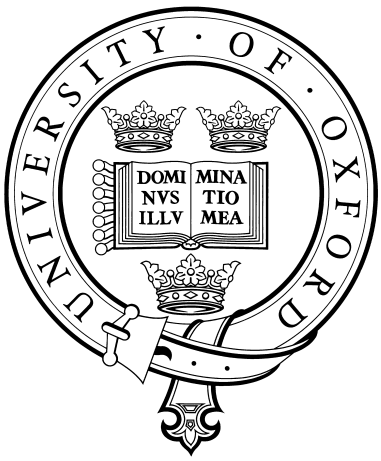
Spatial Stochastic Modelling of Biological Processes



Anna C. Jones

Exeter College

University of Oxford

This thesis is submitted for the degree of

Doctor of Philosophy in Computational Biology

November 2018

Acknowledgements

It is very difficult to name everyone who got me to the point of handing in this thesis over the many, many years, so I'll just list as many as I can think of. I'm not a very mushy person, so I'll keep it short.

Academia: Of course, I have to start by thanking Kevin Burrage & Helen Byrne for putting up with me for so many years. They have provided me with advice and supervision long after I deserved it and I am grateful to them for actually getting me this far. For letting me even start this project, I am grateful to the DTC (especially all the people in Systems Biology DTC) for accepting me onto the programme and nurturing me through the difficult first year. They made it really fun and it felt like a home by the end of our time there. For additional support, I'd like to mention Martin Robinson who helped me with Smoldyn, which was invaluable to my thesis. Finally, I have to thank Dave Gavaghan who has been so extremely supportive. It is not an understatement to say that without Dave's support during the last two years, this thesis never would have been finished.

Friends: There are far, far too many people to mention. Everyone in CompBio, particularly Sara, who was always un-waiveringly kind to everyone in the department and was a great partner for organising events in the department. Everyone from the DTC, particularly Stephen and Debbie for being the best desk-buddies to work between. Federico and Joe, who were always the funniest people to hang out with in DTC projects, at the pub, and at the Aston Street house. Hannah, Anna, Kylie, and Aimee who made the whole process enjoyable, and without whom I would have gone completely insane throughout various stages of my PhD. Kylie, Laura, Georgina, and Matt, the best housemates anyone could ask for. Morrell Avenue was the best place to live and I can't thank you guys enough for inviting me to live with you (double mention for Kylie because her hilarious stories were a reason to come to work every day). From Aston Street, I'd like to thank Matz and

Boozy for cooking for me and letting me stay over with Jaime for so many years. Other mexicans, like Gravy and Andrea, for always being the most easy-going people, always up for a drink and a BBQ in the garden. The best of the rest: Anouk, Charles and Camille, the Ruuuuusty Wires, like Midori, Bea, Majo and Eva, the old salsa crew, like Adam, Justin and DJ, and Petar, who was just there to Skype whenever I needed a little rant. Finally, I can't forget about all my friends from home, like Beard, David (for all the teas and scones), Kate, Damon, Liam, Anna1, Ryan, Paul, Bailey, Kev, Kat, Meg, George and Amy, without whom I would have never believed I was smart enough to even apply for the PhD, and who never made me feel guilty for not visiting home nearly enough. There are many other people to thank, so apologies if I forgot about you.

General thanks to my entire family for always supporting me and giving me the confidence to do this in the first place. I won't be more explicitly nice to them otherwise they might cry (or think that I'm dying).

Last but not least, I would like to thank Jaime for being such a positive, lovely person, who always tells me to do whatever makes me happy. I never would have gotten to this point in my life without him. I am eternally grateful for his continuing support over the last 7 years (and counting).

Abstract

Computational simulation of cellular processes can provide us with important insights into the dynamics of a cell. In particular, stochastic simulation has proven to be an invaluable tool for modelling sub-cellular environments, due to the intrinsic noise of not knowing when reactions will occur in a cell and which molecules will react. This uncertainty is amplified when there are limited numbers of molecules within a cell, as the probability of reactions occurring becomes small.

One aspect of cellular dynamics that is often overlooked in stochastic models is the effect of spatial constraints on the behaviour of a cell. Cellular processes depend not only on the number of molecules present, but also on the distribution of these molecules within the cell and how they move and interact with each other. For this reason, we hypothesised that spatial stochastic simulations of cells would prove to be more effective at recreating experimentally observed cells than spatially-averaged models. To begin testing this hypothesis, we create and compare spatially-dependent models against spatially-averaged models for three biological processes. The first two processes we study involve post-transcriptional gene regulation by small RNAs. An overarching theme across each biological process that we have studied in this thesis is the implication in tumourigenesis. We present examples in which stochasticity and spatial detail detail can produce different modelling results when compared with non-spatial models. We also show that, in many cases, the biological systems we have chosen to model behave somewhat deterministically. We have shown that there are threshold values for certain parameters, for example, molecular concentrations, at which the systems largely follow the dynamics of the deterministic models.

Gene expression is also inherently stochastic; low numbers of genes, proteins and mRNAs are found within many cells, and many genes are controlled by probabilistic interactions between these molecules. In particular, if some molecular species are scarce then the likelihood of these molecules finding a binding partner becomes probabilistic. Therefore, stochastic modelling is an appropriate technique for studying this biological process. In this thesis, we study cellular signalling processes as stochastic processes. Specifically, we built a stochastic model of the canonical Wnt/β-catenin signalling pathway and compared this with existing deterministic models. We present here some preliminary results and a proof-of-concept stochastic. An obvious path for further investigation

would be to expand on this model by including cell-cell adhesion and movement of molecules between cellular compartments in a crowded environment.

Contents

1 Literature Review **1**

1.1 Introduction 1

1.2 Post-Transcriptional Gene Regulation in the Cytoplasm 2

1.2.1 The *PTEN* Tumour Suppressor Gene 3

1.2.2 MicroRNAs 6

1.2.3 Post-transcriptional Regulation of *PTEN* 8

1.3 RNA Interference in the Nucleus 16

1.3.1 RNA Interference Mechanisms 17

1.3.2 Evidence for RNA Interference in the Nucleus 17

1.4 Wnt Signalling 20

1.4.1 The Canonical Wnt/βCatenin Signalling Pathway 22

1.4.2 Mutations in The Wnt Signalling Pathway 26

1.4.3 The Role of βCatenin in Cell-Cell Adhesion 27

1.5 Mathematical and Computational Modelling 27

1.5.1 Deterministic Modelling 28

1.5.2 Stochastic Modelling 28

1.5.3 Spatial Modelling 35

1.5.4 Smoldyn 58

1.5.5 Existing Models of *PTEN* Post-Transcriptional Regulation 62

1.5.6 Existing Models of RNAi 64

1.5.7 Markov Chain Model of *PTEN* Post-Transcriptional Regulation and RNAi . 66

1.5.8 Existing Models of the Canonical Wnt Signalling Pathway 67

2 RNAi In The Nucleus **73**

2.1 Introduction . . . 73

2.2 RNA Escape from the Nucleus . . . 74

2.3 Markov Model . . . 75

2.4 Markov model with copycats . . . 80

2.4.1 Extension to One-Dimensional Markov Model . . . 84

2.5 Conclusion to Markov Model . . . 87

2.6 ODE model . . . 88

2.6.1 ODE Parameter Values . . . 89

2.7 SSA model . . . 90

2.7.1 SSA Parameter Values . . . 91

2.8 Development of Spatial Stochastic Model . . . 91

2.8.1 Smoldyn Model Formulation . . . 92

2.9 Results and Discussion . . . 96

2.9.1 Extension of Markov Model Simulations . . . 96

2.9.2 Effect of Obstacles . . . 97

2.9.3 MSD calculation . . . 101

2.9.4 Effect of Copycats on RNA Degradation (Without Escape) . . . 103

2.10 Conclusion . . . 104

3 PTEN post-transcriptional regulation **109**

3.1 Introduction . . . 109

3.2 ODE model . . . 111

3.2.1 ODE Parameter Values . . . 112

3.3 SSA model . . . 113

3.3.1 SSA Parameter Values . . . 114

3.4 Development of Spatial Stochastic Model . . . 114

3.4.1 Smoldyn Model Formulation . . . 114

3.5 Results . . . 123

3.5.1 Effect of Obstacles on Molecular Diffusion . . . 123

3.5.2 Effect on PTEN mRNA Degradation . . . 135

3.5.3 Effect of Electrostatic Interactions with Obstacles . . . 143

3.5.4 Effect of Copycats, CeRNAs and Binding Site Reversibility 145
3.5.5 SSA and ODE comparisons 152
3.5.6 SSA Results in Isolation 161
3.6 Discussion 163
3.6.1 Hypotheses 1-3: Effect of Macromolecular Crowding 165
3.6.2 Hypotheses 4-5: Effect on RNA Degradation 166
3.6.3 Summary 168

4 The Canonical Wnt/β-catenin Signalling Pathway **171**
4.1 Introduction 171
4.2 Stochastic Model 174
4.2.1 Schlögl model 175
4.2.2 Smoldyn Model Formulation 177
4.3 Results 181
4.4 Discussion 181
4.4.1 Multi-cellular modelling 182

5 Outlook **185**

Appendices **189**

A Additional Figures **190**

B Smoldyn Code **197**

C Maria Bruna Mean Squared Displacement Code **203**

D Maria Bruna Mean Squared Displacement Smoldyn Model **205**

E PTEN SSA Python Model **209**

F PTEN ODE Python Model **235**

List of Figures

1.1 Gene regulatory network and post-transcriptional regulation 3
1.2 Position of *PTEN* at the q23 locus on chromosome 10. 4
1.3 PTEN acting on PI3K-Akt pathway 5
1.4 A schematic diagram of miRNA biogenesis in mammals 7
1.5 A schematic of miRNA−mRNA seed binding 7
1.6 Post-transcriptional regulation with ceRNAs and microRNAs 10
1.7 Switch-like behaviour due to ceRNAs 12
1.8 A schematic of the miRNA sponging by ceRNAs 13
1.9 A schematic of copycat functionality 15
1.10 A schematic of the miRNA-PTEN post-transcriptional regulation 15
1.11 Schematic of nucleus and nuclear pores 16
1.12 Inhibition of miR-15a by miR-709 19
1.13 Wnt gradient in the human intestinal crypt 21
1.14 A Markov chain schematic of the weather prediction model 30
1.15 Excluded volume interactions 37
1.16 Diffusion of small molecules in crowded cellular space 48
1.17 A schematic of the movement within a 2D grid 53
1.18 A schematic of the possible 2D domains for molecule movement 53
1.19 Cellular automaton 55
1.20 A schematic of bi-molecular reactions in Smoldyn 59
1.21 A schematic of diffusion, binding and unbinding of molecules 59
1.22 A schematic of macromolecular crowding in a 2D simulation 61
1.23 A schematic of excluded volume interactions between two molecules 61

1.24 A schematic of the miRNA catalyticity 64

2.1 The one-dimensional transition diagram for siRNA binding to mRNAs 77
2.2 The two-dimensional transition diagram for miRNA and copycat binding to pri-miRNAs 82
2.3 Figure taken from Holcman et. al ((Holcman et al., 2014)). The escape probability as a function of the threshold, T. Reversible binding (black, squares) compared to the irreversible binding (red, dashed). Total binding sites$N_b = 20$ and forward binding rate $k_f = 1$. Reversible binding appears to decrease the probability of escape when copycats are used, particularly for $T = 11$. 85
2.4 A schematic of the nucleus in 2D 92
2.5 Smoldyn set up of nucleus and nuclear pores 97
2.6 Diffusion of siRNAs, RNAs and obstacles for varying crowding fractions and mobile obstacles 99
2.7 Diffusion of siRNAs, RNAs and obstacles for low binding rates $k_f\tau = 0.1$, stationary vs. mobile obstacles, and varying mobile fractions ($\theta = 0.2$ and $\theta = 0.3$) 100
2.8 Mean squared displacement of RNA and siRNA for $\theta = 0.2$ and big obstacles 102
2.9 Mean squared displacement of RNA and siRNA for $\theta = 0.3$ and big obstacles 103
2.10 Mean squared displacement of RNA and siRNA for $\theta = 0.2$ and normal obstacles . . 104
2.11 Fraction of mRNA degraded for varying numbers of copycats and reversibility 105
2.12 Fraction of mRNA degraded for varying number of binding sites and degradation thresholds 106

3.1 A schematic of an miRNA−mRNA complex with seed sequence and a bulge in the nucleotide sequence 110
3.2 A schematic of the miRNA binding sites in the 3′UTR of a PTEN mRNA 115
3.3 A schematic of the level of *PTEN* repression as a result of copycats and ceRNAs . . 118
3.4 Smoldyn simulation screenshots with and without crowding 119
3.5 A schematic of excluded volume reactions, between two non-reactive molecular species and between two obstacles 119
3.6 A schematic of the cell in 2D with microRNAs, mRNAs and ceRNAs 120
3.7 Typical cell schematic for displaying molecule trajectories 124

3.8 Graph showing the typical MSD behaviour for Brownian, anomalous, directed and confined diffusion . . . 125
3.9 MSD behaviour on linear and log-log scale with and without traps . . . 126
3.10 Example trajectories of RNAs and miRNAs in Smoldyn . . . 128
3.11 Example trajectories of RNAs and miRNAs in Smoldyn with faster diffusion . . . 129
3.12 Isolated RNA trajectory from Figure 3.11. . . . 130
3.13 Comparison of a typical Brownian and Lévy flight trajectory . . . 130
3.14 PTEN mRNA MSD under varying volume fractions and fast diffusion . . . 130
3.15 PTEN mRNA MSD under varying volume fractions . . . 131
3.16 PTEN mRNA MSD under varying volume fractions for stationary obstacles . . . 132
3.17 MicroRNA MSD under varying volume fractions for mobile and stationary obstacles 132
3.18 MSD for populations of molecules . . . 133
3.19 Distance travelled by miRNAs and mRNAs over time periods of $\Delta t = 0.15s$ (for time step, $\delta t = 0.0001$s) for varying volume fractions . . . 134
3.20 Distance travelled by miRNAs and mRNAs over time periods of $\Delta t = 0.15s$ (for time step, $\delta t = 0.0001$s) for varying volume fractions and fast diffusion of obstacles . . . 135
3.21 Distance travelled by miRNAs and mRNAs over time periods of $\Delta t = 0.15s$ (for time step, $\delta t = 0.0001$s) for varying volume fractions and stationary obstacles . . . 136
3.22 Cumulative distance travelled by miRNAs and mRNAs . . . 137
3.23 Example trajectories of RNAs and miRNAs ($t_f = 15s$), where miRNAs are "NORMBIG" . . . 138
3.24 Example trajectories of RNAs and miRNAs ($t_f = 15s$), where miRNAs are "SLOWBIG" 139
3.25 Fraction of the PTEN mRNA population degraded comparing volume fraction and obstacle mobility . . . 139
3.26 Fraction of the PTEN mRNA population degraded comparing obstacle size and mobility for a range of volume fractions . . . 140
3.27 Fraction of the PTEN mRNA population degraded comparing miRNA diffusion coefficents for increased miRNA size for a range of volume fractions . . . 140
3.28 Fraction of the PTEN mRNA population degraded for big miRNAs with unchanged diffusion coefficient . . . 141

3.29 Fraction of the PTEN mRNA population degraded for big miRNAs moving with unchanged diffusion coefficient 142
3.30 Fraction of the PTEN mRNA population degraded when both miRNAs and RNAs have $R = 0.15\mu$m, i.e. RNA-RNA interactions 144
3.31 Fraction of the PTEN mRNA population degraded when RNAs may undergo weak binding to obstacles 145
3.32 Fraction of the PTEN mRNA population degraded when ceRNAs are present or absent, and when copycats are present or absent 146
3.33 Fraction of the PTEN mRNA population degraded when $N_b = 3$, ceRNAs are absent, copycats are either present or absent, and binding is either reversible or irreversible . 148
3.34 Fraction of the PTEN mRNA population degraded when $N_b = 3$, ceRNAs are absent, copycats are either present or absent, and binding is either reversible or irreversible . 149
3.35 Fraction of the PTEN mRNA population degraded when ceRNAs are present (with increasing binding sites on ceRNAs) 150
3.36 Fraction of the PTEN mRNA population degraded when ceRNAs are present (with same number binding sites as PTEN mRNAs) 151
3.37 PTEN mRNA degradation rates when miRNA binding is cooperative 153
3.38 Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results, comparing high and low concentrations of RNAs 154
3.39 Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results, comparing slow and fast diffusion 155
3.40 Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results, comparing copycats present and absent 156
3.41 Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results, comparing volume fractions 157
3.42 Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results, comparing volume fraction, binding sites, and number of miRNAs . . . 158
3.43 Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results, comparing volume fraction, binding sites, and number of miRNAs with copycats 159

3.44 Fraction of the PTEN mRNA population degraded comparing SSA and ODE results, comparing volume fraction, and diffusion rates . . . 160
3.45 Fraction of the PTEN mRNA population degraded comparing SSA and ODE results and diffusion rates ($\theta = 0.3$) . . . 161
3.46 Fraction of the PTEN mRNA population degraded and unbound, using the SSA, without copycats . . . 162
3.47 Fraction of the PTEN mRNA population degraded and unbound, using the SSA, with copycats . . . 163
3.48 Fraction of the PTEN mRNA population degraded and unbound, using the SSA, varying copycats, reversibility, and number of binding sites . . . 164

4.1 Schematic for Schmitz et al. model with cell-cell adhesion . . . 173
4.2 Schematic for van Leeuwen et al. model . . . 173
4.3 Screenshot of the three-dimensional Smoldyn simulation including β-catenin, DC, TF and β-catenin/TF complexes . . . 179
4.4 Preliminary van Leeuwen results comparing Hypothesis 1, Hypothesis 2, No Wnt, and long-term analysis . . . 182
4.5 Different Smoldyn multi-cell models . . . 184

A.1 Fraction of the PTEN mRNA population degraded for big miRNAs moving with unchanged diffusion coefficient . . . 191
A.2 Fraction of the PTEN mRNA population degraded for varying miRNA sizes and diffusion coefficents . . . 192
A.3 Fraction of the PTEN mRNA population degraded for each miRNA size and diffusion coefficent . . . 193
A.4 Diffusion of siRNAs, RNAs and obstacles for varying crowding fractions and stationary vs. mobile obstacles . . . 194
A.5 Diffusion of siRNAs, RNAs and obstacles for varying crowding fractions and stationary obstacles . . . 194
A.6 Degradation and escape times for varying numbers of microRNAs . . . 195
A.7 Degradation and escape times for varying crowding fractions with one binding site . 196
A.8 Degradation and escape times for varying crowding fractions with 20 binding sites . 196

Chapter 1

Literature Review

1.1 Introduction

Gene expression is the process by which genes are transcribed into RNA and subsequently translated into proteins. Each step of gene expression is subject to control. Regulation at the gene level is known as transcriptional regulation; this usually involves direct mutations in the gene sequence, or modulation of transcription factors, the molecules tasked with transcribing the gene into RNA. Regulation at the RNA level is known as post-transcriptional, or translational, regulation; this type of regulation usually involves interruptions to the machinery responsible for translating the RNA into protein, or direct interference with the RNA itself. Genes can be up- or down-regulated to produce varying levels of particular proteins depending on the needs of a cell. For example, when a healthy cell becomes mutated, up-regulation of the genes involved in apoptosis (programmed cell death) and/or down-regulation of genes involved in cell proliferation ensures that the mutated cell is removed before it can damage the organism. These gene expression mechanisms, which are important for maintaining cell homeostasis, are often dysregulated in diseased cells. Abberant gene regulation can lead to many different diseases, including Alzheimer's and Huntington's diseases. In this study, however, we focus on the dysregulation of genes that result in carcinogenesis. Genes whose activation may lead to cancer are called proto-oncogenes, whereas genes whose activation can counteract the cancer-causing mechanisms are called tumour suppressor genes.

Tumourigenesis, also known as carcinogenesis or oncogenesis, is the mechanism by which cancerous cells form from normal, healthy cells and undergo uncontrolled growth and proliferation, to form a mass of cancerous cells referred to as a tumour or neoplasm. In 2012, the annual number of

cancer-related deaths was approximately 160,000 in the UK; this number is predicted to rise as the population ages, since the likelihood of cancer increases with age ($\sim 77\%$ of all deaths are attributed to those aged 65 and over) [1]. In normal cells, proliferation and programmed death via apoptosis are tightly regulated via gene expression (Alberts et al., 2002). Activation of proliferation-inducing genes, such as the oncogene *Myc*, or silencing of apoptosis-inducing genes, such as the tumour suppressor gene *p53*, may lead to cancer (Alberts et al., 2002).

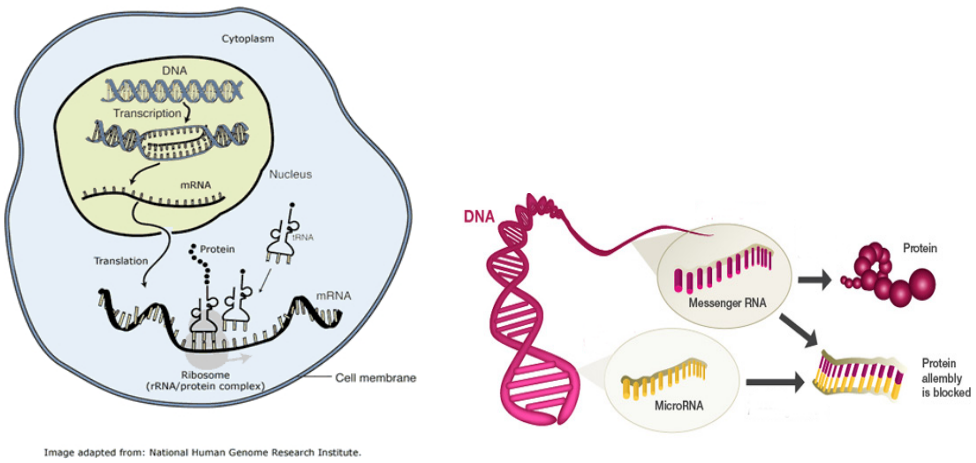
In this thesis, we study two processes involved in gene regulation: Post-transcriptional regulation and Wnt signalling. In Sections 1.3 and 1.2, respectively, we describe the process of RNA interference (a form of post-transcriptional regulation) in the nucleus, and the related topic of post-transcriptional regulation of the tumour suppressor gene, *PTEN* in the cytoplasm of mammalian cells. In Section 1.4, we introduce the the canonical Wnt signalling pathway, an example of transcriptional regulation via modulation of transcription factors. The aim of this work is to determine whether spatial modelling of these processes is necessary to replicate behaviours observed experimentally. To this end, we create and compare a spatially-resolved model of each process against spatially-averaged models, using biological data to validate the results where possible. For each process, we highlight its potential involvement in cancer as this is the focus of the thesis. We end, in Section 1.5, by reviewing the existing literature on spatial and stochastic modelling, with specific mention of existing efforts to model each biological process that we have studied, and an explanation of how the work in this thesis relates to these models.

1.2 Post-Transcriptional Gene Regulation in the Cytoplasm

Gene regulatory networks consist of groups of genes whose function can be positively or negatively affected by various signals and/or molecular interactions, which promote or inhibit protein production (El-Samad et al., 2005). When genes are activated, they are transcribed into messenger RNA (mRNA) molecules, which are in turn translated into proteins (see Figure 1.1(a)) (El-Samad et al., 2005).

Post-transcriptional regulation is a process that occurs between the transcription and translation of a gene (Alberts et al., 2002). There are many types of post-transcriptional regulation. For

[1] www.cancerresearchuk.org/cancer-info/cancerstats/mortality



(a) (b)

Figure 1.1: (a) Gene transcription and translation into protein in the cell. (b) The impact of post-transcriptional regulation by microRNAs.

example, alternative transcript splicing involves "cutting" RNA at different nucleotide positions, to produce different protein products from a single gene. Another example is inhibition of nuclear export, whereby a gene transcript is prevented from entering the cytoplasm to be translated into protein. RNAs can also be regulated by small RNAs, in which target transcripts are bound to small RNAs with complementary nucleotide sequences, leading to transcript degradation or translational repression. We focus on this latter process, as it is known to play an important role in tumourigenesis for particular target tumour suppressor genes and oncogenes. When tumour suppressor genes are inactivated, for example through deletion or mutation of the gene, cancerous cells are more likely to form. In this thesis, we focus on post-transcriptional regulation of the tumour suppressor gene, *PTEN* (phosphatase and tensin homolog), as several recent experimental studies have focused on its involvement in a vast network of competing RNAs that plays a key role in cancer (see Figure 1.1(b)) (Tay et al., 2011; Sumazin et al., 2011; Karreth et al., 2011; Cesana et al., 2011).

1.2.1 The *PTEN* Tumour Suppressor Gene

PTEN (phosphatase and tensin homolog), also known as *MMAC1* (mutated in multiple advanced cancers), was first identified as a novel tumour suppressor gene in 1997 by two independent research groups (Li and Sun, 1997; Steck et al., 1997). The *PTEN* gene is located at the 10q23 locus of the human genome, i.e. at position 23.3 on the long arm (q) of chromosome 10, a frequently mu-

tated genomic region implicated in a variety of human cancers, such as brain and prostate cancer (see Figure 1.2) (Salmena et al., 2008; Di Cristofano and Pandolfi, 2000). In healthy cells, tumour suppressor genes act to maintain healthy cell growth and proliferation, in direct opposition to oncogenes. There are many different tumour suppressor genes in the human genome, acting on a wide range of different pathways and processes within the body, in order to maintain healthy cellular growth. Most tumour suppressor genes act via the gene transcript and/or protein product, affecting cellular processes through direct and indirect mechanisms. For example, the *BRCA1* gene (breast cancer 1, early onset) acts as a tumour suppressor via its protein product BRCA1, which is directly involved in DNA repair. Conversely, the protein product of the *p53* gene acts indirectly as a tumour suppressor. The p53 protein is involved in apoptosis which, if deregulated, may lead to irregular and uncontrolled growth, one of the hallmarks of cancer.

As shown in Figure 1.3, the *PTEN* gene acts via its protein product that antagonises the Phosphoinositide 3-kinase (PI3K)-Akt signalling pathway. PTEN inhibits the production of Akt, a protein kinase that promotes cell proliferation, growth and survival, via antagonism of the PI3K enzyme (Pedrero et al., 2005). *PTEN* accomplishes this by catalysing the dephosphorylation of PtdIns(3,4,5)P3 (phosphatidylinositol 3,4,5-triphosphate) to PtdIns(4,5)P2 (phosphatidylinositol 4,5-bisphosphate) (Georgescu, 2011). Akt promotes cell survival and inhibits apoptosis via a number of pathways (some of which are shown in Figure 1.3). By inhibiting Akt, *PTEN* promotes programmed cell death (or apoptosis). Apoptosis is an important process in healthy adult cells, resulting in destruction of damaged, mutated or cancerous cells (Alberts et al., 2002). Normal *PTEN* expression is, therefore, extremely important in preventing tumour development.

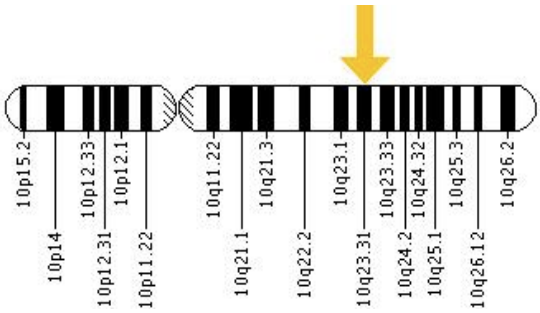


Figure 1.2: Position of *PTEN* at the q23 locus on chromosome 10.

PTEN is one of the most frequently mutated tumour suppressor genes in human cancer patients (Hollander et al., 2011), occurring in approximately 50% of prostate cancers and more than 20% of

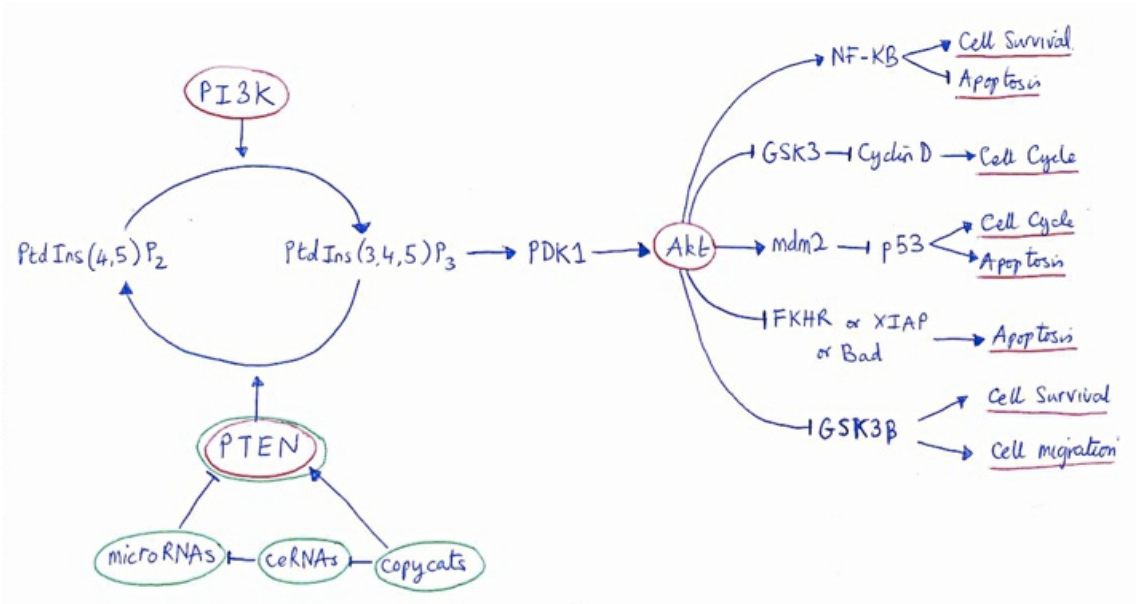


Figure 1.3: PTEN acting on PI3K-Akt pathway and the interaction of microRNAs, ceRNAs and copycats on PTEN (detailed in Section 1.2.3).

primary glioblastomas, with less frequent reports in other types of cancer including head and neck cancer ($\sim$ 12%), lung cancer ($\sim$ 9%), thyroid cancer ($\sim$ 7%) and sporadic breast cancer ($\sim$ 5%) (Chu and Tarnawski, 2004). Loss of *PTEN* is also implicated in other human diseases, such as diabetes and autism (Tamguney and Stokoe, 2007). Typically, monoallelic deletion of *PTEN* is sufficient to initiate tumourigenesis, with the second allele commonly lost in the advanced stages of tumour development and in metastatic cancers (Salmena et al., 2008). Besides gene mutations and deletions, *PTEN* has also been shown to lead to cancer through irregular subcellular compartmentalisation of PTEN (Perren et al., 2000), epigenetic gene silencing during transcription or translation, post-translational modifications of PTEN, and post-transcriptional down-regulation of *PTEN* by microRNAs (Meng et al., 2006; Song et al., 2012). Each of these mechanisms falls into the so-called 'continuum model for tumour suppression', in which PTEN protein production can be fine-tuned, as opposed to the alternative 'two-hit model for tumour suppression' in which genes are regulated via heterozygous (50%) or homozygous (100%) gene knockout (Berger et al., 2011). In particular, the study of post-transcriptional regulation of *PTEN* is of particular importance as it is not yet fully understood. It offers a new area for exploring potential cancer therapies in patients with deregulated *PTEN* expression. In Section 1.2.2 and Section 1.2.3, we explain the nature of microRNAs and how they are involved in the post-transcriptional regulation of *PTEN*.

1.2.2 MicroRNAs

MicroRNAs (often referred to as "miRNAs" or "miRs") are endogenous, non-coding RNAs that play a key role in gene regulation in multicellular organisms (Fabian et al., 2010). In humans, miRNA genes are found in all chromosomes excluding the Y chromosome (Feng et al., 2009). Since the discovery of the first miRNA gene, *lin-4*, in *Caenorhabditis elegans* by Lee et al. in 1993 (Lee et al., 1993), miRNAs have been found in many different species; As of miRBase Release 22 (March 2018), the miRBase Registry [2] currently lists more than 38000 different mature miRNAs in more than 270 species, including around 2650 distinct mature human miRNAs (Griffiths-Jones, 2006). They are small, approximately 22 nucleotides in length, and yet are predicted to target one-third of all genes in the human genome, with each miRNA expected to target around 200 different mRNAs (Hon and Zhang, 2007). MiRNAs have been observed to be both up- and down-regulated in a variety of human cancers (Ambs et al., 2008; Calin et al., 2002; Iorio et al., 2005; Ventura and Jacks, 2009).

As shown in Figure 1.4, biogenesis (or maturation) of miRNAs begins in the nucleus where primary miRNA (pri-miRNA) is transcribed and converted into precursor miRNA (pre-miRNA), a 60-70 nucleotide RNA, by the Drosha-DGCR8 complex, known as Microprocessor. The pre-miRNA is then exported to the cytoplasm where it is processed further into miRNA by another enzyme, Dicer. The result of processing the pre-miRNA is a $\sim$22 nucleotide duplex, which is subsequently separated into two miRNA strands by the enzyme Helicase. Although either strand of the duplex may have the potential to function as an miRNA, usually only one strand is active and binds to a multi-protein complex known as the RNA-induced silencing complex (RISC); the other strand is degraded. The primary component of RISC is an Argonaute protein that is activated upon miRNA-RISC binding and interacts directly with the miRNA (Bartel, 2004). In humans, it is the Argonaute protein, AGO2, that is responsible for degrading complementary mRNAs via endonucleoltyic cleavage (Meister et al., 2004; Liu, 2004).

MiRNAs, along with short interfering RNAs (siRNAs), function primarily in the cytoplasm. They act by base-pairing with complementary nucleotide sequences within messenger RNA (mRNA) (see Figure 1.5). Upon miRNA$-$mRNA binding, mRNA translation rates are altered either by mRNA molecular degradation via endonucleolytic cleavage, or by translational inhibition (Valencia-Sanchez et al., 2006). MiRNAs can also indirectly cause translational activation, for example, by blocking

[2] www.mirbase.org

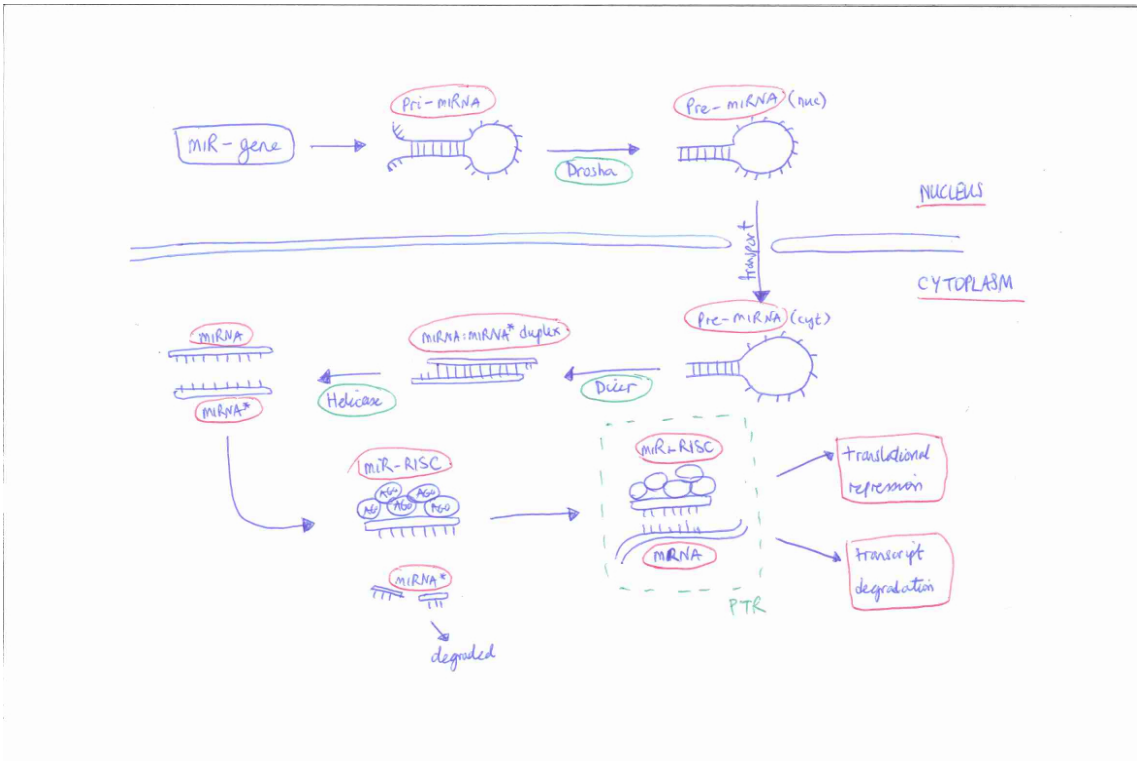


Figure 1.4: A schematic diagram of miRNA biogenesis in mammals, with eventual incorporation into RISC and silencing of target mRNAs.

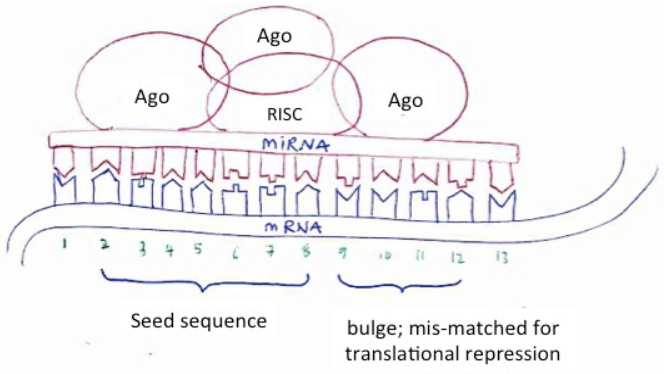


Figure 1.5: A schematic of miRNA−mRNA binding, showing the seed region at 2-9 nucleotides with a bulge at nucleotides 9-12.

adjacent binding sites on mRNAs which, if these adjacent sites are usually targeted by antagonising miRNAs, would indirectly upregulate mRNA translation (Brodersen and Voinnet, 2009)

Over the last decade, research into small-RNA-mediated gene regulatory mechanisms has led to the discovery that mRNAs can be regulated by multiple different miRNAs, and that each type of miRNA can have many (possibly hundreds of) target transcripts (Valencia-Sanchez et al., 2006;

Bartel, 2004). We now expand on the specific example of *PTEN* regulation by miRNAs.

1.2.3 Post-transcriptional Regulation of *PTEN*

In 2008, *PTEN* was shown to be susceptible to post-transcriptional regulation by miRNAs. The initial focus was placed on miR-21, an miRNA that, when over-expressed, can promote the growth of hepatocellular cancer (HCC) in humans via downregulation of *PTEN*. MiRNAs bind to target RNAs via complementary nucleotide sequences known as microRNA response elements (MREs), usually located within the 3′-untranslated region (3′UTR) of the mRNA. Each target transcript can have multiple MREs and the RISC−miRNA complex is guided towards these regions by the miRNA itself. Depending on the level of complementarity between the miRNA and its binding site, the miRNA−mRNA complex is thought to be transported to one of many distinct points in the cytoplasm, called P-bodies, where the mRNA is either degraded or held in a translationally repressed state, until it receives a signal to release the mRNA and resume translation (Macfarlane and Murphy, 2010). Specifically, it is suggested that binding to just the 'seed' region of the miRNA (nucleotides 2-8, counted from the 5′ end) will result in translational repression, whereas additional binding to nucleotides 9-12 is required for mRNA cleavage to occur (Ebert et al., 2007). Primarily, mRNAs are thought to be degraded by miRNAs (Guo et al., 2010), although some reports suggest that translational inhibition is the most prevalent method used by miRNAs (Petersen et al., 2006).

Most mRNAs contain multiple MREs that can be targeted by a variety of miRNA species, and some have been shown to also contain repeated MREs for certain miRNAs. In particular, miR-19b and miR-26a were recently shown to each have two target MREs within the 3′UTR of *PTEN* mRNA (Tian et al., 2013). It is not known whether multiplication of sites serves to increase the number of available targets, or whether these sites act cooperatively, such that multiple miRNAs must bind in order to repress or degrade the mRNA. There is evidence to support both hypotheses, and it is possible that both may operate in particular cases and under a range of conditions. Ragan et al. suggest that miRNAs can only cooperate if their target sites are less than 40 nucleotides apart (Ragan et al., 2011), whilst Hon et al. propose that miRNA cooperativity depends on the length of 3′UTR, the number of binding sites, and the distance between neighbouring sites (Hon and Zhang, 2007).

According to miRTarBase, an online database of published miRNA-target interactions, as many as

27 miRNA species were identified as targeting *PTEN*, with strong experimental evidence obtained from reporter assays, Western blots and quantitative PCR (qPCR) [3]. Studies of glioma (Huse et al., 2009), lung cancer (Zhang et al., 2010) and prostate cancer (Poliseno et al., 2010a) revealed that *PTEN* can be down-regulated by over-expression of a variety of miRNAs including miRNA 26a (miR-26a), miR-21, miR-22 and miR-25 (Leslie and Foti, 2011). Of these, miR-21 is one of the most frequently and vigorously up-regulated miRNAs in human glioma (a form of brain and spinal cancer) (Krichevsky et al., 2003).

It is not yet known whether miRNAs are degraded along with transcripts or are 'recycled' so that they may continue to interact with their targets. Early studies showed that miRNAs are recycled when perfect miRNA−mRNA seed sequence complementarity is achieved, i.e. when transcript degradation occurs (Hutvagner, 2002; Ameres et al., 2007). A more recent study by Baccarini et al. suggested that miRNAs are also recycled in the case of imperfect, or 'bulged', binding sites, and that the rate of decay of miRNAs only increased when all target mRNAs were bound by miRNAs, and excess miRNAs remained in the cytoplasm (Baccarini et al., 2011).

Thus far, it has been shown that *PTEN* is regulated at the mRNA level by a variety of miRNAs, leading to dysregulation of PTEN protein production, and increasing the likelihood of tumourigenesis across a range of different human tissues (Hollander et al., 2011). As an additional level of complexity, PTEN mRNAs form part of a vast network of RNAs that compete for a shared pool of miRNAs.

1.2.3.1 Sponge Modulation

The most recent breakthrough in the study of *PTEN* post-transcriptional regulation came in 2011 when four studies presented evidence of vast networks of RNAs, regulated by competition for a shared pool of miRNAs, with *PTEN* as the focus of three of the studies (Tay et al., 2011; Sumazin et al., 2011; Karreth et al., 2011; Cesana et al., 2011). "Sponges", or "competing endogenous RNAs" (ceRNAs), act by soaking up miRNAs, leaving fewer miRNAs to bind to their intended target mRNAs. This miRNA sequestration usually results in an up-regulation of target gene translation, as miRNA-mediated translational repression and transcript degradation are inhibited (see Figure 1.8) (Salmena et al., 2011). In this capacity, ceRNAs may act as tumour suppressors, preventing *PTEN* down-regulation by sequestering miRNAs away from PTEN MREs. Prior to this, there had

[3] www.mirtarbase.mbc.nctu.edu.tw - Last release: 1st November 2013. Number of microRNA species: 1,232.

miRNA	PTEN	NCOA7
miR-17-5p	1	1
miR-19a	1	2
miR-19b	3	1
miR-20a	1	1
miR-20b	1	1
miR-26a	2	-
miR-26b	2	-
miR-93	1	1
miR-106a	1	1
miR-106b	1	-

Table 1.1: MiRNA binding sites on PTEN mRNA and one putative ceRNA, NCOA7, for comparison.

been one confirmed PTEN ceRNA, discovered by Poliseno et al., the *PTEN* pseudogene, *PTENP1*, which is a non-protein-coding genetic relative of *PTEN* (Poliseno et al., 2010b). Due to its role in regulating normal PTEN levels, it is not surprising that *PTENP1* is commonly lost in human cancers (Poliseno et al., 2010b); the fewer PTENP1 mRNAs that are present in a cell, the greater the down-regulation of *PTEN* and the higher the chance of tumourigenesis. A schematic illustrating this regulation is presented in Figure 1.6 and Figure 1.8.

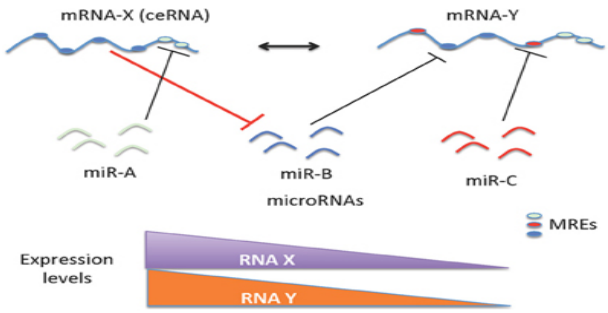


Figure 1.6: Post-transcriptional regulation of multiple competing RNAs by microRNAs (Kartha and Subramanian, 2014).

In (Sumazin et al., 2011), Sumazin et al. used gene expression analysis to identify a network of 13 ceRNAs that are directly involved in the downregulation of *PTEN* via miRNAs, and increase rates of glioma formation. They also detailed a variety of non-sponge mechanisms, which we describe in the next section. They referred to the combined effect of sponge and non-sponge modulation as a "miR-Program mediated Regulatory (mPR) network", and they suggested that *PTEN* has more than 500 such mPR regulators in Glioblastoma, which form part of a larger set of approximately

248,000 miRNAs, 7000 sponges and 150 non-sponge regulators uncovered from a full Glioblastoma gene expression analysis (Sumazin et al., 2011). In this thesis, we develop computational and mathematical models to study a highly idealised system in which PTEN mRNAs crosstalk with one ceRNA species via one miRNA species. Our justification for this simplification is given in Section 3.4.

Karreth et al. demonstrated downregulation of *PTEN* by ceRNAs in melanoma formation in mice. They identified *ZEB2* mRNA as one particular *PTEN* ceRNA (Karreth et al., 2011). This was especially novel as, prior to this, only gene-pseudogene pairs, such as *PTEN* and *PTENP1*, had been considered as potential ceRNAs (Poliseno et al., 2010a). Their results underscore the new interest in pseudogenes and non-coding RNAs as potential cancer therapies, both of which had previously been considered to be "junk" (Karreth et al., 2011).

Tay et al. performed a combined computational and experimental study of *PTEN* crosstalking with other RNAs, declaring a number of miRNAs that bind to PTEN mRNA and the frequency of MREs per transcript (Tay et al., 2011). They presented a summary of 14 PTEN MREs for 10 miRs (one MRE for miR-17-5p, 19a, 20a, 20b, 93, 106a, 106b; two MREs for miR-26a, 26b; three MREs for miR-19b) and a list of seven putative ceRNAs (NCOA7, BCL11B, SERINC1, ZNF460, NUDT13, DRWD2, VAPA) containing a variety of MREs for the same set of miRNAs (with between 1 and 4 MREs per miRNA).

Working independently, Cesana et al. identified a long non-coding RNA (lncRNA), linc-MD1, that behaves as a ceRNA and is involved in the regulation of human and mouse myoblast differentiation (Cesana et al., 2011). They identified binding sites for miR-135 and miR-133 in the linc-MD1 sequence, supporting their claim that linc-MD1 can crosstalk with these miRNAs and their target mRNAs, Myocyte-specific enhancer factor 2C (MEF2C) mRNA and Mastermind-like-1 (MAML1) mRNA, respectively. This crosstalk was shown to induce activation of muscle genes and subsequent differentiation of myoblasts via MEF2C and MAML1 proteins, respectively.

Whilst the above studies focus on naturally occurring ceRNAs, previous studies had proposed that ceRNAs could be used therapeutically to control miRNA expression (Arvey et al., 2010; Ebert et al., 2007; Poliseno et al., 2010a; Seitz, 2009). For example, Ebert et al. (Ebert et al., 2007) describe the construction of ceRNAs (referred to as "miRNA sponges"). They introduce ceRNAs as a method for studying miRNA function *in vivo*. The benefit, they explain, of using ceRNAs rather

than creating genetic knockouts is that some miRNAs are expressed at multiple geneomic loci, which makes the process of gene knockout more difficult. Additionally, utilising ceRNAs for genetic regulation negates the need for genetically engineered mutants, which is more relevant for human therapeutics. Ebert goes into more detail in her doctoral thesis (Ebert, 2010), where she describes how they engineered ceRNAs containing two MREs with perfect complementarity to the miRNA seed region (corresponding to miRNA-mediated transcript degradation), and another type of ceRNA containing seven 'bulged' MREs (corresponding to miRNA-mediated translational repression). The effect of targets and ceRNAs on the level of miRNAs is described as 'molecular titration'. They test the ability of titration to generate thresholds/switch points and demonstrate that the use of perfectly complementary MREs does not produce sensitive transition thresholds, whereas the use of bulged sites does (see Figure 1.7).

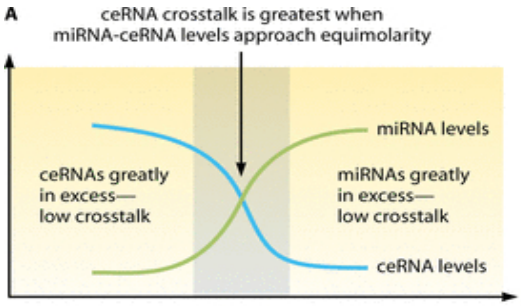


Figure 1.7: The type of switch-like behaviour observed as a result of ceRNAs (de Giorgio et al., 2013).

Ebert et al. added ceRNAs to HEK293T (kidney cancer) cells, expressing an abundance of miR-20, at a ceRNA:target ratio of 8:1. In their experiments, they showed that ceRNAs containing 'bulged' MREs were more effective at repressing miRNA effects than ceRNAs with perfectly complementary MREs. However, they were unable to discern if this difference was due to the greater number of sites or the additional stability of these ceRNAs. In Chapter 3, we use our models to try to distinguish between the effects of MRE complementarity (that is, ceRNA degradation versus repression) and MRE abundance. In another experiment, Ebert et al. transfected cells with 1nM and 5nM of miRNAs and showed a negative 2-2.5-fold change in target RNAs containing a single MRE, which they state is similar to the level of repression observed between endogenous miRNAs and their natural targets, which are also thought to contain a single MRE. In assessing the effect of ceRNAs on miRNAs, they estimated the mRNA copy number to be $\sim 1000 - 2000$ per cell, each containing seven binding sites, which, if fully utilised, may repress $\sim 10^4$ miRNAs per cell, a number consitent

with inhibition of almost all miRNAs in any cell type. They also state how difficult it is to isolate one miRNA without affecting the behaviour of the other miRNAs, something that can be done computationally (Ebert, 2010). In order to build on the work of Ebert et al., we test whether the number of MREs or their complementarity to the miRNAs has the greatest effect on miRNA deregulation in our model.

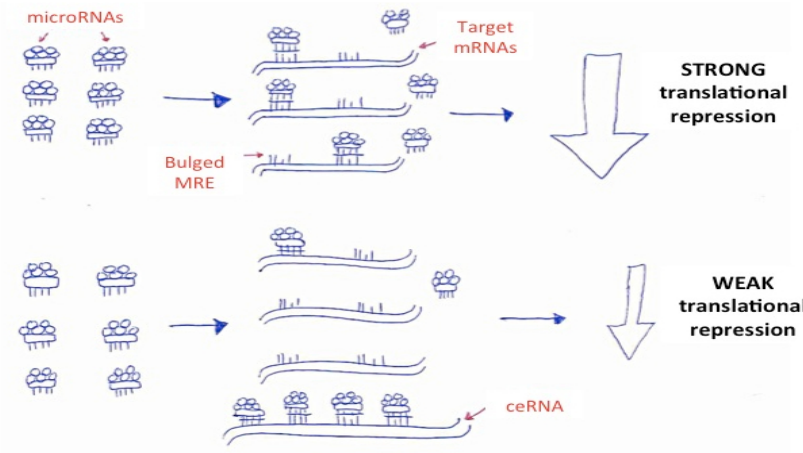


Figure 1.8: A schematic of the miRNA sponging by ceRNAs. In the top half of the figure, we see that if there are no ceRNAs, then there is a greater chance that an miRNA will bind to PTEN mRNAs, which will lead to down-regulation of the target mRNA (e.g. *PTEN*). However, if we introduce a ceRNA with multiple MREs, shown in the bottom half of the figure, then the pool of miRNA will have more sites to which it can bind and there will be less down-regulation of the target mRNA (although some target mRNA may still be repressed or degraded).

