentiation of progenitor cells into erythroid and myeloid/monocytic cell phenotypes. A) Simplified diagram representing two possible end point differentiation outcomes from FDCP-mix progenitor cells: Erythroid (transcription factors GATA1 up regulated and PU.1 downregulated) and Myeloid/monocytic (GATA1 downregulated and PU.1 upregulated). B) High-dimensional trajectory representation of the possible differentiation paths, Progenitor (P) → Erythroid (E) or Progenitor (P) → Myeloid (M), in time. Each colour coded figure represents the pattern generated by a Self-Organized-Map (SOM) application to the gene expression dataⁱ at a point in time (0h, 4h, ..., 168h). For the analysis of the results of (133) log₂-transformed relative expression levels for each gene were used, with expression at 0h being the reference. Each cell type, E or M, as a distinct expression pattern (see last column). Taken from (133).

The concept of attractor landscapes has been extensively explored in cell differentiation studies in both deterministic and stochastic frameworks. Deterministic studies of low order gene regulatory networks such as bistable

ⁱ <http://www.childrenshospital.org/research/ingber/GEDI/gedihome.htm>

switches have been given a thorough theoretical investigation and constructed *de novo* (95). The conditions for existence of multistability are naturally linked to parameters controlling self-stimulation and cross-stimulation (180) and are thought to be a clear framework to understand selective expression in different cell types. The issue of bistability has also been successfully characterized in several stochastic gene regulatory networks, e.g. λ phage decision switch in *Escherichia coli* coupled or uncoupled to the quorum-sensing signalling pathway or the SOS signalling pathway (254). Noise was proven to be essential for switching from attractor to attractor when the system has direct coupling with a signalling module.

There have been generalizations to high dimensional switches of the switching behaviour observed in small circuits. These models were also understood in both deterministic scenarios (55) and noisy ones (270). Regarding the action of noise as an attractor selector, Kaneko and coworkers (270) showed that its strength and duration in globally couple map lattices is sufficient for transition from attractor to attractor in a system capable of exhibiting ordered phases, synchronized phases and turbulent phases. Yet, in subsequent studies they also came to the conclusion that if the circuit is coupled to a signalling pathway, the efficiency of the selection process is increased (145). Noise has also been proved to be essential for optimal attraction selection in evolutionary systems coupling metabolic activity to gene expression pattern in the absence of any adequate signalling external inputs (88). The gene expression pattern selected in integrative metabolic-transcription network simulations corresponds to the maximal cell growth rate.

Cell fate being associated with a high-dimensional attractor or gene expression pattern, as first proposed by Kauffman, has been confirmed experimentally by Huang and coworkers in neutrophil differentiation (131) and in a hematopoietic cell line with convergence of high-dimensional trajectories (133) (see Figure 17). Huang and coworkers developed a pattern oriented visualization tool to characterize cell phenotypes^j (75). This tool applies a self-organized map

^j <http://www.childrenshospital.org/research/ingber/GEDI/gedihome.htm>.

to the expression data and compresses it into a 2D image by clustering genes with similar activities. In this way, by visual inspection one can verify, for example for the case of haematopoiesis (133), that different cell types have distinct expression patterns which are stable when a cell is fully differentiated. These correspond to high-dimensional attractors to which the system converges over time and in response to a specific signal inducing a particular cell type (see Figure 17 B for sequence of patterns). In other studies, “*return of noise-induced deviations of the transcriptome from the border of the basin of attraction back to the attractor state*” has also reinforced the idea of high-dimensional attractors (50).

Let us consider the following potential landscape (Figure 18) as a paradigmatic example representing point attractor selection by external signals (S_1 and S_2 , see Equation 14) in intracellular network dynamics. It will serve as an introduction to the theory explored in chapters 2 and 3.

Equation 14

$$U = \Delta r^2 \times (\Delta r^2 - 2) + (1 - S_1(t)) \times (\Delta r - 1)^2 + (1 - S_2(t)) \times (\Delta r + 1)^2$$

with

$$\Delta r = [TF_Y] - [TF_X]$$

Evolution of the system could be modelled using a Langevin type of approach (see section 1.4.3) by differentiating the potential with respect to Δr and adding a noise term.

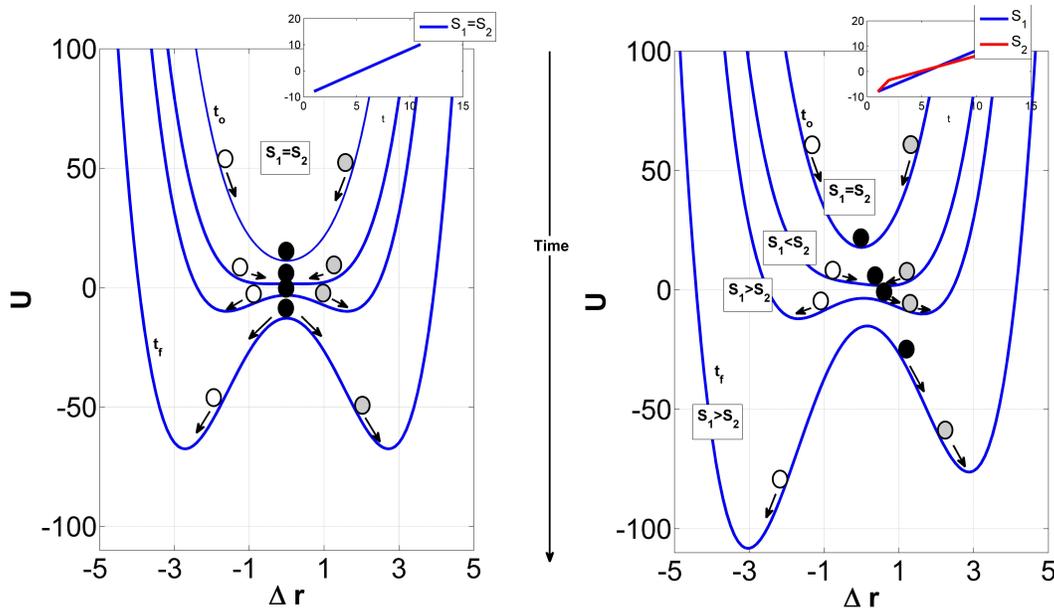


Figure 18 Potential for a hypothetical gene regulatory decision switch (Figure 19) driven by external signals (S_1 and S_2 , see also Equation 14) leading to attractor selection. Left: signals have equal profiles. Right: signals change differently with time. The differences are integrated by the transcriptional system. U , potential function. TF concentration difference = $\Delta r = [TF_Y] - [TF_X]$. The system is represented as full circles, white, grey or black.

This type of potential landscape was already used by Huang and coworkers (133) in understanding the action of signals on a toggle switch of two transcription factors (GATA1 and PU.1) with dimerization and self-regulation. This circuit is involved in bipotent progenitor cell differentiation into erythroid (GATA1 up-regulated, PU.1 down-regulated) or myeloid/monocytic (GATA1 down-regulated, PU.1 up-regulated) states (see Figure 17 A). For a discussion about specific details of the most probable potential shape describing the GATA1 and PU.1 we recommend the original paper (133). Here, we focus on general principles to highlight possible avenues for understanding cell fate decision.

External signals can induce only one attractor at the final instant if the differences between them are sufficiently large. Even if the system undergoes a bifurcation it can return to a scenario of monostability (section 8.6). If so, the only attractor populated is the final attractor. This particular “deterministic-instructive” setting is also known as a progression switch (109). Another avenue for

understanding the action of signals brings a higher contribution to fluctuations in gene expression. A symmetric change (see Figure 18, left), also known as “deterministic-selective”, forces a bistable regime, through a pitchfork bifurcation, with equally large attractor basins for each of the attractor states at t_f . The end attractors are populated as a result of a combination of stochasticity and variations in the position in the basin of attraction at instant t_0 (Figure 18, left). Let us observe the cases represented in Figure 18 (left). The system represented as full coloured circles can either start at the only attractor or deviated from it at t_0 . If the relaxation time to the attractor is slow compared to the changes in the potential shape induced by $S_1(t)$ and $S_2(t)$, then the system can be forced to end in either the attractor at $\Delta=-3$ or at $\Delta r=3$, depending on its initial position. Alternatively, if it starts at the only attractor at t_0 then its final position at t_f will be determined by noise. However, the cell relies on the combination of a wealth of asymmetric external signals to decide which phenotype is more appropriate to the circumstance in hand (96). Therefore, the changes in the potential landscape represented in Figure 18 (right) are a much more realistic scenario. Asymmetries between external signals over time, $\Delta S(t)=S_1(t)-S_2(t)$ (referred to in this thesis as external asymmetry), induce adequate changes in the epigenetic landscape, with the size of basin of attraction and strength of attractor reflecting both $\Delta S(t)$ magnitude and sign. The probability of reaching each of the attractors is not only dependent on initial position but also on $\Delta S(t)$. Therefore, given the right conditions, e.g. fast relaxation times, the final distribution over attractors will have memory of $\Delta S(t)$ (see Figure 18, right, for one example where the system, represented as a black full circle, should end in the attractor at $\Delta r=-3$ but doesn't due to memory of an interval where $S_2>S_1$). In this last scenario, noise is also an important factor in memory robustness. If its amplitude is sufficient then transitions over the potential barrier (hill top in Figure 18) are still possible and any effects of $\Delta S(t)$ are completely blurred, degrading therefore circuit memory.

Another relevant aspect for attractor selection is the speed with which the critical zone is crossed (investigated extensively in chapter 2). This mechanism may also be used to trap the system in one of the basins of attraction (200) as

was briefly explained before. Hence, the circuit memory of past events is dependent on the sweeping process.

The mechanism involving epigenetic landscape structure changes, e.g. by bifurcation, has been proved to be more efficient in selecting attractors with slightly larger basins of attractions in comparison to the system not undergoing bifurcation (191). The differences in basin size and depth of the attractors can be, on the other hand, a result of intrinsic system properties. For instance in cancer point mutations in genes induces a larger proliferation basin of attraction, which is thought to recurrently tilt the system into cell division even if other anti-proliferative signals are present.

1.6 Project goals

This project aims at understanding how cell fate decision emerges from the overall intracellular network connectivity and dynamics. To achieve this goal both small paradigmatic signalling-gene regulatory networks and their generalization to high-dimensional space were tested. Achievement of the project goals included the following tasks:

- Understand the importance of the rate of time-dependent bifurcation parameters in canonical models of non-equilibrium phase transitions (section 2.2);
 - Study numerically the effects of a time-dependent bifurcation parameter and a transient external asymmetry on attractor statistical selection in the presence of fluctuations (section 2.2.5);
 - Devise analytical expressions describing attractor selectivity dependence on critical parameter sweeping speed, transient asymmetry rates and noise strength (section 2.2.4);
- Understand the effects of combinations of external input signals on small gene regulatory network motifs with time-scale separation (section 2.3);
 - Study attractor selectivity induced by a combination of time dependent signals; prove the importance of the rate of sweeping

- through the critical region for cell fate decision in the presence of fluctuations (section 2.3.2);
- Analyse the proposed original mechanism for cell fate decision in the light of non-equilibrium phase transition models (section 2.3.2);
 - Study the effect of time-scale separation on the sensitivity of the genetic decision switch to differences in external signals in the presence of fluctuations (section 2.3.2);
- Study the response of a high-dimensional gene regulatory decision switch to a combination of external input signals (section 3);
 - Analyse path-dependent effects on high-dimensional attractor selection arising from combinations of N external signals (section 3.2);
 - Evaluate the importance of sweeping speed, externally induced asymmetry and noise for high-dimensional attractor selection (section 3.3);
 - Study perturbations induced by non-specific interactions of small molecule kinase inhibitors in real networks under normal and pathological scenarios (section 4);
 - Understand if additional off-target binding induces substantial differences in signalling inputs to gene regulatory networks. For this end a completely parameterized Epidermal Growth Factor Receptor model will be used (section 4.2);
 - Discuss the efficiency of possible multitargeted therapies taking into consideration cross-talk between pathways, integration of signals by gene regulatory networks and noisy attractor selection (4.4);

2 Attractor selection in integrative noisy signalling-transcriptional regulatory networks with time-scale separation

2.1 Introduction

Temporal gene expression profiles of large groups of genes have been obtained for several cell types when stimulated by specific growth factors (e.g. EGF stimulation of HeLa cells (9)) or during execution of the cell cycle program (265). Such data is crucial for a detailed understanding of genetic control of cellular responses to environmental circumstances (39), and execution of transcriptional programs. As was previously described the project focuses on changes in the transcriptional program induced by signalling protein pathways with crosstalk. We will, in further sections, analyse several characteristics of integrated circuits. Three simulation sets will be performed. One on a canonical model used to study non-equilibrium phase transitions (section 2.2). This will be performed with the objective of understanding simple principles that will be extrapolated to a second system, a genetic decision switch with a combination of 2 external signals and time-scale separation (section 2.3). Finally, another set of simulations will be performed on a high-dimensional generalization of the decision switch (section 3) with N external signals.

The integration of cue signals is both performed by the signalling system, whose function is akin to a multilayer perceptron (see Figure 10 D) or combinatorial decoder (122) (see Figure 8), and the transcriptional machine (176). The signalling module generates a classification of the combination of inputs (activated receptors) based on their level (dosage) and timing (122) (see also section 1.2.5), recurring to the multiple interlinked chemical processes (most commonly kinase/phosphatase reactions). The transformation outcome is a combination of activation concentration profiles of the output nodes (S_1 and S_2 in Figure 19), whose shape (signal duration, signal amplitude, signalling rising and

relaxation times) has been correlated through combined experimental and simulation studies with the induced genetic programs. For example, the temporal control of the signalling module containing the I κ B kinase (IKK), its substrate inhibitor of NF- κ B (I κ B), and the fundamental inflammatory transcription factor NF- κ B can induce selective gene activation when stimulated by Tumour Necrosis Factor Receptor (TNFR) and other concurrent pathways (263). Other pathway systems such as the Epidermal Growth Factor Receptor (EGFR) have also been the focus of extensive experimental studies associating signalling output activation with cell phenotype. In these studies, transient versus sustained activation of the output node indicated the competition between two mutually exclusive cell phenotypes, proliferation and differentiation (195), respectively.

Regarding targeting strategies of complex networks, the question arises of how to modulate the biochemical processes as to induce the correct combination of concentration profiles of output nodes and how to interpret this combination. Targeting single nodes belonging to the central processing core of the signalling information processing layer has not been as fruitful as desired (155). The multiple mutation scenario of cancer associated networks (64) turns these abnormal cells into very robust systems (155). Moreover, the inherent redundancy of molecules in cell signalling exhibiting extensive cross-talk (181) also renders the development of targeting strategies a highly complex endeavour. In Figure 19 we represent a paradigmatic global system and a class of possible drugs to be considered in a further section (4.2) dedicated to tuning and modulating of gene expression.

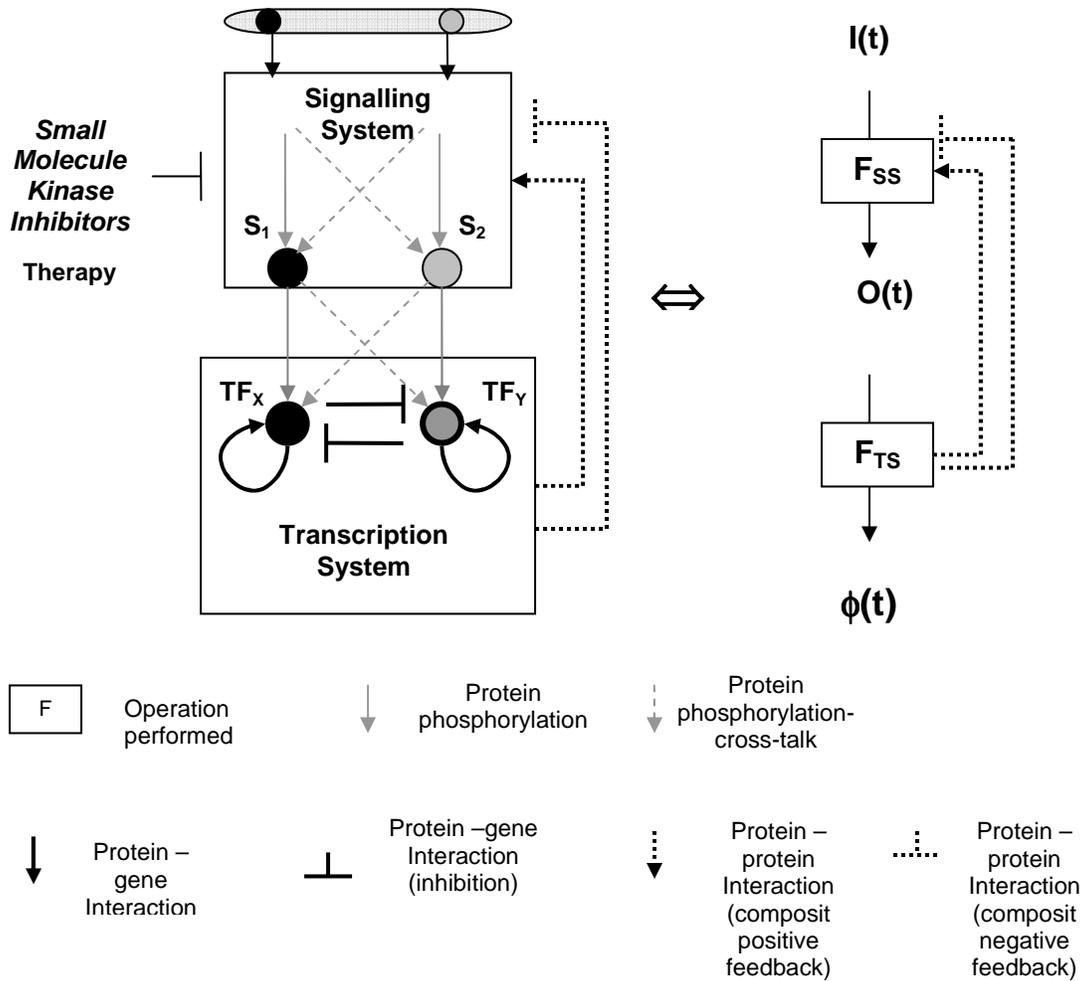


Figure 19 Generic integrative signalling-transcriptional system. Left-representation of a paradigmatic signalling-transcriptional system. $S_{1,2}$ -stands for output nodes of the signalling processing unit, commonly kinases. $TF_{X,Y}$ - represents the transcription factors activated in response to an incoming signal. Right- abstract representation of each of the functions of each of the systems, signalling and transcription. F_{SS} - represents the signalling system. F_{TS} - represents the transcription system.

The function performed by each of the modules (F_{SS} signalling and F_{TS} transcription) is dependent on the wiring of the network, W_{SS} (see Equation 15) and W_{TS} (see Equation 16) (80), and the set of inputs to the module over time. Here the wiring is not only associated with the connectivity matrix between system's nodes but also with the actual parameters (kinetic parameters such as phosphorylation, transcription and degradation rates) determining nuances in the dynamical regimes.

Equation 15

$$O(t) = F_{SS}(I(t), W_{SS}) \circ \phi(t)$$

Equation 16

$$\phi(t) = F_{TS}(O(t), W_{TS})$$

Composite motifs (7) involving feedback loops between the transcriptional and signalling systems will not be considered in the simulation performed ahead (see Figure 19). These can give rise to much more complex behaviours such as damped and sustained oscillations as well as differentiation (188), when delays associated with the macromolecular transport (236, 237), transcription and translation are considered. This aspect has been explored before (188). Delays can be discrete (188) or distributed (218). In conjunction with dimerization reactions and activation of transcription factors by phosphorylation Smolen and coworkers (236, 237) proved that changes in parameters (transcription rates) representing action of external inputs (phosphorylation) and the action of delays induces bi-stability and memory in the system.

A representative shape of the output of the signalling layer (96) (see also Figure 19) can be visualized in Figure 20.

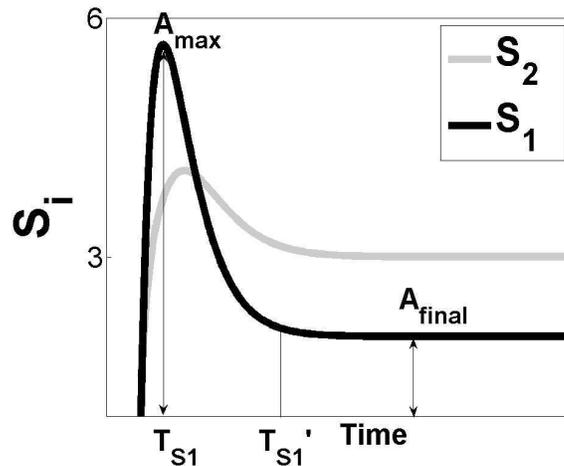


Figure 20 Generic representative signal shapes for incoming signals to the transcriptional network. A_{max} -maximum amplitude. A_{final} - final amplitude. T_{S1} - rising time. $(T_{S1'} - T_{S1})$ -relaxation time. The profiles were generated through MATLAB R2010b gamma distribution function *gmpdf*.

Each of the signal characteristics may have specific effects on the transcriptional machinery, and, consequently, on the attractor/cell phenotype selected. As will be analyzed by bifurcation analysis in section 2.3 (and also in section 8.6), the signal amplitude influences the number of available attractors in the system. See for instance recent contributions by Guantes and coworkers (109), and Pfeuty and coworkers (210), on the importance of self-stimulation for generating multiple dynamical regimes and flexible epigenetic control. The effect of signal duration has also been studied (109). It has influence on the capacity of the circuit to generate persistent regimes or memories. The capacity of bistable switches for frequency selectivity is also an important achievement (108, 236).

One particular characteristic that has never been explored is the rising and relaxation times (see Figure 20). If these are considered the analysis becomes more complex. Significant effects near the bifurcation point may shed some light on additional cell fate decision mechanisms. Pattern induction or selection through testing regimes of parameter sweeping has never been approached in genetic network theoretical studies or in experimental studies to the best of our knowledge. The following section is dedicated to fundamental aspects behind the new mechanism proposed. Models from statistical physics will help to clarify and formalize the main aspects of cell fate decision, as seen from a perspective of bifurcation parameter sweeping experiments.

2.2 Symmetry breaking in parametrically driven far-from equilibrium systems with time-dependent external fields

The bistable potential has been a reference model for phase transitions in Statistical Physics, specifically 2nd order or continuous transitions^k (see Figure 21 C). On the other hand, 1st order or discontinuous transitions^l are usually modelled through a slightly different canonical model (see Figure 21 A).

^k Examples such as the ferromagnetic transition, the superfluid transition, and Bose–Einstein condensation fall under the category of second–order phase transitions

^l several important transitions can be included in this category, including the solid→liquid→gas.

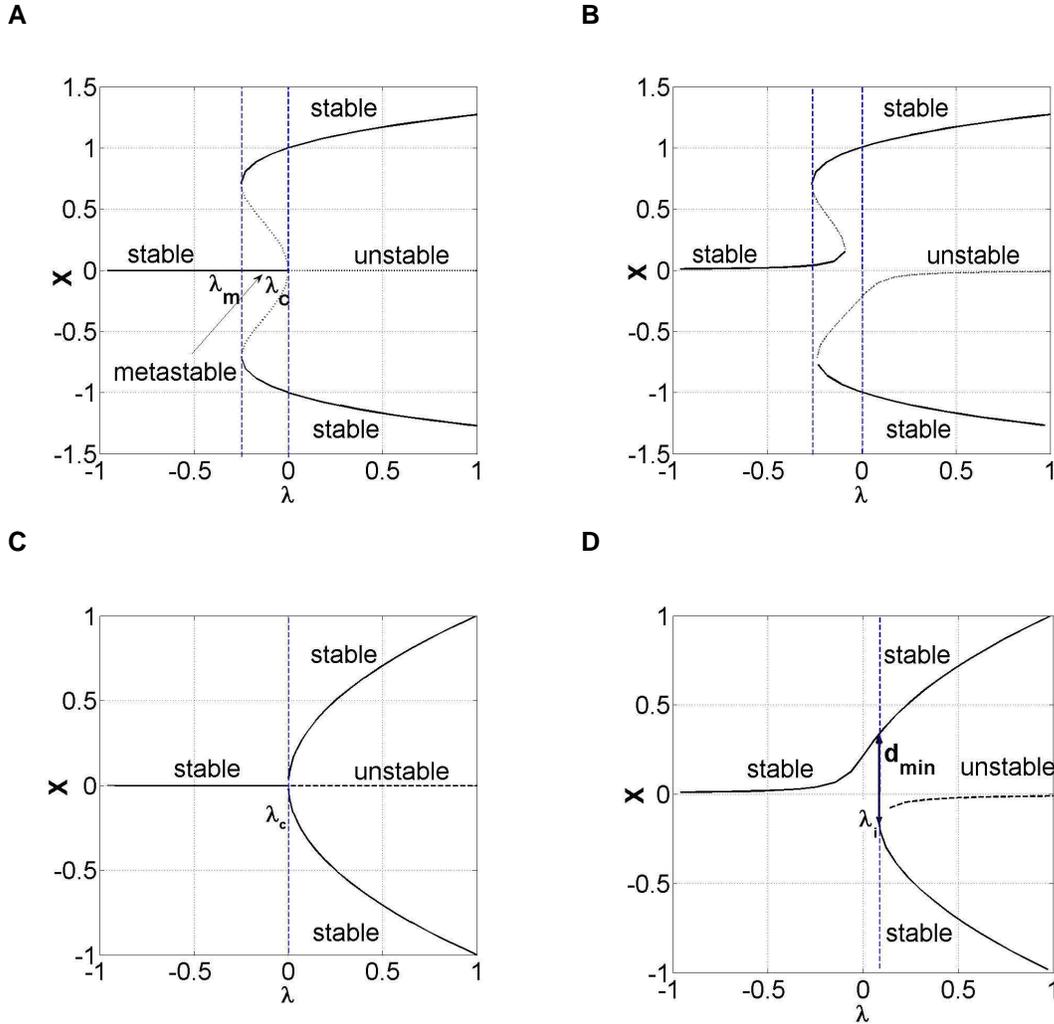


Figure 21 Attractors available for varying values of the critical parameter λ in canonical models of phase transitions. A) Subcritical pitchfork bifurcation diagram with critical value $\lambda_c=0$ (see also Equation 17). B) Imperfect bifurcation corresponding to the subcritical pitchfork bifurcation for $g=0.01$ (see also Equation 17). C) Supercritical pitchfork bifurcation, with the critical value $\lambda_c=0$ (see also Equation 18). D) Imperfect bifurcation resulting from an external field $g(t)=0.01$ (see also Equation 18). The bifurcation diagrams were all created with the software XPPAUT (18). The presence of a constant external field g induces a disconnection of the solution branches (see B and D). In D, d_{\min} represents the minimum distance between branches (see also Equation 19). $d_{\min}=\Delta x=(X_{\text{upper branch}}-X_{\text{lower branch}})$ at $\lambda=\lambda_i$.

1st order phase transitions can be reduced near the bifurcation point to a so called subcritical Pitchfork normal form depending only on the order parameter (see Figure 21 A and Equation 17):

Equation 17

$$\dot{X} = -X^5 + X^3 + (\lambda - \lambda_c)X + g$$

This normal form gives rise to a transition, by varying the critical parameter λ , from a region where only monostability is observed ($\lambda < \lambda_m$) to a region where 3 stable states are encountered ($\lambda_m < \lambda < \lambda_c$). The stable state corresponding to $X=0$ between λ_m and λ_c constitutes a metastable state (Figure 21 A). If fluctuations are considered, the system leaves the metastable state much more easily when compared to the other set of solutions. Beyond λ_c , $X=0$ loses its stability. The stabilizing quintic term present in Equation 17 assures that for larger values of λ the trajectories are not driven to infinity.

In our work we will focus on systems falling under the category of continuous phase transitions. Although experimental work has been performed on comparing the likelihood of the two types of bifurcations occurring, for example, in differentiation of cells in blood cell line commitment (78, 133), we will focus our studies on the possible applicability of phenomena arising in systems described by supercritical Pitchfork bifurcations (see Figure 21 C and D). Specifically, our interest lies in the effects of time varying parameters, such as g and λ , in attractor selectivity under fluctuations, and its applicability to cell fate decision.

Sweeping through the bifurcation point in the presence of external noise was an important contribution to the area of nonequilibrium macroscopic systems coupled to fluctuating environments. Several theoretical endeavours in the 80s and 90s of the last century (161, 191, 202) focused on solving this paradigmatic problem analytically. The initial drive for this particular research arose in the field of prebiological evolution (160, 162, 214). Sugars and amino acids, vitally important for biology, are found only in a particular chiral form. For sugars only right enantiomers are found and amino acids only left. This contrasts clearly with inorganic nature where one can find both chiral forms in equal proportions (160). Some generic models state that this is a result of a supercritical pitchfork bifurcation that lead to spontaneous chiral symmetry breaking at the earliest

stages of prebiological evolution (162). The primary problem is whether the additive external field is enough to cause a clear break in the symmetry or if the effects of the bias are suppressed by fluctuations. Particular applications to noisy electronic circuits (161) have shown that the speed with which certain parameters of a bi-stable system vary is important in creating asymmetries in the distribution of trajectories falling into there several attractors in the system. The normal form for a supercritical Pitchfork bifurcation or a second order phase transition affected by an external field is represented in Equation 18 and depicted in Figure 21 C and D. Observe that the cubic term is now a stabilizing effect (compare with Equation 17).

Equation 18

$$\dot{X} = -X^3 + (\lambda - \lambda_c)X + g$$

In the case of the external field being zero, $g=0$, the previous equation admits for $\lambda < 0$ the unique asymptotically stable steady-state (sst) solution $X_{sst}=0$. For $\lambda > 0$, 3 solutions appear: the asymptotically stable branches given by $X_{sst} = \pm\sqrt{(\lambda - \lambda_c)}$ (see Figure 21A) and the trivial unstable solution $X_{sst}=0$. The steady state branches in the presence of an external field g can be determined by finding the roots to the cubic equation $-X^3 + (\lambda - \lambda_c)X + g = 0$. The analytical solutions are fairly complicated. We are primarily interested in understanding what happens in the critical region. The bifurcation point or intersection disappears, giving place to a connected set of solutions and a disconnected branch (see Figure 21 D). The point (λ_i) where the 3 solutions appear is displaced from the original bifurcation

point by a distance: $\lambda_i - \lambda_c = \frac{3}{2} g^{\frac{2}{3}}$ ^m. The minimum distance, $d_{min}=(\Delta X)_{\lambda=\lambda_i}$

^m The point where the 3 solutions appear can be estimated by recurring to the discriminant (Δ) of a generic cubic equation, $aX^3 + bX^2 + cX + d$, and finding its roots, $\Delta = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd = 4\lambda^3 - 27g = 0$. When the discriminant is zero there is a double real root, which is the situation of interest for the problem in hand.

$=X_{\text{upper branch}} - X_{\text{lower branch}}$ (see Figure 21D), between the two branches is given by Equation 19.

Equation 19

$$d_{\min} = \frac{3}{2^{\frac{1}{3}}} g^{\frac{1}{3}}$$

The unstable branch of solutions far below or above the critical value λ_c can be estimated to be displaced approximately by $-\frac{g}{(\lambda - \lambda_c)}$. The effect of the asymmetry as a state selector, for $g \ll 1$, is expected to be more pronounced near the bifurcation point due to the disconnection between branches. The asymmetry plus noise (e.g. Gaussian) induces a higher selectivity of one of the branches when the bifurcation parameter is varied in time with lower speeds (161, 191). This finding has relevance to the problem of cell fate decision. As was described above, there have been put forward competing approaches (133). One privileges the role of noise in a fixed landscape. Others give emphasis to the action of asymmetric incoming signals. In our opinion, the transcriptional landscape integrates a wealth of input combinations. Signals not only drive transcriptional landscape changes but also create the appropriate asymmetries enhancing the probability of reaching the attractors that encode the adequate evolutionary response. The question arises as to what extend are the combinations associated with only one attractor, and which characteristics are important in attractor selection. If signal shape is relevant, the effect of rate of changes inherent to the signal is going to have specific consequences. The effect of rising times to maximum amplitudes and relaxations times to final amplitudes can be understood through simple canonical models.

The effect of the external asymmetry on selecting a specific branch of steady state solutions can be extended by making the asymmetry depend on time. This tests the effect of the dynamics of parameter driving against system's natural dynamics. The motivation for this particular approach is to understand the effects of a path in phase space on the distribution of trajectories across

branches. Essentially we will try to test memory effects of changing paths in the phase diagram. This problem may help to understand the gene regulatory decision switch to be studied ahead (see section 2.3).

2.2.1 Critical fluctuations and long time-scales

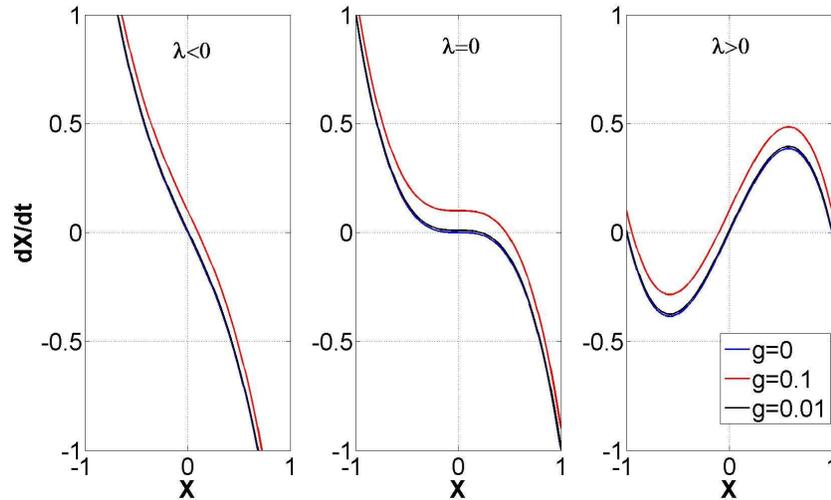
For a supercritical Pitchfork bifurcation, near the critical point one observes both a “slowing down” of the relaxation processes and an amplification of the amplitude of the fluctuations. Intuitively, observing Figure 22 A (central graph), for $g=0$, the origin is still stable, but less strongly, since the linearization vanishes. Solutions of Equation 18 no longer decay exponentially but as a much slower algebraic function of time (see Figure 22 B). This is seen as a signature of a second order phase transition or supercritical Pitchfork bifurcation.

The effect of critical slowing down has been an active field of research for a long time (179). It was generally believed that due to the mentioned effect one would have to wait longer times to achieve relevant observations when investigating second order phase transitions. Another signature of second order phase transitions is the amplitude of the critical precursory fluctuations (239). By analysing the growth of the noise as the order parameter is changed, we are able to deduce if the system is in the vicinity of the critical point. This method has been systematically used in determining, for example, material failure (13, 93), financial crashes (138, 139, 240) and earthquakes (37, 239).

The effect of slow bifurcations acting synergistically with fluctuations is thought to be a prominent effect in selecting the branch of solutions favoured by the external field g near the original critical value λ_c . Here, a distance is present between the point where 3 solutions appear and the original critical value. If the bifurcation parameter is slowly varied in this area, one allows for the system to converge to the lowest point in the potential and average out the effect of the diffusion caused by the thermal noise. Also, even after the point where bistability arises, the distance between branches grows with $\sqrt[3]{g}$. The location of the unstable branch approximates the asymptote $X=0$ and the relative size of the basins of attraction is reduced (see Figure 21 D). Consequently, it is expected

that until a small distance after the 3 solutions arise the branch selection is already completed.

A



B

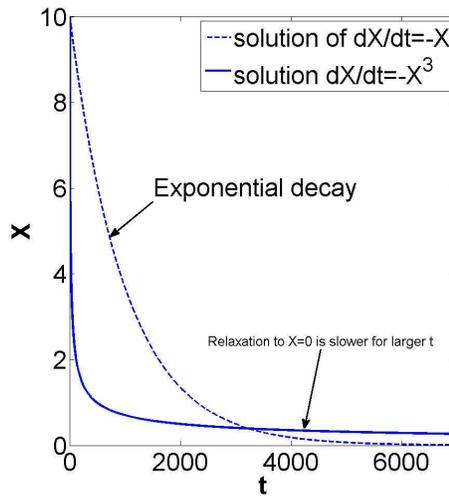


Figure 22 Critical slowing down observed at a supercritical Pitchfork bifurcation or second order phase transition. **A)** Near the critical point $\lambda=0$ the system observes a sluggish evolution due to the inflexion point in the derivative (central picture). Left: $\lambda=-1$. Center: $\lambda=0$. Right: $\lambda=1$. **B)** Time-series of X for exponential decay and the solution of Equation 18 with $\lambda=\lambda_c=0$ and $g=0$. For $t \rightarrow \infty$ the solution of Equation 18 is slower than exponential decay.

2.2.2 Bifurcation delay and dynamical hysteresis

Bifurcation delay is a phenomenon that has been studied comprehensively in a variety of physical systems that exhibit jumps or switching behaviour. Traditionally, these jumps result from slowly changing externally a specific parameter. The main problem lies with the difficulties associated with the prediction of the instant when the jumps occur. It is possible to observe that the switch does not take place at the original bifurcation point but is delayed. This particular delay has motivated the in-depth research of bifurcation problems with control parameters depending on time. The delay depends on the analytical structure of the specific problem in the vicinity of the critical transition point (245). Several experimental studies of transitions through bifurcations arising from time-dependent parameters have had a fundamental part in understanding, for instance, Benard convection (19, 89), oscillatory chemical reactions (246), bistable chemical reactions (244) and lasers (260). Let us refer back to the supercritical Pitchfork normal form represented in Equation 18. Through numerical investigations one observes a bifurcation delay when the critical parameter λ is swept linearly (Equation 20) with a significant speed γ_λ . The bifurcation point is no longer $\lambda_c=0$ but an incremented value λ_c^* (see Figure 23 A). The sample-paths depart the neighbourhood of the unstable state branch after a delay which is proportional to the sweeping speed γ_λ (Figure 23 A). On the other hand, the effect of the critical parameter driving speed is reduced by the presence of the external asymmetry g (see Figure 23 A).

Equation 20

$$\lambda(t) = -1 + \gamma_\lambda t$$

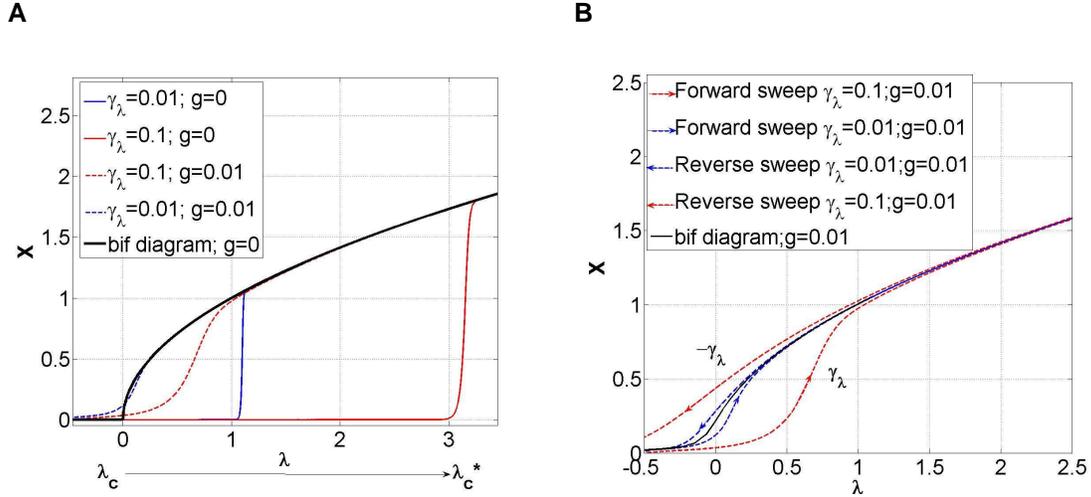


Figure 23 Numerical confirmation of bifurcation delay and dynamical hysteresis for a supercritical Pitchfork bifurcation. A) Bifurcation delay with different critical parameter sweeping speeds γ_λ (Equation 20), with and without constant asymmetry g . B) Dynamical hysteresis observed by sweeping the critical parameter through the bifurcation point in the forward and reverse directions, in the presence of g . All sample-paths, with exception of the original bifurcation diagrams (black lines), were determined numerically by a simple Heun method (section 8.4) applied to Equation 18. Increasing the parameter sweeping speed induces bigger bifurcation delays. Increasing the asymmetry g decreases bifurcation delays.

Additionally, by substituting in Equation 18 the critical parameter sweep speed γ_λ by $-\gamma_\lambda$ we arrive at a dynamical hysteresis phenomenon whereby the bifurcation point is no longer λ_c^* but $-\lambda_c^*$ (107). For an analytical treatment, through asymptotic theory, of bifurcation delay phenomena see the work of Erneux and coworkers (245).

2.2.3 Bistable potential with time-dependent critical parameter and external asymmetry

Assuming that not only the critical parameter λ is time-dependent but also that the external asymmetry g changes according to some rate law, interesting effects may take place that could be extended and applied to the problem of cell fate decision. In section 2.3 we will explore the effects of a combination of two external signals on a paradigmatic circuit known as genetic decision switch. The simulations performed here and in further sections will help to clarify the results of section 2.3 .

A 2D bifurcation diagram, (λ, g) , can be viewed in Figure 24 A. Varying both g and λ with a specific rate law corresponds to changing, accordingly, the path in the diagram represented in Figure 24 A. Entry points into the bistability area (II) equates with a specific distance between the emerging branches (see Figure 24 B and Equation 19).

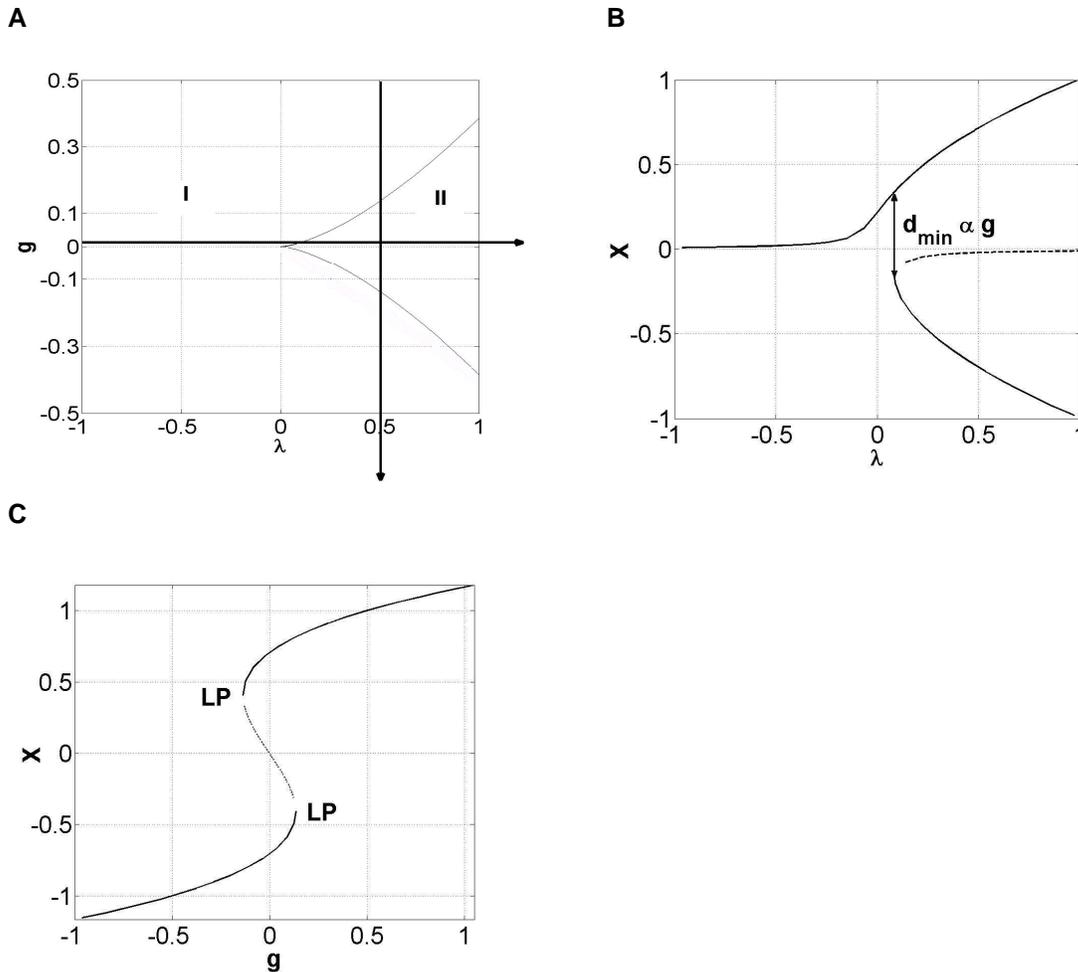


Figure 24 Bifurcation curves for a one dimensional system. A) 2D bifurcation diagram obtained by sweeping both g and λ . The hysteresis borders separate areas in the parameter space where only one asymptotical stable state exists, region I, or bistability arises, region II. B) $g=0.01$. C) $\lambda=0.5$. LP-Limit Point. This diagram shows a saddle-node bifurcation.

2.2.3.1 Trajectories in the deterministic system: effects of a transient asymmetry $g_0(t)$ with maximum at the critical point

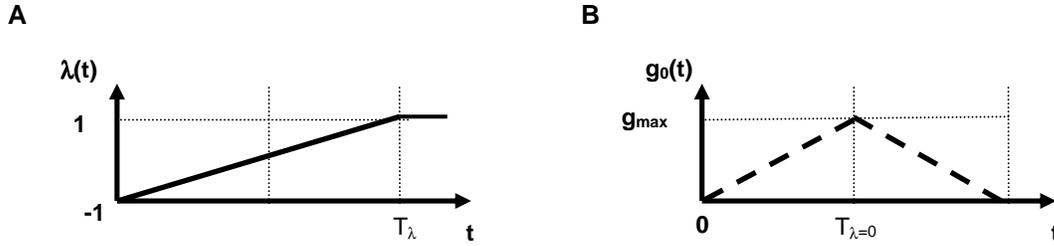


Figure 25 Critical parameter and asymmetry shapes in time. **A)** Profile for $\lambda(t)$ (Equation 20). **B)** Profile for $g_0(t)$ (see also Equation 21).

Determining numerically the sample-paths of a system ruled by Equation 18, the normal form for a supercritical pitchfork bifurcation, we can show the effects of varying both the asymmetry g and the critical parameter λ . The system is initially started at a point very close to $X=0$, which for $-1 < \lambda < 0$ is the only solution of Equation 18. The critical parameter is subsequently varied according to a linear law from -1 to 1 (see Figure 25 A). The asymmetry, on the other hand, follows a piecewise linear law (see Figure 25 B) with the maximum value being reached at the precise instance $T_{\lambda=0}$ when the critical parameter λ reaches 0 . This is the original bifurcation point for $g=0$ (see Figure 21 C). We impose, therefore, a dependence on γ_λ (see Equation 20 and Equation 21).

Equation 21

$$g_0(t) = \begin{cases} g_{\max} \left(1 - \left| \frac{t - T_{\lambda=0}}{T_{\lambda=0}} \right| \right), & 0 \leq t \leq T_\lambda \\ 0 & , t \geq T_\lambda \end{cases}$$

with

$$T_\lambda = \frac{2}{\gamma_\lambda} \quad \text{and} \quad T_{\lambda=0} = \frac{T_\lambda}{2}$$

Throughout this chapter we will identify this particular case as $g_0(t)$. In further sections additional cases will be tested where the maximum asymmetry, g_{\max} , is reached at an instant $t < T_{\lambda=0}$ before the critical parameter λ reaches the

critical point ($\lambda=0$) or at an instant $t_+ > T_{\lambda=0}$ after (see Figure 27 ahead). We will, throughout the thesis, refer to the asymmetry function corresponding to the former case as $g_-(t)$, and the latter as $g_+(t)$.

We chose to test these profiles precisely because they simulate features of a paradigmatic genetic decision switch which will be studied in section 2.3. In the genetic switch the differences between 2 external signals (see Figure 19 and Figure 39), to which we refer to as external asymmetry, perform the same function as g in the canonical model represented in Equation 18. The fact that the signals also change with time and induce a supercritical pitchfork bifurcation is similar to the effect of sweeping the critical parameter λ (Equation 18 and Equation 20). Another important aspect is the dependence of the profile of g on the critical parameter sweeping speed γ_λ . This particular condition is an attempt to establish a parallel between the supercritical pitchfork bifurcation normal form and the effect of the external signals on the genetic switch: they create both the external asymmetry and drive the system through the critical region. Also, the asymmetry g in the canonical model is transient due to the fact that the asymmetry between external signals in the paradigmatic genetic switch is also transient (Figure 39 C). We will have, under this approach, the possibility to establish a parallel between a model motivated by biology and another traditional in statistical physics. The latter allows analytical treatment (2.2.4) while the former, due to non-linearities characteristic of epigenetic regulation, doesn't.

Combinations of independent profiles for the sweeping parameter λ and the external asymmetry have been published before, specifically with the bifurcation parameter being swept linearly and g with an oscillatory and chaotic time-series (199).

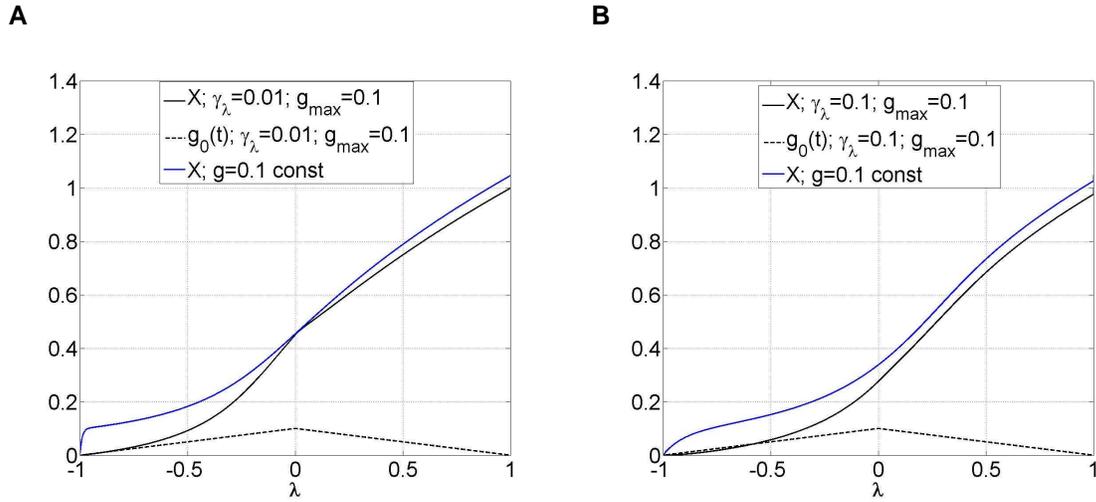


Figure 26 Effects of a time-dependent asymmetry on the trajectories of X. A) X concentration profile for sweeping speed $\gamma_\lambda = 0.01$ and time-dependent asymmetry $g_0(t)$ (see Figure 25 B) with $g_{\max} = 0.1$. Also shown is the trajectory for constant asymmetry $g = 0.1$. B) X concentration profile for sweeping speed $\gamma_\lambda = 0.1$ and time-dependent asymmetry $g_0(t)$ (see Figure 25 B) with $g_{\max} = 0.1$. Also shown is the trajectory for constant asymmetry $g = 0.1$.

Observing the sample-paths obtained by numerical integration (through a simple Heun method, see section 8.4) we notice that switching from $X=0$ occurs at later values of λ if the speed of parameter driving is increased. This had already been seen in a previous section (2.2.2). The switching point is further increased if the asymmetry is swept with a law proportional to the critical parameter sweeping speed (Figure 26 A and B, observe differences in profiles of blue and black lines). Together with higher bifurcation delays, the potential minima of the cases where the asymmetry g is time-dependent are dislocated to lower values of x (observe differences between blue and black lines at $\lambda = 1$, Figure 26). This stems from the fact that the asymmetry returns to zero.

2.2.3.2 Trajectories in the deterministic system: effects of a transient asymmetry with maximum before, $g_-(t)$, or after the critical point, $g_+(t)$

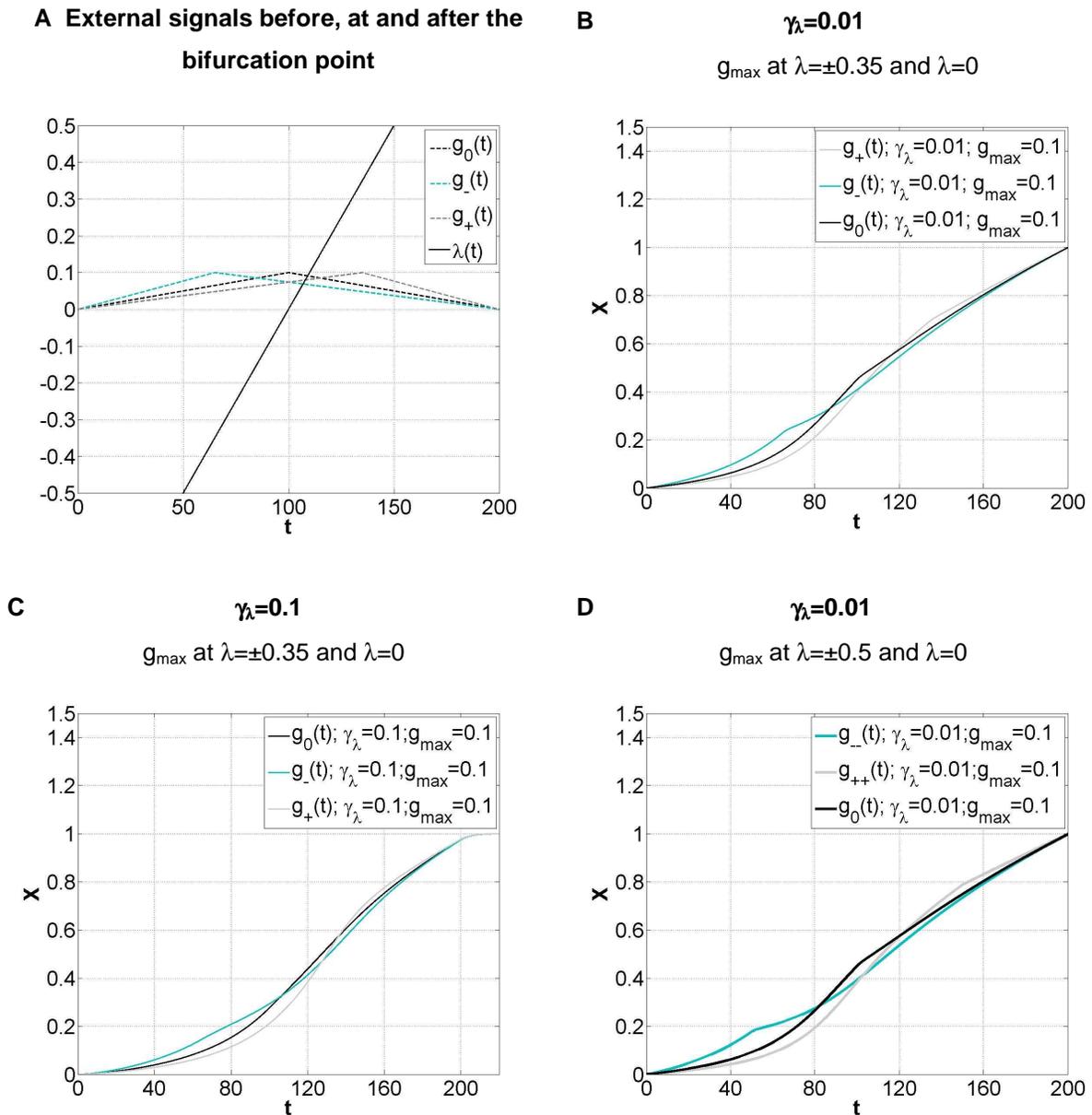


Figure 27 Effects of time dependent asymmetry with unequal ascending and descending rates. **A)** Asymmetry profiles before, $g_-(t)$, at, $g_0(t)$ and after the bifurcation point, $g_+(t)$, and $\lambda(t)$ profiles. **B)** Trajectories for the respective asymmetry profiles, with $\gamma_\lambda=0.01$, and g_{\max} at $\lambda=\pm 0.35$ and $\lambda=0$. **C)** $\gamma_\lambda=0.1$, g_{\max} at $\lambda=\pm 0.35$ and $\lambda=0$. **D)** $\gamma_\lambda=0.01$, g_{\max} at $\lambda=\pm 0.5$ and $\lambda=0$. Smaller critical parameter sweeping speeds induce bigger differences between sample-paths generated with $g_0(t)$, $g_+(t)$ and $g_-(t)$.

In Figure 27 B, we demonstrate the effect of differences in the absolute value for the sweeping rates of $g(t)$, ascending and descending segments, for a maximum amplitude $g_{\max}=0.1$ (see Figure 27 A, grey and light blue lines). The displacement of the maximum asymmetry level g_{\max} instant induces changes in the sample-paths in the interval $[t_0, t_{g(t)=0}]$. If the maximum is displaced to the left, the respective trajectory departs from $X=0$ earlier than if the displacement of the maximum is to the right. This particular behaviour stems from the non-zero distance between $X=0$ and the branch of solutions in a neighbourhood preceding the original bifurcation point $\lambda=0$ (see Figure 28 A). The distance is proportional to the value of $g(t)$ at each instant but not in a linear fashion. The branch of solutions for negative values of the critical parameter λ can be observed for several constant values of g in Figure 28. The larger the value of the asymmetry reached before $\lambda=0$ the greater the displacement from $X=0$ due to faster relaxation to the equilibrium, for the same sweeping speed γ_λ . Nevertheless, the relationship is not linear. Throughout the simulation the relative position of the trajectories of X follow roughly the relative positions of $g_{-,0,+}(t)$ in the interval $[t_0, t_{g(t)=0}]$. For large times, the trajectories saturate to the asymptotically stable solution $\sqrt{\lambda_{final}}$ due to non-linearities, and become indistinguishable.

In Figure 27 C we observe again the interplay between the magnitude of g_{\max} , the sweeping speed and slow time-scales near the critical region. Nevertheless, comparing with Figure 27 B, the differences between the trajectories induced by $g_0(t)$, $g(t)$ and $g_+(t)$ are smaller. This arises, as expected, due to the larger sweeping speeds used in Figure 27 C, which provoke larger bifurcation delays, and consequently reduce the system's sensitivity to the influence of the external asymmetry, particularly before the critical point is reached. If we shift again the position of g_{\max} we observe further differences in the sample-paths (see Figure 27 D). Yet, the effects of g_{\max} position are expected to have maximum relevance in an interval near the bifurcation point, with asymmetry profiles shifted to the left exerting stronger effects than those shifted

to the right. See for example the bifurcation diagrams plotted in Figure 28 A. The distance between each of the upper branch of solutions for two different values of g is maximal just before the original bifurcation point $\lambda=0$ (see Figure 28 B). For very large values or very small values of the critical parameter the solutions don't disagree significantly. Consequently, in studies of attractor selection in the presence of fluctuations (studied in sections 2.2.4.2 and 2.2.5), differences in time-dependent profiles for $g(t)$ are expected to generate observable dissimilarities only if g_{\max} is attained in the vicinity of the instant when the critical point is crossed. Moreover, the presence of strong fluctuations may also play a fundamental role in the capacity for differential processing of $g(t)$ profiles, as they may hinder the system's sensitivity.

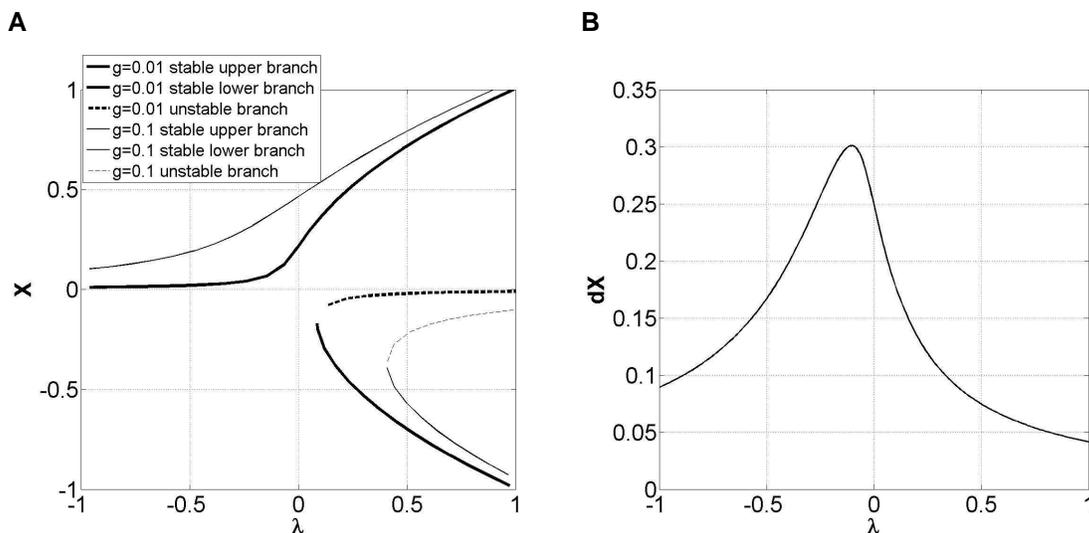


Figure 28 Comparative bifurcation diagrams for different parameters of a supercritical pitchfork bifurcation normal form. A) Bifurcation diagrams for constant $g=0.01$ and $g=0.1$. B) Distance $dX = (X_{\text{upper branch}})_{g=0.1} - (X_{\text{upper branch}})_{g=0.01}$, between the upper branch of solutions of a supercritical pitchfork bifurcation for constant $g=0.01$ and $g=0.1$ (see branches above in A).

2.2.4 Branch selectivity in the presence of additive noise: analytical results

We now augment the previous problem to account for fluctuations. If noise is included how will the selectivity of each of the attractors be affected by both the

speed of parameter λ driving and the maximum asymmetry g_{\max} ? Sample trajectories for a small asymmetry $g=0.01$ and noise intensity $\sigma=0.05$ (see Equation 23) are represented in Figure 29. Both the cases of time-dependent and constant asymmetry are plotted.

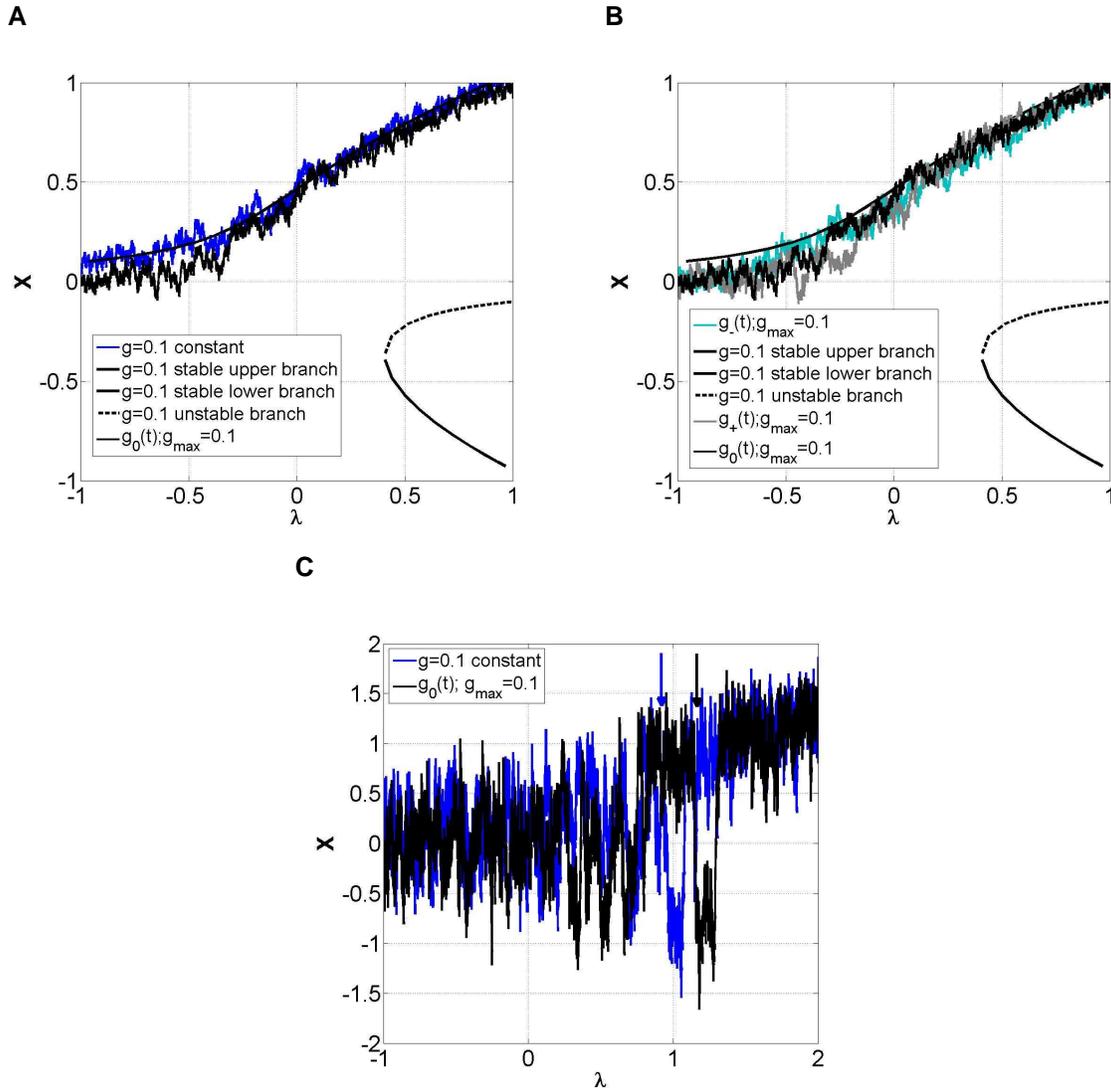


Figure 29 Examples of trajectories in the presence of noise for sweeping of the critical parameter. A) For $g_{\max}=0.1$, both constant (black line) and symmetric time-dependent (blue line), and $\gamma_\lambda=0.1$, in the presence of fluctuations with intensity $\sigma=0.05$. B) For $g_{\max}=0.1$, $g_0(t)$ (black line), $g_+(t)$ (grey line) and $g_-(t)$ (light blue line) (see also Figure 27 A) with $\sigma=0.05$. C) Profile of $x(t)$, for $g_{\max}=0.1$, both constant (black line) and time-dependent (blue line), with $\gamma_\lambda=0.1$ and $\sigma=0.5$. The noise is assumed to be Gaussian. Time-dependent asymmetry induces sample paths closer to the unstable branch even in the presence of noise. Larger noise intensities blur differences generated by time-dependent asymmetries.

It is possible to check that the trajectory with additional time-dependent asymmetry has, as was seen for the deterministic case (see Figure 26), a path closer to the unstable branch of solutions of the bifurcation diagram. This may enhance the probability of jumping across the potential barrier that coincides with the unstable branch (Figure 29 A). In Figure 29 B we can additionally see the effect of the asymmetry reaching its maximum before and after the critical point. These results are as expected from the analysis performed in a previous section (see Figure 27). For higher intensities of noise (Figure 29 C), the role of the asymmetry g as a state selector is decreased. The trajectories are capable of crossing to the lower branch even when the critical parameter has reached a considerable value with which a sufficient depth of the potential is achieved (see Figure 30). The transition dynamics time-scale (τ_{\pm}) (see Equation 22) between minima can be estimated using Kramers classical theory (94):

Equation 22

$$\tau_{\pm} \approx \pi \left(-U''(X_u)U''(X_{\pm}) \right)^{-\frac{1}{2}} e^{\frac{2\Delta U_{\pm}}{\sigma^2}}$$

with

$$\Delta U_{\pm} = U(X_u) - U(X_{\pm})$$

In Equation 22 X_u stands for the unstable solution, X_+ for the upper branch of stable solutions and X_- for the lower branch of solutions of Equation 18. U represents the potential landscape obtained from Equation 18 (see Equation 24). The Kramers formula tells us that if the difference in potential between the minima and the barrier is sufficiently high, we are capable of freezing the system in one of the states, due to the existence of large $\tau_{+/-}$. Nevertheless, if ΔU is not sufficient or if the noise intensity σ is considerable, jumps across the barrier become possible. Interpreting our problem in the light of Equation 22 we see that the action of an external asymmetry is to force one of the minima to become deeper, in our case X_+ , and hold there trajectories that have reached it. Also observe that τ_+ and τ_- are not equal in the presence of g . Therefore, even if a

trajectory jumps to X_- , there is a higher probability of it jumping back to X_+ . Yet, in our case the asymmetry is transient and differences in depth of each of the minima disappear. We have, as a result, several mechanisms to take into consideration in the attractor selection problem:

- **λ sweeping speed influences Kramers transition time.** The transient asymmetry considered in our work has a time-dependent profile that is a function of γ_λ (see Equation 21). Consequently, the time interval the asymmetry is present is proportional to γ_λ . Because the asymmetry is responsible for ΔU and subsequently differences in τ_+ and τ_- , we expect that the longer the asymmetry is present, the higher the probability of a trajectory remaining in the minima X^+ corresponding to the upper branch.
- **Slow passage through the critical region allows for faster relaxation processes in the presence of larger asymmetries.** As was previously seen (Figure 26 and Figure 27) larger asymmetries induce faster relaxation times to the equilibrium. Because the asymmetry is dependent on sweeping speed, the longer larger values are present the quicker the relaxation process and the more efficient will be the potential barrier in deterring jumps to other branches. We can conclude that, at least near the critical region, we need smaller γ_λ .
- **After the critical region faster sweeping freezes trajectories in the selected minimum.** Let us assume that for a particular noise strength σ the system has reached the upper branch. Even if we have passed the critical region there is always a probability of crossing the potential barrier. Due to the fact that ΔU depends on λ (see also Equation 24), after the critical region the best strategy would be increasing γ_λ to enhance the action of the potential barrier and reduce the probability of the system jumping to the lower branch (see Equation 22). We will, nevertheless, not consider piecewise functions for γ_λ in further work presented ahead.

Overall, the critical parameter has to be swept with a sufficiently low speed to induce maximum probability of the system migrating to the branch of solutions favoured by g . Yet, the speed has to be sufficiently high to increase the potential barrier and as a result increase the transition time to the lower branch (200). If these conditions are met we reduce the probability of a jump erasing the effect of the external signal. Before we perform a systematic numerical investigation of the role of asymmetry, sweeping rate and noise on the branch selectivity (section 2.2.5), we devise analytical expressions that will clarify the numerical experiments. For this purpose, we will make use of the Langevin equation (see Equation 23) associated with Equation 18.

Equation 23

$$\dot{X} = -X^3 + (\lambda(t) - \lambda_c)X + g(t) + \xi(t)$$

In what follows $\xi(t)$ will be assimilated to a Gaussian distributed noise, $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = \sigma^2 \delta(t - t')$. Writing Equation 23 in its variational form for further use we have:

Equation 24

$$\dot{X} = -\frac{dU(X,t)}{dX} + \xi(t)$$

with

$$U(X,t) = \frac{X^4}{4} - (\lambda(t) - \lambda_c) \frac{X^2}{2} - g(t)X$$

$U(X,t)$ is the kinetic potential. The shape of the potential for several values of g and the critical parameter λ can be visualized in Figure 30 .

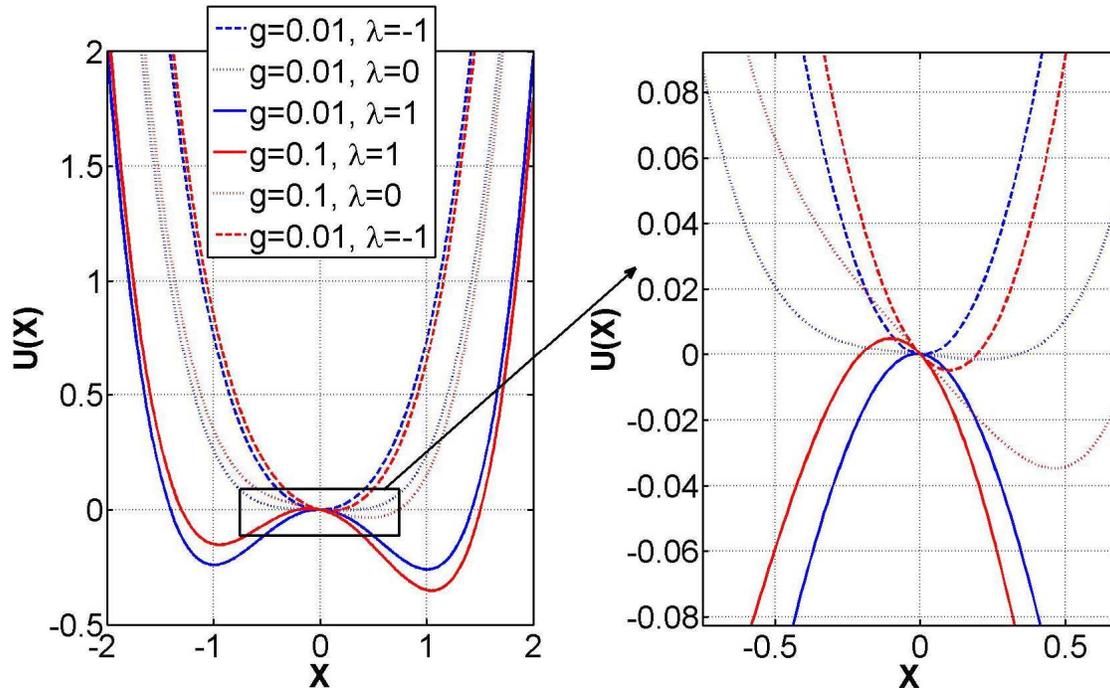


Figure 30 Potential profiles for several values of external asymmetry before, at and after the bifurcation point. Left: $U(X,t)$ profiles for several combinations of (g,λ) . Right: detail near the unstable state.

Varying the critical parameter from -1 to 1 for $g=0.01$ and $g=0.1$ exerts different action on the potential. The higher the asymmetry g , the deeper the potential at the minimum when $\lambda=0$. The dislocation of the position of the potential minima could already be seen in the bifurcation diagrams presented earlier (Figure 21 D).

2.2.4.1 Evolution of the probability density and the process of branch selection

Several assumptions underlie the applicability of the following theoretical approach. The distribution around the branch of solutions before the critical point λ_c is assumed to be Gaussian. As the bifurcation parameter is passed through the critical region the distribution starts to drift towards the branch favoured by the external field $g(t)$. The drift rate is approximately $g(t)$. In this region, as was previously highlighted (section 2.2.1), the relaxation to the equilibrium is slow. At $\lambda=0$, the position of the steady state is approximately $(g)^{1/3}$, which makes the

relaxation time $\sim(g)^{-2/3}$. For very small asymmetries, this relaxation process is extremely slow (162). For illustration purposes, a representative example with constant g is depicted in Figure 31. Concurrently with the drift process, the distribution also suffers spreading due to the fluctuations represented by the noise term in Equation 23. As was previously mentioned, around the critical region the amplitude of the fluctuations is amplified (section 2.2.1). Nevertheless, if a strong external field g is applied the fluctuations can be neglected if the critical parameter λ is slowly changed. However, this situation may be modified if we go rapidly through the critical region. Sweeping with large enough rates allocates a larger importance to the presence of fluctuations, which become significant and determine to a large extent the equilibrium state selected (107).

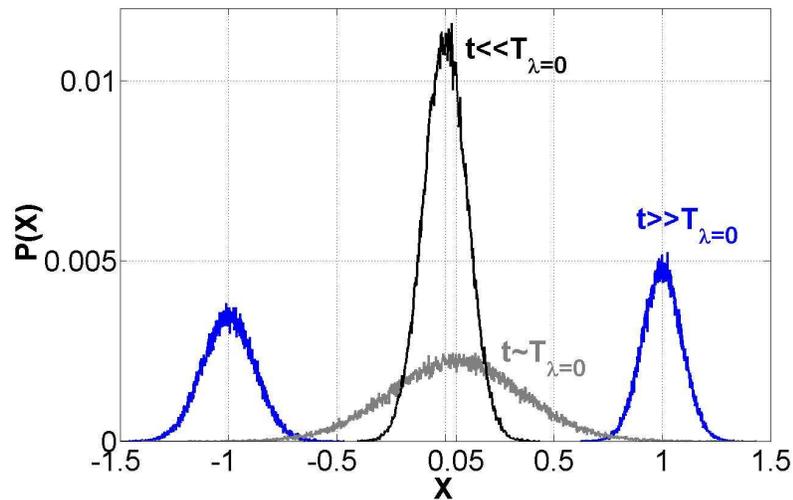


Figure 31 Probability evolution with time in a system with critical parameter sweeping and constant asymmetry $0 < g < 1$. At $t \sim T_{\lambda=0}$ the peak is displaced to the right due to the positive asymmetry $g > 0$.

After the critical point the distribution becomes bimodal with each peak centred on the respective stable state (Figure 31). The assumption of a Gaussian like distribution for the process is only valid if the speed with which the critical parameter is forced to go through the critical region is sufficiently high. If this condition is not met, the distribution relaxes to its non-Gaussian form and the analytical solutions determined in the following section are not sufficiently descriptive (161, 191). It is expected that an optimum selection process occurs if

the critical parameter is changed with a speed that allows for the drift to centre the distribution on a point that shifts most of the area under the curve at $T_{\lambda=0}$ (see Figure 31) beyond the position of the unstable branch. Unlike studies devising analytical approaches based on Kramers theory for the probability density mass transfer across the potential barrier (198, 200), we will assume, for the sake of simplicity that the attractor selection process is complete just after the bifurcation point (161).

2.2.4.2 Attractor selectivity: devising analytical expressions

The probability density function $P(X,t)$ can be described by a Fokker-Planck equation derived for the Langevin equation of the bistable potential represented in Equation 18 (94). It takes the following form (Equation 25):

Equation 25

$$\begin{aligned} \frac{\partial P(X,t)}{\partial t} &= -\frac{\partial}{\partial X} \left[-\frac{\partial U}{\partial X} \times P(X,t) \right] + \frac{\sigma^2}{2} \frac{\partial^2}{\partial X^2} P(X,t) \\ &= -\frac{\partial}{\partial x} \left[(\lambda(t) \times X - X^3 + g(t)) \times P(X,t) \right] + \frac{\sigma^2}{2} \frac{\partial^2}{\partial X^2} P(X,t) \end{aligned}$$

The analysis performed in this section will follow roughly the method devised by Kondepudi and coworkers (161). Nevertheless, our case study involves a time dependent asymmetry $g(t)$, which was not considered before. Here we do not take into account any fluctuations in the λ parameter. Their contribution is thought to be negligible for the calculations to follow (161). The evolution of the probability density function $P(X,t)$ can be calculated by finding how each of the moments changes during the sweeping process.

Multiplying Equation 25 by X^n , neglecting the cubic term (near the bifurcation point, where the selection process occurs, $X \ll 1$) and integrating one obtains:

Equation 26

$$\frac{\partial \langle X^n \rangle}{\partial t} = - \int_0^{X'} X^n \frac{\partial}{\partial X} [(\lambda(t) \times X + g(t)) \times P(X, t)] dX \\ + \int_0^{X'} X^n \frac{\sigma^2}{2} \frac{\partial^2}{\partial X^2} P(X, t) dX$$

Both terms on the right hand side of Equation 26 can be integrated by parts. The probability density function and its derivatives are assumed to decay exponentially with X . Hence, an equation for each of the moments of the probability density function can be easily devised and takes the following form:

Equation 27

$$\frac{dm_n}{dt} = n((\lambda(t) - \lambda_c)m_n(t) + g(t)m_{n-1}(t)) + \frac{\sigma^2}{2} n(n-1)m_{n-2}(t)$$

for $n \geq 2$

Assuming the first two moments are sufficiently descriptive of the evolution of the probability density function $P(X,t)$ near the bifurcation point, we choose to restrict our analysis to Equation 28 (mean) and Equation 29 (variance).

Equation 28

$$\frac{dm_1}{dt} = ((\lambda(t) - \lambda_c)m_1(t) + g(t))$$

Equation 29

$$\frac{dm_2}{dt} = 2((\lambda(t) - \lambda_c)m_2(t) + g(t)m_1(t)) + \sigma^2$$

For a maximum asymmetry value $g_{\max} \ll 1$, the contribution of $g(t)$ near the bifurcation point disappears from the equation for the second moment (Equation 29).

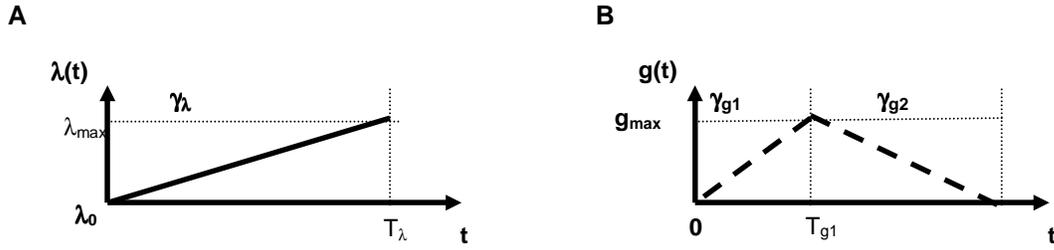


Figure 32 Time-dependent profiles for variables λ and g . A) Critical parameter sweeping function profile. B) Asymmetry sweeping function profile.

For our particular problem, we are interested in the effects of a monotonous linear function for the bifurcation parameter λ (see Equation 30 and Figure 32 A) and for the asymmetry $g(t)$ a piecewise linear function (see Equation 31 and Figure 32B).

Equation 30

$$\lambda(t) = \begin{cases} \lambda_0 + t\gamma_\lambda, & 0 \leq t \leq T_\lambda \\ \lambda_0 + \gamma_\lambda T_\lambda, & t \geq T_\lambda \end{cases}$$

Equation 31

$$g(t) = \begin{cases} t\gamma_{g1}, & 0 \leq t \leq T_{g1} \\ g_{\max} - (t - T_{g1})\gamma_{g2}, & T_{g1} \leq t \leq T_{g2} \\ 0, & t \geq T_{g2} \end{cases}$$

Equation 28 and Equation 29 are linear ordinary differential equations and have an analytical solution given by Equation 32 and Equation 33, respectively.

Equation 32

$$\langle X \rangle (t) = \frac{1}{\mu_{\langle X \rangle}(t)} \left(\mu_{\langle X \rangle}(0) \langle X \rangle (0) + \int_0^t \mu_{\langle X \rangle}(s) g(s) ds \right)$$

with

$$\mu_{\langle X \rangle}(t) = e^{\int -(\lambda(t) - \lambda_c) dt} = e^{-\left(\frac{\gamma_\lambda t^2}{2} - \Delta\lambda t\right)} = e^{-\frac{\gamma_\lambda}{2} \left(\left(t - \frac{\Delta\lambda}{\gamma_\lambda} \right)^2 - \left(\frac{\Delta\lambda}{\gamma_\lambda} \right)^2 \right)}$$

Equation 33

$$\langle (\delta X)^2 \rangle (t) = \frac{1}{\mu_{\langle (\delta X)^2 \rangle}(t)} \left(\begin{array}{l} \mu_{\langle (\delta X)^2 \rangle}(0) \langle (\delta X)^2 \rangle (0) \\ + \\ \sigma^2 \int_0^t \mu_{\langle (\delta X)^2 \rangle}(s) ds \end{array} \right)$$

with

$$\mu_{\langle (\delta X)^2 \rangle}(t) = e^{2 \int -(\lambda(t) - \lambda_c) dt} = e^{-(\gamma_\lambda t^2 - 2\Delta\lambda t)} = e^{-\gamma_\lambda \left(\left(t - \frac{\Delta\lambda}{\gamma_\lambda} \right)^2 - \left(\frac{\Delta\lambda}{\gamma_\lambda} \right)^2 \right)}$$

$$\delta X^2 = (X - \langle X \rangle)^2$$

Extending the integrals of the previous equations we have the expressions represented in Equation 34 and Equation 35.

Equation 34

$$\langle X \rangle (t) = e^{\left(\frac{\gamma_\lambda t^2}{2} - \Delta\lambda t\right)} \langle X \rangle (0) + e^{\frac{(\gamma_\lambda t - \Delta\lambda)^2}{2\gamma_\lambda}} \left(\begin{array}{l} \int_0^{T_{g1}} e^{\frac{-\gamma_\lambda (s - \gamma_\lambda \Delta\lambda)^2}{2}} (g_0 + \gamma_{g1} s) ds \\ + \int_{T_{g1}}^t e^{\frac{-\gamma_\lambda (s - \gamma_\lambda \Delta\lambda)^2}{2}} (g_{\max} - \gamma_{g2} (s - T_{g1})) ds \end{array} \right)$$

Equation 35

$$\begin{aligned} \langle (\delta X)^2 \rangle (t) = & e^{(\gamma_\lambda t^2 - 2 \Delta \lambda t)} \langle (\delta X)^2 \rangle (0) \\ & + e^{\frac{(\gamma_\lambda t - \Delta \lambda)^2}{\gamma_\lambda}} \left(\sigma^2 \int_0^t e^{-\gamma_\lambda \left(s - \frac{\Delta \lambda}{\gamma_\lambda} \right)^2} ds \right) \end{aligned}$$

Through an appropriate change of variables, $t' = \sqrt{\gamma_\lambda} \left(s - \frac{\Delta \lambda}{\gamma_\lambda} \right)$, with $\Delta \lambda = \lambda_c - \lambda_0$, Equation 34 and Equation 35 are reduced straightforwardly to Equation 36 and Equation 37. The initial conditions term for each of the expressions is assumed to be negligible in comparison to the final instance where separation of the initial monomodal distribution is completed, for small enough γ_λ rates.

Equation 36

$$\langle X \rangle (t) \approx \frac{e^{\frac{(\gamma_\lambda t - \Delta \lambda)^2}{2\gamma_\lambda}}}{\sqrt{\gamma_\lambda}} \left(\int_{\frac{-\Delta \lambda}{\sqrt{\gamma_\lambda}}}^{\frac{(T_{g1} \gamma_\lambda - \Delta \lambda)}{\sqrt{\gamma_\lambda}}} e^{\frac{-t'^2}{2}} \left(\gamma_{g1} \left(\frac{t'}{\sqrt{\gamma_\lambda}} + \frac{\Delta \lambda}{\gamma_\lambda} \right) \right) dt' + \int_{\frac{(T_{g1} \gamma_\lambda - \Delta \lambda)}{\sqrt{\gamma_\lambda}}}^{\frac{(t \gamma_\lambda - \Delta \lambda)}{\sqrt{\gamma_\lambda}}} e^{\frac{-t'^2}{2}} \left(g_{\max} - \gamma_{g2} \left(\frac{t'}{\sqrt{\gamma_\lambda}} + \frac{\Delta \lambda}{\gamma_\lambda} - T_{g1} \right) \right) dt' \right)$$

Equation 37

$$\langle (\delta X)^2 \rangle (t) \approx \frac{e^{\frac{(\gamma_\lambda t - \Delta \lambda)^2}{\gamma_\lambda}}}{\sqrt{\gamma_\lambda}} \left(\sigma^2 \int_{\frac{-\Delta \lambda}{\sqrt{\gamma_\lambda}}}^{\frac{(t \gamma_\lambda - \Delta \lambda)}{\sqrt{\gamma_\lambda}}} e^{-t'^2} dt' \right)$$

Let us study the contribution of $g(t)$ to the mean until an instant t' beyond T_{g1} , with $t' \gamma_\lambda = (1 + K) \Delta \lambda$ and $K < 1$. This change of variables allows us to eliminate some of the complexity in the expressions devised above. K is simply a measure of

how far we are from the bifurcation point, after we have crossed it in the sweeping process. Integrating the previous expressions we obtain Equation 38 and Equation 39.

Equation 38

$$\langle X \rangle (t) \approx \frac{(K \Delta \lambda)^2}{2\gamma_\lambda^{\frac{3}{2}}} [T_1 + T_2]$$

with

$$T_1 = \left(2 \left(\gamma_{g1} e^{\frac{-(\Delta \lambda)^2}{2\gamma_\lambda}} + \gamma_{g2} e^{\frac{-(K \Delta \lambda)^2}{2\gamma_\lambda}} - (\gamma_{g1} + \gamma_{g2}) e^{\frac{-(-\gamma_\lambda T_{g1} + \Delta \lambda)^2}{2\gamma_\lambda}} \right) \sqrt{\gamma_\lambda} \right)$$

and

$$T_2 = \sqrt{2\pi} \left(\begin{array}{l} \Delta \lambda \gamma_{g1} \operatorname{erf}\left(\frac{\Delta \lambda}{\sqrt{2\gamma_\lambda}}\right) + (-\Delta \lambda \gamma_{g2} + \gamma_\lambda (g_{\max} + \gamma_{g2} T_{g1})) \operatorname{erf}\left(\frac{K \Delta \lambda t}{\sqrt{2\gamma_\lambda}}\right) \\ - \\ (\Delta \lambda (\gamma_{g2} + \gamma_{g1}) - \gamma_\lambda (g_{\max} + \gamma_{g2} T_{g1})) \operatorname{erf}\left(\frac{\Delta \lambda - \gamma_\lambda T_{g1}}{\sqrt{2\gamma_\lambda}}\right) \end{array} \right)$$

Equation 39

$$\langle (\delta X)^2 \rangle (t) \approx \sigma^2 \frac{e^{\frac{(K \Delta \lambda)^2}{\gamma_\lambda}}}{2} \left(\operatorname{erf}\left(\frac{\Delta \lambda}{\sqrt{\gamma_\lambda}}\right) + \operatorname{erf}\left(\frac{K \Delta \lambda}{\sqrt{\gamma_\lambda}}\right) \right) \sqrt{\frac{\pi}{\gamma_\lambda}}$$

The error function term present in Equation 38 and Equation 39 is given by

$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-z'^2} dz'$ and can be visualized in Figure 33.

Let us analyse Equation 39 first: since the term $\left(\operatorname{erf}\left(\frac{\Delta\lambda}{\sqrt{\gamma_\lambda}}\right) + \operatorname{erf}\left(\frac{K\Delta\lambda}{\sqrt{\gamma_\lambda}}\right) \right) \leq 2$ the dispersion induced by a time-dependent asymmetry will be always smaller or equal to the original case (161) (Equation 41). Also, let us put $\gamma_{g1} = c_1\gamma_\lambda$ and $\gamma_{g2} = c_2\gamma_\lambda$. With these simplifications we arrive at:

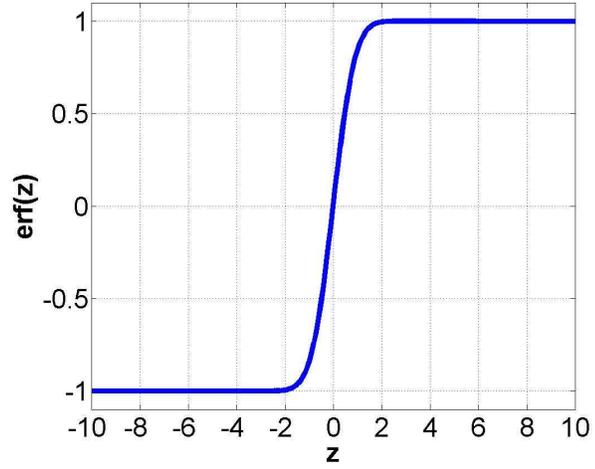


Figure 33 Profile for error function, erf(z).

Equation 40

$$\langle X \rangle (t') \approx e^{\frac{(K\Delta\lambda)^2}{2\gamma_\lambda}} [T']$$

with

$$T' = T_1' + T_2';$$

$$T_1' = \left(c_1 e^{\frac{-(\Delta\lambda)^2}{2\gamma_\lambda}} + c_2 e^{\frac{-(K\Delta\lambda)^2}{2\gamma_\lambda}} - (c_1 + c_2) e^{\frac{-(\Delta\lambda - (g_{\max}/c_1))^2}{2\gamma_\lambda}} \right)$$

and

$$T_2' = \frac{1}{2} \left(c_1^2 \Delta\lambda \operatorname{erf}\left(\frac{\Delta\lambda}{\sqrt{2\gamma_\lambda}}\right) - (c_1 + c_2)(c_1 \Delta\lambda - g_{\max}) \operatorname{erf}\left(\frac{K\Delta\lambda}{\sqrt{2\gamma_\lambda}}\right) + (-c_2 c_1 \Delta\lambda + (c_1 + c_2) g_{\max}) \operatorname{erf}\left(\frac{\Delta\lambda - (g_{\max}/c_1)}{\sqrt{2\gamma_\lambda}}\right) \right) \sqrt{\frac{2\pi}{\gamma_\lambda}} / c_1$$

Equation 41

$$\langle (\delta X)^2 \rangle (t') \leq \sigma^2 e^{\frac{(K\Delta\lambda)^2}{\gamma_\lambda}} \sqrt{\frac{\pi}{\gamma_\lambda}}$$

The differences arise in the mean. Depending on the proportion of the rates γ_λ , γ_{g1} and γ_{g2} we will have different cases:

- **Case1- g₀(t)**

$$T_{g2} = T_{\lambda} \quad \text{and} \quad T_{g1} = \frac{T_{\lambda}}{2} \Rightarrow c_1 = c_2 = g_{\max} \quad \text{with} \quad \Delta\lambda \sim 1 \quad (\text{see also Figure 32})$$

This set of constraints renders the contribution of time-dependent g(t) always smaller or equal to the original case with constant asymmetry (see Equation 42) (161).

Equation 42

$$T' = g_{\max} \left(-2e^{\frac{-(-1+\Delta\lambda)^2}{2\gamma_{\lambda}}} + e^{\frac{-(\Delta\lambda)^2}{2\gamma_{\lambda}}} + e^{\frac{(K\Delta\lambda)^2}{2\gamma_{\lambda}}} \right) +$$

$$\frac{g_{\max}}{2} \left(-2(-1+\Delta\lambda) \operatorname{erf}\left(\frac{-1+\Delta\lambda}{\sqrt{2\gamma_{\lambda}}}\right) + \Delta\lambda \operatorname{erf}\left(\frac{\Delta\lambda}{\sqrt{2\gamma_{\lambda}}}\right) - (-2+\Delta\lambda) \operatorname{erf}\left(\frac{K\Delta\lambda}{\sqrt{2\gamma_{\lambda}}}\right) \right) \sqrt{\frac{2\pi}{\gamma_{\lambda}}}$$

$$\leq g_{\max} \sqrt{\frac{2\pi}{\gamma_{\lambda}}}$$

This particular situation could also be seen in the deterministic simulations performed above (Figure 26).

- **Case2- g₁(t)**

$$T_{g1} = \frac{T_{\lambda}}{4} \quad \text{and} \quad T_{g2} = T_{\lambda} \Rightarrow c_1 = 2g_{\max} \quad \text{and} \quad c_2 = \frac{2}{3}g_{\max} \quad \text{with} \quad \Delta\lambda \sim 1. \text{ This}$$

combination of parameters also centres the distribution on lower values than the original case (161). Additionally, depending on K, one can observe that the path induce slightly above below a threshold or slightly below Case1.

Equation 43

$$T' = \frac{2g_{\max}}{3} \left(-4e^{\frac{-(1-2\Delta\lambda)^2}{8\gamma_\lambda}} + 3e^{\frac{-(\Delta\lambda)^2}{2\gamma_\lambda}} + e^{\frac{-(K\Delta\lambda)^2}{2\gamma_\lambda}} \right) +$$

$$\frac{g_{\max}}{3} \left((2 - 4\Delta\lambda) \operatorname{erf} \left(\frac{-1+2\Delta\lambda}{2\sqrt{2\gamma_\lambda}} \right) + 3\Delta\lambda \operatorname{erf} \left(\frac{\Delta\lambda}{\sqrt{2\gamma_\lambda}} \right) - (-2 + \Delta\lambda) \operatorname{erf} \left(\frac{K\Delta\lambda}{\sqrt{2\gamma_\lambda}} \right) \right) \sqrt{\frac{2\pi}{\gamma_\lambda}}$$

$$\leq g_{\max} \sqrt{\frac{2\pi}{\gamma_\lambda}}$$

- **Case3**

g₊(t)

$$T_{g1} = \frac{3T_\lambda}{4} \quad \text{and} \quad T_{g2} = T_\lambda \quad \Rightarrow \quad c_1 = \frac{2}{3} g_{\max} \quad \text{and} \quad c_2 = 2g_{\max} \quad \text{with} \quad \Delta\lambda \sim 1$$

Finally, when the maximum g_{\max} is shifted to the right of $T_\lambda/2$ the distribution is centred on lower values than the original case (161), but higher values than Case1 previously analysed.

Equation 44

$$T' = \frac{2g_{\max}}{3} \left(-4e^{\frac{-(3-2\Delta\lambda)^2}{8\gamma_\lambda}} + e^{\frac{-(\Delta\lambda)^2}{2\gamma_\lambda}} + 3e^{\frac{-(K\Delta\lambda)^2}{2\gamma_\lambda}} \right) +$$

$$\frac{g_{\max}}{3} \left((6 - 4\Delta\lambda) \operatorname{erf} \left(\frac{-3+2\Delta\lambda}{2\sqrt{2\gamma_\lambda}} \right) + \Delta\lambda \operatorname{erf} \left(\frac{\Delta\lambda}{\sqrt{2\gamma_\lambda}} \right) - 3(-2 + \Delta\lambda) \operatorname{erf} \left(\frac{K\Delta\lambda}{\sqrt{2\gamma_\lambda}} \right) \right) \sqrt{\frac{2\pi}{\gamma_\lambda}}$$

$$\leq g_{\max} \sqrt{\frac{2\pi}{\gamma_\lambda}}$$

The values calculated for the first moment for each of the cases exposed before can be seen in Figure 34. The values for the second moment will be used ahead. As was previously determined by numerical simulations (see Figure 27) the order with which the paths are positioned follows the profiles for the respective asymmetry functions. By choosing K we are effectively assuming that the selection process is completed when the paths are in a certain order.

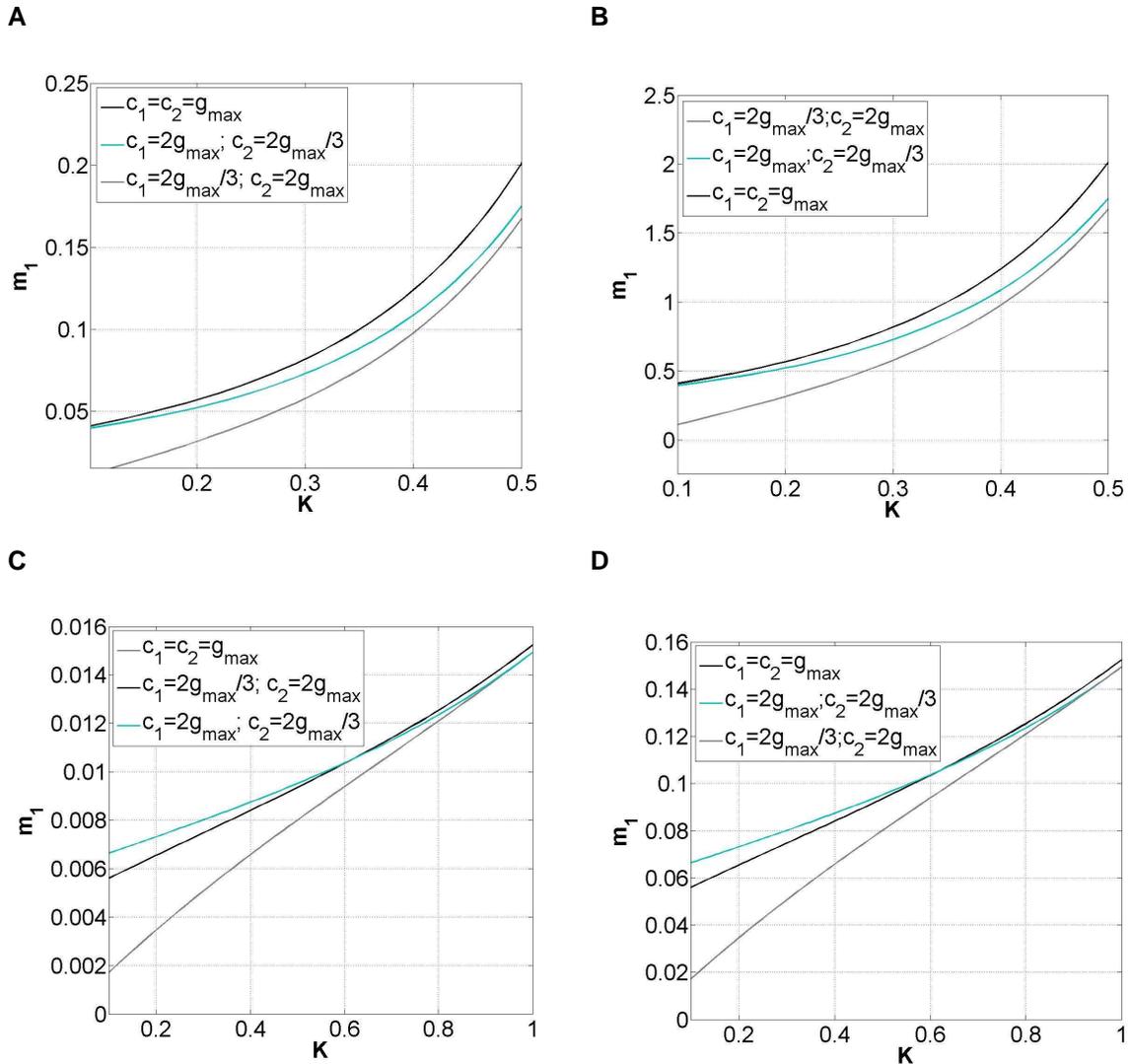


Figure 34 Dependence of the distribution mean on the instant the maximum asymmetry is reached: before, at, and after the critical bifurcation point $\lambda=0$. A) $g_{\max}=0.01$ and $\gamma_{\lambda}=0.1$. B) $g_{\max}=0.1$ and $\gamma_{\lambda}=0.1$. C) $g_{\max}=0.01$ and $\gamma_{\lambda}=1$. D) $g_{\max}=0.1$ and $\gamma_{\lambda}=1$. K is simply a measure of how far we are from the bifurcation point, after we have crossed it in the sweeping process

Nevertheless, the analytical expressions do not approximate well the steady states for large values of K . This stems from the fact that we eliminated the cubic term from the original normal form equation (see Equation 23). For larger values of K there's an explosion for the first moment. We expect that for calculations of selectivity of attractors small values of K are necessary to generate significantly accurate predictions.

Evaluating the selection probability resulting from this sweeping process is, in fact, a way of measuring the memory capacity of the system to transient signals $g(t)$. For $g_{\max} > 0$ the probability of reaching the favoured state is given by Equation 45. Since the original assumption was that the distribution could be approximated by a Gaussian, we use the expressions for the mean and variance calculated before.

Equation 45

$$P_{up} = \frac{1}{\sqrt{2\pi \langle (\delta X)^2 \rangle}} \int_{\frac{-g_{\max}}{(\lambda_{Final} - \lambda_c)}}^{+\infty} e^{\frac{-(X - \langle X \rangle)^2}{2\langle (\delta X)^2 \rangle}} dX$$

The lower limit for the integral in Equation 45 is the estimated position of the asymptotically unstable state, far above or below the critical point, when the asymmetry is constant and equal to g_{\max} . In the problem where the asymmetry is constant, far below or above the critical point, the position of the unstable branch is exactly $X=0$. However, in the vicinity of the critical point the maximum

estimated position for the unstable state is $\frac{-g_{\max}}{(\lambda_{Final} - \lambda_c)}$. Hence, by assuming this as the correct value for the unstable branch we are over-estimating the value of P_{up} with respect to the original problem (constant g), but underestimating with respect to the real value. Through an appropriate change of variables,

$X' = \frac{X - \langle X \rangle}{\sqrt{\langle (\delta X)^2 \rangle}}$, using the integration by substitution ruleⁿ, and the fact that the

Gaussian is symmetric with regards to its mean^o, Equation 45 takes the following simplified form:

$$^n \int_a^b f(g(s))g'(s) = \int_{g(a)}^{g(b)} f(x)dx$$

$$^o \int_{-a}^{+\infty} N(0,1) = \int_{-\infty}^a N(0,1)$$

Equation 46

$$P_{up} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^N e^{-\frac{X'^2}{2}} dX'$$

In Equation 46 N stands for the number of standard deviations that the peak of the distribution is displaced from the unstable branch of solutions for a particular value of λ :

Equation 47

$$N = \frac{\langle X \rangle (t') + \frac{g_{\max}}{(\lambda_F - \lambda_c)}}{\sqrt{\langle (\delta X)^2 \rangle (t')}}}$$

with

$$t' \gamma_\lambda = (1 + K)\Delta\lambda$$

The expression for the selectivities of each of the cases analysed before are presented in section 8.5. Lets us present the expressions for the case $\gamma_{g1} = \gamma_{g2} = c_1 \gamma_\lambda = g_{\max} \gamma_\lambda$. N is given by Equation 48.

Equation 48

$$N = \frac{T' + e^{-\frac{(K\Delta\lambda)^2}{2\gamma_\lambda}} \frac{g_{\max}}{(\lambda_F - \lambda_c)}}{\left(\frac{\sigma^2}{2}\right)^{1/2} \left(\operatorname{erf}\left(\frac{\Delta\lambda}{\sqrt{\gamma_\lambda}}\right) + \operatorname{erf}\left(\frac{K\Delta\lambda}{\sqrt{\gamma_\lambda}}\right)\right)^{1/2} \left(\frac{\pi}{\gamma_\lambda}\right)^{1/4}}$$

with

$$T' = g_{\max} \left(-2e^{-\frac{(-1+\Delta\lambda)^2}{2\gamma_\lambda}} + e^{-\frac{(\Delta\lambda)^2}{2\gamma_\lambda}} + e^{-\frac{(K\Delta\lambda)^2}{2\gamma_\lambda}} \right) +$$

$$\frac{g_{\max}}{2} \left(-2(-1+\Delta\lambda)\operatorname{erf}\left(\frac{-1+\Delta\lambda}{\sqrt{2\gamma_\lambda}}\right) + \Delta\lambda\operatorname{erf}\left(\frac{\Delta\lambda}{\sqrt{2\gamma_\lambda}}\right) - (-2+\Delta\lambda)\operatorname{erf}\left(\frac{K\Delta\lambda}{\sqrt{2\gamma_\lambda}}\right) \right) \sqrt{\frac{2\pi}{\gamma_\lambda}}$$

Since $\Delta\lambda=1$ and $(K\Delta\lambda)^2/2\gamma_\lambda \ll 1$, the second term in the numerator of previous equation can be neglected for small enough sweeping rates. Consequently, N

can be approximated by Equation 49 which is always smaller than the value obtained with constant asymmetry equal to g_{\max} (161).

Equation 49

$$N = \frac{g_{\max} \left(-2 + e^{\frac{-1}{2\gamma_\lambda}} + e^{\frac{-K^2}{2\gamma_\lambda}} \right) + \frac{g_{\max}}{2} \left(\operatorname{erf} \left(\frac{1}{\sqrt{2\gamma_\lambda}} \right) + \operatorname{erf} \left(\frac{K}{\sqrt{2\gamma_\lambda}} \right) \right) \sqrt{\frac{2\pi}{\gamma_\lambda}}}{\left(\frac{\sigma^2}{2} \right)^{1/2} \left(\operatorname{erf} \left(\frac{1}{\sqrt{\gamma_\lambda}} \right) + \operatorname{erf} \left(\frac{K}{\sqrt{\gamma_\lambda}} \right) \right)^{1/2} \left(\frac{\pi}{\gamma_\lambda} \right)^{1/4}}$$

$$\leq \frac{g_{\max}}{\left(\frac{\sigma^2}{2} \right)^{1/2}} \left(\frac{\pi}{\gamma_\lambda} \right)^{1/4}$$

The expression for the probability of attractor selection can thus be computed. If we choose sweeping rates smaller than 1, the error function is approximately 1. The calculation is therefore substantially simplified (see Equation 50). For sweeping rate bigger than 1 we can also simplify the P_{up} calculation by performing asymptotic expansion of the error functions^p (see Equation 51) and retaining only the first term.

Equation 50

$$P_{\text{up}} = \frac{1}{2} \left[1 + \operatorname{erf} \left[\frac{g_{\max}}{\sigma} \left(\left(-2 + e^{\frac{-1}{2\gamma_\lambda}} + e^{\frac{-K^2}{2\gamma_\lambda}} \right) \frac{\sqrt{2}}{2} \left(\frac{\gamma_\lambda}{\pi} \right)^{1/4} + \left(\frac{\pi}{\gamma_\lambda} \right)^{1/4} \right) \right] \right]$$

$$\leq \frac{1}{2} \left[1 + \operatorname{erf} \left[\frac{g_{\max}}{\sigma} \sqrt{2} \left(\frac{\pi}{\gamma_\lambda} \right)^{1/4} \right] \right]$$

for

$$\gamma_\lambda < 1$$

^p $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{n!(2n+1)}$

Equation 51

$$P_{up} = \frac{1}{2} \left[1 + \operatorname{erf} \left[\frac{g_{\max}}{\sigma} \frac{\left(-2 + e^{\frac{-1}{2\gamma_\lambda}} + e^{\frac{-K^2}{2\gamma_\lambda}} \right) \gamma_\lambda^{1/2} + \frac{(1+K)}{2} \left(\frac{\pi}{\gamma_\lambda} \right)^{1/2}}{(1+K)(\pi)^{1/4}} \right] \right]$$

$$\leq \frac{1}{2} \left[1 + \operatorname{erf} \left[\frac{g_{\max}}{\sigma} \sqrt{2} \left(\frac{\pi}{\gamma_\lambda} \right)^{1/4} \right] \right]$$

for
 $\gamma_\lambda \geq 1$

Several profiles for both Equation 50 and Equation 51 can be observed in Figure 35.

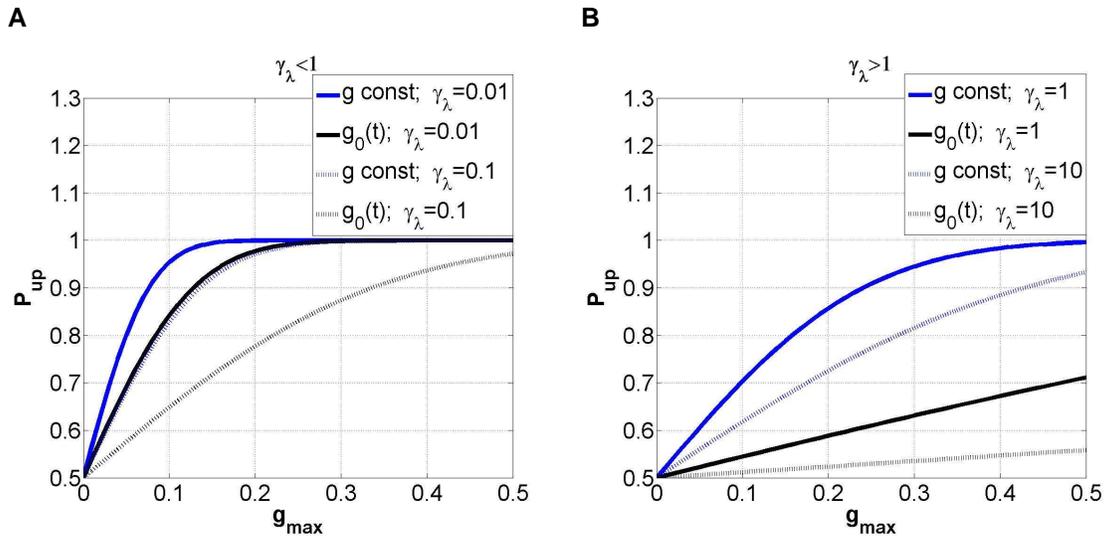


Figure 35 Branch selection probability dependence on asymmetry and sweeping speed. A) $\gamma_\lambda < 1$, $K=0.5$. B) $\gamma_\lambda > 1$, $K=0.5$. For both figures $\sigma=0.05$. Only curves for the external asymmetry reaching its maximum g_{\max} at the instant the critical parameter crosses the point $\lambda=0$ are shown.

Figure 35 A and B show that as the maximum external asymmetry is increased the probability of the system reaching the upper branch is also increased. This was to be expected from the deterministic simulations show before in Figure 26 and the conclusions taken from analysis of the variables involved in minima to

minima transition (see Equation 22) in a bistable potential with additive Gaussian noise. We can also observe speed-dependent attractor selection as was previously predicted: higher rates of passage through the critical region reduce the probability of reaching the branch of solutions favoured by the external asymmetry (compare in each Figure 35 A and B the effects of γ_λ increase and between figures). One interesting aspect of the expressions derived for P_{up} (Equation 50 and Equation 51) is the dependence on the g_{max}/σ ratio. We can see clearly that we will have the same probability P_{up} of reaching the upper branch if both maximum asymmetry and noise intensity are raised or decreased by the same factor. Another dependence not explored above is the inverse of the sweeping speed to time-scale ratio. This will be analysed ahead in section 2.3.2, Figure 44, for the biological equivalent to the supercritical pitchfork bifurcation normal form, in a paradigmatic genetic circuit. The theoretical derivations will be compared with simulation experiments in the following section.

2.2.5 Branch selectivity in the presence of additive noise: numerical results

In this section we test, through extensive numerical experiments (Figure 36), the predictions made before regarding branch selectivity in the presence of time-dependent parameters in a supercritical pitchfork normal form (Equation 18).

For an external field following a piecewise linear function (see Equation 30 and Equation 31) with $T_{g1} = \frac{T_\lambda}{2}$ (see Figure 32), we obtain coherent qualitative results with the explanation previously provided for the selectivity dependence on sweeping speed and maximal asymmetry g_{max} (Figure 36 B).

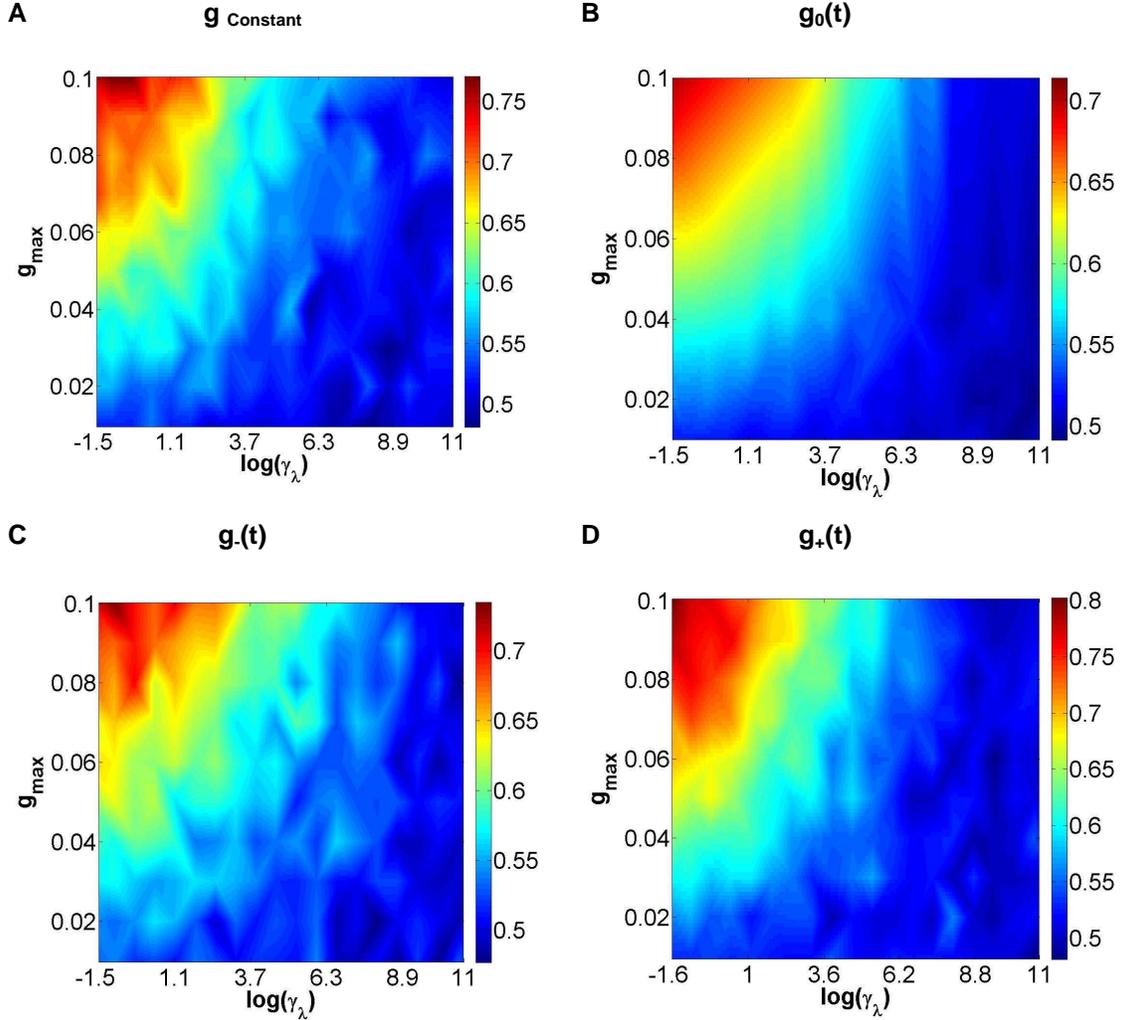


Figure 36 Probability P_{up} of reaching upper branch as a function of external asymmetry and sweeping speed. **A)** Constant g . **B)** $g_0(t)$, $T_{g1} = T_\lambda/2$, $T_{g2} = T_\lambda$. **C)** $g.(t)$, $T_{g1} = T_\lambda/4$, $T_{g2} = T_\lambda$. **D)** $g_+(t)$, $T_{g1} = 3T_\lambda/4$, $T_{g2} = T_\lambda$ (see also Figure 32). Number of simulations for A,C and D, was 1000; B, 10000 simulations. Noise intensity $\sigma=0.5$. The colormaps were created by fitting a surface, through the TriScatteredInterp linear interpolation method in MATLAB R2010b, to the data generated through simulation. For all figures a 100 by 100 grid of $(\gamma_\lambda, g_{max})$ points was sampled.

Observing Figure 36 B one verifies that there exists a minimum asymmetry $g_{max} \sim 0.03$, with $T_{g1} = T_\lambda/2$ and $T_{g2} = T_\lambda$, which synergistically acts with noise and induces a selectivity of the upper branch larger than 50%. Selectivity or P_{up} , (see Equation 52) is equated with the probability of reaching the upper branch.

Equation 52

$$P_{up} = \frac{\# \text{ runs reaching upper branch}}{\# \text{ total runs}}$$

Raising the speed with which the system crosses the critical region increases symmetry between the distributions of the final attractors, as can be seen by the predominance of blue region for larger values of $\log(\gamma_\lambda)$. Comparing the results obtained with constant and time-dependent asymmetry, Figure 36 A and B respectively, we view that although the area where the highest values of selectivity are achieved is similar, the values are higher when g is constant. This outcome supports the analytical work previously performed for the paths of the distribution mean (section 2.2.4.2, Equation 50). Regarding the selection process when the maximum asymmetry is reached before (Figure 36 C) and after (Figure 36 D) the bifurcation point, we observe that the recorded selectivities are not extremely different from those obtained for $g_0(t)$ (Figure 36 B). Yet, the number of trajectories ending in the upper branch is increased for $g_+(t)$ with respect to $g_0(t)$ (check the coloured scale). Concerning $g_-(t)$, the selectivities obtained are not particularly distinguishable from those resulting from the application of $g_0(t)$. It should be emphasized that there is a 10 fold difference in the number of simulations performed for $g_0(t)$. This results in a smoother figure which is easier to interpret. Further simulations were not performed due to time restrictions. Nonetheless, if performed, they are expected to show numerically the results predicted analytically in section 2.2.4.2. Let us also do a comparison between the analytical expressions obtained for the selectivity (see section 2.2.4) with the numerical results (Figure 37). Let us observe how the selectivity varies for a low constant sweeping speed.