It is still not clear whether ceRNAs are degraded when they are bound by miRNAs. CeRNAs may simply sequester miRNAs to prevent them from binding with their target RNAs, with variable binding and unbinding rates of miRNAs possible (de Giorgio et al., 2013). Alternatively, ceRNAs may degrade the miRNAs upon binding, or vice versa. It is likely that ceRNAs behave in a similar way to the target RNAs, in that, ceRNAs with perfectly complementary MREs will be degraded by a binding miRNA, whereas imperfectly matched MREs will stabilise the ceRNA, protecting it from miRNA-mediated degradation, thereby improving its function as an miRNA sponge (Tay et al., 2014). If this is correct, then we can conclude that it is important to design bulged MREs when engineering ceRNAs, as postulated by Ebert et al. (Ebert, 2010).

Another factor that remains unclear is whether miRNAs bind reversibly to MREs. This consideration is of particular importance when co-operativity of ceRNAs and miRNAs is considered. For example, if miRNAs unbind from ceRNAs then they may re-enter the miRNA pool and bind to their target

mRNA. Similarly, if miRNAs bind irreversibly, then the ceRNAs may become saturated and unable to act as miRNA sponges. In the case of miRNA-cooperativity, if multiple miRNAs must be bound for target mRNA repression or degradation, then unbinding of miRNAs may decrease the likelihood of both miRNAs binding together, whereas irreversible binding would increase this likelihood. These factors are particularly challenging to analyse biologically as they may affect both target RNAs and ceRNAs, making it difficult to discern the source of any effects. In Chapter 3, we will model multiple MREs, miRNA cooperativity, and compare reversible and irreversible binding of miRNAs, in order to investigate the potential impact of each scenario on the ability of miRNAs to repress or degrade their target (*PTEN*) mRNAs.

1.2.3.2 Non-Sponge Modulation

Many other mechanisms can affect miRNA−mRNA interactions without directly sponging up the miRNA pool. These mechanisms are collectively known as "non-sponge modulation" (Sumazin et al., 2011). One such mechanism involves depleting the miRNA pool by miRNA degradation (Chatterjee et al., 2011); Chatterjee et al. showed that a mutation in the "seed" region of an miRNA (nucleotides 2-9) causes impaired target binding, rendering the miRNA susceptible to degradation. Although the reason for this has yet to be established, a decrease in the ability of miRNAs to bind to their target transcripts, in conjunction with an increase in miRNA degradation, may up-regulate *PTEN* translation. Another type of non-sponge modulation occurs when miRNA-like molecules prevent miRNAs from binding to their targets by binding to MREs. These miRNA-like molecules are often referred to as "decoys" or "mimics", however ceRNAs are also sometimes referred to as decoys and mimics, so we henceforth refer to these miRNA-like molecules as "copycats" to avoid confusion (Eiring et al., 2010). Copycats can be naturally occurring miRNAs with seed sequences that match a particular, functional miRNA, or they can be designed experimentally. In both cases, the copycats usually do not degrade the RNA target, but simply block the target MREs to prevent functional miRNAs from binding (see Figure 1.9). The impact of copycats is not obvious and will depend on the proportions of miRNAs and ceRNAs. In the case of *PTEN*, if copycats block the binding sites of PTEN mRNAs, then *PTEN* will be up-regulated. However, if the copycats block the binding sites of the ceRNAs, then *PTEN* will be down-regulated, as there will be fewer ceRNAs to soak up the miRNAs and also fewer copycats to block the PTEN mRNA binding sites.

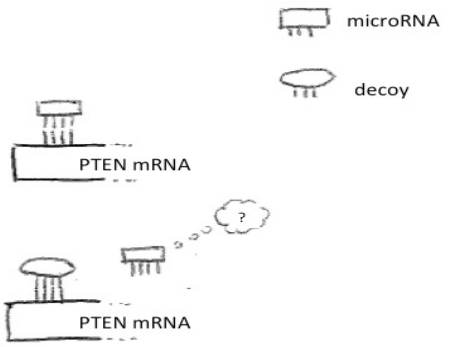


Figure 1.9: A schematic of copycat functionality.

MiRNA copycats have been proposed as a potential therapy for interfering with miRNA activity on *PTEN*. One study in 2011 detailed the factors to consider when designing miRNA copycats for use in therapeutic gene silencing (Wang et al., 2009). Specifically, the investigation reports the use of copycats with imperfectly complementary seed sequences that are unable to degrade target mRNAs, and the ability to design a copycat which is gene specific, in order to reduce off-target effects. This would be particularly useful as it would reduce the likelihood of the copycat binding to a ceRNA. By contrast, it is unclear whether such off-targets events arise with naturally occurring copycats.

In Chapter 3, we present the computational and mathematical models that we use to study miRNA-regulation of *PTEN* with copycats and ceRNAs (see Figure 1.10). Next, we introduce the second biological process that we model, that of RNA interference in the nucleus.

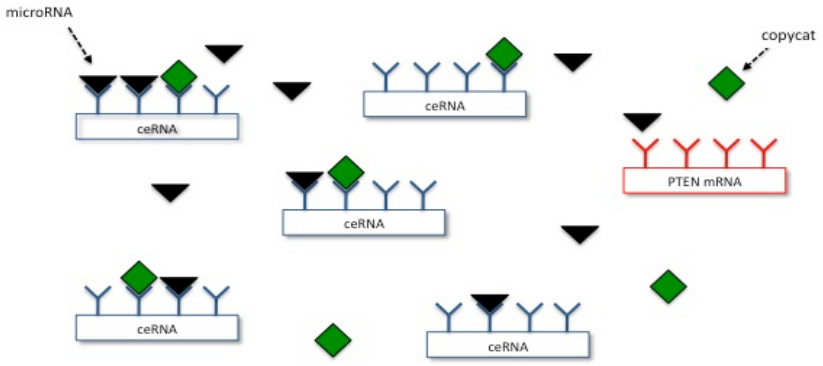
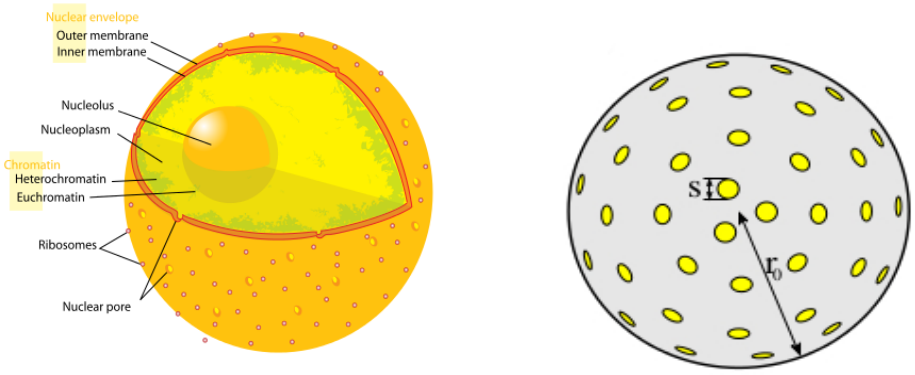


Figure 1.10: A schematic of the miRNA-PTEN post-transcriptional regulatory process; showing sponge modulation and non-sponge modulation, in the form of ceRNAs and copycats (labelled as 'decoys' here), respectively.

1.3 RNA Interference in the Nucleus

In eukaryotes, mRNA transcription occurs within the nucleus. Once transcribed, mRNAs form complexes with proteins, and are dispersed throughout the nucleus as they move by Brownian motion (Vargas et al., 2005). They subsequently exit the nucleus via small pores on its surface (pore diameter $\approx$ 120nm), to be translated into protein in the cytoplasm (see Figure 1.11) (Grünwald and Singer, 2010). RNA interference (RNAi) is a biological process in which this gene expression is regulated via interactions between small non-coding RNA molecules, known as short interfering RNA (siRNA), and target mRNA molecules (Alberts et al., 2002).



(a) Nuclear Cross-Section (b) Nuclear Membrane

Figure 1.11: (a) Schematic of the nucleus, with nucleolus and chromatin crowding the nucleoplasm, and nuclear pores scattered across the membrane (`www.wikipedia.org/wiki/Cell_nucleus`). (b) Schematic of nuclear pores on the membrane (Erban et al., 2007).

In general terms, RNA interference (RNAi) in the nucleus can be viewed as an example of the post-transcriptional regulatory process described in Section 1.2. RNA molecules in the nucleus are targeted for degradation by siRNAs, and may avoid degradation by escaping into the cytoplasm. As with post-transcriptional regulation in the cytoplasm, deregulation of RNAi can lead to disease and tumourigenesis via repression or over-expression of certain target genes. However, the regulatory mechanisms in the nucleus are thought to differ from the so-called "classical" post-transcriptional regulation in the cytoplasm. We describe below the hypothesised mechansims of RNAi in the nucleus and discuss their potential impact on disease.

1.3.1 RNA Interference Mechanisms

In mammalian cells, post-transcriptional regulation is typically localised to the cytoplasm, with nuclear post-transcriptional mechanisms, usually referred to as RNA interference (RNAi), reported only in small animals and plants, such as *Caenorhabditis elegans*, *Drosophila melanogaster* and *Schizosaccharomyces pombe* (Castel and Martienssen, 2013). However, recent evidence suggests that RNAi mechanisms may also occur in the nucleus of mammalian cells (Meister et al., 2004; Hwang et al., 2007; Politz et al., 2006; Liao et al., 2010; Jeffries et al., 2011; Gagnon and Corey, 2012). We note here that other types of epigenetic and genetic nuclear RNAi occur in mammalian cells, including heterochromatin formation, DNA methylation, and small RNA binding directly to DNA (Matzke and Birchler, 2005). However, as we are studying the specific regulatory process of post-transcriptional regulation by small RNAs, we do not consider these in our description of nuclear RNAi.

The main difference between post-transcriptional regulation and RNAi appears to be that post-transcriptional regulation usually describes repression or degradation of mRNAs by any mechanism. Conversely, RNAi usually describes post-transcriptional gene silencing (mRNA repression rather than degradation) by siRNAs. There are alternative descriptions of RNAi, however, for the purpose of this thesis, we continue to refer to nuclear post-transcriptional regulation by siRNAs as RNAi, with clarifications used when necessary. This principal difference in binding of siRNAs compared with miRNAs is thought to result from the difference in the ends of the siRNA and miRNA nucleotide sequences. In particular, the ends of miRNA sequences show little variation, which is thought to contribute to their superior ability to interact imperfectly with target mRNAs resulting in target repression (Carthew and Sontheimer, 2009). On the other hand, siRNAs have a 2 nucleotide overhang in their 3′UTR which is considered to facilitate degradation of their target RNAs, and limit their ability to bind imperfectly to targets (Langlois, 2005).

1.3.2 Evidence for RNA Interference in the Nucleus

The first notable report of RNAi in the nucleus was published in 2002 by Paul et al.; which they suggested that certain siRNAs can suppress pre-mRNAs before they exit the nucleus for translation in the cytoplasm (Paul et al., 2002). More recently, Robb et al. provided evidence in support

of nuclear RNAi by demonstrating the presence of siRNA-induced, RNAi-mediated degradation of target mRNAs in the nucleus of HeLa cells (Robb et al., 2005). This process resulted in down-regulation of target genes, as the number of target transcripts that reach the translation machinery in the cytoplasm was reduced. Robb et al. also stated that, in general, nuclear RNAi is important because some RNAs are exclusively located in the nucleus (Robb et al., 2005).

As with cytoplasmic post-transcriptional regulation, it is thought that the human Argonaute protein, AGO2, facilitates siRNA-mediated target RNA repression, although it is unclear how AGO2 is transported into the nucleus for this purpose (Gagnon et al., 2014). Berezhna et al. use confocal microscopy to image two types of nuclear siRNA and their targets, 7SK siRNA/RNA and U6 siRNA/RNA, in human hepatoma cells (Berezhna et al., 2006). They proposed several potential methods by which RISCs may translocate to the nucleus, including assembly in the nucleus, and active or passive siRNA-mediated transport from the cytoplasm to the nucleus (Berezhna et al., 2006). Further to this, Janowski et al. demonstrated that some AGO-mediated transcriptional repression occurs in the nucleus of mammalian cells. This was one of the first studies to show that a post-transcriptional regulatory pathway similar to that observed in the cytoplasm could be present in the nucleus (Janowski et al., 2006).

In 2012, Tang et al. showed that miR-709, a mouse miRNA found predominantly in the nucleus, can post-transcriptionally regulate the primary-miRNAs, (pri-)miR-15a and (pri-)miR-16-1, in the nucleus by binding directly to an MRE on the target pri-miRNAs (Tang et al., 2011). This interaction enchanced apoptosis by restricting nuclear export, and subsequent maturation, of miR-15a/16-1, which normally acts to post-transcriptionally repress *Bcl-2*, a gene involved in inhibiting apoptosis (see Figure 1.12). This study is of particular importance as it can be considered as being an "upstream" mechanism of the post-transcriptional regulation in the cytoplasm as miRNAs are themselves being post-transcriptionally regulated.

Unfortunately, the study by Tang et al. failed to clarify the mechanism by which miR-709 translocated to the nucleus from the cytoplasm. Hwang et al. suggested that the addition of a specific hexonucleotide sequence (AGUGUU) to the miRNAs could target these molecules for nuclear localisation. They identified two miRNAs, miR-29b and miR-29a, in the nucleus and cytoplasm, respectively, that were identical apart from the addition of this nucleotide sequence to the nuclear miRNA. They added further weight to this hypothesis by showing that adding this motif to another

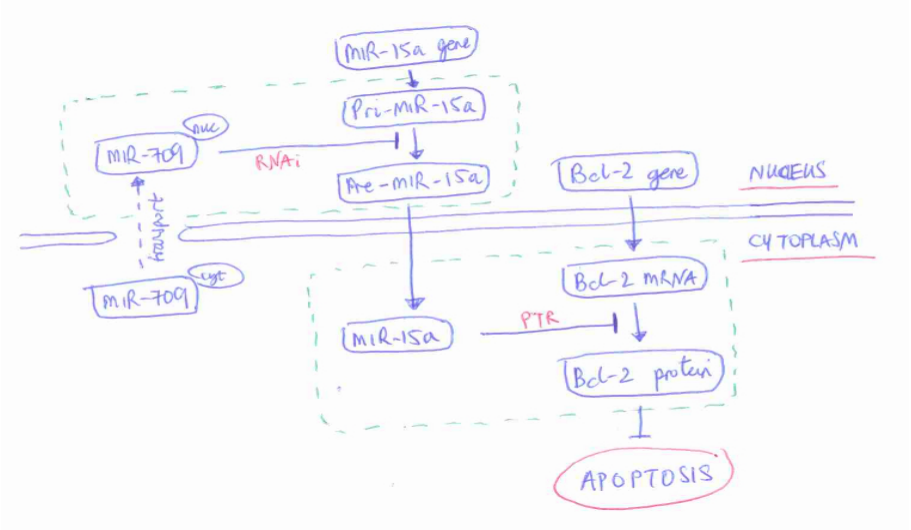


Figure 1.12: Impact of nuclear miR-709 on cellular apoptosis via inhibition of miR-15a maturation (note: the same mechanism holds for miR-16-1).

siRNA biased its subcellular localisation to the nucleus (Hwang et al., 2007).

Hansen et al. investigated the impact of miRNAs on circular anti-sense RNAs in the nucleus, focusing on miR-671-AGO2-mediated degradation of *CDR1* (Cerebellar Degeneration-Related protein 1), a gene whose de-regulation may lead to degeneration of the cerebellum and other diseases known collectively as "paraneoplastic neurological disorders" (Hansen et al., 2011, 2013). They further demonstrated that miR-671 is one of the most nuclear-enriched miRNAs, a result which is in line with a prior study of subcellular miRNA localisation and abundance (Liao et al., 2010). More recent studies have shown that most miRNAs can localise to the cytoplasm and the nucleus, but few have been shown to participate in miRNA-mediated transcript regulation in the nucleus (Jeffries et al., 2011; Gagnon et al., 2014; Roberts, 2014).

More information on the potential uses of RNAi was presented in a review by Liang et al (Liang et al., 2013). They stressed the emerging importance of nuclear RNAi activity, which makes our study all the more pertinent. Despite the growing evidence to support it, the topic of nuclear RNAi remains fraught, with many researchers questioning the validity of the results presented above and speculating over the quality of the experiments. Gagnon et al. attempted to quell these arguments by demonstrating the presence of RNAi factors in the nucleus of human cells, analysing the localisation and interactions between factors involved in RNAi in HeLa cells (Gagnon et al., 2014). They showed that factors involved in post-transcriptional regulation, such as Dicer and AGO2, are present in the

nucleus. They suggest that the mechanisms of small-RNA mediated RNAi differ between cytoplasm and nucleus, as many RISC proteins were not found in the nucleus. However, they did observe siRNA-mediated repression of nuclear RNA targets, including cleavage by AGO2 at complementary binding sites (Gagnon et al., 2014).

RNAi mechanisms in the nucleus have not been widely studied, and, therefore, represent a novel avenue for therapeutic targets. SiRNAs can be exogenously inserted into cells and, as such, RNAi has been used to treat of a variety of human diseases (Li and Rana, 2012). For example, the siRNA drug TKM-PLK1 is being trialled as a therapy for patients with advanced solid tumours, for which current therapies have proven to be ineffective (Burnett et al., 2011). It is also possible that the design of a competing RNA that sponges up miR-709 could be used therapeutically to dampen the apoptotic effect of miR-709.

In Chapter 2, we present a computational model of RNAi. We continue here by presenting the third, and final, biological process studied in this thesis, the canonical Wnt signalling pathway.

1.4 Wnt Signalling

Cell signalling is important for normal development and cell growth. Typically, cells communicate via excretion of proteins from a cell. These proteins bind to complementary receptors on the surface of a cell, leading to signal transduction via internal cellular modifications. The type of signalling molecules involved depends on the type of signalling pathway being implemented by the cell. For example, cells can signal to themselves (autocrine), to neighbouring cells (paracrine and juxtacrine), or to distant cells (endocrine). There are also intracrine signalling pathways, in which the signalling protein is contained entirely within a cell, without the need for binding to extracellular receptors on the cell membrane. The response of cells to various intra- and extra-cellular stimuli depends on the temporal and spatial dynamics of each signalling pathway and their components.

Extracellular signalling ligands primarily bind to transmembrane G protein-coupled receptors (GPCRs). These receptors span the cell membrane seven times, resulting in the presence of both an extracellular and an intracellular terminus. Upon binding of ligands to the extracellular N-terminus of their target GPCRs, the signal is transduced across the cell membrane via a conformational change in the receptor. This results in activation of the intracellular (cytoplasmic) C-terminus of the GPCR

and subsequent interaction with cytoplasmic proteins. The resulting intracellular changes depend on the signalling pathway.

One of the most thoroughly studied signalling proteins is Wnt (Wingless-related integration site), a highly conserved protein involved in maintainance of proper development, growth, and proliferation of cells. Wnt signalling is studied primarily using human, mouse, fly and fish models. Wnt was first discovered, almost simultaneously, by three research groups in 1982; Nusse and Varmus (Nusse and Varmus, 1982) discovered the gene, which they called *Int1*, in the mammary gland tumours of mice, whereas Nüsslein-Volhard and Wieschaus (Nüsslein-Volhard and Wieschaus, 1980), and Sharma and Chopra (Sharma and Chopra, 1976), both discovered the same gene, which they called *Wingless*, in *Drosophila*. Upon realisating that these genes were identical, the two names were combined to give the currently used portmanteau of *Wnt*.

When the Wnt signalling pathway is activated (Wnt "on"), cells are typically proliferative stem-like cells. When Wnt signalling is inactive (Wnt "off"), cells become fully differentiated. For example, in the human intestinal epithelial crypt, Wnt is present in a decreasing spatial morphogen gradient from the base to the tip; Wnt is "on" at the base of the crypt, where cells are continuously proliferating and forcing new cells up the crypt, against the Wnt gradient. When the cells reach the tip of the crypt, where Wnt is "off", cells enter a differentiated state and either undergo apoptosis or are shed into the blood stream for removal (see Figure 1.13). When Wnt genes are mutated, or the pathway is perturbed, the resulting change in dynamics has been shown to have clear implications for tumourigenesis, as well as developmental abnormalities and other diseases, depending on which target genes are over-/under-expressed (Alberts et al., 2002).

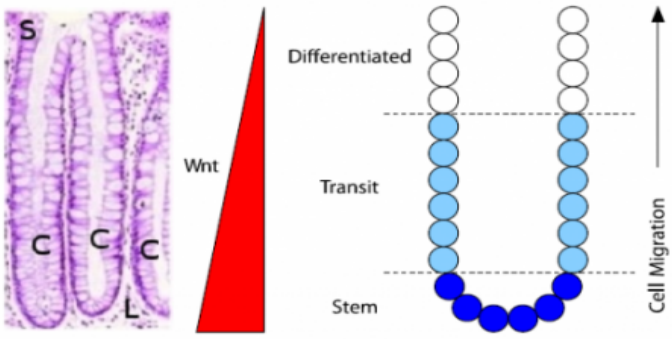


Figure 1.13: Figure showing Wnt gradient in the human intestinal crypt. Taken from Fletcher et al. (2012) (Fletcher et al., 2012).

There are several types of Wnt signalling pathway: The canonical Wnt pathway, the non-canonical planar cell polarity pathway, and the non-canonical Wnt/calcium pathway. As the name suggests, the canonical Wnt pathway (also known as the Wnt/β-catenin pathway) is the most extensively studied, and it is the process on which we focus in this thesis.

In addition to its involvement in Wnt signalling, β-catenin also plays a pivotal role in the cell adhesion. Specifically, molecules of β-catenin may act as a structural adaptor protein, forming complexes with α-catenin and E-cadherin molecules at the cell membrane. It has been demonstrated that down-regulation of this adhesion complex may contribute to tumourigenesis, as cells lose their adhesiveness and become capable of over-proliferation and invasion (Basan et al., 2010).

In this chapter, we introduce the biological details of the signalling pathway, and β-catenin's dual role within the cell, with specific emphasis on the potential for tumourigenesis when these processes are disrupted. An introduction to our work and the hypotheses we wish to test is given in Chapter 4. By building upon existing models, we aim to investigate the effects of spatial localisation and stochasticity on the dynamics of this pathway. First, however, we explain briefly how the pathway acts.

1.4.1 The Canonical Wnt/βCatenin Signalling Pathway

In mammals, there are currently 19 known Wnt proteins that play a pivotal role in normal embryonic development, and tissue maintainance and repair. Each protein is composed of approximately 350-400 amino acids, and is characterised by 23-24 conserved cysteine residues.

In the Wnt/β-catenin signalling pathway, cells communicate via paracrine signalling. Wnt ligands bind to extracellular membrane receptors on the surface of neighbouring cells, causing an accumulation of the protein β-catenin, which plays an important role in tumour formation and metastasis (Cai et al., 2013) via transcription of a variety of target genes, such as those responsible for cell proliferation and apoptosis.

In healthy adult (differentiated) cells, in the absence of a Wnt stimulus (Wnt "off"), β-catenin is phosphorylated by a complex of proteins known as the destruction complex (DC), which labels β-catenin for degradation. Specifically, β-catenin is initially phosphorylated by the kinase CK1 and subsequently phosphorylated further by another kinase, GSK3, labelling the phosphorylated β-

catenin for degradation by the ubiquitin ligase, β-TrCP, in the proteasome, a large protein complex responsible for protein degradation in eukaryotic cells via proteolysis.

In healthy stem or embryonic cells, when cells have not yet formed adult specialised cells, Wnt signalling is active (Wnt "on"), and the pathway proceeds via paracrine signalling, consisting of a network of proteins that transfer information between closely neighbouring cells (Alberts et al., 2008).

1.4.1.1 In The Cytoplasm

The "destruction complex" (DC) is primarily composed of glycogen synthase kinase 3 (GSK3), and two scaffold proteins, Adenomatous polyposis coli (APC) and Axin. It also contains casein kinase 1 (CK1), the E3-ubiquitin ligase β-TrCP, and protein phosphatase 2A (PP2A), however, we focus on the 3 key components in this study as they are the most crucial for DC functionality. They are also often mutated, degraded or sequestered by other molecules when the Wnt signalling pathway is active (in normal and mutated cells).

APC is the largest of the core DC components ($\sim$ 310kDa), conserved across a variety of organisms, and present in the human intestine where it functions as a vital tumour suppressor; Mono-allelic loss of *APC* occurs in $\sim$ 80% of all human colorectal cancers (Rowan et al., 2000; Bienz and Clevers, 2000), whereas bi-allelic loss is typically fatal. Aside from its cytoplasmic role in β-catenin destabilisation, APC can also be found in the nucleus, as well as associating with the plasma membrane and microtubules. APC has also been shown to bind directly to β-catenin and may be capable of transporting β-catenin between the nucleus and the cytoplasm (Bienz and Clevers, 2000). Another possible function of APC, in the Wnt/β-catenin pathway, is in the retention of β-catenin within its current intracellular compartment (nucleus or cytoplasm). It is thought that phosphorylated APC, resulting from the accumulation of β-catenin, is responsible for β-catenin retention and may also cause its dissociation from its nuclear binding partners (discussed in the next section).

Axin is the next largest core component of the DC, compared with APC, with a molecular weight of $\sim$ 110kDa. As well as possessing a binding site for β-catenin, Axin also contains binding sites for the other DC components, GSK3, CK1 and APC. However, when Axin and APC are not bound together, they compete for the same binding site on β-catenin, along with LEF/TCF and E-cadherin,

meaning that if β-catenin is already bound to APC, it cannot be bound to by the DC.

GSK3 exists in two isoforms in mammalian cells, GSK3α and GSK3β. GSK3 is the smallest of the core DC components, with each isoform having a molecular weight of $\sim$ 50kDa, and functions in the Wnt/β-catenin pathway via phosphorylation of β-catenin, labelling it for degradation.

1.4.1.2 At The Membrane

As mentioned previously, the Wnt signalling pathway functions via paracrine signalling. For short-range signalling, the cell in which *Wnt* is activated begins the process of Wnt secretion; Wnts are glycosylated and lipidated in the endoplasmic reticulum (ER) by the acyltransferase Porcupine. Wnts then enter the Golgi body from where they are escorted, in secretory vesicles, to the plasma membrane by a group of transmembrane proteins, called Wntless (Wls). Wnt then enters the extracellular space and Wls is recycled for further use within the cell via endocytosis.

When secreted Wnt reaches a neighbouring cell, the proteins bind to extracellular surface receptor proteins, known as Frizzled (Fzd), and its co-receptor proteins LRP5/6 (low-density lipoprotein receptor-related proteins). Upon activation by Wnt, the cytoplasmic domain of the Fzd-LRP complex is phosphorylated and further complexes with a cytoplasmic protein called Dishevelled (Dsh). This in turn leads to the breakdown of the destruction complex via recruitment of Axin to the membrane, causing β-catenin levels to rise in the cytoplasm (Mirams et al., 2009). β-catenin can then translocate to the nucleus where it can complex with transcription factors (discussed in Section 1.4.1.3), or to the membrane where it can form a complex with cadherins (discussed in Section 1.4.3).

1.4.1.3 In The Nucleus

Within a cell, intercompartmental molecular transport between the cytoplasm and the nucleus, occurs bi-directionally through any of $\sim$ 3000 nuclear pore complexes (NPCs) in the nuclear membrane. The NPC, one of the largest proteins in the eukaryotic cell ($\sim$ 125MDa), controls which molecules can and cannot cross between intracellular the cytoplasm and the nucleus. In order to translocate through an NPC, most macromolecules require a chaperone. For movement into the nucleus, macromolecules are chaperoned by a type of karyopherin, called importins, after tagging with a nuclear localisation signal (NLS). For movement out of the nucleus, macromolecules are chaperoned

by another type of karyopherin known as exportins, after tagging with a nuclear export sequence (NES).

With a pore radius of $\sim$ 5nm, small molecules ($<$ 50kDa) are able to move freely through NPCs. However, despite its larger size of $\sim$ 90kDa, β-catenin is also capable of moving through NPCs without a chaperone. This is likely due to β-catenin expressing its own NLS, similar to that of importins, allowing it to move freely into the nucleus. It has previously been hypothesised that β-catenin is chaperoned into the nucleus by APC and Axin, however, it has also recently been suggested that, instead, APC and Axin may simply retain β-catenin in the cytoplasm or nucleus (Henderson and Fagotto, 2002; Krieghoff, 2006; Schmitz et al., 2013).

The mechanisms behind nuclear export of β-catenin are less clear. In some experiments, β-catenin has been shown to exit the nucleus independent of any chaperone. Other research has suggested that β-catenin is chaperoned out of the nucleus by two of the components of the DC, APC and Axin. Indeed, APC, Axin and GSK3 have all been shown to move between the cytoplasm and the nucleus, with some studies suggesting that the DC may be assembled in the nucleus, and other results demonstrating the capabilities of GSK3 in regulation of β-catenin levels from within the nucleus as well as in the cytoplasm.

However it enters the nucleus, once inside β-catenin acts as a cofactor for the T-cell factor/Lymphoid enhancing factor (TCF/LEF) family of transcription factors (TFs) (Giles et al., 2003), which are responsible for the transcription of Wnt target genes, including c-Myc, Cyclin D1 and TCF-1 (Takebe et al., 2010). Several genes that are controlled by β-catenin have been shown to promote cell growth and proliferation (Takebe et al., 2010); over-expression of such genes by an excess of β-catenin may, therefore, result in tumourigenesis. Some (approximately 14 out of 123) of these target genes also encode for components of the Wnt/β-catenin pathway, such as Axin, Frz and TCF, creating the potential for feedback loops within the pathway.

In the Wnt "off" state, β-catenin's potential binding partners in the nucleus, transcription factors LEF/TCF, are kept in an inactive state by transcription factor inhibitors, such as Groucho. In our models, we use β-catenin/TF complex formation as the output measure of our models, since transcription of the target genes by this complex is the ultimate step in the canonical Wnt pathway, and a direct reflection of the effect of perturbations on the system.

Varying levels of β-catenin may also affect cell-cell adhesion as β-catenin can bind to E-cadherin molecules at cell adherens junctions.

1.4.2 Mutations in The Wnt Signalling Pathway

There are a variety of known mutations that can occur which have a negative impact on the functionality of the Wnt/β-catenin signalling pathway. Irregular Wnt function has been shown to cause Alzheimer's disease, schizophrenia, developmental abnormalities, and cancer. For example, a mutation in the *APC* gene, whose normal role is in the degradation of β-catenin in the cytoplasm, can lead to increased β-catenin levels, resulting in upregulation of Wnt target gene expression in cells where these genes should be silent. Such genes are usually involved in cell growth and proliferation and, when over-expressed, can lead to the kind of uncontrolled cell growth that is indicative of tumourigenesis.

1.4.2.1 Wnt Antagonists

The destruction complex and its constituent components are considered to be intracellular antagonists of the Wnt/β-catenin signalling pathway, as they target β-catenin for degradation, ultimately preventing Wnt target genes from being transcribed in the nucleus. Another known intracellular antagonist, called Chibby (Cby), acts by interfering with β-catenin/TF interaction in the cell nucleus (Kim et al., 2013). There are many other antagonsists to the Wnt/β-catenin pathway, the majority of which are secreted, therefore acting as extracellular antagonists. Extracellular antagonists include Wnt inhibitory factor proteins (WIFs) which bind to secreted Wnts, and secreted Frizzled related proteins (sFRPs) which can bind either to Wnts or to Fzd. Each of these antagonists prevents Wnt/Fzd/LRP complex formation. Similarly, Dickkopf (DKK) proteins can bind to LRP5/6 to prevent Fzd/LRP complex formation (MacDonald et al., 2009). Other examples include Shisa proteins, which act by retaining Fzd in the endoplasmic reticulum so that it is unable to reach the cell membrane; Norrin proteins, which bind directly to Fzd/LRP; and Wise/SOST family of proteins, which can interfere with Fzd/LRP complex formation similarly to DKK proteins.

Whilst we do not explicitly study extracellular or nuclear antagonisation of the Wnt/β-catenin pathway in this thesis, it is important to note the existence of such mechanisms, particularly because

of the links between Wnt/β-catenin deregulation and an increased likelihood of cancer formation.

1.4.3 The Role of βCatenin in Cell-Cell Adhesion

β-catenin was initially discovered via a study into cell adhesion. In addition to binding to certain transcription factors, as well as APC and Axin, β-catenin proteins also contain binding sites for certain cadherins (Kemler, 1993). β-catenin contributes to cell-cell adhesion by forming a complex with E-cadherin, a transmembrane adhesion complex, and an α-catenin protein, linking the entire adhesion complex to the cytoskeleton (Orsulic et al., 1999). Formation of this complex leads to reduced cell motility, therefore, down-regulation of this complex formation can lead to tumourigenesis, as cells become over-proliferative and invasive. Whilst β-catenin is participating in cell ahesion, in a complex with α-catenin and E-cadherin, it cannot participate in the Wnt signalling pathway.

More information about existing mathematical models of the canonical Wnt pathway is presented in Section 1.5.8. Next we discuss the mathematical and computational modelling techniques that can be, and have been, used to aid the study of biological processes.

1.5 Mathematical and Computational Modelling of Biological Systems

As with all biological systems, the processes described above are complex and difficult to study experimentally. It is often difficult to control the state of the cellular system, without initiating off-target effects. One can gain additional information from computational and mathematical modelling, where the state of the system can be tightly controlled by the user, the impact of independent perturbations on the system can be examined, and hypotheses regarding the importance of system components, such as molecular species and intracellular geometry, can be tested.

Intracellular processes, such as biochemical reactions and diffusion of molecules, are usually modelled by systems of differential equations or via stochastic simulations, depending on whether our molecules of interest are abundant or scarce, respectively. We argue here that, particularly when the number of molecules are low, models should account for spatial effects. Throughout this thesis, we compare our spatial-stochastic modelling approach to deterministic and (spatially-averaged) stochastic models, in

order to determine the impact, if any, of stochasticity and spatial detail on the biological processes under investigation.

1.5.1 Deterministic Modelling

Biochemical reactions within the cell are usually considered to behave deterministically. The Law of Mass Action, governing deterministic reactions, states that the future molecular concentrations in a chemcial reaction network can be determined using initial molecular concentrations (Espenson, 2002; Schnell and Turner, 2004). Many biological systems have been modelled successfully using deterministic models. This usually involves creating a system of ordinary differential equations (ODEs), with fixed rates of change of each molecular species over time (Ala et al., 2013; Mukherji et al., 2011). These models work best when the molecules are assumed to be present in large enough quantities to have predictable long-term dynamics, in order to obey the Law of Mass Action, and when the evolution of the system is assumed not to be dependent on the space in which the molecules interact. However, the Law of Mass Action is an increasingly poor assumption in small systems, where molecular concentrations are very low and chemical reaction events become progressively less likely (Gillespie, 1977; Burrage et al., 2011). Cellular processes of this nature should be considered to be inherently stochastic, as discrete molecules randomly diffuse through the crowded intracellular media, undergoing specific and non-specific interactions with other molecules, leading to large cell-cell variability even in genetically identical cells (Mettetal and van Oudenaarden, 2007).

1.5.2 Stochastic Modelling

Stochastic modelling is an important method for modelling biochemical reactions in which one or more of the reactive species is present in low quantities. Arguments in favour of modelling intracellular reactions as stochastic processes focus on the necessity for representation of molecular species as individual molecules rather than concentrations, particularly those with low molecular counts. This is utilised in order to identify the rate-limiting steps of the process, which are the most likely to occur stochastically and, therefore, define the overall rate of the process. In post-transcriptional regulation and RNAi, both miRNAs and their target mRNAs are often present in low copy numbers in human cells. Therefore, their interactions should be modelled stochastically.

In Wnt signalling, most molecular species are present at high concentrations, so we focus primarily on those molecules present in low numbers. In this latter case, a hybrid deterministic-stochastic model would reduce the computational cost of modelling the pathway, however, we leave this for future work. There are numerous methods available for modelling stochastic processes which we describe in this section. For further information about stochastic and spatial modelling of biological processes, see the review article by Burrage et al. (Burrage et al., 2011).

1.5.2.1 Stochastic Differential Equations

Stochastic differential equations (SDEs) are used to describe stochastic processes. They usually comprise a system of ODEs with random noise introduced, and are typically modelled by a set of Brownian or Weiner processes (Jha and Langmead, 2012). Such noise is important when there are random fluctuations in the number of molecules for species that are present in small numbers; in such cases, the system may no longer behave deterministically. SDEs are an intermediate approach between a system of ODEs and the Stochastic Simulation Algorithm (Section 1.5.2.3). SDEs are often written in the "Langevin form" and referred to as a Chemical Langevin Equation, which may be written as

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t, \tag{1.1}$$

where X_t is a stochastic process, μ is the drift coefficient, σ is the diffusion coefficient, and B is a Wiener process (or Brownian motion) (Gillespie, 2000). The noise that is introduced into the system depends on a random number generated from the Normal distribution with mean 0 and variance 1, i.e. $\sim N(0, 1)$, and also depends on $\sqrt{dt}$, such that, in the limit $dt \to 0$, the noise term disappears and the system is best described by a system of ODEs. SDEs can be solved numerically for simple systems, however, biological systems are typically too complex to obtain a simple solution. A simpler method, which is also useful for modelling the stochasticity in the system more explicitly, is to describe the system as a discrete-valued Markov process (i.e. a continuous-time Markov chain) which can be solved using the Stochastic Simulation Algorithm.

1.5.2.2 Markov Chain Model

A stochastic process is described as Markovian if the state of the system at the next time step depends only on the current state of the system, and is independent all previous states of the system. A continuous-time Markov chain (CTMC) refers to the progression through a countable number of states in which each new state (or "link") depends only on the previous state (the last link) in the chain and a transition probability matrix. In other words, a CTMC system can be in one of a countable number of states at time $t + \delta t$, dependent on the state of the system at time t and the transition rate matrix, Q. To determine the probability of being in state j at $t + \delta t$, we isolate all transitions that take the system from the current state, i, to state j in the time step, δt, and sum over the probabilities. The "transition rate matrix", Q, is an $N \times N$ array of all transition probabilities for every state $j \in \{1 \dots N\}$. The elements $p(i, j)$ represents the transition probability from state i to state j, and elements $p(i, i)$ are chosen such that each row of the matrix sums to zero.

As a simple example of a discrete-time Markov chain (DTMC), where the system is allowed remain in the same state, we consider weather prediction: we assume that the weather can be in one of two states, 'sunny' (X_1) or 'rainy' (X_2), and that the weather tomorrow depends only on the weather today (see Figure 1.14).

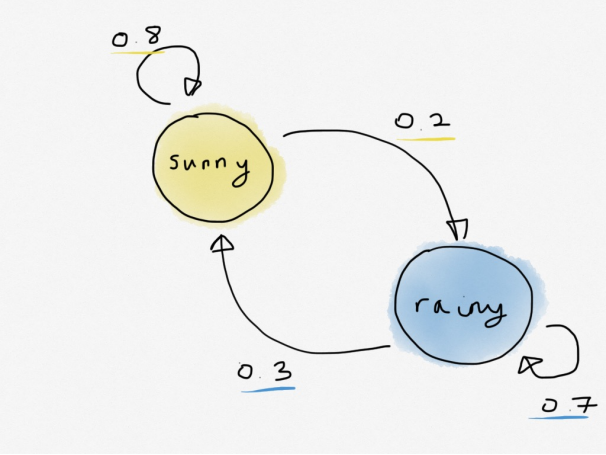


Figure 1.14: A Markov chain schematic of the weather prediction model.

We can formulate this process as a Markov chain, $X = \{x_1, x_2\}$, in order to determine (the state of) the weather tomorrow given (the state of) the weather today. For this example, the transition

matrix takes the form,

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix}, \tag{1.2}$$

where the rows represent the current state of the system (1 = sunny, 2 = rainy) and the columns represent the possible state of the system in the next time step, with $p(i, j)$ representing the probability of transitioning from state i to state j in the time interval $[t, t + \delta t]$. The state of the system, x, at any time, $t + \delta t$, can be calculated by $x(t + \delta t) = x(t)P$. For the weather example in Figure 1.14, the matrix takes the form

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}. \tag{1.3}$$

The rows of the transition matrix are independent of each other and must sum to one as the number of states is finite and the list of states is exhaustive, i.e. the system must be in one of the states at all times. This type of matrix is called "right stochastic" or "row stochastic". For a generic DTMC, $\mathbf{X_t} = \{X_1, X_2, \ldots X_S\}$, of S independent states, the transition matrix is given by

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,S} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,S} \\ \vdots & \vdots & \ddots & \vdots \\ p_{S,1} & p_{S,2} & \cdots & p_{S,S} \end{pmatrix}, \tag{1.4}$$

where $p_{i,j}$ is the probability of a transition $X_i \to X_j$, and

$$\sum_j p_{i,j} = 1, \qquad \forall i \in [1, \ldots, S].$$

The transition matrix over n time steps is given by P^n. Taking the limit of the transitions as $\delta t \to 0$ gives the so-called "Chemical Master equation" (CME), which represents the solution to the

probability density function, $p_s(t)$. For example, the time evolution of sunny days is given by

$$p_s(t+\delta t) = 0.8\delta t \times p_s(t) + 0.3\delta t \times p_r(t)$$

$$\implies \frac{p_s(t+\delta t) - p_s(t)}{\delta t} = -0.2p_s(t) + 0.3p_r(t)$$

$$\overset{\delta t \to 0}{\implies} \frac{dp_s}{dt} = -0.5p_s(t) + 0.3, \tag{1.5}$$

using $p_s(t) = 1 - p_r(t)$ in Equation (1.5). Similarly, since $p_r(t) = 1 - p_s(t)$, the time evolution of rainy days is given by

$$\frac{dp_r}{dt} = -0.5p_r(t) + 0.2. \tag{1.6}$$

Equations (1.5) and (1.6) form the CME, which can be expressed in matrix form as,

$$\frac{dX}{dt} = XP, \tag{1.7}$$

where $X = \{x_s, x_r\}$ is the row vector of states (sunny or rainy) and P is the transition matrix between states. For a set of N first-order differential equations, an analytical solution can be found given $N-1$ boundary conditions (the final state can be solved as 1−Prob(all other states)). So, for the weather scenario, since $p_s(t) = 1 - p_r(t)$, we can solve either of equations (1.5) or (1.6) using an initial condition. For example, knowing only that it is sunny today ($p_s = 1$ at $t = 0$),

$$p_s(t) = \frac{2}{5}e^{-0.5t} + \frac{3}{5}, \tag{1.8}$$

$$p_r(t) = -\frac{2}{5}e^{-0.5t} + \frac{2}{5}. \tag{1.9}$$

This means, for example, that the probability of it being a sunny day at $t = 10$ (days), given that it is sunny today, is

$$p_s(10) \approx 0.6, \tag{1.10}$$

independent of the state of the weather for $1 \leq t \leq 9$ days. We stress, however, that this is only a probability of being in a particular state.

For systems that are more complex, and analytical solution is often unattainable and a solution can be found using numerical methods. For example, consider the kinetic system described by the chemical reactions,

$$\begin{gathered} \emptyset \xrightarrow{k_1} \mathrm{A}, \qquad \emptyset \xrightarrow{k_2} \mathrm{B}, \\ \mathrm{A} + \mathrm{A} \xrightarrow{k_3} \emptyset, \qquad \mathrm{A} + \mathrm{B} \xrightarrow{k_4} \emptyset, \end{gathered} \tag{1.11}$$

where k_i $(i = 1, 2, 3, 4)$ are rate constants. The system has CME of the form

$$\begin{aligned} \frac{dp_{n,m}(t)}{dt} &= k_1 p_{n-1,m}(t) - k_1 p_{n,m}(t) + k_2 p_{n,m-1}(t) - k_2 p_{n,m}(t) \\ &+ k_3(n+2)(n+1)p_{n+2,m}(t) - k_3.n.(n-1)p_{n,m}(t) \\ &+ k_4(n+1)(m+1)p_{n+1,m+1}(t) - k_4.n.m.p_{n,m}(t), \end{aligned} \tag{1.12}$$

where $p_{n,m}(t)$ is the probability that there are n molecules of species A and m molecules of species B at time t. The matrix form of this system would be given by

$$p_{n,m} = \begin{pmatrix} p_{n,m-1} & p_{n-1,m} & p_{n,m} & p_{n+1,m+1} & p_{n+2,m} \end{pmatrix} \begin{pmatrix} k_2 \\ k_1 \\ -(k_1 + k_2 + k_3(n.(n-1)) + k_4.n.m) \\ k_4(n+1)(m+1) \\ k_3(n+2)(n+1) \end{pmatrix} \tag{1.13}$$

Unfortunately, such a system is too complicated to solve analytically. One of the most widely used numerical methods for solving such chemical reaction networks is the Stochastic Simulation Algorithm (SSA).

1.5.2.3 Stochastic Simulation Algorithm

The direct stochastic simulation algorithm (SSA) was developed by Daniel Gillespie (Gillespie, 1977) in 1977. In this paper, Gillespie presented an algorithm to calculate the probabilities of N uniformly distributed molecules reacting in a compartment of finite volume via M discrete chemical reactions. The algorithm proceeds as follows (derivations of the formulae used in steps (2) and (3) can be found in (Gillespie, 1977; Erban et al., 2007)):

1. Describe the system by a set of N reactions.

2. Assign propensity functions, $\alpha_i(t)$ $(i = 1, 2, \ldots, N)$, to each reaction such that $\alpha_i(t)$ is proportional to the number of molecules of each reactant involved in reaction i at time t and rate k_i for reaction i. The quantity $\alpha_i(t)\tau$ gives the probability of reaction i occurring in the time interval $[t, t + \tau]$.

3. Generate a random number, r_1, from the uniform distribution $U[0, 1]$ to determine the time step,

$$\tau = \frac{1}{\alpha(t)} \ln\left[\frac{1}{r_1}\right]. \tag{1.14}$$

4. Generate another uniformly distributed random number, r_2, to determine which reaction, j, will occur and time $t + \tau$. Specifically, for $\alpha(t) = \sum_{i=1}^{N} \alpha_i(t)$, the reaction j is chosen such that

$$\frac{1}{\alpha(t)} \sum_{i=1}^{j-1} \alpha_i(t) \;\leq\; r_2 \;<\; \frac{1}{\alpha(t)} \sum_{i=1}^{j} \alpha_i(t) \tag{1.15}$$

$$\tau = \frac{1}{\alpha(t)} \ln\left[\frac{1}{r_1}\right] \tag{1.16}$$

5. Perform reaction j, updating the number of molecules of each reactant and product, the current simulation time, $t + \tau$, and the propensity functions.

6. Repeat steps 2-4 until the simulation time is exceeded or some 'stop' condition has been satisfied, for example, if no substrate molecules remain in the system.

There are two main drawbacks to this type of stochastic modelling. First, the time steps are often small, making simulations extremely slow. One way this has been addressed is by tau-leaping, presented by Gillespie in (Gillespie, 2001) and by Tian and Burrage in (Tian and Burrage, 2004), who each adapted the SSA to allow for a larger time step, $\tau_1 > \tau$, sacrificing some accuracy in favour of increased simulation speed. This "τ-leap" SSA involves drawing a time step, τ, from a distribution (Gillespie uses a Poisson distribution; Tian and Burrage use a Binomial distribution) in which multiple reactions may occur, thereby simulating $\frac{\tau_1}{\tau}$-iterations of the non-τ-leap SSA simultaneously. This is particularly efficient when simulating a system for long times, or when there are many reactions to consider. However, care is needed to prevent the system entering a physically unrealistic state at time $t + \tau$. For example, if multiple reactions involving the same substrates are simulated in a single time step, then negative numbers of molecules may be predicted. Methods designed to address such issues are discussed in (Tian and Burrage, 2004).

The other main drawback is that the system is assumed to be spatially homogeneous. In reality, two molecules only interact if they are in close proximity; such interactions are highly unlikely in stochastic systems when numbers of molecules are low. Additionally, reactant molecules cannot interact if they reside in different subcellular compartments or are separated by other macromolecules. Equally, they must be within a certain "binding radius" and in the correct orientation in order to react. We address these issues by incorporating spatial effects into the stochastic models that we develop in this thesis (Erban et al., 2007).

1.5.3 Spatial Modelling

There are many mathematical and computational methods for modelling spatial systems. Later in this section, we review some of the available methods and discuss their advantages and disadvantages. First, we explain one of the major factors motivating spatial modelling of biological processes, that is, the impact of macromolecular crowding (Schnell and Turner, 2004).

There have been many studies into the effects of macromolecular crowding on molecular diffusion and reaction kinetics (see reviews by (Schnell and Turner, 2004) and (Mourão et al., 2014b)). The majority of these studies are done in two-dimensional (2D) space on a lattice. We will discuss here the impact of domain features such as dimensionality (2D or 3D), granularity (off-lattice or on-lattice), and periodicity (periodic or reflective boundaries) later in this section.

1.5.3.1 Macromolecular Crowding

In biology, the term "macromolecular crowding" is used to describe molecules, organelles and other intracellular features which restrict or impede the movement of other molecules (those of interest to the researcher) within the cell. The term is primarily used in reference to inert "obstacles", i.e. those which do not themselves interact with the molecules being studied, aside from excluded-volume (steric, usually hard-sphere, repulsive) interactions (Elcock, 2010; Zhou et al., 2008), however, we also consider here the effects of weak force interactions between a fraction of the obstacles and the reactant species whose size is comparable to the size of the obstacles (Mourão et al., 2014a).

It is estimated that anywhere between 5% and 40% of the total intracellular volume is occupied by a variety of macromolecules (Fulton, 1982; Gershon et al., 1985). Macromolecular crowding amplifies the effects of low molecular concentrations: the presence of molecular 'obstacles' makes interactions between two reactive molecular species less likely to occur. Macromolecular crowding is thought to affect reaction and diffusion rates as a result of the volume exclusion effect and potential electrostatic interactions between molecules and obstacles. Reaction kinetics, therefore, depend not only on the number of interacting molecules present, but also on the distribution of these molecules within the cell, and the nature of the crowded intracellular media in which they are diffusing (Alberts et al., 2002).

Excluded-volume interactions occur when two impenetrable particles come into contact, resulting in redirection of the particles' intended trajectory via repulsion (Bruna and Chapman, 2012a,b). These interactions bias the movement of particles towards areas of low particle/obstacle density, which leads to a greater spread of diffusing volume-exclusive particles, which is somewhat counterintuitive. This is particularly true when tracking the diffusion of a single particle undergoing excluded-volume interactions with other particles/obstacles, sometimes known as "self-diffusion". To analyse the self-diffusion coefficient, one needs to label a particular particle, rather than treating all particles as equal. The self-diffusion coefficient can be thought of as a diffusion coefficient intrinsically attached to each particle. This is equivalent to the difference between finding any particle at x compared with finding a particular particle there. Thus the collective diffusion coefficient is not associated with an individual particle which means that it cannot easily be related to the MSD of particles, which is important to consider when considering individual or collective diffusion (Bruna and Chapman, 2012a).

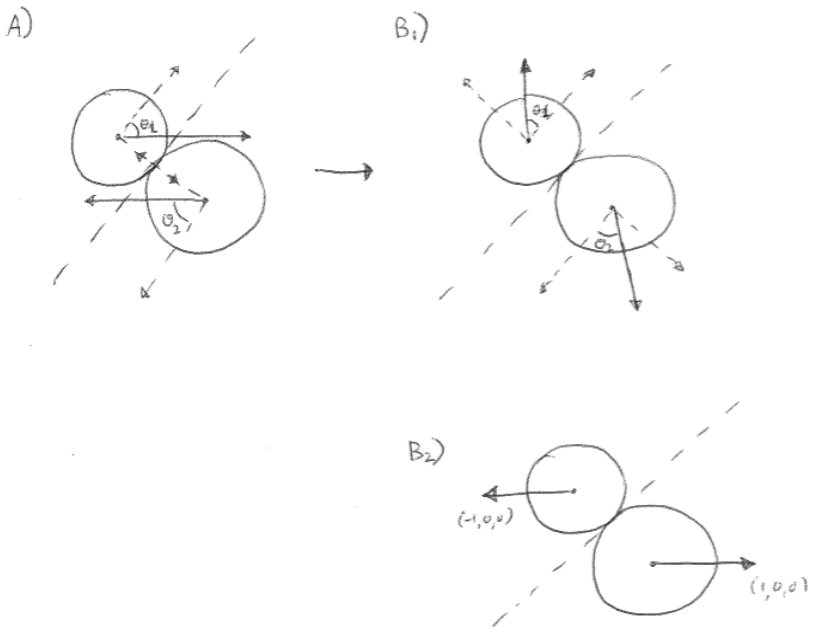


Figure 1.15: Excluded volume interactions. Figure A show two particles coming into contact with each other at a certain angle. Figure B1 show how these two particles would be reflected away from each other following physical laws of excluded-obstacle interaction. Figure B2 shows how Smoldyn handles these two particles interacting, by choosing a pre-defined direction for particle reflection.

In Smoldyn, excluded-volume interactions between particles and/or obstacles (i.e. not boundaries) do not exactly follow physical laws for reflection, that is, they do not reflect as per Figure 1.15B1, but rather a predefined coordinate direction is chosen for particle reflection, regardless of the directional of movement of each particle leading up to the collision. An example is shown in 1.15B2 (Bruna and Chapman, 2012a).

The collective mean squared displacement (MSD) is calculated as

$$MSD(t) = \langle ||\mathbf{X}_i(t) - \mathbf{X}_i(0)||^2 \rangle, \tag{1.17}$$

where $\mathbf{X}_i(t)$ is the position of the ith particle at time t, and angular brackets denote the ensemble average (assuming all particles are identical). In (Bruna and Chapman, 2012a), Bruna et al. study the diffusion of particles ("blue", b) in a domain crowded by obstacles ("red", r). They also look at an individual tagged particle ("blue tagged", b_t). Their ODE model is represented by the matrix system,

$$\frac{\partial}{\partial t}\begin{pmatrix} b^t \\ b \\ r \end{pmatrix} = \nabla \cdot \underline{D} \nabla \begin{pmatrix} b^t \\ b \\ r \end{pmatrix} \tag{1.18}$$

where $\underline{D}$ is a 3×3 diffusion matrix with diagonal elements, D_{ii}, representing the effective diffusion coefficient of the particle group i; The variable $b(x,t)$ is the probability density function of finding any blue particle at x at time t, $r(x,t)$ is the same for red crowders, and $b^t(x,t)$ is the tagged particle. Off-diagonal terms represent non-diffusion terms and are only present for one particle type if the other particle types are not uniformly distributed, e.g. if $\nabla r \neq 0$ then r particles are not uniformly distributed and this will have effects on b and b^t.

The quality Bruna et al. measure when computing self-diffusion is D_{11}, corresponding to "species 1" (i.e. the "blue tagged" particle), given by

$$D_{11} = D_b[1 - (N_b - 1)\epsilon_b^d \gamma_{bb} b - N_r \epsilon_{br}^d \gamma_{br} r], \tag{1.19}$$

where

$$\gamma_{bb} = \frac{\pi}{d}, \tag{1.20}$$

$$\gamma_{br} = \frac{2\pi}{d}\frac{D_b}{D_b + D_r}, \tag{1.21}$$

$$\epsilon_{br} = \frac{\epsilon_b + \epsilon_r}{2}. \tag{1.22}$$

Here, N_b and N_r represents the number of blue and red crowders of size ϵ_b and ϵ_b, respectively. Consequently, ϵ_{br} is the distance at contact between a blue and red particle. If both blue and red are uniformly distributed then the probabilities b and r are both 1 and the equation simplifies to

$$D_{11} = D_b[1 - (N_b - 1)\frac{\pi}{d} - N_r \epsilon_{br}^d \frac{2\pi}{d}\big(\frac{D_b}{D_b + D_r}\big)]. \tag{1.23}$$

The volume fractions can be expressed as

$$\phi_b = \frac{\pi N_b \epsilon_b^d}{2d}, \tag{1.24}$$

$$\phi_r = \frac{\pi N_r \epsilon_r^d}{2d}. \tag{1.25}$$

Whilst there are many factors involved in simulating a crowded intracellular environment (e.g. volume fraction, obstacle size, etc), some of which we explore later, we argue that, no matter how the crowding is initialised, the resulting model will be more representative of the true intracellular conditions than a dilute model system, and is therefore worth including in any spatially-resolved cellular model (Elcock, 2010). Macromolecular crowding impedes the passage of other molecules through the domain, and may result in certain subregions of the domain being completely inaccessible to diffusing molecules. As a result diffusion rates may be slowed and anomalous subdiffusion has been observed for intermediate timescales. It is often assumed that the anomalous diffusion observed in many biological observations is a direct result of macromolecular crowding, with the power law exponent, α, depending on the volume fraction of obstacles, ϕ. However, there is evidence to suggest that diffusion throughout randomly distributed immobile obstacles results only in hindered diffusion (reduced diffusion constant), with anomalous diffusion only occurring with mobile obstacles, and close to the "percolation threshold", although this subdiffusive regime is known to rapidly vanish when obstacles move by Brownian motion.

Note, the percolation threshold here is the limit at which a large connected component forms from independent objects, such that there is a complete flow through the object. It comes from percolation theory, which describes how a connected network forms in random systems, through which substances can "percolate". The percolation threshold is the level of crowding at which infinite connectivity (percolation) occurs. Below the threshold, no such connected network exists. In our case, this would be the point at which all the obstacles are interconnected and molecules can diffuse through the entire available space. In reality, our simulations will never reach this threshold as our obstacles have a fixed size and are hard spheres. This is known as "auto-crowding", that is, obstacles cannot overlap and, therefore, there may be areas of the space which are not occupied by any object, but which cannot be reached by a diffusing molecule.

In our study, we are particularly interested in diffusion-limited reactions, where the two reactant

molecules will react (with almost certainty) when the molecules are within their binding radius. These reactions are, however, limited by the diffusion of each molecular species. In the case of the Smoluchowski-Einstein formula, the rate of diffusion-limited reactions is proportional to diffusion, i.e. $k_1 = 4\pi Dr$, where D and r are the diffusion constant and reaction radius, respectively (Erban, 2014; Kopelman, 1988). Both the diffusion and reaction rates follow a power law distribution in time, given by the power law,

$$k = k_1 t^{-h}, 0 \leq h \leq 1 (t \geq 1) \tag{1.26}$$

$$\implies \log k = -h \log t + \text{constant}, \tag{1.27}$$

where h is the fractal-like kinetic parameter, initial reaction rate, $k_1 \equiv k(t = 1)$, and the new reaction rate coefficient, k. For Brownian motion, $h = 0$. For diffusion-limited reactions, $h \neq 0$ and k=$k(t)$ is no longer constant but varies over time. These types of reactions are also known as "fractal-like kinetics". For reactions between two different molecular species (in 3D space), $h \approx 1/3$ (Kopelman, 1988) (Kopelman, 1988).

For steady states of the system, the rate of change of reactants is, by definition, zero. That is, reactants are created and destroyed at the same rate. The most general case is given by

$$\frac{d[A]}{dt} = R_A - k[A][B] \tag{1.28}$$

$$\frac{d[B]}{dt} = R_B - k[A][B], \tag{1.29}$$

for two reactive species, A and B, that react with rate k and are created at rates R_A and R_B, respectively.

An good example of fractal-like kinetics was given in (Kopelman, 1988) which we paraphrase here. Assume we have two vessels (e.g. beakers) with exactly the same internal and external conditions. However, one contains a certain concentration of molecular species, A, which is initialised at $t = 16$s, and the other is given the same concentration of A but at the earlier time of $t = 1$s. Molecules of A should react at the same rate, k, in both containers (according to $A + A \xrightarrow{k} A$), but they do not when they follow fractal-like kinetics. Why? Assume $k = k_1 t^{-h}$ with $h = 1/2$. In the first container, $k = k_1 \times 16^{-1/2} = k_1/4$, whereas in the second container $k = k_1 \times 1^{-1/2} = k_1$. So the second

container has an initial rate coefficient, k, that is four times greater than the first container. This is possible in a fractal-like reaction system as the distribution of molecules throughout the space tends to become more ordered over time, as diffusing molecules are biased towards, and therefore oscillate around, their initial position (Kopelman, 1988). For fractal-like systems, such as percolation clusters, a segregation of "unlike" molecules (e.g. A and B) and an aggregation of "like" molecules (e.g. A to A, and B to B) may be observed.

Smoldyn operates on Brownian Dynamics, diffusion-limited kinetic simulations (Erban, 2014). The drawback of using such BD models is that they are more computationally expensive and require more parameters. Following Smoluchowski dynamics, Smoldyn determines that a bi-molecular reaction occurs when reactants are closer than their reaction radius,

$$r = \frac{k}{4\pi(D_A + D_B)} \tag{1.30}$$

where D_A and D_B are the diffusion contstants of reactants A and B, and k is the reaction rate of the bi-molecular reaction $A + B \rightarrow C$. From this, one can determine the microscopic reaction radius r from the macroscopic reaction rate k. Since not all reactions will occur within this radius, e.g. due to molecule alignment, activation energies, etc, one can additionally declare a probability, f, that each reaction will occur, given that the reacting molecules are within the radius r.

Experimentally (e.g in crowded *in vitro* and *in vivo* studies), systems have been exhibit a fast transient period of anomalous diffusion at intermediate time scales, with normal Brownian diffusion at small and long time scales (Höfling and Franosch, 2013). To differentiate between anomalous subdiffusion and this transient behaviour, one can measure the exponent, α, of the power law relation $\langle \Delta r^2(t) \rangle \propto t^\alpha$ (Berezhkovskii et al., 2014); For subdiffusion, α is constant, whereas for transient behaviour α depends on time ($\alpha(t)$). There are three ways to measure α: fitting the MSD vs t plot to the line t^α; measuring the area under the MSD curve and using it in the formula,

$$\alpha = \left(\frac{t \langle \Delta r^2(t) \rangle}{\int_0^\infty \langle \Delta r^2(t') \rangle dt'} \right) - 1; \tag{1.31}$$

or differentiating $\langle \Delta r^2(t) \rangle \propto t^\alpha$ with respect to time to get the formula,

$$\alpha = \frac{t}{\langle \Delta r^2(t) \rangle} \frac{\mathrm{d}\langle \Delta r^2(t) \rangle}{\mathrm{d}t} = \frac{\mathrm{d}\ln\langle \Delta r^2(t) \rangle}{\mathrm{d}\ln t} \tag{1.32}$$

For Brownian diffusion, the diffusion coefficient is constant as MSD $\propto t$. Even in hindered diffusion, the MSD remains proportional to time, but the diffusion coefficient would be reduced. In anomalous subdiffusion, MSD $\propto t^\alpha$, where $0 < \alpha < 1$ (for superdiffusion, $1 < \alpha < 2$).

In (Saxton, 2007b), Saxton suggests that mobile obstacles lead to anomalous diffusion only over short times/distances, tending to hindered Brownian diffusion, ie.

$$\langle r^2(t) \rangle = 2dDt, \tag{1.33}$$

over long times/distances, where d is the dimension of the domain (in our case $d = 3$), and $D < D_0$. He suggests that anomalous diffusion occurs close to the percolation threshold, or when there are correlations in diffusive motion, or in the presence of molecular "traps" (e.g. binding sites). Similarly, confined diffusion may also be observed in the cytoplasm due to the finite size of the domain, with diffusion normal at short times/distances, tending to a constant that is proportional to the size of the domain for long times/distances. He represents the time dependence of the diffusion coefficient as

$$\langle r^2(t) \rangle \approx \langle r^2(t) \rangle_0 [1 - \exp(-t/\tau)] \tag{1.34}$$

where τ is the jump time. Saxton additionally performed independent Monte Carlo studies of the effect of obstacles and binding on the presence of anomalous diffusion (Saxton, 1994, 1996).

The power law dependence of diffusion has been shown to persist even in the presence of reflective bounding wall, e.g. the cell membrane, to which momentum can be transferred as the intended molecular trajectories are impeded and altered. In the Lorenz model, a single particle scatters ballistically throughout a random distribution of immobile hard obstacles, resulting in a dependence of the diffusion constant on the density of obstacles, resulting in a power law tail of the form $At^{(d+2)/2}, A < 0$, where d is the dimension of the system (in our case, $d = 3$) (Franosch et al., 2010;

Alder and Wainwright, 1967, 1970).

One model of macromolecular crowding is given the nickname of the "Swiss cheese" model, for obvious reasons. The obstacles are modelled as spheres distributed according to a Poisson process in an $L \times L \times L$ domain. The spheres form "clusters" if they come into contact with, or overlap, neighbouring spheres. The remaining space forms a porous path through which other molecules can diffuse. The density of spheres at which a single cluster spans the entire domain is called the "critical density". There are three common variables for expressing the critical density: the volume fraction ϕ_c, the reduced number density η_c, and the number density n_c. They are related in the following ways (for 3D space):

$$\phi_c = 1 - e^{-\eta_c} \tag{1.35}$$

$$\eta_c = (\pi/6)n_c \tag{1.36}$$

$$\boldsymbol{n}_c = N(2R)^3/L^2, \tag{1.37}$$

where R is the radius of the spheres and N is the number of spheres distributed in the domain (Lorenz and Ziff, 2001).

For 3D systems, most models predict that, at the percolation threshold, $n_c \approx 0.652960$ which can be converted into volume fraction using the above formulae to obtain a value of $\eta_c = 0.341889$ and $\phi_c \approx 0.289573$ (Lorenz and Ziff, 2001; Rintoul and Torquato, 1997). Another value used to represent the critical density is the number of points in a spherical domain $W_c = 4\pi n_c/3$. In the Lorenz model, a Poisson distribution is used to determine how many obstacle centers were in each of the cubes ($L \times L \times L$ cubes of unit volume), where the probability P_n that there are n obstacles in a given volume V is given by

$$P_n = \frac{1}{n!}(\rho V)^n e^{-\rho V} \tag{1.38}$$

where ρ is the density of obstacle centres in each cube of volume, V ($V = 1$ for unit cubes). Another limitation of such simulations is in the effects of a finite-sized domain, which can impact on the accurate interpolation of the critical threshold value. One commonly used method to overcome these finite-size effects is to use periodic boundaries, although these are biologically unrealistic.

Finite-size effects may also have little to no impact if the boundaries are rarely, if ever, encountered by diffusing particles, which can be estimated by measuring the ratio of the MSD of each type of particle to the size of the domain. If the ratio is very small, the particle is unlikely to contact the domain boundaries unless it starts from a position close to the boundary, which can, to some degree, be controlled in most simulations. For example, for a spherical domain of radius, R, and a maximal MSD of m, we can create a compartment of radius, $r = R - m$, within which all of our particles are initialised.

Volume Fraction of Obstacles

The volume fraction of obstacles, θ, has been shown to affect diffusion and reaction rates in simulations. In (Mourão et al., 2014a), increasing the volume fraction generally increases bi-molecular rate for large obstacles. This is valid from $\theta = 0$ (dilute) to $\theta = 0.4$. On the other hand, for small obstacles, increasing density generally reduces bi-molecular rates. The work presented by Mourao et al. is done for 2D space, whereas our models will build upon this work to look at 3D space.

The percolation threshold is the point at which the space first possesses long-range connectivity and global mixing halts (volume fraction, $\phi = 0.2895$, for a sphere of radius R). Below this threshold, anomalous diffusion is observed at short times (Saxton, 2007a). At this point, the MSD scales as $\langle r^2(t) \rangle \propto t^{\alpha}$, where, in three-dimensional space, $\alpha = 0.54$ according to (Nicolau et al., 2007). As the percolation threshold is approached, anomalous diffusion is observed at longer and longer time scales. When the percolation threshold is finally reached, anomalous diffusion is observed over all timescales (Saxton, 2007a).

As θ increases, Mourao et al. show that reversible and catalytic reaction rates fall, and that reactions become less dependent on θ as the size of the obstacles increases. Again, we will examine this in our 3D models.

Size of Obstacles

In (Berry and Chaté, 2014), the effects of "polydispersity", i.e. a range of obstacle sizes, is studied and found to have a very weak influence on the diffusion of molecules throughout the domain. This is useful for spatial modelling as it reduces the complexity of the obstacle initialisation as they can

be created to be uniform in size and diffusion coefficient. However, we must still study the effects of obstacle size as the relative size of obstacles to other diffusing molecules is important; Increasing obstacle size, δ, generally creates more space for diffusion for a fixed volume fraction (in 2D) (Mourão et al., 2014a).

Mobility of Obstacles

Immobile obstacles reduce the available volume for diffusion of particles, but are not very realistic as most macromolecular crowding is caused by other mobile macromolecules. With diffusing obstacles, anomalous diffusion seems to disappear, but if the obstacles themselves move anomalously then subdiffusion returns. One possible way to do this is to have obstacles impeded by another set of molecules.

In vivo measurements of molecular diffusion within cells are consistently recorded as "anomalous", particularly sub-diffusive, such that the mean squared displacement (MSD) of molecules scales as a power law of time, t^α, where $0 < \alpha < 1$, with macromolecular crowding often cited as the primary cause (Berry and Chaté, 2014). In (Berry and Chaté, 2014), Berry and Chaté explain how this conclusion is based on studies that look at macromolecular crowding due to immobile obstacles. However, most obstacles or "crowders" in a cell are mobile, such as other proteins, RNAs, vesicles, etc. As such, simulations with diffusing obstacles were carried out, however, in this case, the observed anomalous subdiffusion vanishes. In order to restore anomalous diffusion in simulations, whilst maintaining the reality of mobile obstacles, Berry proposes the use of an Ornstein-Ulhenbeck regime for obstacle movement, in which the obstacles are themselves inhibited in their movement throughout the space and rarely move far from their initial position. Effectively, this amounts to Brownian motion with the additional condition of confining the obstacle to a restricted subspace of the cell close to its initial position. In more detail, if we consider an obstacle with position $\mathbf{x}_{\text{obs}}(t)$, Brownian motion would update the position of the molecule at time $t + \delta t$ as

$$\mathbf{x}_{\text{obs}}(t + \delta t) = \mathbf{x}_{\text{obs}}(t) + \sqrt{2D_{\text{obs}}\delta t}\xi, \tag{1.39}$$

where ξ is a normally distributed random number with zero mean and unit variance, i.e. $\xi \in N(0, 1)$,

D_{obs} is the diffusion coefficient of obstacles. For OU motion, this formula becomes

$$\mathbf{x}_{\text{obs}}(t+\delta t) = \mathbf{x}_{\text{obs}}(t) + \delta t \frac{\mu - x(t)}{\tau} \sqrt{2D_{\text{obs}}\delta t}\xi, \tag{1.40}$$

where μ is the long-term equilibrium position of the obstacle (in this case, $\mu = \mathbf{x}(0)$), and τ is the relaxation time for the obstacle to reach its equilibrium position, μ ($t = \infty$ corresponds to Brownian motion).

In our simulations, we measure the diffusion of obstacles and see that their movement is indeed hindered. This is likely due to the inclusion of volume excluded reactions between obstacles as well; In Berry's study, the only volume excluded reactions are of the form Obs + X $\rightarrow$ Obs + X, where X can be any other molecule of interest other than another obstacle. This allows obstacles to diffuse freely throughout the space and potentially overlap. In our model, we include X + X $\rightarrow$ X + X, for all molecules including obstacles. This limits the movement of obstacles through space in a way that mimics OU movement (or at least, keeps molecules close to where they were initially placed which is the generic idea behind OU motion).

Dimensionality, Granularity and Periodicity of the Domain

Lattice-based models can be as fine- or coarse-grained as the user chooses, however, computational expense increases with the fineness and the size of the mesh. Usually, the maximum coarseness allows for at most one molecule (of the smallest species) per pixel/voxel, and movement is restricted to 4, 6, or 8 directions. For off-lattice models, spatial coordinates are modelled explicitly, which gives the molecules more freedom of movement but further increases the computational cost. Both types of simulations are limited by the computational time, the size of the domain, the number of molecules to track, and the number of potential reactions that can occur.

Two studies by Kammerer et al. (Kammerer et al., 2008) and Höfling et al. (Höfling et al., 2008) examined the off-lattice movement of point-particles (no explicit size) on large spatial domains and over long time scales. They reported no relationship between α and obstacle density, however they do discuss how diffusion becomes anomalous as the system approaches the percolation threshold; It is the topology of the obstacles rather than their density that determines how easily molecules can diffuse through the space.

The periodicity of the domain also plays a role on the reaction and diffusion dynamics in the system due to "finite-size effects", which further restricts the movement of molecules throughout the domain. Periodic boundaries reduce finite-size effects, but they are unrealistic, particularly in 3D, models of the cell.

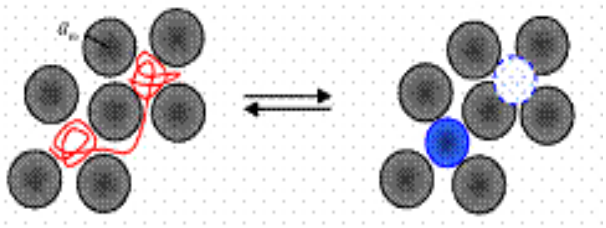
The effects of crowding on cellular thermodynamics and molecular diffusion were first reported in the late 1980s and early 1990s (Zimmerman and Minton, 1993; Luby-Phelps et al., 1986). In particular, Kopelman et al. showed that the Law of Mass Action no longer holds for diffusion-limited reactions in crowded, spatially heterogenous, subcellular environments (Kopelman, 1986, 1988). They also showed that for the enzymatic reaction,

$$\mathrm{E} + \mathrm{S} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{C} \xrightarrow{k_3} \mathrm{E} + \mathrm{P} \ , \tag{1.41}$$

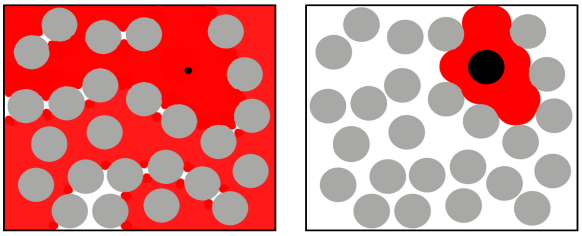
the complex formation rate, k_1, is no longer constant but is said to decay linearly with time in diffusion-limited reactions, i.e. when the probability of enzyme-substrate binding is approximately one, and diffusion becomes the primary source of stochasticity and variability in the process (Kopelman, 1986, 1988). There have been many studies of macromolecular crowding, showing its impact on diffusion coefficients and reaction rates for varying volume fractions (proportion of cellular volume obstructed by obstacles) and obstacle sizes (Schnell and Turner, 2004).

In 1983, Minton proposed that the volume exclusion effect caused by macromolecular crowding had a significant effect on biological processes (Minton, 1983). Minton had commented that crowders (or "background molecules") had the biggest impact on the system when their size was similar to the other molecules of interest, because they generated stronger repulsive forces between molecules and obstacles. Alternatively, if the diffusing molecules are much smaller than the crowders, then they may be able to diffuse between the crowders, negating the volume exclusion effect (see Figure 1.16) (Minton, 2001).

Schnell and Turner reviewed the importance of macromolecular crowding, attributing the breakdown of the Law of Mass Action in inhomogeneous, crowded intracellular media due to the resulting analomalous diffusion and molecular mixing throughout the cell (Schnell and Turner, 2004). More recently, Nicolau and Burrage demonstrated how Monte Carlo simulations (performed using cellular automata) can be used to estimate an "anomalous diffusion" parameter, $\alpha \in [0, 1]$, in order to



(a)



(b)

Figure 1.16: (a) Small molecules may diffuse between crowders whereas molecules of a similar size to the obstacles are much more restricted. (b) The volume avalailable to small molecules diffusing in a highly crowded medium is much greater than for large molecules.

describe the effect of macromolecular crowding on a system (Nicolau and Burrage, 2008). Fickean diffusion has an anomalous diffusion parameter of $\alpha = 1$ and can be characterised by the mean-square displacement (MSD),

$$\left\langle r^2 \right\rangle \propto Dt, \tag{1.42}$$

that is proportional to the constant diffusion coefficient, D, and time, t (Saxton, 2007a). For anomalous diffusion, the MSD becomes

$$\left\langle r^2 \right\rangle \propto t^{\alpha}, \tag{1.43}$$

where $\alpha < 1$ represents sub-diffusion. The corresponding diffusion coefficient is time-dependent and given by

$$D(t) = \frac{D_0}{1 + t^{1-\alpha}}, \tag{1.44}$$

where $D_0 = D(0)$. For a three-dimensional system, with macromolecular crowding approaching the

"percolation threshold" (volume fraction, $\theta \sim 0.4$), the system is reported to have an anomalous diffusion parameter of $\alpha \approx 0.54$ (Nicolau and Burrage, 2008; Wachsmuth et al., 2000). Additionally, reaction rates are found to depend on time, t, and α as a result of macromolecular crowding, such that a bi-molecular reaction which occurs at rate k_0 in dilute systems, occurs at rate

$$k(t) = k_0 t^{-h}, \tag{1.45}$$

where $h = 1 - \alpha$ is termed the "fractal parameter" and $k_0 = k(0)$. In dilute systems in three-dimensional space, $h = 0$ and $k(t) = k_0$ (Grima and Schnell, 2006).

In one of the most recent studies, Mourao et al. propose that increasing obstacle size leads to increased diffusion rates, whereas increasing the volume fraction reduces the rates of diffusion. They also highlight the impact of non-specific electrostatic interactions between molecules of interest and obstacles, in a Michaelis-Menten model of enzyme-substrate dynamics (Equation (1.41)) (Mourão et al., 2014a). Specifically, Mourao et al. talk about "fractal-like kinetics", which can be used as a measure of the deviation from the Law of Mass action. The fractal-like kinetic parameter, h, is shown to increase with volume exclusion (or volume fraction), θ, and they show that as θ increases, the obstacle size, δ, has a greater impact on the fractal-like kinetics.

Matsuda et al. recently compared moving and stationary obstacles in a model of transcription factor binding dynamics (Matsuda et al., 2014). Their study is noteworthy because it specifically addresses macromolecular crowding (by obstacles) in the nucleus. Transcription factors (TFs) search for a binding site on DNA by undergoing a special mechanism known as facilitated diffusion. Facilitated diffusion consists of periods of three- and one-dimensional diffusion within the cell (Mirny et al., 2009). TFs diffuse freely within the cell until they bind non-specifically to DNA, at which point they slide (via one-dimensional diffusion) along the length of the DNA. During the sliding, TFs may find a specific binding site on the DNA to begin transcription. Via this method, TF association to its DNA target site is approximately 1000 times faster than other protein-protein associations within the cell (Mirny et al., 2009). Matsuda et al. estimate that the crowding (volume) fraction is around $\theta_n = 0.3$. They report that mRNA levels, transcribed as a result of transcription factors binding to DNA, are non-monotonically dependent on the volume fraction, with maximum transcription rates observed for the highest volume fraction that they studied ($\theta_n = 0.3$). They compare, what they call, "dilute" conditions (i.e. no crowding $\theta_n = 0$), with crowded conditions, and explain why the

reaction rate changes in the following way:

1. Reaction rates of transcription factors (TFs) binding non-specifically to DNA are reduced due to impaired diffusion in the crowded medium.

2. Stronger bond between TFs and DNA once bound, due to the more favourable entropic energy state of being bound together compared to being separate; The excluded volume of the complex is smaller than the excluded volume of the the two substrates separately.

3. This favourable entropic state also introduces an activation energy barrier for binding and unbinding of TFs to DNA, making it harder to TFs to associate and dissociate from DNA.

As Matsuda et al. state in (Matsuda et al., 2014), mRNA concentration changes non-monotonically because of these contrasting effects of crowding on the dynamics. Namely, crowding causes enhanced binding between TFs and DNA, whilst simultaneously causing impairment in TF diffusion and non-specific binding to DNA. In our models we consider the impact of restricted diffusion, volume exclusion and non-specific obstacle-molecule interactions on the system. However, in particularly crowded media, the binding and unbinding rates will undoubtedly be affected. For example, a complex will be unable to dissociate if there is not enough surrounding volume to contain the separated molecules. Similarly, two reactive molecules will be more likely to end up within their binding radii of each other, thereby forming a complex, if the surrounding volume is limited. Similar to our simulations, Matsuda et al. assume spherical 'crowders' on the order of their molecules of interest (TFs).

As stated earlier, in our simulations we are interested in diffusion-limited reactions, in which reactive molecules interact with probability close to one upon interaction, and the likelihood of contact is limited only by the diffusion of each type of molecule throughout the domain. As such, the ODE is likely to be a good approximation to the mean of the spatial simulations when the diffusion of molecules is normal (taking ODE parameters from the initial output of Smoldyn; note that these results may still diverge from published parameters depending on the diffusion coefficient), and will diverge when diffusion is anomalous. We, therefore, push our simulations to the point of anomalous diffusion (e.g. by reaching the percolation threshold, creating traps through non-specifically binding obstacles, confining the reactions to a finite cellular domain, etc) and study the system divergence from the ODE model at these limits.

From this, it is clear that spatial modelling of cells can have a big impact on the behaviour of chemical kinetics in silico. However, spatial stochastic models are more computationally expensive, particularly in three-dimensional systems. In this thesis we analyse the importance of modelling biochemical reactions with a high level of spatial and stochastic detail by examining the degree of divergence between deterministic modelling and spatial stochastic modelling in prediction of *in vivo* chemical kinetics. The significance of this difference is the most important factor, as the improvement in matching experimentally observed dynamics must be weighted against the increase in computational expense. We hypothesise that spatial and stochastic modelling is worth the additional computational effort, based on the aforementioned studies.

Now, we present an overview of the available spatial modelling methods, give some advantages and disadvantages for each method. The section is roughly ordered with increasing stochasticity and spatial detail.

1.5.3.2 Compartmentalised Ordinary Differential Equations

The first step towards a full spatial model is in the compartmentalisation of a system of ODEs. This is useful when, for example, molecules can move between intracellular compartments, such as the nucleus and the cytoplasm, and when modelling the behaviour in both compartments simultaneously is necessary. In such a situation, these molecules are usually modelled as different species, one for each compartment, such as "cytoplasmic" and "nuclear". Additional transitions are then added to the system to represent molecules that have moved between compartments, although, in reality, the molecules themselves have not physically changed. For example,

$$\mathrm{A}_{cyt} \underset{k_b}{\overset{k_f}{\rightleftharpoons}} \mathrm{A}_{nuc} \tag{1.46}$$

represents the movement of molecules of type A betweent the nucleus and the cytoplasm, although they are modelled separately as A_{cyt} and A_{nuc}, with $\mathrm{A}_{cyt} + \mathrm{A}_{nuc} = \mathrm{A}$. This approach is better than a non-compartmentalised system of ODEs for processes that rely heavily on the transport of molecules between compartments. However, molecules are still assumed to be homogeneously distributed and to behave deterministically, including the transition between compartments which is unrealistic. We look at compartmentalised ODEs in the Wnt signalling pathway in Section 4.

1.5.3.3 Partial Differential Equations

Partial differential equations go one step further in explicitly modelling the evolution of systems in space as well as time. They are akin to ODEs, except that the variable of interest is now represented as a function of space, x, and time, t. PDEs are also much more difficult to solve analytically and one must rely on numerical methods of analysis. They are useful for modelling Brownian diffusion of molecules, for example, however they do not solve the problem of explicitly modelling biochemical reactions that depend on the spatial distribution of each molecule, as the reaction rates still only depend on time. Additionally, the diffusion coefficients of many molecules are unknown, and approximating the coefficients using the molecular weights (in Daltons) is likely to be insufficient as movement of molecules will be affected by macromolecular crowding (Neves and Iyengar, 2009; Fulton, 1982; Minton, 2001). Therefore, we turn to more explicit spatial models. We also address the assumption of molecules behaving deterministically in ODEs and PDEs, and turn to stochastic models for more realistic models of biological processes involving small numbers of molecules.

1.5.3.4 Compartmentalised SSA

The SSA model described previously only modelled biochemical reactions and not the diffusion of molecules. One method for building the "reaction-SSA" into a "reaction-diffusion-SSA" is to compartmentalise the space and have a separate "diffusion-SSA" describing the movement of molecules between neighbouring compartments (see Figure 1.17 and Figure 1.18). For example, movement of molecular species, A, between N neighbouring compartments of length L in a 1D system can be described by the system of chemical reactions

$$A_1 \underset{r}{\overset{r}{\rightleftharpoons}} A_2 \underset{r}{\overset{r}{\rightleftharpoons}} \cdots \underset{r}{\overset{r}{\rightleftharpoons}} A_N, \tag{1.47}$$

where the rate constant, r, is given by $r = D/L^2$, where D is the diffusion coefficient of A. At $t + \delta t$, the compartmental positions of each molecule are determined according to the above algorithm. Then, for each compartment, the reaction-SSA is run, using a count of the number of each type of molecule in each box, to see if any reactions will occur. However, the solution is not trivial for non-linear reactions and compartmentalising the entire cellular space and simulating each voxel manually would be very computationally expensive.

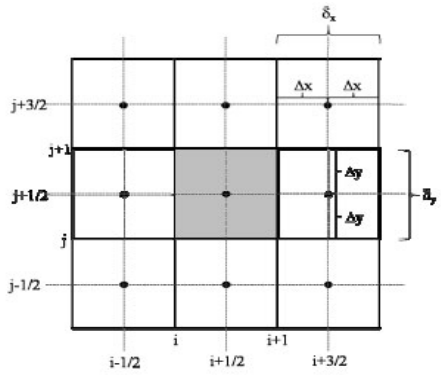


Figure 1.17: A schematic of the movement within a 2D grid.

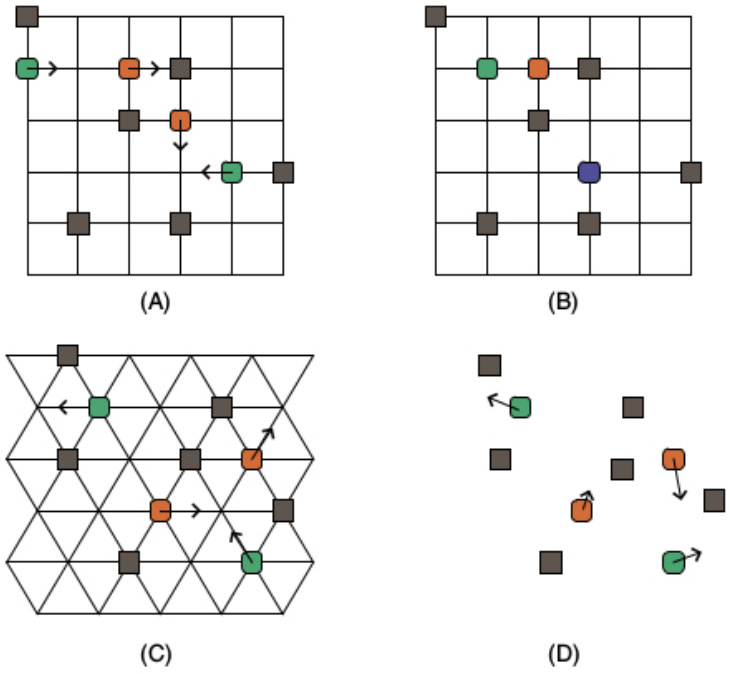


Figure 1.18: A schematic of the possible 2D domains for molecule movement (Grima and Schnell, 2006). (A) and (B) show the movement of molecules on a regular square grid over one time step, in which molecules can move in one of four possible directions. (C) shows movement on a triangular lattice, in which diagonal movement is inclued. (D) shows movement off-lattice.

A more accurate method for modelling the diffusion of molecules is to use the Smoluchowski equation for diffusion of individual molecules (Erban and Chapman, 2009; Erban, 2014). The diffusion of each molecule is described by a diffusion coefficient, D, which is determined by molecular size and the temperature and viscosity of the solution in which the molecule is diffusing, for example, the cytosol for diffusion in the cytoplasm. The new position of the molecule at time $t + \delta t$, for a small time step

δt, in 3D is calculated as follows:

1. Generate three normally distributed random numbers, ξ_x, ξ_y and ξ_z, from $N[0,1]$.
2. Calculate the new x, y and z coordinates at time $t+\delta t$ as

$$X(t+\delta t) = X(t) + \sqrt{2D\delta t}\xi_x, \tag{1.48}$$

$$Y(t+\delta t) = Y(t) + \sqrt{2D\delta t}\xi_y, \tag{1.49}$$

$$Y(t+\delta t) = Y(t) + \sqrt{2D\delta t}\xi_y. \tag{1.50}$$

Combining this algorithm with the reaction-SSA is not trivial, however, this is the underlying mechanism of the spatial-stochastic modelling tool, Smoldyn, which is based on Smoluchoski Dynamics. In Smoldyn, the positions of all molecules are updated at each small time step δt and then the reaction SSA is run on particles which are now determined to have come into contact with each other (those which are within a certain reactive radius of each other). For example, consider a system two reactive molecular species, A and B, of radius r_A and r_b, diffusing through a model of the cytoplasm. They react at a rate of $k_A B$ and diffuse at rates D_a and D_b respectively. At each time step δt the positions of all molecules of A and B are updated according to the above model, and those molecules of A and B whose centres are now within $r_A + r_B$ of each other will react with probability $k_A B$.

We use Smoldyn to simulate our non-linear reaction-diffusion models of post-transcriptional regulation, RNAi and Wnt signalling (see Section 1.5.4). First, we present alternative stochastic simulation tools to compare with the capabilities of Smoldyn. In most cases Smoldyn has distinct advantages over other software, in terms of accuracy, speed or in-built modelling features and capabilities. In some cases, there is little difference between the capabilities of Smoldyn and the software under discussion. In such cases, Smoldyn was ultimately chosen because it is widely used, can run on Linux-based machines, and has more available examples, support and documentation, for example.

1.5.3.5 On-Lattice Models and Cellular Automata

We used Smoldyn to simulate an off-lattice model of our chosen biological processes. Alternatively, an off-lattice model could have been used. For example, Cellular automata (CA) are lattice-based

models in which discrete particles move in a finite number of directions on the grid with interactions possible between neighbouring grid spaces. A recent study published by Mourao et al. detailed the effects of crowding on diffusion and reaction rates in a two-dimensional cellular automaton model for an enzyme-substrate reaction system (Mourão et al., 2014a). They showed that these rates depend not only on the volume fraction but also on the size of the "crowders" and any non-specific binding which may be possible between the crowders and the enzymes. However, Mourao et al. also showed that modelling such reaction-diffusion systems in two dimensions causes some loss in accuracy when observing the kinetics of molecules in a highly crowded system. Creating a three-dimensional CA would be very computationally expensive.

Before we began using Smoldyn, we initially created a cellular automaton for the model of post-transcriptional regulation of *PTEN* (see Figure 1.19). In our two-dimensional CA model, mRNA molecules were defined as a string of pixels of length N_b, representing the MREs. The mRNAs could rotate clockwise or anti-clockwise, diffuse in any of eight spatial directions, and interact with miRNAs (single pixels in the grid) if they came into contact with a free MRE.

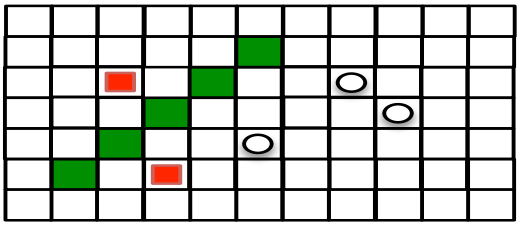


Figure 1.19: A schematic of cellular automaton for PTEN mRNA (green vector) regulation by miRNAs (red circles) and copycats (white circles).

We decided not to continue with this type of spatial modelling technique primarily because the system was proving too computationally expensive. There are a number of different ways to model spatial and stochastic systems, and quite a few modelling software tools available to do so. Aside from the fact that we simply did not have time to create a model in each software for comparison, we describe below the main reasons for choosing Smoldyn over the other available software. We limit our discussion here to methods and tools which are capable of both stochastic and spatial modelling.

WinSSS is a particle-based, discrete space, continuous time, stochastic simulation tool (Krone, 2004; Guan and Krone, 2004). Interactions are modelled via a Markov process on a two-dimensional lattice. The main drawback of this tool was that it is Windows-based and we did not have access to a Windows-enabled computer. Further to this, we decided that continuous space in three dimensions

made the use of Smoldyn more advantageous in terms of accuracy of modelling interacting particles. In particular, it has been suggested that volume fraction has a lower impact on the rate of diffusion with the addition of a third spatial dimension (Vilaseca et al., 2011; Mourão et al., 2014a).

MesoRD is a continuous time, discrete space, stochastic simulation tool capable of modelling in three-dimensional space. The reaction volume is split into sub-volumes in and across which molecules can diffuse and interact according to chemical reaction equations and rate constants. The main disadvantages of this software are the potential for losing information across sub-volumes, the necessity to use Systems Biology Markup Language (SMBL) in order to code up the simulations, and the lack of macromolecular modelling capabilities.

SmartCell is a lattice-based software tool that utilises a compartmentalised reaction-diffusion SSA. The space is discretised into voxels (cubes or triangles), with diffusion, active transport and reactions occurring within and between nearest neighbouring subvolumes. The space is considered to be meso-scale which results in a lower accuracy compared with the micro-scale of Smoldyn and other off-lattice simulators. The primary drawback of this software is the inability to model macromolecular crowding, which is imperative for our spatial models.

E-Cell is a software tool initially designed as a temporal modelling platform which is now capable of modelling microscopic-lattice-based simulations. Two-dimensional and three-dimensional simulations are possible, in which the space is discretised into a hexagonal lattice. The biggest advantage of this software is the capability of simulating macromolecular crowding, which is unusual for lattice-based tools. Smoldyn is also capable of modelling three-dimensional systems with macromolecular crowding but will provide greater accuracy from off-lattice modelling. However, the main reason for not proceeding with E-Cell was that it was only available for use with Window-based operating systems, which we did not have access to.

1.5.3.6 Alternative Off-Lattice Models and Spatial Stochastic Simulators

Aside from a reaction-diffusion model within a simulated cellular compartment, as is used in Smoldyn, there are other methods for modelling biological systems off-lattice (Burrage et al., 2011). There are also several open-source alternatives which we could have used for our off-lattice simulations. Below we describe these different tools in comparison with Smoldyn.

GROMACS is a molecular dynamics software for simulating particle interactions based on Newton's equations of motion using activation energies, electrostatic forces, etc. It has recently been used by Matsuda et al. to model transcription factors binding to DNA in a crowded environment (Matsuda et al., 2014), so its utility for the study of macromolecular crowding in biological systems is unquestionable. Unfortunately, this study was only published in March 2014 so we were unaware of the potential of molecular dynamics simulations for this kind of modelling, and we had focused our search on the identification of a reaction-diffusion modelling tool. Given more time, we would compare GROMACS capabilities versus Smoldyn. We believe the main difference would be the speed up in simulation time given that GROMACS is capable of being parallelised.

ChemCell is a particle-based simulation tool capable of running both spatial and non-spatial simulations (via ODEs and SSAs). It is generally considered easy to use, although it lacks many features compared with other software tools of this kind. The main drawback of using ChemCell for our models was that it does not support zero-order reactions, of the form $0 \rightarrow A$, and when compared with Smoldyn, it has a lower accuracy when modelling second order reactions, such as bi-molecular reactions. These two types of biochemical reactions are the foundations of our models so we decided not to use this tool. Additionally, the user cannot define molecular concentrations in the system and, when a Michaelis-Menten model was simulated and compared with Smoldyn (10,000 molecules for 10 seconds at 1 millisecond time steps), the simulation took twice as long with ChemCell (Andrews et al., 2010).

MCell is also a particle-based simulation tool, similar to ChemCell and Smoldyn, and is one of the most popular tools for modelling biological systems. The main strength of MCell is the superior graphics that it produces, which are important for demonstrating the simulations in publications and presentations. It is primarily used for modelling reactions at the membrane, such as transmission of molecules between synapses in nerve cells. However, as with ChemCell, it is incapable of modelling zero-order reactions and has lower accuracy than Smoldyn when simulating bi-molecular reactions. Similarly, in the Michaelis-Menten test run by Andrews et al., it was around three times slower than Smoldyn.

1.5.4 Smoldyn

We created our spatial stochastic models using Smoldyn, a continuous time and space, particle-based, spatial stochastic cellular modelling too, developed and maintained by Steven S. Andrews (Andrews et al., 2010; Andrews and Bray, 2004). Smoldyn is open source, runs on Mac, Linux and Windows operating systems, and all relevant installation files, documentation and example configuration files can be found at `http://www.smoldyn.org` (Andrews, 2011). The tool takes its name from Smoluchowski Dynamics, a fundamental microscopic kinetic model of collisions between hard spheres moving by a random walk, which can be applied to molecules diffusing and colliding via chemical reactions (Mazo, 2008). Smoldyn has been used to model a variety of biological systems and phenomena to date, such as macromolecular crowding (Marquez-Lago et al., 2012), calcium dependent kinase II in dendritic spines (Khan et al., 2011), and chemotaxis in *E. coli* (Lipkow and Odde, 2008).

Molecules in the system are represented as point particles in a one-dimensional (1D), two-dimensional (2D), or three-dimensional (3D) space. The space is discretised into virtual boxes of equal size and shape. These boxes can be defined by the user, either directly or by defining the maximum number of molecules per box. The size of these boxes only becomes important if the box sizes are smaller than the largest reaction radius, at which point some of the respective bi-molecular reactions may be missed and the reaction rate will therefore be slower than expected. If this was the case, a warning would be flagged at the start of the simulations, giving the user a chance to change the box sizes accordingly.

Molecules can be confined to compartments or surfaces, diffuse, react with other molecules, cross or bounce off surfaces, and behave much like they would in a real system. Reactions between molecules are described via chemical reactions with a specified rate or binding radius. The effective binding radii for each bi-molecular reaction, $A + B \rightarrow C$, are determined using the Smoluchowski formula,

$$k_i = 4\pi(D_A + D_B)\sigma_b \tag{1.51}$$

where k_i is the given reaction rate (taken from the ODE models), D_A and D_B are the diffusion coefficients of species A and B, respectively, and the sum of the molecular radii, σ_b, is the effective binding radius between A and B, within which a reaction between A and B may take place (see

Figure 1.20). The binding radius is called "effective" as a reaction does not always have to occur when two molecules are within this distance, consistent with reality in which factors such as molecular conformations and activation energies also play a part. However, the default reactive probability in Smoldyn is one and they state that a reaction probability less than one has a similar accuracy and results in slower simulations. In our simulations, we use a reactive probability of 0.5 as we are, at this stage, capable of sacrificing some simulation speed for more realistic reactions. In accordance with the binding radius, for reversible reactions, two reactants that unbind are initially separated by an unbinding radius, $\sigma_u > \sigma_b$ so that they cannot re-bind immediately (see Figure 1.21) (Andrews and Bray, 2004).

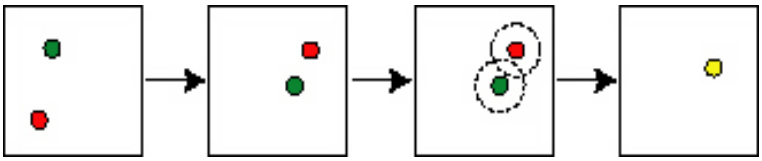


Figure 1.20: A schematic of bi-molecular reactions in Smoldyn.

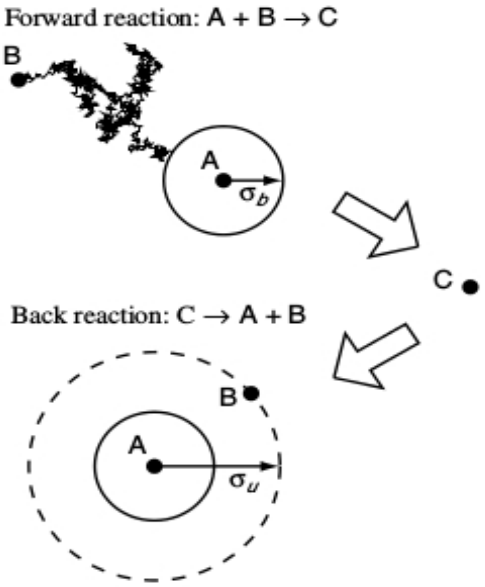


Figure 1.21: A schematic of diffusion, binding and unbinding of molecules (Andrews and Bray, 2004).

In addition to defining a binding radius, sizes can be imposed on point particles by assigning different

diffusion coefficients to molecules of different sizes, using the formula

$$D_i = \frac{3270}{4 \times \sqrt[3]{m_i}} \mu \mathrm{m}^2 \mathrm{s}^{-1}, \tag{1.52}$$

for each molecule, i, of mass m_i Daltons (Da) (Andrews, 2011). This formula derives from the Stokes-Einstein equation, and assumes that diffusion coefficients of molecules moving through eukaryotic cytoplasms and nuclei are one-quarter of their value in water.

In general, bi-molecular reactions between two molecules, A and B, with binding radii r_A and r_B, and diffusion coefficients, D_A and D_B, respectively, have a maximum reaction rate of

$$k_{\max} = 4\pi(D_A + D_B)(r_A + r_B). \tag{1.53}$$

This maximum value is known as the "diffusion-limited" reaction rate and is used for fast bi-molecular reactions with small time steps that have not yet reached steady state. Smoldyn generally defines diffusion-limited reactions as those with an average diffusion step much lower than the binding radius of its reactants. These type of reactions are the foundation of Smoluchowski Dynamics and, therefore, Smoldyn.

Smoldyn also contains a built-in utility program SmolCrowd to model volume exclusion by stationary physical obstacles in the cytoplasm, such as organelles and cytoskeletal filaments (see Figure 1.22). In reality, however, the majority of macromolecular crowding is a result of moving obstacles, so we employ excluded volume reactions with obstacles of size and diffusion rate relative to RNAs. Studies into macromolecular crowding also suggested that simulating moving obstacles rather than stationary obstacles is more important, as adding stationary molecules simply serves to reduce the volume in which molecules can diffuse and react, whereas moving obstacles provides the addition inhibitory role of having a constantly changing and moving excluded geometry. Excluded volume reactions are not implicitly included into Smoldyn, i.e. non-reactive particles may occupy the same position in space, so one must define reactions of the form A + B $\rightarrow$ A + B between all non-reactive species, A and B, in which two such molecules, when separated by less than the sum of their radii, will "bounce" off each other to a distance slightly greater than the sum of their radii, in a similar method to unbinding events between reactive species (see Figure 1.23). We chose RNAs as the

model for obstacles here as anything smaller would require too many additional molecules to model in Smoldyn, considerably increasing the simulation time, and anything larger would have a diffusion coefficient so slow that it could be approximated by a stationary obstacle. Additionally, volume exclusion is thought to have the biggest impact when the reactive molecules are of a similar size to the obstacles (Schnell and Turner, 2004). In our simulations, we look at the differences resulting from the use of stationary or moving obstacles, or a combination of both.

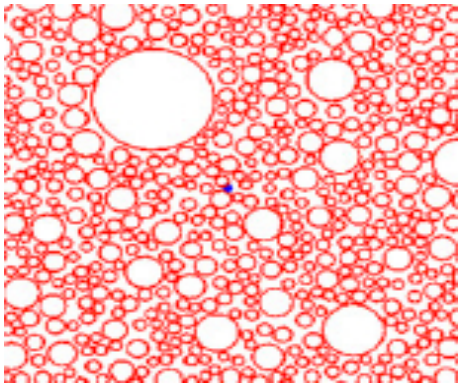


Figure 1.22: A schematic of macromolecular crowding in a 2D simulation (Marquez-Lago et al., 2012).