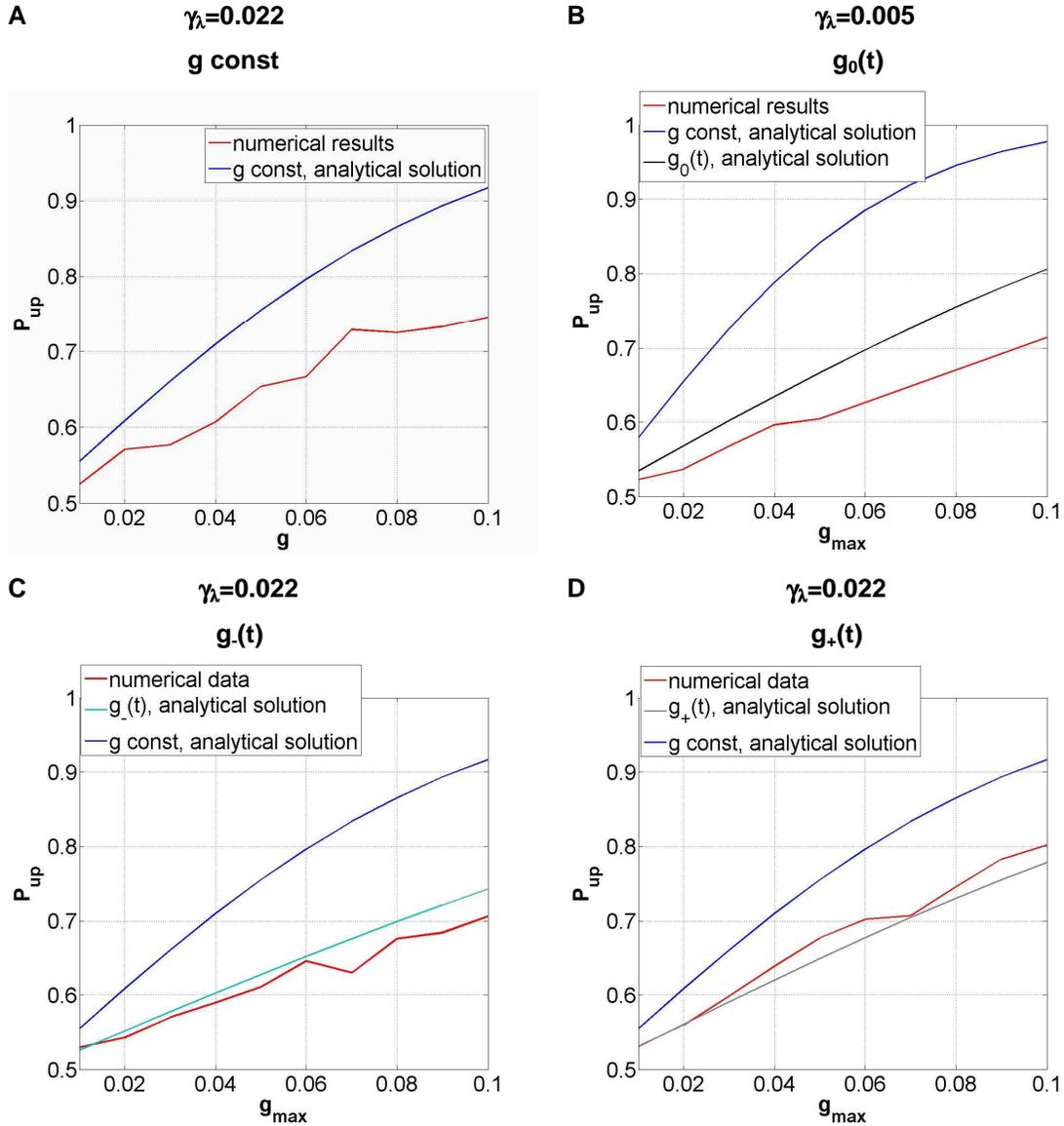


Figure 37 Branch selectivity, theory vs. numerical results. A) Constant asymmetry g , and $\gamma_\lambda = 0.022$. B) $g_0(t)$, $K = 0.01$. $\gamma_\lambda = 0.005$. C) $g(t)$, $K = 1$. $\gamma_\lambda = 0.022$. D) $g_+(t)$, $K = 0.01$. $\gamma_\lambda = 0.022$ (see also Figure 32 and Equation 50 for better understanding of asymmetry profiles and variables K , respectively). Noise intensity $\sigma = 0.5$ for all figures.

As predicted before (see Equation 38-Equation 51), selectivities with constant g are always larger than with time-dependent profiles (see Figure 37). For all the cases studied with exception of that presented in Figure 37 D, the analytical expressions always overestimate the selectivity with respect to the numerical results. This mismatch may arise due the fact that at a noise intensity σ of 0.5 branch to branch transitions may occur frequently, inducing a

considerable increase in symmetry (smaller P_{up}). Simulations with smaller noise intensities should approximate the results to the theoretical predictions. On the other hand, the theoretical predictions determined for $g_{-}(t)$ and $g_{+}(t)$ are much closer to the numerical results than those obtained with $g_0(t)$. This may be due to differences in sweeping speeds used to calculate the theoretical values presented in each figure (see Figure 37). Larger sweeping speeds seem to approximate the theoretical predictions to the numerical results much better. Further analysis is necessary to clarify why the theoretical expressions obtained for $g_{-}(t)$ and $g_{+}(t)$ are much more successful than those devised for $g_0(t)$.

The impact of external signalling on the paradigmatic systems/synthetic biology toggle switch attractor landscape can be understood by extrapolating the findings of the previous section.

2.3 Response of a gene regulatory decision switch to external signalling inputs

Integrated signalling-transcriptional networks are crucial for understanding how adequate genetic programs are induced in diverse environmental circumstances (39, 176). From rapid sensory responses to evolutionary adaptation, the transduction of signals into appropriate expression programs evolves in several timescales. Most signalling mechanisms operate in a much faster timescale than the transcriptional machinery. Extracellular signals usually change transcription factor activities in a sub-second scale, while transcription and translation of target genes may take minutes, and accumulation of protein product minutes to hours, with the additional delay due to macromolecular transport (7) (see Table 1). Usually, models in synthetic biology or systems biology separate these two areas (signalling and transcription) by recurring to quasi-steady state approximations with the advantage of transforming the models into a more manageable size and complexity. This technique may not, however, represent accurately transient dynamics (22). In the physics literature the adiabatic elimination method (112) has been extensively used. Despite its

general acceptance, if the difference in timescales is not sufficiently large, correlations between the two modes (fast and slow) arise (85, 86), thus invalidating the approximation. For this reason, in integrated signalling-transcriptional models the differences in shapes of external inputs to a gene regulatory network may become significant in the process of attractor selection and subsequently cell fate decision. Asymmetries in the distribution across the resultant attractor landscape due to path dependent effects may appear as a result of the interplay between transients and noise. This section focuses on a small integrated stochastic decision switch forced by time dependent external signals. Deterministic studies of low order gene regulatory networks such as bistable switches have been systematically analysed (52), and have been constructed de novo (95). Conditions for multistability have been determined through varying parameters controlling self-stimulation and cross-stimulation.

Regarding the action of external signals on genetic circuits most of the work has focussed on the importance of signal amplitude and signal duration on attractor selection (109) and the capacity of the system to frequency selectivity (108, 236). Important theoretical work has also been performed on the flexibility conferred by the combination of negative and positive links in switching from attractor to attractor, with the switching speed being proportional to link relative strengths (210). Noise is essential for switching from attractor to attractor when the system has an interface with a signalling module. Noise has also been a fundamental player in models and experimental systems for optimal adaptation to fluctuating environment (76). Yet, the existence of signalling structures discriminating extra-cellular signals are more efficient in generating optimal attractor switching (145).

The paradigmatic genetic decision switch has been extensively used to understand cell transition from undifferentiated to differentiated states. There have been published two approaches to this problem (133). In this section we extend the conceptual working hypothesis mentioned for cell fate decision in blood cell line commitment by exploring additional effects of the differences in the rate of sweeping through the critical region in the presence of noise. This will

constitute a test of the results observed for the paradigmatic bistable potential previously analyzed. Also, induction of attractor selection is much more efficient in systems undergoing bifurcation than in systems where the landscape of the decision switch is fixed (161, 191).

We explore through extensive simulation the effects of asymmetries in rising time between each of the input signals representing the end points of an upstream phosphorylation cascade of events (see Figure 39 B). The combinatorial complexity of the set of entry points to the transcriptional machinery is integrated by the decision switch which should classify adequately the message and migrate to a transcriptional attractor which encodes the evolutionary response to the environment (145, 176). The effect of cross-talk between pathways (interrupted gray arrows in Figure 39 B) converging on the phosphorylation of each of the transcription factor proteins present in the cytoplasm and encoded by each of the genes in the circuit will be understood through bifurcation analysis (see Figure 40 A and section 8.6). If cross-talk is not considered the differences between S_1 and S_2 are fully integrated by the decision switch. Each of the transcription factors will only operate on the respective promoter regions when in an activated state. Activation by phosphorylation is one of the most common mechanisms (39). Therefore, although the mechanism chosen for our model is quite simple, it is also completely representative. Different kinetics for phosphorylation and dephosphorylation can allow for different sensitivities to stimulus (236, 237). The signal in this case represents the concentration of a class of proteins known as kinases in an activated/phosphorylated state. These are capable of transferring a phosphate group from a molecule of ATP to a downstream kinase or in our case a transcription factor (see Figure 38). The decision switch (12, 95, 109, 255) circuit to be studied here is a motif representative of master regulator transcription factor networks in development (55, 133, 222) and

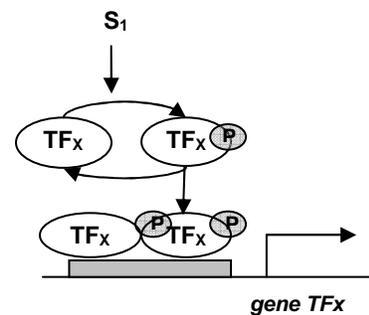


Figure 38 Phosphorylation of transcription factors necessary for initiation of transcription

two component networks exhibiting multistability. Bi-stability is a dynamical feature of core biological systems. Numerous studies, experimental and theoretical, have been devoted to understanding its natural occurrence (52, 118, 255, 267, 268). We will model our simple decision switch with dimerization and self-activation following a mean-field approach commonly used in the literature (11, 109, 222). The effects of each of the signalling inputs should have similar effects on the transcription landscape as the parameter λ and asymmetry g have on the bistable potential (see section 2.2). We will therefore, extrapolate the findings previously described to the paradigmatic signalling-gene regulatory model.

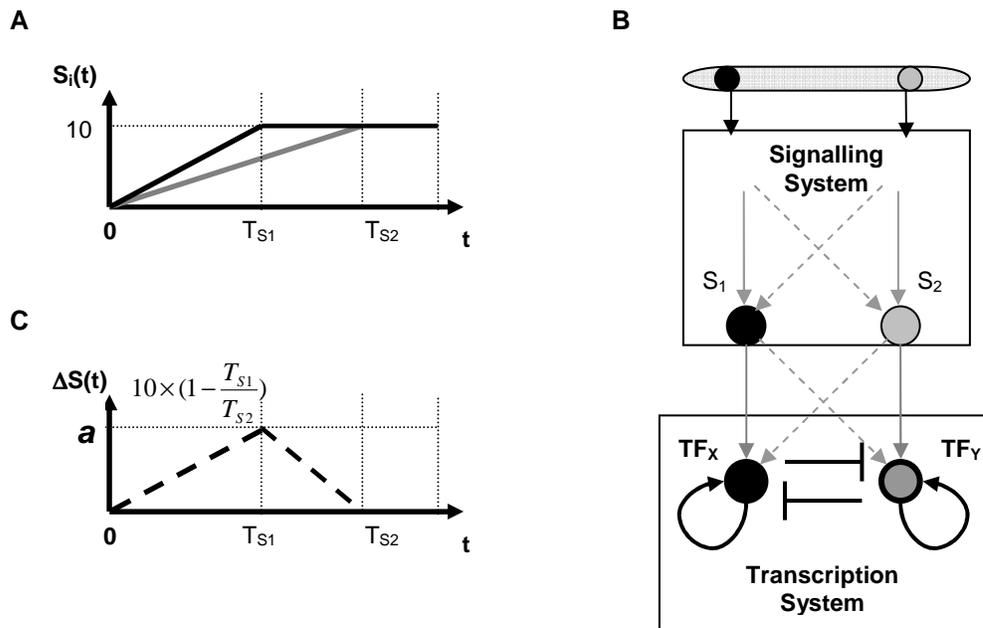


Figure 39 Integrative signalling-gene regulatory network. A) Shape of input signals $S_1(t)$ and $S_2(t)$. S_1 will always have a rising time faster than S_2 . **B)** Generalized schematic representation of the circuit model. Nodes represent proteins, S_1 and S_2 protein kinases and TF_X and TF_Y transcription factors. Reactions between kinases and transcription factors are represented through gray arrows. Cross-talk between pathways is represented through interrupted gray arrows. Reactions between transcription factor and the promoter regions is represented through (\rightarrow) for activation and (\perp) for inhibition. **C)** Amplitude of asymmetry $\Delta S(t) = S_1(t) - S_2(t)$, with $a = \max(\Delta S(t)) = 10(1 - (T_{S1}/T_{S2}))$.

2.3.1 Model

The simple model to be studied here seeks to represent, through a sufficient number of parameters with biological relevance, basic dynamic features of decision circuits.

2.3.1.1 Model assumptions

The design of the model is based on the following assumptions:

- Both transcription factors can act as activators if bound to their own promoter region (positive feedback, see Figure 39 B)
- Transcription initiation is only successful by binding of homodimers of each molecule (cooperativity equals 2)
 - Dimerization or cooperativity larger than one has been proved to be essential for multistability and clustering in this circuit (54, 54, 55, 95, 180).
 - Several types of transcription factor classes act according to this principle: bHLH proteins, leucine zipper factors or certain types of homeodomain proteins, e.g. the POU factor Oct4 or the caudal related protein Cdx2 (109). Human transcription factors are primarily Zn fingers, followed by homeobox and basic helix–loop–helix (bHLH).
- Post-translational regulation is represented by phosphorylation reactions, one of the most common mechanisms for “protein activation” (39).
- The circuit will be assumed to be in a constant volume cell (no cell division).

Homodimers will only form after phosphorylation of each of the monomers. This scenario has already been modelled by Smolen and coworkers (236) before. Nevertheless, the explicitly inclusion of phosphorylation reactions was not used.

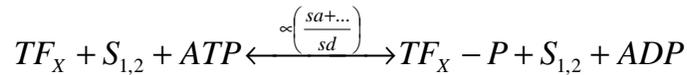
Instead, the transcription rates of each of the genes involved were changed according to some rate law.

2.3.1.2 Model equations

Under the assumptions previously outlined a set of chemical reaction equations can be written (Equation 53-Equation 59). The equations outlined are restricted to one of the elements (TF_x). For the other an equivalent set of equations is necessary for full model description.

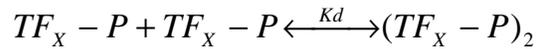
(Phosphorylation-dephosphorylation)

Equation 53



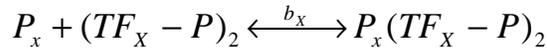
(Dimerization)

Equation 54



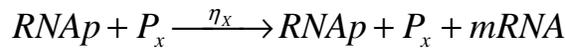
(Binding of dimer to promoter region P_x)

Equation 55



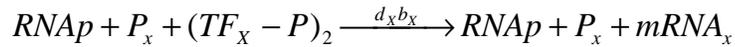
(Basal transcription)

Equation 56



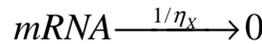
(Transcription)

Equation 57



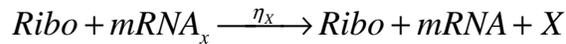
(mRNA degradation)

Equation 58



(Translation)

Equation 59



The dynamics of the decision genetic switch can be modelled by following the reaction schemes represented in Equation 53 to Equation 59. The effect of external signals $S_{1,2}$ (see Figure 39 B) will be modelled by phosphorylation of

transcription factors (Equation 53) according to the mass-action law (Equation 60 and Equation 61, with $[TF_{X,Y}] = X, Y$ and $[TF_{X,Y}-P] = X^a, Y^a$, see also section 1.4.1 for understanding modelling framework). The input function to each of the promoter regions (see P_X for example in Equation 55, see also section 1.2.6 for clarification of mechanisms) will be modelled according to the formalism developed in (11, 109). The methodology in both papers for derivation of the input functions is based on a mean-field approach to promoter site occupation (33, 34, 40). The concentration of each of the transcription factors depends not only on transcription and translation of new molecules (X, Y) but also on the decay of activated/phosphorylated ones (X^a, Y^a) (see Equation 62 and Equation 63). We will not consider degradation of activated molecules.

The model equations were devised by assuming a quasi-steady state approximation of mRNAs with respect to protein accumulation (Equation 55 to Equation 59) and under a “mean-field” approach to promoter site occupation for initiation of transcription (11) (Equation 62 and Equation 63). It is assumed that each transcription factor homodimer has an independent promoter site. The difference in time scales between phosphorylation reactions and the transcription and translation processes is generically represented by a set of parameters $\tau_{S_{X,Y}}$ and $(1/\tau_{TF_{X,Y}})$ (Equation 60-Equation 63).

Equation 60

$$\tau_{S_x} \frac{dX^a(t)}{dt} = [X(t) \times (sa + k_{x1}S_1(t) + k_{x2}S_2(t))] - (X^a(t) \times sd)$$

Equation 61

$$\tau_{S_y} \frac{dY^a(t)}{dt} = [Y(t) \times (sa + k_{y1}S_1(t) + k_{y2}S_2(t))] - (Y^a(t) \times sd)$$

Equation 62

$$\frac{dX}{dt} = \frac{1}{\tau_{TF_x}} \left[\eta_x \frac{1 + d_x b_x (X^a(t))^2}{1 + b_x (X^a(t))^2 + g_x (Y^a(t))^2} - X \right] - \frac{dX^a(t)}{dt} + \sigma_x \xi_x(t)$$

Equation 63

$$\frac{dY}{dt} = \frac{1}{\tau_{TF_y}} \left[\eta_y \frac{1 + d_y b_y (Y^a(t))^2}{1 + b_y (Y^a(t))^2 + g_y (X^a(t))^2} - Y \right] - \frac{dY^a(t)}{dt} + \sigma_y \xi_y(t)$$

In the model equations sa stands for a residual or spontaneous transition from inactive to active state and models mechanisms of activation not contemplated in this simple model (see Equation 53). The rate sd represents natural decay of a phosphorylated protein to its non-active state or the action of phosphatases whose function is to remove the phosphate group and attenuate the signal. We assume a constant concentration for this type of proteins (see Equation 53). This is a common approach in pathway modelling (51, 230). Parameters d_i stand for the ratio between rate of expression of the respective gene when X^a or Y^a homodimers are bound to the promoter region with respect to basal transcription (see also Equation 56 and Equation 57). Parameters b_i and g_i represent ratios between binding and unbinding of dimers to promoter regions for self-activation and cross-inhibition, respectively (11) (see also Equation 54 and Equation 55). This parameter can be seen as a threshold for gene transcription. Parameters η_i correspond to a measure of promoter strength multiplied by translational efficiency, i.e., basal transcription rate multiplied by translation rate divided by mRNA degradation rate (see also Equation 56 and Equation 58). The functions $\xi_{X,Y}(t)$ represent Gaussian noise with zero mean and correlation $\langle \xi_X(t) \xi_Y(t') \rangle = \sigma_{X,Y}^2 \delta_{X,Y} \delta(t-t')$ and models the contribution of intrinsic random fluctuations inherent to transcription and translation processes (243). The consequence of extrinsic source of noise in circuit behaviour, such as fluctuations in kinase number, will not be approached here. Guantes and coworkers (109) found that fluctuations of an additive external signal may help to increase circuit efficiency in classifying differences in S_1 and S_2 (Figure 39 B) signal duration. The multiplicative noise term is interpreted in the Ito sense (94). This is the appropriate stochastic interpretation for a realistic noise arising from stochastic binding events (135, 258). $\sigma_{X,Y}$ is the strength of the noise which follows Equation 64. This expression is the correct form for chemical reaction events taking place in time as a Poisson process (258). We'll use a simple Heun method for integrating the differential equations (209) (see section 8.4).

Equation 64

$$\sigma^2_X = \left[\frac{1}{\tau_{TF_X}} \times \left[\eta_X \frac{1 + d_X b_X (X^a(t))^2}{1 + b_X (X^a(t))^2 + g_Y (Y^a(t))^2} + X \right] + \frac{dX^a(t)}{dt} \right]$$

For more accurate biological representation a delay should be included. This would represent in a condensed way the processes of elongation, splicing, processing and export of primary gene transcripts. These intermediate steps are very complex and time consuming (16, 23, 257). An extreme of this process is the dystrophin gene taking up to 16 hours (247). Overall, the average delay is between 10 to 20 min between the initiation induced by a transcription factor on a specific gene's promoter region and the appearance of mature mRNA in the cytoplasm. This delay includes transport. Translation has also a typical delay of 1 to 3 min. These time-scale differences with signalling processes will be simplified and modelled by the constants $\tau_{S,X,Y}$ and (see Equation 62 and Equation 63).

2.3.2 Symmetry breaking and attractor selectivity by time-dependent external signals in a fast/slow stochastic cell fate decision regulatory network: numerical results

With the intent of determining the number of attractors available in each combination of input amplitudes, a bifurcation analysis (see Figure 40 A and section 8.6) of the circuit was performed using the software XPPAUT (18). As was previously mentioned, the type of bifurcation chosen to model the process of cell fate decision is fundamental in understanding how specific expression programs exist. The two fundamental types of bifurcation used extensively to study differentiation are the subcritical pitchfork bifurcation and the supercritical Pitchfork bifurcation (78, 133). Several bifurcation studies have been performed on paradigmatic differentiation problem, e.g. lineage-commitment in bipotent blood progenitor cells (see Figure 17). We will draw from this study and explore further the action of external signals on the genetic circuit. The type of bifurcation more often thought as describing the several stages of blood cell line commitment is the subcritical pitchfork bifurcation. Very insightful work into the effects of a metastable state (the progenitor cell) on the transition rates to the

final attractors (differentiated cells) could inform on peculiar dynamical aspects of cell differentiation and differential processing of combinatorial external signal complexity (201). The theoretical efforts behind that type of endeavour demand a further use of first exit time theory which is beyond the scope of this project. Moreover, we are interested in generic features of cell fate decision: the importance of synergistic effects of asymmetries, fluctuations and speed-dependent cellular decision making. Therefore, in order to establish a strong link with analytical solutions previously devised (Equation 50 and Equation 51), and understand their informative potential on basic principles not yet studied for genetic circuits, we decided to work with the simplest canonical supercritical Pitchfork bifurcation. In the simulations performed we will resort to a simplified circuit with no cross-talk. The possible effect of cross-talk between S_1 and TFy, for example, results in a contraction of the bistability region (see section 8.6.2.1).

The bifurcation cusp ($(S_1, S_2) = (0.5, 0.5)$) resulting from varying signal input amplitudes is visible in Figure 40 A. The area corresponding to bistability is located between the two hysteresis borders. The combination of signals S_1 and S_2 (see Figure 39 B) forces a supercritical Pitchfork bifurcation on the transcription circuit, in a specific configuration of parameters. The differences in amplitude in time force an asymmetry on the decision switch (section 8.6) that is qualitatively similar to the effect of the external field $g(t)$ in the bistable potential canonical problem analysed previously (see section 2.2). To test if the decision genetic system obeys similar principles we performed a set of simulations with several combinations of S_1 and S_2 . In the course of the simulations performed (without cross-talk), the external signals S_1 and S_2 were changed linearly according to $S_i(t) = (10t/T_{Si})$ from point P_i ($(S_1, S_2) = (0, 0)$) to point P_m and finally to point P_f ($(S_1, S_2) = (10, 10)$) on the bisectrix (see Figure 40 A). Point P_m will be different for each combination of T_{Si} . Along the direct path $\overline{P_i P_f}$, a slice of the X manifold (or Y) would reveal a bifurcation diagram with a perfect supercritical pitchfork bifurcation (completely connected branches). Along a path $\overline{P_i P_m}$ in the codimension-2 bifurcation diagram (Figure 40 A) the disconnection between branches is revealed (Figure 41). The minimal distance between branches d

(Figure 41) is proportional to the maximal difference between S_1 and S_2 during the simulation. Asymmetries caused by differences in S_1 and S_2 rising times provoke proportional differences in the size of the basins of attraction at each time step. The basin of attraction border moves according to the external signal differences, with noise being responsible for crossing the boundary or potential barrier. The speed with which the basin border moves may also play a role. Since the regulatory landscape is responding to the external signals the path which the system follows to appropriately commit the cell to a certain outcome is relevant. Signals arriving at the nucleus are encoded not only by amplitude, duration (109) and sequence (84) but also how fast they induce changes in the transcriptional landscape. The response of the switch studied in this work to more complicated combinations of input signals will be reported elsewhere.

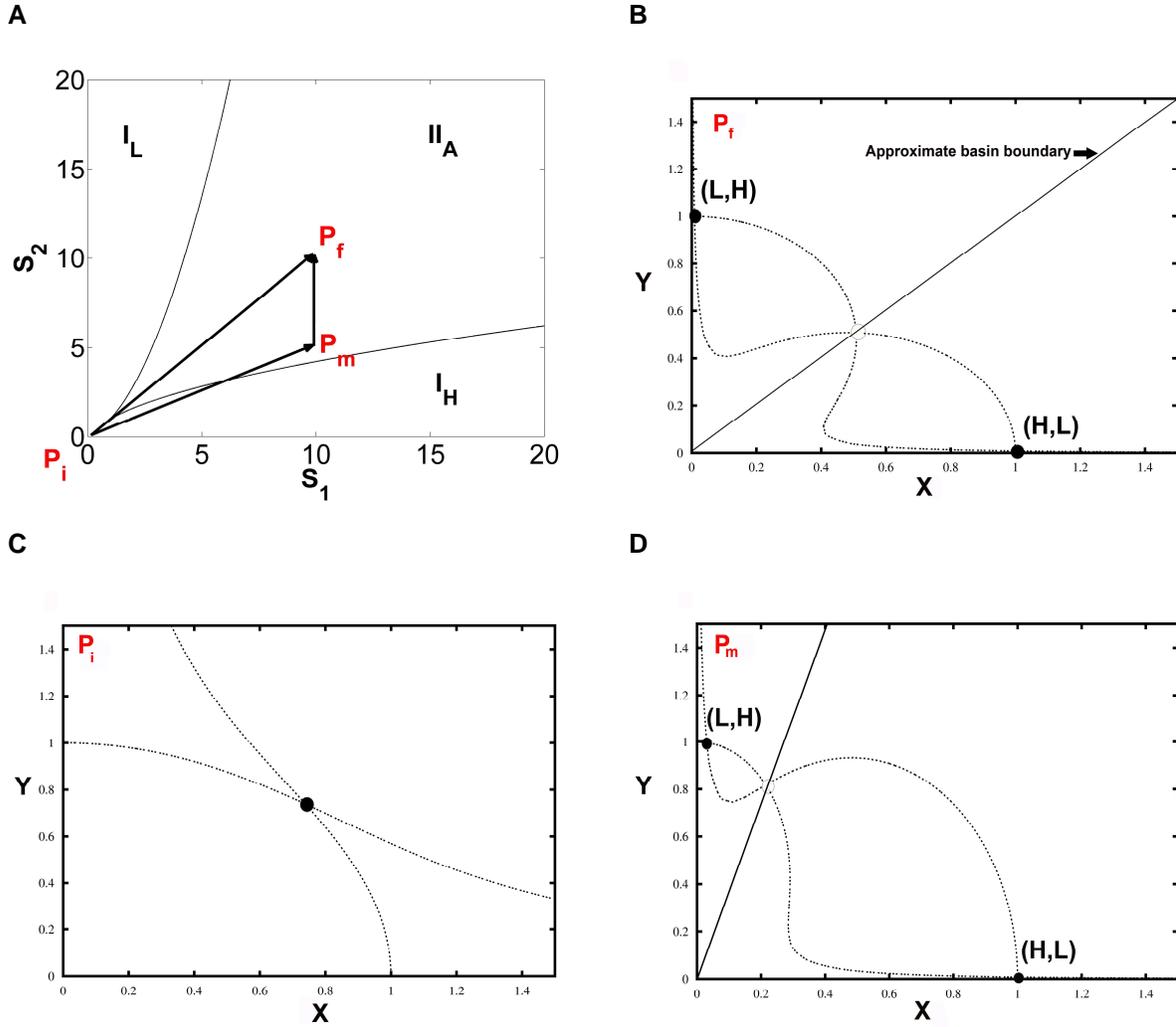


Figure 40 Bifurcation diagrams and nullclines for the decision genetic switch. A) Two dimensional bifurcation diagram (for component X) for the circuit observed in Figure 39 B with no cross-talk ($k_{x2}=k_{y1}=0$, see Equation 60 and Equation 61). The diagram was obtained by varying the amplitude of each of the input signals. Point P_i corresponds to the initial amplitudes of S_1 and S_2 which are set to 0. The steady state values of each of the variables in the circuit at P_i are the following, $(X^a, Y^a, X, Y)_A = (0.739, 0.739, 0.739, 0.739)$. Point P_f corresponds to the final amplitude achieved during the simulation $(S_1, S_2) = (10, 10)$. Point P_m corresponds to an intermediary point at $t = T_{S1}$ (see Figure 39C) achieved in the path between P_i and P_f resulting from the differences in sweeping rates between each of the input signals. B) Nullclines at P_f . C) Nullclines at P_i . D) Nullclines at P_m . Parameters $d_x=d_y=1$; $b_x=b_y=1$; $g_x=g_y=1$; $\eta_x=\eta_y=1$; $sa=sd=1$; $\sigma=0$ (see Equation 60 and Equation 64). The approximate basin of attraction boundary represented in B is a drawn projection of the 4 dimensional system studied here. It was determined by inspection in XPPAUT (18).

An example of the evolution of concentrations of each of the transcription factors in the circuit is shown in Figure 42 A. The distribution of trajectories (corresponding to 1000 trajectories), for each of the steady states, for 3 time

points is also presented. Initially the distribution is monomodal and starts to broaden until the bifurcation point is reached. Just before the bifurcation an amplification of noise due to the broadening of the potential is expected (279). The bifurcation extends the broadening until the onset of a bimodal distribution.

In Figure 42 B the corresponding trajectories in phase space (X,Y) are shown. Both the stochastic trajectory and the numerical deterministic solutions are presented for two transcription time-scales ratios: $\tau_{TFX,Y}/\tau_{SX,Y} = 1$ and 10. When the time-scale separation between phosphorylation reactions and transcription processes is higher, the trajectories spend naturally more time near the basin border (section 2.3.2), which corresponds to the location of the unstable solution or potential barrier. As will be proved ahead this will have clear impact on the capacity for the system to discriminate between differences in external signals S_1 and S_2 . It can also be seen that the variation around the deterministic path is higher for smaller time-scales. This particular phenomenon is a clear consequence of the noise term chosen (Equation 64), which follows the standard form for fluctuations of chemical reactions arising as a Poisson process (258).

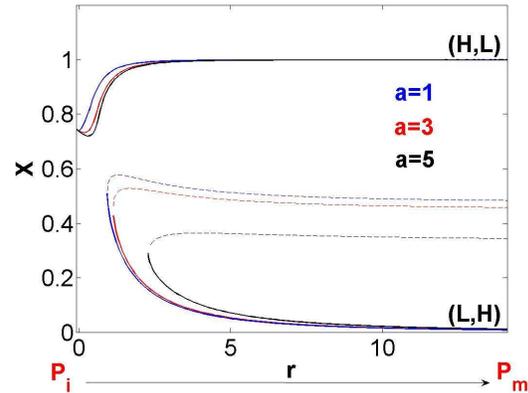
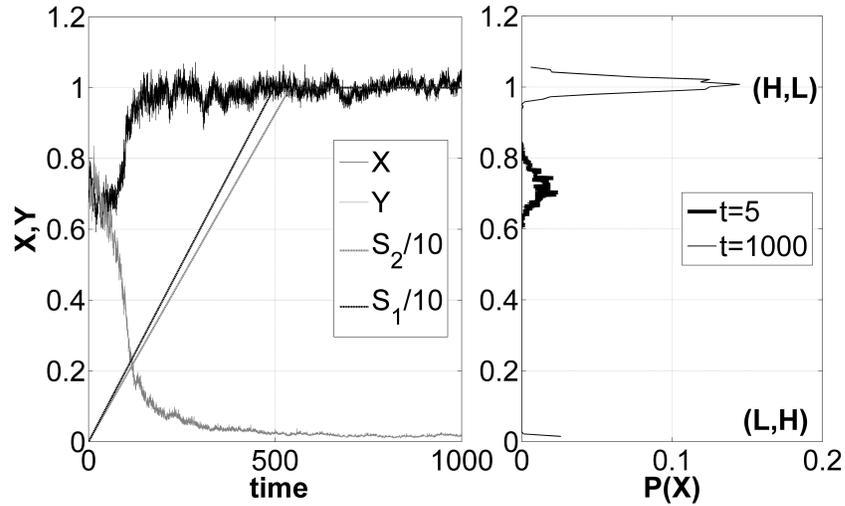


Figure 41 Bifurcation diagram for component X with different parameterization, from P_i to P_m (see Figure 40 A). a stands for the maximum asymmetry between external signals observed during the simulation (see Figure 39C). r stands for the distance between the origin and a point along the path $\overline{P_i P_m}$. It can be calculated by the following expressions, $r = S_1 / \cos(\arctg((10-a)/10)) = S_2 / \sin(\arctg((10-a)/10))$. Parameters: $d_x = d_y = 1$; $b_x = b_y = 1$; $g_x = g_y = 1$; $\eta_x = \eta_y = 1$; $sa = sd = 1$ (see also Equation 60-Equation 64).

A



B

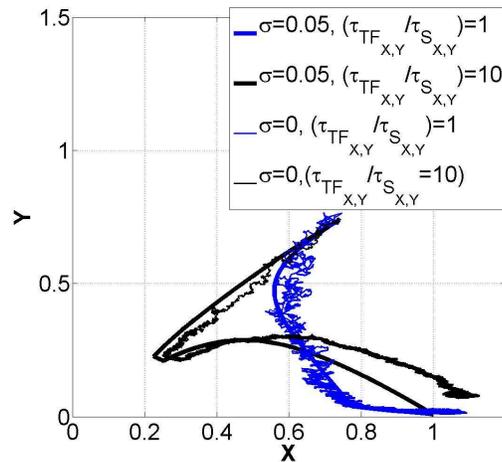


Figure 42 Examples of stochastic and deterministic trajectories for the decision genetic switch. A) Left, evolution in time of the concentration of transcription factors X and Y for a generic combination of input signals, $(\tau_{TF_{X,Y}}/\tau_{S_{X,Y}})=1$ and $\sigma=0.05$ (see also Equation 60-Equation 64). Right, initial and final distributions for 1000 cells for transcription factor X. B) Trajectories in phase space for $\sigma=0.05$ and two time-scales ratios, $(\tau_{TF_{X,Y}}/\tau_{S_{X,Y}})=1$ and $(\tau_{TF_{X,Y}}/\tau_{S_{X,Y}})=10$. Parameters: $d_X=d_Y=1$; $b_X=b_Y=1$; $g_X=g_Y=1$; $\eta_X=\eta_Y=1$; $s_a=s_d=1$ (see also Equation 60-Equation 64).

Another conspicuous aspect observed in Figure 42 B is the fact that when noise is considered there is a substantial mismatch between the final attractor of the trajectories that was not observed in the deterministic scenario. This may arise due to the fact noise is multiplicative and its mean constituent contributes to the deterministic part of the dynamics. This will not, on the other hand, alter

substantially the selectivity studies performed in the following section. Although the final attractor is displaced, the trajectory commitment is final following the switching during the critical sweeping process (see Figure 42 B). Attractor-attractor transition occurrences are completely quenched.

2.3.2.1 Genetic decision switch attractor selectivity dependence on sweeping speed and external signal asymmetry

To understand the capacity of the decision switch to discriminate between differences in sweeping rates and consequently paths in the codimension-2 bifurcation diagram, a 100 by 100 grid of combinations of maximum asymmetry and sweeping speed was selected according to Equation 65 :

Equation 65

$$T_{S2} = \frac{T_{S1}}{\left(1 - \frac{a}{10}\right)}$$

$$a = \max_{t \in [0, t_{final}]}(\Delta S(t))$$

The sensitivity of the decision switch to small asymmetries in the presence of fluctuations, even in the case of $a \ll \sigma_{X,Y}$, should indicate how likely the shape of the external signals influences cell fate decision. The selectivity in the final simulation step for each of the input combinations was determined by calculating the percentage of number of runs in the vicinity of each of the attractors. The results varying both noise strength σ (Equation 64) and transcription plus translation timescale $\tau = \tau_{TFX} = \tau_{TFY}$ (Equation 62 and Equation 63), are presented in Figure 43.

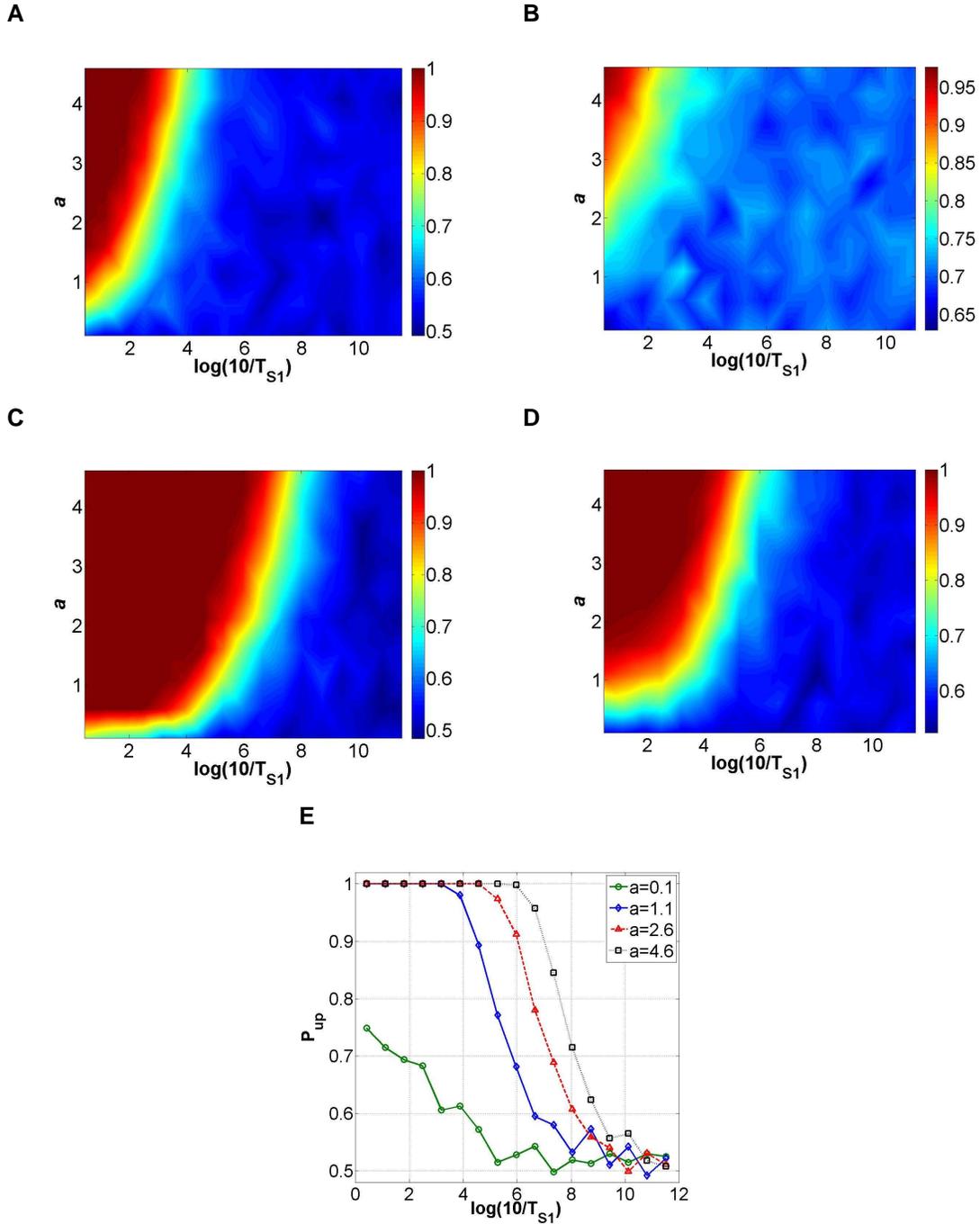


Figure 43 Selectivity vs. asymmetry and sweeping speed for several noise and time-scale ratios. A) $\sigma=0.01$ and $(\tau_{TFX,Y}/\tau_{SX,Y})=10$. B) $\sigma=0.05$ and $(\tau_{TFX,Y}/\tau_{SX,Y})=10$. C) $\sigma=0.01$ and $(\tau_{TFX,Y}/\tau_{SX,Y})=1$. D) $\sigma=0.05$ and $(\tau_{TFX,Y}/\tau_{SX,Y})=1$. The pictures are a result of fitting a surface to the simulated data through the TriScatteredInterp linear interpolation method (MATLAB R2010b). Selectivity or P_{up} is represented in a coloured scale. E) Selectivity P_{up} for several asymmetries a , for $\sigma=0.01$ and $(\tau_{TFX,Y}/\tau_{SX,Y})=1$. Parameters $d_x=d_y=1$; $b_x=b_y=1$; $g_x=g_y=1$; $\eta_x=\eta_y=1$; $s_a=s_d=1$ (see also Equation 60-Equation 64).

As in the original findings in the bistable potential (161, 191) and the numerical simulations of section 2.2.4, the speed ($\log(10/T_{S1})$) with which the system crosses the critical region influences clearly the sensitivity to small asymmetries. A set of cross-sections of Figure 43 C for several external asymmetries and variable sweeping speeds show exactly this speed-dependent cell fate decision mechanism (Figure 43 E). Observe that for higher sweeping speeds the cross-sections obtained become indistinguishable (Figure 43 E).

As can be observed in Figure 41, the asymmetry depends on the crossing point between the hysteresis lines in the 2D bifurcation diagram and the path $\overline{Pi Pm}$. Although the system regains symmetry, it momentarily is exposed to asymmetric signals. As in the bistable potential, larger asymmetries generate larger selectivities and larger intensities of noise blur clear asymmetries between signals. Observe that the border between high (red region) and low (blue region) selectivities shifts towards the left and upwards. According to the theoretical derivations performed on 1 dimensional systems in section 2.2, the selectivity of the connected branch depends on the signal g_{max} to noise ratio but also on ratio between the inverse of sweeping rate (Equation 66). Due to the dimension of the decision genetic switch with external stimulation, solving the Fokker-Planck equation by the moments method (94) with a multiplicative noise term given by Equation 64 is a cumbersome task. We chose, due to time restrictions, to qualitatively evaluate the selectivity dependence on the signal to noise ratio and the sweeping speed to time-scale ratio. The time-scale difference between phosphorylation and transcription processes ($\tau_{TFX,Y}$ in Equation 62 and Equation 63) is the additional dependence not explored above in one dimensional systems but demonstrated by Kondepudi and coworkers in electronic circuits (Equation 66) (161, 191).

Equation 66

$$P_{up} \propto \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{g_{max}}{\sigma} \times F \left(\left(\frac{1}{\gamma_{\lambda} \tau} \right)^{\alpha} \right) \right) \right] \quad \text{with } \alpha < 1$$

In Equation 66 function F should depend on the set of equations describing the system undergoing a dynamic bifurcation. See for example the calculations performed in section 2.2.4.2. Yet, the scale $\left(\frac{T_{S1}}{10\tau}\right)^\alpha$ dependence can still be checked by evaluating the numerical data obtained. When the time-scale ratio is raised from $\tau_{TFX,Y}/\tau_{SX,Y} = 1$ to 10, for a noise amplitude equal to 0.01, the area where selectivity one is observed is reduced (Figure 43). This could be justified, at a first glance, because the system spends more time near the unstable basin border with jumps to the disconnected branch of attractors (L,H) being more probable (see Figure 42 B). Following the parameter dependencies implicit in Equation 66, if similar quantitative aspects underlie the genetic switch, we should, for a specific signal to noise ratio, observe similar selectivities if the sweeping rate to time-scale ratio is maintained.

For the numerical data acquired, for example for noise strength equal to 0.01 and a variation of $(\tau_{TFX,Y}/\tau_{SX,Y})=1$ by a factor of ten, several selectivity profiles as a function of g_{max} are represented in Figure 44. For $(\tau_{TFX,Y}/\tau_{SX,Y})=10$, the closest curves to the selectivity obtained for $(\tau_{TFX,Y}/\tau_{SX,Y})=1$ are both associated with a decrease in sweeping speed of approximately 1 order of magnitude. Graphically, this is equivalent to cross sections P_{up} vs a in Figure 43 A being similar to those observed in Figure 43 C but shifted to the left by approximately $\log(1/10)$. If the number of runs performed were higher, smoother profiles would

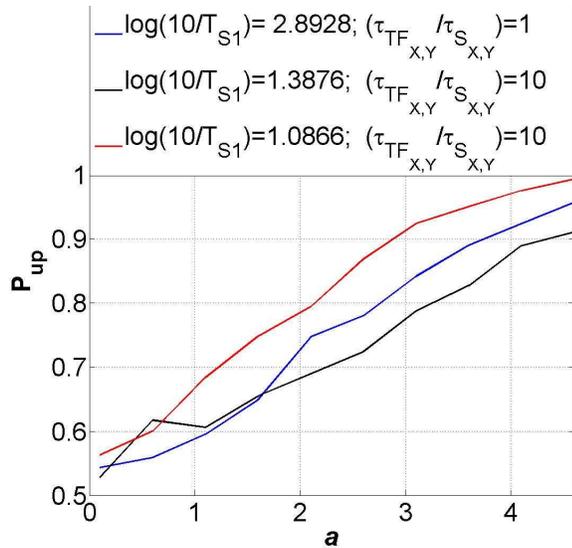


Figure 44 Probability P_{up} of reaching the upper branch (numerical data) as a function of maximum asymmetry a for several time-scale ratios $(\tau_{TFX,Y}/\tau_{SX,Y})$ and sweeping speeds $(10/T_{S1})$. Parameters $d_x=d_y=1$; $b_x=b_y=1$; $g_x=g_y=1$; $\eta_x=\eta_y=1$; $sa=sd=1$ (see also Equation 60-Equation 64)

have been obtained and possibly better accordance between numerical results in Figure 44.

The shape of the external signal has been explored in connection with the cell outcome induced. Transient versus sustained activation is usually correlated with cell fate decision, although there is a significant variance between cell lines and pathways, e.g. Epidermal Growth Factor (EGF) competing with the Tumour Necrosis Factor (TNF) activation of output kinases (91, 194, 195, 259). It is likely that transcription factor networks controlling the expression of other proteins and the network itself clusters or integrates a combination of inputs with different shapes into attractors with the most appropriate response to the environmental signals. Time-scale separation effects should have a more profound effect in the capacity for gene regulatory circuits to discriminate between time-dependent signals. Only external signals held at a level for an interval comparable to the transcription plus translation processes' time-scale have a high correlation coefficient with the attractor selected. Traditionally, in these cases quasi-steady state or adiabatic approximations are a practical simplification technique for understanding long-term feature of dynamical systems. Normally, in gene regulatory networks this technique arrives at much faster convergence times to the slow manifold. Other techniques have been developed which approximate considerably better the transient dynamics (22). Systems with well separated time-scales can usually be cast in fast-slow canonical forms (24). In other situations, most notably if the systems admit bifurcation points, several dynamical aspects may not be accurately described by slow-fast paradigmatic forms: phenomena such as jumps between distinct parts of the slow manifold, bifurcation delays and dynamical hysteresis loops (see section 2.2.2), and relaxation-oscillation regimes.

2.4 Summary

We have shown in both canonical models in Physics and Synthetic/Systems Biology the effects of the speed of bifurcation parameter sweeping in the presence of fluctuations and small asymmetries: even in the presence of strong fluctuations, slow passage through the critical region increases sensitivity to external asymmetries. These results will have further impact in investigations of genetic circuits with high dimension and undergoing more complex types of bifurcation. It is a fundamental mechanism that completes the set of characteristics of external signals studied so far in the literature. The impact on understanding how cell networks should be perturbed is also expected to be original. Most of the therapies currently used focus on the signalling system. Therefore, we are bound to the effect of time-dependent protein concentrations inducing distinguishable attractor selection.

The results for the canonical supercritical Pitchfork bifurcation normal form helped us framing the findings for both time-dependent critical parameter and external asymmetry. In fact, there's equivalence between this model and the decision genetic switch. Further analytical work on the genetic switch will also help us in the analysis of the effects of multiplicative versus additive noise in the attractor selection problem. This problem was not discussed here. Also, further development of the analytical expressions devised for the canonical model, should be extrapolated to the decision switch and help us understand if similar effects of reaching the maximum asymmetry before and after the critical region is relevant for cell fate decision.

3 Path-dependent pattern selection effects in high-dimensional integrative signalling-gene regulatory intracellular networks

3.1 Introduction

Induction of a transcriptional program by external signalling inputs is a crucial aspect of intracellular network functioning. As previously reviewed (section 1.5) the theoretical concept of attractors representing particular transcriptional programs is reasonably adapted to experimental observations of “whole-genome” expression profiles (131). These can be associated either with developmental outcomes such as differentiation into specific types of cells, or maintenance of cell functioning such as proliferation or apoptosis. An example of a gene expression pattern selection generated during differentiation can be viewed in Figure 17 B. We recommend the reader to consult the data generated by Huang and coworkers for the cyclic pattern observed during HeLa cell cycle^q. Here we will examine the clustering of external signals by a highdimensional switch circuit (see Figure 45). The crucial characteristics of input signals such as the speed with which they induce high-dimensional transcription landscape changes will be tested. The combination of external signals effectively induces visible changes in the trajectories in phase space. Due to effects of fluctuations in conjunction with external signal characteristics, the paths taken in high-dimensional space will change the probability of attractor selection.

^q <http://www.childrenshospital.org/research/ingber/GEDI/gedihome.htm>

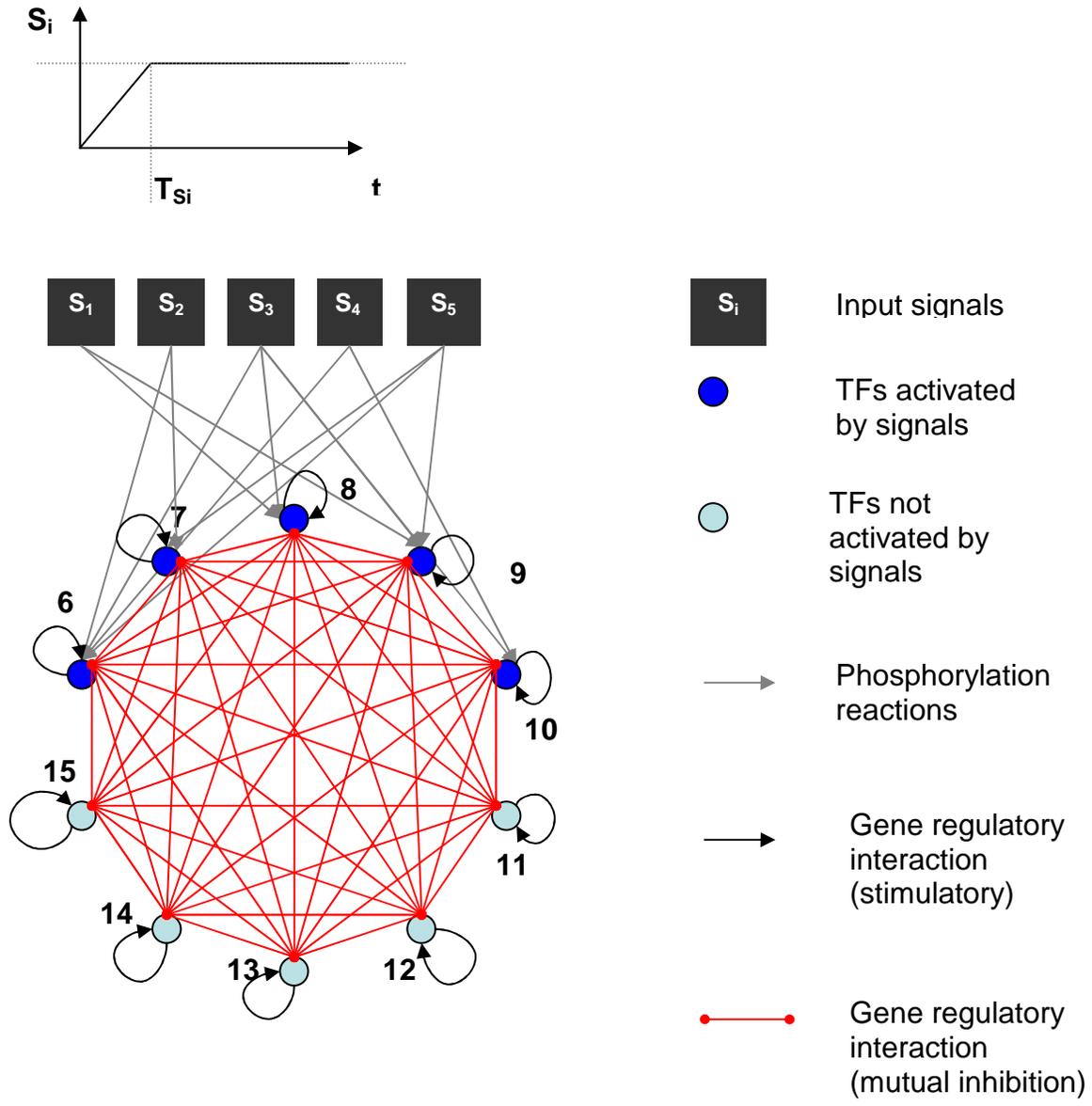


Figure 45 Representation of a highdimensional genetic decision switch with 10 transcription factors (nodes 6 to 15) and 5 input signals. Only nodes 6 to 10 need to be activated (phosphorylated) to act on any promoter region of the rest of the transcription factors in the network. Each transcription factor reinforces its own expression and represses all other nodes. The connectivity matrix between the set (S_1, \dots, S_5) and nodes 6 to 10 is represented in Table 2.

The highdimensional switch has been explored before by Cinquin and coworkers in modelling generalized, switch-like competitive basic Helix-Loop-Helix (bHLH) heterodimerization networks in the context of differentiation (53-55).

A set of rules for the clustering capacity of this type of network was devised as a function of competition between synthesis, degradation and complex formation rates of different elements. Here we will assume a specific type of network parameters that induce multistability but in a different class of models from Cinquin's work (53-55). Essentially, we will employ a generalized version of the "mean-field" model analyzed before for the bistable decision genetic switch (11) (see section 2.3.1). The sequences (S_1, \dots, S_N) (see Figure 45) will only differ on their rising time T_{Si} . As with the bistable switch previously studied, the differences in rising times impose time-dependent asymmetries which are processed by the network. Unlike the low order decision genetic switch, here we additionally consider an extra layer of nodes (TF_{11}, \dots, TF_{15}) (see Figure 45) that should respond to the activity of the "genomic gateway" set of nodes (TF_6, \dots, TF_{10}) (see Figure 45). We chose to work with five inputs because it stands as the number of nodes most often associated in the literature with competing attractor selection by signals. Usually, the external signals studied are (122):

- Akt-activity correlated with apoptosis
- Erk –activity linked with proliferation
- Rac -correlated with cytoskeletal regulation
- Sapk and p38-stress outputs

3.1.1 Modelling the highdimensional switch

For each TF-TF connection, associated with a protein-gene interaction or regulatory process, we will resort to a generic representation shown in Equation 67. It represents a mean-field approximation to the combinatorial logic operating on the promoter region of a gene. All regulatory interactions to any gene are replaced with an average or effective interaction, taking into account the repression, activation and multimerization mechanisms inherent to epigenetic regulation. This formalism follows Andrecut and coworkers paper (11). If we recall the set of equations previously used in section 2.3 for the small genetic switch, they stand as a generalization that takes into account all possible

reactions between nodes. Previously, only formation of homodimers was allowed. Here, we will allow for hetero and homodimers.

Equation 67
$$\frac{dX_i(t)}{dt} = \frac{1}{\tau_i} [-X_i(t) + F(X(t))] - \frac{dX_i^a(t)}{dt} + \xi_i,$$

$$X(t) = [X_i(t)], i, j = 1, \dots, n$$

$$F_i(X(t)) = \eta_i \frac{1 + \overbrace{\sum_j \gamma_j^i \beta_j^i c_j^i a_j^i X_j(t)}^m + \overbrace{\sum_j \sum_{j'} \mu_{jj'}^i \lambda_{jj'}^i \alpha_{jj'}^i d_{jj'}^i b_{jj'}^i X_j(t) X_{j'}(t)}^d + \dots}{1 + \sum_j \beta_j^i a_j^i X_j(t) + \sum_j \sum_{j'} \lambda_{jj'}^i \alpha_{jj'}^i b_{jj'}^i X_j(t) X_{j'}(t) + \dots}$$

$$\langle \xi_i(t) \rangle = 0$$

$$\langle \xi_i(t) \xi_i(t') \rangle = \sigma^2 \sqrt{X_i} \times \delta(t - t')$$

In Equation 67 we only show terms associated with monomer (m) and dimer (d) formation. Other terms generate a very long and complicated formulation. In Equation 67 we also only show terms associated X_j . The complete expression including also X_j^a would be too long. For that reason only we omitted them. During the simulation studies this species is also considered. As in the small dimension switch (see Equation 53 - Equation 59) parameters have specific meaning:

- c_j^i stands for the ratio between rates of expression of the respective gene and basal transcription when monomers j are bound to the promoter region.
- $d_{jj'}^i$ stands for the ratio between the rate of expression of the respective gene when dimers ($j-j'$) are bound to the promoter region with respect to basal transcription.
- a_j^i represents ratios between binding and unbinding of monomers to promoter regions.

- $b_{jj'}$ represents ratios between binding and unbinding of dimers ($j-j'$) to promoter regions. This parameter can be seen as a threshold for gene transcription.
- η_i correspond to a measure of promoter strength multiplied by translational efficiency, i.e., basal transcription rate multiplied by translation rate divided by mRNA degradation rate.
- The matrices α , β , γ , λ and μ represent binary selection switches for each reaction involved. α stands for the formation of dimer complexes prior to binding to the promoter region. β represents a reaction of a transcription factor in monomer stage with a downstream gene. The product $\lambda\alpha$ controls the inclusion of repression reactions exerted by dimers. γ will be 0 if a monomer formed represses the expression of a downstream gene. μ will be 0 if the dimer formed exerts repressive action on the downstream gene.

This formalism establishes a connection between a Boolean representation of the links between nodes and a continuous framework. Equation 67 has been proved, on the other hand, to be equivalent to a much more simplified version contemplating multimers up to order M (11) (see F_i in Equation 68). The simplified version for the circuit in question takes the following form when all the binary tensors, α , β , etc, take the associated values to the highdimensional switch. We chose to use the simplified version due to its compact way of dealing with the complex set of reactions inherent to the transcription initiation process.

Equation 68

$$F_i(X(t)) = \left[\eta_i \frac{\left(c^i a^i X_i^a(t) \right)^{M+1} - 1}{\left(a^i X_i^a(t) + \sum_{j \neq i} a^j X_j^a(t) + \sum_{k \neq i} a^k X_k(t) \right)^{M+1} - 1} \right]^A$$

with

$$A = \frac{\left(a^i X_i^a(t) + \sum_{j \neq i} a^j X_j^a(t) + \sum_{k \neq i} a^k X_k(t) \right) - 1}{\left(c^i a^i X_i^a(t) \right) - 1}, \quad j = 6, \dots, 10 \text{ and } k = 11, \dots, 15$$

Equation 68 represents the dynamics of the set of nodes activated by external signals (TF_6, \dots, TF_{10}). For the rest of the nodes, (TF_{11}, \dots, TF_{15}) the term $-\frac{dX_i^a(t)}{dt}$ is not included and the numerator in $F_i(x(t))$ depends on X and not X^a .

Equation 68 incorporates all possible combinations of multimers up to order M between all input species. The larger the multimer order, the larger the cooperativity between input species. Depending on the order M of multimers allowed to be formed, several regimes can be generated by combining both negative and positive links between transcription factors: multiple clustering attractors ($M < 6$), oscillations ($5 < M < 8$) and chaotic regimes ($M > 8$) (11). In the case of the highdimensional switch chosen for our work, $M=2$, only a high density of multiple stable states are observed.

Regarding the contribution of random fluctuations we assumed it to be multiplicative and proportional to $\sqrt{X_i}$. This multiplicative noise term is interpreted in the Stratonovich sense, which is the correct interpretation for a realistic source of noise showing small temporal autocorrelations (135).

An important feature of the model is the fact that only half of the transcription factors need to go through an activation reaction before being able to act on a downstream promoter region. This models generically the action of signalling molecules on Immediate Early Gene products (IEGs). Examples of the action of signalling molecules on IEGs, such as *c-jun*, *c-fos* and *c-myc* were

described in detail in section 1.2.5. The rest of the transcription factors operate even if no signal is present. They stand for delayed early genes products (DEGs), the second wave of transcription initiated by the signal (see Figure 46). Although this scenario is a simplistic approach to modelling the interface between the signalling module and the transcriptional machinery, it serves quite well our objective: observe and generalize the effects of parameter sweeping speed and transient external asymmetries in the process of highdimensional attractor selection.

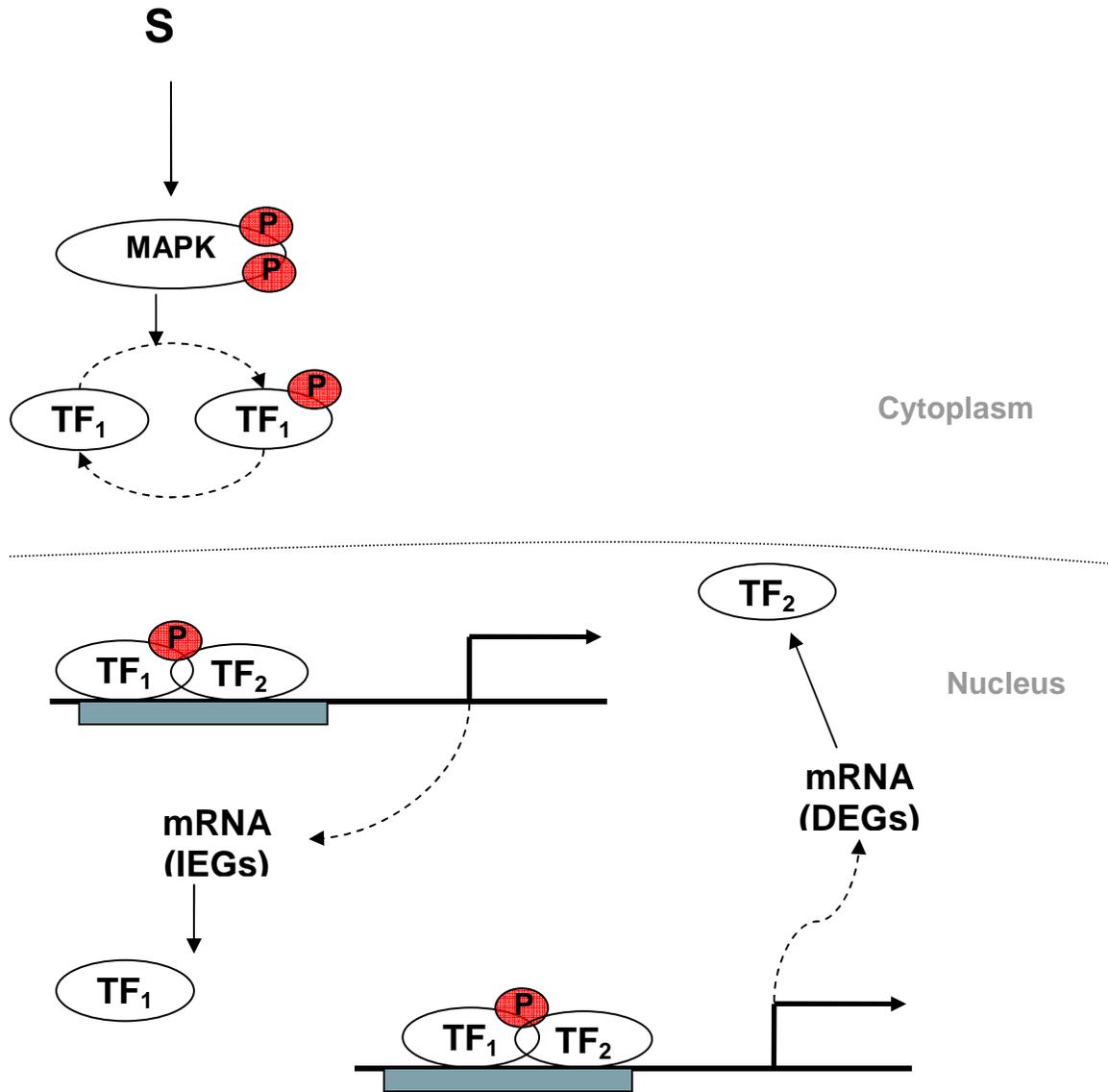


Figure 46 Generic representation of signal induction of immediate early gene products (IEGs) and subsequent delayed early gene products (DEGs). The signal **S** initiated at a receptor phosphorylates the output node of the signalling module MAPK or Erk which directly activates transcription factor 1 (TF₁). Examples of IEGs are the AP-1 components FOS and JUN. These are the gateway for genomic response and instigate a second wave. DEGs encode a variety of proteins such as transcriptional repressors, e.g. NAB2, RNA-binding proteins, e.g. ZFP36, and MAPK phosphatases.

The action of signals on 50% of the transcription factors will be assumed to follow mass-action kinetics (see Equation 69 and Figure 46 and section 1.4.1 for introductory notes).

Equation 69

$$\tau^S_i \frac{dX_i^a(t)}{dt} = \left[(X_i(t)) \times (sa + \sum_{j=1}^{k_{in}} S_j(t)) \right] - sd \times [X_i^a(t)],$$

$$X^a(t) = [X_i^a(t)] \text{ activated form , } i = 6, \dots, 10, \quad j = 1, \dots, 5$$

In Equation 69, $S_j(t)$ stands for the signal profile (see Figure 45) and sa and sd represent a basal level of activation and decay.

An extensive study of all sets of parameters and S_i -TF $_i$ connectivity matrices was performed. We selected the network that exhibited the highest number of attractors in order to generate maximum discrimination between combinations of inputs. The connectivity chosen between the set of signalling inputs (S_1, \dots, S_5) and the set of transcription factors activated is represented in Table 2.

	S₁	S₂	S₃	S₄	S₅
TF₆	0	1	1	1	1
TF₇	0	1	1	0	1
TF₈	1	0	0	0	0
TF₉	1	0	1	0	1
TF₁₀	0	0	1	1	0

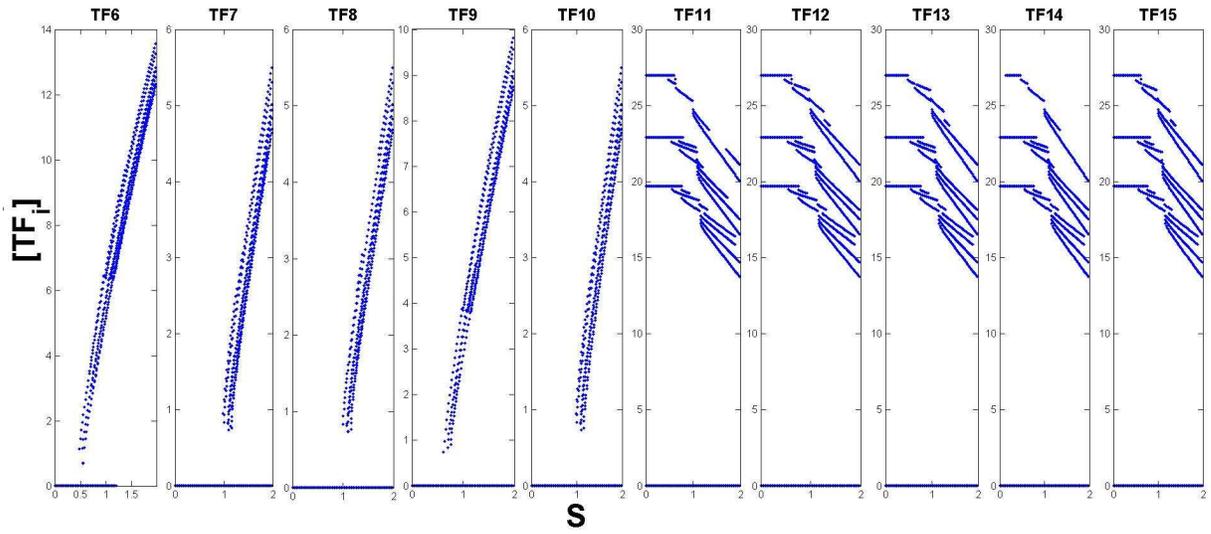
Table 2 Connectivity matrix between signalling inputs and transcription factors. See also Figure 45.

The inspection of the number attractors was performed numerically by generating bifurcation diagrams for each set of parameters (see Figure 47). Initial bifurcation diagrams were generated by assuming $S=S_1=S_2=S_3=S_4=S_5$ and investigating for each value of critical S the attractors emerging from initiating the system at 100 random initial conditions. In this chapter we present only the results for the selected network set of parameters and topology. This network will be maintained throughout the numerical simulations in this chapter.

Analyzing Figure 47 A it is possible to clearly verify the existence of multiple attractors for all network nodes. For the set activated by the external signals, only when the amplitude crosses a certain threshold ($S \sim 1$) do multiple

attractors above zero become clear. Actually, even before the amplitude reaches this point there's a very fine set of states very close to zero (see Figure 47 B). For the set of transcription factor nodes that do not directly interact with any S_i , the existence of multiple high concentration steady states is clear even for low values of signal amplitude. There is also a very fine set of attractors very close to zero. As the signal value is raised the nodes TF_6 to TF_{10} tend to show higher and higher steady state concentrations. Nevertheless, a set of low concentration steady states is still observed for all values of S and for all nodes with the exception of TF_6 . Regarding the nodes TF_{11} to TF_{15} , higher levels of S reduce the steady state concentration levels (Figure 47 A). The finer structure of steady states close to zero is also maintained for this set of nodes (Figure 47 B). The parameters chosen for this circuit were maintained throughout the chapter: $c_i^i=20$, $a_i^i=0.1$ (self-activation) and $a_j^i=0.1$ (cross-repression), for $i, j=6, \dots, 15$. The bifurcation diagrams show that for this set of parameters the system seems to go through a sub-critical type of bifurcation, due to the disconnection between emerging branches. Indeed, this set of parameters have shown to induce in 2 dimensional genetic switches a transition between a region of 1 stable state with low values (I_L) and another with 3 stable states with high concentration values (III_H) (109). Although the model in (109) was slightly different, if a similar process is present in our circuit then the disconnection is indeed caused by a sub-critical type of bifurcation. On the other hand, the disconnection may arise due to limited number of simulations sampling the distribution of attractors in phase space, and indeed the type of bifurcation present is supercritical. Further investigations are necessary to clarify this point.

A



B

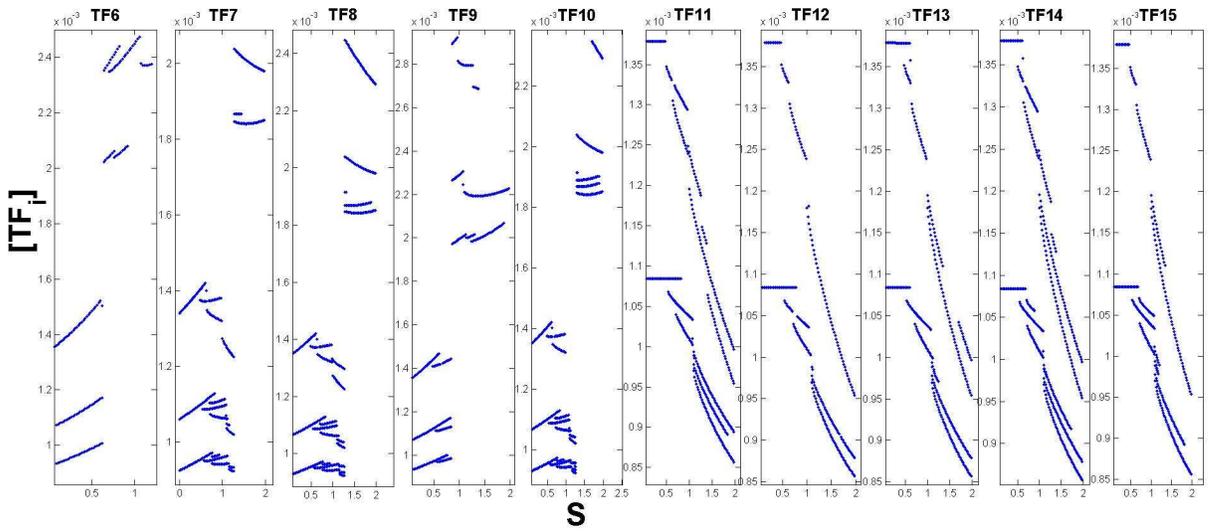


Figure 47 Bifurcation diagram for each of the transcription factors for $S=S_1=S_2=S_3=S_4=S_5$. **A)** Complete bifurcation diagram. **B)** Amplification of lower part of the bifurcation diagram represented in A). Parameters: $M=2$, $\eta_i=0.1$, $c_i=20$, $a_i=1$ (self-activation) and $a_j^i=0.1$ (cross-repression), $sa=0$, $sd=0.3$, $\tau_i^S=\tau_i^T=0.001$, for $i,j=6,\dots,15$ (see Equation 68 and Equation 69). S is the horizontal axis for all the figures, from TF_6 to TF_{15} .

For the time-dependent signals studied ahead, the asymmetries $\Delta S_{ikim}(t)$ between each input influence the available attractors in the system at each time step. Further ahead we will focus on 3 specific input combinations. Their

bifurcation diagrams show small differences in the distribution of attractors (compare Figure 90, Figure 91 and Figure 92)

3.2 Clustering of input signal combinations by the highdimensional decision switch

In order to understand if differences in time-dependent input signal S_i profiles induce different network attractors, we tested the response of the decision switch to a batch of 100 combinations of inputs $I_k=(S_1,\dots,S_5)_k$ generated by randomly selecting T_{S_i} 's for each input S_i (see Figure 45). The maximum amplitude allowed for each signal S_i was 2. This restriction was chosen by extensively testing several input amplitudes and selecting the one that allowed for the highest number of attractors.

For each I_k combination the system was initiated at 100 initial conditions randomly, with $X_i(0)\in [0, \eta_i]$ (see Equation 68). The asymptotic stable states were recorded for each of the combinations I_k and each of the initial conditions. The set of initial conditions was exactly the same for each I_k . In order to understand if there are substantial differences in the set of attractors induced by each combination, the average euclidean distance (AED, see Equation 70) between the set of concentrations in the limit of large times, when the system is at an asymptotically stable-state (sst), was compared for all possible pairs $(I_k, I_{k'})$ and averaged over the number of initial conditions tested.

Equation 70

$$AED_{(I_k, I_{k'})} = \frac{1}{N_{ic}} \sum_{m=1}^{N_{ic}} \sum_{l=1}^{10} (X^{I_k}_l - X^{I_{k'}}_l)^2_{sst_k},$$

with $N_{ic} = 100$ (number of initial conditions)

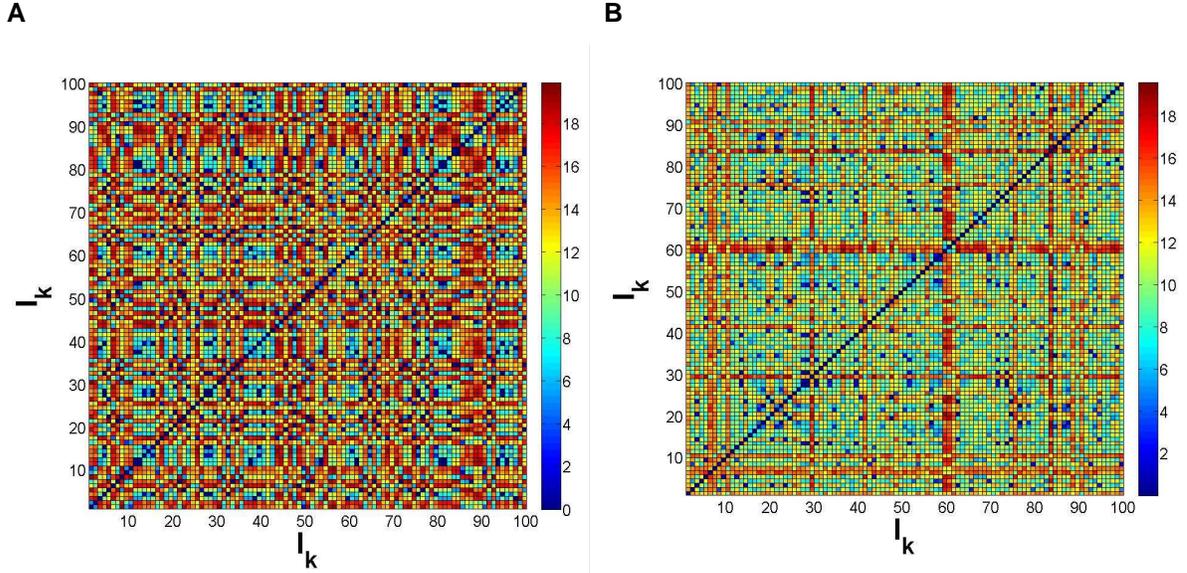


Figure 48 Pair-wise average distance over 100 runs (each corresponding to a different initial condition) between asymptotically stable states induced by input combinations.

A) Results for time-scale ratio $\frac{\tau^T}{\tau^S} = 1$. **B)** Results for time-scale ratio $\frac{\tau^T}{\tau^S} = 5$. τ^S_i is equal for all nodes, as well as τ^T_i . See Equation 67. Parameters: $M=2, \eta_i=0.1, c_i^i=20, a_i^i=1$ (self-activation) and $a_j^i=0.1$ (repression), $sa=0, sd=0.3$ (see Equation 68 and Equation 69), for $i, j=6, \dots, 15$.

In Figure 48 A, the results obtained from the application of the average distance metric AED (Equation 70) can be visualized for two time-scale ratios $\frac{\tau^T}{\tau^S}$, with $\tau^T_i = \tau^T$ and $\tau^S_i = \tau^S$, for $i=6, \dots, 15$. Because the matrices are symmetric we need only to observe values below the diagonal. In both matrices one can verify that certain combinations I_k induce substantially different end attractors (red pixels, higher AED distance). Others, for the same initial conditions selected exactly the same attractors, on average (blue pixels, lower AED). This indicates that certain combinations I_k combinations of signals S_i are clustered together due to the incapacity of the network to distinguish them.

Observing Figure 48 B we see that the AED distance for each pair of input combinations is decreased if the time-scale difference between phosphorylation and transcriptional processes is raised. This effect had been seen already in the decision genetic switch; for larger time-scales differences, trajectories stay near

the unstable branch for longer periods before the switching to one of the attractors occurs (Figure 42 B). In real biological systems the differences in time-scale magnitude can be substantial (see Table 1). If genetic circuits are not sensitive to slight differences between driving external signals when time-scale separation is significant, then integration of signals is only successful when very pronounced asymmetries occur. In the limit, only considerable differences in amplitude held for an interval compared to the characteristic relaxation time-scale of the system will be discriminated efficiently.

3.3 Attractor selectivity in the presence of multiplicative noise

In order to prove the existence of path-dependent effects in attractor selection in the presence of fluctuations, we first analyzed the inter-trajectory distance for every pair $(I_k, I_{k'})$ generating the same end attractors (see Figure 48, dark blue squares) in a deterministic simulation. For this calculation we used a correlation (r in Equation 72) based distance metric (Equation 71 and Equation 72).

Equation 71
$$ITD_{(I_k, I_{k'})}(t) = 1 - r_{(I_k, I_{k'})}(t)$$

Equation 72

$$r_{(I_k, I_{k'})}(t) = \frac{\sum_i \left(x_i^{I_k}(t) - \overline{x^{I_k}(t)} \right) \left(x_i^{I_{k'}}(t) - \overline{x^{I_{k'}}(t)} \right)}{(n-1) s_{x^{I_k}(t)} s_{x^{I_{k'}}(t)}}$$

with

$\overline{x^{I_k}(t)}$, $\overline{x^{I_{k'}}(t)}$ **being the sample mean**

and

$s_{x^{I_k}(t)}$, $s_{x^{I_{k'}}(t)}$ **the sample standard deviation**

In Figure 49 the results for each of the pairs extracted from Figure 48, with zero AED (see Equation 70), can be observed as a function of time. There are a number of ITD(t) profile shapes. Some pairs have maximum trajectory

differences for short times and return immediately to zero. Others show more distributed differences. We will focus on only one characteristic of the inter-trajectory distance generated by $(I_k, I_{k'})$: the maximum value, $\max[ITD(t)_{t \in [t_0, t_{final}]}]$. Several other simulations will be reported elsewhere comparing the influence of other characteristics of the inter-trajectory distance profile such as the duration and average ITD over time. This should reveal extra information on differences between distributions over attractors:

- Influence of the instant the maximum $\max[ITD(t)_{t \in [t_0, t_{final}]}]$ is achieved
- Influence of the time the trajectories in phase space are different

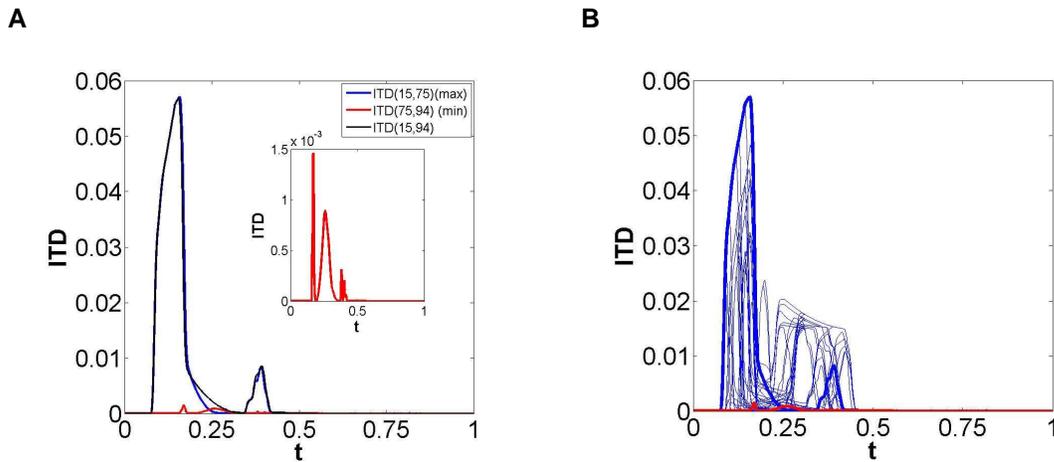


Figure 49 Inter-trajectory distance for pairs $(I_k, I_{k'})$ inducing the same attractors. **A)** Pairs exhibiting the highest value for $\max(ITD(t))$ (Equation 71) and the lowest value for $\max(ITD(t))$. **B)** All pairs $(I_k, I_{k'})$ with zero AED (Equation 70). See also Figure 48.

The pair $(I_k, I_{k'})$ with input combinations inducing the same end attractors that had, at a particular instant, the highest maximum for the inter-trajectory distance ITD (Equation 71) amongst all the pairs was (I_{15}, I_{75}) . On the other hand, the pair exhibiting the smallest maximum was (I_{75}, I_{94}) (see Figure 49). The time-dependent profile for I_{15}, I_{75} , and I_{94} can be visualized in Figure 50 A. A typical trajectory in time can also be observed in Figure 50 B. The trajectory presented corresponds to the evolution of the system by applying I_{15} . Usually, the trajectories converge very rapidly to high or low concentration values (see Figure 50 A). Subsequently, for nodes migrating to low concentration values there a

further reorganization of states. In the vicinity of the instant when all S_i 's have reached their maximum amplitude there's once more reorganization of states with 2 nodes reaching intermediate concentration values between 4 and 10 (see Figure 50 B right).

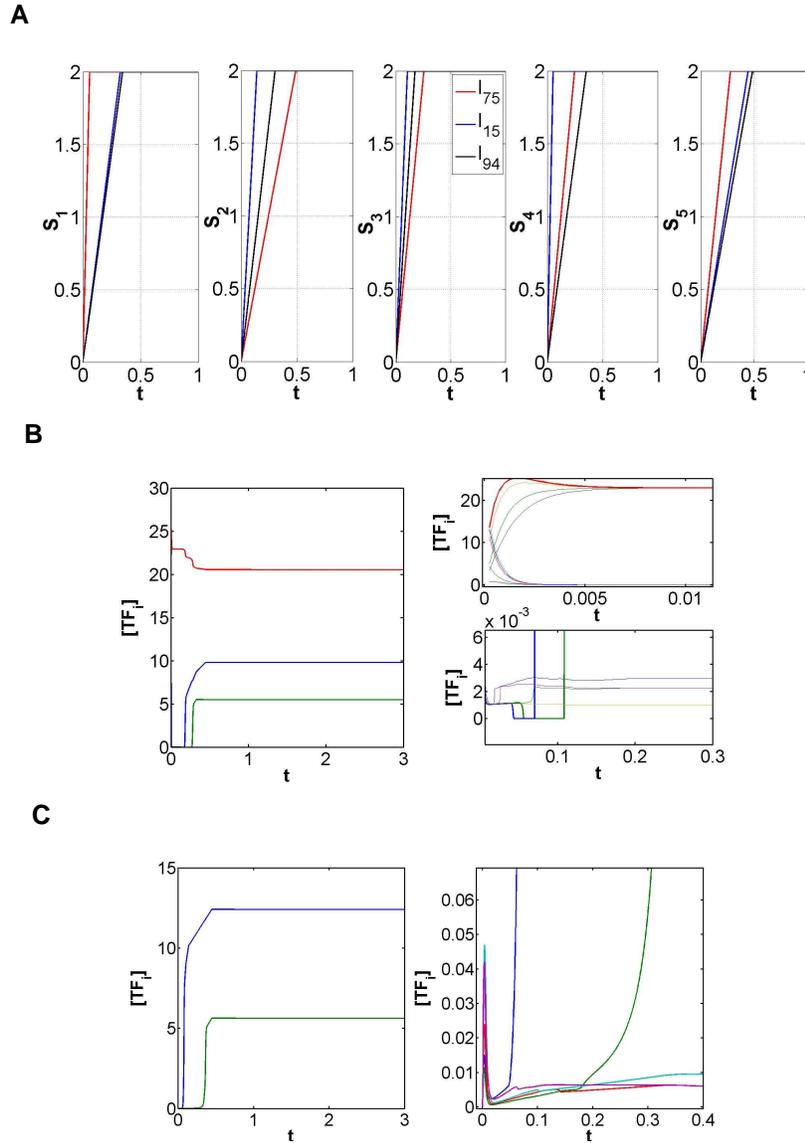


Figure 50 Time-dependent evolution of the highdimensional decision switch circuit. A) Time-dependent profile for each input S_i for 3 input combinations I_{15} , I_{75} , I_{94} . B) Typical evolution of concentrations for all the nodes TF_i , $i=6, \dots, 15$. This particular trajectory was generated by applying I_{15} . C) Typical evolution of concentrations for activated nodes TF_6 – TF_{10} . This particular trajectory was generated by applying I_{15} . Parameters: $M=2$, $\eta_i=0.1$, $c_i=20$, $a_i=1$ (self-activation) and $a_j=0.1$ (repression), $sa=0$, $sd=0.3$, $\tau_i^S=\tau_i^T=0.001$ (see Equation 68 and Equation 69), for $i,j=6, \dots, 15$.

The attractors with the highest probability, selected by the 3 input combinations (I_{15}, I_{75}, I_{94}), in a deterministic scenario, can be seen in Figure 51. Remember that all input combinations have the same amplitude of each S_i at the initial and final instant of the sweeping process.

One can observe that TF_6 to TF_{10} show propensity to converge to attractors with intermediate and high concentrations (see also the bifurcation diagrams in Figure 90, Figure 91 and Figure 92). TF_7, TF_8 and TF_{10} also show some probability to reach attractors close to zero when the external inputs S_i are at their maximum value. These attractors are very close to each other. For nodes corresponding to the DEG layer (not activated by signals) higher selectivity frequencies are registered for attractors with higher concentrations. However, there are still trajectories with asymptotic states near zero (Figure 51).

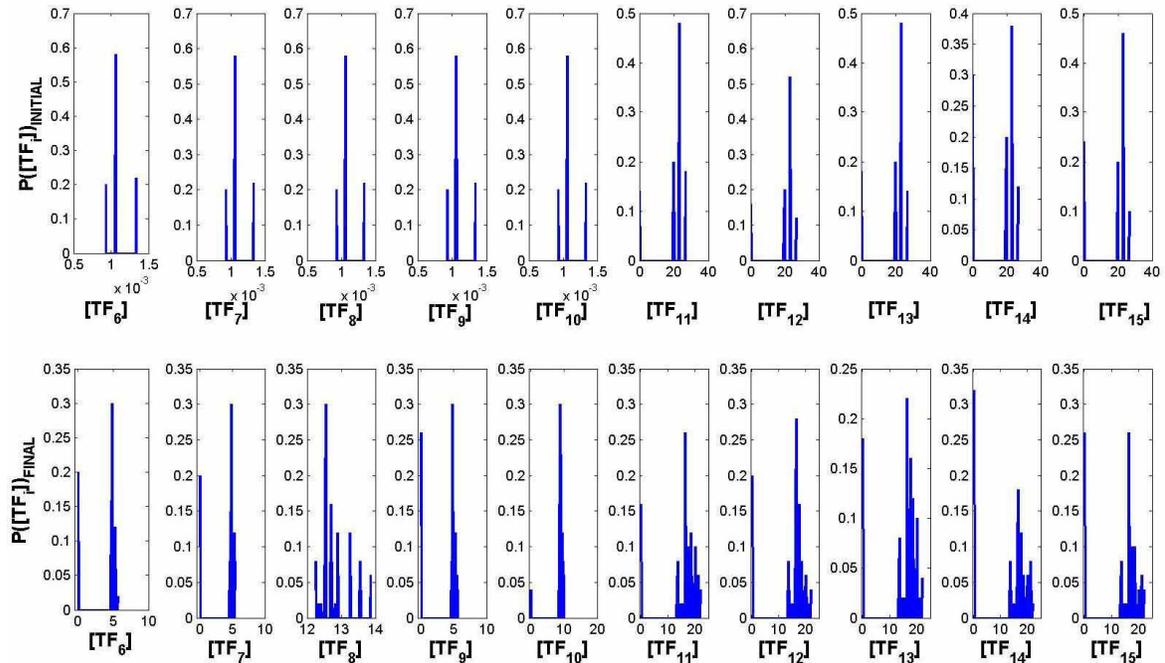


Figure 51Initial and final attractors' positions induced by I_{15} and respective frequency. Since I_{15}, I_{75} and I_{94} have initial and final zero ITD, the position and frequency of attractors represented here is equal for the 3 input combinations. Parameters: $M=2, \eta_i=0.1, c_i^i=20, a_i^i=1$ (self-activation) and $a_j^i=0.1$ (repression), $sa=0, sd=0.3, \tau_i^S=\tau_i^T=0.001$ (see Equation 68 and Equation 69), for $i, j=6, \dots, 15$.

These 3 input combinations were once again applied to the circuit but in the presence of fluctuations (see below, section 3.3.1). An example of a trajectory in the presence of noise is shown in Figure 52. Noise tends to blur the low concentration steady states.

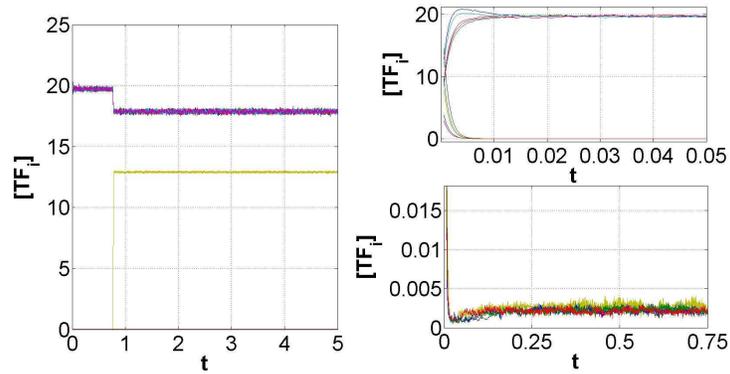
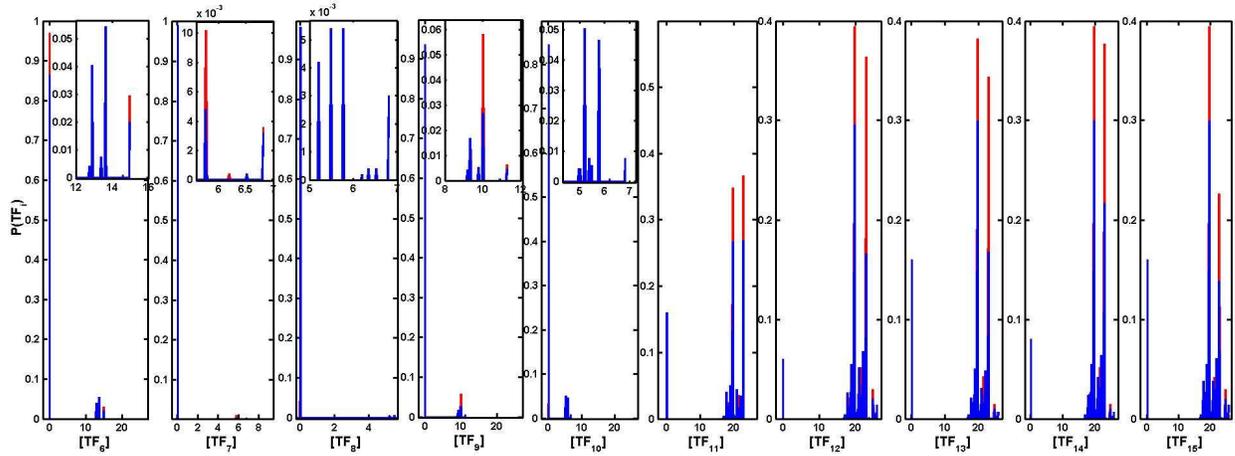


Figure 52 Typical evolution of concentrations for all the nodes TF_i . This particular trajectory was generated by applying I_{15} with noise intensity 0,05. Parameters: $M=2$, $\eta_i=0.1$, $c_i=20$, $a_i=1$ (self-activation) and $a_j=0.1$ (repression), $sa=0$, $sd=0.3$, $\tau^S_i=\tau^T_i=0.001$ (see Equation 68 and Equation 69), for $i, j=6, \dots, 15$.

3.3.1 Inter-distribution distance depends on noise

A



B

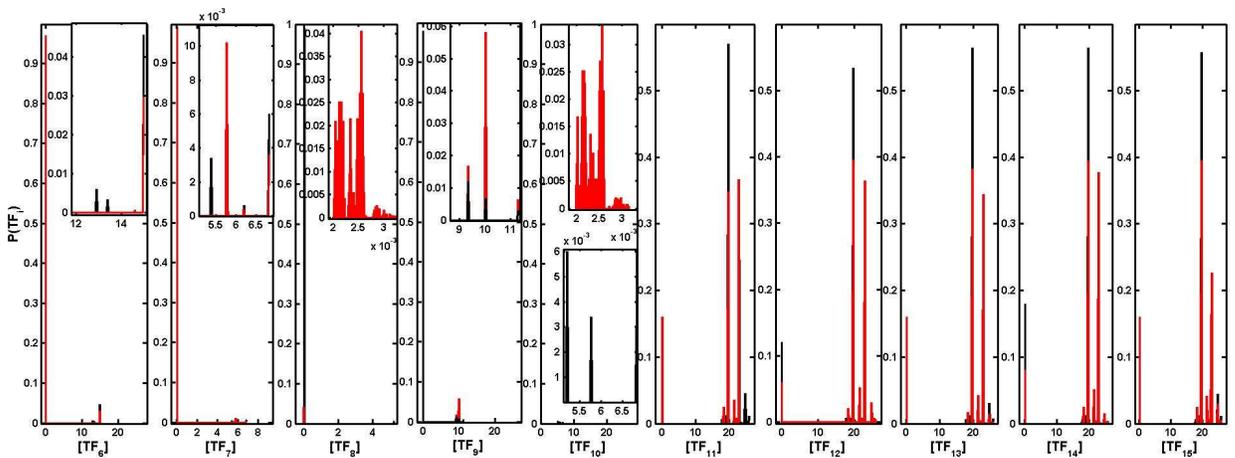


Figure 53 Distributions across attractors in the presence of noise intensity $\sigma=0.05$. **A)** I_{15} (blue) and I_{75} (red). **B)** I_{75} (red) and I_{94} (black). Parameters: $M=2$, $\eta_i=0.1$, $c_i=20$, $a_i=1$ (self-activation) and $a_j=0.1$ (repression), $sa=0$, $sd=0.3$, $\tau_i^S=\tau_i^T=0.001$ (see Equation 68 and Equation 69), for $i, j=6, \dots, 15$. Histograms for each input combination generated from 5000 trajectories.

The final distributions across attractors generated by applying each of the input combination of signals highlighted in the previous section, but in the presence of noise ($\sigma=0.05$), can be viewed in Figure 53 (see also Figure 93 and Figure 94 for final distributions obtained with $\sigma=0.01$ and $\sigma=0.5$, respectively). Overall, the data from 5000 trajectories was collected including several starting points in phase space. There are observable changes in the attractors selected when fluctuations

are considered. For the IEG layer of transcription factors (TF_6 - TF_{10}) there is a considerable transfer of probability mass to states located near zero. These were not identified as being very probable in the deterministic scenario (see Figure 51). The addition of noise leads the system to jump across potential barriers to stronger attractors which, in this case, are closer to zero. As was seen in the one dimensional canonical model (section 2.2.4), according to Kramer's classical theory (see Equation 22) the transition time for a system in one dimension to jump across the potential barrier is decreased with noise intensity. Yet, in higher dimensions, and especially in our system, the distribution of attractor in phase space and their basin of attraction structure is quite complex (see Figure 90, Figure 91 and Figure 92). There are several aspects to the attractor selection process occurring here:

- First, let us recall the probability distribution shown in Figure 51. These results are dependent only on differences in attractor basins and number of initial conditions tested since they were obtained in a deterministic simulation. The basin of attraction in dynamical system theory is taken as the percentage of points converging onto a specific attractor. 100 initial points chosen randomly may not have probed completely the phase space. Higher sampling could reveal finer aspects of attractor basins.
- A second aspect of the selection process may arise as a function of the fact that different external signals are exerting different changes on the attractor landscape. If the probability mass transfer to attractors located near zero was mainly a consequence of the combination of input signals, then the differences observed in Figure 53 should be more evident. The only clear differences recorded have very low probabilities. Also, we had seen before that the bifurcation diagrams also do not show a striking difference when each of the input combinations is applied (see Figure 90, Figure 91 and Figure 92). We can conclude from these observations that the high frequency found for low concentration values for the set of nodes TF_6 to TF_{10} is closely related to the concept of strength of an attractor in

the face of perturbations. This concept is defined as the minimum size of a perturbation that results in very low probability of return (143). The attractors close to zero are probably stronger. Nevertheless, they seem to have a smaller basin of attraction which led to lower frequencies in the deterministic case.

Regarding the attractors found for the DEG layer of transcription factor nodes, the distribution does not differ considerably in terms of location from that generated in the deterministic scenario.

The differences between applying each pair of combinations, (I_{15}, I_{75}) or (I_{75}, I_{94}) , which had highest maximum ITD amplitude and the lowest maximum ITD amplitude, respectively, do not show striking differences in Figure 53 . Mostly, the differences in frequency occur in the same set of attractors at high concentration values. Applying one or another input combination shifts the probability maximum to an attractor in the vicinity. Given the fact that the pair (I_{15}, I_{75}) generated an inter-trajectory distance that reached a maximum higher than the pair (I_{75}, I_{94}) , it would be expected that the differences between the respective final distributions would reflect this fact.

We also evaluated the distance (see Equation 72) between distributions for several noise intensities (Figure 54) to understand if, as in the small integrated signalling-gene regulatory decision switch (see section 2.3), noise increases symmetry between distribution across attractors or if it causes new attractors to be populated according to the changes exerted by each I_k .

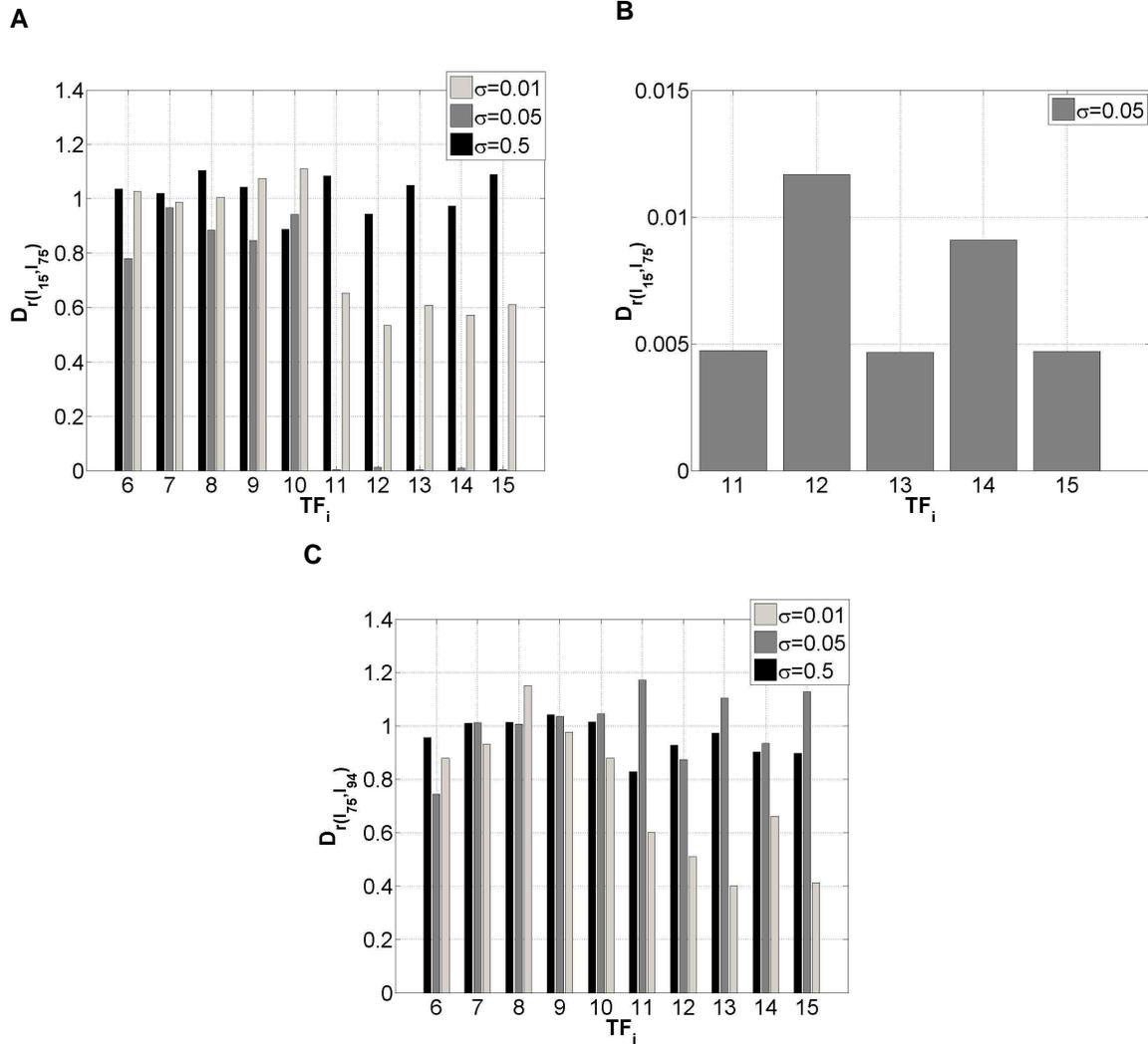


Figure 54 Distance between final distributions generated by different pairs of input combinations (l_k, l_k') in the presence of fluctuations. A) Pair (l_{15}, l_{75}). B) Pair (l_{15}, l_{75}), focus on TF_{11} to TF_{15} . C) Pair (l_{75}, l_{94}). Parameters: $M=2$, $\eta_i=0.1$, $c_i=20$, $a_i=1$ (self-activation) and $a_j=0.1$ (repression), $sa=0$, $sd=0.3$, $\tau^S_i=\tau^T_i=0.001$ (see Equation 68 and Equation 69), for $i, j=6, \dots, 15$. See also Figure 53, Figure 93 and Figure 94 for distribution histograms. σ stands for noise intensity.

The distance metric $D_{r(l_k, l_k')}$ (see Figure 54) applied in this task is a correlation based metric similar to Equation 71 and Equation 72.

For the pair (l_{15}, l_{75}), the most noticeable fact when we raise noise intensity from 0.01 to 0.05, is the relative proximity of the distributions for the DEG node layer (Figure 54 A and B). The 5 fold increment seems to force the system to jump to the strongest attractors. Effectively, comparing by visual inspection the distribution obtained with noise intensity 0.01 and 0.05, Figure 93 (section 8.8.1)

and Figure 53 respectively, we verify that for noise 0.05 essentially the maximum frequencies occur at the same attractors. For IEG nodes (TF_6 to TF_{10}) the same observation stands although it is not as evident (Figure 54 A). Further rising of noise intensity increases the distance between final distributions, which would be expected due to the increased capacity to cross potential barriers and as a result populate different attractors, or just as a consequence of the dispersion induced by noise (Figure 54 A).

For the pair of input combinations (I_{75}, I_{94}) that, as was calculated before, had a very small difference between the trajectories in phase space (Figure 49), the tendency observed for the distance calculated between distributions when noise intensity is increased from 0.01 to 0.5 is similar (also compare histograms represented in Figure 93 and Figure 94). Also, for these noise intensities $D_{r(I_{15}, I_{75})}$ is higher than $D_{r(I_{75}, I_{94})}$, which is consistent with the fact that $\max(\text{ITD}(I_{15}/I_{75})) > \max(\text{ITD}(I_{75}, I_{94}))$ (see Figure 49). Nevertheless, for noise amplitude equal to 0.05 the tendency observed for (I_{15}, I_{75}) is not maintained. At this noise intensity, instead of an optimal attractor selection that approximates the distributions, the opposite effect is present.

The numerical results reported above indicate that, for the circuit chosen for this chapter, there is an optimal intensity of noise that increases the convergence of trajectories to the same attractors, when the differences between trajectories induced by each I_k is large. When the differences in phase space trajectory are small the noise optimality effect observed before reverses its role and increases inter-distribution distance.

3.3.2 Inter-distribution distance depends on sweeping speed

To test the ideas developed in other chapters (see section 2.2.5 and 2.3.2) related to attractor selectivity as a function of bifurcation parameter sweeping rate we extended the simulation experiments for noise intensity $\sigma=0.5$. We did not perform experiments for the other noise intensities due to time restrictions. On the other hand, we chose to perform the extra simulations with the maximum

noise intensity to understand if the sweeping speed could override the strong effects of noise.

The original selected combinations, I_{15} , I_{75} , I_{94} , were changed such that the maximum asymmetry between each of the inputs was maintained but the sweeping speed was decreased by the following algorithm:

1. For input S_1 of the original combination calculate the maximum asymmetry reached between S_1 and S_i recurring to Equation 73;
2. Increase T_{S_1} by n steps and calculate the necessary T_{S_i} (Equation 73) for each of the inputs that maintain the maximum asymmetries a between signals (recall Figure 39).

Equation 73

$$\left(1 - \frac{a}{2}\right) = \frac{T_{S_1}}{T_{S_i}}$$

$$a = \max_{t \in [0, t_{final}]} (\Delta S(t))$$

This strategy secures that the signals induce similar changes in the transcriptional landscape as the original combinations, but at a smaller speed. The distance between the final distributions was calculated again by applying a correlation based distance metric to 3 extra cases: same input combinations but 100, 300 and 500 time-steps slower. The results are shown in Figure 55. In light of the results obtained for the small genetic decision switch we expected that the differences between final distributions across attractors induced by each pair (I_k, I_k') would be increased if the speed with which the signals S_i are changed is reduced. Figure 55 A shows that, overall, the path-dependent effects registered before for the pair of input combinations (I_{15}, I_{75}) are less clear if we perform the sweeping process at lower rates. Nevertheless, decreasing the sweeping speed through the bifurcation region (by imposing for example S_i 's 500 time-steps slower) seems to have, for most of the transcription factors, an effect which brings the distributions induced by I_{15} and I_{75} closer together. For the other sweeping speed experiments (Figure 55 A, 100 and 300 steps slower) there

seems to be a tendency for the pair (I_{15}, I_{75}) to induce closer and closer final distributions as we decrease the sweeping speed. Yet, this occurs in a non-monotonous fashion. This observation contrasts with the findings of speed-dependent decision making in the bistable decision genetic switch (see section 2.3) where slower sweeping rates increased the sensitivity to external asymmetries. Hence, the differences in the final distributions arising from the respective paths in phase space should have been more pronounced. On the other hand, we do observe reasonably clear speed-dependent effects for the highdimensional switch. Further simulation studies (for $\sigma=0.01$ and $\sigma=0.05$) are necessary to clarify the synergistic effects of sweeping speed and noise intensity in highdimensional phase space with less regular attractor landscapes.

Regarding the other input combination pair, (I_{75}, I_{94}) (see Figure 55 B), a considerable reduction in sweeping speed (500 time steps slower) induces exactly the opposite effect observed for (I_{15}, I_{75}) . This tendency to observe opposite effects in the input combination pairs used throughout this chapter is quite intriguing and should be investigated with the complete set of pairs $(I_k, I_{k'})$ with same end attractors (see Figure 49). Overall, we observe that slower sweeping speeds induce a higher sensitivity of the highdimensional circuit to external signals when the differences between the respective paths in highdimensional phase space, induced by each pair $(I_k, I_{k'})$, are smaller.

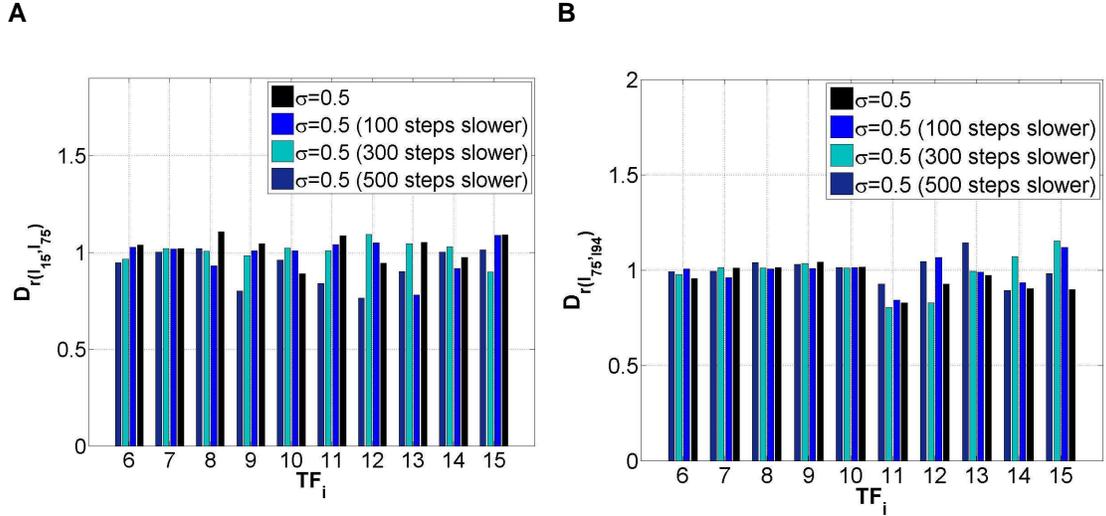


Figure 55 Inter-distribution distance dependence on sweeping speed. A) Inter-distribution distance between the attractors induced by combination I_{15} and I_{76} . B) Inter-distribution distance between the attractors induced by combination I_{75} and I_{94} . D_r stands for the distance metric based of the correlation between distributions (similar to Equation 71 and Equation 72). Parameters: $M=2$, $\eta_i=0.1$, $c_i=20$, $a_i=1$ (self-activation) and $a_j=0.1$ (repression), $sa=0$, $sd=0.3$, $\tau^S_i=\tau^T_i=0.001$ (see Equation 68 and Equation 69), for $i, j=6, \dots, 15$.

3.4 Summary

The generalization of the parameter sweeping mechanism to high-dimensional space developed in this chapter proved that it is strongly dependent on phase space structure and on the efficiency of noise to induce transitions across potential barriers. The capacity of the highdimensional genetic circuits to integrate a combination of complex signals is closely linked to the initial condition chosen. The effect of sweeping speed is considerable more complex due to the irregular distribution of attractors. Nevertheless, we have shown that even in situations that the deterministic system does not discriminate between external signals, the effect of noise may render the differences between signals more pronounced. Although not as evidently we have also shown that slower sweeping speeds are the most effective in enhancing the capacity of highdimensional genetic circuits to distinguish between signals, in the presence of fluctuations, when the original external signals result is similar paths in phase space.

4 Non-specific effects of small molecule kinase inhibitor therapies: insights into synergistic effects of multiple targeting

4.1 Introduction

In many cancer treatments, multi-component combinations have become the main strategy (see section 1.3.3). A problematic aspect that may arise in combinatorial approaches stems from the additional targets, also known as non-specific interactions, which may or may not be relevant in a particular tumour molecular profile. Indeed, several studies of binding affinity against panels of kinases (79) have revealed that a wealth of additional targets are inhibited by currently used small molecule kinase inhibitors. This uncertainty increases the probability for toxicity. Alternatively, complete disruptions of a network by a set of specific drugs may not be the most successful strategy. Studies of large networks (3) have shown that multiple weakly modulated nodes have a higher impact on network efficiency, which is a “*global measure of network integrity related to the shortest path length between each pair of elements within the network*” (3). This has motivated therapeutical approaches looking for synergistic effects of drug combinations with reduced toxicity (169). In this chapter we will study the effects of non-specific interactions of small molecule kinase inhibitor therapies on a fully parameterized model of the paradigmatic Epidermal Growth Factor Receptor pathway (EGFR). This is an attempt to understand if, indeed, multitargeting is absorbed by the network or if it has additional effects that can be used to clarify how one should target a network. All small molecules will be tested independently. Hence, the high-order targeting nature of the investigations will not come from understanding the action of multiple drugs but by clarifying the effect of multiple weakly bound targets by one agent, in a data oriented fashion. Recall also the discussion of section 1.3.3.

Additionally, we will devise modelling approaches to study the problem of multitargeting in an extended version of the Human Signalling Network (64) (see section 4.3.2).

4.2 Targeting the Systems Biology paradigmatic Epidermal Growth Factor Receptor Pathway with documented small molecule kinase inhibitor therapies

This section will be dedicated to identifying the effects of additional off-target interactions characteristic of small molecule kinase therapies. We will focus on the biological impact of each kinase inhibited and its possible connection to the EGFR. Additionally, correlation with experimental findings reported in the literature will also help estimate the impact of non-specific interactions in the pathway output behaviour.

The EGFR pathway has been considered a fundamental example in the development and application of mathematical tools in the area of Systems Biology. The wealth of data on this network is sufficiently vast to allow for several lines of research to take place: proving the existence of structural and functional (dynamical) modules, testing the importance of concepts from control theory such as feedback (269), and developing easily manageable modelling tools such as the stochastic π -calculus (211, 215). The introductory discussion of this section will be very useful to understand how documented non-specific interactions exerted by targeted cancer therapies contribute to cell fate decision seen from the traditional perspective of the concentration profile of a single species.

4.2.1 EGFR/MAPK pathway dynamics: focus on Schoeberl et al (2002) model.

Schoeberl and coworkers model (230) reconstructs a complete signalling cascade by mathematical description of the EGF receptor induced Mitogen Activated Protein (MAP) kinase pathway, including receptor internalization. It is

one the largest kinetic models based on ODEs developed so far. An extension to this model has been published recently that includes reactions involving all possible members of the ERBB receptor family (51). In 2003, Resat and coworkers (220) developed one of the largest models of the EGFR cellular network but using a probability weighted-dynamic Monte Carlo stochastic simulation. Their model is an integration of both trafficking and signalling aspects of the EGFR system. It encompasses hundreds of distinct endocytic compartments and considers around 13000 reactions over a considerable spatio-temporal range. EGFR signalling sequence is well known but the kinetic behaviour and critical signalling events that are responsible for cell phenotype such as cell growth, survival, or differentiation is still to be clarified.

The components involved in Schoeberl's model (230) and its connection are represented in Figure 57. The model calculates the concentration of 94 compounds after EGF stimulation by computing the changes according to a mass action law type of approach (Equation 74). It also includes the role of receptor internalization. It is assumed that receptor at the cell surface and internalized receptors in endosomal compartments bring about identical signalling downstream cascades. One of the extension to the Schoeberl and coworkers' model will be used throughout this chapter (125). It contains 148 cellular molecular processes and 103 molecular species. It is an augmented version of previous work (150, 230). Here we will focus on the EGFR or Erbb1 pathway and study the effect of drugs targeting this receptor (section 4.2.2).

Equation 74

$$\frac{d[C_i]}{dt} = \sum v_p - \sum v_c$$

In Equation 74 v_p stands for the reaction rate of production and v_c for reaction rate of consumption of particular species included in the model. In this model the cell was idealized as a sphere with diameter of 15 μ m resulting in a volume of 1 $\times 10^{-12}$ L. The estimated radius of the endosome is 100nm giving rise to a volume of 4.2 $\times 10^{-18}$ L (151). The cell line used was HeLa (230). Understanding what parts

of the pathway exert a stronger influence on the activation of ERK is an important step towards the dissection of the network into its functional parts. Regarding cell fate decision the response evoked by the MAPK pathway is crucial. Magnitude and duration of the signal (transient/proliferation versus sustained/differentiation) of ERK activation are fundamental (58, 184). A characteristic concentration profile for some of the pathways components is represented in Figure 56.

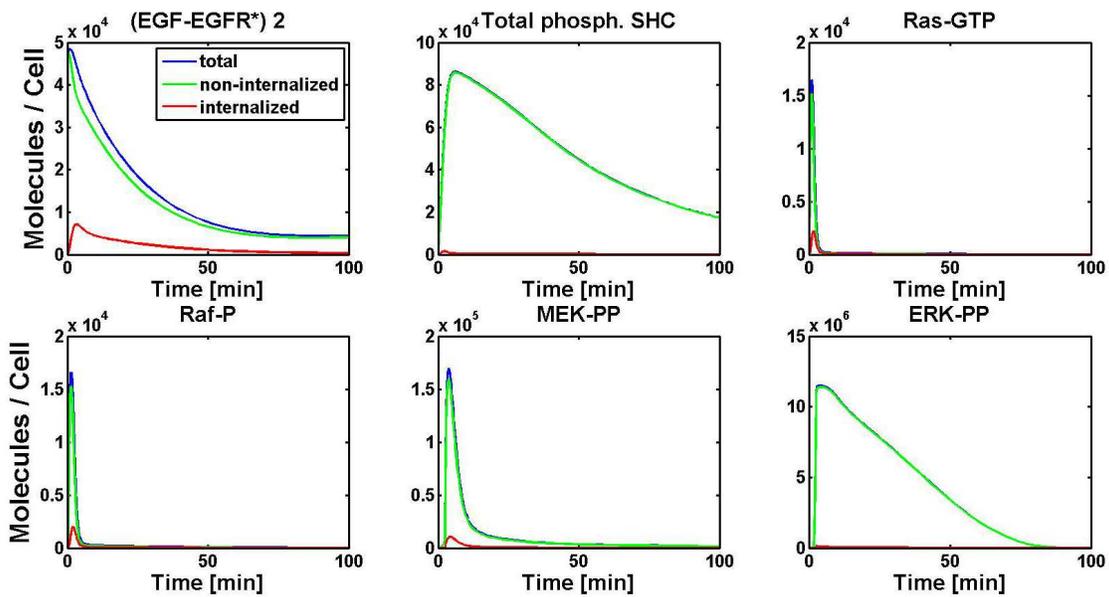


Figure 56 Concentration profiles for EGFR pathway. EGF concentration-50ng/ml. EGFR-50000. The rest of the initial conditions and parameters as in (230). For a constant ligand concentration the response transiently to the stimulus.

The results obtained by Schoeberl and coworkers (230) give us further insight into the mechanisms of signal transduction by the EGFR pathway. During signal transduction EGF concentration and binding kinetics affect the velocity of the EGFR activation and therefore the flux of information from the outside to the inside of the cell. EGFR autophosphorylation suffers a decrease in its initial velocity activation with decreasing EGF concentration. The impact of lower ligand concentrations is also seen in Shc phosphorylation although not as strong as with EGFR. Ras-GTP activation is also dependent on ligand concentration; it shows always lower peaks in amplitude and earlier in time with decreasing [EGF].