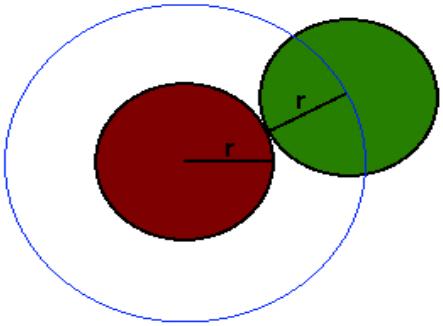


Figure 1.23: A schematic of excluded volume interactions between two molecules of radius r. The outlined blue circle represents the volume excluded region for the red molecule.

As explained in Section 1.5.3, the effects of macromolecular crowding on diffusion and reaction rates of other molecules is primarily as a result of non-specific eletrostatic interactions between the molecules of interest and the obstacles. We model this by changing the volume exclusion reaction rates to allow these two molecular species to bind for certain times, and compare this to 'inert' obstacles which simply bounce off other molecules.

The output of Smoldyn simulations is a text file, in which the user can choose to record, for example, molecule count in each compartment at every n time steps, or the exact times at which a particular reaction occurs. The user can also set conditions under which certain commands will be executed during the simulation. For example, the 'stop' command will allow the user to stop the simulation if a certain molecular species is less than or greater than a pre-defined quantity; the 'kill' command can be used to remove molecules of a certain type from the simulation, if they are undesirable or do not need to be modelled; etc. This is useful as it gives the user some control over the state of the system once the simulation has already begun.

Diffusion-limited reactions further prevent the problem of having two molecules potentially reacting when they are separated by other molecules or obstacles, an obviously undesirable trait in a highly crowded subcellular environment. If reactions are approaching an activation-limited state, reducing the time step is the simplest solution. As a result, we use a short time step in order to keep our simulations as accurate as possible, which does have the undesired effect of causing our simulations to run relatively slowly. If we wished to speed up our simulations, we could sacrifice some accuracy and extend the time step, but since we are mostly interested in the time before molecules reach steady state at this stage, we can maintain our accuracy and ignore long-term dynamics. Alternatively, we can take a hybrid approach, combining the Smoldyn simulations for molecules present in small quantities, with an ODE model for molecular species which are present in large quantities (Klann et al., 2012). We consider this approach for our model of the Wnt signalling pathway but leave this as part of our future work (see Section 5).

Before we expand upon the design of our spatial stochastic models, let us first present the existing models for each biological process that we have studied.

1.5.5 Existing Models of *PTEN* Post-Transcriptional Regulation

Several deterministic models of post-transcriptional regulation have already been created; Most of these models involve modulation of several ceRNAs by a single miRNA species, including transcription, translation, and degradation mechanisms. One of the earliest deterministic models of miRNA−mRNA interactions is a mathematical model published by Khanin and Higham in 2007 (Khanin and Higham, 2007). In their study, and in the expanded computational study published by the pair two years later, they create a system of ordinary differential equations (ODEs) to examine

the effect of a single miRNA on two mRNA species by modelling the fold-change in the target mRNA and protein levels (Khanin and Higham, 2009). They study the long-term steady-state levels of the two target mRNAs and show that there is indeed a cross-talk between the mRNAs via the pool of miRNA. In particular, they note that if one mRNA species is over-expressed due to some unspecified transcriptional control mechanism, then the pool of miRNA will be depleted in the subsequent titration effect of binding to this mRNA, causing up-regulation of all other targeted mRNAs due to limited miRNA-mediated repression. This observation emphasises the sensitivity of the dynamics to both miRNA and mRNA levels, and supports the hypothesis that fine tuning of gene expression via mRNAs will only occur when mRNAs and miRNAs are delicately balanced, most likely when they are present in equal concentration (Ragan et al., 2011).

Building on this, Wang et al. published a two-compartment ODE model of post-transcriptional regulation (Wang et al., 2010). They modelled a single mRNA species regulated by a single miRNA species, including miRNA biogenesis (across the nucleus and cytoplasm), miRNA−mRNA binding, and RNA degradation, using a collation of published parameter values. Wang et al. further compare this ODE model to an SSA model, citing the measurement of limited numbers of certain species of mRNA and miRNA (by Lee et al. (2008)) as the primary justification for stochastic modelling of this process (Lee et al., 2007; Wang et al., 2010).

More comprehensive studies have since been reported and introduce the impact of multiple ceRNAs and miRNA species on target-miRNA regulation in the cytoplasm. One such model, by Ala et al. (2013), includes the effects of multiple MREs (binding sites) and ceRNAs (Ala et al., 2013). In this model, a generalised $M \times N$ system of ODEs, representing M miRNAs interacting with N target mRNAs, is presented. Importantly, Ala et al. discuss the effect of the catalyticity, α, of miRNA−mRNA interations, defined as the probability of both the miRNA and mRNA degrading upon miRNA−mRNA complex degradation. They propose that, in order to observe any crosstalk between competing RNAs and miRNAs, α must be non-zero. For clarity, $\alpha = 0$ corresponds to a purely catalytic interaction between miRNA and mRNA, such that the miRNA simply acts as a catalyst for mRNA degradation and is not itself degraded during the reaction. Conversely, $\alpha = 1$ corresponds to a purely stoichiometric reaction in which mRNA degradation always results in degradation of the bound miRNA (see Figure 1.24) (Ala et al., 2013). This ODE model forms the basis of our deterministic comparison as it can easily be modified for our model of two ceRNAs

crosstalking via two miRNA species.

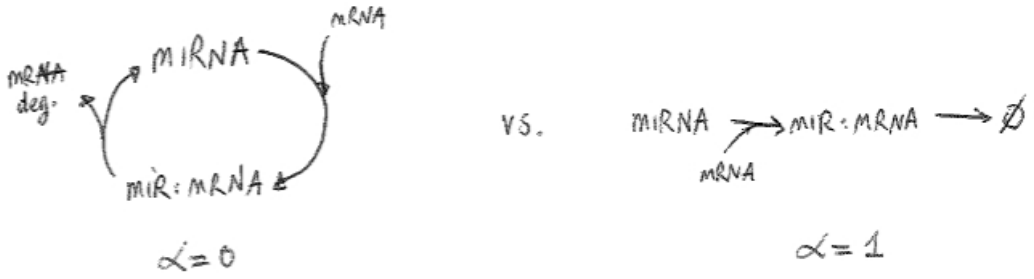


Figure 1.24: A schematic of the miRNA catalyticity.

In the same year, Bosia et al. published a study presenting a generalised SSA model of M miRNAs regulating N mRNAs, presenting the chemical master equation and a series of results showing a variety of simulated ceRNA networks (Bosia et al., 2013). They describe threshold, T, at which, if one ceRNA species is above T, all other ceRNAs are 'driven' above T, a behaviour that is typical of such titration mechanisms. It is proposed that this phenomenon arises if one mRNA species is highly up-regulated, resulting in miRNAs binding preferentially to this more abundant target, allowing other ceRNAs to increase in number as they are targeted by fewer miRNAs. The systems they simulated showed hypersensitivity around such thresholds, resulting in a switch-like behaviour (see Figure 1.7).

The impact of introducing an additional ceRNA species or a single miRNA copycat species has, thus far, not been studied. Nor has the impact of spatial modelling been examined. Both of these have been tackled in this thesis, as well as testing other specific hypotheses such as the effect of macromolecular crowding, miRNA cooperativity, additional MREs, reversible binding, enzymatic versus stoichiometric models, and mRNA repression versus degradation.

In Section 1.5.7, we present our collaborative model of miRNA-mediated regulation of *PTEN* in the presence of one competing miRNA species. First, we present existing models of the second process that we study: RNAi.

1.5.6 Existing Models of RNAi

As far as we are aware, until Holcman's model was published (Holcman et al., 2014), no modelling has been done on RNAi in the nucleus, as this is a fairly new idea. However, the concepts behind the

post-transcriptional regulation and RNAi processes are very similar, so any model of *PTEN* post-transcriptional regulation can also be applied to the RNAi process. The main differences between the post-transcriptional regulatory models when used as a models of RNAi in the nucleus are:

- There is no known sponge modulation in the nucleus.
- The nuclear volume is smaller than the cytoplasm, with a ratio of around 1:3.
- The nucleus has pore complexes on its membrane through which molecules can escape into the cytoplasm.
- The degree of macromolecular crowding is reported to be lower in the nucleus ($\sim 30\%$ (Matsuda et al., 2014)).
- RNAi and post-transcriptional regulation are not synonymous; There are minor differences in the mechanisms of repression and translation, most likely due to differences between siRNAs in the nucleus and miRNAs in the cytoplasm.

An ODE model of RNAi in the cytoplasm of a mammalian cell was pubished several years ago by Cuccato et al., which we use as the basis for our deterministic model of RNAi in the nucleus (Cuccato et al., 2011). As stated above, there are no ceRNAs in the model, however, Cuccato et al. do include the possibility of siRNA cooperativity via multiple MREs. In particular, they describe three biochemical models of RNAi: (1) a stoichiometric model, (2) a stoichiometric model with siRNA cooperativity, and (3) an enzymatic model (Cuccato et al., 2011). In Model (1), the stoichiometric model, only one siRNA must bind to a target binding site in order to degrade both the mRNA and the siRNA, i.e.

$$\text{siRNA} + \text{mRNA} \rightleftharpoons \text{siRNA}{-}\text{mRNA} \rightarrow \emptyset. \tag{1.54}$$

Notice here that siRNA$-$mRNA complex formation is assumed to be reversible. We test the importance of this reversibility in our simulations. In Model (2), the stoichiometric model with cooperativity, the stoichiometric model is adapted to include the requirement that a threshold number of siRNAs must bind to a target mRNA before degradation is possible, i.e.

$$\text{siRNA}_1 + \cdots + \text{siRNA}_T + \text{mRNA} \rightleftharpoons T^*\text{siRNA}{-}\text{mRNA} \rightarrow \emptyset. \tag{1.55}$$

The above notation is a simplification of the process. In reality, this is a series of reversible reactions of the form

$$\text{siRNA} + k^*\text{siRNA}-\text{mRNA} \rightleftharpoons (k+1)^*\text{siRNA}-\text{mRNA}, \tag{1.56}$$

where $k = T - 1$ is the final stoichiometric reaction. Finally, in Model (3), in the enzymatic model, siRNA molecules are assumed to be recycled following the degradation of the target mRNA, thereby acting only as an enzyme for the reaction. In this model, a threshold number of siRNAs must also be reached. However, this threshold, T, can be anywhere in the range $[1, N_b]$, where N_b is the number of siRNA binding sites, or siRNA response elements (SREs). In other words,

$$\text{siRNA} + (T-1)^*\text{siRNA}-\text{mRNA} \rightleftharpoons T^*\text{siRNA}-\text{mRNA} \rightarrow \text{siRNA}_1 + \cdots + \text{siRNA}_T. \tag{1.57}$$

In all of the above models, mRNAs are assumed to be ultimately degraded. For our model of RNAi in the nucleus, this assumption is supported by a lack of P-bodies in the nucleus, in which translationally repressed RNAs might be stored. However, in our cytoplasmic models of post-transcriptional regulation, we study the impact of target repression compared with target degradation.

As far as we are aware, our models will be the first to be used for the computational and mathematical study of RNAi in the nucleus, although the ODE and SSA models themselves will primarily be adaptations of existing models of post-transcriptional regulation and RNAi in the cytoplasm. First, let us present Holcman's model of RNAi (Holcman et al., 2014).

1.5.7 Markov Chain Model of *PTEN* Post-Transcriptional Regulation and RNAi

In a collaboration between Professor David Holcman and Dr Khan Dao Duc (Ecole Normale Supérieure, France), and Professor Helen Byrne, Professor Kevin Burrage and myself (University of Oxford, UK), we created Markov chain and SSA models of PTEN post-transcriptional regulation in the cytoplasm, and RNAi in the nucleus (Holcman et al., 2014). We took a common approach to modelling each of these processes by considering them to consist of two interdependent problems: the mean time to threshold (MTT) and the narrow escape problem (NEP).

1.5.7.1 Narrow Escape Problem

The process of RNAi in the nucleus is a well-described form of the narrow escape problem, whereby a Brownian particle (a diffusing RNA molecule) confined to a domain (the cell nucleus) may escape its domain via small traps (nuclear pores). In general, the narrow escape problem in the nucleus is more clearly defined than in the cytoplasm, as this compartment has definitive exit points in the form of nuclear pores, through which RNA molecules can escape to the cytoplasm. This is in contrast to PTEN 'escape' from the cytoplasm, which we define as binding to a ribosome for subsequent transport to the endoplasmic reticulum (ER).

1.5.7.2 Mean First Passage Time

In general, the mean first passage time (MFPT) is used to estimate the time taken for an miRNA/siRNA to locate a target binding site on an mRNA. For multiple binding sites which must be filled, this extends to the mean time to threshold (MTT). Studying post-transcriptional regulation and RNAi using the MTT and Narrow escape problem (NEP) lends itself to more quantitative measurement of the effect of miRNAs on RNAs, by measuring the proportion of RNAs which can escape from their current subcellular compartment, before degradation can occur.

Post-transcriptional regulation and RNAi are well-defined as Markov processes, as the progression of the system (the number of each type of molecule in the cell following transcription, translation and bi-molecular reactions) depends only on the current state of the intracellular environment e.g. the amount of each molecular species. The state transition diagrams for post-transcriptional regulation in the cytoplasm and RNAi in the nucleus are given in Figure 2.1 and Figure 2.2, respectively. For the most general cases, we cannot solve our CME analytically. Therefore, we proceed primarily using the SSA and give analytical solutions for some reduced versions of the system. We describe the formulation of each CME and the subsequent analysis in Chapters 3 and 2.

1.5.8 Existing Models of the Canonical Wnt Signalling Pathway

Most of the quantitative models of the canonical Wnt pathway build upon the pioneering work of Lee et al., who created the first comprehensive ODE model of the pathway (Lee et al., 2003). They created a coupled system of ODEs that describe the temporal change in concentrations of β-catenin,

TCF, and the three core DC components, Axin, APC and GSK3 (the GSK3β isoform), in response to a Wnt stimulus (via Dsh activation). They include protein synthesis and degradation, as well as GSK-mediated phosphorylation of β-catenin, Axin and APC. The system is assumed to be well-stirred and, as such, forward and backward bi-molecular reaction rates obey mass action kinetics, with the forward rate assumed to be much faster than the backward rate such that complexes are in rapid equilibrium (Schmitz, 2014).

Lee et. al obtained experimental data from *Xenopus* oocyte extracts, primarily using Western blot, and selected the model parameters to match experimental observations. They observed constant levels of Dsh, GSK3β, APC and TCF in their experiments. In addition they performed a parameter sensitivity analysis which showed that the values chosen were robust to variation. The primary contribution of their model to the study of the Wnt/β-catenin pathway, aside from the number of studies that it motivated, was in the identification of a new role for APC and Axin in the degradation of β-catenin, independent of the DC. In particular, they found that when APC levels are low, Axin assumes the role of primary β-catenin regulator, and vice versa. The primary drawback of this model is in the use of data from *Xenopus* oocyte cells for parameterisation; Wnt signalling my not behave in exactly the same way in *Xenopus* cells as mammalian cells, and the concentrations of molecular species can be quite different (see work by Tan et al. described later in this chapter), making it unsuitable for direct application to the study of mammalian cellular systems.

One such extension to the Lee et al. model is the "compartmentalised" ODE model by van Leeuwen et al., which incorporates the role of β-catenin in three subcellular compartments - the cytoplasm, the nucleus, and the cell membrane (cell-cell adhesion) - although they still assume that all reactions take place within a single cellular compartment within their model (van Leeuwen et al., 2007). They model the effect on the system in two ways. In hypothesis 1 (H1), E-cadherin simply competes with the destruction complex (DC) and transcription factors (TFs) for β-catenin binding. In hypothesis 2 (H2), Wnt signalling results in the formation of a "closed-form" of β-catenin, which is unable to form adhesion complexes, and, therefore, binds preferentially to TFs in the nucleus. Their main finding was that these two hypotheses cannot be distinguished by measuring gene expression alone, but that changes in E-cadherin levels following Wnt signal induction must also be measured. Specifically, reduction in E-cadherin levels following Wnt stimulus can only be replicated under H2. We aim to use our model to test this hypothesis.

Another more recent model is an ODE system presented by Schmitz et al., which explicitly models cellular compartments and intercompartmental shuttling of molecules between the cytoplasm and the nucleus (Schmitz et al., 2013). This compartmentalized model incorporates nucleo-cytoplasmic shuttling of β-catenin and the primary components of the DC. They first present a model called "The Retention Model" which assumes that DC components are capable of retaining β-catenin within a given intracellular compartment, thereby promoting or inhibiting the formation of β-catenin/TCF complexes depending on whether β-catenin is retained in the nucleus or cytoplasm, respectively. The main finding of this paper was that the distribution of molecules between compartments has a major impact on the regulation of β-catenin and Wnt target gene transcription. More specifically, the DC has limited influence on β-catenin when the two molecular species are separated by compartments (i.e. by the nuclear membrane). This supports our hypothesis that spatial modelling of the canonical Wnt pathway is important, as the compartmental localisation of reacting molecules in the pathway determines the rate of target gene transcription and subsequent signal transduction. They expand on this model to include potential degradation of β-catenin by the DC in both the nucleus and cytoplasm, and refer to this model as "The Degradation Model" (Schmitz, 2014). In our study, we combine the models by Schmitz et al. and van Leeuwen et al., to create a compartmentalised model of the canonical Wnt pathway with cell-cell adhesion.

One of the most recent models, by MacLean et al., is presented as a parameter-free model in an attempt to circumvent the errors that result from differing parameter values measured from *in vivo*, "dilute" *in vitro*, and "crowded" *in vitro* experiments (MacLean et al., 2015). They create a novel "shuttle" model that explores the emergence of bistable switching via intercompartmental shuttling and degradation of proteins. They further use multistability (injectivity) analysis to demonstrate that this type of model is the only published model capable of emulating the multiple steady states observed experimentally. In more detail, MacLean et al. study the effects of the intracellular localisation of the core DC components, as well as β-catenin and Dsh, and a more comprehensive mechanism for DC/β-catenin interaction, whereby active DC can degrade β-catenin upon binding, whereas inactive DC can merely sequester β-catenin without affecting its stability. Their mathematical analysis is beyond the scope of this thesis, however, they do summarise a number of conditions imposed on model parameter values that are necessary to observe multistationarity, which we discuss later in addition to the aforementioned shuttling and degradation model components.

There have been other extensions to the Lee et al. model, which we do not build upon in our thesis, but some of which are worth discussing for a full picture of the available models. One such model, by Cho et al., included a negative feedback loop in the form of Axin synthesis by β-catenin/TCF-mediated transcription of the *Axin2* gene in the nucleus (Cho et al., 2006). They perform steady state analysis on their model, examining the effects of APC mutations on β-catenin levels. They determine that the optimal level of β-catenin occurs when the increase in β-catenin levels due to Wnt activation balances with the decrease in β-catenin levels due to increased Axin transcription.

A similar model published by Wawra et al. the following year included an additional negative feedback loop, involving the Dsh inhibitor, Dickkopf-1 (Dkk-1) (Wawra et al., 2007). In their delay-differential equation model, they exhibited the existence of oscillatory behaviour with the addition of such feedback loops, matching some experimentally observed oscillations, for example, during embryonic development, which is unobserved in models without such loops or delays. However, it as been noted in other reviews of this paper that the concentrations of Axin and β-catenin needed to observe such oscillations were "unrealistically high", although it is a step closer to a Wnt model capable of matching observed oscillations in β-catenin in certain experiments (Schmitz, 2014). In our study, we also aim to explore the required conditions for observing oscillations of β-catenin in the system, which we discuss in Chapter 4.

The first stochastic model of the Wnt/β-catenin signalling pathway was created by Mazemondet et al., which also dealt with compartmentalisations of molecules within the cytoplasm and nucleus and cell cycle asynchrony (Mazemondet et al., 2012). They demonstrated the stochasticity that results from the low levels of Axin measured in the Xenopus egg extracts presented in the Lee et al. model (Lee et al., 2003). They state that these effects, which can not be observed in the deterministic models, contradict experimentally observed data, and suggest that this might be because Axin levels are higher than expected, as per mammalian measurements (Tan et al., 2012), or that an important mechanism might have been overlooked, such as a feedback loop (Wawra et al., 2007). They provide alternative (increased) molecular concentration values that reduce the amount of stochasticity observed *in silico* such that it more closely matches values observed *in vitro*. They further demonstrate that the impact of the cell cycle, at least for their chosen cell line, is negligible. It is also shown that cells exhibit biphasic peaks in β-catenin expression, and hypothesise that the initial peak is the result of Wnt activation, and that the second peak is due to "self-induced" Wnt

signalling, i.e. without an external stimulus, that may be a result of growth factor removal that occurs when cells begin differentiation/specialisation.

Another model, by Kogan et al., focuses more on receptor binding and antagonsim by sFRP and Dkk-1, and assumes that DC concentration is in equilibrium (Kogan et al., 2012). They obtain data from mouse fibroblasts which do not contain cadherins, negating the need to model β-catenin's dual role. For a comprehensive review of other models of the Wnt/β-catenin pathway, we recommend the recent review by Lloyd-Lewis et al. (Lloyd-Lewis et al., 2013).

In each of the aforementioned models, the output is measured as the changes in absolute levels of β-catenin. However, there are other models that support the measurement of fold-changes of both free and complexed β-catenin. This was first proposed by Goentoro et al. who showed, via simulation and experimentation in mammalian cells, that the system is more robust to small perturbations when fold changes are measured rather than absolute levels (Goentoro and Kirschner, 2009). Since perturbations are normal for inherently noisy biological systems, we adopt this recommendation and measure fold-changes in β-catenin/TCF when perturbing our system, i.e. we still measure absolute values over time, but use fold changes for comparing Wnt "on" and "off" (e.g. Schmitz, 2014).

Our plan for the Wnt model was to build upon existing models and examine the effects of a higher degree of spatial and stochastic detail; Building a spatial stochastic version of Schmitz et al.'s ODE model, then modelling the effects of macromolecular crowding and volume exclusion.

Chapter 2

RNA Interference in the Nucleus

2.1 Introduction

Until recently, RNA interference (RNAi) was thought to be confined to the cytoplasm. However, reports are emerging that RNAi mechanisms are active in the nucleus, with some differences to RNAi in the cytoplasm. As a result, nuclear RNAi represents a new avenue for potential therapeutic discovery and development, for which modelling is an important tool, particularly where data is limited. In this chapter, we focus on the RNAi process of primary-miRNA (pri-miRNA) regulation by miRNAs (Tang et al., 2011). Pri-miRNAs diffusing in the nucleus are targeted by miRNAs, but may escape degradation by exiting the nucleus via pores in the nuclear membrane. If pri-miRNAs were able to exit the nucleus, they would be capable of forming mature miRNAs and begin binding to their target mRNAs in the cytoplasm (Tang et al., 2011). This latter cytoplasmic component is similar to the *PTEN* regulatory model outlined in Section 1.2 and, as such, RNAi in the nucleus could be considered to be a form of upstream post-transcriptional regulation to this pathway. As part of future work, these two processes could be combined to a larger miRNA-ceRNA network occurring across the entire cell.

Our model of pri-miRNA regulation can also be applied to the regulation of 7SK small nuclear RNA (snRNA) by siRNAs in the nucleus, an alternative RNAi mechanism proposed by Robb et al. (Robb et al., 2005), via generalisation of this system to the study of RNAs escaping via a nuclear pore complexes before degradation by siRNAs (Schuss et al., 2007). In either case, understanding the conditions under which RNAs may avoid degradation by siRNAs could provide valuable information on how to maximise the effectiveness of siRNAs designed for therapeutic treatment.

In the remainder of this chapter, we present a Markov model of nuclear RNAi (Holcman et al., 2014) - formulated as a narrow escape problem (NEP) and a mean time to threshold (MTT) - along with our expansion of the analytical solutions already obtained. We go on to describe deterministic, stochastic, and spatial stochastic models of the process for comparison. We believe that spatial modelling will provide us with valuable insights into the possible sources of stochasticity and heterogeneity in the system, and the importance of spatial detail on estimating rates of pri-miRNA degradation or escape. Before we go into detail of the modelling approaches, let us first explain what we mean by the narrow escape problem and mean time to threshold, which we discuss in reference to the Markov model and in particular (Holcman et al., 2014).

2.2 RNA Escape from the Nucleus

The nucleus of mammalian cells has around 3000 nuclear pore complexes distributed across the membrane surface, through which proteins and RNA molecules can be transported (actively or passively) between the nuclear and cytoplasmic compartments (Alberts et al., 2008). In particular, RNA molecules are usually transported from the nucleus to the cytoplasm following transcription in order to locate the translational machinery necessary for protein production. For our model of RNAi, we primarily consider miRNA-mediated regulation of pri-miRNAs before they have formed mature miRNAs. In the absence of any RNAi, pri-miRNAs would diffuse throughout the nucleus, subsequently entering the cytoplasm via nuclear pore complexes to continue their biogenesis to mature miRNAs. Specifically, with the addition of miR-709, some molecules of pri-miR-15a/16-1 will be directly bound by an miR-709 molecule, preventing nuclear export and halting pri-miRNA maturation. It is unclear exactly how this repression occurs, i.e. whether pri-miRNAs are degraded in the nucleus or simply prevented from entering the cytoplasm to complete maturation. In our models, we do not consider the mechanisms following miR-709−pri-miRNA binding or pri-miRNA escape, and leave elucidation of the exact repression mechanism for future work (Tang et al., 2011).

The process of being targeted for degradation by miRNAs can be modelled using the mean time to threshold (MTT), whereby a threshold number of miR-709 must bind to a pri-miRNA for degradation/retention to occur. Similarly, escape of pri-miRNAs from the nucleus via small holes in the nuclear membrane can be modelled as a narrow escape problem (NEP) (Holcman et al., 2014). We

do this via a Markov model, in which RNAs have a certain probability of reaching a nuclear pore, thereby escaping degradation, or else binding to/unbinding from an siRNA.

2.3 Markov Model

In (Holcman et al., 2014), we collaborated with Holcman et al. to present a Markov model of RNAi in the nucleus and post-transcriptional regulation in the cytoplasm (presented in Chapter 3). The model was created by Holcman et al. but we made several primary contributions to the publication. We created our own SSA model of the system in order to run simulations and create Figure 1 and Figure 8 in the paper. We provided the solution to the matrix system described in Section 2.2.2 and Figure 8 of the paper, showing escape probabilities as a function of the threshold, T. We provided the background biological study in Section 1.1 and 1.2 of the paper, used as motivation for both the PTEN and RNAi models to which the Markov model was applied. We replicated and validated Professor Holcman's results using our own models, and made several corrections to the mathematical formulae were necessary (exact corrections not known). We also made some of the final conclusions drawn from the study, in particular surrounding Figure 8 and 9 of the paper. Finally, we made some of the propositions for future work, in particular regarding the recommendation of a spatial model of the RNAi system which we present in this thesis. We furthermore present an extension to this publication with an explicit solution to the case of reversible binding in Section 2.4.1 of this thesis.

As part of its application to the RNAi system, the model from (Holcman et al., 2014) was utilised to calculate the probability, P_a, of an mRNA escaping from the nucleus (via a nuclear pore) before it is degraded by siRNAs (analogous to pri-miR-15a/16-1 escaping degradation by miR-709). We consider degradation to occur if a threshold number, T, of siRNAs are able to bind to an mRNA before it is able to escape from the nucleus.

RNAs in the nucleus are assumed to diffuse via Brownian motion, modelled according to the formula for the motion of a particle, $\mathbf{x}(t)$, in a confined domain, Ω, i.e.

$$d\mathbf{x} = \sqrt{2}\mathbf{B}(x)d\mathbf{w}(t), \tag{2.1}$$

where $\mathbf{B}(x)$ is a diffusion tensor for each RNA, and $\mathbf{w}(t)$ is a vector of independent Wiener processes.

As well as having a particular position in space, each mRNA can be described as being in a particular state, s, representing the number of siRNAs that are bound. Hence, each mRNA can be described by the probability density function (pdf),

$$\text{Prob}\{\mathbf{x}(t) \in \mathbf{x} + d\mathbf{x}, s(t) = s\} = p_s(\mathbf{x}, t)d\mathbf{x}, \tag{2.2}$$

which is the probability that an mRNA is located at position $\mathbf{x}$ at time t with s siRNAs bound.

To represent holes in the domain through which pri-miRNAs may escape, i.e. nuclear pores, we split the domain Ω into two parts, Ω_a and Ω_r such that $\Omega = \Omega_a + \Omega_r$. Here, Ω_a represents the absorbing sections of the domain boundary (N nuclear pores, ϵ_j for $j = 1 \dots N$), and is given by the sum over all of the pores, $\sum_{j=1}^{N} \Omega_{\epsilon_j}$. Whereas Ω_r represents the reflective sections of the domain boundary (the nuclear membrane). Therefore, the narrow escape problem (NEP) of mRNA escape from the nucleus can be formulated as a boundary value problem (BVP) with the following mixed Dirichlet-Neumann boundary conditions (Cheviakov et al., 2010),

$$p_s(\mathbf{x}, t) = 0 \qquad \text{for } \mathbf{x} \in \Omega_a, \tag{2.3}$$

$$\frac{\partial \mathbf{B}\mathbf{B}^T(x) p_s(\mathbf{x}, t)}{\partial \mathbf{n}(\mathbf{x})} = 0 \qquad \text{for } \mathbf{x} \in \Omega_r, \tag{2.4}$$

where p_s represents an mRNA with s miRNAs bound to it. Equation (2.3) represents the zero-probability of finding an intact mRNA at one of the nuclear pores, as it is assumed that these molecules will escape from the domain with probability one (almost surely). Equation (2.4) is a zero-flux boundary condition, demonstrating the reflective quality of the boundary around the pores, given by Ω_r. These equations are identical to Equations 3 and 4 in the Holcman et al. paper where $d = 0$ (d in the paper represents another molecular species called copycat miRNAs, which are introduced in Chapter 3) (Holcman et al., 2014).

As mentioned in Section 1.5.2.2, the Chemical Master equation (CME) describes the continuous-time evolution of a Markov chain. However, for a Markov Chain in time and space, we need a partial differential form of the CME. For this, we formulate the Fokker-Planck equation, a second-order

partial differential equation (PDE) given by

$$\begin{aligned}\frac{\partial}{\partial t}p_s(\mathbf{x},t) = \mathbf{B}\mathbf{B}^T(\mathbf{x})\nabla p_s(\mathbf{x},t) - k_f(N_b - s)p_s(\mathbf{x},t) \\ - k_b p_s(\mathbf{x},t) + k_b(s+1)p_{s+1}(\mathbf{x},t) \\ + k_f(Nb - s + 1)p_{s-1}(\mathbf{x},t),\end{aligned} \tag{2.5}$$

where $0 \leq s < T$, and k_f and k_b are the forward and backward binding rates, respectively, of siRNAs to mRNAs. Each term in the equations represents a possible transition to the state s in a single time step. A summary of possible transition probabilities can be demonstrated by the transition matrix,

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,T} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ p_{T,1} & p_{T,2} & \cdots & p_{T,T} \end{pmatrix}, \tag{2.6}$$

where $p_{i,j}$ is the probability of an $i \to j$ transition, i.e. the probability of j siRNAs being bound at time $t + \delta t$ given i siRNAs bound at time t, and the rows sum to one. This is summarised by the one-dimensional transition diagram shown in Figure 2.1.

$$0 \rightleftharpoons 1 \rightleftharpoons \cdots \rightleftharpoons i-1 \rightleftharpoons i \rightleftharpoons i+1 \rightleftharpoons \cdots \to T$$

Figure 2.1: The one-dimensional transition diagram for siRNA binding to mRNAs. The final state, T, is absorbing as a result of mRNA degradation, hence, transitions to this state are irreversible.

We assumed here that binding of siRNAs is reversible, i.e. $k_b \neq 0$. In this model, siRNAs can only bind to mRNAs one at a time, resulting in a tridiagonal transition matrix, P,

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & 0 & \cdots & 0 \\ p_{2,1} & p_{2,2} & p_{2,3} & \ddots & \vdots \\ 0 & p_{3,2} & p_{3,3} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & p_{T-1,T} \\ 0 & \cdots & 0 & p_{T,T-1} & p_{T,T} \end{pmatrix}, \tag{2.7}$$

To remove the dependence of the Fokker-Planck equation on space (thereby converting it into a CME), we can simplify the model in order to study the "survival probability" of an mRNA, defined as the probability of the mRNA being found intact inside the domain at time t. The probability of an mRNA being found inside the domain Ω in position x at time t is,

$$p_s^S(t) = \int_\Omega p_s(\mathbf{x}, t) d\mathbf{x} \tag{2.8}$$

.

Using this definition, we can ignore the spatial dimension and instead model the probability of an mRNA locating a nuclear pore. This is given by the outflux probability in the small hole limit,

$$J_s(t) = \oint_{\Omega_a} \frac{\partial \mathbf{B}\mathbf{B}^T(x) p_s(\mathbf{x}, t)}{\partial \mathbf{n}(\mathbf{x})} d\mathbf{x}. \tag{2.9}$$

Using the fact that reaching a nuclear pore is rare, since they occupy only a small fraction of the domain, the flux can be approximated in the small hole limit

$$J_s(t) = \frac{1}{\tau} p_s^S(t), \tag{2.10}$$

where $\frac{1}{\tau}$ is the mean arrival rate of an mRNA to a small nuclear pore, and the superscript S indicates the "survival" probability of the mRNA. The time to reach a nuclear pore, τ, is approximated by the time it takes an RNA molecule to diffuse (with diffusion coefficient, D) from the centre of the nucleus (approximated by a sphere of radius, R) to any of N small pores (of radius, ϵ) on the membrane. A similar formula is used to calculate the probability of siRNAs reaching any of N small targets on an RNA molecule. Substituting this back into Equation (2.5), we can ignore the mRNA position vector $\mathbf{x}$ and arrive at the CME,

$$\begin{aligned} \frac{d}{dt} p_s^S(t) = &-\frac{1}{\tau} p_s^S(t) - k_f(N_b - s) p_s^S(t) - k_b p_s^S(t) + k_b(s+1) p_{s+1}^S(t) \\ &+ k_f(N_b - s + 1) p_{s-1}^S(t). \end{aligned} \tag{2.11}$$

This is the same as the equation in Section 2.2 of (Holcman et al., 2014) where $d = 0$. This is why there is no "2" in front of the k_f term, as there is only one forward binding reaction of miRNAs; In

Holcman et al., there is also the forward binding reaction of "copycats" (introduced in the following section) which have the same rate k_f. To find a solution to Equation 2.11, we integrate over all time, $t \in [0, \infty)$, and the entire domain, Ω. We define the probability of mRNA escape through a nuclear pore before degradation as

$$P_a = \frac{1}{\tau} \sum_{k=0}^{T-1} a_s, \tag{2.12}$$

where a_s is the probability of an mRNA, with s siRNAs bound, being found intact within the domain at any time $t \in [0, \infty)$, i.e.

$$a_s = \int_0^\infty p_s^S(t)dt. \tag{2.13}$$

Substituting a_s into Equation (2.11) gives,

$$0 = -\frac{1}{\tau}a_s - k_f(N_b - s)a_s - (k_b)a_s + k_b(s+1)a_{s+1} + k_f(Nb - s + 1)a_{s-1}, \tag{2.14}$$

where the left hand side vanishes because

$$p_s^S(0) = 0, \qquad \forall s \neq 0, \tag{2.15}$$

$$\lim_{t\to\infty} p_s^S = 0, \qquad \forall s \neq T. \tag{2.16}$$

Certain boundary conditions result from Equation (2.14), namely

$$-1 = -\frac{1}{\tau}a_0 - k_f(N_b)a_0 + k_b(1)a_1, \tag{2.17}$$

$$0 = -\frac{1}{\tau}a_{T-1} - k_f(N_b - T + 1)a_{T-1} + k_f(Nb - T + 2)a_{T-2}, \tag{2.18}$$

$$p_T^S(\infty) = k_f(Nb - T + 1)a_{T-1}, \tag{2.19}$$

showing transitions to states $s = 0$, $s = T - 1$ and $s = T$, respectively, and using the fact that

$$p_0^S(\infty) = 0 \quad \text{and} \quad p_0^S(0) = 1. \tag{2.20}$$

For very long times ($t \to \infty$), we know that an mRNA will either have escaped, with probability

P_a, or have been degraded, with probability p_T^S. Thus,

$$P_a + p_T^S(\infty) = 1. \tag{2.21}$$

From (2.21), the CME, and the boundary conditions, we determine that the probability that an mRNA escapes the nucleus intact is given by

$$P_a(T, N_b, x) = \sum_{s=0}^{T-1} x^s \frac{N_b!}{(N_b - s)!} \prod_{j=0}^{s} \frac{1}{1 + (N_b - j)x}, \tag{2.22}$$

where N_b is the total number of binding sites on each mRNA. Here, $x = \tau k_f$, where τ is the mean arrival time of an mRNA to a nuclear pore, and k_f is the forward binding rate of the siRNA to the site.

2.4 Markov model with copycats

Our Markov model was also used to study the two-dimensional system of RNAi regulated by miRNAs and copycats. To avoid confusion, we continue to refer to copycats as such, although they are referred to as "decoys" in the publication (Holcman et al., 2014). As with the one-dimensional Markov model, each pri-miRNA can be described as being in a particular state, $\{s, d\}$, representing the number of miRNAs and copycats that are bound, respectively. Hence, each pri-miRNA can be described by the joint probability density function (jpdf),

$$\text{Prob}\{\mathbf{x}(t) \in \mathbf{x} + d\mathbf{x}, s(t) = s, d(t) = d\} = p_{s,d}(\mathbf{x}, t)d\mathbf{x}, \tag{2.23}$$

which is the probability that a pri-miRNA is located at position $\mathbf{x}$ in the domain at time t with s miRNAs and d copycats bound. Therefore, the narrow escape problem (NEP) can be re-formulated as

$$p_{s,d}(\mathbf{x}, t) = 0 \qquad \text{for } \mathbf{x} \in \Omega_a, \tag{2.24}$$

$$\frac{\partial \mathbf{B}\mathbf{B}^T(x) p_{s,d}(\mathbf{x}, t)}{\partial \mathbf{n}(\mathbf{x})} = 0 \qquad \text{for } \mathbf{x} \in \Omega_r. \tag{2.25}$$

Using the state transition graph in Figure 2.2, we get the CME,

$$\begin{aligned}\frac{d}{dt}p^S_{s,d}(t) = &-\frac{1}{\tau}p^S_{s,d}(t) - 2k_f(N_b - s - d)p^S_{s,d} - (k^s_b + k^d_b)p^S_{s,d} \\ &+ k^s_b(s+1)p^S_{s+1,d}(t) + k^d_b(d+1)p^S_{s,d+1}(t) \\ &+ k_f(Nb - s - d + 1)p^S_{s-1,d}(t) + k_f(Nb - s - d + 1)p^S_{s,d-1}(t).\end{aligned} \tag{2.26}$$

where $0 \leq s < T$; k_f is the forward binding rate for both miRNAs and copycats; and k^s_b and k^d_b are the backward binding rates of miRNAs and copycats, respectively. Here we used the same methods applied in the case of $d = 0$, but we repeat them here for clarity. The probability of an mRNA with s miRNAs and d copycats bound ($p_{s,d}$) being found inside the domain Ω in position x at time t is,

$$p^S_{s,d}(t) = \int_\Omega p_{s,d}(\mathbf{x}, t)d\mathbf{x} \tag{2.27}$$

.

The outflux probability in the small hole limit is now given by,

$$J_{s,d}(t) = \oint_{\Omega_a} \frac{\partial \mathbf{B}\mathbf{B}^T(x)p_{s,d}(\mathbf{x}, t)}{\partial \mathbf{n}(\mathbf{x})}d\mathbf{x}. \tag{2.28}$$

Using the fact that reaching a nuclear pore is rare, since they occupy only a small fraction of the domain, the flux can be approximated in the small hole limit

$$J_{s,d}(t) = \frac{1}{\tau}p^S_{s,d}(t), \tag{2.29}$$

where $\frac{1}{\tau}$ is the mean arrival rate of an mRNA to a small nuclear pore,

Each term in the equations represents a possible transition from the current state of the system to the state $\{s, d\}$ in a single time step. A summary of possible transitions is shown in the two-dimensional transition diagram shown in Figure 2.2. We assumed here that binding of miRNAs and copycats is reversible, i.e. $k_b \neq 0$.

To find a solution to this equation, we integrate over all time, $t \in [0, \infty)$, and the entire domain, Ω.

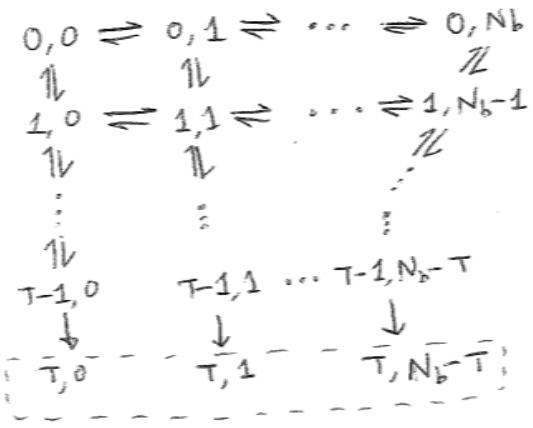


Figure 2.2: The two-dimensional transition diagram for miRNA and copycat binding to pri-miRNAs. The final state, $\{T, k\}$, for $k \in [1, \ldots, N_b - T]$, is absorbing as a result of pri-miRNA degradation/retention, hence, transitions to this state are irreversible.

We define the probability of pri-miRNA escape before degradation as

$$P_a = \frac{1}{\tau} \sum_{s=0}^{T-1} \sum_{d=0}^{N_b - T} a_{s,d} + \frac{1}{\tau} \sum_{d=N_b-T+1}^{N_b} \sum_{s=0}^{N_b-d} a_{s,d}, \tag{2.30}$$

where $a_{s,d}$ is the probability of an mRNA, with s miRNAs and d copycats bound, being found intact within the domain at any time $t \in [0, \infty)$, i.e.

$$a_{s,d} = \int_0^\infty p_{s,d}^S(t) dt. \tag{2.31}$$

Substituting $a_{s,d}$ into Equation (2.26) gives,

$$\begin{aligned} 0 = &-\frac{1}{\tau} a_{s,d} - 2k_f(N_b - s - d) a_{s,d} - (k_b^s s + k_b^d d) a_{s,d} \\ &+ k_b^s (s+1) a_{s+1,d} + k_b^d (d+1) a_{s,d+1} \\ &+ k_f (Nb - s - d + 1) a_{s-1,d} + k_f (N_b - s - d + 1) a_{s,d-1}, \end{aligned} \tag{2.32}$$

where the left hand side vanishes because

$$p_{s,d}^S(0) = 0, \qquad \forall s \neq 0 \qquad \text{and} \tag{2.33}$$

$$\lim_{t \to \infty} p_{s,d}^S = 0, \qquad \forall s \neq T. \tag{2.34}$$

Certain boundary conditions result from Equation (2.32). For $s = T - 1$ and $N_b - T > d > 0$,

$$0 = -\frac{1}{\tau}a_{T-1,d} - 2k_f(N_b - T + 1 - d)a_{T-1,d} - (k_b^s(T-1) + k_b^d d)a_{T-1,d} + k_b^d(d+1)a_{T-1,d+1} + k_f(Nb - T - d + 2)a_{T-2,d} + k_f(N_b - T - d + 2)a_{T-1,d-1}. \quad (2.35)$$

For $T - 1 > s > 0$ and $d = 0$,

$$0 = -\frac{1}{\tau}a_{s,0} - 2k_f(N_b - s)a_{s,0} - k_b^s s a_{s,0} + k_b^s(s+1)a_{s+1,0} + k_b^d a_{s,1} + k_f(Nb - s + 1)a_{s-1,0}. \quad (2.36)$$

For $T > s > 0$ and $d = N_b - s$,

$$0 = -\frac{1}{\tau}a_{s,N_b-s} - (k_b^s + (N_b - s)k_B^d)a_{s,N_b-s} + k_f a_{s-1,N_b-s} + k_f a_{s,N_b-s-1}. \quad (2.37)$$

For $s = 0$ and $Nb > d > 0$,

$$0 = -\frac{1}{\tau}a_{0,d} - 2k_f(N_b - d)a_{0,d} + k_b^d d a_{0,d} + k_b^s a_{1,d} + k_b^d(d+1)a_{0,d+1} + k_f(N_b - d + 1)a_{0,d-1}. \quad (2.38)$$

For $s = 0$ and $d = 0$,

$$-1 = -\frac{1}{\tau}a_{0,0} - 2k_f N_b a_{0,0} + k_b^s a_{1,0} + k_b^d a_{0,1}. \quad (2.39)$$

Finally, for $s = 0$ and $d = N_b$,

$$0 = -\frac{1}{\tau}a_{0,N_b} - k_b^d N_b a_{0,N_b} + k_f a_{0,N_b-1}, \quad (2.40)$$

using the fact that

$$p_{s,d}^S(\infty) = 0 \quad \text{and} \quad p_{0,0}^S(0) = 1. \quad (2.41)$$

For very long times ($t \to \infty$), we know that a pri-miRNA will either have escaped, with probability

P_a, or have been degraded, with probability $p^S_{T,d}$, thus

$$P_a + p^S_{T,d}(\infty) = 1, \tag{2.42}$$

for $0 \leq d \leq N_b - T$.

In the case of irreversible binding, where $k^s_b = k^d_b = 0$, we can determine the probability that a pri-miRNA is degraded before it can escape. As in (Holcman et al., 2014), we compute and sum all possible states that lead to degradation. Using the CME and the boundary conditions, the probability of degradation before escape is given by

$$P_d = \sum_{k=0}^{N_b} \binom{k-1}{T-1} \left(\frac{1}{2}\right)^k \prod_{i=0}^{k-1} \frac{k_f(N_b - i)}{k_f(N_b - i) + 2/\tau}, \tag{2.43}$$

where N_b is the total number of binding sites on each pri-miRNA. Here, $x = \tau k_f$, where τ is the mean arrival time of a pri-miRNA to a nuclear pore, and k_f is the forward binding rate of the miRNA to a pri-miRNA.

In Holcman et. al, they showed the escape probability as a function of the threshold, T, comparing both the reversible and irreversible case (we did not contribute directly to this figure but it is an important finding). Using $N_b = 20$ and $k_f = 1$, they showed that reversible binding decreases the probability of escape when copycats are used. Particular attention was drawn to the case where the threshold $T = 11$, in which a two-fold decrease in the probability of escape was observed for the reversible case (probability of 0.3 compared with 0.6). We concluded that this is likely due to the unbinding of copycats, increasing the likelihood of the T being reached as miRNAs have another chance to bind (see Figure 2.3).

2.4.1 Extension to One-Dimensional Markov Model

In (Holcman et al., 2014), we state that a simple, explicit solution for the reversible binding case is not possible to obtain. However, we show that a solution is attainable, although the algebra involved is messy. Our solution is as folows:

Rearranging Equation (2.14) to give a_{s+1} in terms of a_s and a_{s-1}, with $s \in [1, T-1]$, we can use induction from the boundary conditions to can obtain a general formula of the kind given in the

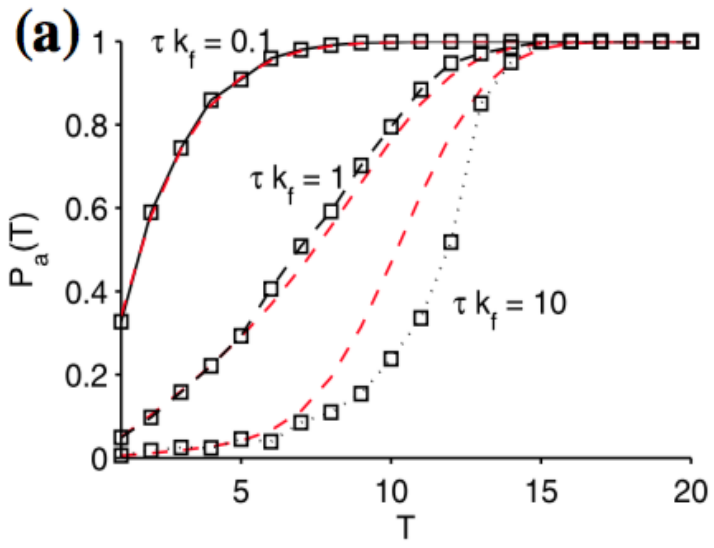


Figure 2.3: Figure taken from Holcman et. al ((Holcman et al., 2014)). The escape probability as a function of the threshold, T. Reversible binding (black, squares) compared to the irreversible binding (red, dashed). Total binding sites$N_b = 20$ and forward binding rate $k_f = 1$. Reversible binding appears to decrease the probability of escape when copycats are used, particularly for $T = 11$.

irreversible case (Equation (2.22)). We use the same boundary conditions from Equations (2.17) and (2.18), and use the assumption that all RNAs are in state a_0 initially (i.e. no siRNAs bound). We are further able to test the accuracy of our solution by setting the unbinding rate, k_b, equal to zero and comparing our solution to the irreversible solution.

We can re-write the general solution (2.14) as

$$a_{s+1} = \left[\frac{1}{\tau} - k_f(N_b - s)a_s - k_b s\right] a_s - \frac{k_f}{k_b(s+1)}(N_b - s + 1)a_{s-1} \tag{2.44}$$

$$= \theta_s a_s + \phi_s a_{s-1}, \tag{2.45}$$

where,

$$\theta_s = \frac{1}{\tau} - k_f(N_b - s)a_s - k_b s \quad \text{and } \phi_s = -\frac{k_f}{k_b(s+1)}(N_b - s + 1). \tag{2.46}$$

Similarly, from (2.17), we can write

$$a_1 = \frac{1}{k_b}\left[\frac{1}{\tau} + k_f N_b\right] a_0 - \frac{1}{k_b} \tag{2.47}$$

$$= \theta_0 a_0 + \phi_0 a_{-1}. \tag{2.48}$$

Combining these two formulae, we can calculate every subsequent a_n in terms of a_0 and $\phi_0 a_{-1}$. For example,

$$a_2 = \theta_1 a_1 + \phi_1 a_0 \tag{2.49}$$

$$= (\theta_1\theta_0 + \phi_0)a_0 + \theta_1\phi_0 a_{-1}. \tag{2.50}$$

We use $\phi_0 a_{-1}$ here as a_{-1} is not defined on its own. That is $\phi_0 a_{-1} = \dfrac{-1}{k_b}$. The general solution is given by

$$a_{n+1} = \left(\sum_{k=0}^{\lfloor \frac{n+1}{2} \rfloor} \sum_{\#\phi=k} F_n\right) a_0 + \left(\phi_0 \sum_{k=0}^{\lceil \frac{n+1}{2} \rceil} \sum_{\#\phi=k} G_n\right) a_{-1}, \tag{2.51}$$

where

- $\#\phi$ is the number of ϕ_i's in F_n and G_n;
- F_n is the product of ϕ_i's and θ_j's for all combinations of ϕ_i $(i = 1, \ldots, n)$, such that $\theta_{i-1} = 1$ for $i - 1 \geq 0$ and $\phi_{j-1} = 1$ for $j - 1 \geq 0$; and
- G_n is the product of ϕ_i's and θ_j's with all combinations of ϕ_i $(i = 2, \ldots, n)$, such that $\theta_{i-1} = 1$ for $i - 1 > 0$ and $\phi_{j-1} = 1$ for $j - 1 > 0$.