Regarding the MAPK cascade the tendency is the same. Nevertheless, a crucial observation is that 70% of ERK's maximum amplitude activation is still attained at lower values of ligand concentration. Therefore, high sensitivity is observed over a relatively broad EGF concentration range. This particular outcome is a result of the MAPK cascade structure which has the capacity for ultra-sensitive behaviour and amplification (228).

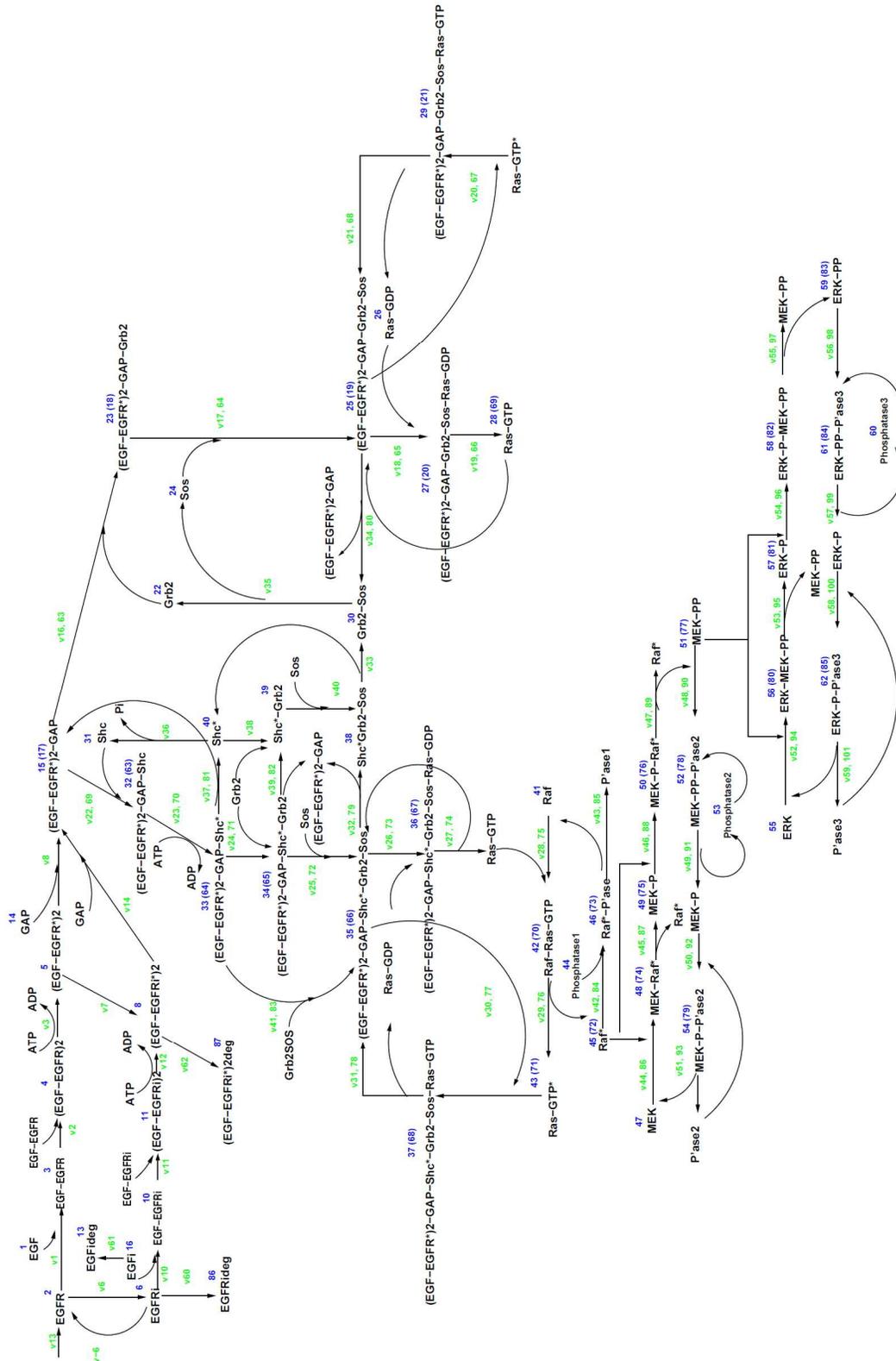


Figure 57 Diagram of the molecular species involved in Schoeberl et al model (230). Kinetic parameters are indicated in green. For numerical values reference (230) should be consulted. Each molecular species has a number in blue. For those with two numbers, the second is associated with internalized processes.

One important result of Schoeberl and coworkers' paper (230) is the fact that attaining the maxima is irrelevant for signal transfer to the next protein in the signal cascade, since the maximal amplitude of activation of downstream proteins is reached before the maxima of the preceding protein. The relevant parameter is EGF-EGFR affinity. As it decreases the initial velocity of EGFR activation also decreases and ERK activation suffers a delay. This is a similar effect to that observed for decreasing EGF concentrations.

Regarding receptor internalization represented in Figure 57, it was concluded that it protects the cell by attenuating signalling at high EGF concentration. Nevertheless, it amplifies the signal after internalization at low EGF concentrations, and probably protects the cell from apoptosis resulting from insufficient growth stimulation. This conclusion has to be taken with due moderation given that interaction between internalized and external receptors may be different for different pathways (42). The contribution of the internalized part of the model to total ERK-PP concentration is at most 1.2%.

The test performed by Schoeberl and coworkers (230) involving EGF receptor number and signal duration showed that although maximum amplitude is maintained when EGFR numbers increase the duration of ERK rises (see Figure 58). This could be related to the process of internalization given the fact that with high numbers of EGFR the machinery of internalization is saturated and its effect on modulation (endocytosis) of the signal along the cascade is not felt (230). Schoeberl's model doesn't involve the negative feedback mechanism from activated ERK (ERK-PP) to SOS. This feedback mechanism results in the dissociation of the Shc-Grb2-SOS complex and therefore stops the activation of Ras and the MAPK cascade. The transient ERK response is then elicited. Instead, in Schoeberl's model the transient response is created by an incorrect modelling of deactivation of Ras. Consulting Figure 57 one can verify that there is a built up of an extra species Ras-GTP* (species 43). Differences in speed in creation and removal of this extra component assure the correct transient ERK-PP temporal profile (206). This inactive Ras-GTP* is not known to exist. This

study also concluded that the simulation results hold for a significant change in the parameters.

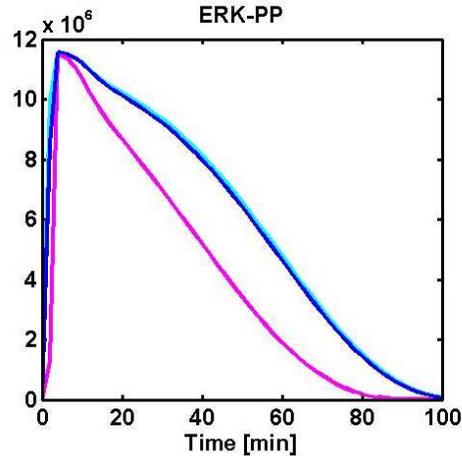


Figure 58 Temporal profile of ERK-PP (total) activation with the presence of the negative feedback loop from ERK-PP to SOS. Reproduction of the extended version (125) of EGFR Schoeberl's model (230). EGF concentration-50ng/ml. EGFR-50000 (purple); 500000 (dark blue); 5000000 (light blue). Parameters as in (125).

Another important aspect of the Schoeberl's model is the function of Shc. The adaptor protein Shc-dependent pathway is redundant and seems to be preferentially used (105). Shc plays an important part only at low EGF concentrations (223).

Disturbances of the normal regulation of cell division and differentiation plays a central role in tumourigenesis (221). Growth factors such as the EGF are crucially involved in cell proliferation. When absent, cell cycle inhibition or arrest may occur and cells may also undergo apoptosis. Constitutive activation of the MAPK cascade, activated by the EGF receptor pathway, may contribute to malignant progression of many human cancers. Although the causes of MAPK activation differ across the variety of tumours, constitutive signalling from the cell surface tyrosine kinase receptors contributes in many cancers to excessive activation of the Ras-Raf-MEK-MAPK central pathway. In breast cancer cells the EGFR is overexpressed up to 20% (73).

The temporal pattern of ERK activation determines cell fate. Besides the mechanisms reviewed in section 1.2.5, another reason for this could be that sustained ERK activation causes ERK to translocate to the nucleus. Transient ERK activation does not show massive translocation to the nucleus (149, 184). This might also be coherent with observations of protein p27, which is activated by mitogenic signal transduction pathways like the Ras-activated MAPK cascade. The protein p27 is a G1 to S phase cell-cycle regulator. During G0 and early G1, the protein p27 binds tightly and inhibits cyclin E1-cdk2. During G1 to S phase progression, proteolysis of p27 is increased, leading to its loss as cells enter S phase. The process of proteolysis is regulated by p27 phosphorylation. It is also known that MAPK can phosphorylate p27. In many cancers p27 degradation is increased (73). For the case of human breast cancer cells it was shown that constitutive MEK/MAPK activation changes significantly p27 phosphorylation and therefore protein levels. This stimulates tumour growth (73).

4.2.2 Tyrosine Kinase Inhibitor non-specific interactions and its influence on EGFR pathway output.

The kinome is made of around 500 kinases. Protein kinase networks work by transferring a gamma phosphate from an adenosine tri-phosphate (ATP) molecule to a hydroxyl group, attaching it covalently (Equation 75).

Equation 75



Protein kinases phosphorylate mostly serine, threonine and tyrosine residues. A class of drugs has been developed to target specific kinases and enable modulation of signalling networks (280). A great percentage of kinase inhibitors

target the ATP binding site and competitively bind the affinity pocket (Figure 59) (280).

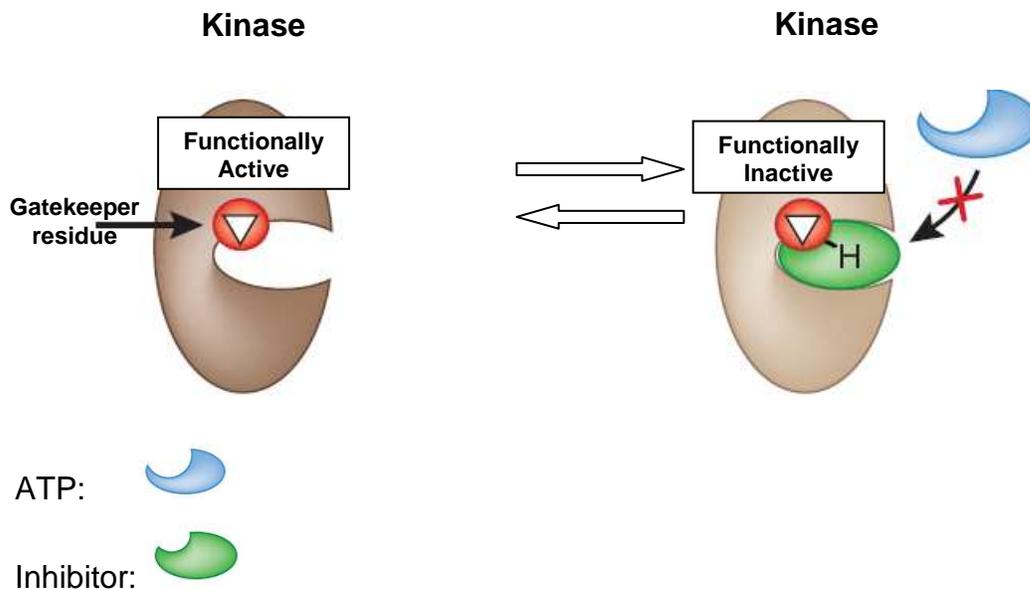


Figure 59 Representation of competitive ATP and Inhibitor binding to a kinase. The interaction represented in the figure holds the enzyme in an 'active' conformation, but does not allow any access to ATP. This functionally inactivates the kinase. H, hydrogen bond. Adapted from (29)

Kinases have the capacity to distinguish between different substrates by resorting to several mechanisms among which are slight differences in catalytic structures, which allows for the creation of somewhat selective inhibitors. In the case of successful binding of the inhibitor to the kinase, the access to ATP is not allowed, leaving the kinase functionally inactive (Figure 59), and the phosphorylation of a downstream kinase impossible. The ATP site is heavily conserved between all the different protein kinase families. This leads to the possibility of high cross-reactivity and non-specific interactions between the kinases and the inhibitors. Fabian and coworkers (79) used an ATP site-dependent competition binding assay to test specificity of 20 kinase inhibitors on

a panel of 119 protein kinases. It was shown that many of them had non-specific effects (79).

Six of the inhibitors (Figure 60) tested by Fabian and coworkers (79) targeted the EGFR: Iressa/Gefitinib, Tarceva/Erlotinib, ZD-6474, CI-1033, GW-2016 and EKB-569. For 5 of them we constructed simplified inhibition signatures that will be shown throughout this chapter. Nevertheless, for the simulation studies we will not work with ZD-6474 (Figure 81 A) or CI-1033 (Figure 81 B).

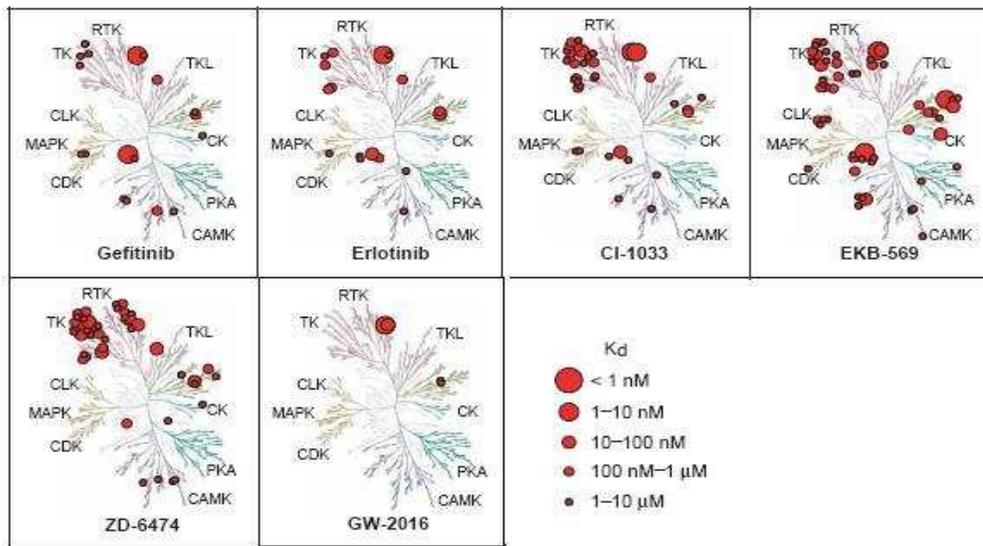


Figure 60 Cross reactivity across the kinome for six small molecule kinase inhibitors that target the EGFR. Degree of binding correlates with kinase inhibition. The smaller the binding constant the strongest the binding and clinical inhibition (79). GW-2016 is the most specific inhibitor. EKB-569 is the inhibitor exhibiting the highest cross-reactivity.

Gefitinib and Erlotinib have already been approved for clinical use in cases of non-small cell lung cancer (NSCLC) (1). The binding characteristics of each of the six small molecule kinase inhibitors targeting the EGFR are represented in Figure 60 and Table 6.

In order to understand the downstream effects on cell fate decision of the inhibition performed by each small molecule kinase inhibitor, a reproduction (in MATLAB R2010b) of an extended version (125) of the kinetic model developed by Schoeberl and coworkers (230) was targeted according to the binding assay

data available. Several attempts have been made to understand the correlation between dynamics observed in cancer cells and targeted therapeutic agents (see (2) for a review). Among them is a targeting study of a mathematical model of the EGFR pathway performed previously by Araujo and coworkers (14). Its main goal was to understand downstream “additive effects” of multiple node targeting. Nevertheless, the mathematical model used involved a lower number of molecular species (14, 151) and did not consider the internalization pathway. Yet, it involved the PLC γ pathway not considered in our reproduction of Schoeberl and coworkers model.

Although the map represented in Figure 60 includes kinases not involved in the kinetic model created in this work, we use the information on those common to both representations to evaluate the consequence on the activity of ERK-PP which, as stated before, correlates with cell fate decision (see also section 1.2.5). In the pathological case of excessive proliferation (tumour growth) the ERK-PP activity should reflect the situation. The kinetic model will be target according to the binding constants taken from reference (79). In this study we are interested in the “signature” of each small molecule kinase inhibitor across the kinases involved in our kinetic model. The absolute values for each K_d constant from Fabian and coworkers paper (79) will be used to perturb the kinetic equations (see also Table 6). Still, we want to maintain the qualitative relation between the absolute value of each constant and the degree of inhibition. We choose a generic simplified version of all the processes involved in inhibitor binding. With that purpose in mind we will resort to function $\varepsilon(\alpha, n)$ (see Equation 76). This generic function reflects the main aspect of inhibitor performance:

- The higher the affinity of the small molecule to the kinase (reflected in the lower values for the dissociation constant K_d) the higher the inhibition.
- The higher the inhibition the lower will be the value of the parameter ε (Equation 76).

Kinase inhibitors that target the ATP binding site work by competitive inhibition which arises from the fact that they also are able to interfere with the affinity

pocket (29). This is a conserved motif near the ATP binding site. For competition binding assay tests the kinetic constants are affected by the concentration of the inhibitors I and the dissociation constants according to a negative Hill type function (241):

Equation 76

$$v = \varepsilon(\alpha, n) \times k \times [C_k] \times [C_j] = \left[\frac{1}{1 + \left(\frac{\alpha}{n} \right)^h} \right] \times k \times [C_k] [C_j]$$

$$\text{with } \alpha = \left(\frac{[I]}{K_{d \min}} \right)$$

$$\text{and } n = \left(\frac{K_d}{K_{d \min}} \right)$$

$$h = 1$$

Normally for competitive binding Michaelis-Menten dynamics is assumed and the Michaelis-Menten constant for binding of the ligand to the protein to be inhibited is affected by $1/\varepsilon(\alpha, n)$ (Equation 77). Other more complex types of mathematical representation could be used for the interaction of the inhibitor with the kinase (241). Nevertheless, we aim at modelling the simplest case that shows the main mechanism at play.

Equation 77

$$v_i = k[C_k] \frac{[C_j]/K_{M'}}{1 + [C_j]/K_{M'}}$$

$$K_{M'} = K_M \times \frac{1}{\varepsilon(\alpha, n)} = K_M \times \left[1 + \left(\frac{[I]/K_{d \min}}{K_{d_k}/K_{d \min}} \right)^h \right] = K_M \times \left[1 + \left(\frac{\alpha}{n} \right)^h \right]$$

$$h = 1$$

Given the fact that we don't have access to those extra constants, and that we only have partial information on the processes the drugs affect, we continue to

assume mass-action dynamics as being the most appropriate option. For profiles of the inhibition function $\varepsilon(\alpha, n)$ see Figure 61 and Equation 76.

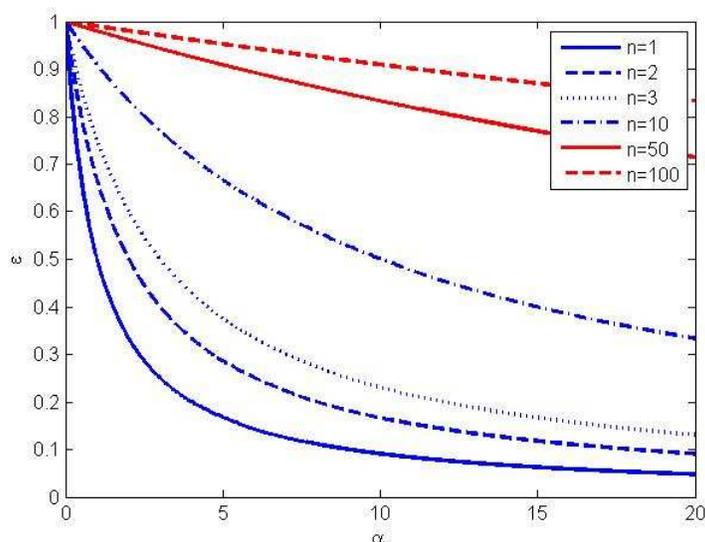


Figure 61 Profiles for inhibition constants $\varepsilon(n, \alpha)$. The higher the concentration of the inhibitor α the higher the inhibition. Greater values of n are associated to greater dissociation constants and therefore with lower affinity and efficiency of the inhibitor (smaller ε).

We should recall that we are interested in studying inhibition signatures or overall distributed action of drugs. This entails knowledge of quantitative action of inhibitors on network reactions. Nevertheless, parameters describing the interaction of the collected set of kinases from Fabian and coworkers' paper with the EGFR dynamic model haven't been published. We are therefore bound to explore the possible impact on EGFR processes of inhibited kinases not in a fully mechanistic fashion but through a generic approach such as that represented in Equation 76. This represents a possible impact on the pathway and not an accurate modelling of all types of processes.

For our generic analysis we will only analyze the cases of Gefitinib, Erlotinib and EKB-569 of the 5 drugs that targeted the EGFR tested by Fabian and coworkers (79). All of them have an inhibition profile which can be tested, to a certain extent, in the kinetic model used in this work: Gefitinib and Erlotinib

inhibit significantly, apart from the EGFR, the Cyclin G-associated Kinase (GAK) protein (also known as auxilin II) (281).

EKB-569 is the most promiscuous small molecule. One other protein inhibited by this drug which could be seen as being represented in the kinetic model belongs to the family of MAPK kinases: MAP3K4. Instead of MAP3K4 we have Raf, also a MAP3K. We'll invest in modulation of the processes where this protein is involved instead, since it serves as an input to the MAPK kinase cascade and could be crucial for amplitude and signal duration.

Our numerical simulation will focus on measuring the changes in the output of the dynamic model. The main output signal (ERK-PP) characteristics to be measured in the perturbative studies are represented in Equation 78 - Equation 80.

Equation 78

ERK-PP Signal Amplitude

$$S = \frac{I}{2v}$$

Equation 79

ERK-PP Signalling time

$$\tau = \frac{T}{I} \quad \text{with} \quad I = \int_0^{\infty} ERK - PP(t) dt \quad \text{and} \quad T = \int_0^{\infty} t ERK - PP(t) dt$$

Equation 80

ERK-PP Signal duration

$$v = \sqrt{\frac{Q}{I} - \tau^2} \quad \text{with} \quad Q = \int t^2 ERK - PP(t) dt$$

In Figure 62 we provide a geometric representation of each of the signal characteristics to be studied.

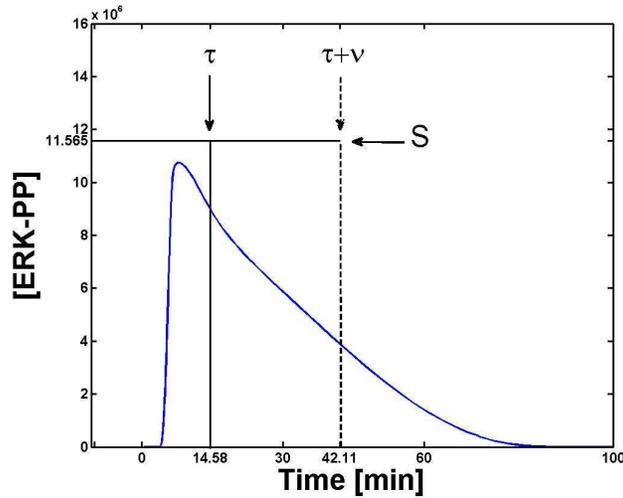


Figure 62 Geometric representation of the signal characteristics. τ -signalling time. ν -signal duration. S -signal amplitude. S corresponds to the height that a rectangle with base 2ν should have to equalize the area under the ERK-PP(t) curve (blue) (see also Equation 80).

We interpret the output signal as a distribution and calculate its mean in time (τ), its standard deviation (ν) and a measure of the area under the curve (S) (see Equation 78 to Equation 80 and Figure 62). S corresponds to the height that a rectangle with base 2ν should have to equalize the area under the ERK-PP(t) curve represented in Figure 62 (see also Equation 80).

4.2.2.1 Gefitinib and Erlotinib: inhibition signature maps and consequences for cell fate decision

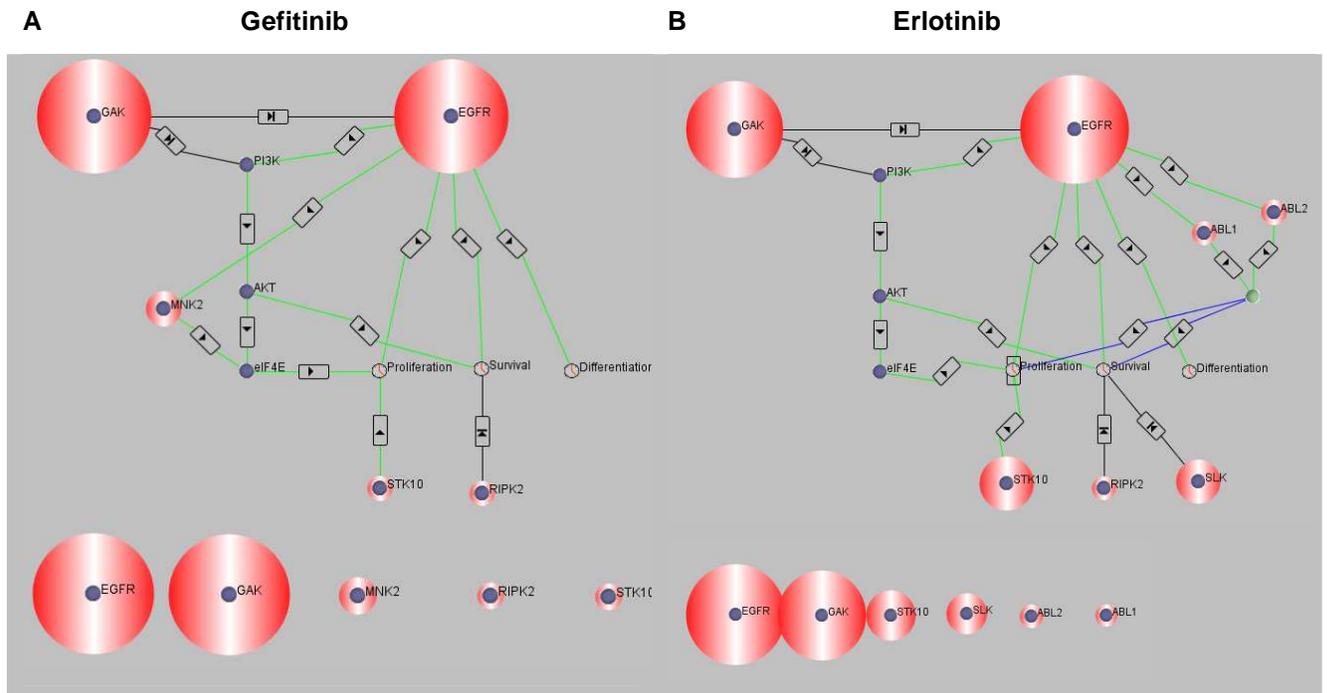


Figure 63 Inhibition signature of small molecule kinase inhibitors Gefitinib and Erlotinib. Diameter of red halo is proportional to degree of inhibition. **A)** Gefitinib. **B)** Erlotinib. The network representation involves kinases collected from several databases. Biocarta, CellSignal, Kinase.com, SwissProt and OMIM were used to link the kinases tested by Fabian and coworkers (79) to relevant phenotypic manifestations in Cancer. The simple map gives us an overall picture of the distribution of off-target interactions. This map was developed in Scipath by Dr.Sylvia Nagl and collaborators. Green arrows stand for stimulation and black arrows for inhibition. Additionally, also cellular responses (proliferation, survival, and differentiation) to upstream kinases are represented.

For the small molecule kinase inhibitors represented in Figure 63 there is no explicit representation, apart from the EGFR, of other component in the dynamic model that has significant inhibition. Nevertheless, the inhibition of the receptor and the conclusions taken from Schoeberl and coworkers (230) study, regarding the effects of the internalization pathway, may be informative on the possible effects of the inhibition, for instance, of GAK (Figure 63 A and B).

The protein GAK is a serine/threonine kinase involved in the uncoating of clathrin coated vesicles and thus regulates receptor trafficking (281) (Figure 64). This protein is relevant to the function of the EGFR pathway due to the protective

fundamental part that the internalization pathway has on signalling control. After ligand binding, the EGF receptors enter a process of internalization partly regulated through clathrin mediated endocytosis. The receptor is then either degraded or recycled. Non-specific inhibition of GAK affects, as can be seen in Figure 63, the internalization of receptors. Hence, a first non-desirable effect arises.

Additionally, it has been observed that down regulation of GAK is associated with an increase in cell proliferation and survival. The levels of receptor expression and tyrosine kinase activity are increased by a factor of more than 50 and the downstream signaling machinery is considerably affected. Amongst several important changes, the levels of activated extracellular signal-regulated kinase 5 and Akt suffer a substantial increase (281). The inhibitory action over Akt represented in Figure 63 A may then be a partial explanation for the enhanced proliferation recorded. Akt phosphorylates and inactivates the protein BAD, which is associated with programmed cell death. Inhibition of GAK has the opposite effect on Akt. Through the just described alternative route an additional non-desirable effect takes place as a result of Gefitinib and Erlotinib non-specific targets. GAK is consequently thought to have tumour suppressor potential.

Moreover, it should be noted that Cbl^f-mediated EGFR degradation pathway is considerably altered in cells with the GAK gene inactivated. The binding is made through its phosphotyrosine binding domain, which subsequently induces receptor ubiquitination.

In GAK knockdown cells, c-Cbl is in fact notably down-regulated. GAK probably affects c-Cbl stability by direct phosphorylation or through other convoluted routes (281).

Gefitinib and Erlotinib are relatively specific when compared to the maps for the inhibitors ZD-6474 (Figure 81 A) and EKB-569 (see Figure 73). Gefitinib has been shown to be most effective in attenuating proliferation in cell types which show a mutated or constitutively active forms of EGFR (113) and so only

^f Cas-Br-M Murine Ecotropic Retroviral Transforming Sequence Homolog

patients that exhibit this particular mutation will benefit from the therapy. Gefitinib's lateral effect on GAK may be one of the reasons. We have, therefore, an antagonistic combination of two effects generated by the same drug.

From Schoeberl's model (125, 230) it was concluded that the internalized part of the model revealed that its function may be signal amplification and therefore protection from apoptosis at low EGF concentrations and signal attenuation at high EGF concentration, which one could expect to be relevant for protection from excessive proliferation. Hence, at normal EGF stimulation the internalization pathway when perturbed may not be performing its attenuation function and, therefore, contributing to enhanced proliferative signals. Also, since an increase in receptor number also increases the duration and therefore the integral of ERK-PP signal (amplitude is maintained, Figure 58) the inefficient operation of internalization is further revealed.

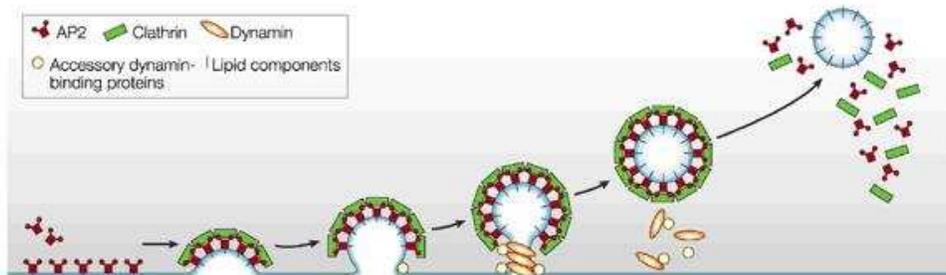


Figure 64 Vesicle formation and uncoating of clathrin vesicles. GAK is involved in the last process. GAK is also thought to be associated with CBL related degradation of EGFR (281). Taken from (110)

In reference (125) a study based on an adaptation of the Metabolic Control Analysis (MCA) method to intracellular signalling was performed. This method was based on analysing response coefficients of the form $C_i^A = d \log(A) / d \log(a_i) = (dA/A) / (da_i/a_i)$, where changes in a specific characteristic (e.g. amplitude A, (Figure 58)) of the output signal (ERK-PP) are analysed with respect to perturbations in a parameter a_i (e.g. reaction rates). The analysis in reference (125) revealed that small and $\pm 50\%$ perturbations on the parameters

regulating internalization and degradation in an extension of Schoeberl's model had interesting effects on the characteristics of the dynamic profile of ERK-PP (see Figure 58). Although the control is highly distributed, there is a core pathway that has a stronger influence (Figure 65).

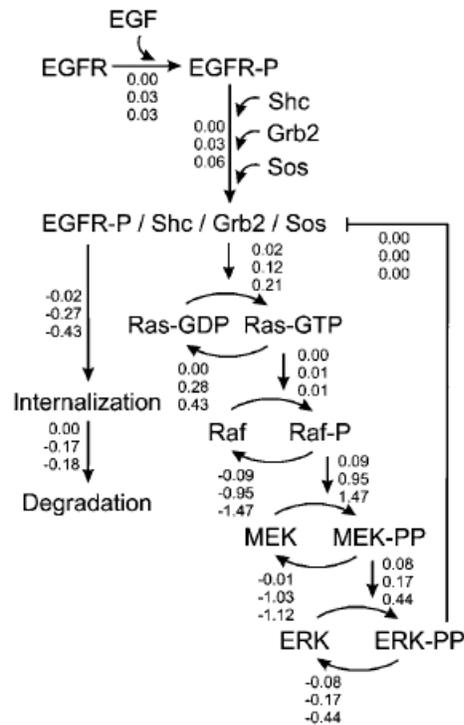


Figure 65. Condensed representation of control in the EGFR pathway. Each of the 148 reactions in the kinetic model was assigned, according to Hornberg and coworkers (125), to a particular class represented in this figure by an arrow. The numbers stand for the control (sum of the response coefficients) exerted upon the 3 characteristic of the output ERK-PP signal by each class of biological processes: amplitude (top), duration (middle), and area under the curve or integrated response (lower). This figure corresponds to the small perturbations in the parameters. For the remaining values see Table 4. Taken from (125).

We see that reactions leading to internalization of the EGFR and respective protein complexes have negative control both on the duration and integrated response of the ERK-PP signal. The amplitude for the coefficients is, nevertheless, significantly less than the more important processes in the system, those associated with Raf and MEK. Even though signaling still continues after internalization, changes in the rate of this process are reflected in ERK activation.

Since GAK is involved in this process, one expects that its inhibition is felt downstream.

Large negative perturbations on the internalizations and degradation processes do affect significantly the output component. Larger negative effects are felt on duration and integral response. Gefitinib inhibition of the EGFR (cross-phosphorylation reaction) and GAK (internalization) (see Figure 63 A) is similar to summing the respective response coefficients in Figure 65 (see also Table 4). The inhibition performed does affect significantly the output. In large perturbations (-0.5), signal duration decreases with the concerted action on receptor activation, internalization and degradation. Regarding the area of the signal (integrated response) the extra perturbation on the degradation process seems to give rise to an increment in this characteristic.

To account for each small molecule signature on the kinetic model we inhibited the respective kinetic parameters according to Equation 76 (see sections 4.2.2.2, 4.2.2.3 and 4.2.2.4). The parameters to be used are equal to those used in reference (230) and (125). This study is different from that performed by Hornberg and coworkers (125). We do not perturb parameter by parameter but use the small molecule kinase “inhibition signature” to study the possible “additive” effects of the non-specific cross-interactions. It is an approach which is motivated by the experimental literature but intrinsically it constitutes a multiparametric perturbation method that could be related to global sensitivity analysis methods developed recently (217).

4.2.2.2 Gefitinib non-specific interactions and EGFR pathway output

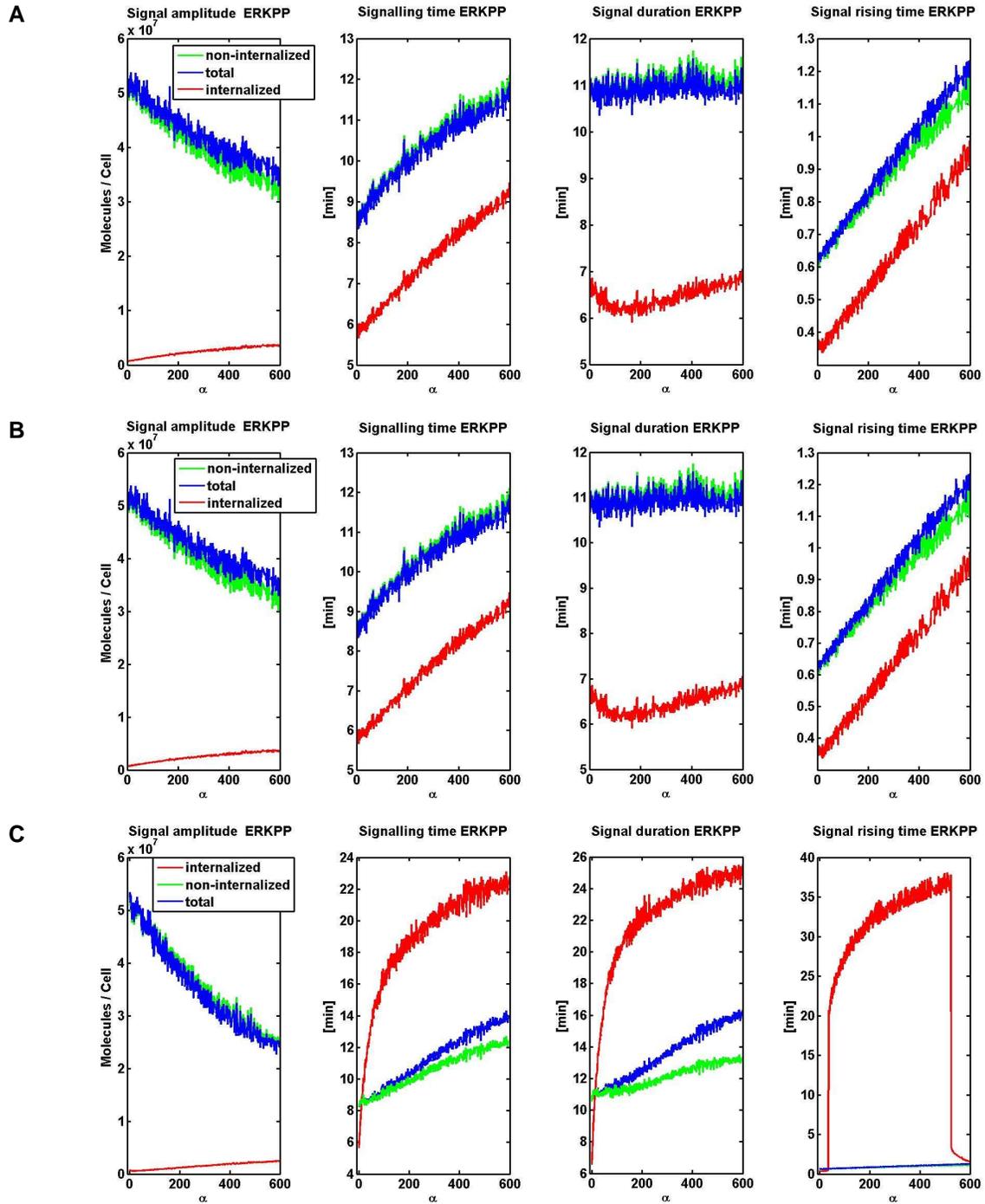


Figure 66 ERKPP response for Gefitinib. Normal expression of EGFR=50000, normal EGF concentration=50ng/ml. A) Only EGFR inhibition (cross-phosphorylation, rate v_3 (Figure 57). B) Cross inhibition-EGFR+internalization rate (v_3+v_6). C) Cross inhibition-EGFR+internalization+degradation ($v_3+v_6+v_{60}$). ODE model parameters are as in (125). See also Equation 78-Equation 80 for signal characteristics definition.

From the previous figure (Figure 66 A and B) we can observe that the higher the relative inhibition level (α) (see also Equation 76), the lower the output signal in the signal amplitude S (Equation 78) response. The signalling time (Equation 79) shows a gradual increase which means that the peak of ERK activation is reached at later and later times (Figure 67).

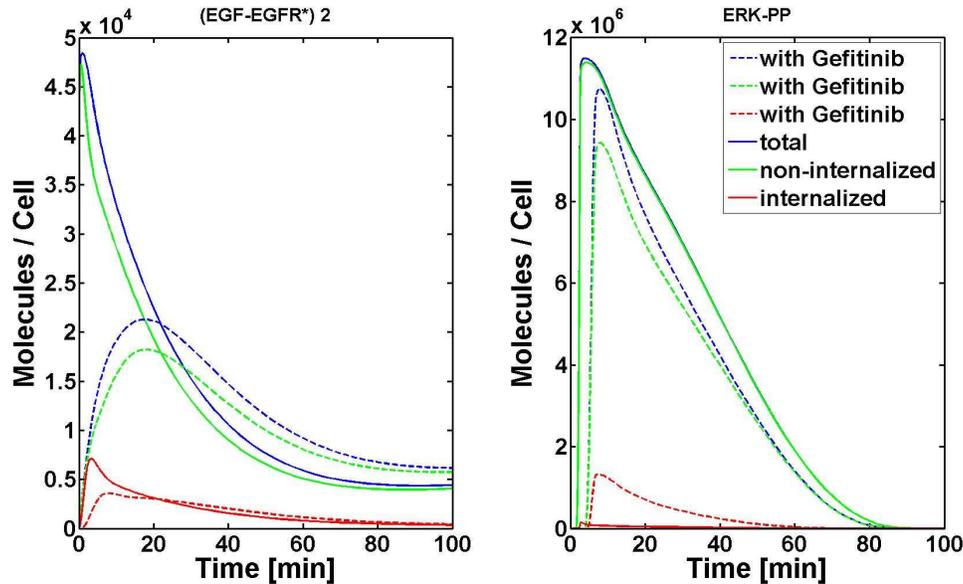


Figure 67 EGFR and ERK-PP profiles for EGFR Inhibition with Gefitinib vs. no Inhibition. The inhibition values correspond to the last recorded value for α (see Figure 66 B). ODE model parameters are as in (125).

The substantial decrease in amplitude was to be expected given that the pathway was simulated with EGF ligand at a constant level and the only reaction inhibited was the cross-phosphorylation between monomers EGFR. Regarding the contribution of the internalized pathway (red lines in Figure 66 and Figure 67) we observe an increasing contribution from this protective pathway to all the measured ERK-PP signal characteristics. The additional cross-inhibition of the internalization rate (Figure 66 B), representing the action of GAK, does not induce any further effects on any of the ERK-PP output signal. We also measured the signal rising time, which corresponds to the time necessary to go from 10% of maximum amplitude to 90% of maximum amplitude. This value has an overall tendency to go up as the inhibitor concentration rises (see Figure 66 A and Figure 67). At normal EGF concentration the perturbation of the

internalization machinery together with EGFR inhibition doesn't seem to contribute to an enhanced cell proliferative potential.

On the other hand, the extra inhibition of the degradation rate by the same amount as that used for the internalization rate gives us a different output profile (see Figure 66 C). The internalized part of the pathway has an interesting contribution. The signal duration, signalling time and signal rising time evolution rate with the inhibition level is much more pronounced than that of the surface receptors, even surpasses it at an α level around 20. The abnormal degradation machinery functioning contributes to the percentage with which the internalized pathway affects whole pathway functioning. It enhances the possibility of deregulated proliferation by reinforcing the signal (see differences between curves regarding the internalized pathway with and without Gefitinib applied, in Figure 68). After a considerably high level of α , the signal rising time is reduced drastically.

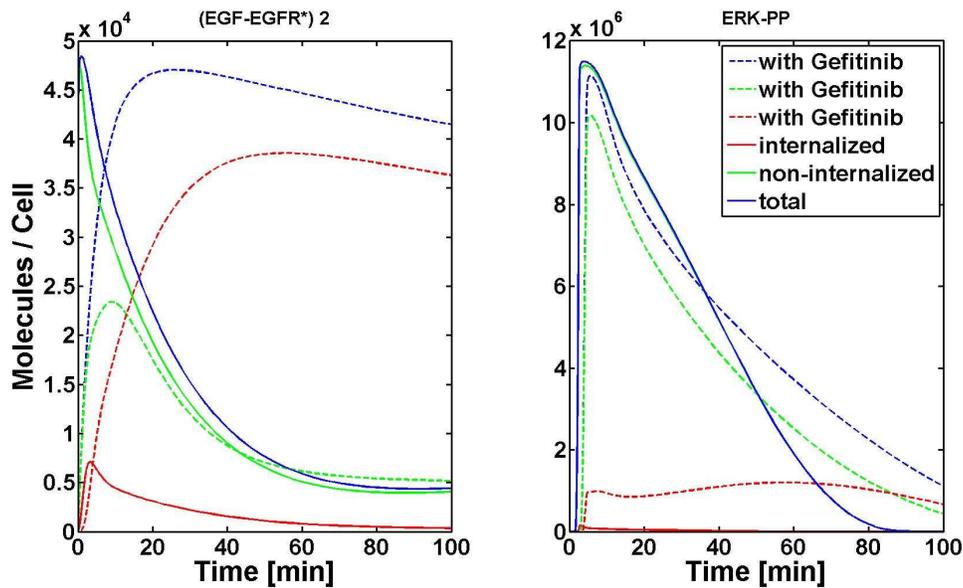


Figure 68 Differences in profiles EGFR when the internalization rate and the degradation rate are inhibited. Inhibition parameter $\alpha=400$ (see Figure 66C). ODE model parameters are as in (125).

These results show us that an additional effect of the cross-reactivity of Gefitinib may counterbalance its initial purpose. The overall performance of the internalized pathway increased always with inhibition concentration.

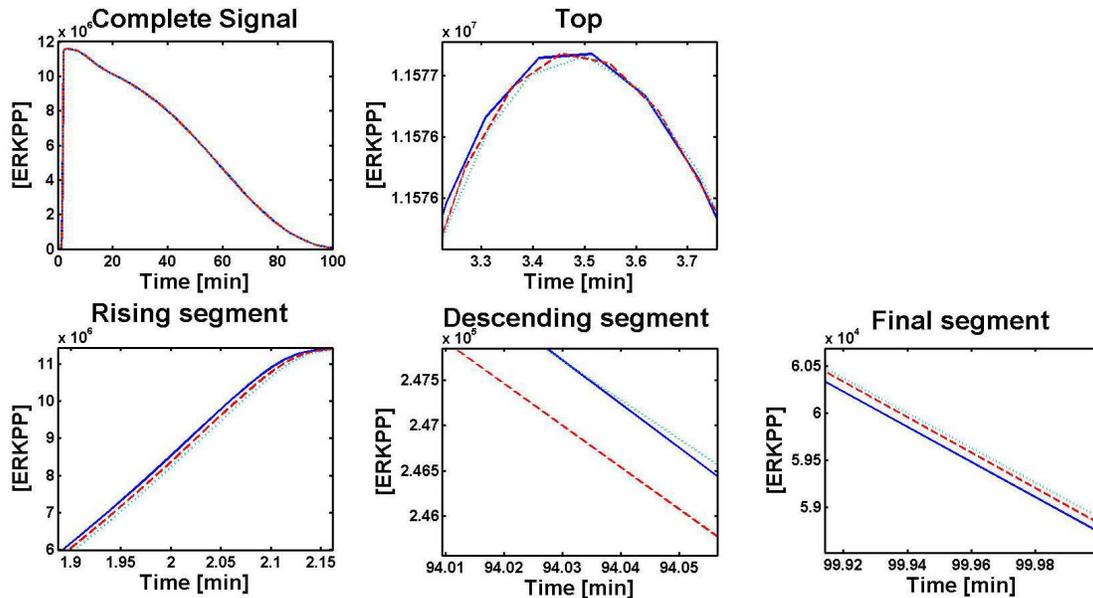


Figure 69 Relative order of signals for 3 consecutive α . See also Equation 76 and Figure 61.

We should make an observation about the irregular shape observed for all the signal characteristics measured. This stems from slight differences in signal shape that arise when we increase the inhibition parameter α . Signals starting above at initial steps with regards to other signals, when measured for consecutive α 's do not maintain their relative position (see Figure 69).

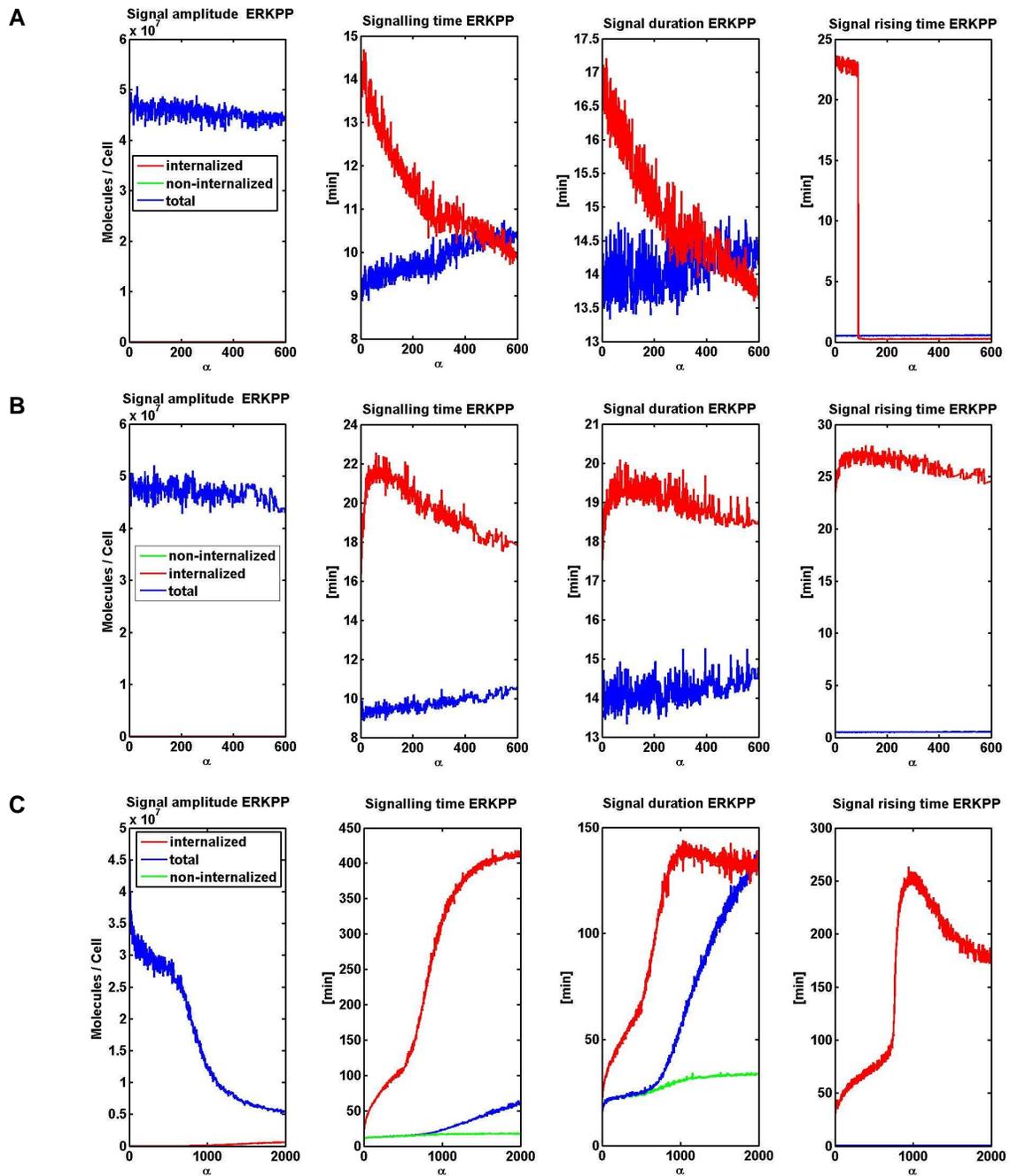


Figure 70 Gefitinib ERKPP response to inhibition α , under EGFR overexpression. EGFR=500000, EGF=50ng/l. A) Only EGFR inhibition (cross-phosphorylation, rate v3 (Figure 57)). B) Cross inhibition-EGFR+internalization rate (v3+v6). C) Cross inhibition-EGFR+internalization+degradation (v3+v6+v60). D) No inhibition. ODE model parameters are as in (125). See also Equation 78-Equation 80 for signal characteristics definition.

The simulation results performed by increasing the number of EGFR receptors by ten fold shows us, as expected, that although the signal amplitude is reduced

only slightly, the signalling time, duration and rising time increase with respect to the normally expressed EGFR simulation (Figure 70 A). If both EGFR and internalization rate are inhibited, only after application of substantial doses of α the signal starts to get attenuated. An even more contrasting result is that obtained for the simulation with inhibition of EGFR, internalization and degradation (Figure 70 C). After a steep rise in signal duration for the internalized part of the pathway, the output starts a gradual descend. The higher doses eventually perform the designed purpose of reduction of signal. Hence, we observe that the combined action of non-specific interactions of this particular small molecule are dependent on the applied dose. This observation shows, to an extent, the success of Gefitinib in patients that show particular mutations in EGFR that enhance signalling (discussed above).

4.2.2.3 Erlotinib non-specific interactions and EGFR pathway output

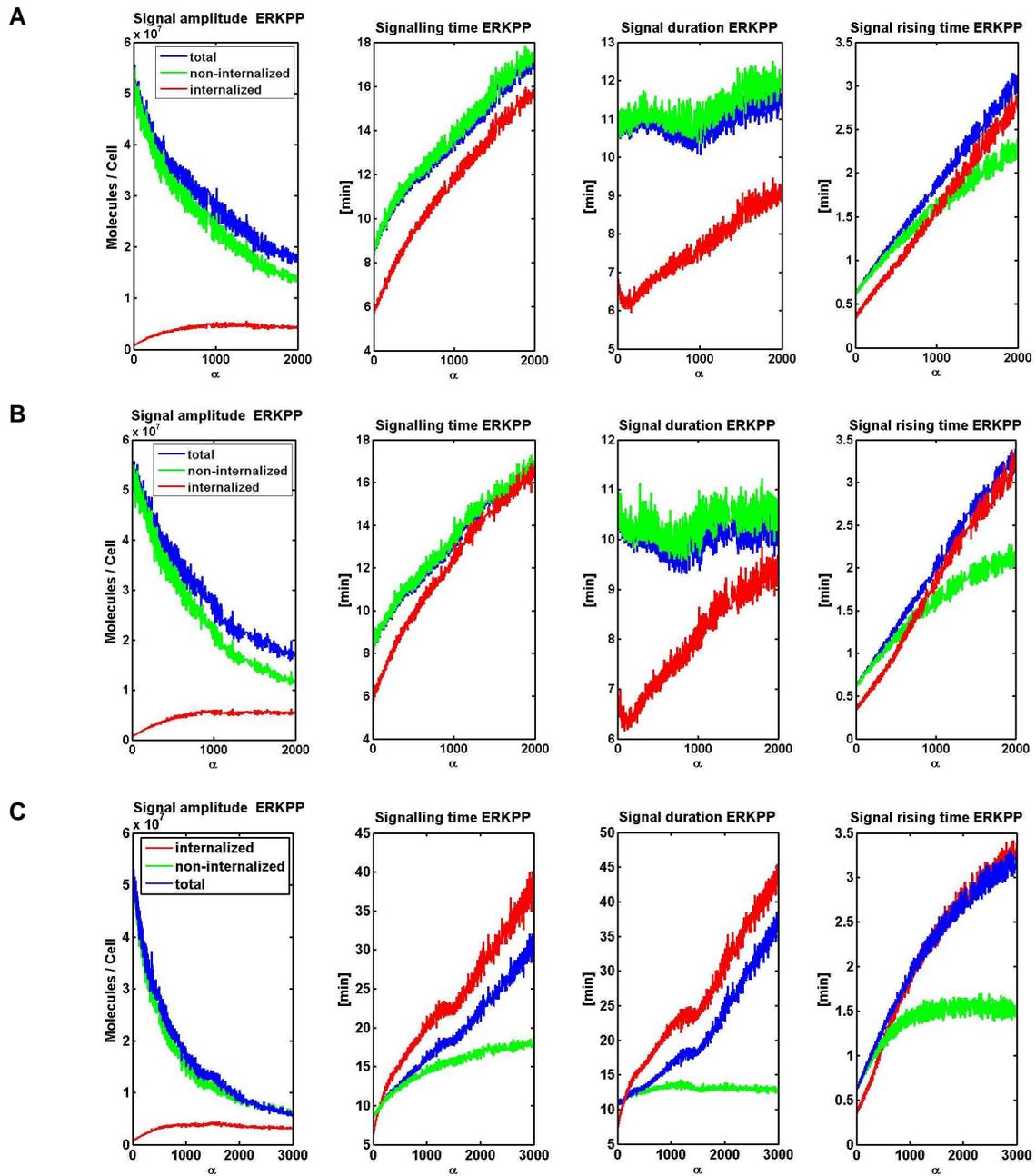


Figure 71 Erlotinib ERKPP response to inhibition α , under normal expression of EGFR. EGFR=50000, EGF=50ng/ml. **A)** Only EGFR inhibition (cross-phosphorylation, rate v_3 (Figure 57). **B)** Cross inhibition-EGFR+internalization rate (v_3+v_6). **C)** Cross inhibition-EGFR+internalization+degradation ($v_3+v_6+v_{60}$). ODE model parameters are as in (125). See also Equation 78-Equation 80 for signal characteristics definition.

Erlotinib has a stronger binding to EGFR than Gefitinib but a weaker effect on the internalization pathway (see Figure 63 B). The output profiles for each of the signal characteristics presented in Figure 71 don't show a very pronounced difference from the tendency of the output generated with Gefitinib inhibition. We should make a remark, on the other hand, that given Erlotinib's weaker action on the internalized pathway due to the differences in the dissociation constant (K_d) values regarding GAK (see Table 6), when EGFR inhibition, internalization and degradation rates are concomitantly affected, Gefitinib forces a higher contribution from the internalized pathway. This particularity has even more striking effects with overexpressed EGFR (Figure 70 C, Figure 72 C). The decrease in signal that was possible when substantial values of α were applied with Gefitinib, do not have the same effect with Erlotinib. The lower the dissociation constant K_d , the faster the response of ε (Equation 76) to changes in inhibitor concentration and the stronger the effects on output. Yet, the relationship is not linear (see Figure 61).

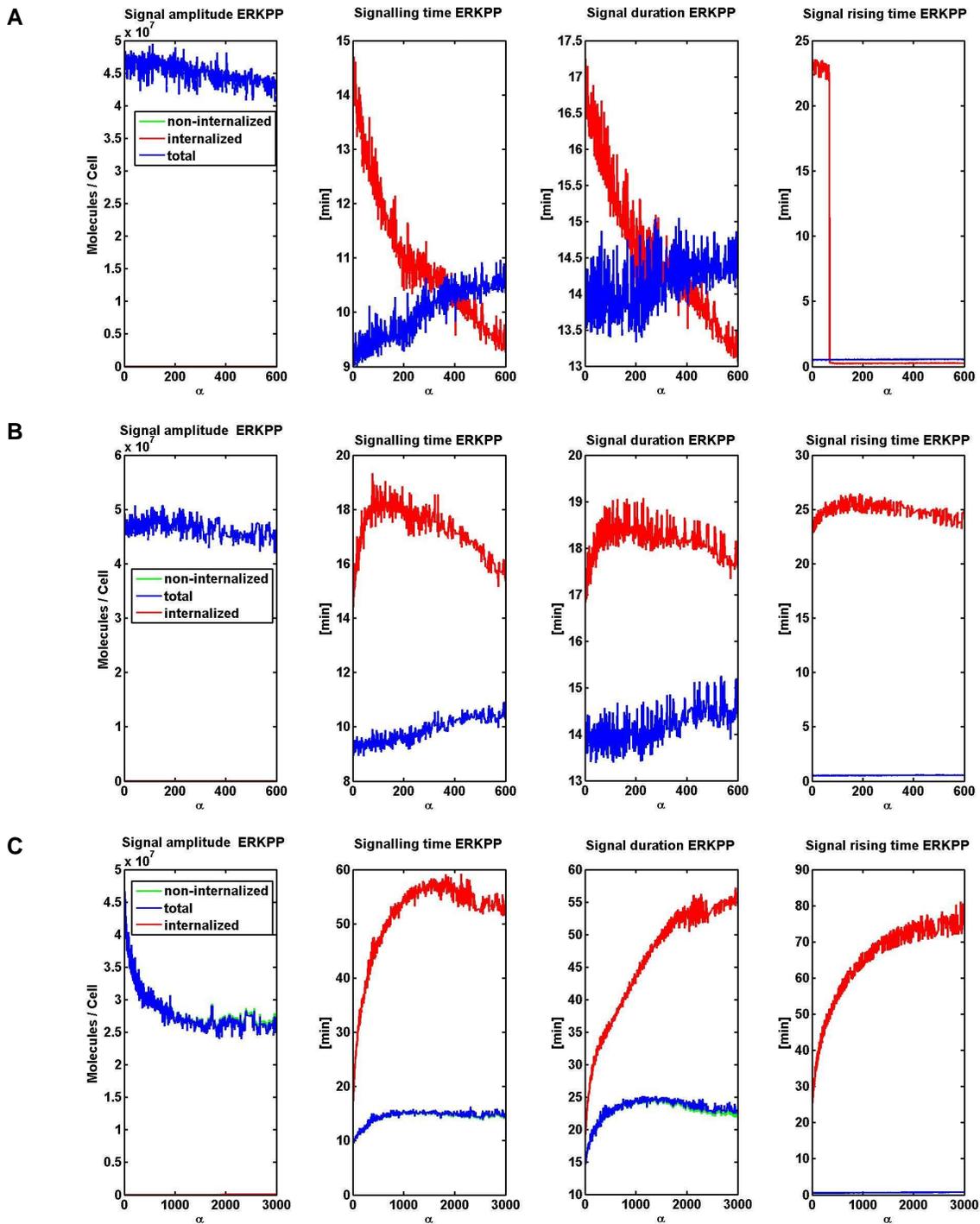


Figure 72 Erlotinib ERKPP response to inhibition α , under over expression of EGFR. EGFR=500000, EGF=50ng/ml. A) Only EGFR inhibition (crossphosphorylation, rate v3, Figure 57). B) Cross inhibition-EGFR+internalization rate (v3+v6). C) Cross inhibition-EGFR+internalization+degradation (v3+v6+v60). See also Equation 78-Equation 80 for signal characteristics definition.

4.2.2.4 EKB-569 inhibition signature map and EGFR pathway output

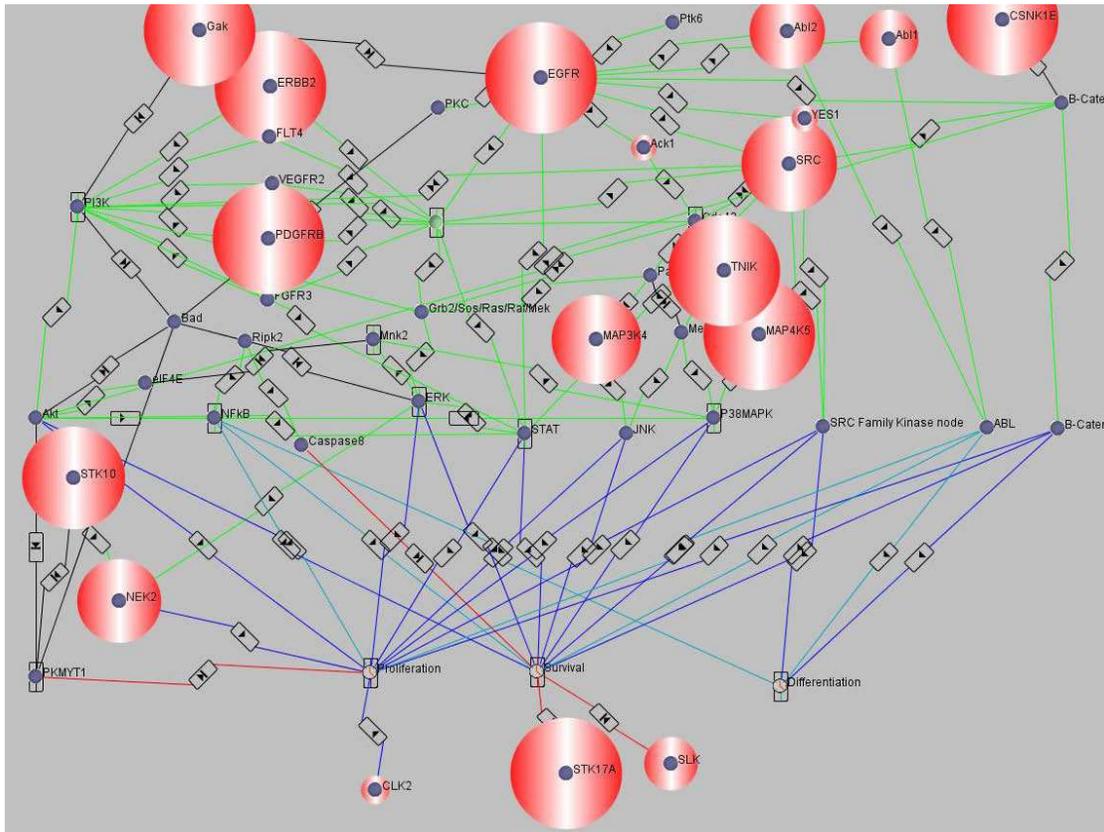


Figure 73 Inhibition signature for EKB-569. This is the most “promiscuous” small molecule kinase tested. The network representation involves kinases collected from several databases. Biocarta, CellSignal, Kinase.com, SwissProt and OMIM were used to link the kinases tested by Fabian and coworkers (79) to relevant phenotypic manifestations in Cancer. The simple map gives us an overall picture of the distribution of off-target interactions. This map was developed in Scipath by Dr.Sylvia Nagl and collaborators. Green arrows stand for stimulation and black arrows for inhibition. Additionally, also cellular responses (proliferation, survival, and differentiation) to upstream kinases are represented.

The action of this small molecule kinase inhibitor has an extra inhibition contributor in the kinetic model used in this work. As previously mentioned, the effect of the cross-reactivity with MAP3K4 of EKB-569 was simulated by inhibiting the process of activation of Raf (see Figure 75). The results in Figure 74 D show that this extra inhibition doesn't have any particular additional effects on any of the signal characteristics tested for this small molecule.

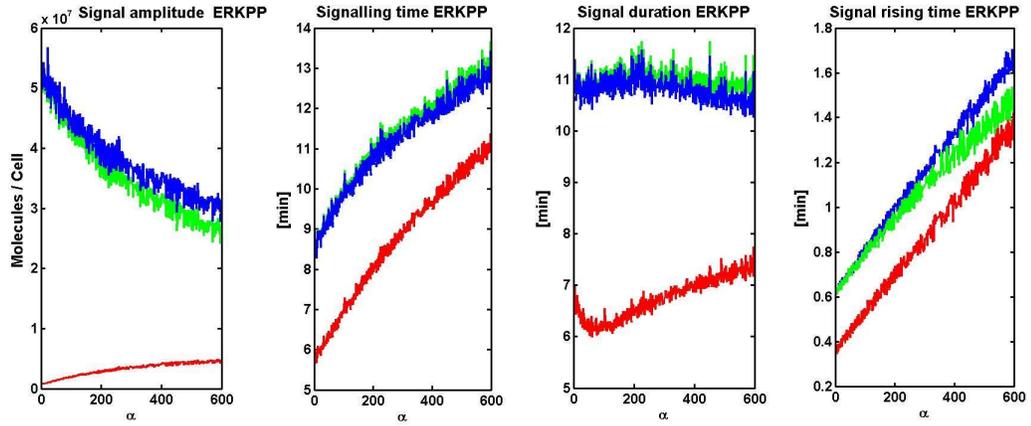
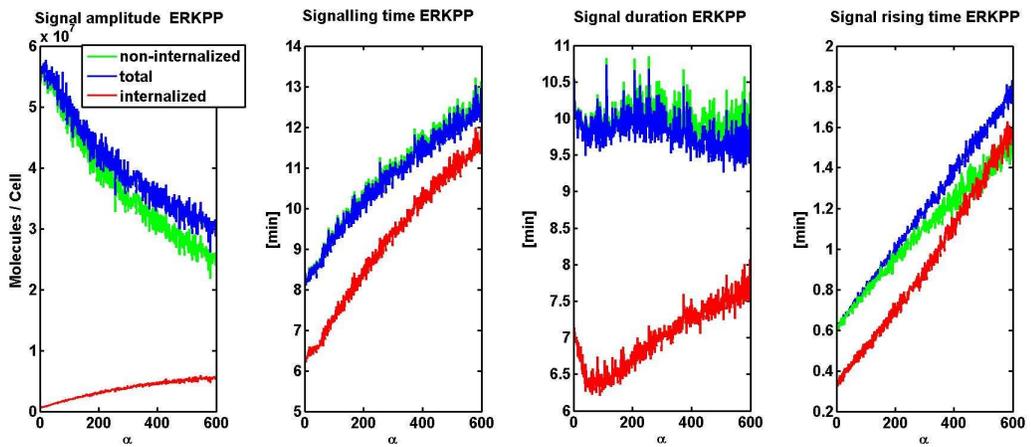
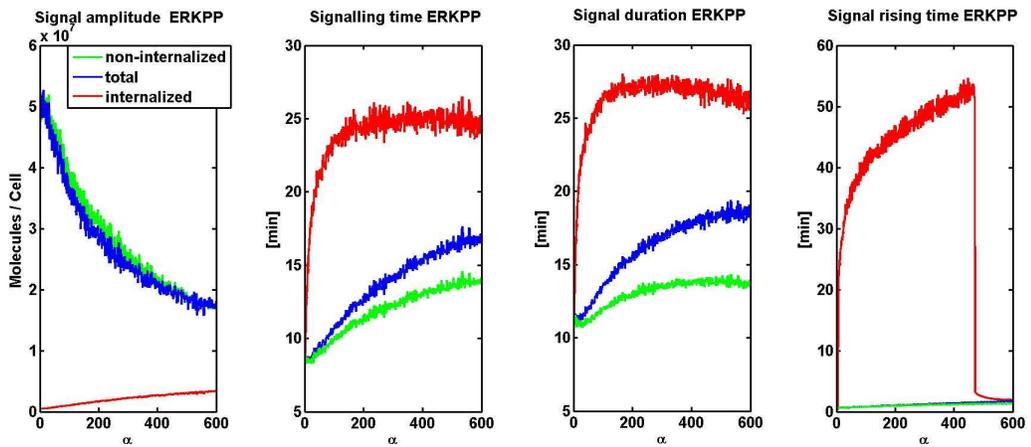
A**B****C**

Figure caption on the next page

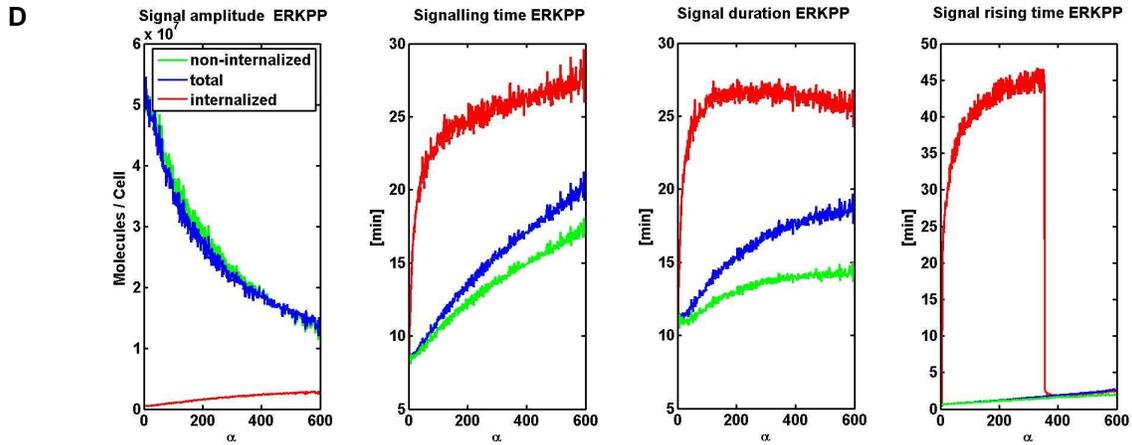


Figure 74 EKB-569 ERKPP response to inhibition α , under EGFR normal expression. EGFR=50000, EGF=50ng/ml. A) Only EGFR inhibition (cross-phosphorylation, rate v_3 , Figure 57). B) Cross inhibition-EGFR+internalization rate (v_3+v_6). C) Cross inhibition-EGFR+internalization+degradation ($v_3+v_6+v_{60}$). D) Cross-inhibition-EGFR+internalization+degradation+Raf ($v_3+v_6+v_{60}+v_{28}+v_{29}$). ODE model parameters are as in (125). See also Equation 78-Equation 80 for signal characteristics definition.

This may come from the fact that according to (79) the dissociation constant is particularly high ($K_d=1300 \times 10^{-4} \mu\text{M}$) with respect to the minimum ($K_{d_{\min}}=10 \times 10^{-4} \mu\text{M}$). Hence, ε is relatively small.

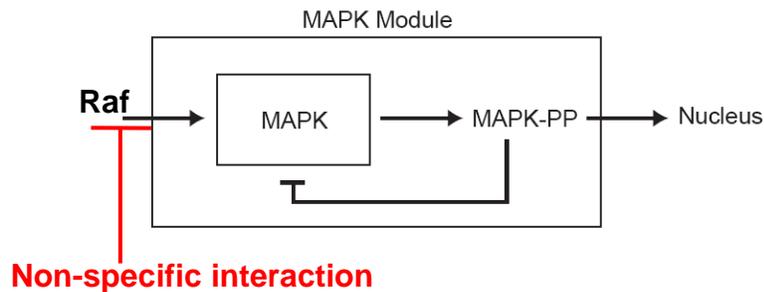


Figure 75 Generic representation of the extra inhibitory interaction considered for the EKB-569 with respect to the MAPK module. According to certain views brought from control theory, the MAPK 3 layer cascade is resilient to perturbations inside the negative feedback loop (228).

If the extra inhibitory interaction had been sufficiently strong the input signal to the MAPK cascade module would have been hindered and, as a result, we would have expected a substantial attenuation in ERK-PP output signal.

If perturbations had been performed inside the module, its performance as a feedback amplifier (228) may have compensated for the inhibition used: the effects would have been absorbed (position defended in reference (228)). Results from the adaptation of the MCA analysis to signalling performed in reference (125) prove that perturbations inside the module (Table 4, Processes 6,7,8,9,10) do have strong effects! Actually the strongest! Moreover, the presence of the feedback loop doesn't affect as much as expected the ERK-PP output. This is, apparently, in contrast with the predictions from the use of control theory principles in understanding embedded network motifs such as those highlighted in (228). The feedback loop from ERK-PP to SOS had been omitted from the original kinetic model developed by Schoeberl and coworkers (230) as was previously mentioned. Its inclusion is crucial for the correct ERK-PP output concentration profile as stated in (206). If perturbations on this feedback loop don't exert considerable influence (Figure 65 and Table 4) on the signal characteristics studied in this report, its action on the pathway may not be as strong as anticipated and the MAPK cascade module is exposed to perturbations on its components Raf, MEK, and ERK.

The results obtained by applying the same perturbation strategy as for the previous cases give similar results to Gefitinib. Their constants K_d are similar for the same species inhibited. The difference with respect to Gefitinib arises once again in the overexpression scenario (see Figure 76). EKB-569 is not as efficient in decreasing output signal when EGFR, internalization rate and degradation rate are inhibited. This comes as a surprise as EKB-569 binds more strongly than Gefitinib both to EGFR and to GAK (see Table 6). Apparently, Gefitinib has the adequate balance between the two targets to induce the changes in output signal reported above.

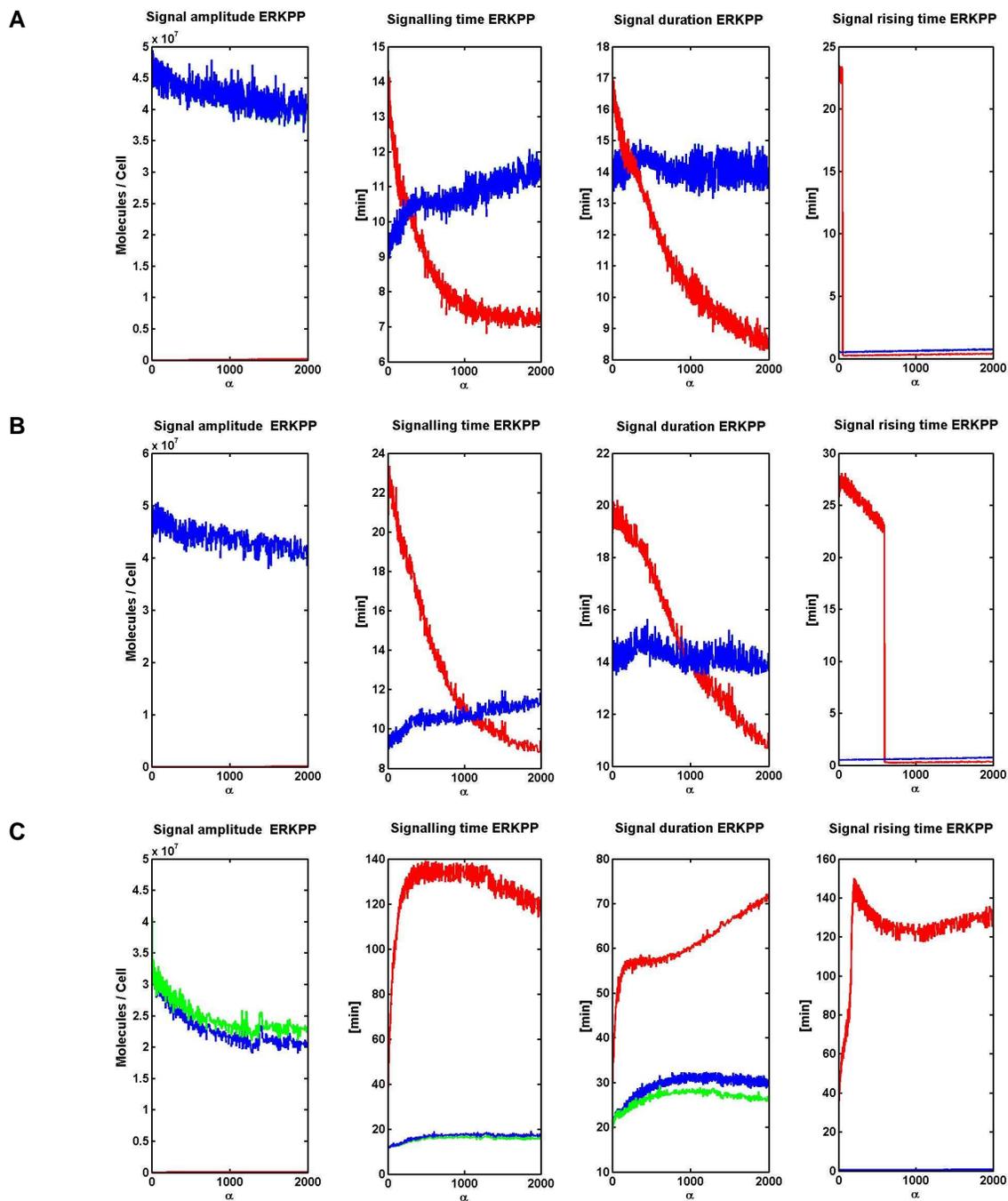


Figure caption on the next page

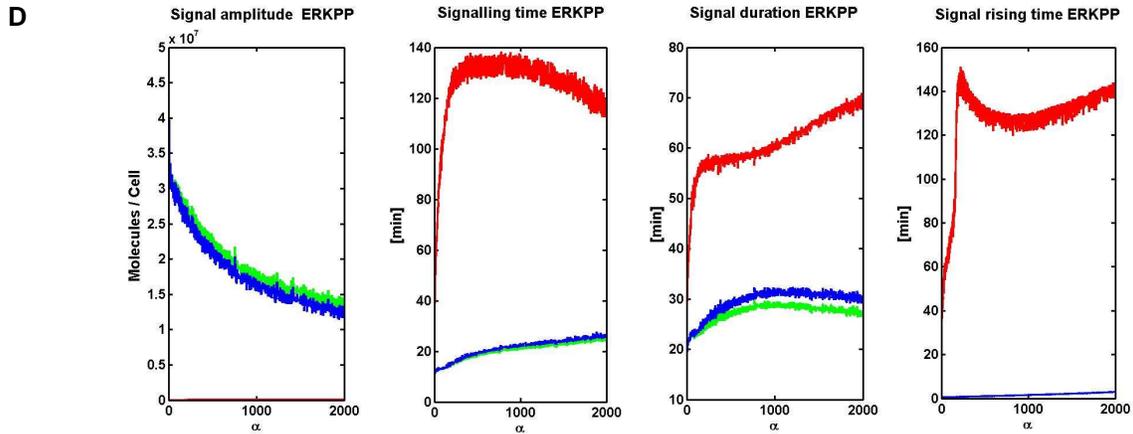


Figure 76 EKB-569 ERKPP response to inhibition α , under overexpression of EGFR. EGFR=500000, EGF=50 ng/ml. A) Only EGFR inhibition (cross-phosphorylation, rate v3 (Figure 57)).B) Cross inhibition-EGFR+internalization rate (v3+v6).C) Cross inhibition-EGFR+internalization+degradation (v3+v6+v60).D) Cross-inhibition-EGFR+internalization+degradation+Raf (v3+v6+v60+v28+v29). ODE model parameters are as in (125). See also Equation 78-Equation 80 for signal characteristics definition.

4.3 Extending the study of non-specific small molecule kinase interactions to a larger integrative network: the Human Signalling network and the FANTOM4 regulatory network

To extend the link between non-specific interactions of therapies and the necessity for a multitargeting approach, we sought to assemble a network incorporating both signalling proteins and transcription factors exerting feedback on the signalling module. Due to time restrictions, we did not use the network in extensive perturbative studies as was performed before in this chapter on the EFGR ODE kinetic model. Nevertheless, the modelling approach developed for integrating both qualitative data collected from the literature and network structure will be described ahead. It combines methods described in section 1.4. and constitutes an original approach to modelling integrative systems' dynamics.

The network chosen should incorporate at least 50% of the kinases assayed by Fabian and coworkers (79). After extensive search we came across two networks fulfilling the requirements for perturbative studies with identifiable global phenotypes or dynamical patterns: the Map of human cancer signalling

and the transcription factor network developed under the program FANTOM4 (248).

The original human cellular signaling network was constructed by consulting the signaling pathways from BioCarta database. The curated dataset included gene names, functions, cellular locations and functional relationships between genes (17, 64). The information collected from Biocarta was also merged with another literature-mined signaling network involving approximately 500 proteins (178). As a result, the published network was made of around 1600 nodes and 5000 links. Around 190 nodes correspond to Transcription factors, 254 to kinases and 31 to phosphatases. An initial evaluation of the basic network connectivity properties (in-degree, and out-degree distribution) in and out-degree can be visualized in Figure 77.

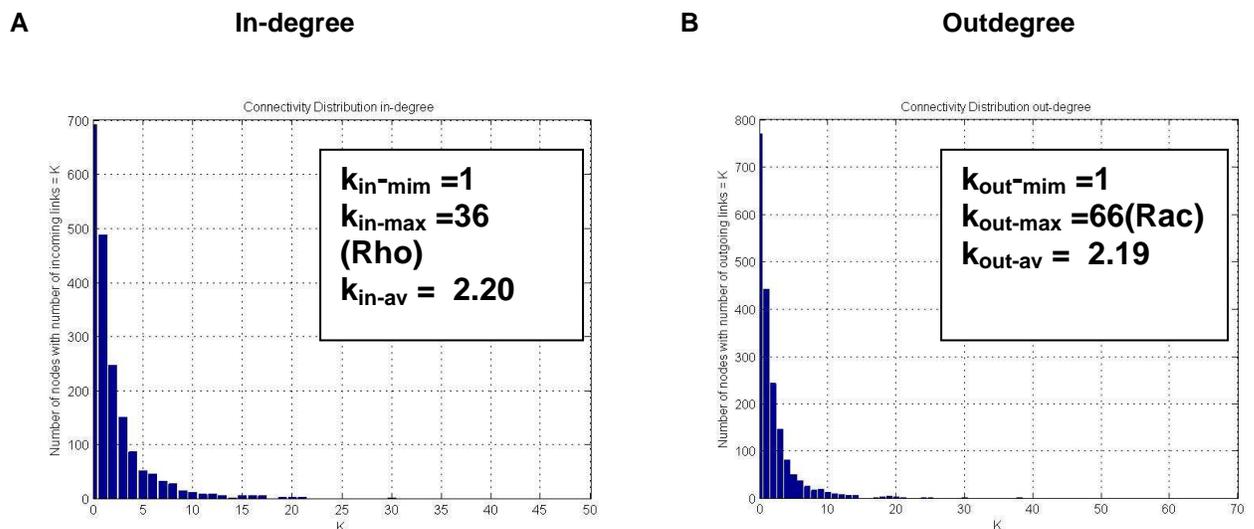


Figure 77. Connectivity distribution (in-degree) for the original human signalling network (64). A) In-degree ($k_{in}^{mim} = 1$, $k_{in}^{max} = 36$, $k_{in}^{av} = 2.20$). B) Out-degree ($k_{out}^{mim} = 1$, $k_{out}^{max} = 66$, $k_{out}^{av} = 2.19$).

As can be seen in the previous picture, most of the nodes have 2 connections. An extensive study was performed on the Map of human cancer signalling including most common network motifs, the existence of structural network communities and modularity and the cellular phenotype most commonly

controlled by each. We recommend the reader the paper of Cui and coworkers (64).

4.3.1.1 Extension of the Human signalling network

The Human signalling network was originally constructed to include transcription factors activated by upstream signalling proteins and participating in multimerization or complex formation necessary for transcription initiation. All the interactions between transcription factors were represented by protein-protein links. No protein-gene regulatory interactions were represented. The extended version presented in this work includes also regulatory interactions between transcription factors devised under the FANTOM4 project, and presented in a paper by Suzuki and coworkers (242). In this paper the transcriptional network that controls growth arrest and differentiation in a human myeloid leukaemia cell line was analysed.

An extension has been performed to the Human signalling network, within our project, by including some of the kinases assayed by Fabian and coworkers (79) on a small molecule kinase inhibitor study of non-specific interactions across the kinome and the core network of transcription factors devised under the FANTOM Consortium and the Riken Omics Science Center (<http://fantom.gsc.riken.jp/4/>). The method for adding to the original network was similar to the one followed originally in its construction. Approximately 40-45% of the panel of 119 kinases assayed in Fabian and coworkers' paper (79) was already represented in the Human signalling network. For some of the kinases assayed and not represented in the Human signalling network we tried to identify all the possible connections by consulting Biocarta (www.biocarta.com), Cell Signal (www.cellsignal.com), Kinase.com (www.kinase.com), OMIM (<http://www.ncbi.nlm.nih.gov/omim/>) and using Cytoscape (<http://www.cytoscape.org/>).

Phosphorylation events were represented through stimulatory edges ($s=+1$). To any connections representing transcription factor complex formation necessary

for transcription initiation a neutral link (0) is attributed. Stimulatory and inhibitory connections are associated with $s=+1$ and $s=-1$ respectively. The nature of each connection ($s=\pm 1$) was extracted from the FANTOM4 database, EDGE EXPRESS DB centre view (<http://fantom.gsc.riken.jp/4/edgeexpress/view/index.php#5558263>), by analyzing the respective node perturbation edge weight matrix devised by siRNA. We focussed on the core regulatory network devised in (248) of 55 highly trusted edges among 30 core motifs. This study measured the “ *genome-wide dynamics of transcription-start-site usage in the human monocytic cell line THP-1 throughout a time course of growth arrest and differentiation*” (248). The list of nodes and the respective connections will be provided with the electronic version of the thesis (see attached CD).

4.3.2 Modelling the extended Human signalling network

The information available about each functional connection between nodes/proteins in the extended network is in the form of a connectivity matrix with stimulatory (+1), inhibitory (-1) and neutral links (physical interactions). A stimulatory connection (+1) would be for example a phosphorylation reaction between 2 kinases. An inhibitory link could represent again a phosphorylation inactivating the downstream protein or the action of a phosphatase. Neutral links represent reactions associated with the necessary binding between proteins in complex formation. In order to model dynamic aspects of network behaviour we will resort to generic forms of representing the class of biochemical processes taking place. The dimension of the network and consequently the computation time would not allow more detailed approaches. Although there are several processes in the literature, e.g. EGFR pathway (51, 230), TNF pathway that have been parameterized, most of the links have no information at all.

A representation under the Boolean formalism for approximately 130 nodes in the present network has been previously published (122). Given that some nodes have very high connectivity (Figure 77), extending this formalism to

the whole network would result in impractical Boolean input function tables. Therefore, we chose to transform the existing Boolean functions into a continuous representation (104, 213) maintaining the same functional properties and that can be applied in the future to the whole network dynamics.

4.3.2.1 Transforming qualitative Boolean expressions into their algebraic equivalent

Let b_i be a Boolean variable, i.e., $b_i \in \{0,1\}$ (4). The complements of b_i are represented as $\overline{b_i}$. The objective is to define an algebraic function $F(X_1, \dots, X_n)$ depending on the algebraic equivalents to the Boolean variables b_i , and that represents the logical Boolean tables ($B(b_1, \dots, b_n)$) with continuous variables (X_1, \dots, X_n) (see Equation 81).

$$\text{Equation 81} \quad F(X_1, \dots, X_n) = \sum_{\alpha \in A} m_{\alpha}(X_1, \dots, X_n) = B(b_1, \dots, b_n) \text{ if } X = b$$

Take for example the following Boolean function:

$$\text{Equation 82} \quad b_1 + b_2 = \overline{b_1}b_2 + b_1\overline{b_2} + b_1b_2$$

Each Boolean function can be represented by the sum of its minterms (rows in the Boolean table where the input function equals 1, m_{α} in Equation 81 and Table 3).

Minterm	b_1	b_2	$b_3=b_1+b_2$
	0	0	0
$m_1=\bar{b}_1b_2$	0	1	1
$m_2=b_1\bar{b}_2$	1	0	1
$m_3=b_1b_2$	1	1	1

Table 3 Boolean function table corresponding to the logical OR function with the corresponding minterms. Minterms are rows where $B(b_1,b_2)=1$ (true).

By substituting each of the Boolean variables in Equation 82 with the algebraic equivalent, $b_j \rightarrow X_j$ and $\bar{b}_j \rightarrow (1 - X_j)$, we get the equivalent algebraic equation $F(x)$ (Equation 81). The transformation follows Equation 83. The matrix $b=[b_{kj}]$ stands for the matrix composed of the b_1 and b_2 columns in Table 3. $B^3=[B_k^3]$ stands for the values of the Boolean function for each of the lines of matrix $b=[b_{kj}]$ for variable or node X_3 in the network.

Equation 83 $F(X_3) =$

$$\begin{aligned}
 & \overbrace{[(1-b_{21}) \times (1-X_1) + b_{21} \times X_1] \times [(1-b_{22}) \times (1-X_2) + b_{22} \times X_2]}^{\text{min term 1}} \times B_2^3 \\
 & + \\
 & \overbrace{[(1-b_{31}) \times (1-X_1) + b_{31} \times X_1] \times [(1-b_{32}) \times (1-X_2) + b_{32} \times X_2]}^{\text{min term 2}} \times B_3^3 \\
 & + \\
 & \overbrace{[(1-b_{41}) \times (1-X_1) + b_{41} \times X_1] \times [(1-b_{42}) \times (1-X_2) + b_{42} \times X_2]}^{\text{min term 3}} \times B_4^3
 \end{aligned}$$

This approach allows us to compute the algebraic expression for each of the 130 nodes of the subnetwork for which there is an associated Boolean table developed by Helikar and coworkers (122). The generalized form is expressed in Equation 84 .

Equation 84

$$F(X_i) = \begin{bmatrix} \prod_{j \in I} [(1 - b_{1j}) \times (1 - X_j) + b_{1j} \times X_j] \\ \vdots \\ \prod_{j \in I} [(1 - b_{2^n j}) \times (1 - X_j) + b_{2^n j} \times X_j] \end{bmatrix} \mathbf{B}^i$$

In Equation 84 I stands for the set of inputs or links influencing variable/node X_i . Additionally, each X_i can follow a sigmoid that assures that the variables are always smaller than 1.

4.3.2.2 Modelling the remaining signalling and transcription processes

For each of the signalling nodes where no Boolean tables are available a generic mass action law will be applied (Equation 85):

Equation 85

$$\tau_i \frac{dX_i^a(t)}{dt} = \left[(X_i(t)) * (sa_i + \delta_{i1}L + \sum_{j=1}^{k_{in}} k_{ij} X_j^a(t) + \dots) \right] - \left[X_i^a(t) * (sd_i + \sum_{j=1}^{k_{in}} l_{ij} X_j(t) + \dots) \right],$$

$$X^a(t) = [X_i^a(t)] \text{ activated form, } i, j = 1, \dots, n$$

L stands for ligand. sa and sd stand for a basal level of activation and decay. The transformed Boolean tables into polynomials is equivalent to having higher order terms of the form, $k_{ijj} X_j X_j$ in Equation 85. These stand for nonlinear interaction between binding sites in a protein or between proteins in the formation of complex structures in order to activate species X_i .

The modelling approach for the dynamics of the FANTOM4 network will be based on a mean-field approximation for promoter occupation (11) which was explored in this thesis.

The extended Human signalling network includes both protein-protein interactions and protein- interactions. Its integrative nature allows for exploring network dynamics in the presence of signalling-transcription separation of time-scales, initiated in chapter 2, but in high-dimensional space. Moreover, due to the fact that the network is motivated by data collected from experimental studies, and we have information on the distribution of targets of a class of therapies, small molecule kinase inhibitors, the study performed for the EGFR pathway could be greatly expanded. We have, therefore, established a perfect ground to develop and test the concept of “inhibition signature” affecting gene expression dynamics and, consequently, pattern or attractor selection. The optimal inhibition signatures should integrate time-scale separation between signalling and transcriptional processes, as well as network structure information.

4.4 Developing inhibition multitargeting signatures inducing the appropriate expression patterns

Systematic biological perturbation approaches call for quantitative phenotypes monitoring system's function. A global phenotype such as cellular proliferation has been frequently used (277). As in the working hypothesis used in chapters 2 and 3 this phenotype can be identified with specific genetic programs or systems attractors, inducing a particular cellular response can be equated as a gene expression pattern selection optimization problem. The use of integrative phenotypes allows for simultaneous examination of system's functions with few measurements. Thus, it is expected that for large-scale investigations of high-order combinations global phenotypes will be most practical.

Multitargeting should be formalized as a basin hopping optimization problem in a multiple fixed attractor landscape or in an induced landscape by external inputs to the transcriptional system ($O(t) \rightarrow F_{TS}$, see Figure 78). In the latter scenario external signals induce epigenetic bifurcations which result in the appropriate attractor being selected (recall the theoretical work performed in both section 2.2 and section 2.3). The end attractors can be point attractors or dynamic attractors with the initial state in phase space determining the accessibility to each of the modes. If the system, for instance in a pathological scenario like cancer, is initially in attractor P (e.g. proliferation in the case of a cancer cell) (see Figure 79, upper figure) and we wish to force it into attractor A

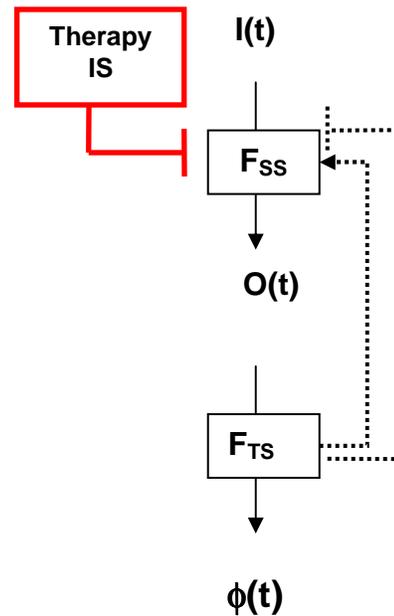


Figure 78 Functional representation of integrated signalling-gene regulatory systems. IS-inhibition signature. F_{SS} - function performed by the signalling system. F_{TS} -function performed by the transcription system. $O(t)$ -output combination of nodes' profile of the signalling system, e.g. IEGs (see Figure 46). $\phi(t)$ - N dimensional gene expression pattern. The negative and positive feedback represent the action of composite motifs (7), not explored above in other sections.

(apoptosis) or D (differentiated cell), a multiple signalling targeting strategy (IS_{Therapy} , see Equation 88) needs to be developed that induces the appearance of only A, akin to a “progressive” switch, erasing P and D completely (Figure 79, left lower figure) or making it highly improbable, i.e. with a small basin of attraction (Figure 79, right lower figure). The problem can be formulated by minimizing the “distance” (Equation 90) between the current state, phenotype ϕ^P , and the target states $\phi^{A,D}$. We defend a multitargeting approach as opposed to a single target approach, due to the wealth of data published on the connectivity of intracellular network, which as shown that cross-talk between pathways renders single target therapies inefficient (see section 1.3.2) and that cells integrate multiple signalling cues at any point of its life-time (96).

A specific phenotype, A, D or P is characterized by a gene expression pattern ϕ , which is induced by specific time-dependent inputs $I(t)$ (see Figure 78). This phenotype arises as a function of the connectivity matrix associated with both the signalling system W_{SS} and the transcription system W_{TS} (see Equation 87).

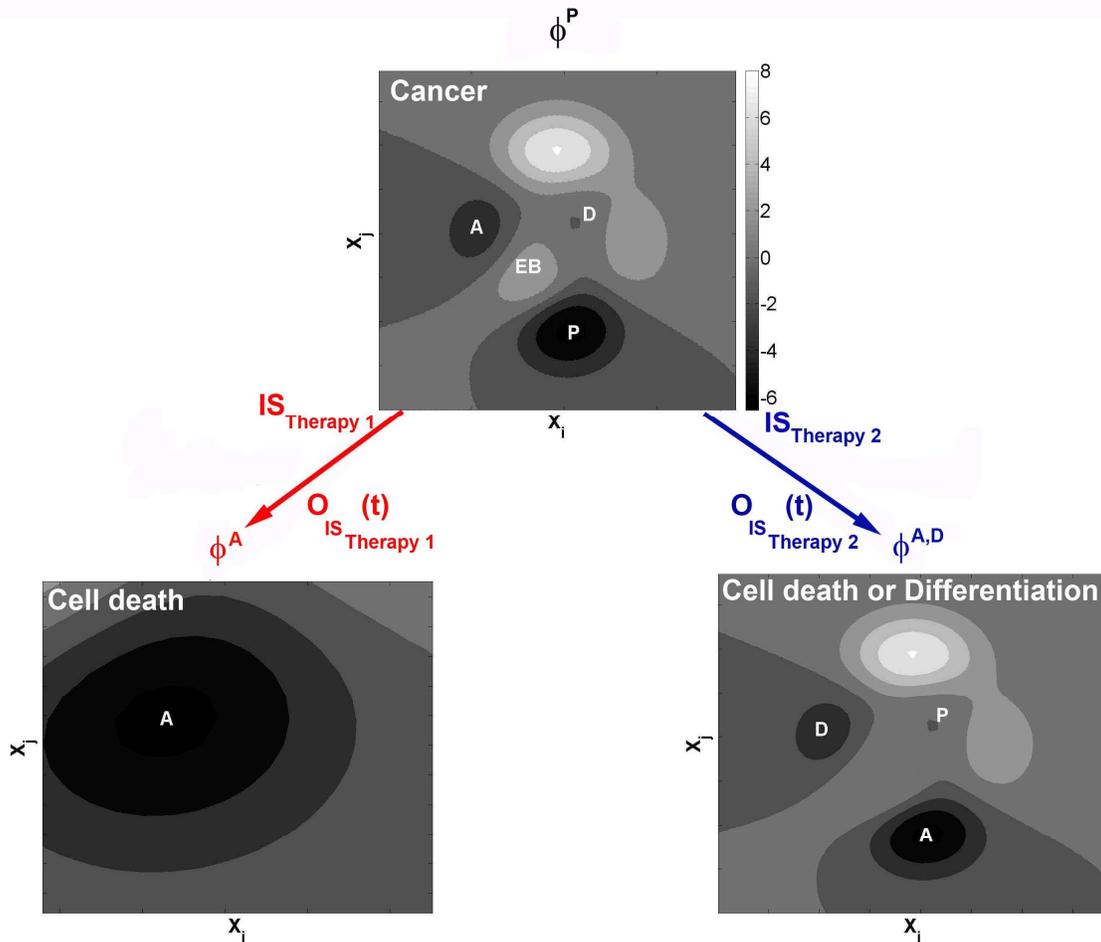


Figure 79 Simplified representation of the possible effects of therapies on the phase space of a cancer cell. A: apoptosis attractor. D-differentiation attractor (one of several possible cell types). P: proliferation attractor (here this point attractor may interpreted as the first stage in the cell cycle). EB-epigenetic landscape barrier. Shades of gray are associated with depth of the epigenetic landscape (see gray scale in upper figure). Each figure is a simple representation of a projection of the high-dimensional concentration phase space (X_1, \dots, X_n) of the proteins involved in the system. Observe that by applying therapy 2 attractor P becomes less probable (smaller basin of attraction and lower strength), but also the epigenetic barrier EB disappears. Therefore, a system in either attractor D or P has access to A. $IS_{Therapy i}$ -inhibition signature therapy. Figures were generated with function *peaks* (MATLAB R2010b).

The distance minimization problem referred to before can be achieved by simulated annealing approaches minimizing the mean square error function (141, 142) or maximizing the mutual information between the initial expression pattern (P) and the target state (A,D). The parametric changes could be performed by targeting the signalling module according to Equation 89. By studying the capacity of a biologically motivated network such as the extended Human

signalling network useful insight could be gained regarding the efficiency of a multiple targeting therapy with variable ε_j 's inducing visible changes at the transcription level, which in a “global phenotype” framework is the most suitable indicator of cell phenotype. The simulations would reveal the importance of connectivity (k_{out}) for message relaying to the nucleus, and the impact of inhibiting specific nodes, with particular activity levels and number of connections, in conjunction with others. This study could extend approaches based only on structural properties that have concluded that a set of weakly modulated nodes are much more efficient in disrupting a network than complete removal of a small number of targets (3, 62), e.g. the hub proteins (see also section 1.3.3).

Subsequently, given the optimal inhibition signature determined by stochastic optimization methods, one could consult information on available drugs, their non-specific interactions or “inhibition signature” (79), and select those closest to the solution determined theoretically. Experimental testing could be performed either with a single agent binding multiple targets or with several. Application of several drugs would, nevertheless, increase the risk of toxicity.

Equation 86
$$O(t) = F_{SS}(I(t), W_{SS}, IS_{Therapy}) \circ F_{TS}(O(t), W_{TS})$$

Equation 87
$$\phi(t) = F_{TS}(W_{TS}, O(t))$$

Equation 88
$$IS_{Therapy} = IS_{Therapy}(\varepsilon_1, \dots, \varepsilon_N, k_{out1}, \dots, k_{outN})$$

Equation 89
$$\varepsilon_i = \frac{[S_j] / K_M'}{1 + [S_j] / K_M'}, \quad K_M' = K_M \left(1 + \frac{[I]}{K_{d_i}}\right),$$

with

I – inhibitor

S_j – substrate j activated by the inhibited kinase i

Equation 90

$$d\phi = \phi^{A,D} - \phi^P$$

In Equation 89 IS_{Therapy} represents the “inhibition signature” of the therapy, which in previous sections was based on small molecule kinase inhibitors. Each inhibition constant ε_i in Equation 89 follows the traditional representation for competitive reversible binding (see also discussion in section 4.2.2). In this expression K_d represents the binding affinity of the drug to a specific protein network node (79), K_M stands for the Michaelis-Menten constant. Also, k_{out} represents the number of output connections of each of the nodes. The drug shifts the threshold K_M to higher concentrations, with the increment depending on the dose $[I]$. We choose as the main connectivity measure k_{out} due to the very nature of how we propose to change the links between components in the optimization process. As was discussed in section 4.2.2 (see Figure 59), the action of an inhibitor such as a small molecule like Gefitinib is on a protein that loses its capacity to transfer a phosphate group to a downstream component. Hence, it's its capacity as an effector that is diminished. Consequently, the number of nodes it connects to (k_{out}) is fundamental.

In the therapy optimization problem, node activation dependent on extracellular inputs and measures of “pathway specificity” (20), which is associated with mean-node activation functions like those represented in Equation 78, should also be incorporated. Recently, a compendium of signals operating through cross-talk between EGF and TNF pathways have contributed extensively to whole network understanding of cell phenotype (96).

Several comments regarding the multitargeting approach with “global phenotype” should be made. First, if a cancer cell has the configuration of attractors in phase space shown in Figure 79 (upper figure), and if a particular multitargeting approach induces a change in this phase space which is favourable, not only the end point but also the intermediate steps during the transformation are relevant. Remember that by choosing an inhibition signature IS_{Therapy} we are in fact selecting the appropriate time-dependent output $O(t)$ which

selected the phenotype ϕ . As we saw in chapter 3, there are path-dependent effects arising from taking into account the signal shape of each of the $O_i(t)$. There, not only noise but the speed at which external signals induce cell fate decision are important for selecting desired attractors. We expect these original mechanisms for cell fate decision to be also relevant for multitargeting therapies. The intermediate steps from the cancer epigenetic landscape to another favouring either phenotype A or D, induced by $O(t)$, which arises from the optimal I_{Therapy} (see Equation 86), should be such that the desired attractors are chosen more frequently in the face of fluctuations (remember the observations of both section 2.2.4 and 2.3). Furthermore, the therapy should also contemplate the effects of timescale separation previously studied in the gene regulatory switch (section 2.3.2). Large differences between signalling processes and transcriptional ones could render the perturbations envisaged by the multitargeting therapies completely irrelevant (86). Recalling our example of a potential undergoing a bifurcation (see Figure 18) and the findings of section 2.3, if the time-scale separation between the evolution of signals S_i and expression dynamics is considerable, the system is expected to remain near the top of the potential barrier for longer. Consequently, the asymmetries arising from the external signals are not reflected in the final distribution over the attractors. On top of the potential barrier noise has the effect of recovering symmetry. As a result, devising a multitargeting therapy is intrinsically linked to choosing appropriate cell fate decision paths in high dimension.

It has been proved that a necessary condition for any perturbation to be translated between fast and slow time-scales is the existence of a mechanism known as “bifurcation cascade”, whereby the perturbation induced on faster elements induce bifurcations that percolate down to slower system’s elements (85). In our paradigm, the optimal inhibition signature at signalling level, fast reactions, should induce a bifurcation at transcriptional level, slower reactions.

One should also comment on the consideration of composite motifs between slow and fast processes (see Figure 78). This renders the problem even harder. Interesting aspects of this type of coupling have shown behaviour such

as adaptation and chaotic itinerancy in systems where metabolism is coupled to signalling (88).

Multiparametric approaches generalizing the theory behind control metabolic analysis have been proposed recently by Rand and coworkers (217). These are based on finding the directions in parameter state space that are stiff, and consequently inform successfully strategic combinations of parameters to affect model behaviour. Also, extensive Monte Carlo simulations have also provided insights into pathway functioning, specifically on the reactions involved in the transient vs. sustained activation of Erk, in an early version of the Epidermal Receptor Pathway (278). These approaches have not, on the other hand, been related back to therapy design. Also, they have not explored mechanisms such as speed-dependent cellular decision making (197), which is one of the main tenets of our work. Yet, as in the work developed Rand and colleagues, which is based on the notion of minimization of a pseudo-energy function or a global quantity, our concept also directs its main strength at finding possible nodes and inhibition constants that synergistically induce “global phenotype” change.

A natural extension to multitargeting approaches is “timed-multitargeting” therapies. This refers to the administration of drugs sequentially. Adding an extra component to the problem might be more successful in dealing with problems where selection between complicated dynamical regimes is necessary. Another aspect that timed-therapies might address is tumour resistance to targeted cancer therapies deriving from mutations. Foo and Michor (83) created a methodology optimizing treatment protocols, through both continuous and pulsed drug administration, inducing minimal risk of resistance. This methodology takes drug toxicity and side effects as model constraints.

On the other hand, selecting the appropriate parameters ε_j (see Equation 89), applied concurrently and not sequentially, is in fact inducing the appropriate combination of profiles $O(t) = (O_1(t), \dots, O_n(t))$ which, as was proved in section 2.3 and 3, is one of driving forces behind attractor selection in noisy genetic circuits. Therefore, we expect a considerable overlap between the two approaches.

4.5 Summary

From the simulation results with the three small molecule kinase inhibitors (Gefitinib, Erlotinib and EKB-569) targeting the Epidermal Growth Factor Receptor (EGFR) we may see that overall non-specific interactions do have additional effects that may be undesired, e.g. rise in ERK-PP signal duration, and that may contribute to proliferative potential. Correlating this with the possible cell fate decision may be considerably more elaborate than the traditional transient/sustained paradigm. Also, not always does the cell line react according to the duality transient/sustained ERK activation. We need, therefore, an overall network behaviour including genetic network activation to understand properly circuit pathway logic and integration of multiple signals in pathways functioning concurrently.

Extensive perturbative studies should be performed on the extended Human signalling network constructed for this project. Information on mutated nodes published by Cui and coworkers (64) on the Map of Human Cancer Signalling could also reveal further aspects of the interplay between mutations and non-specific interactions. Optimization of theoretical inhibition signatures through stochastic search methods, taking into account network connectivity, node inhibition and cellular response, are expected to generate fruitful results regarding the best combination of small molecule kinase inhibitors for efficient drug therapy design.

5 Summary of key results and discussion

This thesis was divided into two main areas dedicated to the effects of external signals on gene regulatory networks and to the effects of multitargeting therapies upstream of the inputs to the genetic program. We will discuss the main aspects of the results extensively analyzed above and link them to wider scientific areas that will benefit from the investigations performed.

5.1 Sensitivity to transient external signals in the presence of noise depends on parameter sweeping rates.

In this thesis we demonstrated both in canonical phase transition models (section 2.2) and in the genetic decision switch the effects of bifurcation parameter sweeping rates and transient external asymmetries on statistical attractor selection in small networks (section 2.3) and high-dimensional ones (section 3.3).

5.1.1 Transient external asymmetries in supercritical pitchfork bifurcations induce higher selectivities in the vicinity of the critical region

For the simple supercritical Pitchfork normal form original findings were reported. The effect of a transient asymmetry with a maximum occurring before, at and after the critical point have showed that the system is capable of retaining memory of the occurrence of an external signal in the presence of strong fluctuations. Fundamentally, transient signals taking place in the vicinity of the bifurcation point at slow enough rates induce higher selectivities (sections 2.2.4, 2.2.5 and 8.5). Moreover, reaching asymmetry maximum amplitudes far below or above the critical region become indistinguishable scenarios (section 2.2.4).

5.1.2 Combination of slow varying external signals induce symmetry breaking and increase attractor selectivity in signalling-gene regulatory decision networks

Combinations of external inputs with different rising times are sufficient to break the symmetry in the genetic circuits. Effectively, the external signals drive the systems through a bifurcation critical region and concurrently create a transient asymmetry favouring the selection of certain attractors. In the presence of fluctuations, time-dependent external signals inducing a slow passage through the critical region enhance attractor selectivity (section 2.3.2). The findings in the canonical model and the genetic regulatory model open very important questions for problems associated with cell fate decision not yet reported in the literature. They are expected to generate high impact in experimental naturally occurring and synthetic genetic circuits with coexistence of different dynamical regimes (see also section 7.2). Likewise, they add to the set of parameters thought to play a role in externally driven genetic network cell fate decision: signal rate, amplitude, duration, frequency and sequence in time.

5.1.3 Time-scale separation decreases sensitivity to differences in external signals in the presence of fluctuations

One of the key aspects identified for the action of external signals on decision genetic circuits is the time-scale separation between phosphorylation and transcription reactions. In line with the observations for the original canonical supercritical Pitchfork bifurcation model, the decision genetic switch with external stimulation also induces selectivities which are dependent on the inverse sweeping speed to time-scale ratio. We have thus a scaling factor, and not an independent relationship, for the response of gene regulatory networks to the exterior. Consequently, faster signals will induce the same selectivity in systems with fast genetic processes as do slow signals in systems with slow transcription reactions (section 2.3.2.1, Figure 44).

5.1.4 Paradigmatic cell fate decision models for crucial mechanism identification

The use of paradigmatic simple models was incredibly useful to isolate successfully inherent properties that underlie attractor selection (section 2.2).

Specifically, the representation of the effect of external signalling inputs to gene regulatory networks was extensively simplified by generating similar characteristics in canonical models. This increased both the speed of numerical simulations and the analytical solvability, and helped deep analysis of the processes in question. Further analytical derivations will link the decision genetic switch and the representative supercritical Pitchfork bifurcation. This will prove to be a strong base for the investigation of parameter sweeping effects in more complex models of regulatory networks. Also, further numerical experiments proving the effects of dynamical hysteresis loops (section 2.2.2) in the genetic switch will be an invaluable addition to systems exhibiting multistability. In multistable systems exhibiting bifurcation cascades, going through the critical region in the forward or reverse direction may select a different outcome from the original starting point. Hence, signals with more complex shapes, e.g. with rising and relaxation times and different maximum and final amplitudes, are expected to show interesting effects and inform on experimentally designed synthetic systems behaviour. Furthermore, experimental differentiation studies, with special emphasis on pattern formation, constitute also a viable avenue that is expected to reveal interesting relationships between the speed with which the system grows, and the organized complexity permitted. Investigation of pattern formation under time-dependent parameters in physics (61) will further provide sound results for applications in symmetry understanding in morphogenesis .

5.1.5 Highdimensional attractor selectivity depends strongly on structure of phase space

The generalization of the parameter sweeping mechanism to high-dimensional space (section 3) proved that it is strongly dependent on phase space structure and the efficiency of noise to induce transitions across potential barriers. Moreover, the capacity of highdimensional genetic circuits to integrate a combination of complex signals is closely linked to the initial condition chosen. Also, it was clearly shown that input combinations that generate the same attractors in a deterministic system have significant differences in final distributions when noise is taken into account (section 3.2 and 3.3). Hence, path-

dependent effects exerted by different complex signals and noise are relevant for attractor selectivity and cell fate decision in highdimensional systems. Furthermore, we expect that in real systems not only the combinatorial complexity of amplitudes of driving signals are important, but also a finer time-dependent discrimination plays a role.

Further studies on the connectivity properties such as the number and nature of the input links to each node should show a fundamental role in the interplay between attractor selection, circuit information storage capacity (72) and the characteristics of external driving signals. Also, further simulations studies are necessary to understand attractor selection principles dependence on parameter sweeping speed, in systems with coexistent of different dynamical regimes. This endeavour will constitute an interesting extension and contribute to the clarification of real selectivity mechanisms present in cells that execute competing differentiation, proliferation or apoptosis programs.

5.2 Effects of non-specificity of drugs may inform the development of relevant multitargeting therapies

5.2.1 Epidermal Growth Factor Receptor (EGFR) pathway output is affected by documented non-specific interactions

Small molecule kinase inhibitors have been successfully applied in Non-Small Cell Lung Carcinoma treatment. Nevertheless, their success has been dependent on a set of mutations being present. We provided a possible proof of principle for the limited success of certain drugs based on numerical simulations of drug “inhibition signatures”. Numerical studies of non-specific interactions on the EGFR pathway dynamics showed that Gefitinib, a drug already approved for Non Small Cell Lung Carcinoma (NSCLC) treatment, has a conflicting action on output (Erk) activation. Gefitinib additionally perturbs a kinase (Gyclin G associated kinase (GAK)) which is involved in internalization of vesicles and consequently of receptors. The EGFR internalization pathway under normal concentration of ligand protects the cell from excessive signalling. Additional

GAK inhibition perturbs this protective pathway, counterbalancing the intended positive inhibition (section 5.2.1). This result shows, nevertheless, avenues of research with emphasis on multitargeting that could enhance the response to therapy.

5.2.2 Multitargeting strategies are expected to induce the correct expression program

Multitargeting therapy design strategies integrated with whole network response patterns may increase correct expression. Several observations should be put forward regarding the distribution of targets in a specific network. Information on targets' connectivity, average activity and correlation with expression pattern should be taken always into consideration.

Whole system targeting strategies based on the premise that the "whole is bigger than the sum of its parts" have been put forward (282). Moreover, the development of databases such as the CMAP, which is dedicated to the collection of gene expression signatures of drugs such as those tested in this thesis, should also reveal significant effects taking place downstream of the drug target. Interesting attempts linking the set of mutations present in specific types of cancers and the connectivity of nodes they strike is also fundamental in selecting optimal strategies in cancer therapy (64). However, although all these research avenues provide us with statistical understanding of the end points of systems disease perturbations, they do not correlate dynamics and structure successfully. The concepts and basic models introduced in this thesis show possible dynamical mechanisms that can demonstrate biases in, for example, cancer attractor selection.

6 Conclusion

This thesis addressed the problem of cell fate decision in the context of network connectivity and dynamics. Several paradigmatic signalling-gene regulatory networks, such as the genetic decision switch, were crucial in identifying never before reported mechanisms. Additionally, canonical models from statistical physics were successfully used in devising analytical expressions for attractor selectivity in systems with time-dependent parameters. The most striking feature of our findings is the crucial impact of the rate with which the time-dependent bifurcation parameters in genetic systems are changed. In the presence of small asymmetries and fluctuations, slow passage through the critical region increases substantially specific attractor selection. This has strong implications for the cell fate decision problem since cell phenotype in stem cell differentiation, cell cycle progression, or apoptosis studies, has been successfully identified as an attractor of a whole network expression process induced by signalling events.

7 Future work

7.1 *Decision genetic switch with external stimulation*

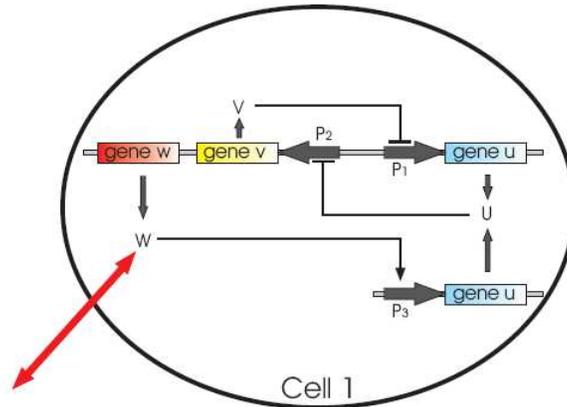
- **Understand the effects of combinations of external inputs, with representative shapes, on the paradigmatic decision switch (109, 197), with time-scale separation, and undergoing bifurcation.**

The combination of signals (S_1 and S_2 , Figure 19) will follow representative shapes observed in vitro and characterized by rising time, maximum and final amplitudes and relaxation time (see Figure 20). This will increase the scope of the work developed in this thesis, where only rising time was studied. The signal shapes to be tested should give rise to a phenomenon known as dynamical hysteresis (107), which has not been explored in genetic systems and could have important consequences for cell fate decision. Further simulation experiments on this paradigmatic circuit will be performed by combining one of the inputs, S_1 or S_2 , with a periodic signal. The results of this step will inform on the possibility of an effect known as stochastic resonance which is observed in canonical models of phase transitions in statistical physics (198), and is expected to be fundamental for gene regulatory network attractor selection.

A different approach to treat the system analytically will also be performed by transforming the original system of equations (see Equation 60-Equation 63) (109, 197) into a simpler two dimensional trigger equation system and separate it into two branches as it has been done by Lindner (174).

7.2 Highdimensional networks with coexistence of dynamical and point attractors

A



B

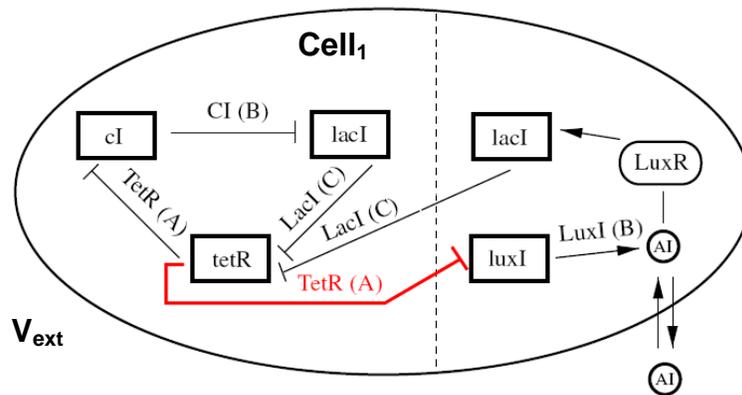


Figure 80 Highdimensional synthetic oscillatory genetic networks. A) Decision genetic switch with external simulation (cell to cell coupling) and relaxation oscillation properties. For details see (165). B) Repressilator with cell-cell phase repulsive coupling. The modified version of the original repressilator (77) involves an inhibitory connection between *tetR* and *luxI* genes which leads to a positive loop of TetR on itself. For details see (256).