We can simplify this to

$$a_{n+1} = \mathbb{F}_n a_0 + \mathbb{G}_n a_{-1}. \tag{2.52}$$

As an example,

$$a_4 = \mathbb{F}_3 a_0 + \mathbb{G}_3 a_{-1}, \tag{2.53}$$

where

$$\mathbb{F}_3 = \theta_3\theta_2\theta_1\theta_0 + \theta_3\theta_2\phi_1 + \theta_3\phi_2\theta_0 + \phi_3\theta_1\theta_0 + \phi_3\phi_1, \tag{2.54}$$

$$\mathbb{G}_3 = \phi_0\left(\theta_3\theta_2\theta_1 + \theta_3\phi_2 + \phi_3\theta_1\right). \tag{2.55}$$

Using Equation (2.52), together with the end boundary condition in Equation (2.18)

$$\begin{aligned} 0 &= -\frac{1}{\tau}a_{T-1} - k_f(N_b - T + 1)a_{T-1} + k_f(N_b - T + 2)a_{T-2} \\ &\implies \theta_{T-1}a_{T-1} = \phi_{T-1}a_{T-2}, \end{aligned} \tag{2.56}$$

and the fact that at $t = 0$ and $a_s = \delta_s$, we can, given any threshold value, T, write a general formula for the probability of escape, using

$$P_a = \frac{1}{\tau}\sum_{k=0}^{T-1} a_s. \tag{2.57}$$

2.5 Conclusion to Markov Model

In this section, we presented the Markov Model from (Holcman et al., 2014) that we collaborated on with Holcman et al. It was designed as mathematical framework for studying RNAi and the repression of mRNAs by miRNAs. One key finding of the paper was that, when the miRNA binding process is reversible rather than irreversible, the probability of escape is lower. In this thesis, we expanded upon the paper to provide the solution to the reversible binding case. Despite this advancement of the Markov model for RNAi, we believe it is important to study the NEP and MTT in a realistically restricted environment, where molecule counts are low and nuclear pores are small. Therefore, we focus on the expansion of the current Markov chain into a spatial stochastic model using Smoldyn, as described in the following section. In the Markov model, a very small number of pore complexes ($N = 10$), and large threshold values ($1 < T < 20$), are assumed, which are unrealistic values for a mammalian cells. In our Smoldyn model, we simulate around 3000 nuclear pore complexes on the surface of the nucleus and study the actual time for mRNAs to diffuse and locate one of these complexes, before it can be degraded. We also assume $T = 1$, as the nucleotide sequences of miRNAs and pri-miRNAs are approximately the same length.

2.6 ODE model

As nuclear RNAi is a fairly new topic of research, there are no existing models to build upon directly. However, RNAi can be thought of as a subset of post-transcriptional regulation so we can adapt such models for use here. One of the most recent ODE models of post-transcriptional regulation, which we also use in Chapter 3, is by Ala et al. and includes multiple MREs and ceRNAs (Ala et al., 2013). We adapt this ODE model for comparison against our SSA and spatial-stochastic models. Our model consists of pri-miRNAs (P), miR-709 (M), and one species of copycat miRNA (D).

$$(1)\quad \emptyset \underset{k_-^m}{\overset{k_+^m}{\rightleftharpoons}} \mathrm{M} \qquad (2)\quad \emptyset \underset{k_-^d}{\overset{k_+^d}{\rightleftharpoons}} \mathrm{D} \qquad (3)\quad \emptyset \underset{k_-^p}{\overset{k_+^p}{\rightleftharpoons}} \mathrm{P}$$

$$(4)\quad \mathrm{P} + \mathrm{D} \underset{k_f^{p,d}}{\overset{k_b^{p,d}}{\rightleftharpoons}} \mathrm{C}^{p,d} \qquad (5)\quad \begin{cases} \text{if } \alpha = 0, & \mathrm{P} + \mathrm{M} \xrightarrow{k_f^{p,m}} \mathrm{C}^{p,m} \xrightarrow{k_\emptyset^{p,m}} M \\ \text{if } \alpha = 1, & \mathrm{P} + \mathrm{M} \xrightarrow{k_f^{p,m}} \mathrm{C}^{p,m} \xrightarrow{k_\emptyset^{p,m}} \emptyset \end{cases}$$

From this, we get the following system of ODEs:

$$\frac{dM}{dt} = k_+^m - Mk_-^m - MPk_f^{p,m} + (1-\alpha)(C^{p,m}k_\emptyset^{p,m}), \tag{2.58}$$

$$\frac{dD}{dt} = k_+^d - Dk_-^d - DPk_f^{p,d} + C^{p,d}k_b^{p,d}, \tag{2.59}$$

$$\frac{dP}{dt} = k_+^p - Pk_-^p - P(Mk_f^{p,m} + Dk_f^{p,d}) + C^{p,d}k_b^{p,d}, \tag{2.60}$$

$$\frac{dC^{p,m}}{dt} = PMk_f^{p,m} - C^{p,m}k_\emptyset^{p,m}, \tag{2.61}$$

$$\frac{dC^{p,d}}{dt} = PDk_f^{p,d} - C^{p,d}k_b^{p,d}, \tag{2.62}$$

Here, we assume that miR-709 (M) and copycats (D) form complexes with pri-miRNAs (P) with association rates $k_f^{p,m}$ and $k_f^{p,d}$, and dissociation rates $k_b^{p,m}$ and $k_b^{p,d}$, where lowercase letters correspond to matching uppercase letters. Transcription rates are given by k_+^i and degradation rates are given by k_-^i, where $i \in \{p, m, d\}$. We assume that only pri-miRNA−miR-709 complexes ($C^{p,m}$) can degrade, at rate $k_\emptyset^{p,m}$, as a result of nuclear retention or miRNA-mediated degradation. The parameter $\alpha \in [0, 1]$, in the first ODE, represents the stability of miRNAs, i.e. the probability that miRNAs are degraded when the transcript to which they are bound is degraded. When $\alpha = 1$, miRNAs are always degraded upon pri-miRNA−miR-709 complex degradation, and the pri-miRNA−miR-

709 reaction is said to be "stoichiometric". When $\alpha = 0$, miRNAs are always recycled back into the miRNA pool upon pri-miRNA−miR-709 complex degradation, and the pri-miRNA−miR-709 reaction is said to be "catalytic" or "enzymatic". We do not have any data on this parameter so we assume, for simplicity, that $\alpha = 1$ such that miR-709 is always retained/degraded in the complex.

2.6.1 ODE Parameter Values

For comparison with our Smoldyn model, we take our parameter values from Table 3.2. These parameters, plus additional ODE-specific parameters are listed in Table 3.1 below. We solve our ODE using MatLab's built in 'ode45' function and our results are given in Section 3.5.

Param.	Description	Value	Ref.
V	Volume of the nucleus	572.66 μm^3	(Ala et al., 2013)
$C_m(0)$	Initial concentration of miR-709	[1]	
$C_d(0)$	Initial concentration of copycats	[1]	
$C_p(0)$	Initial concentration of pri-miRNA	[1]	
k_+^M	Rate of miR-709 production	0.2 s^{-1}	(Ala et al., 2013)
k_-^M	Rate of miR-709 degradation	0.0003 s^{-1}	(Ala et al., 2013)
k_+^M	Rate of miRNA copycat production	0.2 s^{-1}	(Ala et al., 2013)
k_-^M	Rate of miRNA copycat degradation	0.0003 s^{-1}	(Ala et al., 2013)
k_+^P	Rate of pre-mRNA production	0.155 s^{-1}	(Ala et al., 2013)
k_-^P	Rate of pre-mRNA degradation	0.0004 s^{-1}	(Ala et al., 2013)
k_b^{PM}	Rate of pri-miRNA−miR-709 unbinding	0.0003 s^{-1}	(Ala et al., 2013)
k_b^{PD}	Rate of pri-miRNA−copycat unbinding	0.0003 s^{-1}	(Ala et al., 2013)
α	Catalytic constant	1	(Ala et al., 2013)

Table 2.1: ODE Parameters.

[1] Calculated using the formulae $C_i(0) = \frac{N_i(0)}{(N_A \times V)}$, where $N_i(0)$ is the corresponding number of molecules of species i from Table 2.2, $N_A = 6.022 \times 10^{23}$ is Avogadro's constant, and V is the volume of the subcellular compartment which, in this case, is the nucleus.

2.7 SSA model

To study the importance of stochastic and spatial effects on RNAi, we create a purely stochastic model to compare with our ODE and Smoldyn models. As explained in Chapter 1, we use the Stochastic Simulation Algorithm (SSA) as a basis for our stochastic model (Gillespie, 1977). The general SSA algorithm is as follows:

1. **Initialisation ($t = 0$)**: Initialise the number of molecules in the system and the rate parameters.
2. **Monte Carlo step ($t = t + \Delta t$)**: Use randomly generated numbers to determine which molecules will react and the time step, Δt.
3. **Update**: Increase the time step and update the numbers of each molecule in accordance with the biochemical reactions.
4. **Iterate**: Repeat steps 2-3 unless no reactions can occur or the simulation time has been exceeded.

In more detail, we assign propensity functions to all biochemical reactions, including transcription and bi-molecular reactions, dependent on their reaction rates and the concentration of substrate. For the case of miR-709, M, regulating pri-miRNA, P, via one MRE leading to degradation of both pri-miRNA and miR-709, we have

$$(1)\quad \emptyset \xrightarrow{k_+^m} \mathrm{M} \qquad (2)\quad \emptyset \xrightarrow{k_+^p} \mathrm{P} \qquad (3)\quad \mathrm{M} \xrightarrow{k_-^m} \emptyset \qquad (4)\quad \mathrm{P} \xrightarrow{k_-^p} \emptyset \qquad (5)\quad \mathrm{M} + \mathrm{P} \xrightarrow{k_\emptyset^d} \emptyset.$$

This is a simplified version of the reaction system shown in Section 2.6 after removing copycats. Their respective propensity functions would be,

$$\alpha_1(t) = k_+^m, \qquad \alpha_2(t) = k_+^p, \qquad \alpha_3(t) = k_-^m M(t), \qquad \alpha_4(t) = k_-^p P(t), \qquad \alpha_5(t) = k_\emptyset^d M(t)P(t).$$

From this we calculate $\alpha(t) = \sum_{i=1}^{5} \alpha_i(t)$ to use in the remainder of the SSA. The full algorithm is as follows (general version given in Section 1):

1. Initialise $t = 0, M(0)$ and $P(0)$.

2. Calculate the propensity functions, $\alpha_i(t)$, for each reaction, i, as above.

3. Generate two random numbers, r_1 and r_2, from the uniform distribution, $U[0,1]$.

4. Compute $\alpha(t) = \sum_{i=1}^{5} \alpha_i(t)$ and determine at what time, τ, the next reaction will occur as

$$\tau = \frac{1}{\alpha(t)} \ln\left[\frac{1}{r_1}\right]. \tag{2.63}$$

5. Determine which reaction, j, will occur by

$$\frac{1}{\alpha(t)} \sum_{i=1}^{j-1} \alpha_i(t) \leq r_2 < \frac{1}{\alpha(t)} \sum_{i=1}^{j} \alpha_i(t). \tag{2.64}$$

6. Perform the reaction, j, updating the number of each reactant and product, the current simulation time, $t + \tau$, and the propensity functions.

7. Repeat steps $2 - 5$ until the simulation time is exceeded or some 'stop' condition has been satisfied.

2.7.1 SSA Parameter Values

Unless otherwise stated, we use the same parameter values as listed in Table 2.1, and use the initial number of each molecular species given in Table 2.2. We simulate our SSA, including random number generation, using MATLAB. Our results are shown in Section 2.9.

2.8 Development of Spatial Stochastic Model

We built a spatial stochastic model of RNAi using the Smoldyn simulation tool, in order to study the effects of spatial heterogeneity, and to test some of the assumptions made in our Markov model regarding nuclear escape probabilities and mean times to siRNA binding (Holcman et al., 2014).

2.8.1 Smoldyn Model Formulation

We modelled the mammalian nucleus as a sphere, of radius R_{nuc}. Our reactions are considered to be taking place in the nucleus, therefore, we do not model the cytoplasm or any other part of the cell. We assume that there are initally P pri-miRNA and and M miR-709 molecules diffusing through the domain. Pri-miRNA molecules can be degraded if they are bound by a molecule of miR-709 (assuming the threshold value, $T = 1$). Alternatively, pri-miRNA molecules can escape from the nucleus, with probability P_a, if they come into contact with any of N nuclear pore complexes on the nuclear membrane (see Figure 2.4). For simplicity, we will also be ignoring the nucleolus here, in order to create the simplest possible model for comparison with the Markov model. We look at a more complex geometries in the PTEN model.

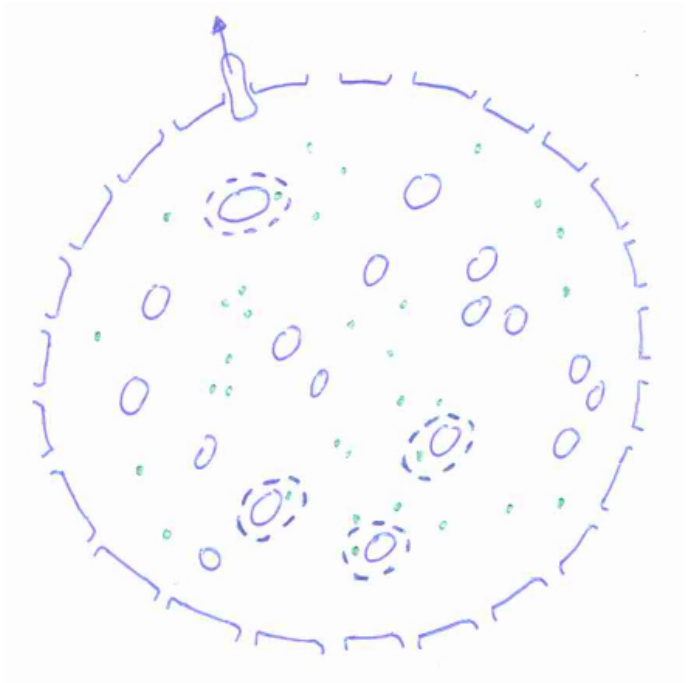


Figure 2.4: A schematic of the nucleus in 2D. Large blue circles represent pri-miR-15a/16-1 and small green dots represent miR-709. Nuclear pores are shown all around the surface, through which pri-miRs can escape into the cytoplasm, as shown at the top of the figure. Potential reactions are shown here surrounded by a dotted line. In the next time step, these reactions would be realised, resulting in changes in the number of each molecular species involved.

Initially, we assume that binding of siRNAs and degradation of RNAs are both irreversible, and assume a stoichiometric model, in which miRNAs are retained/degraded with their complexed target

mRNAs. The entire chemical reaction network is as follows:

$$(1)\quad \emptyset \underset{k_-^m}{\overset{k_+^m}{\rightleftharpoons}} \mathrm{M} \qquad (2)\quad \emptyset \underset{k_-^p}{\overset{k_+^p}{\rightleftharpoons}} \mathrm{P} \qquad (3)\quad \mathrm{P} + \mathrm{M} \xrightarrow{k_f^{p,m}} \mathrm{C}^{p,m} \xrightarrow{k_\emptyset^{p,m}} \emptyset,$$

where P represents pri-miRNAs and M represents miR-709.

As mentioned above, we consider RNAs to be capable of escaping from the nucleus via nuclear pores, thereby avoiding degradation by entering the cytoplasm, to potentially be translated into protein. In the case of the Markov chain model, the probability that a pri-miRNA molecule can escape from the nucleus is modelled as a narrow escape problem (Cheviakov and Zawada, 2013; Reingruber et al., 2009; Schuss et al., 2007). This makes the assumption that the likelihood of finding a nuclear pore is solely based on the frequency of holes and how an RNA diffuses from the centre of the nucleus to the membrane. In order to include more detail, we look at the effects of unevenly distributed nuclear pores on the membrane, and non-uniform distribution of RNAs within the nucleus. We also look at the impact of macromolecular crowding on the ability of a pri-miRNA to locate a nuclear pore before it is located by a miR-709 molecule, by introducing stationary and/or moving obstacles, O.

We only ever model one pri-miRNA species targeted by one miRNA species, as it is unclear whether ceRNA networks are active in the nucleus. We examine the impact of reversible siRNA binding as well as varying the threshold, T.

Repression/degradation of pri-miRNAs is assumed to be irreversible here, i.e. we do not consider the case of reversible transcript repression. Our model also takes into account the excluded volume interactions, such that any two non-reactive species which interact within their given binding radii will be moved away from each other to two new points within the space, separated by slightly more than the sum of their binding radii.

As with *PTEN* regulation, the study of RNAi in the nucleus has potential benefits for novel cancer therapeutics in the form of copycats introduced into the nucleus for interference with miRNA-mediated regulation of nuclear pri-miRNAs. Although we do not have data to validate the model, we hope to understand the underlying mechanisms and the importance of spatial modelling of such processes.

2.8.1.1 Smoldyn Parameter Values

We primarily used the work of Fusco et al. for calculating the binding radii and diffusion coefficients of molecular species in our simulations (Fusco et al., 2003). In their publication, Fusco et al. detail calculations for RNA radii and diffusion coefficients that use the length of the nucleotide sequences and the assumption that these molecules are roughly spherical. They utilise the Stokes-Einstein equation,

$$D = \frac{k_B T}{6\pi\eta r},$$

where:

- $k_B = 1.38 \times 10^{-23}$ Nm/K is Boltzman's constant;
- $T = 310$K is absolute temperature; and
- $\eta = 8.1 \times 10^{-3}$ Nsm^{-2} is the cytoplasmic viscosity.

The radius of the RNA, r, is approximated to be circular using $l = 2\pi r$, where l is the length of the RNA nucleotide sequence in meters (1 nucleotide (nt) $\sim$ 0.3 nm). Additionally, Fusco et al. assume that RNAs travel in complexes with radii roughly four times the size of the miRNA radius, resulting in adjusted diffusion coefficients, although the binding radii remain the same. The calculation for miRNAs is as follows:

1. Length of miRNA, l:

$$l = 22 \text{ nt} = 6.6 \text{ nm (using 1 nt} \sim 0.3 \text{ nm)}.$$

2. Circular miRNA radius, r:

$$2\pi r = 6.6 \text{ nm} \implies r = 1.0504 \times 10^{-9} \text{ m } (1.0504 \times 10^{-3}\,\mu\text{m}).$$

3. Complexed miRNA radius, R:

$$R = 4 \times r = 4.202 \times 10^{-9} \text{ m}.$$

4. Diffusion coefficient, D:

$$D = k_B T / 6\pi\eta r = 6.67\ \mu\text{m}^2\ \text{s}^{-1}.$$

We also use this above method to find the diffusion coefficients of mobile obstacles within the system (See Table 3.2 for values).

Param.	Description	Value	Ref.
R_{nuc}	Radius of the nucleus	5.67 μm	
V_{nuc}	Volume of the nucleus	763.55 μm^3	
R_p	Binding radius of each pri-miRNA	0.0015 μm	(Fusco et al., 2003)
R_m	Binding radius of each miRNA	0.0015 μm	(Fusco et al., 2003)
R_o or δ	Radius of mobile crowders	0.3 μm [1]	(Fusco et al., 2003)
D_p	Diffusion coefficient of each pri-miRNA	6.67 μm^{2}s^{-1}	(Fusco et al., 2003)
D_m	Diffusion coefficient of each miRNA	6.67 μm^{2}s^{-1}	(Fusco et al., 2003)
D_o	Diffusion coefficient of each mobile crowder	0.093 μm^{2}s^{-1} [2]	(Fusco et al., 2003)
θ	Volume fraction	0.3	(Mourão et al., 2014a)
k_+^p	Rate of pri-miRNA production	0.2 s^{-1}	(Ala et al., 2013)
k_-^p	Rate of pri-miRNA degradation	0.0003 s^{-1}	(Ala et al., 2013)
k_+^m	Rate of miRNA production	0.2 s^{-1}	(Ala et al., 2013)
k_-^m	Rate of miRNA degradation	0.0003 s^{-1}	(Ala et al., 2013)
k_b^{pm}	Rate of pri-miRNA-miRNA unbinding	0.0003 s^{-1}	(Ala et al., 2013)
$N_p(0)$	Initial number of pri-miRNA molecules	See text.	
$N_m(0)$	Initial number of miRNA molecules	See text.	
H	Number of nuclear pore complexes	3000	

Table 2.2: Smoldyn RNAi Parameters.

2.9 Results and Discussion

2.9.1 Extension of Markov Model Simulations

From Figure 7. in (Holcman et al., 2014), we see the probability of escape of an RNA molecule for varying thresholds ($T \in [1, 20]$) and escape times (τ). We see that, for $T = 10$, the escape probabilites are 0.5, 0.75, 1 for increasing escape times ($\tau k_f = 0.1, 1, 10$). We would expect this as

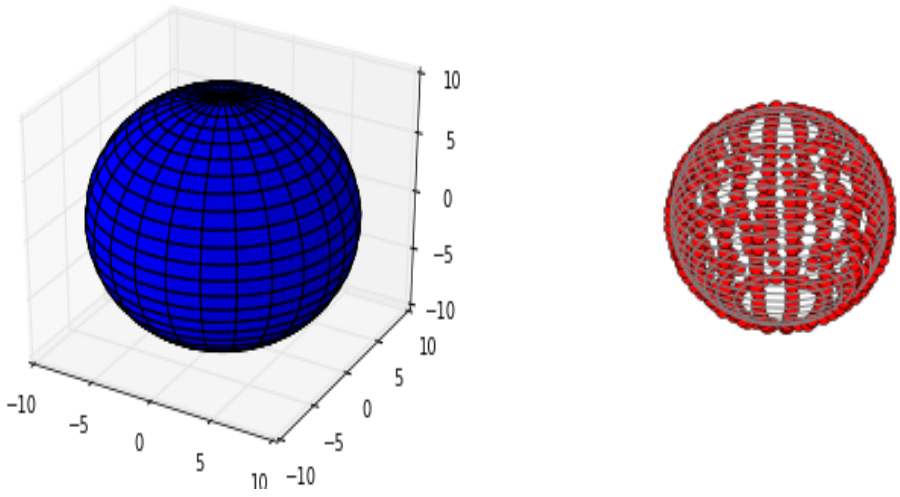
[1] $R_O \div 2$ for small crowders, $R_O \times 1.5$ for large crowders.
[2] $D_O \times 2$ for small crowders, $D_O \div 1.5$ for large crowders.

decreasing the escape time, τ, whilst fixing the siRNA-RNA binding rate (e.g. $k_f = 1s^{-1}$), increases the probability of escape.

Here we expand on the results presented in our paper (Holcman et al., 2014), beginning with the effects of obstacles on escape and degradation of molecules. First, let us demonstrate the spatial system in which our simulations are carried out.

2.9.1.1 Smoldyn Setup

Figure 2.5 here shows the set up of the spherical domain and the holes on the surface of the nucleus, which we create for all of the following Smoldyn simulations. As per Table 2.2, we added 3000 holes to the surface of the nucleus. According to the Smoldyn user manual, "the molecules are placed randomly on the surface with a constant density, on average" (`http://www.smoldyn.org/SmoldynUsersManual.pdf`).



(a) A (b) B

Figure 2.5: Smoldyn set up showing the spherical domain of the nucleus and the pores on the surface.

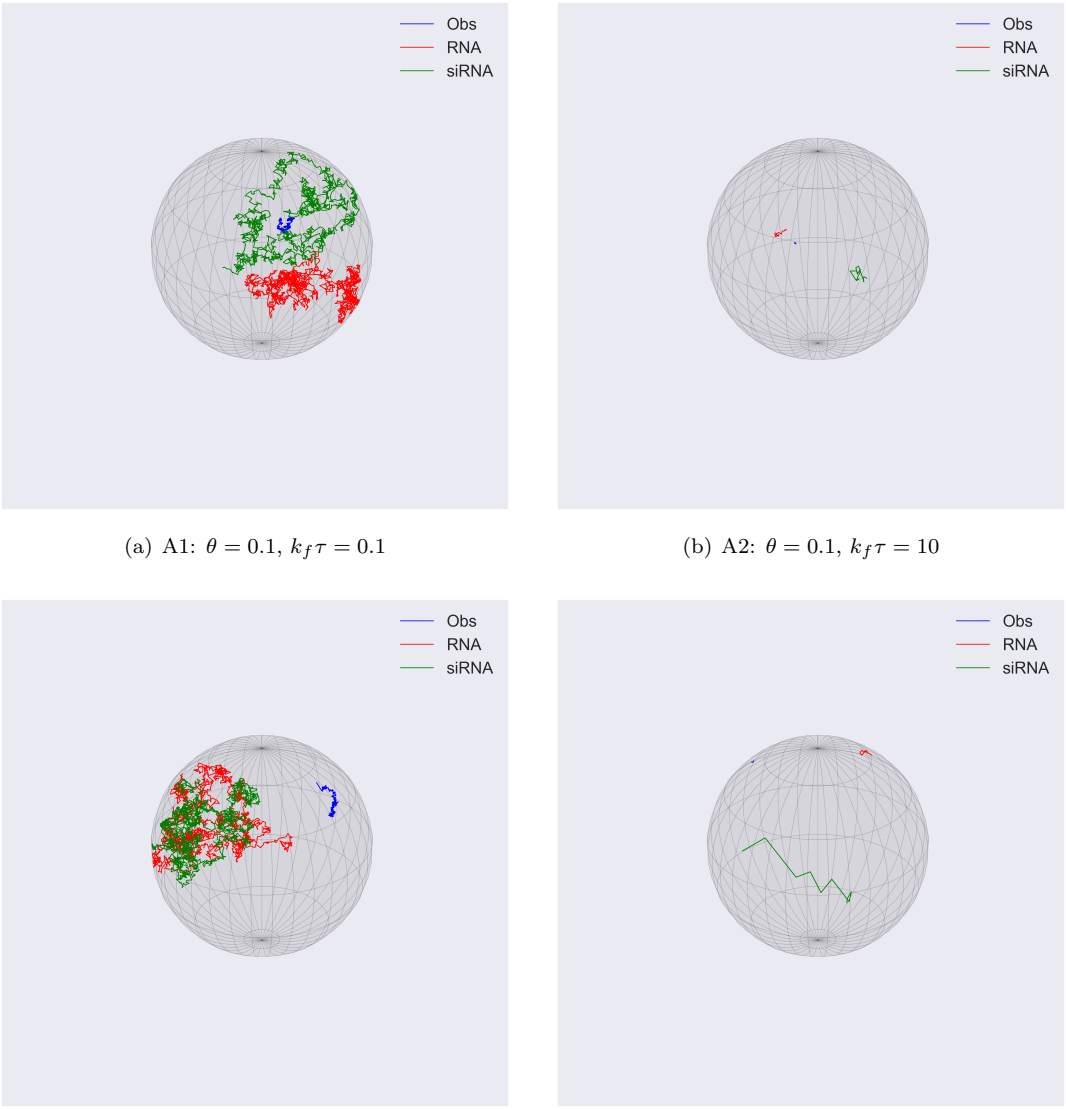
2.9.2 Effect of Obstacles

Before we begin the discussion of varying volume fractions (macromolecular crowding proportions), we should clarify that the typical obstacle (crowder) radius in the nucleus is 3nm. However, if we

were to use this value, we would need more than 1 billion obstacles for volume fraction $\theta = 0.25$, for example. These simulations would be too difficult to simulate in a reasonable time frame, so we increase the radius to reduce the number of obstacles to several hundred or thousand whilst maintaining the volume fraction.

Here (Figure 2.6 and 2.7), we look at the diffusion of siRNAs, RNAs and obstacles for varying crowding fractions. We do not allow molecules to escape or react, and simply look at the impact of crowding on the mobility of molecules in the space. This gives us an insight into the effects that macromolecular crowding and spatial simulations could have on the escape and binding rates. We can see that, in fact, increasing the volume fraction from $\theta = 0.1$ to $\theta = 0.2$ has little effect on the ability of molecules to diffuse through the space. However, increasing the value of $k_f\tau$ drastically inhibits the motion of all molecules. We should note that, in Smoldyn, increasing $k_f\tau$ is implemented by increasing the radius of molecules, thereby increasing the binding rate, k_f. So in this case, it is not the increased binding rate which impedes the movement of molecules, it is the increased size of the molecules and their interactions with obstacles.

From Figure 2.7, with the same volume fractions and molecule radii, we can see that the effect of crowding is much more heightened for stationary obstacles (Figures 2.7(b)) compared with mobile obstacles (Figures 2.7(a)). The most interesting result is that, for stationary obstacles, a higher volume fraction can actually increase the diffusion of molecules (comparing Figures 2.7(b) and 2.7(c)). It is not clear why this phenomenon occurs, but perhaps the molecules are being propelled through the space, as their radius is small enough and the binding-unbinding process results in swifter movement through the crowded environment. However, we were not able to find any experimental evidence to validate this finding. Additional results for $k_f = 10$ are given in the Appendix Figures A.4 and A.5, however the differences are much less apparent.



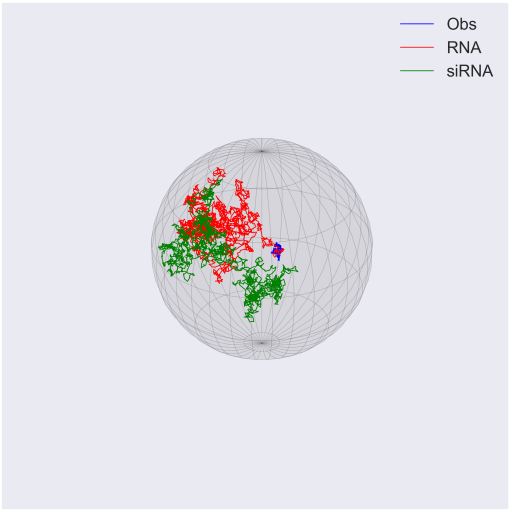
(a) A1: $\theta = 0.1$, $k_f\tau = 0.1$

(b) A2: $\theta = 0.1$, $k_f\tau = 10$

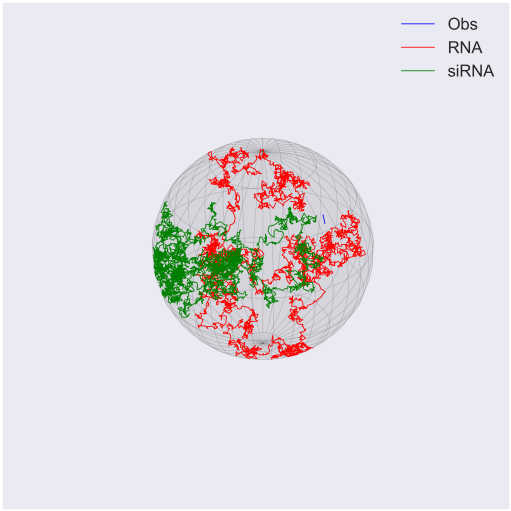
(c) B1: $\theta = 0.2$, $k_f\tau = 0.1$

(d) B2: $\theta = 0.2$, $k_f\tau = 10$

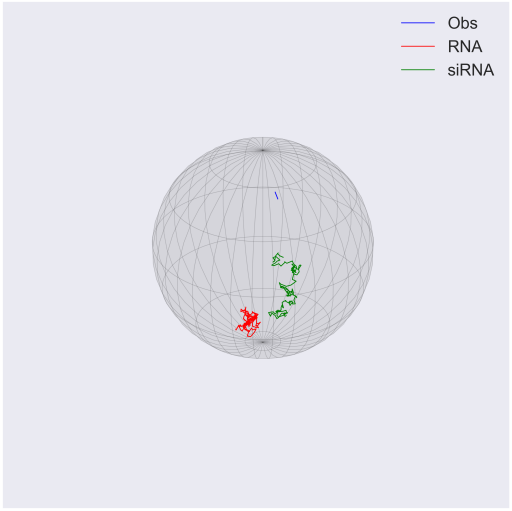
Figure 2.6: Diffusion of siRNAs, RNAs and obstacles for varying crowding fractions and mobile obstacles. Results are from a single simulation but similar results are seen when the simulations were repeated (data not shown). We also show variations in the binding rate, $k_f\tau$. We see here that increasing the volume fraction from $\theta = 0.1$ (Figures 2.6(a) and 2.6(b)) to $\theta = 0.2$ (Figures 2.6(c) and 2.6(d)) does not impede molecular diffusion through the available space. However, increasing the value of $k_f\tau$ from 0.1 (Figures 2.6(a) and 2.6(c)) to 10 (Figures 2.6(b) and 2.6(d)) drastically inhibits molecular diffusion.



(a) C1: Mobile Obstacles, $\theta = 0.3$, $k_f\tau = 0.1$



(b) D1: Stationary Obstacles, $\theta = 0.3$, $k_f\tau = 0.1$



(c) E1: Stationary Obstacles, $\theta = 0.2$, $k_f\tau = 0.1$

Figure 2.7: Diffusion of siRNAs, RNAs and obstacles for stationary vs. mobile obstacles, and varying mobile fractions ($\theta = 0.2$ and $\theta = 0.3$). Results are from a single simulation but similar results are seen when the simulations were repeated (data not shown). For a low binding rate, $k_f\tau = 0.1$, stationary obstacles (2.7(b)) can actually increase the diffusion of molecules compared to mobile obstacles (2.7(a)). A higher volume fraction is also shown to be able to increase the diffusion of molecules (2.7(b) vs. 2.7(c))

2.9.3 MSD calculation

In the previous section we have spoken about the impact of obstacles on molecular diffusion, showing examples of Smoldyn simulations with various volume fractions. Here, we combine this with the work done by Bruna et al. on mean square displacement (MSD) to calculate the MSD of molecules in crowded environments (Bruna and Chapman, 2013). We are comparing all of the following results to a formula by Bruna et al. which, in each of the following cases, is calculated as follows (Bruna and Chapman, 2013):

- L = 13.81 → This is $(4/3) \times \pi \times L^3$ (volume of the domain)
- $D_{b_0} = 0.17$ → Diffusion coefficient of the moving particles of interest e.g. RNA
- $D_{r_0} = 0.17$ → Diffusion coefficient of Obstacle
- $D_b = 1$ → Non-dimensional diffusion coefficient of RNA ($D_{b_0} \times [(1/L^2)/(D_{b_0}/L^2)] = D_{b_0}/D_{b_0} = 1$)
- $D_r = 1$ → Non-dimensional diffusion coefficient of Obs
- $Eps_b = 0.6/\mathrm{L}$ → Non-dimensional diameter of RNA (a point particle would be 0)
- $Eps_r = 0.6/\mathrm{L}$ → Non-dimensional diameter of Obs (divide by L)
- $N_b = 1$ → Number of RNA molecules
- $N_r = 4050$ → Number of Obs molecules
- d = 3 → Dimension of system

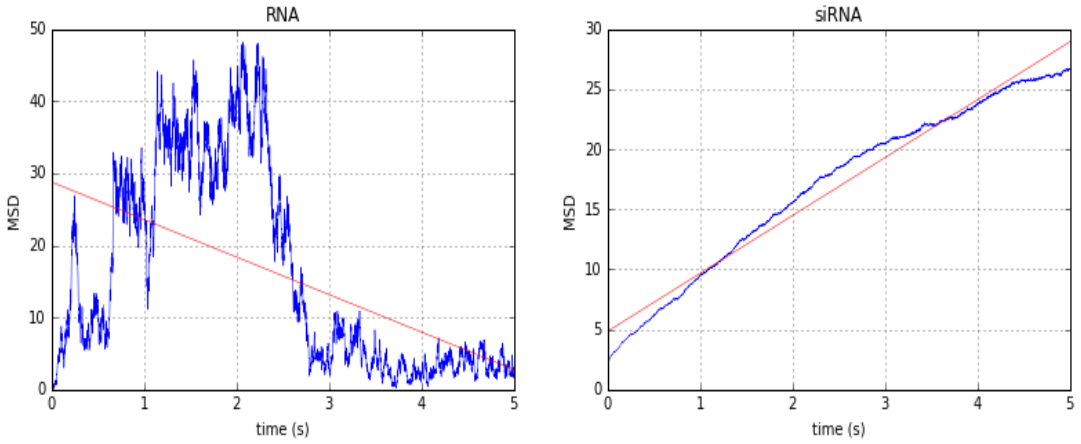
$$D11 = D_b \times (1 - (N_b - 1) \times \varepsilon_b^d \times (\pi/d) - N_r \times (((\varepsilon_b + \varepsilon_r)/2)^d) \tag{2.65}$$

$$\times (2 \times \pi/d) \times (D_b/(D_b + D_r))) \tag{2.66}$$

$$D_{eff} = D11 \times D_{b_0} \rightarrow \quad \text{re-dimensionalised effective diffusion coefficient} \tag{2.67}$$

For Bruna et al's formula, the effective diffusion coefficient is $D_{effMB} = 0.111$.

In Figures 2.8 and 2.9, we can see that the diffusion is positive. This is likely due to the fact that



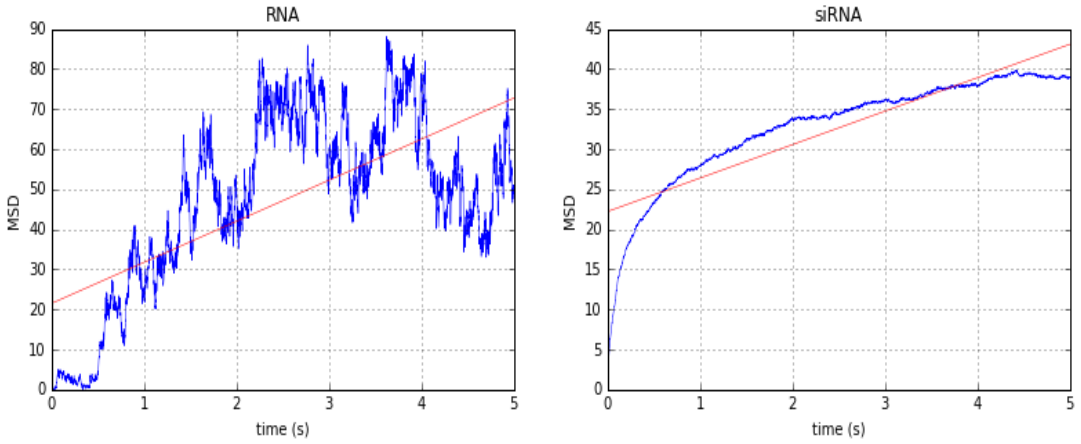
(a) RNA: $\theta = 0.2$, large mobile obstacles

(b) siRNA: $\theta = 0.2$, large mobile obstacles

Figure 2.8: Mean squared displacement of RNA and siRNA for $\theta = 0.2$ and big obstacles. (a) D_{eff} = -0.866, (b) D_{eff} = 0.802. The negative diffusion coefficient observed in (a) may be as a result of RNA trapping between obstacles in the crowded environment due to the RNA's larger size. Conversely, siRNAs are freely diffusing due to their smaller size.

the initial diffusion coefficients of molecules in the Smoldyn simulations are not the actual diffusion rates (depending on the interactions with obstacles and other molecules). Curiously, we observed a negative diffusion coefficient for the case of RNAs diffusion at $\theta = 0.2$ (Figure 2.8(a)). This may be because the RNA has become trapped after some time and the diffusion is restricted. On the other hand, siRNAs are not impeded due to their smaller size.

Increasing the volume fraction generally decreases the MSD. In the case of $\theta = 0.3$, the MSD increases as the molecules move more quickly throughout the space (which we have seen in the previous figures of molecular diffusion). At this point, however, we do start to see some anomalous diffusion in the case of siRNAs. Again, this is most likely due to the obstacles and their interactions with the diffusing molecules.



(a) RNA: $\theta = 0.3$, large mobile obstacles

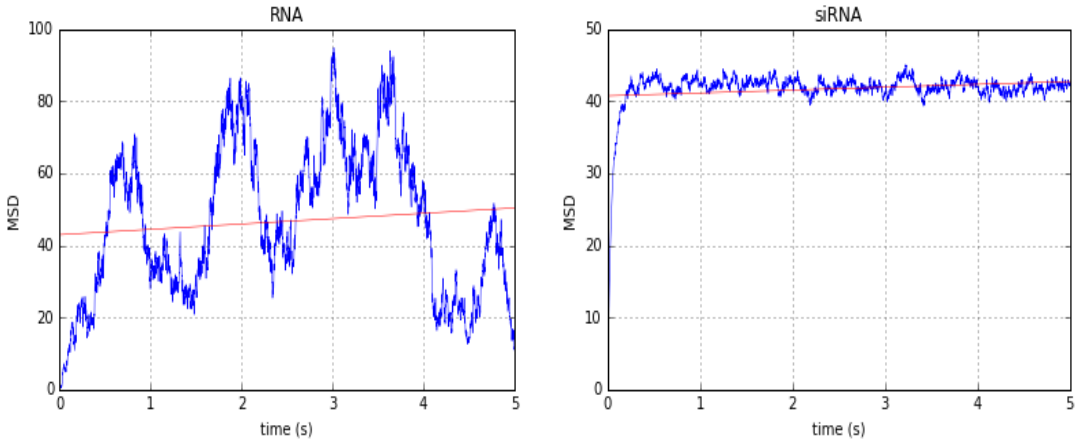
(b) siRNA: $\theta = 0.3$, large mobile obstacles

Figure 2.9: Mean squared displacement of RNA and siRNA for $\theta = 0.3$ and big obstacles. (a) $D_{eff} = 1.707$, (b) $D_{eff} = 0.695$. Comparing to Figure 2.8, increasing the volume fraction $\theta = 0.2$ to $\theta = 0.3$ generally decreases the MSD. We are beginning to see some anomalous diffusion in the case of siRNAs, most likely due to the obstacles interacting with the diffusing molecules.

In Figure 2.10, we look at stationary obstacles. We can see that the MSD is greatly reduced compared to mobile obstacles. It is likely that this reduction in MSD is due to some degree of volume exclusion from the initial geometry of the stationary obstacles, meaning that both siRNAs and RNAs are unable to diffuse freely from their initial position.

2.9.4 Effect of Copycats on RNA Degradation (Without Escape)

Next we look at the effects of copycats on RNA degradation. We begin with the ODE model of the system (where $N_b = 1, T = 1$) in each case. In these figures we can see that, as we might expect, the presence of copycats decreases the amount of RNA degradation. This is simply because the copycat molecules compete with siRNAs for RNA binding sites and, for both reversible and irreversible binding, some sites can be filled up so that the RNA can never be degraded.



(a) RNA: $\theta = 0.2$, NORMAL

(b) siRNA: $\theta = 0.2$, NORMAL

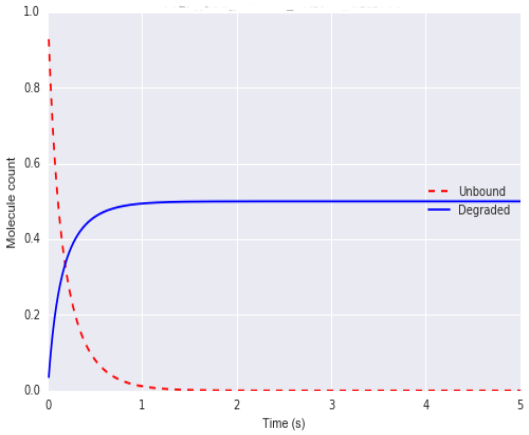
Figure 2.10: Mean squared displacement of RNA and siRNA for $\theta = 0.2$ and normal obstacles. (a) $D_{eff} = 0.247$, (b) $D_{eff} = 0.0665$. We see here the lowest MSD thus far, which is likely due to the stationary obstacles limiting mobility throughout the space via some degree of volume exclusion. This would mean that both siRNAs and RNAs are unable to diffuse freely within the domain.

For the SSA model, we look at varying the number of binding sites (N_b) on each RNA as well as the threshold T, whilst maintaining the ratio ($T = N_b/2$). We can see that, as we might expect, it affects the rate and number of degraded RNAs. We ran 100 simulations but we see little variation across the results.

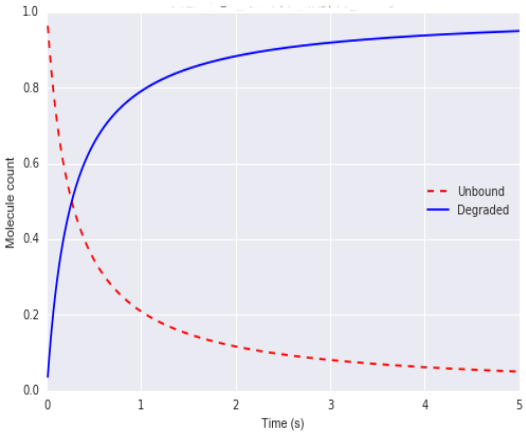
2.10 Conclusion

In this chapter, we have expanded on the work completed in our paper with Holcman et al. to include copycats and spatial heterogeneity and complexity. As we will support in the following chapter, we have shown that RNAs diffusing in a highly crowded confined domain can undergo anomalous (sub-)diffusion, but only when (si)RNAs are small and diffusing fast enough. If (si)RNAs are diffusing slowly or are too large, then the molecule is trapped by the highly crowded environment. We also studied the effects of crowding on the mean square displacement (MSD) of molecules compared to the work of Bruna et al. This supported our findings of molecular diffusion and demonstrated the impact of spatial computational modelling compared to deterministic mathematical formulae.

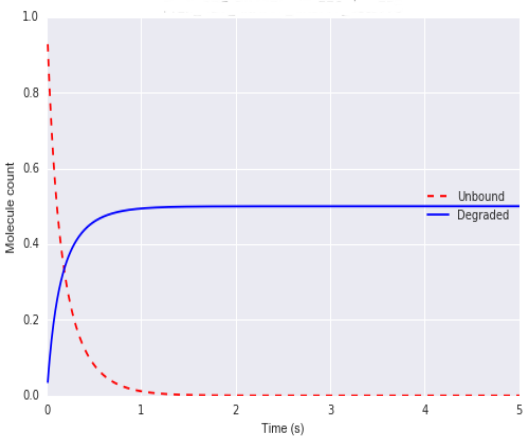
Our spatial model can be used to "reverse engineer" the rate constants and probabilities for SSA



(a) A: RNAs = siRNAs = copycats = 776

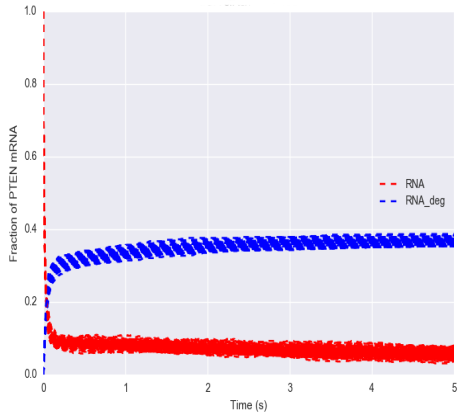


(b) B: RNAs = siRNAs = 776, copycats = 0

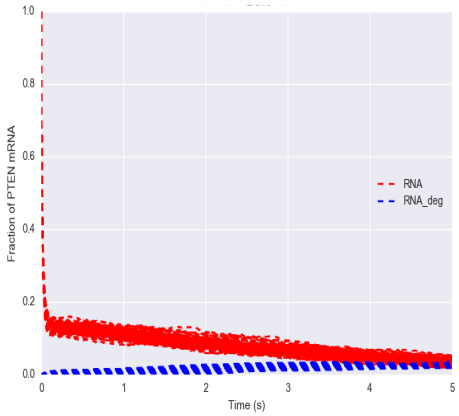


(c) C: RNAs = siRNAs = copycats = 776, reversible

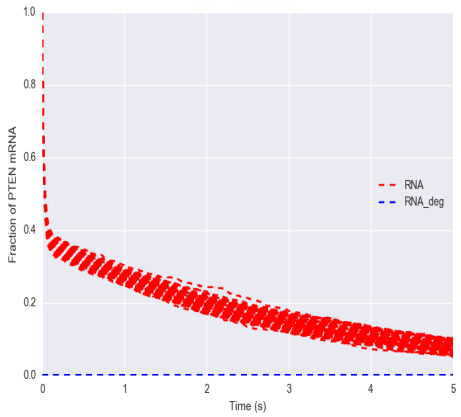
Figure 2.11: Fraction of mRNA degraded for varying numbers of copycats and reversibility. N_b, $T = 1$. The presence of copycats (2.11(a) and 2.11(c)) decreases the amount of RNA degradation. Copycats compete with siRNAs for RNA binding sites and, for both reversible and irreversible binding, some sites can be filled up so that the RNA can never be degraded.



(a) A: $N_b = 4$, $T = 2$



(b) B: $N_b = 10$, $T = 5$



(c) C: $N_b = 20$, $T = 10$

Figure 2.12: Fraction of mRNA degraded for varying number of binding sites and degradation thresholds. We can see, from 2.12(a) to 2.12(c), that increasing the number of binding sites, whilst maintaining a degradation threshold ratio of $T = N_b/2$, affects the rate and number of degraded RNAs. We ran 100 simulations but see little variation across the results (data not shown).

or ODE models using molecular weights and sizes, to better determine coefficients for use with deterministic and stochastic models.

We briefly modelled the effects of copycats, in that they act primarily to compete with siRNAs for RNA binding, and to block sites to protect RNAs from degradation. However, these interactions become much more complex when ceRNAs are also considered, which we model in the following chapter.

Chapter 3

Post-Transcriptional Gene Silencing of PTEN in the Cytoplasm

3.1 Introduction

As detailed in the Section 1.2, understanding post-transcriptional regulation is important for the development of novel cancer therapeutics. One prevalent example is the PTEN−PI3K−Akt pathway, in which PTEN mRNAs are targeted for post-transcriptional regulation by microRNAs.

Each PTEN mRNA nucleotide sequence contains one, or multiple, binding site(s), known as miRNA-response elements (MREs), to which complementary miRNAs can bind and regulate PTEN mRNA function via transcript degradation or endonucleolytic cleavage of the mRNA (degradation by breaking the bonds between nucleotides). Whether a transcript is repressed or degraded depends on the level of complementarity between the MRE and the seed sequence of the miRNA (nucleotides 2-9). Specifically, bulges in the MRE opposite miRNA nucleotide positions 9 (see Figure 3.1) and 13-15 have been shown to prevent transcript degradation in favour of repression.

Each miRNA can target many RNA species, resulting in competition between RNAs for binding to miRNAs. Competitive endogenous RNAs (ceRNAs) act by "sponging" up miRNAs, interfering with the ability of miRNAs to bind to other targets, causing upregulation of the remaining RNAs. As such, ceRNAs have great therapeutic potential in cancer prevention (de Giorgio et al., 2013; Giza et al., 2014). For tumour suppressor genes, addition of ceRNAs may lead to upregulation of protein production (e.g. PTEN), via increased translation, thereby reducing the likelihood of tumour formation and growth.

In addition to sponge-type mechanisms, non-sponge mechanisms are also involved in the post-transcriptional regulation of *PTEN*. In particular, miRNA "copycats" are small RNAs with nucleotide sequences that match functioning miRNAs, but without the ability to degrade or repress the target transcript directly. Instead, copycats act by binding to MREs, thereby blocking access to these sites by miRNAs. An additional level of complexity is introduced if copycats can also bind to ceRNAs. In our models, we include both ceRNAs and copycats in an attempt to determine their impact on PTEN mRNA levels, when present in isolation and in combination.

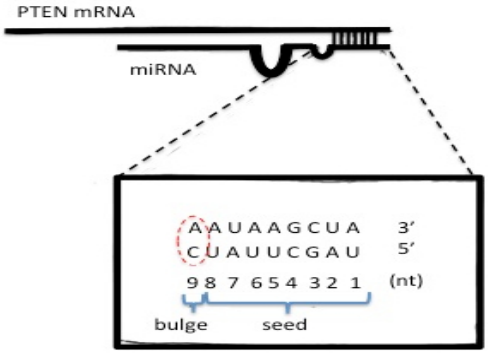


Figure 3.1: A schematic of an miRNA−mRNA complex with seed sequence (nucleotides 2-8) and a bulge opposite nucleotide position 9 caused by a mis-match between an adenine base ('A') in the mRNA sequence, and a cytosine base ('C') in the miRNA sequence. In RNA−RNA binding, cytosine is the complementary base to guanine ('C' binds to 'G') and adenine is complementary to uracil ('A' binds to 'U').

In addition to studying the pathway itself, we also explore the influence of stochasticity and spatial detail by comparing deterministic, stochastic and spatial-stochastic models. As described in Section 1 and Section 2.3, we created a Markov Chain model of the *PTEN* post-transcriptional regulatory pathway, investigating the effect of copycats on miRNA−PTEN-mRNA binding. We formulated the process in terms of a Narrow Escape Problem (NEP) − for PTEN mRNAs to escape miRNA-mediated degradation by translocating to the Endoplasmic Reticulum (ER) − and a Mean Time to Threshold (MTT) (Holcman et al., 2014). However, the model does not account for the effect of ceRNAs which is essential for our study of *PTEN*. As discussed at the start of this chapter and in Section 1.2, PTEN mRNAs compete with other ceRNAs for a limited pool of shared miRNAs (and copycats), so it is important to study this competition for a more realistic representation of the system. Therefore, in this chapter we focus only on the MTT and leave the comparisons with the Markov Model to the RNAi model in Chapter 2.

We begin with a description of the ODE model and SSA, which we developed from existing models. We then present the spatial-stochastic model that we created using the simulation tool, Smoldyn. We end the chapter with a comparison of each modelling approach and a discussion of our findings in Sections 3.5 and 3.6.

3.2 ODE model

There are several deterministic models of post-transcriptional regulation (Ala et al., 2013; Mukherji et al., 2011; Wang et al., 2010). The most recent model was presented by Ala et al. (2013) and includes the effects of multiple ceRNA species (Ala et al., 2013). We use this ODE model for comparison against our SSA and spatial-stochastic models. As per our Smoldyn model, our ODE model consists of PTEN mRNAs (P), one species of ceRNA (S), and one species of miRNA (M), and we include the addition of one species of copycat miRNA (D). With a single binding site on each mRNA/ceRNA, the full reaction network takes the form:

$$(1)\quad \emptyset \xrightarrow{k_+^m} \mathrm{M} \qquad (2)\quad \emptyset \xrightarrow{k_+^d} \mathrm{D} \qquad (3)\quad \emptyset \xrightarrow{k_+^p} \mathrm{P} \qquad (4)\quad \emptyset \xrightarrow{k_+^s} \mathrm{S}$$

$$(5)\quad \mathrm{M} \xrightarrow{k_-^m} \emptyset \qquad (6)\quad \mathrm{D} \xrightarrow{k_-^d} \emptyset \qquad (7)\quad \mathrm{P} \xrightarrow{k_-^p} \emptyset \qquad (8)\quad \mathrm{S} \xrightarrow{k_-^s} \emptyset$$

$$(9)\quad \mathrm{S} + \mathrm{D} \underset{k_f^{s,d}}{\overset{k_b^{s,d}}{\rightleftharpoons}} \mathrm{C}^{s,d} \qquad (10)\quad \mathrm{P} + \mathrm{D} \underset{k_f^{p,d}}{\overset{k_b^{p,d}}{\rightleftharpoons}} \mathrm{C}^{p,d}$$

$$(11)\quad \mathrm{S} + \mathrm{M} \underset{k_f^{s,m}}{\overset{k_b^{s,m}}{\rightleftharpoons}} \mathrm{C}^{s,m} \qquad (12)\quad \begin{cases} \text{if } \alpha = 0, & \mathrm{P} + \mathrm{M} \xrightarrow{k_f^{p,m}} \mathrm{C}^{p,m} \xrightarrow{k_\emptyset^{p,m}} M \\ \text{if } \alpha = 1, & \mathrm{P} + \mathrm{M} \xrightarrow{k_f^{p,m}} \mathrm{C}^{p,m} \xrightarrow{k_\emptyset^{p,m}} \emptyset \end{cases}$$

Here, we assume that miRNAs (M) and copycats (D) form complexes with PTEN mRNAs (P) and ceRNAs (S) with association and dissociation rates $k_f^{i,j}$ and $k_b^{i,j}$, respectively, where $i = \{p, s\}$ and $j = \{m, d\}$ (lowercase letters correspond to matching uppercase letters). Transcription rates are given by k_+^i and k_+^j, and degradation rates are given by k_-^i and k_-^j. We do not consider translation

of PTEN mRNA into protein, and assume that miRNA−PTEN-mRNA complexes ($C^{p,m}$) degrade, at rate $k_{\emptyset}^{p,m}$, as a result of miRNA-mediated target degradation. Later, we expand on this system to include the case when either PTEN mRNAs or ceRNAs (or both) contain multiple binding sites.

From the reaction network, we get the following system of ODEs :

$$\frac{dM}{dt} = k_+^m - Mk_-^m - M(Pk_f^{p,m} + Sk_f^{s,m}) + C^{s,m}k_b^{s,m} + (1-\alpha)(C^{p,m}k_{\emptyset}^{p,m}), \tag{3.1}$$

$$\frac{dD}{dt} = k_+^d - Dk_-^d - D(Pk_f^{p,d} + Sk_f^{s,d}) + C^{p,d}k_b^{p,d} + C^{s,d}k_b^{s,d}, \tag{3.2}$$

$$\frac{dP}{dt} = k_+^p - Pk_-^p - P(Mk_f^{p,m} + Dk_f^{p,d}) + C^{p,d}k_b^{p,d}, \tag{3.3}$$

$$\frac{dS}{dt} = k_+^s - Sk_-^s - S(Mk_f^{s,m} + Dk_f^{s,d}) + C^{s,m}k_b^{s,m} + C^{s,d}k_b^{s,d}, \tag{3.4}$$

$$\frac{dC^{p,m}}{dt} = PMk_f^{p,m} - C^{p,m}k_{\emptyset}^{p,m}, \tag{3.5}$$

$$\frac{dC^{p,d}}{dt} = PDk_f^{p,d} - C^{p,d}k_b^{p,d}, \tag{3.6}$$

$$\frac{dC^{s,m}}{dt} = SMk_f^{s,m} - C^{s,m}k_b^{s,m}, \tag{3.7}$$

$$\frac{dC^{s,d}}{dt} = SDk_f^{s,d} - C^{s,d}k_b^{s,d}. \tag{3.8}$$

As per the RNAi model, the parameter, α, represents the probability that miRNAs are degraded along with the transcript to which they are bound. Ala et al. propose that no crosstalk can be observed if $\alpha = 0$, so we assume that $\alpha \neq 0$. In fact, in our model, we assume that $\alpha = 1$ so that miRNAs always degrade along with their targets and are never recycled.

3.2.1 ODE Parameter Values

In order to compare the ODE/SSA models with our Smoldyn model, we take our parameter values from the Smoldyn output file, which shows the representative reaction rates based on the reaction radii and diffusion coefficents (summarised in Table 3.2). A summary of these ODE parameters are given in Table 3.1 below. We use MATLAB's built in 'ode45' function, which implements an explicit Runge-Kutta method, for solving the ODE system, and present our results in Section 3.5.

Param.	Description	Value	Ref.
V	Volume of the cytoplasm	$2290.08\mu\text{m}^3$	(Ala et al., 2013)
$C_m(0)$	Initial concentration of miRNA	[1]	
$C_d(0)$	Initial concentration of copycats	[1]	
$C_p(0)$	Initial concentration of PTEN mRNA	[1]	
$C_s(0)$	Initial concentration of ceRNAs	[1]	
k_b^{PM}	Rate of miRNA-PTEN mRNA unbinding	0.0003s^{-1}	(Ala et al., 2013)
k_b^{PD}	Rate of copycat-PTEN mRNA unbinding	0.0003s^{-1}	(Ala et al., 2013)
k_b^{SM}	Rate of miRNA-ceRNA unbinding	0.0003s^{-1}	(Ala et al., 2013)
k_b^{SD}	Rate of copycat-ceRNA unbinding	0.0003s^{-1}	(Ala et al., 2013)
α	Catalytic constant	1	(Ala et al., 2013)

Table 3.1: Table of parameters for the ODE model.

3.3 SSA model

To study the importance of stochastic effects on the system, we create a purely stochastic model to compare with our ODE and Smoldyn models. As explained in Section 1, we use the Stochastic Simulation Algorithm (SSA) as a basis for our stochastic model (Gillespie, 1977). The general SSA can be found in Section 1, and we use the same chemical reaction network given for the ODE model in Section 3.2. We assign propensity functions to all biochemical reactions, including transcription and bi-molecular reactions, dependent on their reaction rates and the concentration of substrate, as per Section 2.7.

[1] Calculated using the formulae $C_i(0) = \frac{N_i(0)}{(N_A \times V)}$, where $N_i(0)$ is the initial number of molecules of species i ("HIGH" = 700; "LOW" = 100); $N_A = 6.022 \times 10^{23}\text{mol}^{-1}$ is Avogadro's constant, and $V\mu m^3$ is the volume of the subcellular compartment which, in this case, is the cytoplasm.

3.3.1 SSA Parameter Values

We use the same parameter values as listed in Table 3.1. We use the initial number of each molecular species given in Table 3.2 and use rate constants in units of μs^{-1}. We simulate our SSA using MATLAB and present our results in Section 3.5.

3.4 Development of Spatial Stochastic Model

In the following section, we describe our spatial stochastic model of the system. The primary hypotheses that we wish to test in this chapter are as follows:

$\mathcal{H}1$: Stationary crowders produce similar dynamics to dilute (non-crowded) conditions in a confined domain.

$\mathcal{H}2$: Reactive mobile obstacles have a bigger impact on diffusion and reaction rates than inert (non-reactive) mobile obstacles.

$\mathcal{H}3$: The size of the obstacles is as important as the total volume fraction.

$\mathcal{H}4$: Adding ceRNAs to the system causes a greater upregulation of PTEN mRNAs than the addition of copycat miRNAs or a combination of both ceRNAs and copycats (if copycats can bind to ceRNAs).

$\mathcal{H}5$: Reversibility of miRNAs or copycats binding to PTEN mRNAs or ceRNAs does not affect PTEN mRNA degradation.

3.4.1 Smoldyn Model Formulation

We focus our study on eukaryotic cells, modelled as a sphere of radius R_{cell}, containing a spherical nucleus of radius R_{nuc}. All reactions are assumed to take place in the cytoplasm, therefore, the nuclear membrane is impermeable. We assume that there are initially $P_0(= P(0))$ PTEN mRNAs and $M_0(= M(0))$ miRNAs. To simplify the model, we assume that there are is no birth/death of the miRNAs or mRNAs (aside from miRNA-driven degradation/repression). We assume that there are p MREs on each PTEN mRNA, and degradation occurs when a threshold number of miRNAs, T, are bound to the mRNA. Initially, we assume that binding of miRNAs is irreversible,

and assume miRNAs are degraded with their complexed target mRNAs. From this, we arrive at a much simplified version of the reaction network presented in Section 3.2 where $\alpha = 1$:

$$\text{PTEN}_i + \text{miRNA} \rightarrow \text{PTEN}_{i+1} \quad \text{(Irreversible binding)}, \tag{3.9}$$

$$\text{PTEN}_{T-1} + \text{miRNA} \rightarrow \emptyset, \tag{3.10}$$

where $i \in [0, T-1]$. The threshold value, T, is another factor that we study, as it is unclear what the exact value is *in vivo*, and indeed it is likely to depend on the miRNA species. For example, Tian et al. presented a study of four miRNA species, miR-26a, miR-92a, miR-19b, and miR-23b. These miRNAs also have between one and three MREs on the 3′UTR of PTEN mRNA (see Figure 3.2) (Tian et al., 2013). Prior to this, as mentioned in Section 1.2, Tay et al. demonstrated the presence of 14 PTEN MREs for 10 different miRNA species, with each having between one and three binding sites in the 3′UTR of PTEN mRNA, and between one and four sites on each of seven putative ceRNAs (Tay et al., 2011). The ability of miRNAs to act cooperatively is another factor

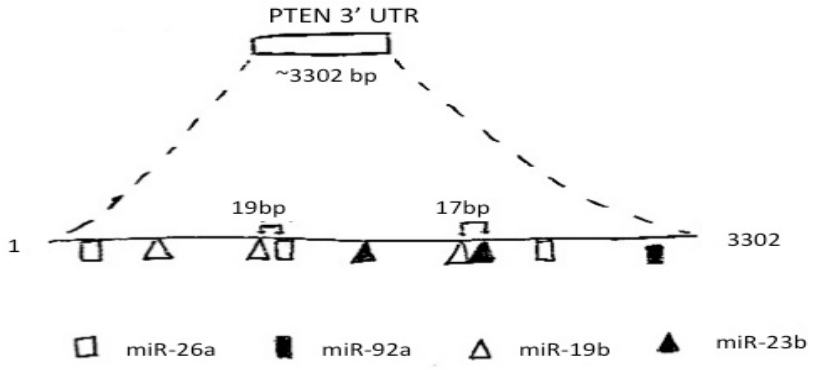


Figure 3.2: A schematic of the miRNA binding sites in the 3′UTR of a PTEN mRNA (adapted from (Tian et al., 2013)).

that must be addressed; If miRNAs are cooperative then the binding of one miRNA may affect the ability of subsequent miRNAs to bind, either positively or negatively, and the total number bound may affect the strength of regulation, i.e. whether targets are repressed or degraded. Consequently, if miRNA cooperativity enhances subsequent miRNA binding, for example, then we model this by increasing the effective binding radius of RNAs as miRNAs bind.

As explained in Section 1.2, PTEN mRNAs form part of a network of ceRNAs (ceRNET). We simplify this network by considering the miRNA pool (miR_P) to be represented by one miRNA

species, and similarly for the network of ceRNAs (ceR_N). We justify this simplification as follows. *PTEN* is said to be part of a ceRNA network of 14 RNAs (including PTEN mRNA) (Sumazin et al., 2011), and PTEN mRNAs are predicted to be targeted by 14 miRNAs. Therefore,

$$\text{miR}_P = \sum_{i=1}^{14} [\text{miR}_i] \leq 14 \times \max_{i=1,\ldots,14} \{\text{miR}_i\}, \tag{3.11}$$

where [*] denotes concentration. Similarly,

$$\text{ceR}_N = \sum_{j=1}^{14} [\text{ceR}_j] \leq 14 \times \max_{j=1,\ldots,14} \{\text{ceR}_j\} \tag{3.12}$$

where ceR_N includes PTEN mRNA. As the networks are of the same order, and the concentrations of miRNAs and ceRNAs are not thought to vary much between RNA families, we assume the simplest case of two ceRNAs (PTEN mRNA plus one ceRNA) modulated by two miRNA species. This speeds up our simulations, as we are using a stochastic simulation tool that tracks individual molecules, whilst still allowing a degree of ceRNA crosstalk. The impact of ceRNAs was not considered in (Holcman et al., 2014), although it is an important factor to consider following the discovery of the network of ceRNAs involved in *PTEN* regulation (Sumazin et al., 2011; Tay et al., 2011; Karreth et al., 2011; Cesana et al., 2011). As such, our spatial stochastic model should provide us with greater insights into the dynamics of *PTEN* post-transcriptional regulation than our Markov model.

With the introduction of S ceRNAs, with s available binding sites on each ceRNA, our biochemical system is now of the form

$$\begin{aligned} \text{PTEN}_i + \text{miRNA} &\rightarrow \text{PTEN}_{i+1}, \\ \text{PTEN}_{T-1} + \text{miRNA} &\rightarrow \emptyset, \\ \text{ceRNA}_j + \text{miRNA} &\rightarrow \text{ceRNA}_{j+1}, \end{aligned} \tag{3.13}$$

for $i \in [0, T-1]$ and $j \in [0, s-1]$, with $s \geq 1$. Later, we will include the possibility of reversible miRNA binding as well as introducing D copycat miRNAs and stationary/mobile obstacles, O, into the system.

Since we do not consider miRNA cooperativity at this stage, we assume that the rate of binding to each site is equal. Later we investigate the case where bound miRNAs increase the likelihood that

subsequent miRNAs will bind. Once a threshold value, T, of miRNAs are bound, the PTEN mRNA is assumed to be degraded so that,

$$\text{PTEN}_{T-1} + \text{miRNA} \to \emptyset, \tag{3.14}$$

whereas the final state for a ceRNA is simply ceRNA_s. We assume here that the miRNAs bound to the PTEN mRNA are also degraded ($\alpha = 1$ in ODE model) (Ala et al., 2013).

In reality, not all miRNAs downregulate their targets, and there are many examples of miRNAs that promote RNA stability. Similarly, copycat miRNAs can be designed experimentally to antagonise miRNAs by this method of direct competition. It is important to study the impact of copycats independent of, and in conjunction with, ceRNAs to determine if there is any negative interference via binding to ceRNAs. When copycats are included, reactions (3.13) and (3.14) are supplemented by

$$\begin{aligned}
\text{PTEN}_{i,k} + \text{miRNA} &\to \text{PTEN}_{i+1,k}, \\
\text{PTEN}_{T-1,k} + \text{miRNA} &\to \emptyset, \\
\text{PTEN}_{i,k} + \text{copycat} &\to \text{PTEN}_{i,k+1}, \\
\text{ceRNA}_{j,l} + \text{miRNA} &\to \text{ceRNA}_{j+1,l}, \\
\text{ceRNA}_{j,l} + \text{copycat} &\to \text{ceRNA}_{j,l+1},
\end{aligned} \tag{3.15}$$

for $i \in [0, T-1]$, $k \in [0, p-i]$, $j \in [0, s-1]$, and $l \in [0, s-j-1]$. As mentioned previously, miRNA copycats have been proposed as a potential therapy for interfering with miRNA-mediated regulation of PTEN. However, as copycats can also bind to ceRNAs, it has not yet been studied whether the addition of copycats would increase or decrease miRNA activity (see Figure 3.3).

We study this in our simulations to try to ascertain the conditions under which copycats have the greatest impedence on miRNA−PTEN-mRNA binding as opposed to miRNA−ceRNA binding. For

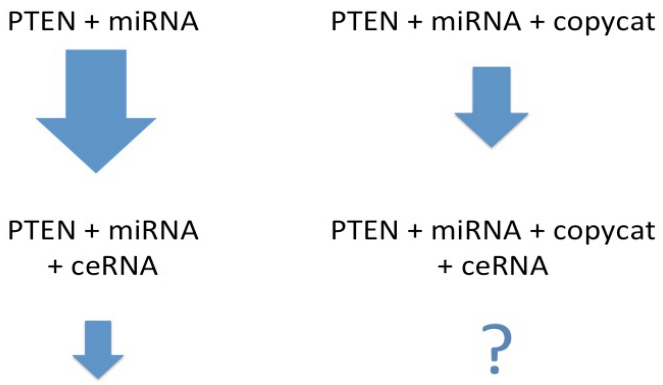


Figure 3.3: A schematic of the level of *PTEN* repression as a result of copycats and ceRNAs. Whilst the effect of ceRNAs and copycats independently is fairly clear, it is difficult to know how they will affect the system together.

reversible binding, the full system is given by

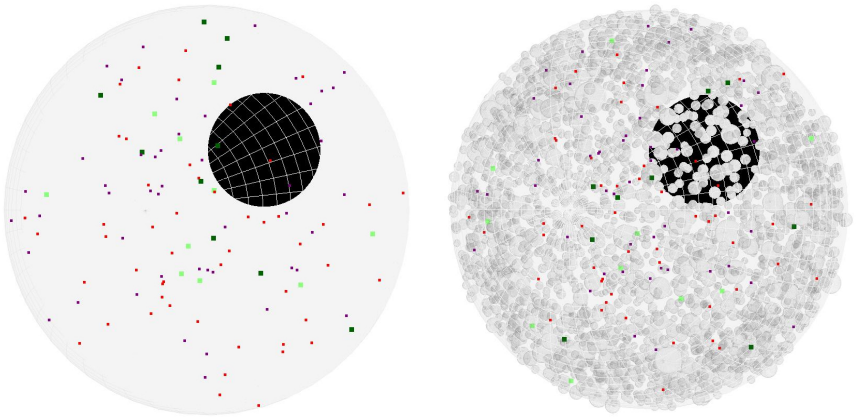
$$\begin{aligned} \text{PTEN}_{i,k} + \text{miRNA} &\rightleftharpoons \text{PTEN}_{i+1,k}, \\ \text{PTEN}_{T-1,k} + \text{miRNA} &\to \emptyset, \\ \text{PTEN}_{i,k} + \text{copycat} &\rightleftharpoons \text{PTEN}_{i,k+1}, \\ \text{ceRNA}_{j,l} + \text{miRNA} &\rightleftharpoons \text{ceRNA}_{j+1,l}, \\ \text{ceRNA}_{j,l} + \text{copycat} &\rightleftharpoons \text{ceRNA}_{j,l+1} \end{aligned} \tag{3.16}$$

for $i \in [0, T-1]$, $k \in [0, p-i]$, $j \in [0, s-1]$, and $l \in [0, s-j-1]$. This is of course for the case without macromolecular crowding (see Figure 3.4(a)).

For our model to be realistic, we add macromolecular obstacles or "crowders" to the system at a specified volume fraction, θ, which represents the fraction of the volume that is obstructed (see Figure 3.4(b)). These reactions are represented as

$$\text{O} + \text{X} \to \text{O} + \text{X}, \tag{3.17}$$

where O represents any obstacle, generated in the same way as other molecules in the system, with a given diffusion coefficient, D_o, and radius, R_o; and X $\in$ {PTEN mRNA, ceRNA, miRNA, copycat}. These types of reactions are necessary in Smoldyn to ensure volume exclusion, and are also used to model interactions between all non-reactive species (e.g. PTEN + ceRNA $\to$ PTEN + ceRNA).



(a) (b)

Figure 3.4: (a) A screenshot of a Smoldyn simulation without crowding. (b) A screenshot of a Smoldyn simulation with crowding. The nucleus is shown as a black opaque sphere. Obstacles, shown in grey, are generated using the Smoldyn extension, SmolCrowd. Mobile obstacles would be represented in a similar way but with uniform size. Dark green and light green squares represent PTEN mRNAs and ceRNAs, respectively. Red and orange squares represent miRNAs and copycats, respectively. The nucleus is shown as a black opaque sphere.

Any two non-reactive species that interact within their given binding radii will be relocated to two new points within the space, separated by slightly more than the sum of their binding radii (see Figure 3.5). A full schematic model is given in Figure 3.6.

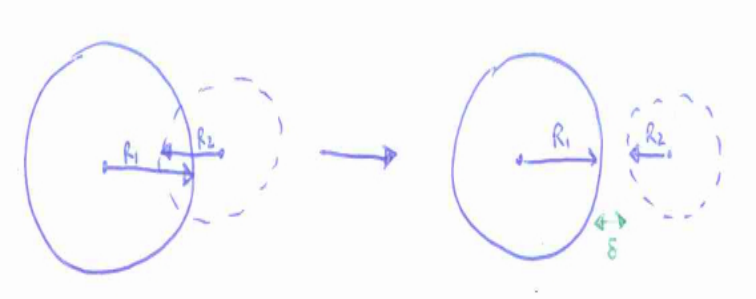


Figure 3.5: A schematic of excluded volume reactions, between two non-reactive molecular species, between two obstacles. For example, when an obstacle, of radius R_1, and a PTEN mRNA, of radius $R2$, come within $R = (R_1 + R_2)\mu m$ of each other, they are relocated to a distance equal to their original separation, R, plus an additional separation, δ. This ensures that rebinding will not occur immediately.

Repression of PTEN mRNAs is assumed to be irreversible throughout, that is, we assume permanent translational repression or degradation of PTEN mRNAs. In reality, some stress-induced translational repression mechanisms are reversible (Bhattacharyya et al., 2006); miRNA−mRNA complexes are stored in processing bodies (P-bodies) and released when the source of cellular stress

is no longer present. We ignore this case for simplicity as there is no data to support this process occurring for PTEN mRNAs. In any case, there is no available information suggesting how long mRNAs are stored in these P-bodies, and the storage time is likely to be considerably longer than our time-scale of interest.

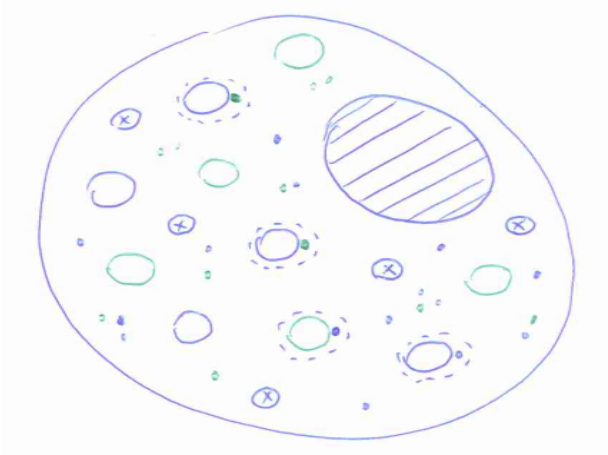


Figure 3.6: A schematic of the cell in 2D. Large blue and green circles represent PTEN mRNAs and ceRNAs, respectively. Small blue and green dots represent miRNAs and copycats, respectively. Circles with a cross represent newly degraded mRNAs. The nucleus is represented by the large shaded circle. Potential reactions are shown here surrounded by a dotted line. In the next time step, these reactions would be realised, resulting in changes in the number of each molecular species involved.

3.4.1.1 Smoldyn Parameter Values

When considering miRNA-RNA interactions, it is important to take into account the number of each molecular species, as well as the relative ratios of each species. Ragan et al. postulate that the ratio miRNA:mRNA is likely to be of the order of 10:1, but suggest that the rate of reactions and overall level of target repression would be different if the relative concentrations were 1μM:100nM compared with 100nM:10nM, for example (Ragan et al., 2011). Biological data on the concentrations of miRNAs and mRNAs in mammalian cells is lacking, but as we are performing a stochastic simulation, we use some of the smaller values recorded in the literature. In particular, Hudson et al. suggest that miR-25, an miRNA known to target PTEN mRNAs, is present in around 776 molecules per cell ("HIGH") (Hudson et al., 2012). We also consider the case with fewer molecules (100 molecules of miRNA; "LOW") to study stochasticity more clearly.

As with the RNAi regulatory model, we follow Fusco et al. (Fusco et al., 2003) when estimating

the binding radii and diffusion coefficients of the molecular species, using the length of the RNA nucleotide sequence and the assumption that molecules are roughly spherical. It is also assumed in (Fusco et al., 2003) that mRNAs and miRNAs travel in complexes roughly four times their size (e.g. miRNA-RISC), resulting in larger diffusion coefficients, although the binding radii remain the same. The calculation for PTEN mRNAs, using the nucleotide sequence length, $l = 3300$nt, gives the mRNA binding radius, $R = 6.28 \times 10^{-7}$m, and diffusion coefficient, $D = 0.0446\text{m}^2\text{s}^{-1}$. We also use this above method to find the diffusion coefficients of mobile obstacles (see Table 3.2). Note: In the table we mention set the reaction probability to be 0.5. This parameter was chosen to be less than 1 as a way to account for the fact that colliding particles require a high enough kinetic energy to bind; With this parameter, when two reactive molecules come into contact, they will not always react. The specific value of 0.5 was chosen arbitrarily between 0 and 1. No other values in this range were experimented with.

Param.	Description	Value	Ref.
N:C	Ratio of nuclear to cytoplasmic volume	1:3	
R_{nuc}	Radius of the nucleus	5.67μm	
R_{cell}	Radius of the cell	9μm	
V_{nuc}	Volume of the nucleus	$763.55\mu\text{m}^3$	
V_{cell}	Volume of the cell	$3053.63\mu\text{m}^3$	
V_{cyt}	Volume of the cytoplasm	$2290.08\mu\text{m}^3$	
R_m	Binding radius of each miRNA	0.0015μm	(Fusco et al., 2003)
R_d	Binding radius of each miRNA copycat	0.0015μm	(Fusco et al., 2003)
R_p	Binding radius of PTEN mRNA	0.157μm	(Fusco et al., 2003)
R_s	Binding radius of ceRNA	0.157μm	(Fusco et al., 2003)
R_o or δ	Radius of mobile crowders	0.3μm (3)	(Fusco et al., 2003)
D_m	Diffusion coefficient of each miRNA	$6.67\mu\text{m}^2\text{s}^{-1}$	(Fusco et al., 2003)
D_d	Diffusion coefficient of each miRNA copycat	$6.67\mu\text{m}^2\text{s}^{-1}$	(Fusco et al., 2003)
D_p	Diffusion coefficient of each PTEN mRNA	$0.044\mu\text{m}^2\text{s}^{-1}$	(Fusco et al., 2003)
D_s	Diffusion coefficient of each ceRNA	$0.044\mu\text{m}^2\text{s}^{-1}$	(Fusco et al., 2003)
D_o	Diffusion coefficient of each mobile crowder	$0.093\mu\text{m}^2\text{s}^{-1}$ (4)	(Fusco et al., 2003)
θ	Volume fraction	0.3	(Mourão et al., 2014a)
k_b^{pm}	Rate of miRNA-PTEN mRNA unbinding	0.0003s^{-1}	(Ala et al., 2013)
k_b^{pd}	Rate of copycat-PTEN mRNA unbinding	0.0003s^{-1}	(Ala et al., 2013)
k_b^{sm}	Rate of miRNA-ceRNA unbinding	0.0003s^{-1}	(Ala et al., 2013)
k_b^{sd}	Rate of copycat-ceRNA unbinding	0.0003s^{-1}	(Ala et al., 2013)
p	Reaction probability	0.5	
$N_m(0)$	Initial number of miRNA molecules	See text.	
$N_d(0)$	Initial number of copycat molecules	See text.	
$N_p(0)$	Initial number of PTEN mRNA molecules	See text.	
$N_c(0)$	Initial number of ceRNA molecules	See text.	

Table 3.2: PTEN Smoldyn Model Parameters.

3.5 Results

In this section we present results from our Smoldyn simulations, and comparisons to ODE and/or SSA results where relevant and possible. Our aim is to demonstrate the impact of stochasticity and spatial resolution, as well as parameters such as volume fraction (θ) and miRNA binding sites (MREs), on the outputs of the system, such as molecular diffusion and the rate of PTEN mRNA degradation.

We begin by looking at the impact of the volume fraction, size, and mobility of obstacles on the diffusion coefficients of miRNAs and mRNAs, to make inferences about any potential anomalous or confined diffusion. We then look at the impact of copycats, ceRNAs, MREs, and binding site reversibility, on the reaction rates and fraction of degraded PTEN mRNAs. We chose this output as it is indicative of the likelihood for tumour suppression due to PTEN levels.

We initially took values for molecular concentrations from the literature, such as (Ragan et al., 2011), which resulted in relative concentrations of [miRNA] $\approx 10\times$ [mRNA]. However, the imbalance of miRNAs to mRNAs resulted in a complete degradation of the pool of mRNAs almost immediately. We therefore modified this condition and proceeded with [miRNA] $\approx$ [mRNA]. In such cases, concentrations of PTEN mRNA will still be affected by miRNAs, but with degradation of mRNAs occurring over a longer time scale. We investigate the impact of ceRNAs and copycats on tumourigenic systems, and at what limit they might reverse the trend of PTEN mRNA degradation to restore its function as a tumour suppressor, thereby preventing the formation of cancerous cells. We introduce copycats and ceRNAs in "reasonable" concentrations ([copycats] $\approx$ [miRNAs]; [ceRNAs] $\approx$ [mRNAs]) that may be capable of affecting the system dynamics.

3.5.1 Effect of Obstacles on Molecular Diffusion

Here, we display typical molecule trajectories within a cell under a variety of different crowding conditions. We do not display the nucleus in each figure, but an example of the full cell without molecules is shown in Figure 3.7 for clarity. We also do not show the obstacles in the system, as this would obscure the trajectories, but crowded figures are labelled appropriately with their volume

[3] $R_O \div 2$ for small crowders, $R_O \times 1.5$ for large crowders.
[4] $D_O \times 2$ for small crowders, $D_O \div 1.5$ for large crowders.

fraction values. The purpose of these figures is to give a qualitative view of the range and freedom of movement of each molecule type depending on the cellular geometry. In all cases, trajectories in red, blue and green represent obstacles, mRNAs, and miRNAs, respectively.

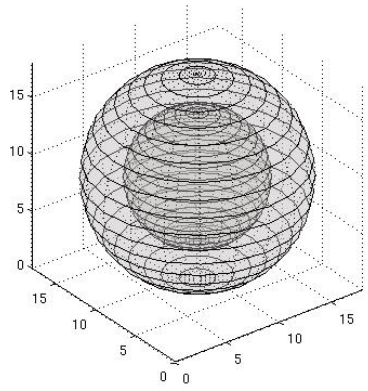


Figure 3.7: Typical cell schematic for displaying molecule trajectories.

For most parameters, we show the trajectories of one molecule of miRNA and mRNA in the cytoplasm for three different volume fractions: $\theta = 0, 0.3$ and 0.6. We match the qualitative trajectories to more quantitative MSD figures for a single molecule and the average over a population of molecules.

Our simulations are most valuable for complex geometries, for which analytical solutions can not easily be found. Even the cell, which can, in the simplest case of cytoplasm, nucleus and infinitely thin membrane, be approximated as a sphere within a sphere (Figure 3.7), is a complicated domain to consider analytically. In the case of Bruna et al.'s analytical solution for the effective diffusion coefficient, one must assume the use of an infinitely large domain or periodic boundary conditions in order to be able to ignore the effects of confinement (Bruna and Chapman, 2013).

Regarding the MSD, if a molecular species moves by Brownian motion in three dimensions, then the MSD, $\langle x^2 \rangle$, satisfies

$$\langle x^2 \rangle = 6Dt, \tag{3.18}$$

where the diffusion coefficient, D, is a constant, i.e. we expect the average MSD over a large number of trajectories or particles to scale linearly with time. For an individual particle, however, the MSD behaviour is inherently random, although we expect the particle to move away from its initial position

at long times. This is, however, for an unconfined system, with no boundaries or periodic boundary conditions. In our model, the system (a cell) is confined by the cellular and nuclear membranes, resulting in a limitation in the maximal possible MSD (see Figure 3.8). In order to compare our results to those obtained for unbounded systems, we would need to look at final simulation times, t_f, which satisfy the condition,

$$t_f \ll \frac{(\mathrm{R}_{cell} - \mathrm{R}_{nuc})^2}{6D}, \tag{3.19}$$

such that the molecules could not reach either the cell or nuclear membrane by $t = t_f$.

3.5.1.1 Comparison figures

Before we show our results, we reproduce figures from publications to demonstrate the expected behaviour of the MSD for normal, directed, anomalous, and confined diffusion, on a linear and log-log scale. We present these as a comparison to our own results, to try to discern the type of diffusion that molecules are undergoing as a result of different crowding conditions.

Figure 3.8, reproduced from (Saxton, 2007b), highlights the behaviour of the MSD for confined diffusion, on a linear and log-log scale. As we know that our system is confined, it is important to be aware of the graphical features of such diffusion in order to be able to identify it.

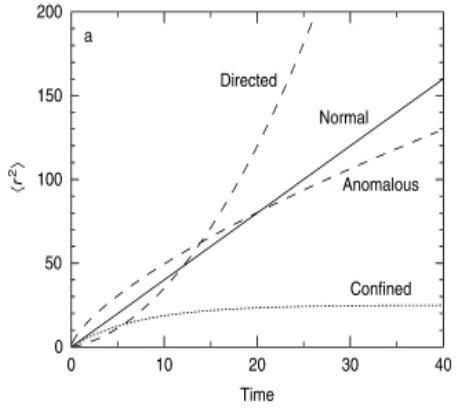


Figure 3.8: Graph showing the typical MSD behaviour for Brownian, anomalous, directed and confined diffusion (Saxton, 2007b).

Figure 3.9, reproduced from (Saxton, 2007a), highlights the behaviour of the MSD for anomalous diffusion, on a linear and log-log scale. Anomalous diffusion is most easily identifiable from the log-

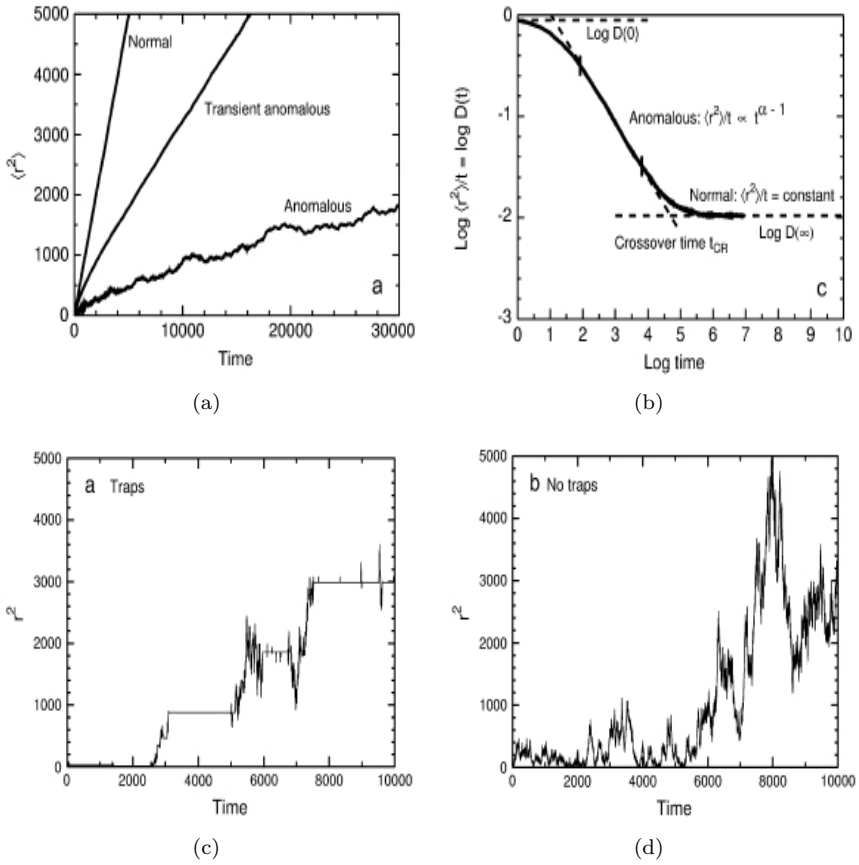


Figure 3.9: (a) and (b) show the typical MSD behaviour on a linear and log-log scale (Saxton, 2007a). (c) and (d) show the MSD behaviour for a system with and without traps.

log plot of MSD/t ($D(t)$) against time, t, (Figure 3.9(b)) as it is shown by a negative linear slope, with gradient $\langle r^2 \rangle / t = t^{\alpha - 1}$. Normal diffusion is shown by a horizontal line. The initial value is log $D(0)$ and the intersection of the anomalous and long-term normal diffusion slopes gives the so-called "crossover time", t_{CR}, from anomalous to normal diffusion. Figure 3.9(c) is very similar to the behaviour presented in Figure 3.15(c) and Figure 3.16(b) seen later in this section, suggesting that obstacles may be acting as "traps" for RNAs in the domain. Figure 3.9(d) highlights the behaviour of the MSD with no trapping or confinement, and is similar to the other trajectories (for example, Figures 3.14 and 3.16), showing that no trapping is occurring in any other case.

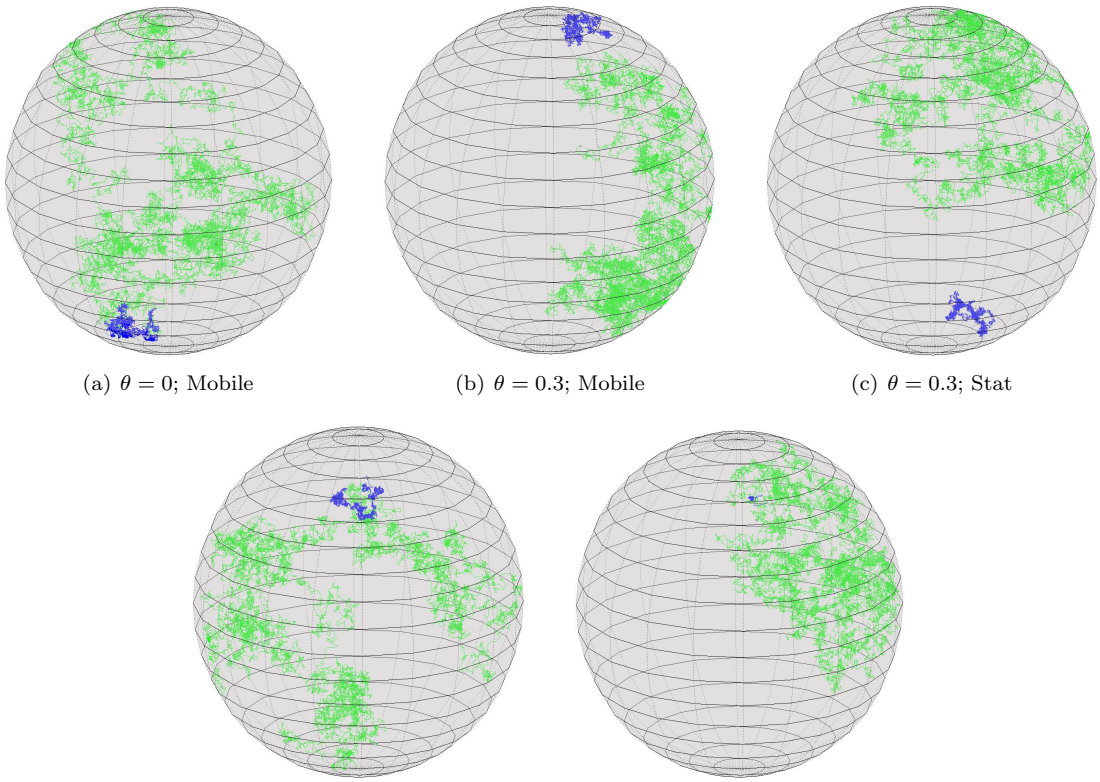
3.5.1.2 Effect of Volume Fraction and Obstacle Mobility

In Figure 3.10, we display typical trajectories (over 15 seconds) of a single miRNA and single mRNA moving throughout the cytoplasm. Figure 3.10(a) shows the case with no obstacles, i.e. the miRNA and mRNA can move freely throughout the space. Figures 3.10(b) and Figure 3.10(c) show the case for 30% crowding (i.e. $\theta = 0.3$) by "large" mobile and stationary obstacles, respectively. The obstacles are labelled as "large" because their radius is twice that of mRNAs and 200-times that of miRNAs. Figures 3.10(d) and Figure 3.10(e) show the mobile and stationary cases, respectively, for $\theta = 0.6$. As we can see, miRNAs have complete freedom of movement due to their size and diffusion coefficient, whereas RNAs are extremely restricted for the same reasons, being larger and slower than miRNAs. Molecules appear to be slightly more confined when obstacles are stationary, although this is only a sample trajectory so it is difficult to say definitively from this qualitative measurement.

The results in Figure 3.11 have the same initial conditions, in the same order (from (a)-(e)), as Figure 3.10 except that, in this case, PTEN mRNAs have the same diffusion coefficient as miRNAs ($D_p = D_m$). This means that mRNAs are able to diffuse faster, although their size remains the same. We can see from this figure that mRNAs have a much greater freedom of movement than in Figure 3.10, but that they are still somewhat restricted in the crowded cases due to their larger size, making them less able to diffuse between obstacles than miRNAs.

We can see more clearly in Figure 3.12, which is a magnified version of Figure 3.11(d), that, when the space is highly crowded, mRNAs appear to be moving via Lévy flight, with periods of Brownian motion interspersed by periods of long-range relocation (see Figure 3.13 for a comparision of Lévy with Brownian motion). This is likely due to mRNAs being "trapped" by the high volume of obstacles (leading to the periods of Brownian motion), followed by a short period where mRNAs have enough space, and a large enough diffusion coefficient, to move a significant distance away from its current location before becoming trapped again. This Lévy-type motion is somewhat clearer in the case with mobile obstacles (Figure 3.11(d) compared to Figure 3.11(e)); mRNAs have a higher chance of breaking free of traps as they "bump" obstacles upon collision, potentially clearing a path through which to diffuse, dependent on the diffusion coefficients of both mRNAs and obstacles.

In the following figures, we show the MSD plots that go with the above trajectories, beginning with



(a) $\theta = 0$; Mobile (b) $\theta = 0.3$; Mobile (c) $\theta = 0.3$; Stat

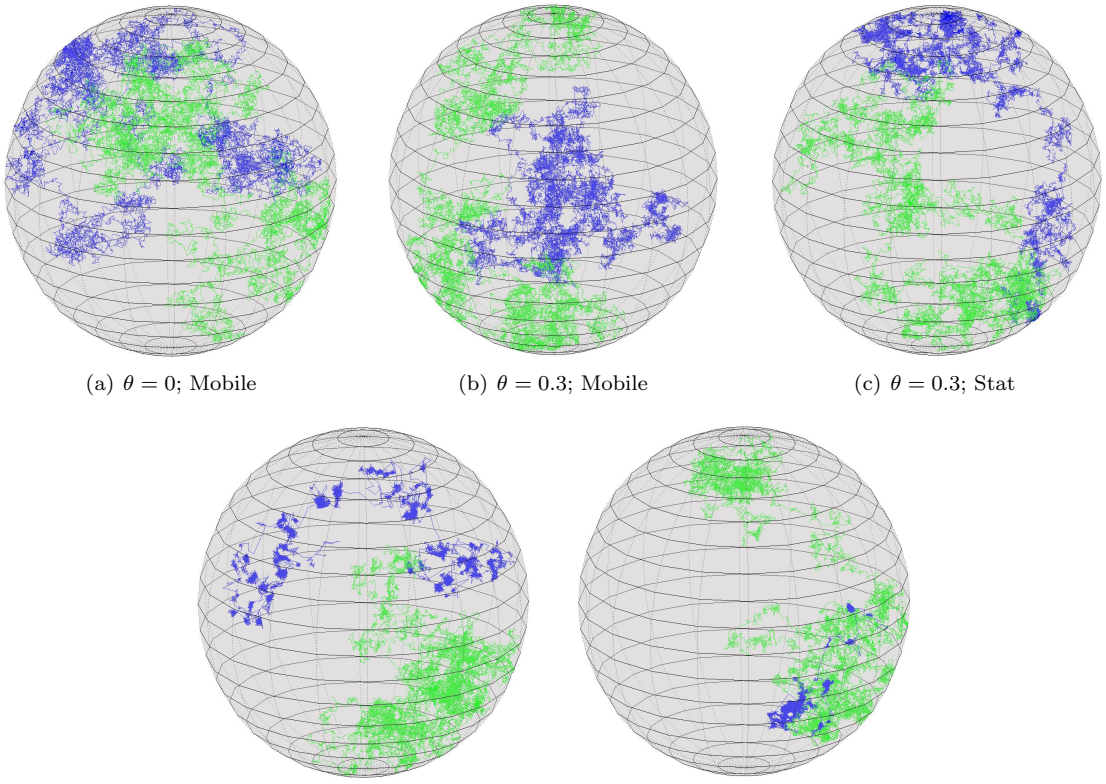
(d) $\theta = 0.6$; Mobile (e) $\theta = 0.6$; Stat

Figure 3.10: Example trajectories of RNAs and miRNAs ($t_f = 15s$) when the only molecules present are a single miRNA, a single mRNA, and obstacles where appropriate. Blue = RNA; Green = miRNA. $D_R = 0.04\mu m^2 s^{-1}$, $D_m = 6.67\mu m^2 s^{-1}$.

mobile obstacles. Please note the differing y-axis scales for each set of figures. We show MSD graphs for populations of molecules later in this section, where we also discuss anomalous diffusion and confinement in relation to the current model.

In Figures 3.14 and 3.15, we compare the MSD of the individual mRNAs from Figures 3.10 and 3.11. The first point to notice is that the mean of the MSD is larger in Figure 3.15, compared to Figure 3.14, when mRNAs have a bigger diffusion coefficient ($D_p = D_s$). The second point to notice is that the mean of the MSD is much larger for $\theta = 0.6$ compared with $\theta = 0.3$. This is not an obvious result to predict, as we would expect the MSD to decrease with volume fraction. In particular, we can see features of the Lévy flight shown as clear steps in the MSD plots in Figure 3.15(c) (and in Figure 3.16(b) for stationary obstacles).

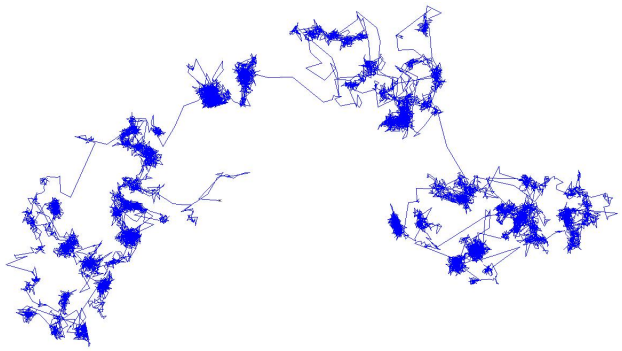
In Figure 3.16, we compare the MSD of trajectories from Figures 3.10 and 3.11, for stationary



(a) $\theta = 0$; Mobile (b) $\theta = 0.3$; Mobile (c) $\theta = 0.3$; Stat

(d) $\theta = 0.6$; Mobile (e) $\theta = 0.6$; Stat

Figure 3.11: Same conditions as Figure 3.10 but in this case PTEN mRNAs have the same diffusion coefficient as miRNAs, although their size remains the same. $D_R = 6.67\mu m^2 s^{-1} = D_m$.



(a) 60% stationary Obs

Figure 3.12: Isolated RNA trajectory from Figure 3.11.

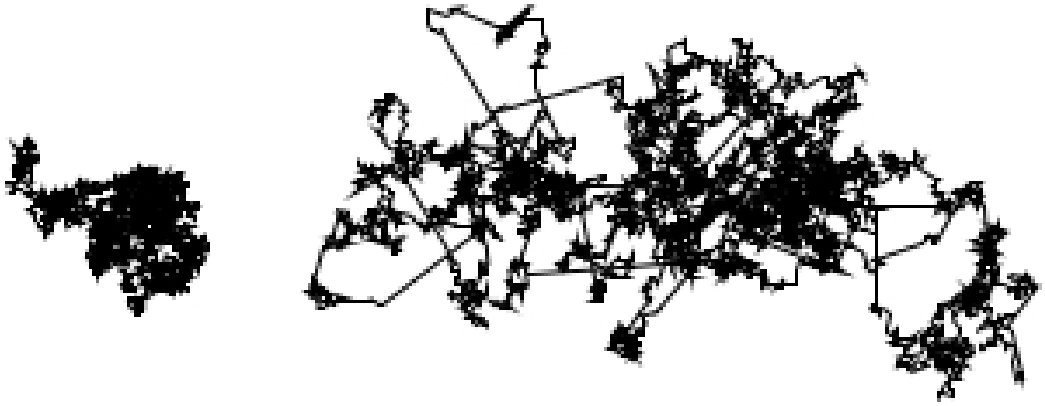
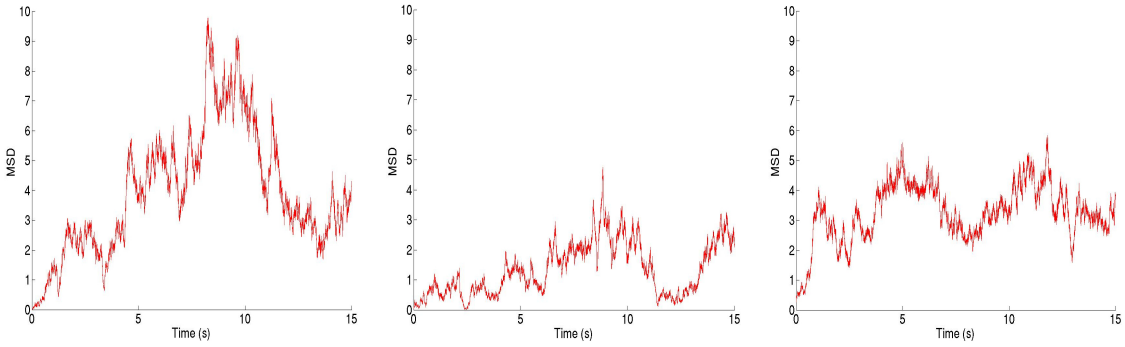
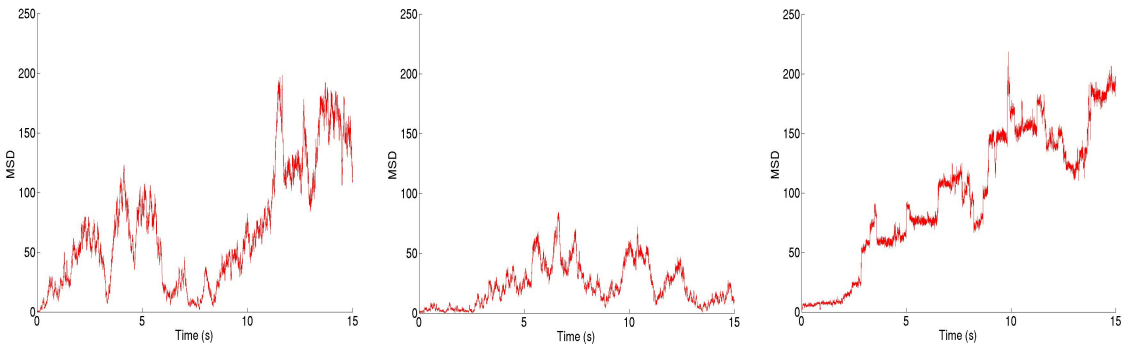


Figure 3.13: Comparison of a typical Brownian trajectory (left) with a typical Lévy flight trajectory (right).