- **Identify the effects of parameter sweeping with ramp like and/or periodic profile, on attractor selectivity, in coupled synthetic gene network motifs with multistability and coexistence of dynamical and point attractors. Preliminary study cases:**
 - **Decision genetic switch with external simulation (cell to cell coupling) with relaxation oscillator properties (164)**

AND

- **Repressilator with cell-cell phase repulsive coupling (256).**

Koseska and Zaikin studied a model of a hysteresis based relaxation oscillator by combining two engineered genetic networks, the toggle switch (lacI gene and cl857 gene from *E.coli*, u and v in Figure 80 A, respectively) and an intercellular communication system (quorum sensing system from *Vibrio fischeri*, w in Figure 80 A). Studies of the influence of coupling strength concluded that intercellular communication is essential in generating a wealth of regimes: anti-phase oscillations, asymmetric oscillations, inhomogeneous oscillations and multiple oscillatory and steady-state clustering. The coexistence of several types of regimes under parameter variations calls for the question of how should these parameters be varied as to induce the desired attractor. In natural biochemical systems parameters don't change in a step like fashion. Varying coupling strength according to ramp like signals and/or periodic signals will allow us to understand how the system reacts to external changes, and how the selected attractor is dependent on the bifurcation inherent to the phase transition occurring in the system (recall the example previously outlined of the double well potential associated to a supercritical Pitchfork bifurcation). Combined effect of ramp like and periodic parameter sweeping has been published for a bistable potential with the occurrence of an interesting effect similar to stochastic resonance (198).

The other motif identified in the literature as an essential paradigm is the repressilator (256). Ullner and Zaikin observed that repressive coupling of repressilators (lacI gene from *E.coli*, tetR gene from the tetracycline-resistant

transposon Tn 10 and *cl* gene from λ bacteriophage, see Figure 80 B) through the quorum sensing pathway also revealed coexisting regimes under different strengths. The repressilator is a transformed version of the original repressilator model created by Elowitz and coworkers (77).

Possible investigations:

- Parameter sweeping experiments with different speeds will be performed in both models by changing the diffusion of the auto inducer (AI, see Figure 80) molecule across the membrane, varying the number of cells in the system or parameters controlling expression rates. Varying the number of cells is a way of representing action of cell division in the highdimensional system. The intensity of noise is known to be related with the population size. It will be interesting to understand how the speed of cell number increase influences the noise in the system and consequently the dynamical regimes selected. This will be a considerable new approach to the problem of selectivity of each of the coexistent dynamical regimes in the presence of noise, with all parameter variations being motivated by the experimental literature. Parameter sweeping experiments driven by oscillatory signals may also reveal important mechanisms for chronotherapy or cell cycle regulation, and contribute with invaluable insight to treatment strategies for the so-called “dynamical diseases”.
- One characteristic of integrative signalling-transcription biological models is the existence of considerable time-scale separation. We plan to test the translation of signal characteristics between fast and slow time-scales. It has been proven that there is a limit and a mechanism known as “bifurcation cascade” for the successful perturbation translation (86). This translation of perturbations is an essential aspect to take into consideration in devising targeting

therapies. Not only a multitargeting approach should be taken into account (196), but it should be devised as to circumvent the intrinsic lack of sensitivity between signalling and gene expression networks when time-scale separation is large. The simulations will be performed under a Langevin type of approach, with multiplicative noise, and using the Gillespie approach.

- Small differences between the parameters mentioned above for each cell will allow us to test the selectivity of each of the dynamical regimes in the presence of intrinsic asymmetries, when the systems are undergoing a bifurcation, and in the presence of noise. Several bifurcation types are inherent to the appearance of each of the dynamical regimes found for both the systems represented in Figure 80: limit point, hopf (163, 164) and torus (163) bifurcation. Further bifurcation analysis will be performed with the purpose built software packages XPPAUT^s and MATCONT^t.
- **Identify the dynamical regimes of a 30 node transcription factor network created under the FANTOM4^u initiative. Highlight the relevance of parameter sweeping experiments induced by external forcing, under small asymmetries and noise, in the selection of expression programs or cell fate decision attractors.**

The FANTOM4 program is dedicated to understanding mechanisms in monocytic differentiation in the acute myeloid leukaemia cell line THP-1. The transcription factor network responsible for monocytic differentiation will be a natural departure from the synthetic biology examples previously highlighted, and

^s <http://www.math.pitt.edu/~bard/xpp/xpp.html>

^t <http://www.matcont.ugent.be/>

^u <http://fantom.gsc.riken.jp/4/>

an invaluable addition to the type of attractors present in a mammalian cell network.

All simulations planned for this particular network and the previously discussed examples will be performed both under a Langevin type of approach to stochastic differential equations (with multiplicative noise scaling as the square root of the concentration) and using the Gillespie algorithm approach.

The modelling approach for the dynamics of the FANTOM4 network will be based on a mean-field approximation for promoter occupation (11) which was explored in this thesis. This will allow for several mean activation functions through manipulation of simple parameters. Identification of signalling entry points to the transcription factor network will be performed by matching the nodes published through the FANTOM4 initiative and the nodes of the Human signalling network (64).

8 Annex

8.1 Control coefficients for the Epidermal Growth Factor Receptor Network

	Small perturbations (10^{-6})			Large perturbation (-0.5)			Large perturbations (0.5)		
	Amplitude	Duration	Area	Amplitude	Duration	Area	Amplitude	Duration	Area
1	0.0002	0.0257	0.0264	0.0003	0.0289	0.0292	0.0001	0.0233	0.0245
2	-0.001	0.0322	0.0626	0.0016	0.1265	0.179	-0.0021	-0.0055	0.0122
3	0.0236	0.1197	0.2127	0.085	0.2358	0.4065	0.012	0.0792	0.1403
4	0.0026	0.2846	0.4267	0.0073	0.3595	0.5294	0.0012	0.2488	0.3778
5	0.0004	0.0051	0.0075	0.0022	0.01	0.0155	0	0.0034	0.0047
6	-0.0877	-0.9487	-1.4714	-0.0491	-1.7354	-2.6498	-0.6217	-0.6079	-1.1578
7	0.0894	0.9486	1.4726	1.5186	0.9727	2.1294	0.0411	0.8454	1.2942
8	-0.008	-1.0257	-1.124	-0.0085	-1.5643	-1.6993	-0.0077	-0.7765	-0.8511
9	0.0809	0.1721	0.4389	0.7294	0.2552	1.1425	0.0478	0.1141	0.3115
10	-0.0803	-0.1718	-0.4353	-0.0614	-0.1555	-0.5006	-0.1642	-0.1507	-0.4173
11	0	-0.0029	-0.0039	0	-0.0029	-0.0039	0	-0.0029	-0.0039
12	-0.0209	-0.2677	-0.4279	-0.021	-0.3469	-0.5411	-0.021	-0.2219	-0.3602
13	0.0008	-0.1713	-0.1846	0.0008	-0.1713	-0.1892	0.0008	-0.1654	-0.1766
SUM (1,12)	0.0029	-0.1223	-0.1888	0.0643	-0.0822	-0.1054	-0.0089	-0.1194	-0.1954
SUM (1,12,13)	0.0037	-0.2936	-0.3734	0.0651	-0.2535	0.5192	-1.00E-04	0.0779	0.2134

Table 4 Control coefficients for amplitude, duration and integrated response signal (area) for the processes involved in the system. Taken from (125). See also Table 5.

Process

1	Receptor activation (v1,v2,v3,v10,v11,v12,v13)
2	Recruitment (v16,v17,v22-25,v32-41,v63,v64,v69-772,v79-83)
3	Ras activation (v18,v19,v26,v27,v65,v66,v73,v74)
4	Ras inactivation (v8,v9,v14,v20,v21,v30,v31,v67,v68,v77,v78)
5	Ras activates Raf (v28,v29,v75,v76)
6	Raf dephosphorylation (v42,v43)
7	Raf phosphorylates MEK (v86-89)
8	MEK dephosphorylation (v90-93)
9	ERK phosphorylation (v94-97)
10	ERK dephosphorylation (v98-101)
11	Negative feedback loop (v126-131, v143-148)
12	Internalization (v5,v6,v7,v102-v125)
13	Degradation (v60,v61,v62)

Table 5 Description of the processes and respective rates (125, 230) associated with each line in Table 4. See also Figure 57.

8.2 Inhibition signature maps for small molecule kinase inhibitors

ZD-6474

CI-1033

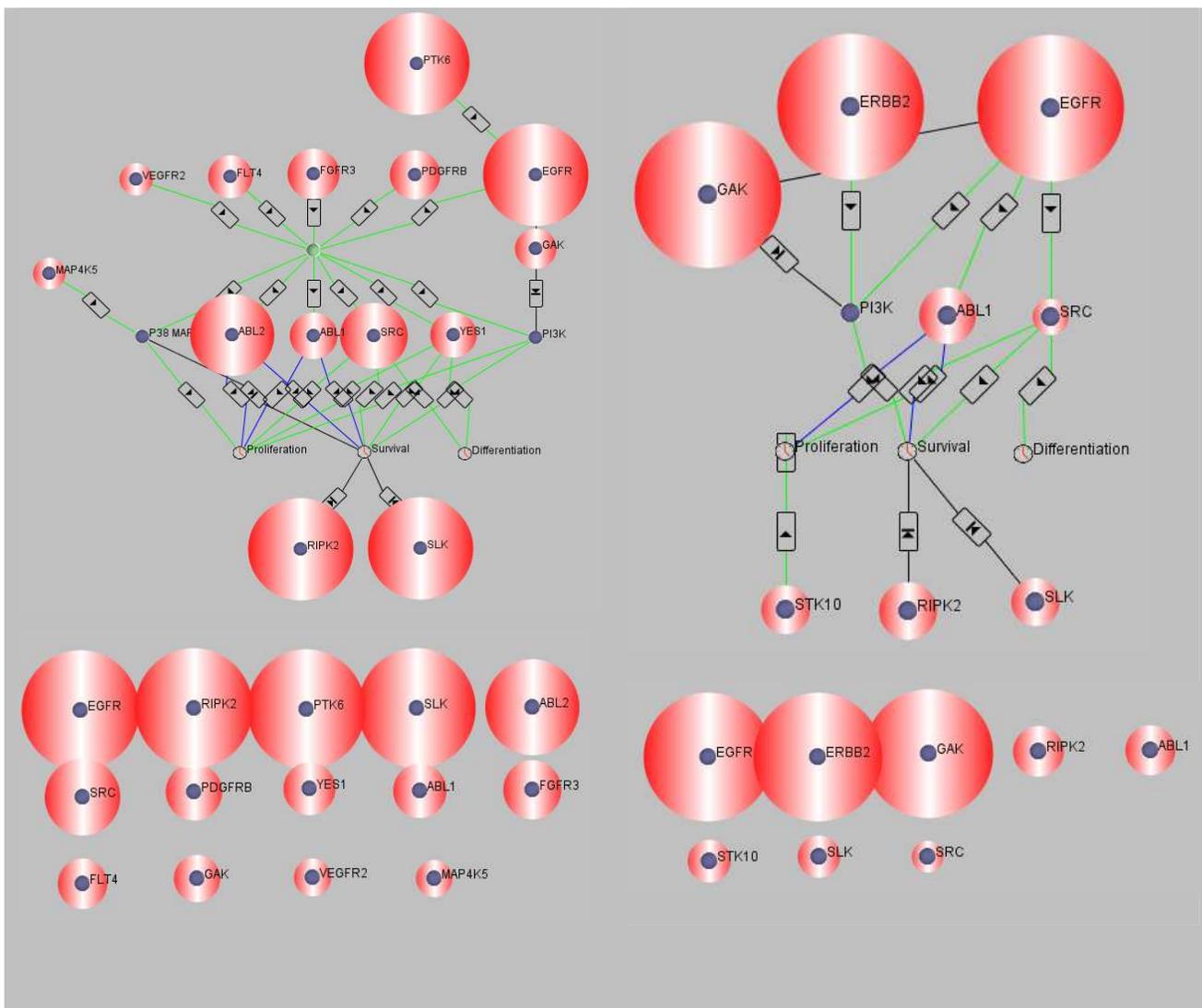


Figure 81 Inhibition signatures of 2 other small molecule kinase inhibitors targeting the EGFR. Red halo is proportional to inhibition strength. Values extracted from (79). Diagrams were constructed with the Scipath software developed by Dr.Sylvia Nagl and collaborators. The network representation involves kinases collected from several databases such as Biocarta, CellSignal, Kinase.com, SwissProt and OMIM. These were used to link the kinases tested by Fabian and coworkers to relevant phenotypic manifestations in Cancer. The simple map gives us an overall picture of the distribution of off-target interactions. Green arrows stand for stimulation and black arrows for inhibition. Additionally, cellular responses to upstream kinases are represented as clocks and links in blue.

8.3 Binding affinities for small molecule kinase inhibitors

	Kinase Inhibitor					
	IRESSA /Gefitinib	TARCEVA /Erlotinib	ZD- 6474	CI-1033	GW- 2016	EKB- 569
ABL1	0	7700	2700	3400	0	2200
ABL2	0	3000	1300	0	0	1600
ACK1	0	0	0	0	0	8900
BTK	0	0	0	7500	0	0
CLK2	0	0	0	0	0	7000
CSNK1E	0	0	0	0	0	970
EGFR	18	14	170	14	55	10
ERBB2	0	0	0	84	110	770
FGFR3	0	0	2400	0	0	0
FLT4	0	0	3000	0	0	0
GAK	70	400	3300	440	0	15
MAP3K4	0	0	0	0	0	1300
MAP4K5	0	0	5100	0	0	37
MNK2	3600	0	0	0	0	0
NEK2	0	0	0	0	0	1400
PDGFRB	0	0	2500	0	0	0
PKMYT1	0	0	0	0	0	9300
PTK6	0	0	330	0	0	0
RIPK2	8000	4100	310	3300	0	0
SLK	0	1100	960	4400	0	2500
SRC	0	0	1700	7600	0	1200
STK10	8700	830	0	4300	0	1100
STK17A	0	0	0	0	0	570
TNIK	0	0	0	0	0	450
VEGFR2	0	0	4700	0	0	0
YES1	0	0	2800	0	0	8400

Table 6 Binding constants for the six EGFR inhibitors to a panel of kinases. Binding Constant numbers (Kd) x 10⁻⁴ (uM). Values extracted from (79).

8.4 Numerical integration method

Numerical simulation of the equations presented in this project is performed through a Heun method (209). This assumes the following form:

$$\begin{aligned} \text{Equation 91} \quad X_i(t + \Delta t) &= X_i(t) + \frac{\Delta t}{2} \times [F(\tilde{X}_i(t)) + F(X_i(t))] \\ \tilde{X}_i(t + \Delta t) &= X_i(t) + \Delta t \times F(x(t)) \end{aligned}$$

If we take the Langevin form of the equations the predictor will be naturally affected by the diffusion term $G(x(t-\tau))$.

$$\begin{aligned} \text{Equation 92} \quad X_i(t + \Delta t) &= X_i(t) + \frac{\Delta t}{2} \times [F(\tilde{X}_i(t)) + F(X_i(t))] \\ &+ \frac{1}{2} \times [G(\tilde{X}_i(t + \Delta t)) + G(X_i(t))] \Delta W_i(t) \\ \tilde{X}_i(t + \Delta t) &= \tilde{X}_i(t) + \Delta t \times F(\tilde{X}_i(t)) + G(\tilde{X}_i(t)) \times \Delta W_i(t) \end{aligned}$$

Δt is the time step and ΔW_i are Gaussian random numbers. The first two moments of the Gaussian process are given by Equation 93.

$$\text{Equation 93} \quad \langle \Delta W_i(t) \rangle = 0, \quad \langle \Delta W_i(t) \Delta W_j(t) \rangle = 2D \delta_{ij} \Delta t$$

Two aspects motivate the use of the Heun scheme for the numerical integration of the SDEs in sections 2 and 3 (209):

- The Heun method gives us Stratonovich solutions of the SDEs without the deterministic drift term being modified.
- The deterministic terms are numerically integrated to a second order accuracy in Δt , thus turning the chosen method into a more stable alternative than the simplistic Euler approach.

Additionally, the Heun method is quite simple compared to more elaborate schemes. It provides, nevertheless, accurate results (209).

8.5 Selectivity expressions for the supercritical pitchfork bifurcation for specific cases of external asymmetry $g(t)$ and critical parameter $\lambda(t)$

Here we determine the expressions for the selectivity of the upper branch of steady state solutions of a bistable potential undergoing bifurcation, for two additional cases to the one that was solved in the main text (see section 2.2.4.2). Both the bifurcation parameter and the external field asymmetry are time-dependent. Additionally, the intensity σ of an additive source of noise is also constant.

Equation 94

$$P_{up} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^N e^{-\frac{x'^2}{2}} dx'$$

8.5.1 Maximum asymmetry before the bifurcation point

For $\gamma_{g1} = c_1 \gamma_\lambda = 2g_{\max} \gamma_\lambda$; $\gamma_{g2} = c_2 \gamma_\lambda = \frac{2}{3} g_{\max} \gamma_\lambda$, the number of standard deviations that the peak of the distribution is displaced from the position of the unstable branch of solutions is given by Equation 95.

Equation 95

$$\begin{aligned}
 & \frac{2g_{\max}}{3} \left(-4e^{\frac{-1}{8\gamma_\lambda}} + 3e^{\frac{-1}{2\gamma_\lambda}} + e^{\frac{-K^2}{2\gamma_\lambda}} \right) + \\
 N = & \frac{\frac{g_{\max}}{3} \left(-2\operatorname{erf}\left(\frac{1}{2\sqrt{2\gamma_\lambda}}\right) + 3\operatorname{erf}\left(\frac{1}{\sqrt{2\gamma_\lambda}}\right) + \operatorname{erf}\left(\frac{K}{\sqrt{2\gamma_\lambda}}\right) \right) \sqrt{\frac{2\pi}{\gamma_\lambda}}}{\left(\frac{\sigma^2}{2}\right)^{1/2} \left(\operatorname{erf}\left(\frac{1}{\sqrt{\gamma_\lambda}}\right) + \operatorname{erf}\left(\frac{K}{\sqrt{\gamma_\lambda}}\right) \right)^{1/2} \left(\frac{\pi}{\gamma_\lambda}\right)^{1/4}} \\
 \leq & \frac{g_{\max}}{\left(\frac{\sigma^2}{2}\right)^{1/2}} \left(\frac{\pi}{\gamma_\lambda}\right)^{1/4}
 \end{aligned}$$

The probability of reaching the upper branch is consequently given by Equation 96.

Equation 96

$$\begin{aligned}
 P_{up} = & \frac{1}{2} \left[1 + \operatorname{erf} \left[\sqrt{2} \frac{g_{\max}}{\sigma} \left(\left(2e^{\frac{-1}{2\gamma_\lambda}} - 4e^{\frac{-1}{8\gamma_\lambda}} + e^{\frac{-K^2}{2\gamma_\lambda}} \right) \frac{1}{3} \left(\frac{\gamma_\lambda}{\pi} \right)^{1/4} + \frac{\sqrt{2}}{3} \left(\frac{\pi}{\gamma_\lambda} \right)^{1/4} \right) \right] \right] \\
 \leq & \frac{1}{2} \left[1 + \operatorname{erf} \left[\frac{g_{\max}}{\sigma} \sqrt{2} \left(\frac{\pi}{\gamma_\lambda} \right)^{1/4} \right] \right] \quad \text{for } \gamma_\lambda < 1
 \end{aligned}$$

As verified in section 2.2.4.2 for the case when the maximum asymmetry is reached exactly at the critical point, also here the selectivity P_{up} is smaller than when asymmetry g is constant.

8.5.2 Maximum asymmetry after the bifurcation point

For $\gamma_{g1} = c_1 \gamma_\lambda = \frac{2}{3} g_{\max} \gamma_\lambda$; $\gamma_{g2} = c_2 \gamma_\lambda = 2 g_{\max} \gamma_\lambda$, the number of standard deviations that the peak of the distribution is displaced from the position of the unstable branch of solutions is given by Equation 97.

Equation 97

$$\begin{aligned}
 & \frac{2g_{\max}}{3} \left(-4e^{\frac{-1}{8\gamma_\lambda}} + 3e^{\frac{-1}{2\gamma_\lambda}} + e^{\frac{-K^2}{2\gamma_\lambda}} \right) + \\
 & \frac{g_{\max}}{3} \left(-2\operatorname{erf}\left(\frac{1}{2\sqrt{2\gamma_\lambda}}\right) + 3\operatorname{erf}\left(\frac{1}{\sqrt{2\gamma_\lambda}}\right) + \operatorname{erf}\left(\frac{K}{\sqrt{2\gamma_\lambda}}\right) \right) \sqrt{\frac{2\pi}{\gamma_\lambda}} \\
 N = & \frac{\left(\frac{\sigma^2}{2}\right)^{1/2} \left(\operatorname{erf}\left(\frac{1}{\sqrt{\gamma_\lambda}}\right) + \operatorname{erf}\left(\frac{K}{\sqrt{\gamma_\lambda}}\right) \right)^{1/2} \left(\frac{\pi}{\gamma_\lambda}\right)^{1/4}}{\left(\frac{\sigma^2}{2}\right)^{1/2} \left(\frac{\pi}{\gamma_\lambda}\right)^{1/4}} \\
 \leq & \frac{g_{\max}}{\left(\frac{\sigma^2}{2}\right)^{1/2}} \left(\frac{\pi}{\gamma_\lambda}\right)^{1/4}
 \end{aligned}$$

The probability of reaching the upper branch is consequently given by Equation 98.

Equation 98

$$\begin{aligned}
 P_{up} = & \frac{1}{2} \left[1 + \operatorname{erf} \left[\sqrt{2} \frac{g_{\max}}{\sigma} \left(\left(e^{\frac{-1}{2\gamma_\lambda}} - 4e^{\frac{-1}{8\gamma_\lambda}} + 3e^{\frac{-K^2}{2\gamma_\lambda}} \right) \frac{1}{3} \left(\frac{\gamma_\lambda}{\pi}\right)^{1/4} + \frac{\sqrt{2}}{3} \left(\frac{\pi}{\gamma_\lambda}\right)^{1/4} \right) \right] \right] \\
 \leq & \frac{1}{2} \left[1 + \operatorname{erf} \left[\frac{g_{\max}}{\sigma} \sqrt{2} \left(\frac{\pi}{\gamma_\lambda}\right)^{1/4} \right] \right] \quad \text{for } \gamma_\lambda < 1
 \end{aligned}$$

Again, the same observations hold for sweeping through the critical region when the maximum external asymmetry is reached after the bifurcation point. The probability of reaching the upper branch of solutions, P_{up} , is smaller than the probability attained with constant g .

8.6 Parameter space analysis of the gene regulatory decision switch

The available states to the regulatory circuit studied in section 2.3 can be fully characterized through exploration of combinations of parameters chosen to model epigenetic regulation: magnitude of basal production rate (η_i), strength of auto-activation link (d_i), magnitude of threshold of auto-activation (b_i) and cross-repression (g_i) (see system Equation 60–Equation 63 and Figure 82). Prediction of circuit behaviour through bifurcation analysis is a very useful tool and can be fundamental in experimental studies (15, 95, 166, 207). Bistability or multistability in genetic circuits has been seen as a fundamental property for flexible signal detection and classification; either for efficient differential processing of signal duration and amplitude (109), or for frequency selectivity (236). Hence, understanding how each state space scenario arises from parameter analysis is one step closer to clarifying the selection mechanisms previously investigated, i.e. speed-dependent decision making (section 2.3). All bifurcation analysis studies reported ahead were conducted in XPPAUT^v. Although in the simulations reported before (see section 2.3) we used the simplest bifurcation possible for the genetic circuit, it is important to show the multitude of scenarios available.

8.6.1 Isolated genetic decision switch

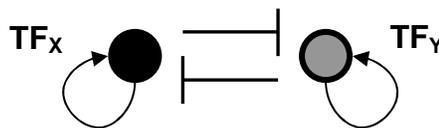


Figure 82 Isolated decision genetic switch. See also Figure 39.

In a first iteration of the bifurcation analysis, the existence of multistability will be tested against parameters characterising the circuit without external stimulation S_1 or S_2 . On the other hand, we will retain all other assumptions regarding circuit regulation, highlighted in section 2.3.1.1. Therefore, although

^v <http://www.math.pitt.edu/~bard/xpp/xpp.html>

dimerization reactions still need to take place, the resultant dimers binding to the promoter regions are X^2 and Y^2 , and not $(X^a)^2$ and $(Y^a)^2$ (see Equation 62 and Equation 63). In section 8.6.2 we return to the full circuit and study the effects of cross-talk between pathways. Also, in all the subsequent sections a considerable percentage of the bifurcation diagrams created are motivated by an attempt to understand the effect of external stimulation on the isolated circuit.

8.6.1.1 Symmetric circuit: large b_i/g_i ratios induce multistability

In biology, we mostly encounter asymmetry between strengths of system's links. Yet, studying a symmetric circuit allows us to understand the magnitude necessary for the chosen parameters to generate several regimes.

Combinations of parameters $d=d_X=d_Y$, $b=b_X=b_Y$, $g=g_X=g_Y$ in a symmetric circuit, are capable of enabling bistability and higher order stable states. As can be seen in Figure 83, choosing either d or b as the bifurcation parameter, it is possible to induce in the system a supercritical Pitchfork bifurcation (Figure 83 A), or a subcritical Pitchfork bifurcation (Figure 83 B), respectively. Parameter g also induces a supercritical transition, although the stable states arising from this have lower concentration values. In our analysis, we show only diagrams for variable X . Diagrams for Y would lead to the same conclusions.

The action of signals S_1 and S_2 in the original system with external stimulation (see Figure 39) is equivalent to assuming b and g dependent on time. Hence, signals drive the circuit through the bifurcation points. Throughout our work in previous sections, we assumed a multiplicative action of the signals S_1 and S_2 on the system's parameters. This was performed through the phosphorylation of X/Y (see Equation 60 and Equation 61). Differences between signals, i.e. external asymmetry, induce changes in the bifurcation diagram according to their respective time-dependent profile.

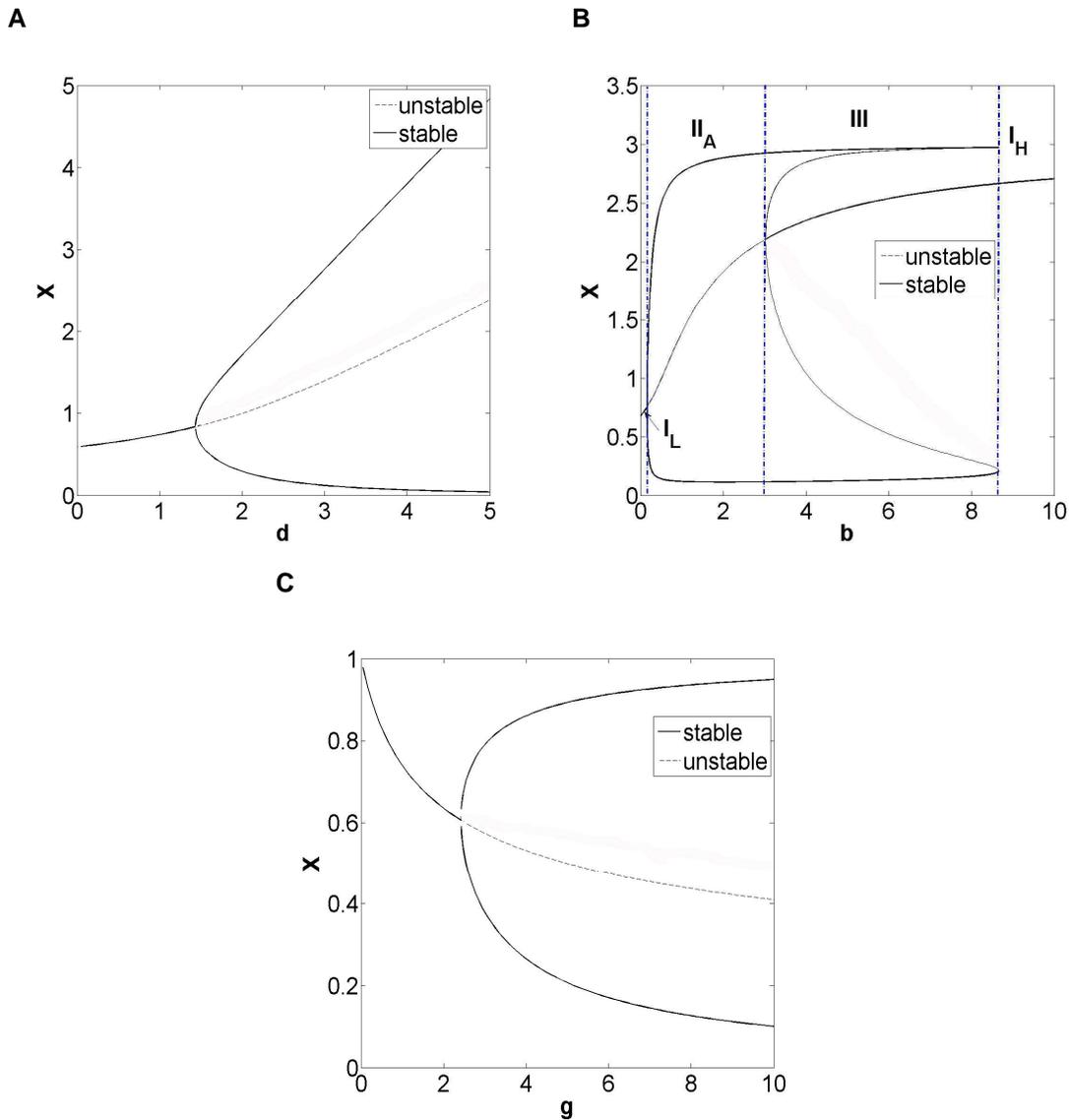


Figure 83 Bifurcation analysis of the symmetric genetic decision switch. A) For $d_x=d_y \rightarrow d$ with $\eta_x=\eta_y=1$, $b_x=b_y=1$ and $g_x=g_y=1$ (see Equation 60 to Equation 63) showing a supercritical Pitchfork bifurcation. **B)** For $b_x=b_y \rightarrow b$ with $\eta_x=\eta_y=1$, $d_x=d_y=3$ and $g_x=g_y=2$ showing a subcritical Pitchfork bifurcation. II_A: bistability region with anti-symmetric states (H,L) and (L,H) branches, where H corresponds to high concentration values, of X or Y, and L to low values. III: tristability region. I_{L,H}: monostability region with low/high X concentration values, respectively. **C)** For $g_x=g_y \rightarrow g$ with $\eta_x=\eta_y=1$, $b_x=b_y=1$ and $d_x=d_y=1$, also showing a subcritical transition. $[TF_x]=X$.

Extending the analysis shown in Figure 83 to 2 dimensions it is possible to further understand the regimes available to the circuit (Figure 84). Exploring the available states for several combinations of parameters one observes that for

multistability (III, see Figure 84 A) to occur it is necessary for both the magnitude of the auto-activation parameter ($d=d_X=d_Y$) and the threshold for initiation of transcription ($b=b_X=b_Y$) to be high. Alternatively, if we assume d to be in the region of bistability (see Figure 83 A), b needs to be relatively large when compared to g to induce multistability (III, see Figure 84 C). This region in state space induces 3 stable states, two with approximately anti-symmetric concentrations, (H,L) and (L,H), and a third with symmetric but high concentrations of both X and Y (see Figure 84 C). Other combinations of b and d , for constant g , either induce bistability (II_A), with anti-symmetric (H,L) or (L,H) concentrations for (X,Y), or monostability with either low symmetric concentration values (I_L) or high (I_H) (see also Figure 83 B). The basal production level η (see Equation 62 and Equation 63), when raised, tends to destroy stable states with low concentration values (109) (figure not shown).

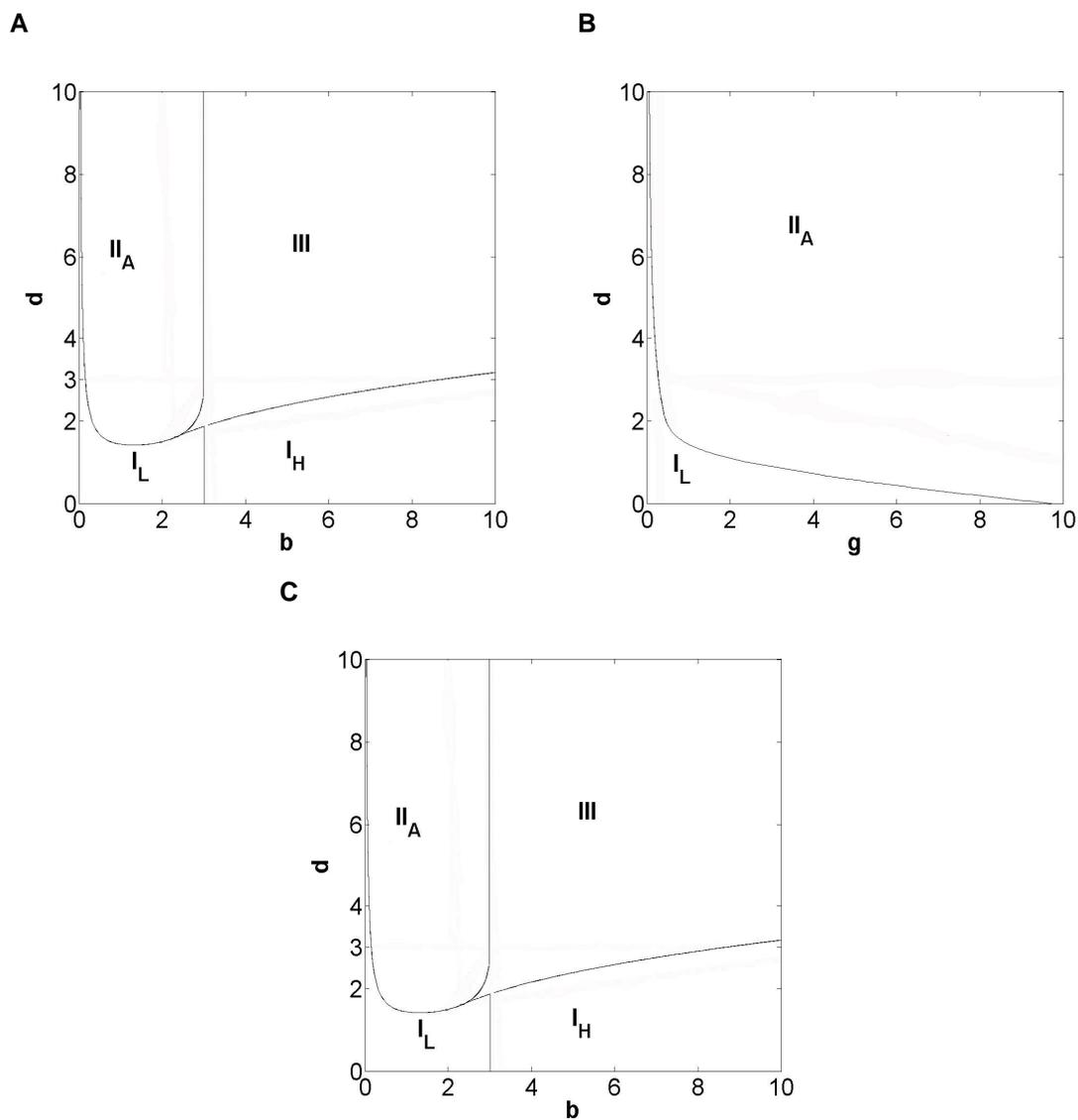


Figure 84 Bifurcation diagrams formed by extending the diagrams in Figure 83. A) For $b_x=b_y \rightarrow b$ and $d_x=d_y \rightarrow d$ with $g_x=g_y=1$ and $\eta_x=\eta_y=1$. B) For $g_x=g_y \rightarrow g$ and $d_x=d_y \rightarrow d$ with $b_x=b_y=1$ and $\eta_x=\eta_y=1$. C) For $b_x=b_y \rightarrow b$ and $g_x=g_y \rightarrow g$ with $\eta_x=\eta_y=1$, $d_x=d_y=3$. II_A: bistability with anti-symmetric stable states being (H,L) and (L,H), where H corresponds to high concentration values, of X, and L to low values. III: tristability region. I_{L,H}: monostability region with low/high X concentration values, respectively. $[TF_x]=X$.

8.6.1.2 Asymmetric circuit: large b_i/g_i generate multistability

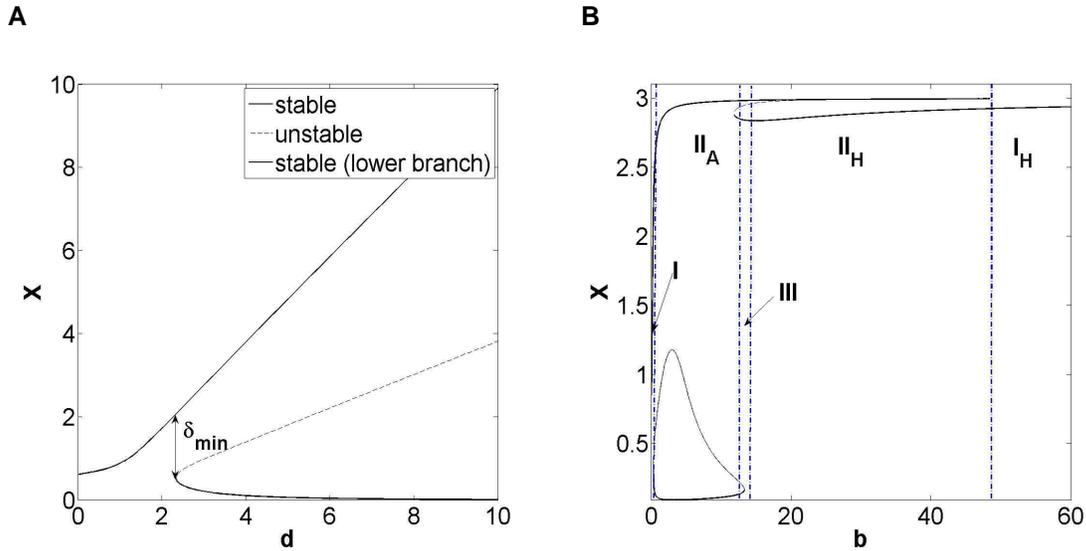


Figure 85 Bifurcation diagrams in the presence of asymmetric basal rates. A) For $d_x=d_y \rightarrow d$ with $\eta_x=1$, $\eta_y=0.8$, $b_x=b_y=1$ and $g_x=g_y=1$. $\delta_{\min} = (X_{\text{upper branch}} - X_{\text{lower branch}})$ = minimum distance between branches at the critical point. **B)** For $b_x=b_y \rightarrow b$ with $\eta_x=1$, $\eta_y=0.8$, $d_x=d_y=3$ and $g_x=g_y=2$. III: tristability region. $I_{L,H}$: monostability region with low/high X concentration values, respectively. $II_{A,H}$: bistability region with anti-symmetric states or high concentration states, respectively. $[TF_x]=X$.

Asymmetry in the original circuit can arise as imposed momentarily by the external signals S_1 and S_2 (considered in section 2.3), as internal differences in parameters η_i , d_i , b_i or g_i , or a combination of both. As can be observed in Figure 85 A, internal asymmetries between η_x and η_y induce a disconnection between branches in the diagram with $d=d_x=d_y$ as the bifurcation parameter. The minimum distance between branches (δ_{\min}) at the critical point is proportional to the ratio η_x/η_y (data not shown). If we choose to analyse the effects of asymmetries between basal expression rates when b is the critical parameter, we observe that additionally to the disconnection between branches, the region III is shrunk considerably (compare Figure 83 B with Figure 85 B).

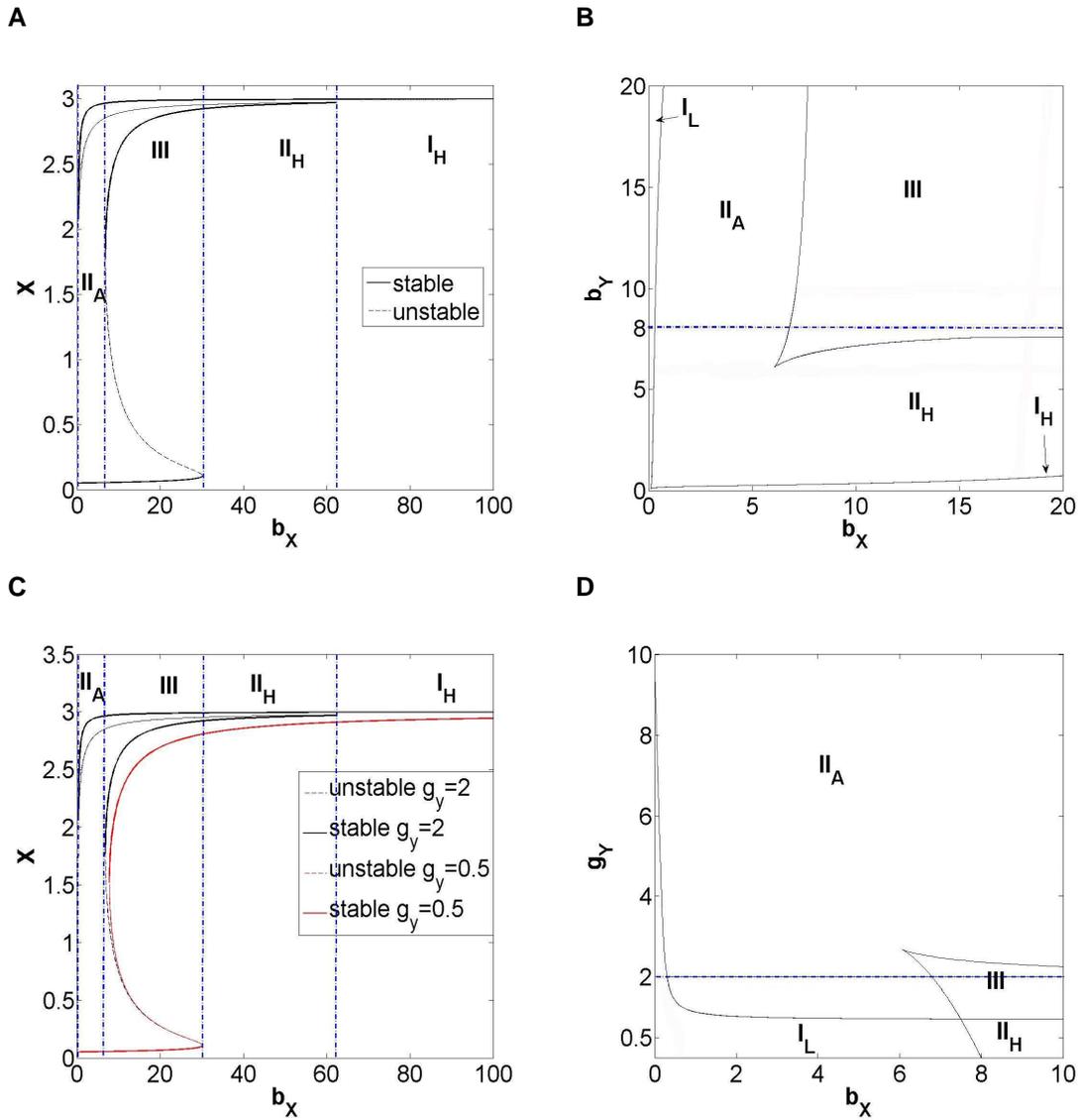


Figure 86 Bifurcation diagrams for asymmetric thresholds of self-activation and cross-inhibition parameters. A) For b_x with $b_y=8$ and $g_x=g_y=2$. B) 2D diagram corresponding to an extension of A). C) For b_x with $b_y=8$. Two cases are shown: $g_x=g_y=2$ and $(g_x,g_y)=(2,0.5)$. D) 2D diagram corresponding to an extension of C). For all figures $\eta_x=\eta_y=1$; $d_x=d_y=3$ (see Equation 60 to Equation 63). The transition (III \rightarrow I_H) to monostability in both A and D stems from further developments in the position of the borders in B and D, respectively, after $b_x=20$ (not shown). III: tristability region. I_{L,H}: monostability region with low/high X concentration values, respectively. II_{A,H}: bistability region with anti-symmetric states or high concentration states, respectively. $[TF_x]=X$.

Because the external signals S_1 and S_2 , in the original circuit (section 2.3), operate multiplicatively on parameters b_x , b_y , g_x and g_y , we will restrict our analysis, until the end of the section, to this group. Asymmetries in parameters b_x

and b_Y also generate multistability (Figure 86 A and B). Again, it is necessary to have high b_X/g_X and b_Y/g_Y ratios, with $(8/20) < b_X/b_Y < (20/8)$, approximately, to create tristability. Furthermore, the diagram g_X versus g_Y (not shown) also confirmed the necessity to have high b_i/g_i to induce multistability. Yet, if we want to understand which regimes the external signals are forcing the system to go to, Figure 86 B is not sufficient. Remember that if S_1 and S_2 are not constant in time they change b_X , b_Y , g_X and g_Y (see Equation 62 and Equation 63) concurrently. Let us then think of a scenario where we have a simpler combination of signals. Let us study the state space generated by varying b_X and g_Y . This would entail having one of the signals varying (S_1) and the other constant in time. This situation is represented in Figure 86 D. In this case, it is necessary that b_X/g_Y be large to achieve multistability.

8.6.2 Genetic decision switch stimulated by a combination of signalling input amplitudes

The decision genetic switch with external stimulation studied before in section 2.3 could be extended to generate a higher number of attractors and inducing other types of bifurcation. We chose, nevertheless, to study the simplest bifurcations possible to clarify speed-dependent cell fate decision. A transition from monostability to bistability served this purpose (Figure 87).

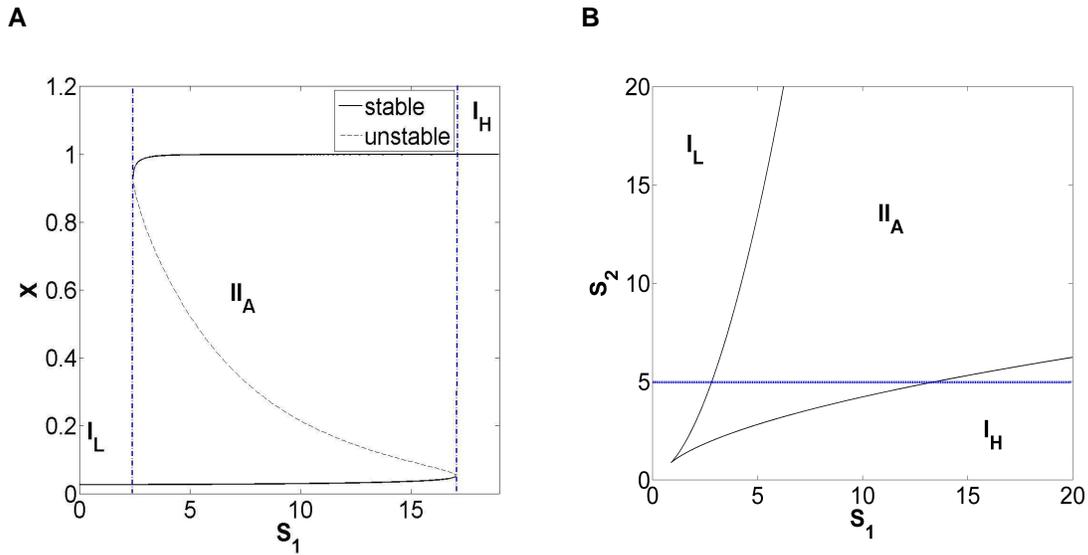


Figure 87 Bistability for the genetic decision switch with external stimulation and no cross-talk. A) For S_1 with $S_2=5$. B) 2D extension of A. For both figures $k_{Y1}=k_{X2}=0$, $k_{X1}=k_{Y2}=1$, $sa=sd=1$, $\eta_X=\eta_Y=1$, $d_X=d_Y=1$, $b_X=b_Y=1$, $g_X=g_Y=1$. II_A - bistability region with anti-symmetric states. $I_{L,H}$ - monostability region with low/high X values. $[TF_X]=X$.

Because S_1 and S_2 act concurrently, all the fact that we need high b_i/g_i ratios to induce multistability, the set of parameters chosen for Figure 87 can only generate bistability (see also Figure 86 B and D). By choosing a set of parameters d_i , b_i and g_i , that generate multistability (see Figure 83 B) we can also induce in the system a transition from monostability, to bistability, to tristability, and back to bistability (see Figure 88).

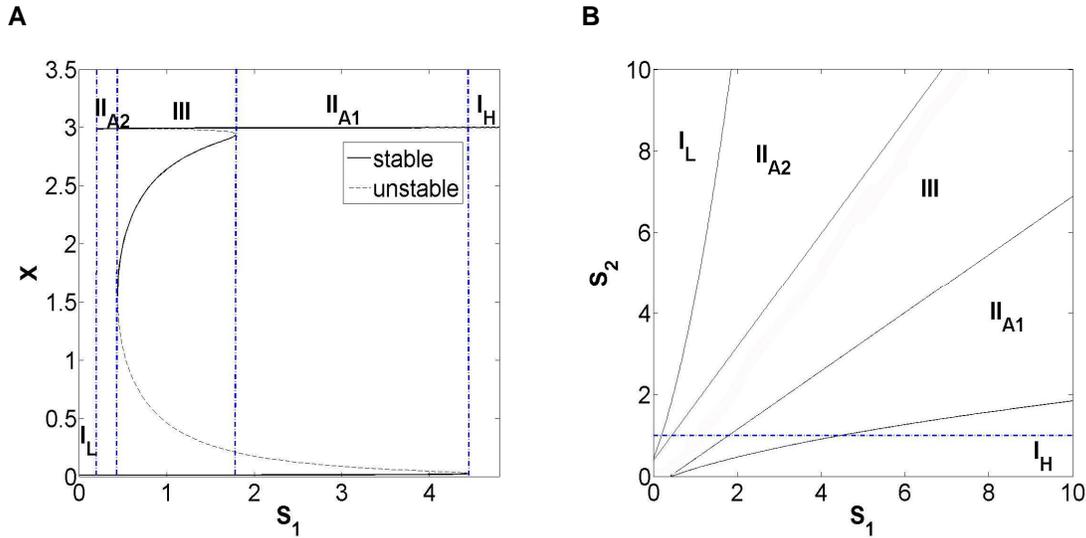


Figure 88 Multistability for the genetic decision switch with external stimulation and no cross-talk. A) Bifurcation diagram for S_1 . B) 2D diagram by extending A. Parameters: $\eta_X=\eta_Y=1$; $d_X=d_Y=3$; $b_X=b_Y=15$; $g_X=g_Y=2$; $k_{X2}=k_{Y1}=1$; $k_{X2}=k_{Y1}=0$; $sa=sd=1$ (see Equation 60 to Equation 63). $I_{L,H}$ -monostability region with low/high X concentration values. II_{A1} -bistability region with anti-symmetric states. III-tristability region. $[TF_X]=X$.

8.6.2.1 Effects of cross-talk: bistability region contracts

Although in the course of the simulations on the decision genetic switch (section 2.3) we assumed that cross-talk was inexistent, its effects can be important in attractor selection processes. Let us consider for the sake of simplicity that only cross-talk between S_1 and TF_Y exists, e.g. $k_{Y1}=0.05$ and $k_{X2}=0$ (see Equation 60-Equation 63). We observe that the bistability region is shrunk and tilted towards the S_1 axis (Figure 89 A). As the cross-talk parameter k_{Y1} strength (see Equation 60-Equation 63) is increased, for a fixed S_2 , the region where bistability exists is also reduced (see Figure 89). The effect of S_1 through k_{Y1} counterbalances the inhibitory action of TF_X over TF_Y (see also Figure 39).

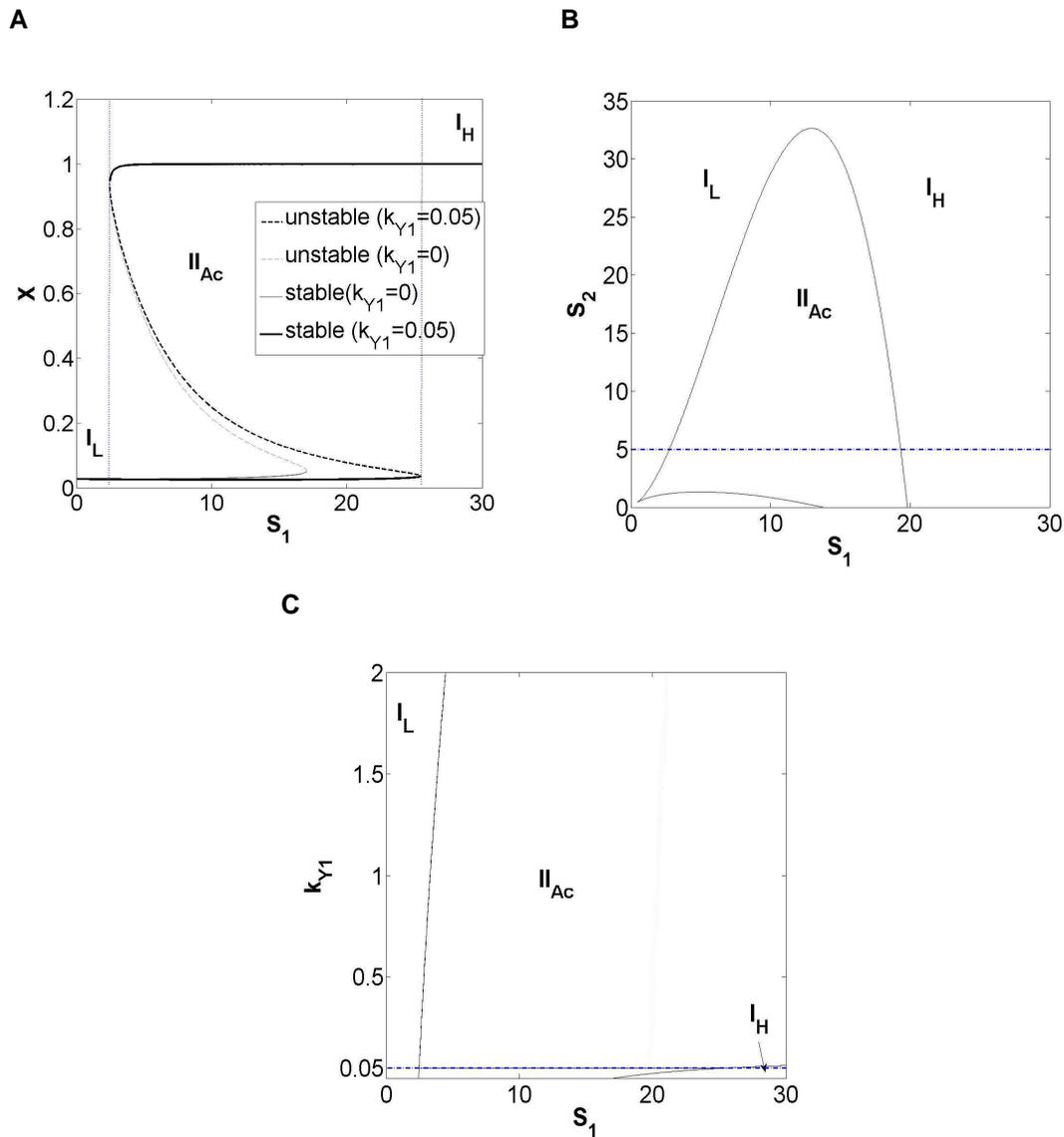


Figure 89 Effects of cross-talk on bistability. A) Bifurcation diagram with $k_{x2}=0$; $k_{y1}=0.05$ and $S_2=5$. Also shown is the diagram of Equation 83 . B) Codimension-2 bifurcation diagram with $k_{x2}=0$; $k_{y1}=0.05$. B) Codimension-1 bifurcation diagram with $S_2=4$. All other parameters for all figures are the following: $\eta_x=\eta_y=1$; $d_x=d_y=1$; $b_x=b_y=1$; $g_x=g_y=1$; $k_{x2}=0$; $k_{x1}=k_{y2}=1$; $sa=sd=1$ (see Equation 60 to Equation 63). $I_{L,H}$ - monostability with low/high X values. II_{Ac} - bistability. $[\text{TF}_x]=X$

Further combinations of cross-talk and internal asymmetries between links characterizing self-activation and cross-repression would create bifurcation diagrams of increased complexity. We observed, nonetheless, that the combinations of parameters explored in this section at most are able to create

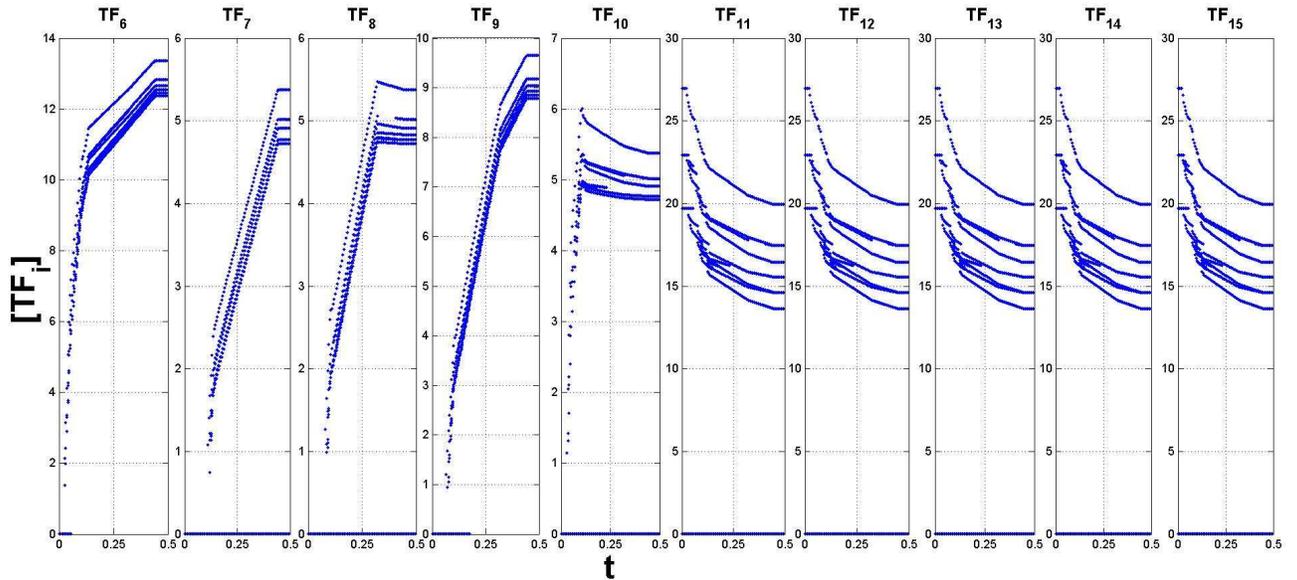
situations where 3 stable states arise. For the conditions necessary for the appearance of higher number of stable states we recommend the work of Macia and colleagues (180).

8.7 Bifurcation diagrams for specific input combinations to the highdimensional decision switch

The set of figures presented here is referred to in section 3.1.1, page 145. They are obtained by putting each of the inputs S_i in Figure 45 according to the characteristic combination of amplitudes of each of the selected input combination I_{15} , I_{75} , I_{94} over time (Figure 50). Figure 90, Figure 91 and Figure 92 show the available attractors for each input combinations as the system is swept through the critical region.