(a) $\theta = 0$ (b) $\theta = 0.3$ (c) $\theta = 0.6$

Figure 3.14: PTEN mRNA MSD under varying volume fractions and fast diffusion. PDifc = NORM. Obstacles, where present, are mobile.



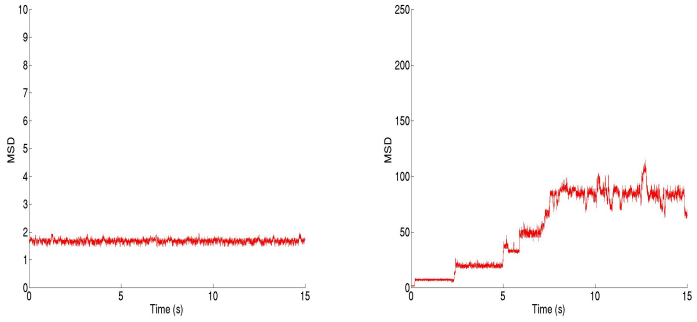
(a) $\theta = 0$ (b) $\theta = 0.3$ (c) $\theta = 0.6$

Figure 3.15: PTEN mRNA MSD under varying volume fractions. PDifc = FAST. Obstacles, where present, are mobile.

obstacles. The most obvious difference between mobile and stationary obstacles is shown in Figure 3.16(a) and Figure 3.14(c); In Figure 3.16(a), mRNAs appear to be unable to move between a high volume of stationary obstacles when mRNA diffusion is normal. This is not a surprising result, but it does highlight the impact of mobile obstacles in allowing some movement of molecules throughout the domain as the excluded volume is somewhat "fluid" and obstacles/molecules can be moved upon collision with other obstacles/molecules. When mRNAs are able to diffuse much faster, we see trajectories similar to the case with mobile obstacles except that the total distance travelled is lower for $\theta = 0.6$ (Figure 3.16(b)). It should be noted here that the maximum of an MSD curve only shows the maximal distance that a particle was able to travel away from its initial position, and does not show the total distance travelled by the molecule, which we demonstrate in Section 3.5.1.2. From Figure 3.16(b), we see that the MSD steps are more defined and the trapping persists for longer compared with Figure 3.15(c), and that the maximum MSD from the initial position is smaller as a result of this trapping.

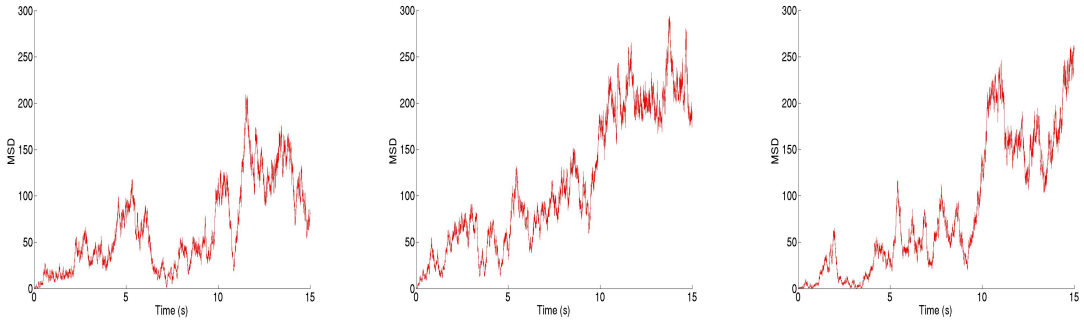
In Figure 3.17, showing miRNA MSD, we see no major difference for varying volume fraction ($\theta = 0, 0.3$ or 0.6) with mobile obstacles. This behaviour is expected as miRNAs are small enough, and diffuse quickly enough, to pass freely between even a high volume of obstacles. The same is true for stationary obstacles, which do not show any definitive signs of trapping. Again, we must clarify that these graphs show the MSD for a single trajectory, and that the particle is also affected by confinement due to the finite-sized impermeable domain.

Figure 3.18 shows the average MSD over 10,000 trajectories for volume fraction values of $\theta = 0$ and



(a) PDifc = NORM (b) PDifc = FAST

Figure 3.16: PTEN mRNA MSD under varying volume fractions for stationary obstacles. $\theta = 0.6$, stationary obstacles.

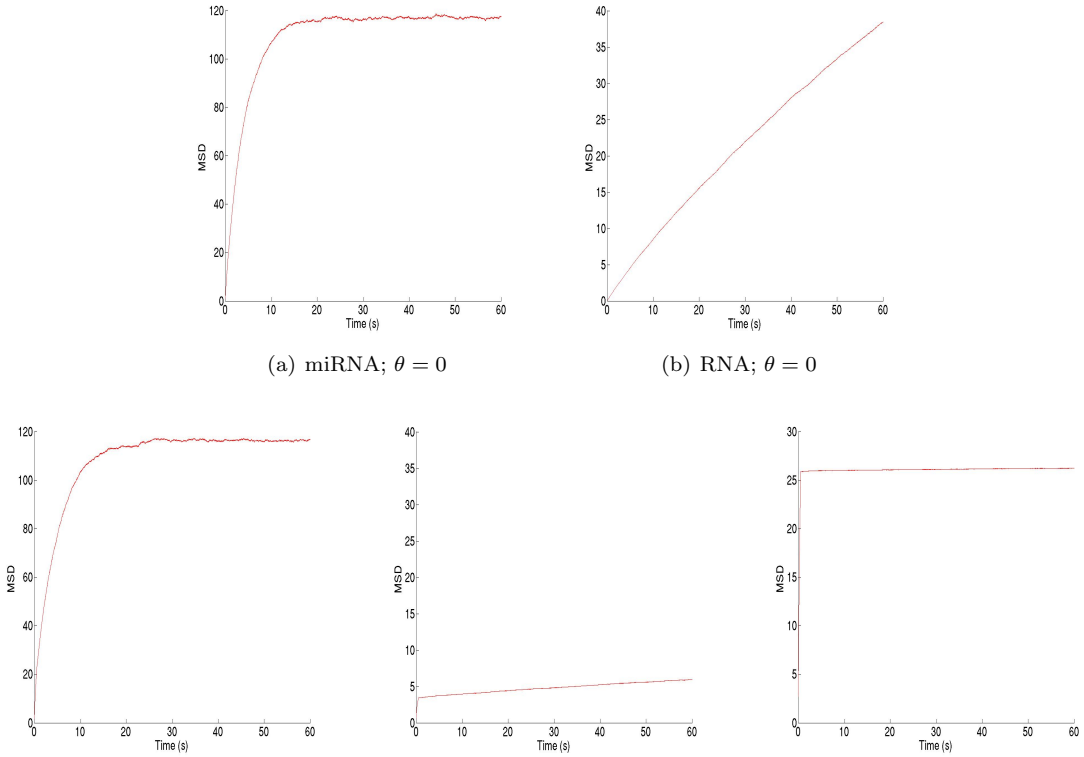


(a) $\theta = 0$; Mobile (b) $\theta = 0.6$; Mobile (c) $\theta = 0.6$; Stat

Figure 3.17: MicroRNA MSD under varying volume fractions for mobile and stationary obstacles.

0.6. Again, please note the varying scales on the y-axis. The purpose of these graphs is to show the (approximately) linear behaviour of the MSD for RNAs and crowders, and the more anomalous and confined behaviour of the MSD for miRNAs.

Figure 3.19 shows the distance travelled by miRNAs and mRNAs over time periods of $\Delta t = 0.15s$ (for time step, $\delta t = 0.0001$s) for each of the volume fractions. For miRNAs (Figures (3.19(a))-(3.19(c))), we see a mean step length of approximately $27 \pm 5\mu$m. For mRNAs, when diffusion is normal, we see a mean step length of approximately $4.25 \pm 0.5\mu$m. When mRNA diffusion is fast we now see a mean step length of approximately $25.5 \pm 5\mu$m in most cases, i.e. similar to miRNA movement. However, in the cases displaying Lévy flight motion, i.e. $\theta = 0.6$ when mRNAs are diffusing fast, we see the mean step length is reduced to 21μm, whereas the range is doubled to $\pm 10\mu$m. This is as we would expect for Lévy motion, as the particle undergoes long periods of compact diffusion, followed



(a) miRNA; $\theta = 0$ (b) RNA; $\theta = 0$

(c) miRNA; $\theta = 0.6$ (d) RNA; $\theta = 0.6$ (e) Obs; $\theta = 0.6$

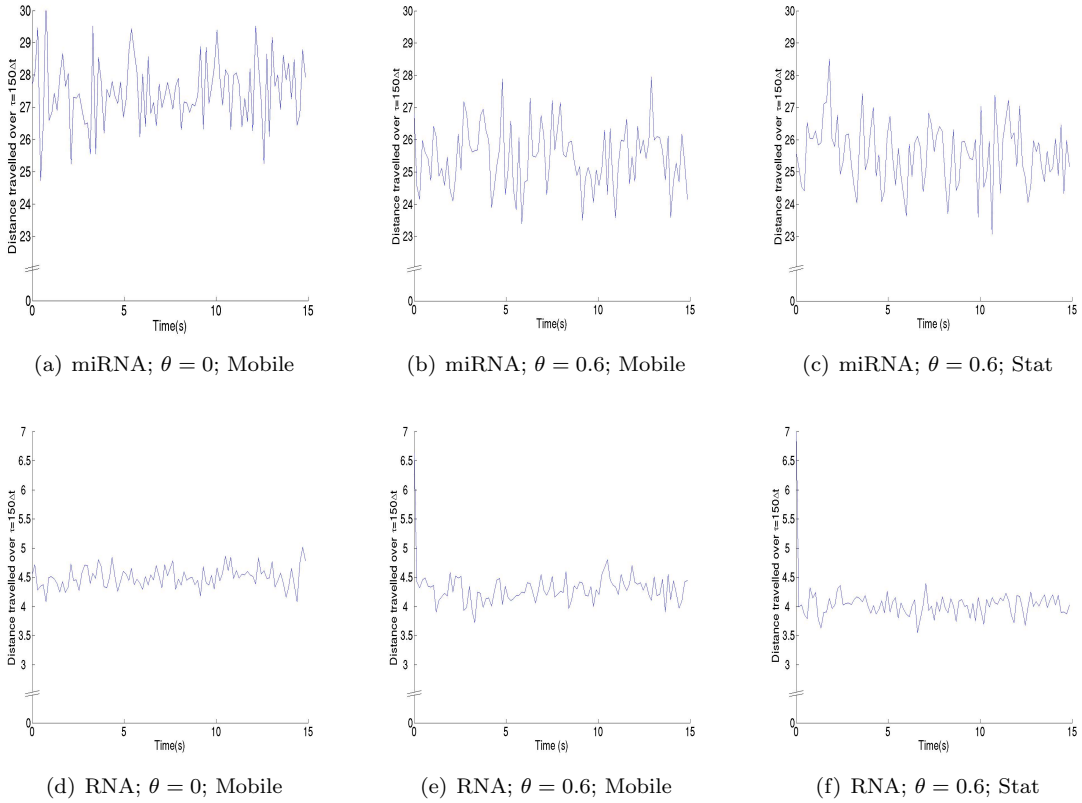
Figure 3.18: MSD for populations of molecules. Obstacles, where present, are mobile.

by long-range relocations.

Figure 3.22 is presented to demonstrate the importance in using small time steps when calculating MSD and the total difference travelled. In Figures 3.22(a) and 3.22(b), we see the total distance travelled by an miRNA and an mRNA for the case of normal diffusion in a 30% crowded environment for $\Delta t = 0.15$. We see that the total distance is measured as around 150-times smaller than the distance measured for $\Delta t = 0.001$, shown in Figures 3.22(c) and 3.22(d), as we might expect from using 150-times fewer data points in the calculations.

3.5.1.3 Effect of Obstacle Size and Molecule Size

In Figure 3.23, we show the trajectories of miRNAs when their size is increased (10-fold) but their original diffusion coefficient is maintained (we call this "NORMBIG" for normal diffusion but big size). It is worth noting here that although the miRNA size has increased 10-fold, they are still 10-



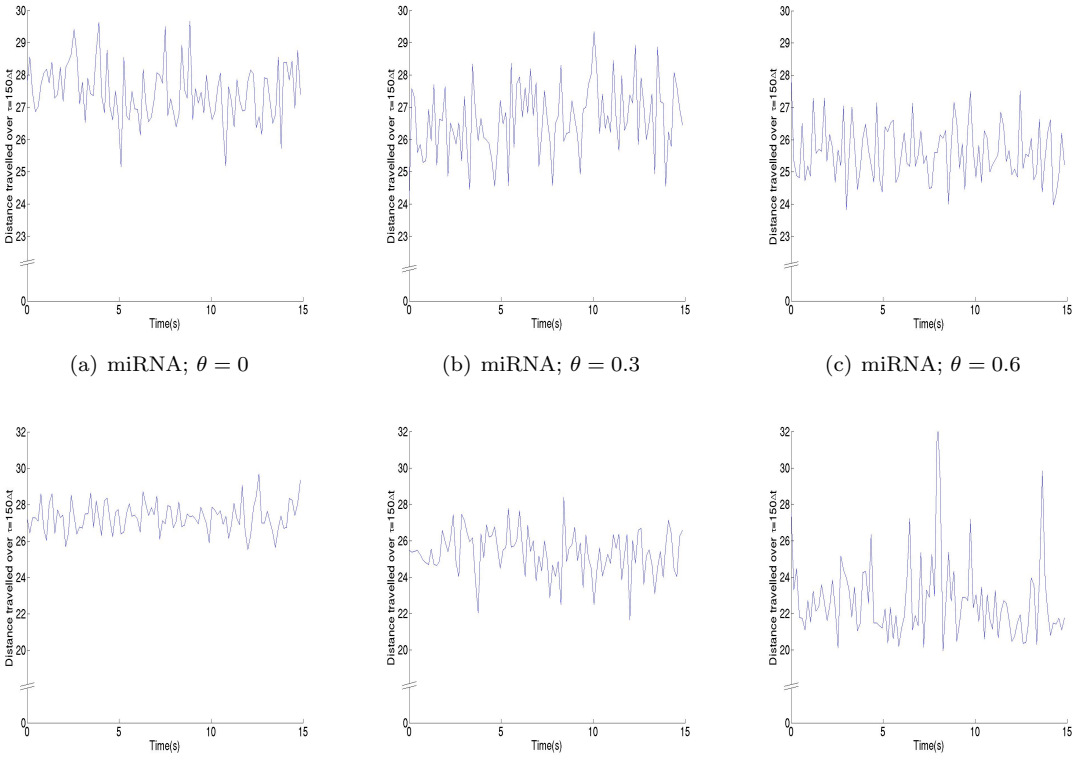
(a) miRNA; $\theta = 0$; Mobile (b) miRNA; $\theta = 0.6$; Mobile (c) miRNA; $\theta = 0.6$; Stat

(d) RNA; $\theta = 0$; Mobile (e) RNA; $\theta = 0.6$; Mobile (f) RNA; $\theta = 0.6$; Stat

Figure 3.19: Distance travelled by miRNAs and mRNAs over time periods of $\Delta t = 0.15s$ (for time step, $\delta t = 0.0001$s) for varying volume fractions.

fold smaller than mRNAs and 20-fold smaller than obstacles. We can see that miRNA movement is more restricted when the space is crowded ($\theta = 0.6$), particularly for stationary obstacles, although still less restricted than mRNAs. As a result of the increased miRNA size, miRNAs become trapped for short periods of time, tending towards a less-defined Lévy-type motion.

In Figure 3.24, miRNAs are again 10-fold larger in size, but this time the diffusion coefficient is representative of its larger size (we call this "SLOWBIG" for slow diffusion and big size). We can see here a drastic reduction in the freedom of miRNA movement for both crowded and non-crowded environments. This scenario of miRNAs being the same size as mRNAs is not biologically realistic, however it demonstrates that the spatial simulations can be pushed to a limit where the volume fraction effects on molecular binding are more prominent. Results of these "SLOWBIG" simulations could be applied to other biological systems where the two reactive molecular species are of a similar size.



(a) miRNA; $\theta = 0$ (b) miRNA; $\theta = 0.3$ (c) miRNA; $\theta = 0.6$

(d) RNA; $\theta = 0$ (e) RNA; $\theta = 0.3$ (f) RNA; $\theta = 0.6$

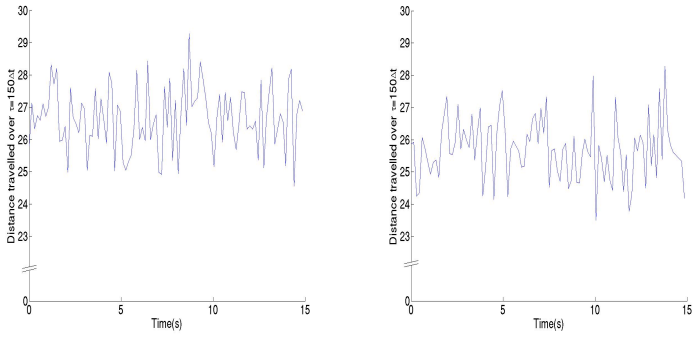
Figure 3.20: Distance travelled by miRNAs and mRNAs over time periods of $\Delta t = 0.15s$ (for time step, $\delta t = 0.0001$s) for varying volume fractions and fast diffusion of obstacles. PDifc = FAST. Obstacles, where present, are mobile.

3.5.2 Effect on PTEN mRNA Degradation

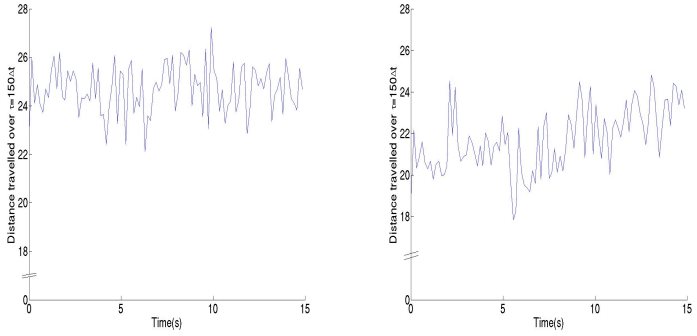
The following graphs show simulation results using the number of degraded PTEN mRNAs at the end of simulation ($t = 5$s) as a measurement for the tumourigenic behaviour of the system (fewer PTEN mRNAs $\implies$ greater tumourigenicity).

3.5.2.1 Effect of Obstacle Size and Molecule Size

In order to model the effects of obstacle size in Smoldyn, we fix the size of obstacles and instead vary the size of miRNAs. By doing this, we prevent increasing the computational cost by keeping the number of obstacles constant. If we were to decrease the size of obstacles, we would require a much greater number of small obstacles in order to maintain the volume fraction.



(a) miRNA; $\theta = 0.3$ (b) miRNA; $\theta = 0.6$



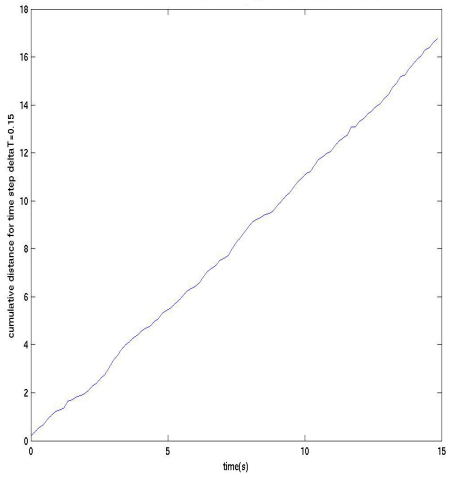
(c) RNA; $\theta = 0.3$ (d) RNA; $\theta = 0.6$

Figure 3.21: Distance travelled by miRNAs and mRNAs over time periods of $\Delta t = 0.15s$ (for time step, $\delta t = 0.0001$s) for varying volume fractions and stationary obstacles. PDifc = FAST. Stationary obstacles.

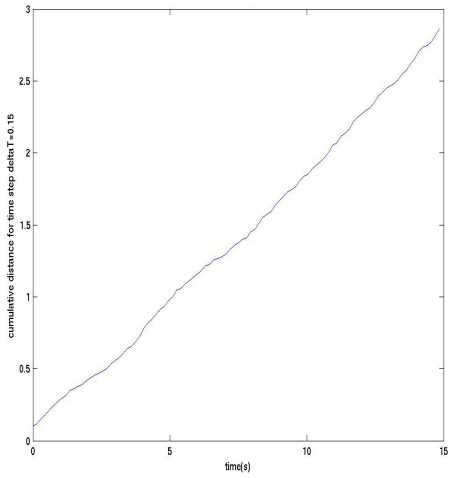
We consider obstacles to affect molecules in three ways:

- by reducing the space for molecules to diffuse, preventing free passage through the domain;
- by interacting ballistically with other molecules, thereby relocating each molecule to a distance proportional to the relevant diffusion coefficients;
- via electrostatic interactions with molecules of a similar size, temporarily preventing them from participating in other reactions.

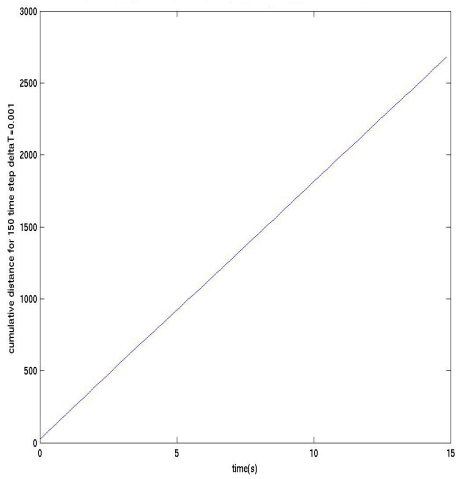
In Figure 3.25(a), we compare the effect of macromolecular crowding on mRNA degradation. We see that most mRNA molecules have been degraded by $t = 2.5$s, in both the dilute and highly crowded ($\theta = 0.6$) environment. This is most likely due to the freedom of miRNA movement throughout the domain, despite the obstacles (shown in the MSD results). The same is true for stationary compared with mobile obstacles, as shown in Figure 3.25(b). This is particularly true for this case



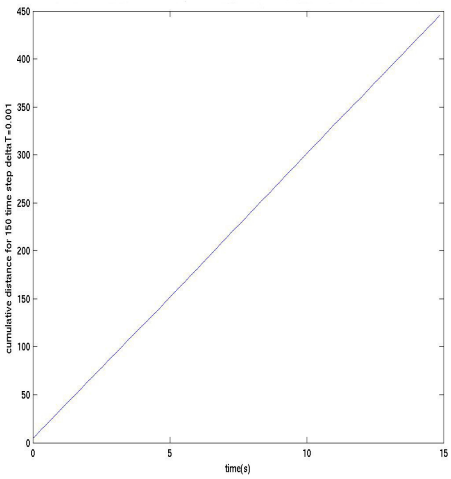
(a) miRNA, $\theta = 0.3$, mobile obs, mRNA fast difc: δt=0.15



(b) RNA, $\theta = 0.3$, mobile obs, mRNA fast difc: δt=0.15

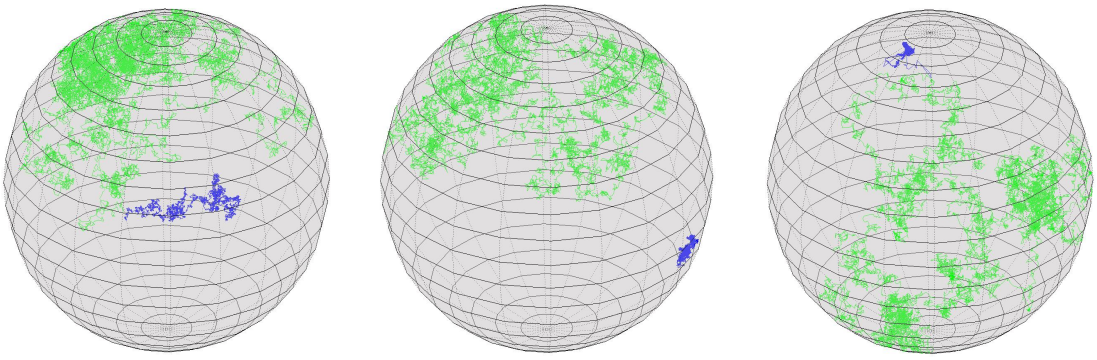


(c) miRNA, $\theta = 0.3$, mobile obs, mRNA fast difc: δt=0.001



(d) RNA, $\theta = 0.3$, mobile obs, mRNA fast difc: δt=0.001

Figure 3.22: Cumulative distance travelled by miRNAs and mRNAs for $\Delta t = 0.15$ (top row) and $\Delta t = 0.001$ (bottom row). Notice the different y-axis scales.

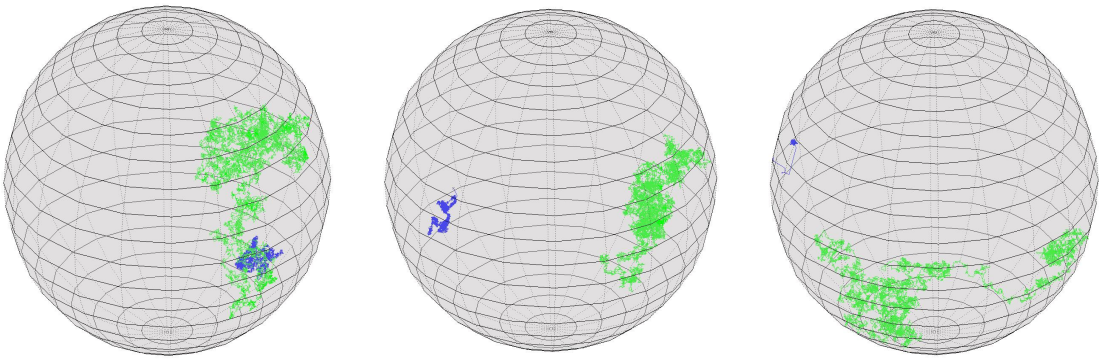


(a) $\theta = 0$ (b) $\theta = 0.6$; Mobile (c) $\theta = 0.6$; Stat

Figure 3.23: Example trajectories of RNAs and miRNAs ($t_f = 15s$), where miRNAs are "NORMBIG". Shown are a single miRNA and a single mRNA. In (b) and (c), obstacles ($\theta = 0.6$) are present, but not shown. Blue = RNA; Green = miRNA. $D_R = 0.04\mu m^2 s^{-1}$, $D_m = 6.67\mu m^2 s^{-1}$. $R_m = R_p = 0.15\mu$m. Movement of miRNAs is slightly restricted when their size is increased but their diffusion coefficient is unchanged.

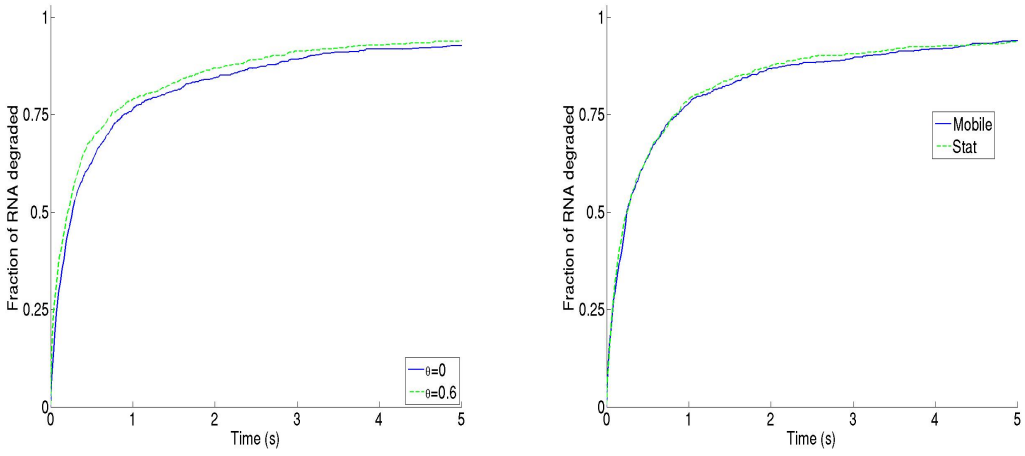
with a "HIGH" molecular count ($N_p = N_m = 776$). It can also be seen for "LOW" molecular count ($N_p = N_m = 100$) with a slight stochasticity in the behaviour, which we will see in the remainder of our results. In order to see any effect of crowding on the number of degraded mRNAs, the movement of miRNAs must be impeded. We induce this by increasing miRNA size and/or decreasing the miRNA diffusion coefficient.

In Figure 3.26(a), we compare the effects of "huge" obstacles ($R_o = 1\mu$m) with the normal (although throughout the thesis we refer to them as "large") obstacles ($R_o = 0.32\mu$m). We do not see any noticeable difference, most likely because miRNAs are considered to be "NORMSMALL" and can still diffuse freely in the space between obstacles (we call this "NORMSMALL" for normal diffusion and small size). Figure 3.27 shows the mRNA degradation reaction rates when the size of miRNAs is increased and the diffusion coefficient is either the same as before ("NORMBIG"; Figure 3.27(a)) or adjusted for its new size ("SLOWBIG"; Figure 3.27(b)). We can see from this figure that increasing the size of miRNAs and reducing its diffusing coefficient ("SLOWBIG"), reduces the number, and rate of, mRNA degradation (Figure 3.27(b)). We also show the effect of adding copycats to the system, which further reduces the number of mRNAs degraded. This is particularly clear when big miRNAs are diffusing normally ("NORMBIG"; Figure 3.27(a)), but less so when big miRNAs are diffusing slowly ("SLOWBIG"). This is likely due to the fact that copycats also diffuse more slowly.



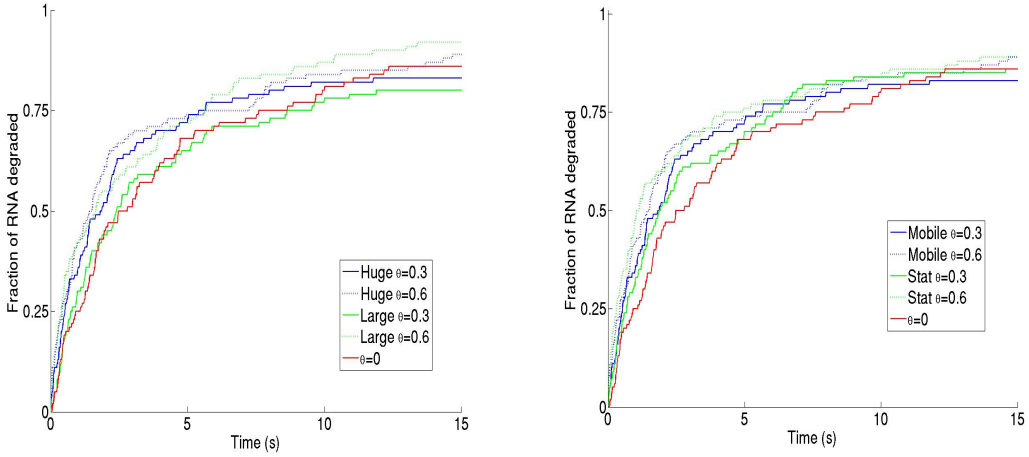
(a) $\theta = 0$ (b) $\theta = 0.6$; Mobile (c) $\theta = 0.6$; Stat

Figure 3.24: Example trajectories of RNAs and miRNAs ($t_f = 15s$), where miRNAs are "SLOW-BIG". Shown are a single miRNA and a single mRNA. Obstacles ($\theta = 0.6$) are present, but not shown, in Figures 3.24(b) and 3.24(c). Blue = RNA; Green = miRNA. $D_m = 0.04\mu m^2 s^{-1} = D_R$. $R_m = R_p = 0.15\mu$m. Movement of miRNAs is much more restricted when their size is increased and their diffusion coefficient is reduced.



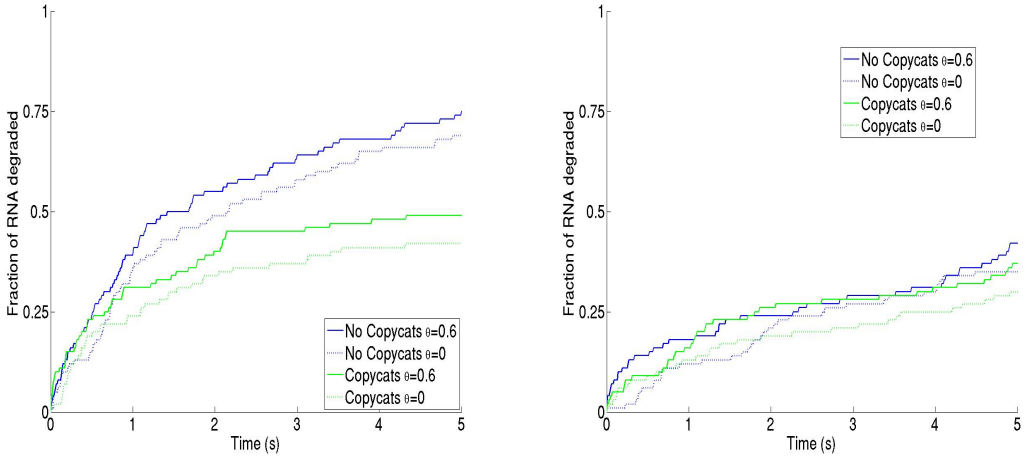
(a) $\theta = 0$ vs. $\theta = 0.6$ (b) Mobile vs. Stat: θ=0.6

Figure 3.25: Fraction of the PTEN mRNA population degraded comparing volume fraction (3.25(a)) and obstacle mobility 3.25(b). PTEN = miRNA = "HIGH".



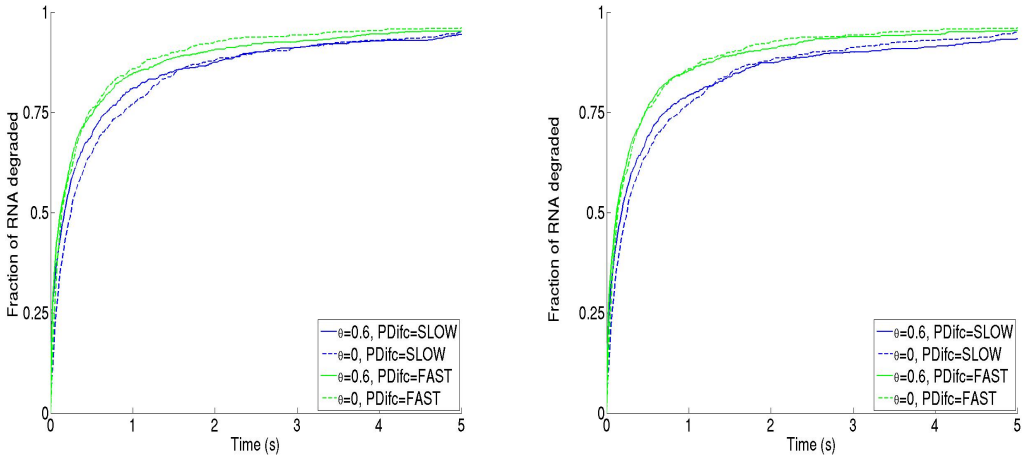
(a) Huge vs. Large (b) Mobile vs. Stat

Figure 3.26: Fraction of the PTEN mRNA population degraded comparing obstacle size (3.26(a)) and mobility (3.26(b)) for a range of volume fractions ($\theta = 0, 0.3$ and 0.6). PTEN = miRNA = "LOW". miRNAs = "NORMSMALL". Data shown is the average of 20 simulations.



(a) miR difc = "NORM" (b) miR difc = "SLOW"

Figure 3.27: Fraction of the PTEN mRNA population degraded comparing miRNA diffusion coefficents for increased miRNA size (1.1 "NORMBIG"; 1.1 "SLOWBIG") for a range of volume fractions ($\theta = 0$ and 0.6). PTEN = siRNA = "LOW". $R_m = R_p = 0.15\mu$m. PTEN = miRNA = "LOW". Data shown is the average of 20 simulations.



(a) Stat (b) Mobile

Figure 3.28: Fraction of the PTEN mRNA population degraded for big miRNAs with unchanged diffusion coefficient ("NORMBIG"; $D_m = 6.67\mu m^2 s^{-1}$; $R_m = R_p = 0.15\mu$m). PTEN mRNA diffusion differs between normal ("SLOW"; $D_p = 0.04\mu m^2 s^{-1}$) and fast ("FAST"; $D_p = 6.67\mu m^2 s^{-1}$). PTEN = miRNA = "HIGH". $\theta = \{0, 0.3, 0.6\}$. Data shown is the average of 20 simulations.

Figure 3.28 shows the mRNA degradation rate when the size of miRNAs is increased, but the diffusion coefficient remains unchanged ("NORMBIG"). Figures 3.28(a) and 3.28(b) compares stationary and mobile obstacles for both fast ($N_p = N_m$) and slow mRNA, with changing volume fraction. We see a slight increase in mRNA degradation rate, and the total amount degraded, when mRNAs are moving faster, which we expect as they are capable of exploring more of the domain, thereby increasing the chance of interaction with miRNAs. We do not see much difference between stationary and mobile obstacles.

Figure 3.29 shows the same as Figure 3.28 with the addition of copycats ($N_d(0) = N_m(0)$). Figure3.29 demonstrates the difference in degradation rate for either fast or slow moving PTEN mRNAs. We see from Figure 3.29 that, when mRNAs are diffusing normally, there is a slightly higher reaction rate for mobile over stationary obstacles, and for 0% over 60% crowding. However, when mRNAs are moving more rapidly ($D_p = D_m$), this trend reverses such that stationary obstacles and a bigger volume fraction lead to higher degradation rates. Whilst this difference is not large, it raises the question of how mobile obstacles interfere with the diffusion, and subsequent reaction, of molecules. Two possible explanations for hindered movement by mobile obstacles are that ballistic interactions between obstacles and mRNAs interrupt their intended trajectories, or that mobile obstacles lead

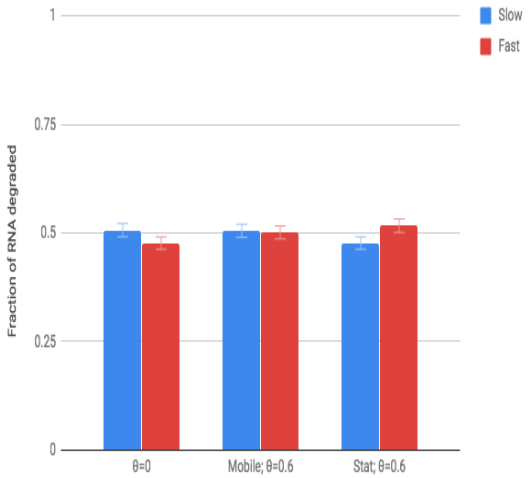


Figure 3.29: Fraction of the PTEN mRNA population degraded for big miRNAs moving with unchanged diffusion coefficient ("NORMBIG"; $D_m = 6.67\mu m^2 s^{-1}$). We compare PTEN mRNA diffusion rates that are normal ("SLOW"; $D_p = 0.04\mu m^2 s^{-1}$) and fast i.e. the same as miRNAs ("FAST"; $D_p = 6.67\mu m^2 s^{-1}$). We show two different volume fractions ($\theta = 0$ and $\theta = 0.6$) and compare mobile and stationary obstacles. We see that, when mRNAs are diffusing normally, the number of degraded PTEN mRNAs is decreased when obstacles are stationary compared to mobile or the case of $\theta = 0$. However, when mRNAs are moving more rapidly ($D_p = D_m$), this trend is reversed such that the number of degraded PTEN mRNAs is higher with stationary obstacles. Data shown is the average of 20 simulations. The error bars shown are $\pm 3\%$ as this was calculated as the approximate range of values from the 20 simulations (data not shown).

to a more complex, "fluid" geometry through which the mRNAs are unable to navigate effectively. Appendix Figures A.1(c) and A.1(d) show these differences by isolating the stationary and mobile obstacle scenarios for both fast and slow mRNA. In both cases, a dilute environment leads to the fewest number of degraded mRNAs.

In Table 3.5.2.1, we demonstrate the effect of changing the size and/or diffusion coefficient of miRNAs for $\theta = 0, 0.3, 0.6$. In each case, decreasing the diffusion coefficient ("SLOW*") decreases the rate and total amount of mRNA degradation, most likely as a result of the reduced freedom of miRNA diffusion. Although the differences are somewhat small, for normal small miRNAs ("NORMSMALL") a highly crowded environment leads to the greatest amount of mRNA degradation, whilst $\theta = 0.3$ leads to the lowest degradation level. Furthermore, the total amount degraded RNA is highest for normal small miRNAs compared with all other scenarios. For slow small miRNAs (we call this "SLOWSMALL" for slow diffusion and small size), this trend is reversed completely and a highly crowded environment leads to the greatest amount of mRNA degradation. For slow big miRNAs

("SLOWBIG"), a dilute space is found to be most favourable, with $\theta = 0.3$ again producing the lowest degradation levels. For big miRNAs diffusing normally ("NORMBIG"), there is almost no difference between the three volume fractions. In Appendix Figures A.2, we further demonstrate these effects of changing the size and/or diffusion coefficient of miRNAs by looking at each case in isolation for $\theta = 0, 0.3, 0.6$. Data shown is the average of 20 simulations. Average minimum and maximum values are given by -1.52% and $+1.41\%$ of the average value, so the differences between "NORM*" and "SLOW*" values do not overlap.

Volume Fraction	"NORMBIG"	"NORMSMALL"	"SLOWBIG"	"SLOWSMALL"
$\theta = 0$	94.33%	93.30%	84.79%	83.63%
$\theta = 0.3$	93.69%	93.04%	84.54%	83.51%
$\theta = 0.6$	93.81%(	91.75%	83.76%	85.18%

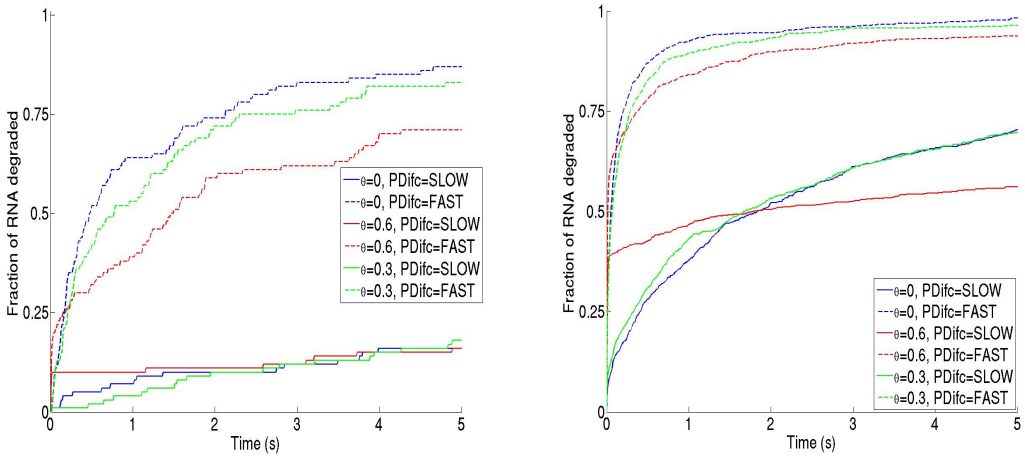
In Figure 3.30, we show the reaction rates when miRNAs have the same size, diffusion coefficient, and binding radius as mRNAs. We see a visible effect of volume fraction on the amount of PTEN mRNA degradation when PTEN mRNAs and miRNAs are diffusing quickly ("FAST"), particularly between $\theta = 0$ and $\theta = 0.6$. These effects are less clear when diffusion is "SLOW", where reaction events are less likely. Interestingly, we see an incredibly fast initial rate of reaction when $\theta = 0.6$. We believe this is due to the initial close proximity of miRNAs and mRNAs, due to the limited available space for such large molecules. Therefore, these molecules will interact within the first few time steps until there is more free space for the molecules to move around.

3.5.3 Effect of Electrostatic Interactions with Obstacles

In Figure 3.31, we examine the effect of weak electrostatic interactions between obstacles (O) and PTEN mRNAs (P), such that

$$O + P \underset{k_b}{\overset{k_f}{\rightleftharpoons}} OP. \tag{3.20}$$

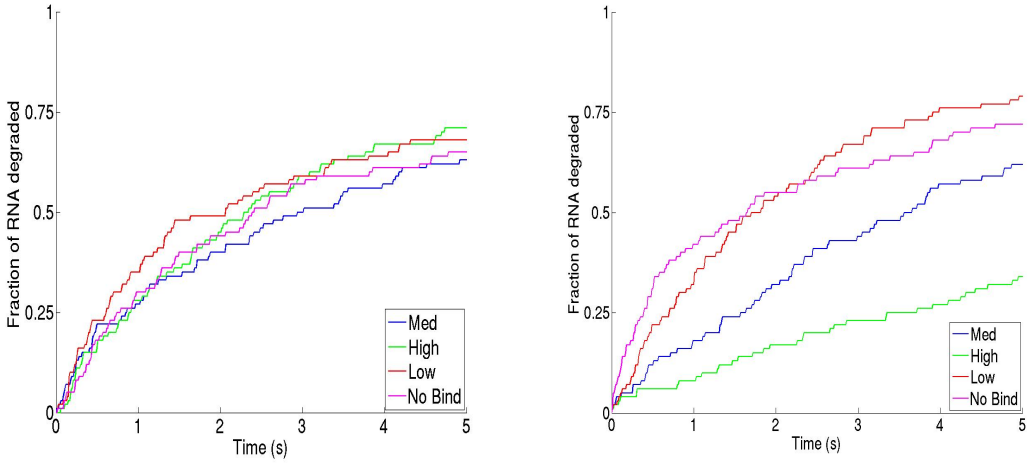
We model the weak electrostatic interactions by applying a fast binding and unbinding rate, k_f and k_b respectively, temporarily preventing the molecules involved from participating in other reactions. For $\theta = 0.3$, we see a slight reduction in RNA degradation when molecules can unbind with the same



(a) "LOW" (b) "HIGH"

Figure 3.30: Fraction of the PTEN mRNA population degraded when both miRNAs and RNAs have $R = 0.15\mu$m, i.e. RNA-RNA interactions. Diffusion coefficents are equal and are either normal ("SLOW"; $D = 0.04\mu m^2 s^{-1}$) or fast ("FAST"; $D = 6.67\mu m^2 s^{-1}$). Molecular count is either "LOW" (as in 3.30(a)) or "HIGH" (as in 3.30(b)). $\theta = \{0, 0.3, 0.6\}$. Data shown is the average of 20 simulations.

probability as they unbind (Med $\implies k_f = 0.01 = k_b$). Conversely, the rate of RNA degradation increases when there is a higher unbinding probability (Low $\implies k_f = 0.01$, $k_b = 0.1$). By increasing the number of obstacles with which RNAs can bind ($\theta = 0.6$), we see these trends more clearly. In general, the higher the trapping probability, the lower the amount, and rate, of RNA degradation. For the highest trapping probability (High $\implies k_f = 0.1$, $k_b = 0.01$), we see a reduction by nearly 50% in the final amount of degraded RNA, presumably because fewer RNAs are available for degradation by miRNAs. An unexpected result in both cases, is that the rate of degradation for low trapping probabilities actually exceeds that of the case with no binding. The cause for this is unclear at this stage and will require futher investigation. It is possible RNAs are moved around the space by binding to obstacles and then unbinding elsewhere in the domain, thereby reducing the time spent trapped between obstacles. RNAs are therefore able to move around the space by binding inside a "trap" and unbinding outside of the trap, making interactions with miRNAs more likely.



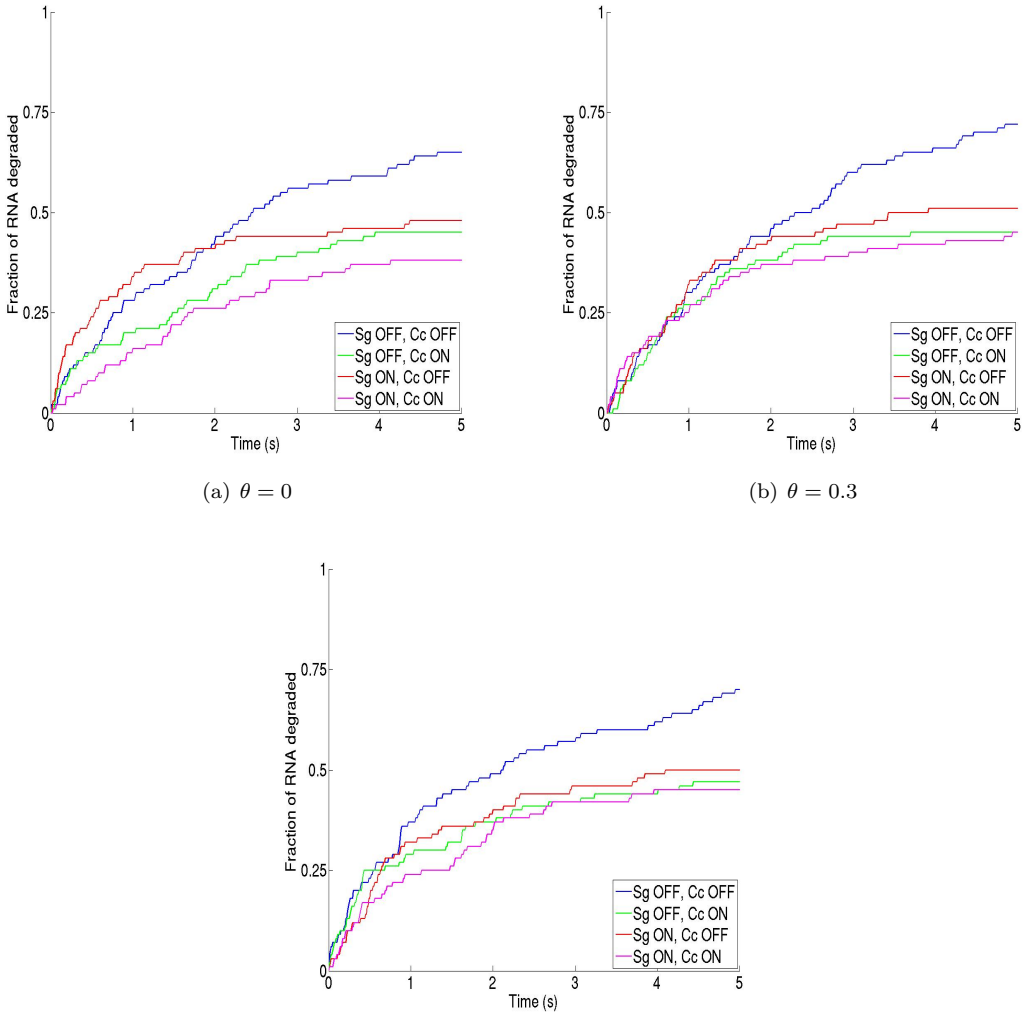
(a) $\theta = 0.3$ (b) $\theta = 0.6$

Figure 3.31: Fraction of the PTEN mRNA population degraded when RNAs may undergo weak binding to obstacles. Low $\implies$ $k_f = 0.01$, $k_b = 0.1$; Med $\implies$ $k_f = 0.01 = k_b$; High $\implies$ $k_f = 0.1$, $k_b = 0.01$. $\theta = 0.3$ and 0.6. PTEN = miRNA = "LOW". Data shown is the average of 20 simulations. The final values at $t = 5s$ varies by around $\pm 3\%$ over these 20 simulations, so the difference that we see between the No binding, Mid, and High binding cases is clear, whereas the Low binding and No binding cases are not clearly different.