8.7.1 Input combination I_{15}

A



B

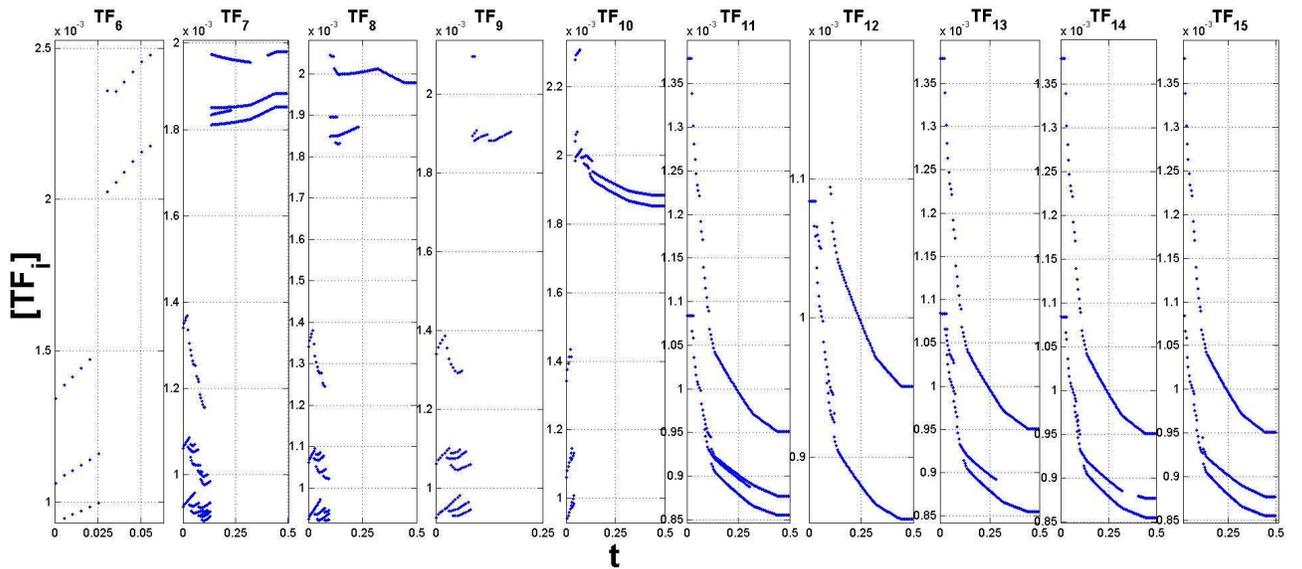
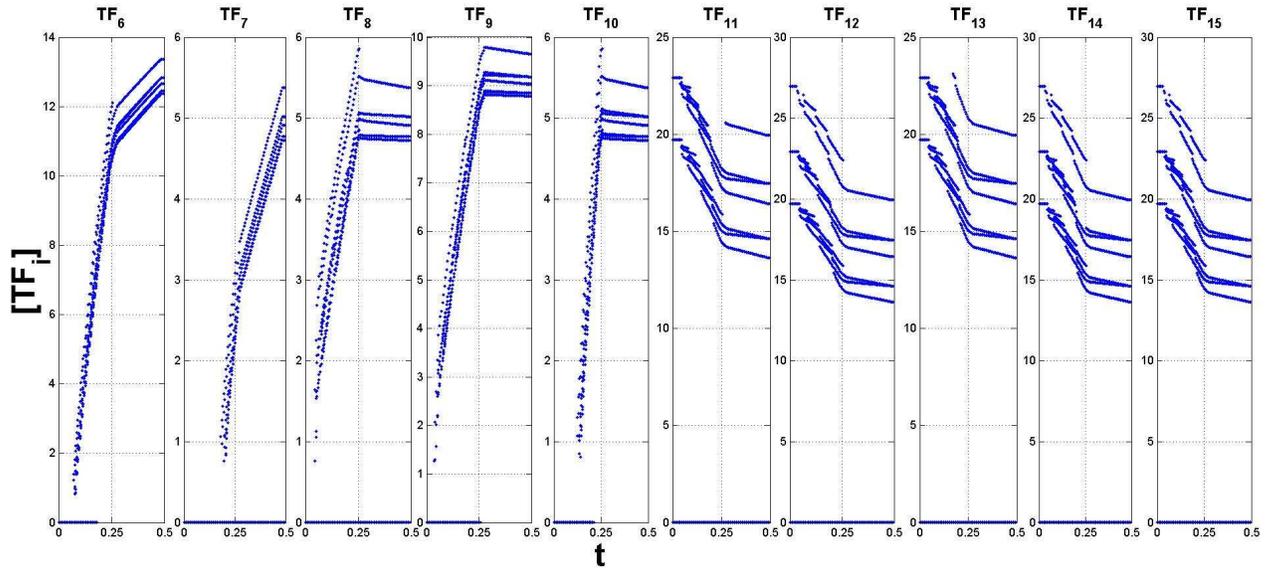


Figure 90 Bifurcation diagram obtained by setting the parameters S_i following the combination of amplitudes inherent to $I_{15}(t)$. A) Complete bifurcation diagram. B) Amplification of lower part of the bifurcation diagram represented in A). Parameters: $M=2$, $\eta_i=0.1$, $c_i=20$, $a_i=1$ (self-activation) and $a_j=0.1$ (cross-repression), $sa=0$, $sd=0.3$, $\tau_i^S=\tau_i^T=0.001$ (see Equation 68 and Equation 69) for $i, j=6, \dots, 15$. The available attractors at specific times can be visualized. The input combination changes the attractor landscape with respect to the original bifurcation diagram with $S=S_1=S_2=S_3=S_4=S_5$ (see Figure 47) and the other input sequences I_{75} and I_{94} . t is the horizontal axis variable for all the figures, from TF_6 to TF_{15} .

8.7.2 Input combination I_{75}

A



B

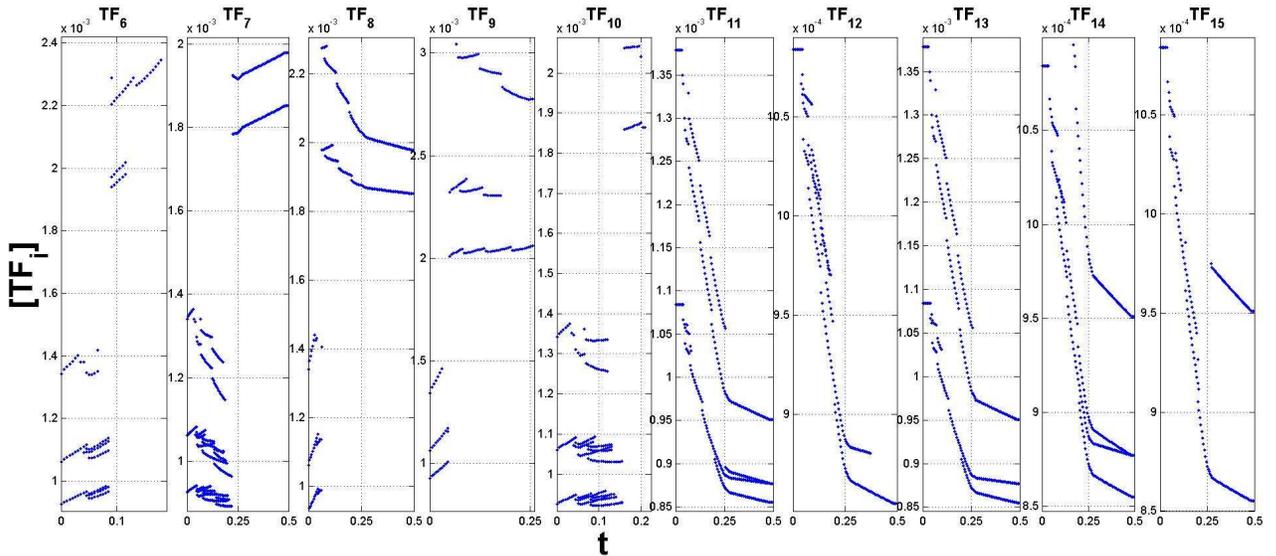
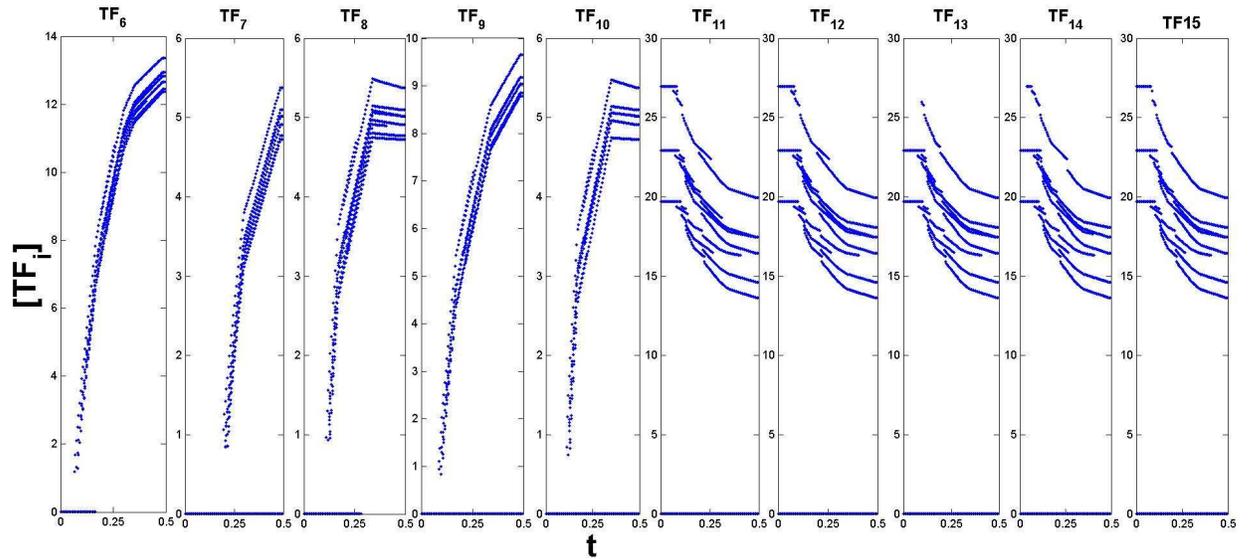


Figure 91 Bifurcation diagram obtained by setting the parameters S_i following the combination of amplitudes inherent to $I_{75}(t)$. A) Complete bifurcation diagram. B) Amplification of lower part of the bifurcation diagram represented in A). Parameters: $M=2$, $\eta_i=0.1$, $c_i^i=20$, $a_i^i=1$ (self-activation) and $a_j^i=0.1$ (cross-repression), $sa=0$, $sd=0.3$, $\tau_i^S=\tau_i^T=0.001$ (see Equation 68 and Equation 69) for $i,j=6,\dots,15$. The available attractors at specific times can be visualized. The input combination changes the attractor landscape with respect to the original bifurcation diagram with $S=S_1=S_2=S_3=S_4=S_5$ (see Figure 47) and the other input sequences I_{15} and I_{94} . t is the horizontal axis variable for all the figures, from TF_6 to TF_{15} .

8.7.3 Input combination I_{94}

A



B

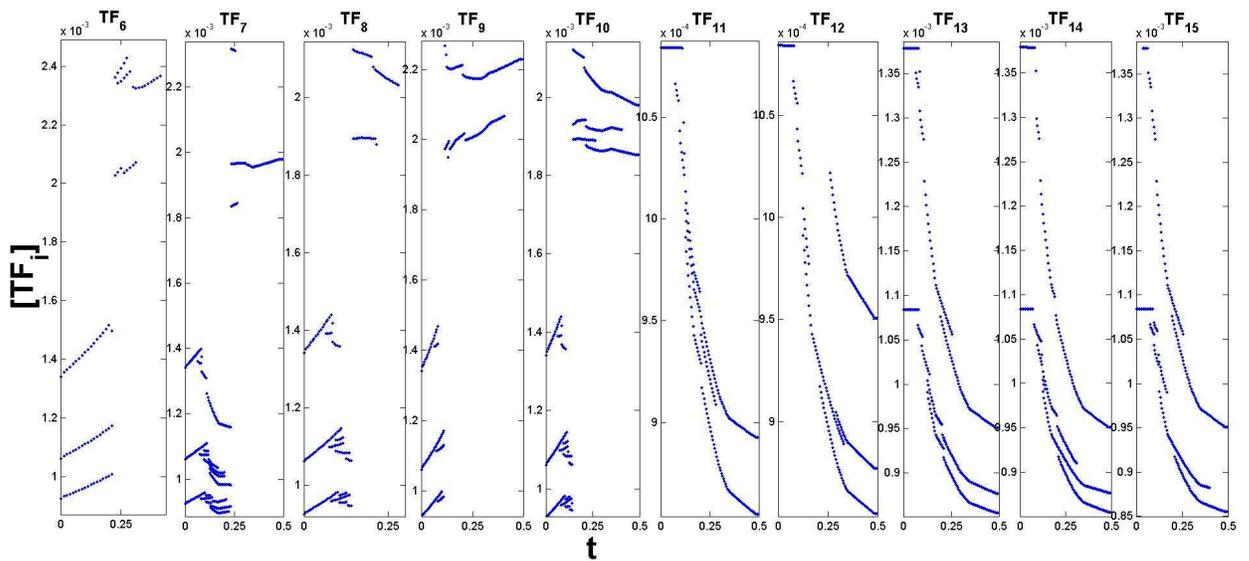


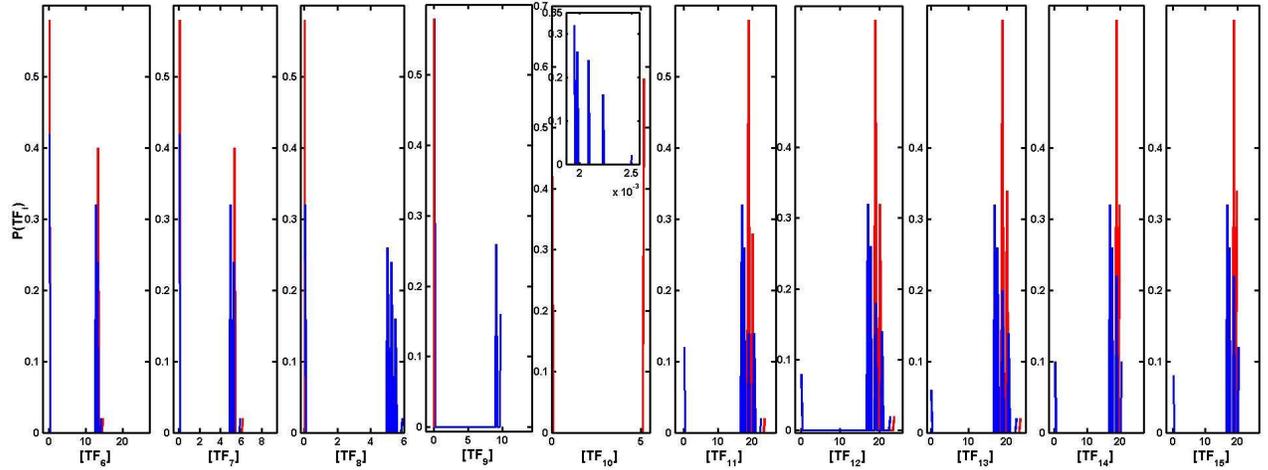
Figure 92 Bifurcation diagram obtained by setting the parameters S_i following the combination of amplitudes inherent to $I_{94}(t)$. A) Complete bifurcation diagram. B) Amplification of lower part of the bifurcation diagram represented in A). Parameters: $M=2$, $\eta_i=0.1$, $c_i=20$, $a_i=1$ (self-activation) and $a_j=0.1$ (cross-repression), $sa=0$, $sd=0.3$, $\tau_i^S=\tau_i^T=0.001$ (see Equation 68 and Equation 69) for $i,j=6,\dots,15$. The available attractors at specific times can be visualized. The input combination changes the attractor landscape with respect to the original bifurcation diagram with $S=S_1=S_2=S_3=S_4=S_5$ (see Figure 47) and the other input sequences I_{15} and I_{75} . t is the horizontal axis variable for all the figures, from TF_6 to TF_{15} .

8.8 Final distributions for the highdimensional switch for input combination I_{15} , I_{75} , I_{94} in the presence of fluctuations

The set of figures presented here corresponds to the analysis of the effects of noise on final distributions of trajectories over attractors, when the highdimensional genetic decision switch (see section 3) is stimulated with input combinations I_{15} , I_{75} , I_{94} . As was seen in section 3.3 (Figure 49) these combinations induce the same attractors when starting at the same initial condition when noise is not considered. These figures also help understanding the differences between final distributions of trajectories summarized in Figure 54. Comments to the figures presented below are made in section 3.3.1.

8.8.1 Distribution across attractors for I_{15} , I_{75} , I_{94} with noise intensity $\sigma = 0.01$

A



B

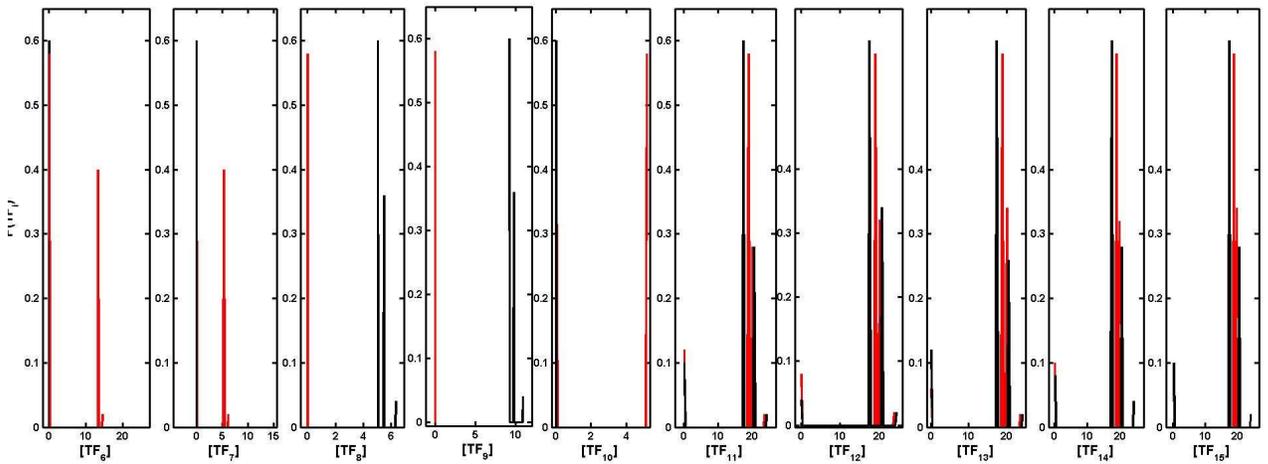
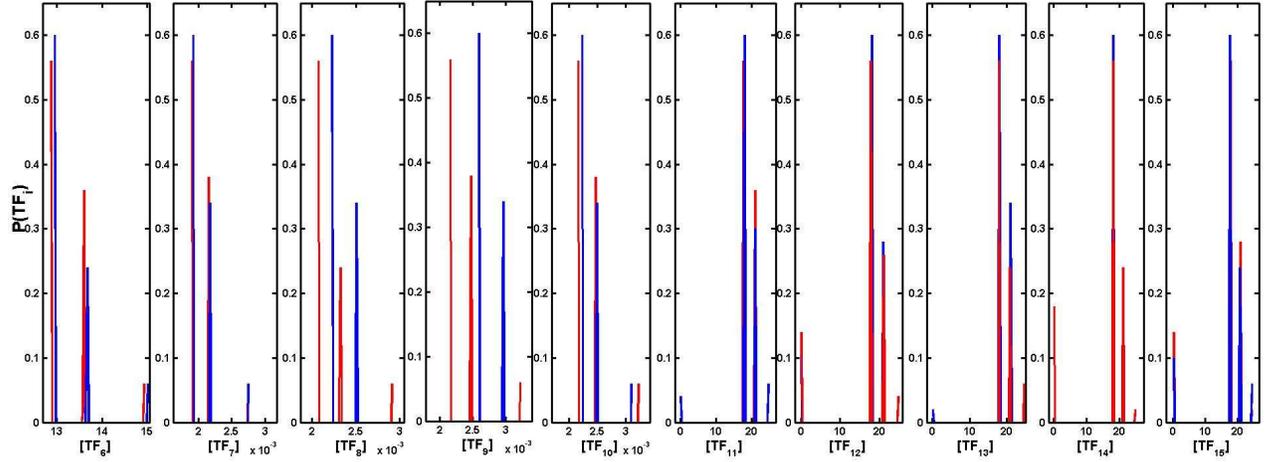


Figure 93 Distributions across attractors in the presence of noise intensity $\sigma=0.01$ (see Equation 68). A) I_{15} (blue) and I_{75} (red). B) I_{75} (red) and I_{94} (black). Parameters: $M=2$, $\eta_i=0.1$, $c_i=20$, $a_i^i=1$ (self-activation) and $a_{ij}^i=0.1$ (cross-repression), $sa=0$, $sd=0.3$, $\tau_i^S=\tau_i^T=0.001$ (see Equation 68 and Equation 69) for $i, j=6, \dots, 15$. Histograms for each input combination generated from 5000 trajectories.

8.8.2 Distribution across attractors for I_{15} , I_{75} , I_{94} with noise intensity $\sigma=0.5$

A



B

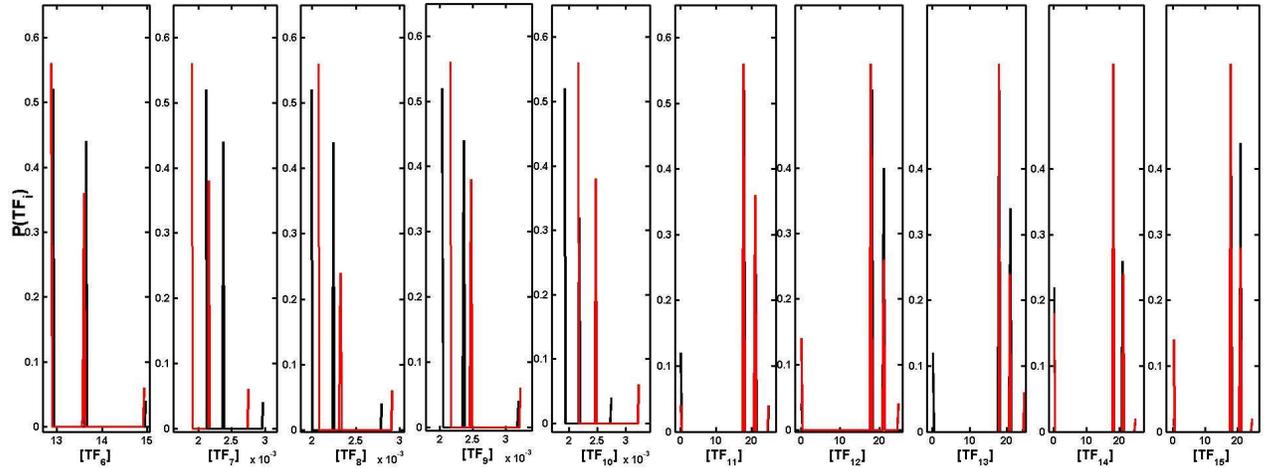


Figure 94 Distributions across attractors in the presence of noise intensity $\sigma=0.5$ (see Equation 68). A) I_{15} (blue) and I_{75} (red). B) I_{75} (red) and I_{94} (black). Parameters: $M=2$, $\eta_i=0.1$, $c_i^{-1}=20$, $a_i^i=1$ (self-activation) and $a_j^i=0.1$ (cross-repression), $sa=0$, $sd=0.3$, $\tau^S_i=\tau^T_i=0.001$ (see Equation 68 and Equation 69) for $i, j=6, \dots, 15$. Histograms for each input combination generated from 5000 trajectories.

9 Reference List

1. Pharmaprojects database. 2006. **PJB Publications**.
Ref Type: Report
2. Abbott, L. H. & Michor, F. (2006) *British Journal of Cancer* **95**, 1136-1141.
3. Agoston, V., Csermely, P. & Pongor, S. (2005) *Phys. Rev E Stat. Nonlin. Soft Matter Phys.* **71**, 051909.
4. Albert, R. & Barabasi, A. L. (2002) *Reviews of Modern Physics* **74**, 47-97.
5. Albert, R., Jeong, H. & Barabasi, A. L. (2000) *Nature* **406**, 378-382.
6. Alberts, B., Johnson, A., Lewis, J., Raff, M., Keith, R. & Walter, P. (2002) *Molecular biology of the cell* (Garland Pub, New York).
7. Alon, U. (2007) *An Introduction to Systems Biology-design principles of biological circuits* (Taylor & Francis Group,LLC).
8. Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D. & Levine, A. J. (1999) *Proc. Natl. Acad. Sci. U. S A* **96**, 6745-6750.
9. Amit, I., Citri, A., Shay, T., Lu, Y., Katz, M., Zhang, F., Tarcic, G., Siwak, D., Lahad, J., Jacob-Hirsch, J. *et al.* (2007) *Nat Genet.* **39**, 503-512.
10. Amit, I., Wides, R. & Yarden, Y. (2007) *Mol. Syst. Biol.* **3**, 151.
11. Andrecut, M. & Kauffman, S. A. (2006) *New J. Phys.* **8**.
12. Andrecut, M. & Kauffman, S. A. (2007) *J. Comput. Biol.* **14**, 973-983.
13. Anifrani, J.-C., Le Floc'h, C., Sornette, D. & and Souillard, B. (1995) *J. Phys. I France* **5**, 631-638.
14. Araujo, R. P., Petricoin, E. F. & Liotta, L. A. (2005) *Biosystems* **80**, 57-69.
15. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. (2003) *Cell* **113**, 597-607.
16. Audibert, A., Weil, D. & Dautry, F. (2002) *Mol. Cell Biol.* **22**, 6706-6718.
17. Awan, A., Bari, H., Yan, F., Moksong, S., Yang, S., Chowdhury, S., Cui, Q., Yu, Z., Purisima, E. O. & Wang, E. (2007) *IET Syst. Biol.* **1**, 292-297.
18. B.Ermentrout (2002) *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to Xppaut for Researchers and Students (Software, Environments, Tools)* (SIAM, North-Holand,Amsterdam).
19. B.S.Jhaveri & and G.M.Homsy (1982) *J. Fluid Mech.* **114**, 251-260.

20. Bardwell, L., Zou, X., Nie, Q. & Komarova, N. L. (2007) *Biophys. J.* **92**, 3425-3441.
21. Behar, M., Dohlman, H. G. & Elston, T. C. (2007) *Proc. Natl. Acad. Sci. U. S A* **104**, 16146-16151.
22. Bennett, M. R., Volfson, D., Tsimring, L. & Hasty, J. (2007) *Biophys. J.* **92**, 3501-3512.
23. Bentley, D. (2002) *Curr. Opin. Cell Biol.* **14**, 336-342.
24. Berlung, N. & Gentz, B. (2006) *Noise-Induced Phenomena Phenomena in Slow-Fast Systems: A Sample-Paths Approach* (Springer).
25. Bhalla, U. S. (2003) *Prog. Biophys. Mol. Biol.* **81**, 45-65.
26. Bhalla, U. S. & Iyengar, R. (1999) *Science* **283**, 381-387.
27. Bhalla, U. S., Ram, P. T. & Iyengar, R. (2002) *Science* **297**, 1018-1023.
28. Bianco, R., Melisi, D., Ciardiello, F. & Tortora, G. (2006) *Eur. J. Cancer* **42**, 290-294.
29. Bilanges, B., Torbett, N. & Vanhaesebroeck, B. (2008) *Nat Chem. Biol.* **4**, 648-649.
30. Bild, A. H., Yao, G., Chang, J. T., Wang, Q., Potti, A., Chasse, D., Joshi, M. B., Harpole, D., Lancaster, J. M., Berchuck, A. *et al.* (2006) *Nature* **439**, 353-357.
31. Binder, B. & Heinrich, R. (2004) *Genome Inform.* **15**, 13-23.
32. Binder, B. & Heinrich, R. (2005) *Genome Inform.* **16**, 164-173.
33. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T. & Phillips, R. (2005) *Curr. Opin. Genet. Dev.* **15**, 125-135.
34. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J. & Phillips, R. (2005) *Curr. Opin. Genet. Dev.* **15**, 116-124.
35. Bishop, C. M. (1994) *Review of Scientific Instruments* **65**, 1803-1832.
36. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. U. (2006) *Physics Reports-Review Section of Physics Letters* **424**, 175-308.
37. Bowman, D. D., Ouillon, G., Sammis, C. G., Sornette, A. & Sornette, D. (1998) *J. Geophys. Res.* **103**, 24359-24372.
38. Brightman, F. A. & Fell, D. A. (2000) *FEBS Lett.* **482**, 169-174.
39. Brivanlou, A. H. & Darnell Jr., J. E. (2000) *Science* **295**, 813-818.
40. Buchler, N. E., Gerland, U. & Hwa, T. (2003) *Proc. Natl. Acad. Sci. U. S A* **100**, 5136-5141.
41. Burack, W. R. & Shaw, A. S. (2000) *Curr. Opin. Cell Biol.* **12**, 211-216.
42. Burke, P., Schooler, K., Wiley, H. S. & Lauffenburger, D. A. (2001) *Mol. Biol. Cell* **12**, 1897-1910.

43. Butcher, E. C. (2005) *Nat Rev Drug Discov.* **4**, 461-467.
44. Butcher, E. C., Berg, E. L. & Kunkel, E. J. (2004) *Nat Biotechnol.* **22**, 1253-1259.
45. Calder, M., Gilmore, S. & Hillston, J. (2006) *Transactions on Computational Systems Biology VII* **4230**, 1-23.
46. Calzolari, D., Bruschi, S., Coquin, L., Schofield, J., Feala, J. D., Reed, J. C., McCulloch, A. D. & Paternostro, G. (2008) *Plos Computational Biology* **4**.
47. Carlson, J. M. & Doyle, J. (2000) *Phys. Rev Lett.* **84**, 2529-2532.
48. Carlson, J. M. & Doyle, J. (2002) *Proc. Natl. Acad. Sci. U. S. A* **99 Suppl 1**, 2538-2545.
49. Carlson, J. M. & Doyle, J. (1999) *Phys. Rev E Stat. Phys. Plasmas. Fluids Relat Interdiscip. Topics.* **60**, 1412-1427.
50. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. (2008) *Nature* **453**, 544-547.
51. Chen, W. W., Schoeberl, B., Jasper, P. J., Niepel, M., Nielsen, U. B., Lauffenburger, D. A. & Sorger, P. K. (2009) *Molecular Systems Biology* **5**.
52. Cherry, J. L. & Adler, F. R. (2000) *J. Theor. Biol.* **203**, 117-133.
53. Cinquin, O. & Demongeot, J. (2002) *C R. Biol.* **325**, 1085-1095.
54. Cinquin, O. & Demongeot, J. (2005) *J. Theor. Biol.* **233**, 391-411.
55. Cinquin, O. & Page, K. M. (2007) *Bull. Math. Biol.* **69**, 483-494.
56. Citri, A. & Yarden, Y. (2006) *Nat Rev Mol. Cell Biol.* **7**, 505-516.
57. Cobb, M. H., Xu, S. C., Hepler, J. E., Hutchison, M., Frost, J. & Robbins, D. J. (1994) *Cellular & Molecular Biology Research* **40**, 253-256.
58. Cook, S. J., Aziz, N. & McMahon, M. (1999) *Mol. Cell Biol.* **19**, 330-341.
59. Cowan, K. J. & Storey, K. B. (2003) *J. Exp. Biol.* **206**, 1107-1115.
60. Crampin, E. J., Schnell, S. & McSharry, P. E. (2004) *Progress in Biophysics & Molecular Biology* **86**, 77-112.
61. Cross, M. C. & Hohenberg, P. C. (1993) *Reviews of Modern Physics* **65**.
62. Csermely, P., Agoston, V. & Pongor, S. (2005) *Trends Pharmacol. Sci.* **26**, 178-182.
63. Csete, M. & Doyle, J. (2004) *Trends Biotechnol.* **22**, 446-450.
64. Cui, Q., Ma, Y., Jaramillo, M., Bari, H., Awan, A., Yang, S., Zhang, S., Liu, L., Lu, M., O'Connor-McCourt, M. *et al.* (2007) *Mol. Syst. Biol.* **3**, 152.
65. Cui, Q., Purisima, E. O. & Wang, E. (2009) *BMC Syst. Biol.* **3**, 21.
66. Dancey, J. E. & Chen, H. X. (2006) *Nat Rev Drug Discov.* **5**, 649-659.

67. de, J. H. (2002) *J. Comput. Biol.* **9**, 67-103.
68. de, J. H., Gouze, J. L., Hernandez, C., Page, M., Sari, T. & Geiselmann, J. (2004) *Bull. Math. Biol.* **66**, 301-340.
69. Dee, D. & Ghil, M. (1984) *Siam Journal on Applied Mathematics* **44**, 111-126.
70. Del, V. D., Ninfa, A. J. & Sontag, E. D. (2008) *Mol. Syst. Biol.* **4**, 161.
71. Deng, X., Geng, H. & Matache, M. T. (2007) *Biosystems* **88**, 16-34.
72. Dominguez, D., Koroutchev, K., Serrano, E. & Rodriguez, F. B. (2007) *Neural Comput.* **19**, 956-973.
73. Donovan, J. C. H., Milic, A. & Slingerland, J. M. (2001) *The Journal of Biological Chemistry* **276**, 40888-40895.
74. Doyle, J. & Carlson, J. M. (2000) *Phys. Rev Lett.* **84**, 5656-5659.
75. Eichler, G. S., Huang, S. & Ingber, D. E. (2003) *Bioinformatics* **19**, 2321-2322.
76. Eldar, A. & Elowitz, M. B. (2010) *Nature* **467**, 167-173.
77. Elowitz, M. B. & Leibler, S. (2000) *Nature* **403**, 335-338.
78. Enver, T., Pera, M., Peterson, C. & Andrews, P. W. (2009) *Cell Stem Cell* **4**, 387-397.
79. Fabian, M. A., Biggs, W. H., III, Treiber, D. K., Atteridge, C. E., Azimioara, M. D., Benedetti, M. G., Carter, T. A., Ciceri, P., Edeen, P. T., Floyd, M. *et al.* (2005) *Nat Biotechnol.* **23**, 329-336.
80. Fernandez, P. & Sole, R. V. (2007) *Journal of the Royal Society Interface* **4**, 41-47.
81. Fischer, O. M., Hart, S., Gschwind, A. & Ullrich, A. (2003) *Biochem. Soc. Trans.* **31**, 1203-1208.
82. Fitzgerald, J. B., Schoeberl, B., Nielsen, U. B. & Sorger, P. K. (2006) *Nat Chem. Biol.* **2**, 458-466.
83. Foo, J. & Michor, F. (2009) *PLoS Comput. Biol.* **5**, e1000557.
84. Fritz, G., Buchler, N. E., Hwa, T. & Gerland, U. (2007) *Syst. Synth. Biol.* **1**, 89-98.
85. Fujimoto, K. & Kaneko, K. (2003) *Chaos* **13**.
86. Fujimoto, K. & Kaneko, K. (2003) *Physica D* **180**, 1-16.
87. Fujimoto, K. & Kaneko, K. (2004) *Physica D* **196**, 67-89.
88. Furusawa, C. & Kaneko, K. (2008) *Plos Computational Biology* **4**.
89. G.Ahlers, M.C.Cross, P.C.Hohenberg & S.Safran (1981) *J. Fluid Mech* **110**, 297-334.
90. Gagneur, J. & Casari, G. (2005) *FEBS Lett.* **579**, 1867-1871.

91. Galanis, A., Yang, S. H. & Sharrocks, A. D. (2001) *J. Biol. Chem.* **276**, 965-973.
92. Galcheva-Gargova, Z., Derijard, B., Wu, I. H. & Davis, R. J. (1994) *Science* **265**, 806-808.
93. Garcimartin, A., Guarino, A., Bellon, L. & Ciliberto, S. (1997) *Phys. Rev. Lett.* **79**, 3202-3205.
94. Gardiner, C. W. (2004) *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*.
95. Gardner, T. S., Cantor, C. R. & Collins, J. J. (2000) *Nature* **403**, 339-342.
96. Gaudet, S., Janes, K. A., Albeck, J. G., Pace, E. A., Lauffenburger, D. A. & Sorger, P. K. (2005) *Mol Cell Proteomics* **4**, 1569-1590.
97. Ghil, M., Zaliapin, I. & Coluzzi, B. (2008) *Physica D-Nonlinear Phenomena* **237**, 2967-2986.
98. Gillespie, D. T. (2000) *Journal of Chemical Physics* **113**, 297-306.
99. Gillespie, D. T. (2001) *Journal of Chemical Physics* **115**, 1716-1733.
100. Gillespie, D. T. (2002) *Journal of Physical Chemistry A* **106**, 5063-5071.
101. Gillespie, D. T. (2007) *Annual Review of Physical Chemistry* **58**, 35-55.
102. Gillespie, D. T. (1977) *Journal of Physical Chemistry* **81**, 2340-2361.
103. Gillespie, D. T. & Petzold, L. R. (2003) *Journal of Chemical Physics* **119**, 8229-8234.
104. Gjuvslund, A. B., Plahte, E. & Omholt, S. W. (2007) *BMC Syst. Biol.* **1**, 57.
105. Gong, Y. & Zhao, X. (2003) *FEBS Lett* **554**, 467-472.
106. Gotzmann, J., Fischer, A. N. M., Zojer, M., Mikula, M., Proell, V., Huber, H., Jechlinger, M., Waerner, T., Weith, A., Beug, H. *et al.* (2006) *Oncogene* **25**, 3170-3185.
107. Grossmann, S. & Mikhailov, A. S. (1990) *Z. Phys. B-Condensed Matter* **78**, 1-10.
108. Guantes, R. & Poyatos, J. F. (2006) *PLoS Comput. Biol.* **2**, e30.
109. Guantes, R. & Poyatos, J. F. (2008) *PLoS Comput. Biol.* **4**, e1000235.
110. Gundelfinger, E. D., Kessels, M. M. & Qualmann, B. (2003) *Nature Reviews Molecular Cell Biology* **4**, 127-139.
111. Gupta, S., Barrett, T., Whitmarsh, A. J., Cavanagh, J., Sluss, H. K., Derijard, B. & Davis, R. J. (1996) *Embo Journal* **15**, 2760-2770.
112. Haken, H. (1977) *Synergetics* (Springer, Berlin).
113. Han, S. W., Kim, T. Y., Jeon, Y. K., Hwang, P. G., Im, S. A., Lee, K. H., Kim, J. H., Kim, D. W., Heo, D. S., Kim, N. K. *et al.* (2006) *Clin. Cancer Res.* **12**, 2538-2544.

114. Han, S. W., Kim, T. Y., Jeon, Y. K., Hwang, P. G., Im, S. A., Lee, K. H., Kim, J. H., Kim, D. W., Heo, D. S., Kim, N. K. *et al.* (2006) *Clinical Cancer Research* **12**, 2538-2544.
115. Hanahan, D. & Weinberg, R. A. (2000) *Cell* **100**, 57-70.
116. Harding, A., Tian, T., Westbury, E., Frische, E. & Hancock, J. F. (2005) *Curr. Biol.* **15**, 869-873.
117. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. (1999) *Nature* **402**, C47-C52.
118. Hasty, J., Isaacs, F., Dolnik, M., McMillen, D. & Collins, J. J. (2001) *Chaos* **11**, 207-220.
119. Hatakeyama, M., Kimura, S., Naka, T., Kawasaki, T., Yumoto, N., Ichikawa, M., Kim, J. H., Saito, K., Saeki, M., Shirouzu, M. *et al.* (2003) *Biochemical Journal* **373**, 451-463.
120. Heijnen, J. J. (2005) *Biotechnol. Bioeng.* **91**, 534-545.
121. Heinrich, R., Neel, B. G. & Rapoport, T. A. (2002) *Mol. Cell* **9**, 957-970.
122. Helikar, T., Konvalina, J., Heidel, J. & Rogers, J. A. (2008) *Proc. Natl. Acad. Sci. U. S. A* **105**, 1913-1918.
123. Holter, N. S., Mitra, M., Maritan, A., Cieplak, M., Banavar, J. R. & Fedoroff, N. V. (2000) *Proc. Natl. Acad. Sci. U. S. A* **97**, 8409-8414.
124. Hood, L. & Perlmutter, R. M. (2004) *Nat Biotechnol.* **22**, 1215-1217.
125. Hornberg, J. J., Binder, B., Bruggeman, F. J., Schoeberl, B., Heinrich, R. & Westerhoff, H. V. (2005) *Oncogene* **24**, 5533-5542.
126. Hoshino, R., Chatani, Y., Yamori, T., Tsuruo, T., Oka, H., Yoshida, O., Shimada, Y., Ari-i, S., Wada, H., Fujimoto, J. *et al.* (1999) *Oncogene* **18**, 813-822.
127. Huang, S. (2001) *Pharmacogenomics* 203-222.
128. Huang, S. (1999) *J. Mol. Med* **77**, 77,469-77,480.
129. Huang, S. (2002) in *Gene Regulation and Metabolism: Post-Genomic Computational Approach*, ed. Collado-Vides, J. a. H. R. (MIT Press).
130. Huang, S. (2005) in *Computational Systems Biology*, eds. Kriete, A. & Elis, R. (Elsevier Academic Press).
131. Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. E. (2005) *Phys. Rev Lett.* **94**, 128701.
132. Huang, S., Ernberg, I. & Kauffman, S. (2009) *Semin. Cell Dev. Biol.* **20**, 869-876.
133. Huang, S., Guo, Y., May, G. & Enver, T. (2007) *Developmental Biology* **305**, 695-713.
134. Itadani, H., Mizuarai, S. & Kotani, H. (2008) *Curr. Genomics* **9**, 349-360.
135. J.Garcia-Ojalvo and J.M.Sancho (1999) *Noise in Spatially Extended Systems* (Springer, New York).
136. Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. (2001) *Nature* **411**, 41-42.

137. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. (2000) *Nature* **407**, 651-654.
138. Johansen, A., Ledoit, O. & Sornette, D. (2000) *Int. J. Theor. Applied Finance* **3**, 219-255.
139. Johansen, A. & Sornette, D. (1999) *Risk* **12**, 91-94.
140. John Ross, Igor Schreiber & Marcel O Vlad (2006) *Determination of Complex Reaction Mechanisms: Analysis of Chemical, Biological, and Genetic Networks* (Oxford University Press).
141. Kaluza, P., Ipsen, M., Vingron, M. & Mikhailov, A. S. (2007) *Phys. Rev E Stat. Nonlin. Soft Matter Phys.* **75**, 015101.
142. Kaluza, P., Vingron, M. & Mikhailov, A. S. (2008) *Chaos* **18**, 026113.
143. Kaneko, K. (1998) *Physica D-Nonlinear Phenomena* **124**, 322-344.
144. Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., Chan, K. W., Ciceri, P., Davis, M. I., Edeen, P. T. *et al.* (2008) *Nat Biotechnol.* **26**, 127-132.
145. Kashiwagi, A., Urabe, I., Kaneko, K. & Yomo, T. (2006) *Plos One* **1**.
146. Kauffman, S. A. (1993) *The Origins of Order: Sel-Organization and Selection in Evolution* (Oxford University Press).
147. Keith, C. T., Borisy, A. A. & Stockwell, B. R. (2005) *Nat Rev Drug Discov.* **4**, 71-78.
148. Kerszberg, M. (2004) *Curr. Opin. Genet. Dev.* **14**, 440-445.
149. Kholodenko, B. N. (2000) *European Journal of Biochemistry* **267**, 1583-1588.
150. Kholodenko, B. N., Demin, O. V., Moehren, G. & Hoek, J. B. (1999) *J. Biol. Chem.* **274**, 30169-30181.
151. Kholodenko, B. N., Hoek, J. B., Westerhoff, H. V. & Brown, G. C. (1997) *FEBS Lett.* **414**, 430-434.
152. Kitano, H. (2001) *Foundations of Systems Biology* (MIT Press).
153. Kitano, H. (2004) *Nat Rev Genet.* **5**, 826-837.
154. Kitano, H. (2007) *Mol. Syst. Biol.* **3**, 137.
155. Kitano, H. (2007) *Nat Rev Drug Discov.* **6**, 202-210.
156. Klemm, K. & Bornholdt, S. (2005) *Phys. Rev E Stat. Nonlin. Soft Matter Phys.* **72**, 055101.
157. Kohn, K. W. & Aladjem, M. I. (2006) *Mol. Syst. Biol.* **2**, 2006.
158. Kohn, K. W., Aladjem, M. I., Weinstein, J. N. & Pommier, Y. (2006) *Mol. Biol. Cell* **17**, 1-13.

159. Komarova, N. L., Zou, X., Nie, Q. & Bardwell, L. (2005) *Mol. Syst. Biol.* **1**, 2005.
160. Kondepudi, D. K. & Asakura, K. (2001) *Acc. Chem. Res.* **34**, 946-954.
161. Kondepudi, D. K., Moss, F. & McClintock, P. V. E. (1986) *Physica D* **21**, 296-306.
162. Kondepudi, D. K. & Nelson, G. W. (1983) *Phys. Rev Lett* **50**, 1023-1026.
163. Koseska, A., Ullner, E., Volkov, E., Kurths, J. & Garcia-Ojalvo, J. (2010) *J. Theor. Biol.* **263**, 189-202.
164. Koseska, A., Volkov, E., Zaikin, A. & Kurths, J. (2007) *Phys. Rev E Stat. Nonlin. Soft Matter Phys.* **75**, 031916.
165. Koseska, A., Zaikin, A., Garcia-Ojalvo, J. & Kurths, J. (2007) *Phys. Rev E Stat. Nonlin. Soft Matter Phys.* **75**, 031917.
166. Laslo, P., Spooner, C. J., Warmflash, A., Lancki, D. W., Lee, H. J., Sciammas, R., Gantner, B. N., Dinner, A. R. & Singh, H. (2006) *Cell* **126**, 755-766.
167. Lauffenburger, D. A. (2000) *Proc. Natl. Acad. Sci. U. S A* **97**, 5031-5033.
168. Le, N. N. & Shimizu, T. S. (2001) *Bioinformatics* **17**, 575-576.
169. Lehar, J., Krueger, A., Zimmermann, G. & Borisy, A. (2008) *Mol. Syst. Biol.* **4**, 215.
170. Lehar, J., Krueger, A. S., Avery, W., Heilbut, A. M., Johansen, L. M., Price, E. R., Rickles, R. J., Short, G. F., III, Staunton, J. E., Jin, X. *et al.* (2009) *Nat Biotechnol.* **27**, 659-666.
171. Lehar, J., Krueger, A. S., Zimmermann, G. R. & Borisy, A. A. (2009) *Discov. Med* **8**, 185-190.
172. Lehar, J., Stockwell, B. R., Giaever, G. & Nislow, C. (2008) *Nat Chem. Biol.* **4**, 674-681.
173. Lewis, T. S., Shapiro, P. S. & Ahn, N. G. (1998) *Adv. Cancer Res.* **74**, 49-139.
174. Lindner, B., Garcia-Ojalvo, J., Neiman, A. & Schimansky-Geier, L. (2004) *Phys. Rep.* **392**, 321.
175. Liu, Y. Y., Slotine, J. J. & Barabasi, A. L. (2011) *Nature* **473**, 167-173.
176. Lopez-Maury, L., Marguerat, S. & Bahler, J. (2008) *Nat Rev Genet.* **9**, 583-593.
177. Louderbough, J. M. & Schroeder, J. A. (2009) *Clinical & Experimental Metastasis* **26**, 899.
178. Ma'ayan, A., Jenkins, S. L., Neves, S., Hasseldine, A., Grace, E., Dubin-Thaler, B., Eungdamrong, N. J., Weng, G., Ram, P. T., Rice, J. J. *et al.* (2005) *Science* **309**, 1078-1083.
179. Ma, S.-K. (1976) *Modern theory of critical phenomena* (W.A.Benjamin, Reading).
180. Macia, J., Widder, S. & Sole, R. (2009) *J. Theor. Biol.* **261**, 126-135.

181. Madhusudan Natarajan, Keng-Mean Lin, Robert C.Hsueh, Paul C.Sternweis & Rama Ranganathan (2006) *NATURE CELL BIOLOGY* **8**.
182. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. (2002) *Science* **298**, 1912-+.
183. Markevich, N. I., Moehren, G., Demin, O. V., Kiyatkin, A., Hoek, J. B. & Kholodenko, B. N. (2004) *Syst. Biol. (Stevenage.)* **1**, 104-113.
184. Marshall, C. J. (1995) *Cell* **80**, 179-185.
185. McClean, M. N., Mody, A., Broach, J. R. & Ramanathan, S. (2007) *Nat Genet.* **39**, 409-414.
186. Mendoza, L., Thieffry, D. & varez-Buylla, E. R. (1999) *Bioinformatics* **15**, 593-606.
187. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. (2002) *Science* **298**, 824-827.
188. Momiji, H. & Monk, N. A. M. (2009) *Physical Review e* **80**.
189. Morton-Firth, C. J. & Bray, D. (1998) *J. Theor. Biol.* **192**, 117-128.
190. Morton-Firth, C. J., Shimizu, T. S. & Bray, D. (1999) *J. Mol. Biol.* **286**, 1059-1074.
191. Moss, F., Kondepudi, D. K. & McClintock, P. V. E. (1985) *Physics Letters* **112A**, 293-296.
192. Motter, A. E., Gulbahce, N., Almaas, E. & Barabasi, A. L. (2008) *Mol. Syst. Biol.* **4**, 168.
193. Murphy, L. O. & Blenis, J. (2006) *Trends Biochem. Sci.* **31**, 268-275.
194. Murphy, L. O., MacKeigan, J. P. & Blenis, J. (2004) *Mol. Cell Biol.* **24**, 144-153.
195. Murphy, L. O., Smith, S., Chen, R. H., Fingar, D. C. & Blenis, J. (2002) *Nat Cell Biol.* **4**, 556-564.
196. N.R.Nene & A.Zaikin. (2010) *Proceedings of the BIOSIGNAL 2010 conference. 14th - 16th July 2010 Berlin, Germany.*
197. Nene, N. R., Garcia-Ojalvo, J. & Zaikin, A. (2011) *submitted to Physical Review Letters.*
198. Nicolis, C. & Nicolis, G. (2005) *New Journal of Physics* **7**.
199. Nicolis, C. & Nicolis, G. (2003) *Phys. Rev E Stat. Nonlin. Soft Matter Phys.* **67**, 046211.
200. Nicolis, C. & Nicolis, G. (2000) *Physical Review e* **62**.
201. Nicolis, G. & Nicolis, C. (2005) *Physica A* **351**, 22-39.
202. Nicolis, G. & Prigogine, I. (1981) *Proc. Natl. Acad. Sci. U. S A* **78**, 659-663.
203. Nishikawa, T., Gulbahce, N. & Motter, A. E. (2008) *PLoS. Comput. Biol.* **4**, e1000236.
204. Noble, M. E. M., Endicott, J. A. & Johnson, L. N. (2004) *Science* **303**, 1800-1805.

205. Oktem, H., Pearson, R. & Egiazarian, K. (2003) *Chaos* **13**, 1167-1174.
206. Orton, R. J., Sturm, O. E., Vyshemirsky, V., Calder, M., Gilbert, D. R. & Kolch, W. (2005) *Biochem. J.* **392**, 249-261.
207. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & van, O. A. (2004) *Nature* **427**, 737-740.
208. Papin, J. A., Hunter, T., Palsson, B. O. & Subramaniam, S. (2005) *Nat Rev Mol. Cell Biol.* **6**, 99-111.
209. Peter E.Kloeden & Eckhard Platen (1999) *Numerical solution of stochastic differential equations* (Springer).
210. Pfeuty, B. & Kaneko, K. (2009) *Phys. Biol.* **6**, 046013.
211. Philipps, A. C. L. A correct abstract machine for the stochastic picalculus. Proc.of BIO-CONCUR'04. Electronic Notes in Theoretical ComputerScience. 2004. Elsevier. Ref Type: Conference Proceeding
212. Pinney, J. W., Westhead, D. R. & McConkey, G. A. (2003) *Biochem. Soc. Trans.* **31**, 1513-1515.
213. Plahte, E., Mestl, T. & Omholt, S. W. (1998) *J. Math. Biol.* **36**, 321-348.
214. Plasson, R., Kondepudi, D. K., Bersini, H., Commeyras, A. & Asakura, K. (2007) *Chirality* **19**, 589-600.
215. Priami, C., Regev, A., Shapiro, E. & Silverman, W. (2001) *Information Processing Letters* **80**, 25-31.
216. Quintaje, S. B. & Orchard, S. (2008) *Molecular & Cellular Proteomics* **7**, 1409-1419.
217. Rand, D. A. (2008) *J. R. Soc. Interface* **5 Suppl 1**, S59-S69.
218. Rateitschak, K. & Wolkenhauer, O. (2007) *Math. Biosci.* **205**, 163-179.
219. Raymond, E., Faivre, S. & Armand, J. P. (2000) *Drugs* **60**, 15-23.
220. Resat, H., Ewald, J. A., Dixon, D. A. & Wiley, H. S. (2003) *Biophys. J.* **85**, 730-743.
221. Robert A.Weinberg (2007) *The Biology of Cancer* (Garland Science, Taylor & Francis Group,LLC).
222. Roeder, I. & Glauche, I. (2006) *J. Theor. Biol.* **241**, 852-865.
223. Saez-Rodriguez, J., Kremling, A., Conzelmann, H., Bettenbrock, K. & Gilles, E. D. (2004) *Control Systems Magazine, IEEE* **24**, 35-52.
224. Sams-Dodd, F. (2005) *Drug Discov. Today* **10**, 139-147.
225. Sanchez, L. & Thieffry, D. (2001) *J. Theor. Biol.* **211**, 115-141.
226. Sasagawa, S., Ozaki, Y., Fujita, K. & Kuroda, S. (2005) *Nat Cell Biol.* **7**, 365-373.

227. Sauro, H. M. (2008) *Mol. Syst. Biol.* **4**, 166.
228. Sauro, H. M. & Kholodenko, B. N. (2004) *Prog. Biophys. Mol. Biol.* **86**, 5-43.
229. Sawyers, C. (2004) *Nature* **432**, 294-297.
230. Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D. & Muller, G. (2002) *Nature Biotechnology* **20**, 370-375.
231. Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D. & Friedman, N. (2003) *Nat Genet.* **34**, 166-176.
232. Seger, R., Rodeck, U. & Yarden, Y. (2008) *IET Syst. Biol.* **2**, 1-4.
233. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. (2002) *Nat Genet.* **31**, 64-68.
234. Shmulevich, I., Dougherty, E. R., Kim, S. & Zhang, W. (2002) *Bioinformatics* **18**, 261-274.
235. Shvartsman, S. Y., Muratov, C. B. & Lauffenburger, D. A. (2002) *Development* **129**, 2577-2589.
236. Smolen, P., Baxter, D. A. & Byrne, J. H. (1998) *Am. J. Physiol* **274**, C531-C542.
237. Smolen, P., Baxter, D. A. & Byrne, J. H. (1999) *Am. J. Physiol* **277**, C777-C790.
238. Somsen, O. J., Siderius, M., Bauer, F. F., Snoep, J. L. & Westerhoff, H. V. (2002) *J. Theor. Biol.* **218**, 343-354.
239. Sornette, D. (2006) *Critical phenomena in natural sciences: chaos, fractals, self-organization and disorder.* (Springer-Verlag).
240. Sornette, D., Johansen, A. & Bouchaud, J.-P. (1996) *J. Phys. I France* **6**, 167-175.
241. Stan Tsai, C. (2002) *An introduction to Computational Biochemistry.* (**Wiley-Liss, Inc, New York**).
242. Suzuki, H. e. al. (2009) *Nature Genetics* **41**, 553-562.
243. Swain, P. S., Elowitz, M. B. & Siggia, E. D. (2002) *Proc. Natl. Acad. Sci. U. S A* **99**, 12795-12800.
244. T.Erneux & and J.P.Laplante (1989) *J. Chem. Phys.* **90**, 6129-6134.
245. T.Erneux, E.L.Reiss, L.J.Holden & M.Georgiou (1990) in *Dynamic bifurcations*, ed. Benoit, E. (Springer-Verlag, pp. 14-28.
246. T.T.Tsotsis, R.C.Sane & T.H.Lindstrom (1988) *AIChE J.* **34**, 383.
247. Tennyson, C. N., Klamut, H. J. & Worton, R. G. (1995) *Nat Genet.* **9**, 184-190.
248. The FANTOM Consortium and the Riken Omics Science Center (2009) *Nat Genet.* **41**.
249. Thieffry, D. & Thomas, R. (1995) *Bull. Math. Biol.* **57**, 277-297.
250. Thomas, R. (2006) *lee Proceedings Systems Biology* **153**, 140-153.

251. Thomas, R. & d'Ari, R. (1990) *Biological Feedback* (CRC Press, Boca Raton, FL).
252. Thomas, R. & Kaufman, M. (2001) *Chaos* **11**, 180-195.
253. Thomas, R. & Kaufman, M. (2001) *Chaos* **11**, 170-179.
254. Tian, T. & Burrage, K. (2006) *Proc. Natl. Acad. Sci. U. S A* **103**, 8372-8377.
255. Tyson, J. J., Chen, K. C. & Novak, B. (2003) *Current Opinion in Cell Biology* **15**, 221-231.
256. Ullner, E., Zaikin, A., Volkov, E. I. & Garcia-Ojalvo, J. (2007) *Phys. Rev Lett* **99**, 148103.
257. Uptain, S. M., Kane, C. M. & Chamberlin, M. J. (1997) *Annu. Rev Biochem.* **66**, 117-172.
258. Van Kampen, N. G. (1992) *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).
259. Ventura, J. J., Hubner, A., Zhang, C., Flavell, R. A., Shokat, K. M. & Davis, R. J. (2006) *Mol. Cell* **21**, 701-710.
260. W.J.Scharpf, M.Squicciarini, D.Bromley, C.Green, J.R.Tredicce & L.M.Narducci (1987) *Opt. Commun.* **63**, 344-348.
261. Waddington, C. H. (1957) *The strategy of the genes: A discussion of some aspects of theoretical biology* (Allen & Unwin, London).
262. Watters, J. W. & Roberts, C. J. (2006) *Mol. Cancer Ther.* **5**, 2444-2449.
263. Werner, S. L., Barken, D. & Hoffmann, A. (2005) *Science* **309**, 1857-1861.
264. Whitehurst, A., Cobb, M. H. & White, M. A. (2004) *Mol. Cell Biol.* **24**, 10145-10150.
265. Whitfield, M. L., Sherlock, G., Saldanha, A. J., Murray, J. I., Ball, C. A., Alexander, K. E., Matese, J. C., Perou, C. M., Hurt, M. M., Brown, P. O. et al. (2002) *Mol. Biol. Cell* **13**, 1977-2000.
266. Whitmarsh, A. J. (2007) *Biochim. Biophys. Acta* **1773**, 1285-1298.
267. Widder, S., Macia, J. & Sole, R. (2009) *PLoS One.* **4**, e5399.
268. Widder, S., Schicho, J. & Schuster, P. (2007) *J. Theor. Biol.* **246**, 395-419.
269. Wiley, H. S., Shvartsman, S. Y. & Lauffenburger, D. A. (2003) *Trends Cell Biol.* **13**, 43-50.
270. Willeboordse, F. H. & Kaneko, K. (2005) *Physical Review e* **72**.
271. Wolf, D. M. & Arkin, A. P. (2003) *Curr. Opin. Microbiol.* **6**, 125-134.
272. Wolkenhauer, O., Ullah, M., Kolch, W. & Cho, K. H. (2004) *IEEE Trans. Nanobioscience.* **3**, 200-207.
273. Wong, P. K., Yu, F. Q., Shahangian, A., Cheng, G. H., Sun, R. & Ho, C. M. (2008) *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5105-5110.

274. Yamamoto, T., Ebisuya, M., Ashida, F., Okamoto, K., Yonehara, S. & Nishida, E. (2006) *Curr. Biol.* **16**, 1171-1182.
275. Yang, S. H., Sharrocks, A. D. & Whitmarsh, A. J. (2003) *Gene* **320**, 3-21.
276. Yarden, Y. & Shilo, B. Z. (2007) *Cell* **131**, 1018.
277. Yeh, P., Tschumi, A. I. & Kishony, R. (2006) *Nat Genet* **38**:489-494 **Nat Genet**, 489-494.
278. Yoon, J. & Deisboeck, T. S. (2009) *Plos One* **4**, e4560.
279. Yu.L.Klimontovich (1991) *Turbulent Motion and Structure of Chaos* (Kluwer Academic Publishers, Dordrecht).
280. Zhang, J., Yang, P. L. & Gray, N. S. (2009) *Nat Rev Cancer* **9**, 28-39.
281. Zhang, L., Roberts, T. M. & Gjoerup, O. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10296-10301.
282. Zimmermann, G. R., Lehar, J. & Keith, C. T. (2007) *Drug Discov. Today* **12**, 34-42.