3.5.4 Effect of Copycats, CeRNAs and Binding Site Reversibility

In Figure 3.32, we show the impact of adding copycats or ceRNAs (or both) to the original system with "HIGH" ($N_*(0) = 776$) and "LOW" ($N_*(0) = 100$) molecule count. We see clearly that adding either of these regulatory molecules results in a decrease in the number of degraded RNAs. The best scenario, in terms of lowest RNA degradation, comes from the addition of ceRNAs and copycats together. If copycats are not present in the system, introducing them in isolation decreases the levels of RNA degradation in all cases. Furthermore, adding ceRNAs without copycats also reduces the total number of degraded RNAs (although the initial rate of degradation may increase as in Figure 3.32(a)). These additions are particularly pronounced for the dilute system, whereas for the crowded system ($\theta = 0.3$ or 0.6), the differences between adding ceRNAs, copycats or both are less clear.

At this point, we introduce the effects of multiple binding sites, beginning with three binding sites on mRNAs ($N_b = 3$) with a threshold value of $T = N_b$, meaning that all of the mRNA sites must be bound to by miRNAs in order to be degraded. We include copycats and study the effects of changing the binding site reversibility. Later we include ceRNAs with multiple binding sites as well.



(a) $\theta = 0$

(b) $\theta = 0.3$

(c) $\theta = 0.6$

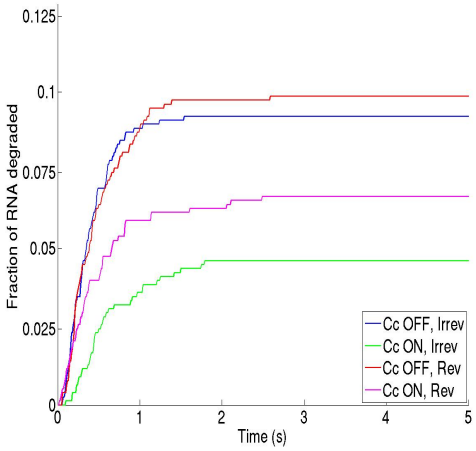
Figure 3.32: Fraction of the PTEN mRNA population degraded when ceRNAs are present (Sp ON) or absent (Sp OFF), and when copycats are present (Cc ON) or absent (Cc OFF) for $\theta = 0, 0.3$ and 0.6. PTEN = miRNA = "LOW". Data shown is the average of 20 simulations. Given that the range of values over these simulations is around $\pm 4\%$, the difference seen between the case when "Sg = OFF and Cc = OFF" and the other three cases is the only one that would be considered notable.

In Figure 3.33(a), reversibility leads to slightly more RNA degradation in the dilute case ($\theta = 0$) without copycats (Cc OFF), with reversibility more clearly preferred in the case with copycats (Cc ON). With the inclusion of crowding (Figures 3.34(a) and 3.33(b)), reversibility is preferred in all cases and the presence of copycats always results in a reduction of total RNA degradation by approximately 50%. Overall, with $N_b = T = 3$, the number of degraded RNAs is reduced approximately 10-fold compared to $N_b = 1$.

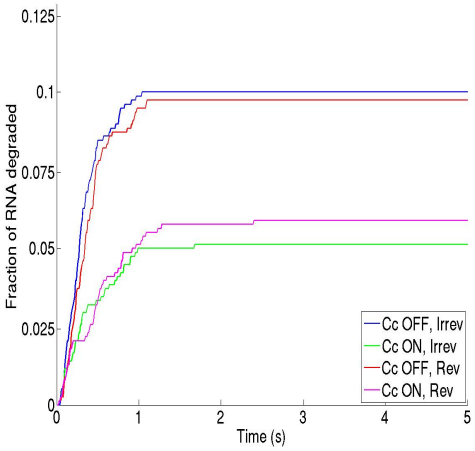
In Figure 3.35, we introduce ceRNAs with two MREs ($N_bC = 2$), whilst mRNAs only have one MRE ($N_b = 1$). In this case, reversibility of binding results in greater RNA degradation, which is a difficult result to predict when copycats are present because they can also unbind. We also see here that the total number of degraded mRNAs is similar to the case for one binding site on both mRNAs and ceRNAs ($N_bC = 1 = N_b$), which suggests that increasing the number of binding sites on ceRNAs does not alter mRNA degradation, except in the case of reversible binding sites. It also confirms that the large decrease in the number of degraded RNAs seen in Figure 3.33 and Figure 3.34 (and later in Figure 3.36) is due to the increase in the threshold T for binding to PTEN mRNAs. We can also see here, particularly in Figures 3.35(c) and 3.35(d), that there seems to be a reduced amount of degradation when the space is crowded. This, again, is likely due to the restrictions on molecular diffusion.

We now look at the case where both mRNAs and ceRNAs have two binding sites ($N_b = N_bC = 2$). Figure 3.36(a), with "HIGH" molecule count, shows a clear distinction between the amount of degraded RNA when adding copycats and changing reversibility. Specifically, without copycats (Cc OFF), reversibility is preferred. This is most likely because unbinding of miRNAs decreases the chance that the threshold value ($T = 2$) will be reached. When copycats are present (CC ON), the number of degraded RNAs is drastically reduced, particularly for the irreversible case. This is expected as binding sites can be permanently blocked by copycats making the threshold value even harder to reach. For "LOW" molecule count, shown in Figure 3.36(b), this distinction is less clear, and it appears that irreversibility is now favoured when copycats are absent. It should be noted that the total number of degraded RNAs is small, therefore this difference is only a small proportion of the total available mRNAs.

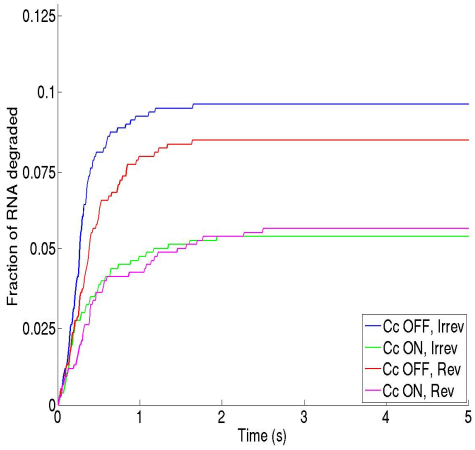
For a very crowded environment (Figure 3.36(c)), the effects of reversibility and irreversibility appear to switch for the case with "HIGH" molecule count, with irreversibility now slightly preferred in the



(a) "HIGH"; $\theta = 0$

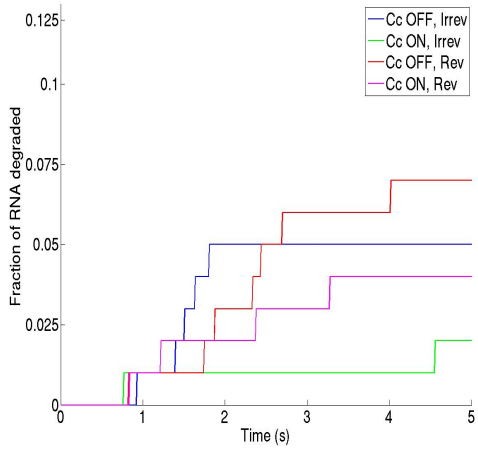


(b) "HIGH"; $\theta = 0.3$

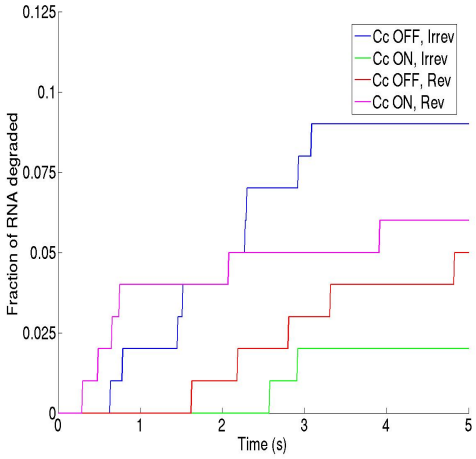


(c) "HIGH"; $\theta = 0.6$

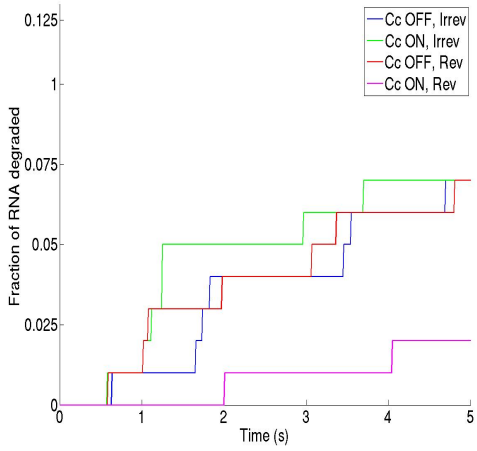
Figure 3.33: Fraction of the PTEN mRNA population degraded when $N_b = 3$, ceRNAs are absent, copycats are either present (Cc ON) or absent (Cc OFF), and binding is either reversible (Rev) or irreversible (Irrev). PTEN = miRNA = "HIGH". $\theta = 0, 0.3$ and 0.6. Data shown is the average of 20 simulations.



(a) "LOW"; $\theta = 0$

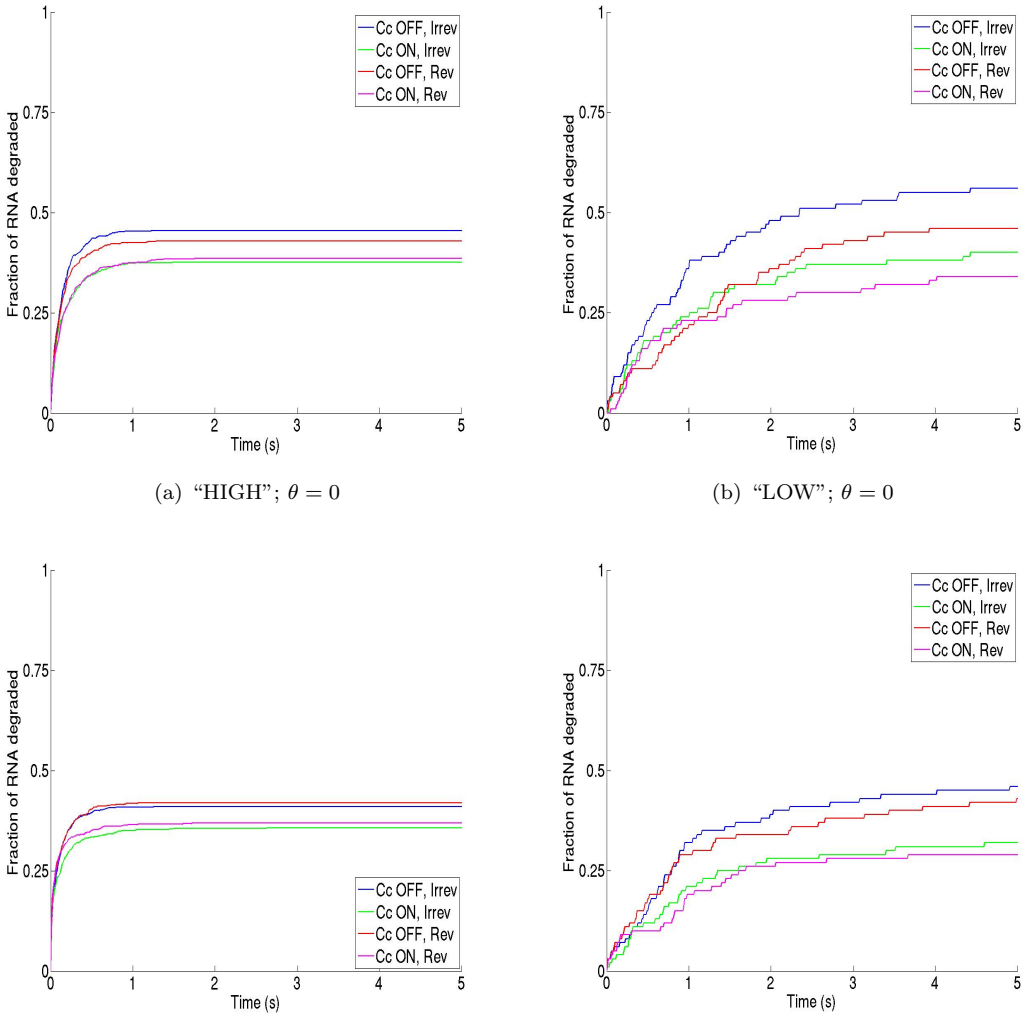


(b) "LOW"; $\theta = 0.3$



(c) "LOW"; $\theta = 0.6$

Figure 3.34: Fraction of the PTEN mRNA population degraded when $N_b = 3$, ceRNAs are absent, copycats are either present (Cc ON) or absent (Cc OFF), and binding is either reversible (Rev) or irreversible (Irrev). PTEN = miRNA = "LOW". $\theta = 0, 0.3$ and 0.6. Data shown is the average of 20 simulations.



(a) "HIGH"; $\theta = 0$

(b) "LOW"; $\theta = 0$

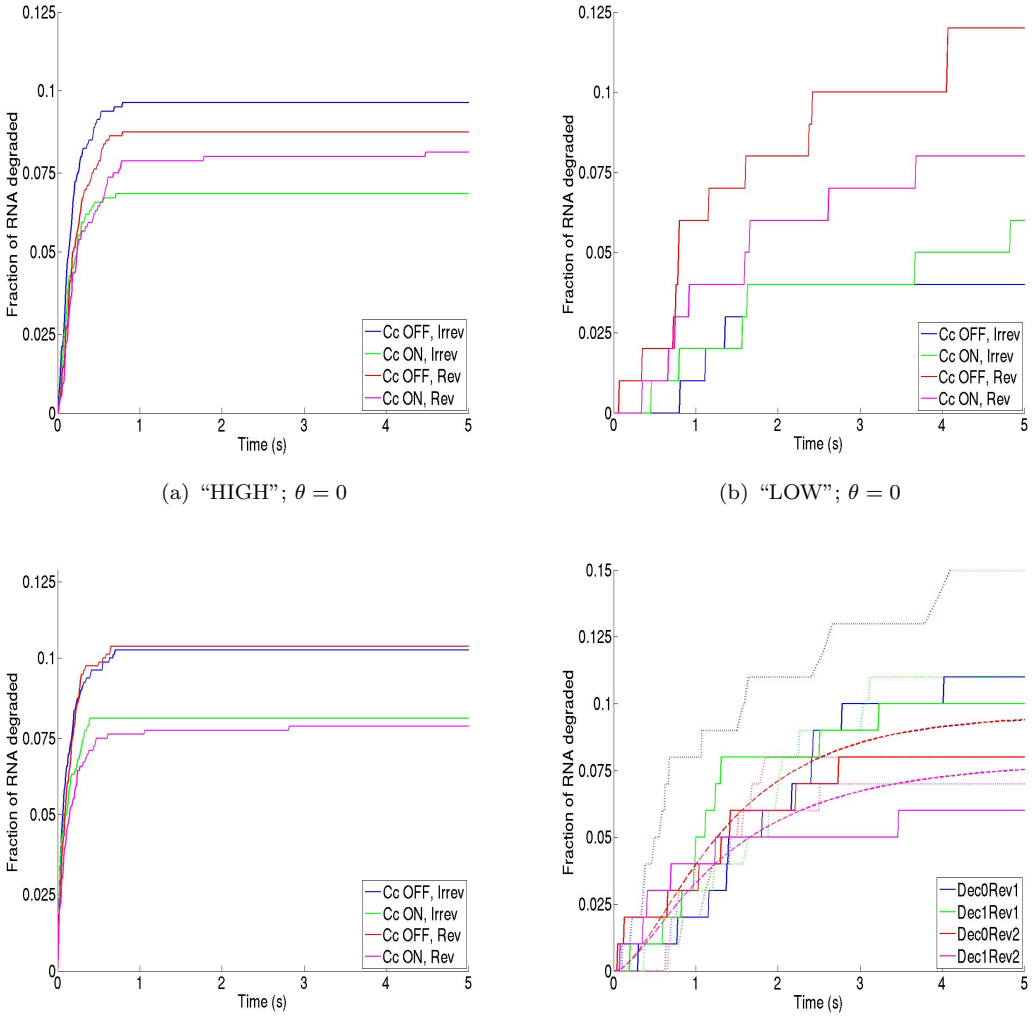
(c) "HIGH"; $\theta = 0.6$

(d) "LOW"; $\theta = 0.6$

Figure 3.35: Fraction of the PTEN mRNA population degraded when ceRNAs are present, $N_b = 1$, $N_bC = 2$, copycats are either present (Cc ON) or absent (Cc OFF), and binding is either reversible (Rev) or irreversible (Irrev). Figures 3.35(a) and 3.35(c) are for PTEN = miRNA = "HIGH". Figures 3.35(b) and 3.35(d) are for PTEN = miRNA = "LOW". $\theta = 0$ or 0.6. Data shown is the average of 20 simulations.

"Cc OFF" case, and reversibility preferred in the "Cc ON" case. It is unclear why this is but it is possible that, in a highly crowded environment, miRNAs/copycats that unbind will rebind almost immediately as there is limited space for such miRNAs/copycats to diffuse away, meaning that reversible and irreversible cases are almost identical.

In Figure 3.37, we show PTEN mRNA degradation rates when miRNA binding is cooperative. Each mRNA and ceRNA has two binding sites ($N_b = 2 = N_bC$) for both miRNAs and copycats and,



(a) "HIGH"; $\theta = 0$

(b) "LOW"; $\theta = 0$

(c) "HIGH"; $\theta = 0.6$

(d) "LOW"; $\theta = 0.6$

Figure 3.36: Fraction of the PTEN mRNA population degraded when ceRNAs are present, $N_b = N_b C = 2$, copycats are either present (Cc ON) or absent (Cc OFF), and binding is either reversible (Rev) or irreversible (Irrev). Figures 3.36(a) and 3.36(c) are for PTEN = miRNA = "HIGH". Figures 3.36(b) and 3.36(d) are for PTEN = miRNA = "LOW". $\theta = 0$ or 0.6. Data shown is the average of 20 simulations.

when one small RNA is bound to a transcript, the binding radius is increased for subsequent small RNAs, effectively increasing the likelihood of the second MRE being bound to. We compare irreversible and reversible binding rates, "NORMSMALL" and "SLOWBIG" small RNAs, and dilute and crowded domains. We see that irreversible binding is preferred when miRNAs are "NORMSMALL", particularly when $\theta = 0$ and the molecule count is "HIGH". This is expected as, although irreversible binding to PTEN mRNAs favours their degradation, irreversible binding of copycats can

permanently block binding sites and decrease the likelihood of reaching $T = 2$. This trend is only inverted when miRNAs are "SLOWBIG", at which point reversible binding is favoured when $\theta = 0.6$ but not when $\theta = 0$. This result is difficult to predict and may be due to the impedance of small RNA movement through the crowded domain, making copycat binding events less likely. Although we could argue that miRNAs are similarly impeded, it is possible that, with binding events even less likely in a crowded environment for "SLOWBIG" small RNAs, any binding event increases the probability of PTEN mRNAs being degraded. Overall, cooperativity only slightly increases the rate of RNA degradation when "NORMSMALL" results are compared with "Cc ON" results from Figure 3.36.

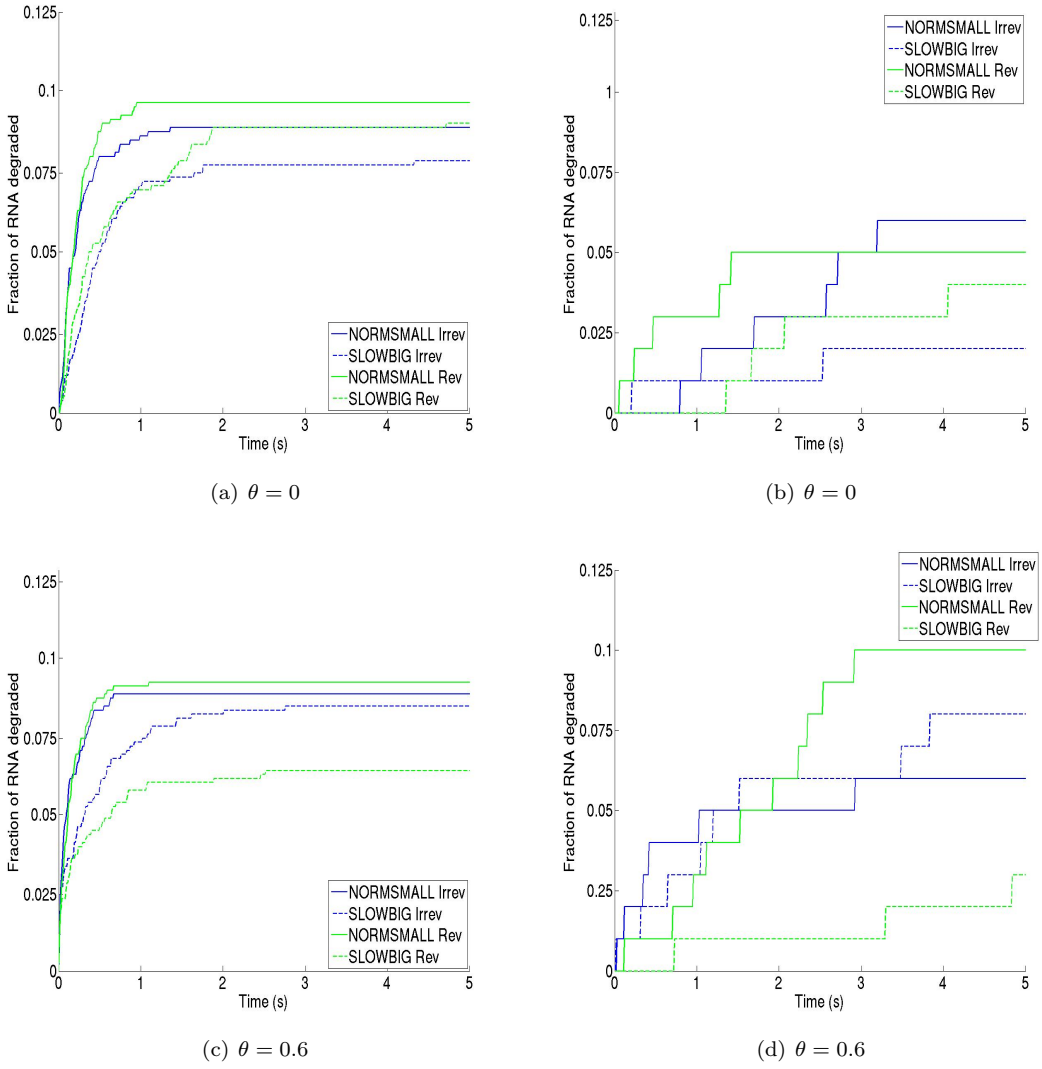
3.5.5 SSA and ODE comparisons

In this section, we repeat some of the above figures with SSA (dotted lines) and ODE (dashed lines) comparisons to Smoldyn (solid lines). As we will see, the SSA and ODE results agree fairly well with Smoldyn results in most cases, with SSA results providing a slightly better match to the Smoldyn results in the case with "LOW" molecular count, as we would expect.

Figure 3.38 shows that there is no difference between the Smoldyn, ODE or SSA model in the simplest case where there are "HIGH" numbers of molecules present and only one irreversible MRE ($N_b = 1$; Figure 3.38(a)). They also show no difference between mobile and stationary obstacles (Figure 3.38(b)), or between "huge" ($R_{obs} = 1\mu$m) and "large" obstacles ($R_{obs} = 0.32\mu$m), for either $\theta = 0$, 0.3 or 0.6 crowding (Figure 3.38(c)).

Figures 3.39 and 3.40 again show that there is no difference between the Smoldyn, ODE or SSA models, for the parameters that we test, in the simplest case when only miRNAs, mRNAs and copycats are present, and only one irreversible MRE ($N_b = 1$).

In Figure 3.41, for the ODE results (dashed lines), the case where copycats are present (Cc ON) but ceRNAs are absent (Sp OFF) completely overlaps with the case where copycats are absent but ceRNAs are present (blue dashed line), showing that ceRNAs and copycats are essentially performing the same function in the ODE model, which is unlikely in reality. However, overall, the SSA and ODE results provide a good match to the Smoldyn results, and show that the best case scenario is one where both ceRNAs and copycats are present. This is difficult to predict, as copycats are able



(a) $\theta = 0$

(b) $\theta = 0$

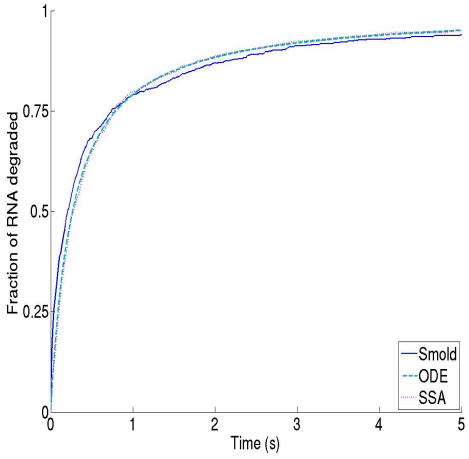
(c) $\theta = 0.6$

(d) $\theta = 0.6$

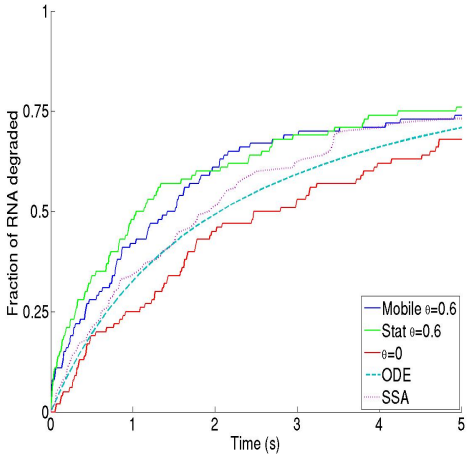
Figure 3.37: Fraction of the PTEN mRNA population degraded when ceRNAs and copycats are present, $N_b = N_b C = 2$, miRNAs vary in size ("*SMALL" and "*BIG") and diffusion coeffients ("NORM*" and "SLOW*"), and binding is either reversible (Rev) or irreversible (Irrev). Figures 3.37(a) and 3.37(c) are for PTEN = miRNA = "HIGH". Figures 3.37(b) and 3.37(d) are for PTEN = miRNA = "LOW". $\theta = 0$ or 0.6. Data shown is the average of 20 simulations.

to bind to ceRNAs as well as mRNAs.

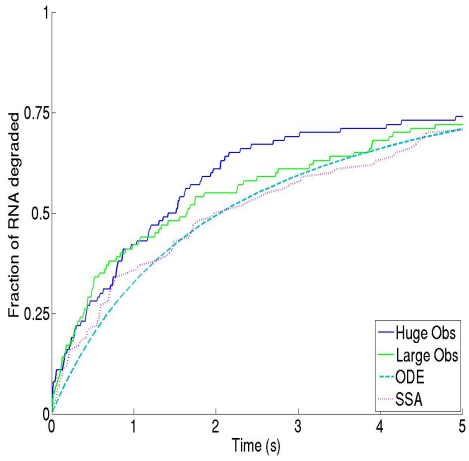
We model additional binding sites on PTEN in the Smoldyn, ODE, and SSA models by creating "new" molecular species to represent the intermediate states of mRNA-miRNA binding. We show this explicitly in the SSA code given in the appendix (Appendix E). Although we did not provide the same code for the ODE model, it is easy to infer the changes needed from the SSA model. In Figure



(a) "HIGH"; $\theta = 0.6$

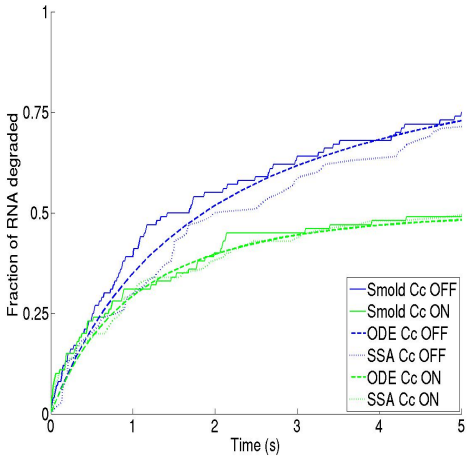


(b) "LOW"; Stationary vs. Mobile

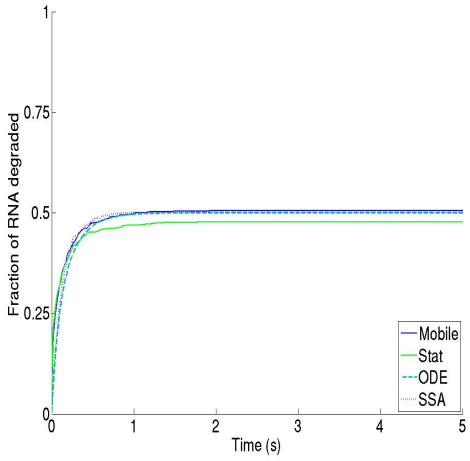


(c) "LOW"; Huge vs. Large

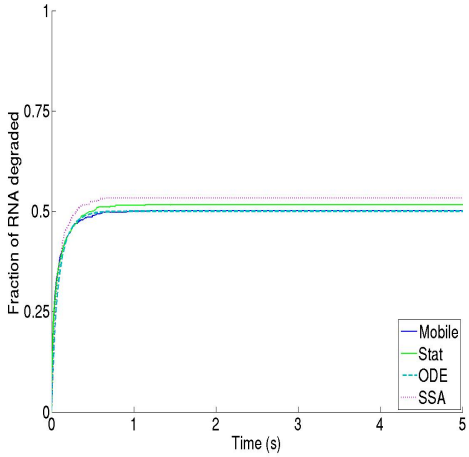
Figure 3.38: Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results. ODE results are shown with a thick dashed line, SSA results are shown with a fine dotted line, and Smoldyn results are shown with a solid line. $N_b = 1$, miRNAs = "NORMSMALL", and binding is irreversible. Figure 3.38(a) is for PTEN = miRNA = "HIGH" and $\theta = 0.6$. Figures 3.38(b) and 3.38(c) are for PTEN = miRNA = "LOW" and compare the size, mobility, and volume fraction of obstacles. Data shown is the average of 20 simulations. We see no visible differences for the simple case without copycats, sponges, multiple MREs, cooperativity, or binding to obstacles.



(a) "LOW" $\theta = 0.6$

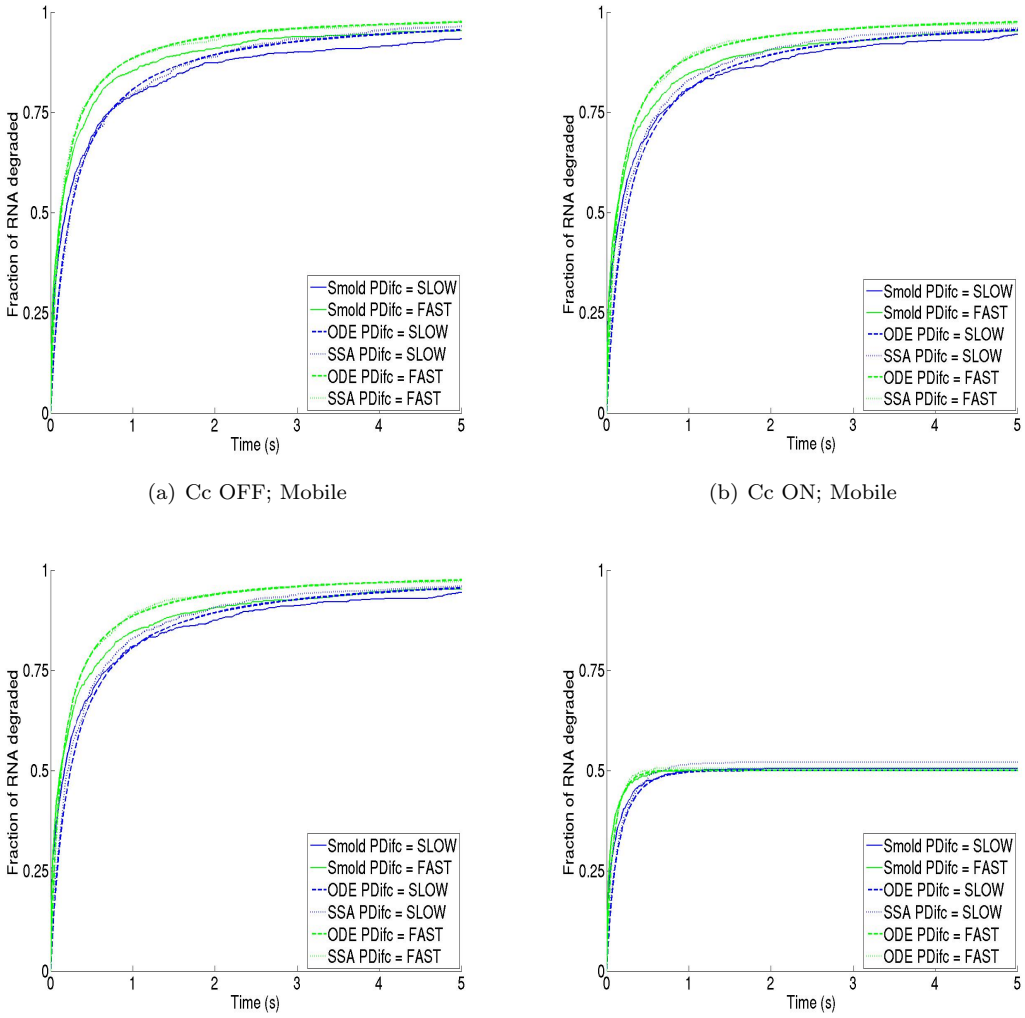


(b) "HIGH"; PDifc = NORM



(c) "HIGH"; PDifc = FAST

Figure 3.39: Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results. ODE results are shown with a thick dashed line, SSA results are shown with a fine dotted line, and Smoldyn results are shown with a solid line. $N_b = 1$, miRNAs = "NORMBIG", and binding is irreversible. All figures have PTEN = miRNA = "LOW", copycats present, and $\theta = 0.6$. Figures 3.39(b) and 3.39(c) compare when the diffusion of PTEN mRNAs is "SLOW" or "FAST", respectively. Data shown is the average of 20 simulations. We still see no visible differences for the simple case with copycats.



(a) Cc OFF; Mobile

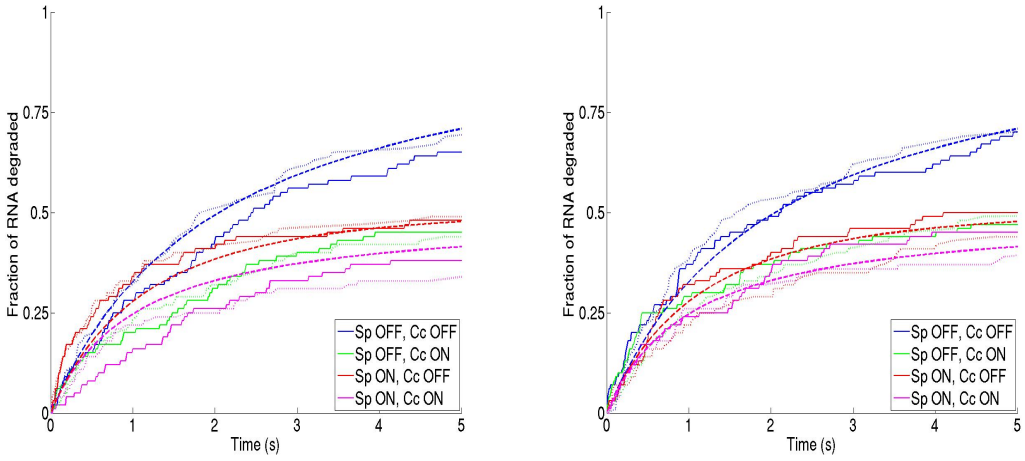
(b) Cc ON; Mobile

(c) Cc OFF; Stat

(d) Cc ON; Stat

Figure 3.40: Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results. ODE results are shown with a thick dashed line, SSA results are shown with a fine dotted line, and Smoldyn results are shown with a solid line. $N_b = 1$, miRNAs = "NORMBIG", and binding is irreversible. PTEN = miRNA = "HIGH", $\theta = 0.6$. Figures 3.40(a) and 3.40(c) are for "Cc OFF" and Figures 3.40(b) and 3.40(d) are for "Cc ON". Data shown is the average of 20 simulations. We compare the diffusion of PTEN, the effect of copycats, and obstacle mobility. We still see no visible differences.

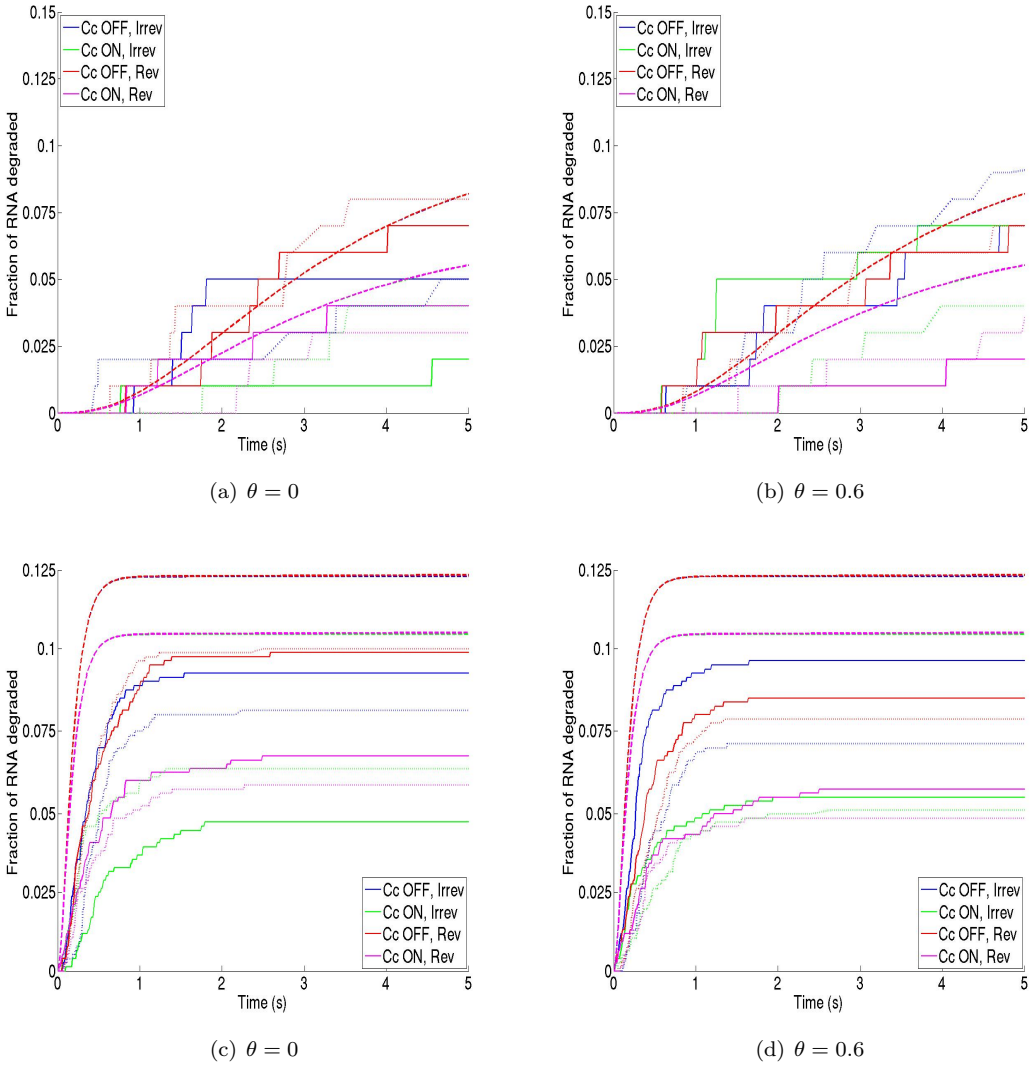
3.42, we see a divergence of the ODE results from the SSA results when the number of binding sites increases from $N_b = 1$ to $N_b = 3$. This is particularly clear in Figures 3.42(c) and 3.42(d), in which we can see that both the rate and final number of degraded RNAs is much higher for the ODE than the SSA and Smoldyn results. The SSA results are also slightly divergent from what we



(a) $\theta = 0$ (b) $\theta = 0.6$

Figure 3.41: Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results. ODE results are shown with a thick dashed line, SSA results are shown with a fine dotted line, and Smoldyn results are shown with a solid line. $N_b = 1$, miRNAs = "NORMSMALL", and binding is irreversible. PTEN = miRNA = "LOW". Data shown is the average of 20 simulations. We compare the cases with copycats and sponges "ON" or "OFF" in different combinations, with and without obstacles. We still see no visible differences.

would expect to see from Smoldyn, particularly when there are no obstacles present in the Smoldyn simulation. However, we are, in general, more interested in the comparisons between the "crowded" Smoldyn results and the SSA/ODE results (Figures 3.42(b) and 3.42(d)), for which we see only a slight difference between SSA and Smoldyn results. Again, it is difficult to conclude anything from the case with "LOW" numbers of molecules, as binding events are so rare in all cases.

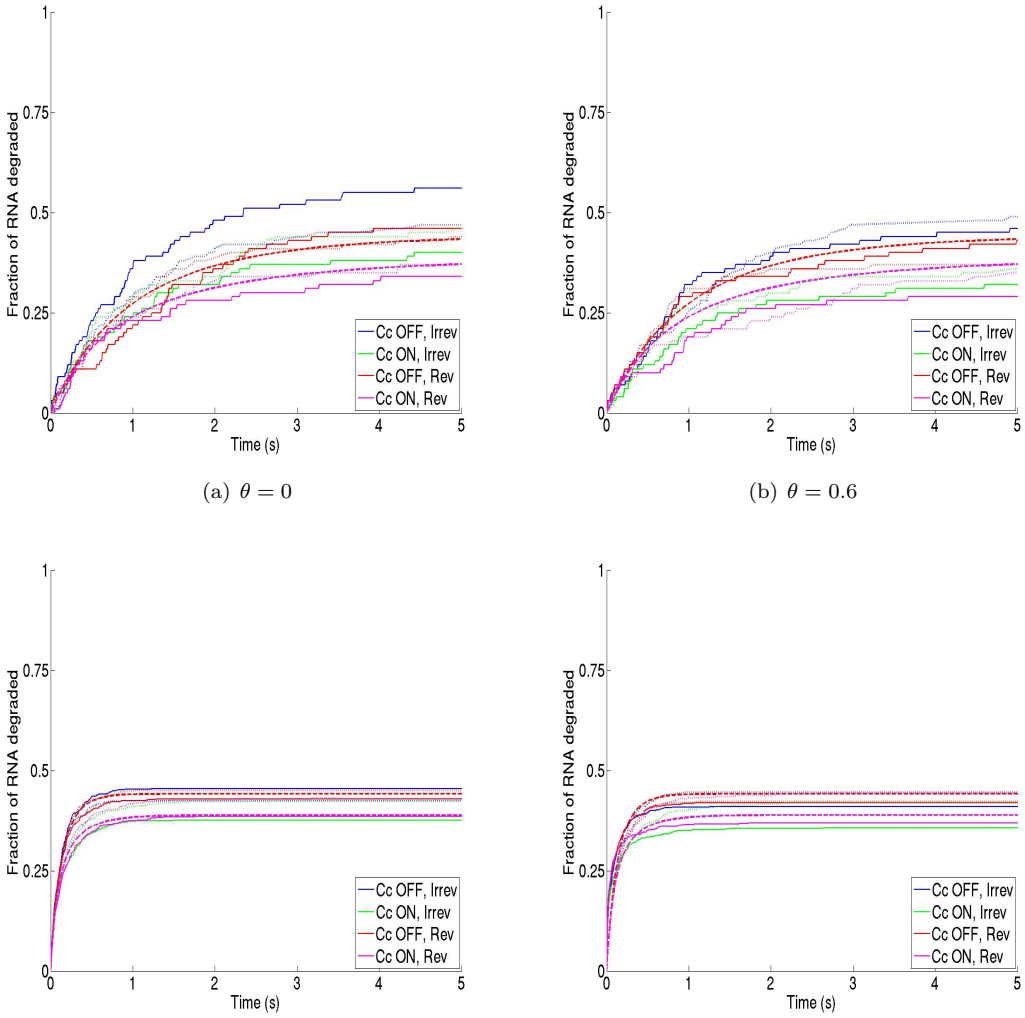


(a) $\theta = 0$ (b) $\theta = 0.6$

(c) $\theta = 0$ (d) $\theta = 0.6$

Figure 3.42: Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results. ODE results are shown with a thick dashed line, SSA results are shown with a fine dotted line, and Smoldyn results are shown with a solid line. $N_b = 3$, miRNAs = "NORMSMALL". We compare the cases with "LOW" and "HIGH" molecule count, copycats "ON" or "OFF", binding reversible ("Rev") or irreversible ("Irrev"), and $\theta = 0$ or 0.6. We finally see some divergence of the Smoldyn, SSA and ODE results. In particular, when the molecule count is "HIGH", the ODE overestimates the amount of RNA degradation and the SSA underestimates it, in terms of the total amount and the rate. Data shown is the average of 20 simulations.

In Figure 3.43, we can see that the ODE results for the cases with either copycats present (Cc ON) or absent (Cc OFF) match exactly, such that reversibility plays no part in the number of degraded RNAs (dashed blue and magenta lines). However, the Smoldyn and SSA results show a greater

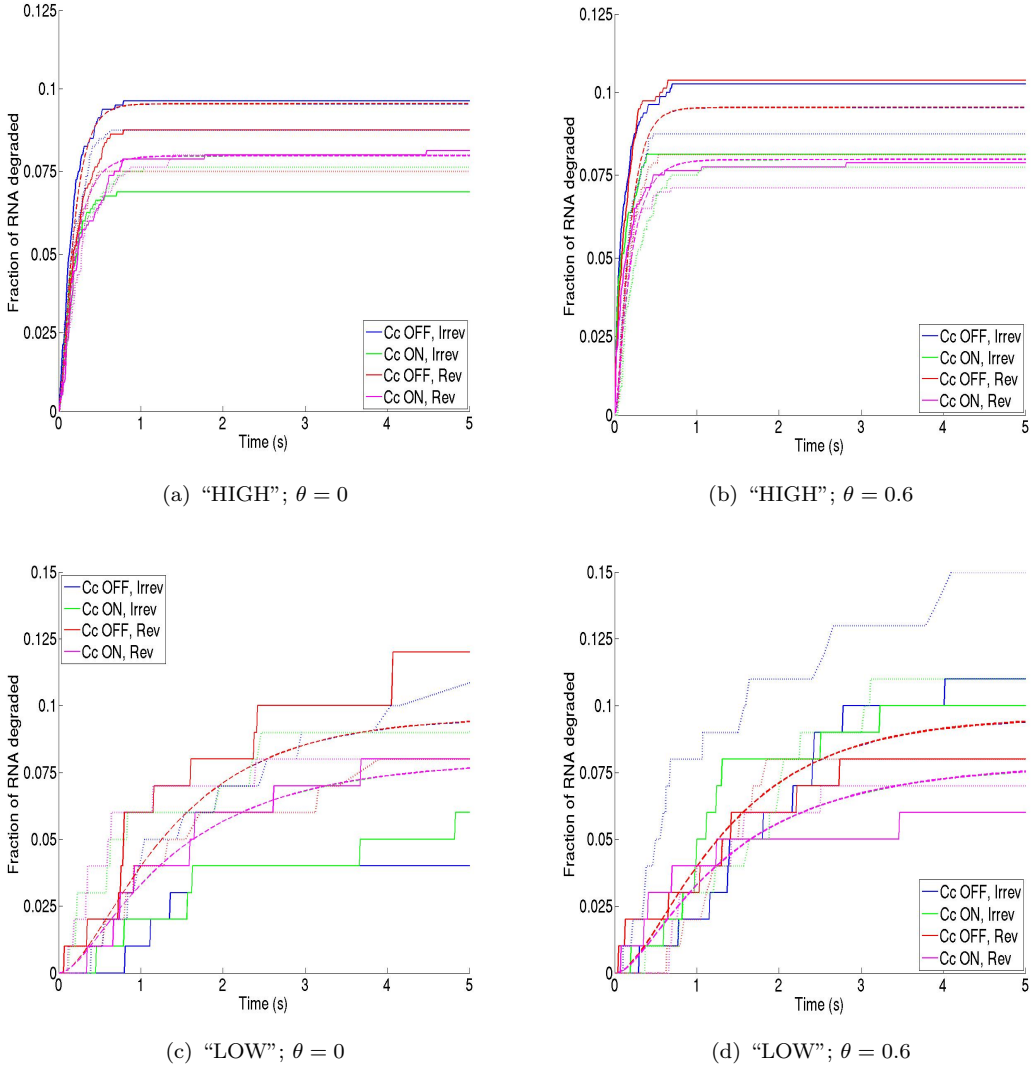


(a) $\theta = 0$ (b) $\theta = 0.6$

(c) $\theta = 0$ (d) $\theta = 0.6$

Figure 3.43: Fraction of the PTEN mRNA population degraded comparing Smoldyn, SSA and ODE results. ODE results are shown with a thick dashed line, SSA results are shown with a fine dotted line, and Smoldyn results are shown with a solid line. $N_b = 1$, $N_bC = 2$; miRNAs = "NORMSMALL"; ceRNAs are present. We compare the cases with "LOW" and "HIGH" molecule count, copycats "ON" or "OFF", binding to ceRNAs either reversible ("Rev") or irreversible ("Irrev"), and $\theta = 0$ or 0.6. We no longer see any divergence of the SSA or ODE results from the Smoldyn results. It is possible that this is because the rate and amount of degradation is 5-fold larger in this case compare to the case with $N_b = 3$ and $N_bC = 0$ (see Figure 3.42). Data shown is the average of 20 simulations.

amount of RNA degradation with irreversible binding in the case with a "LOW" molecule count, and also show a slight increase with a "HIGH" molecule count, although this difference is much less clear.



(a) "HIGH"; $\theta = 0$

(b) "HIGH"; $\theta = 0.6$

(c) "LOW"; $\theta = 0$

(d) "LOW"; $\theta = 0.6$

Figure 3.44: Fraction of the PTEN mRNA population degraded comparing SSA and ODE results. ODE results are shown with a thick dashed line, and SSA results are shown with a fine dotted line. $N_b = 2 = N_bC$; miRNAs = "NORMSMALL"; ceRNAs are present. We compare the cases with "LOW" and "HIGH" molecule count, copycats "ON" or "OFF", binding to PTEN mRNAs and ceRNAs either reversible ("Rev") or irreversible ("Irrev"), and $\theta = 0$ or 0.6. We again see some divergence of the SSA and ODE results. It is likely that this is because the rate and amount of degradation is 5-fold lower in this case compared to the case with $N_b = 1$ and $N_bC = 2$ (see Figure 3.43). Data shown is the average of 20 simulations.

In Figures 3.44(a) and 3.44(b), we see a big divergence in SSA results from the ODE and Smoldyn results for the cases with copycats. In particular, the number of degraded RNA is almost twice what we would expect from either the ODE or Smoldyn results when binding is irreversible. This is observed with and without obstacles when there are "HIGH" numbers of molecules. These effects

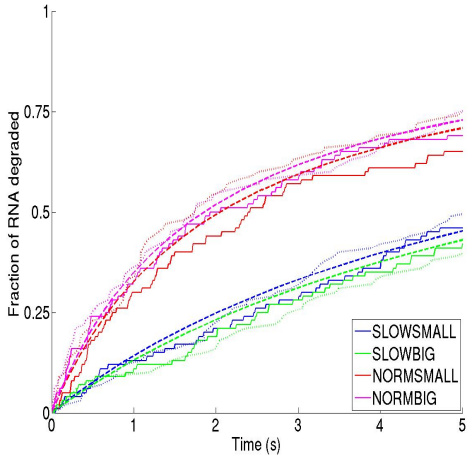
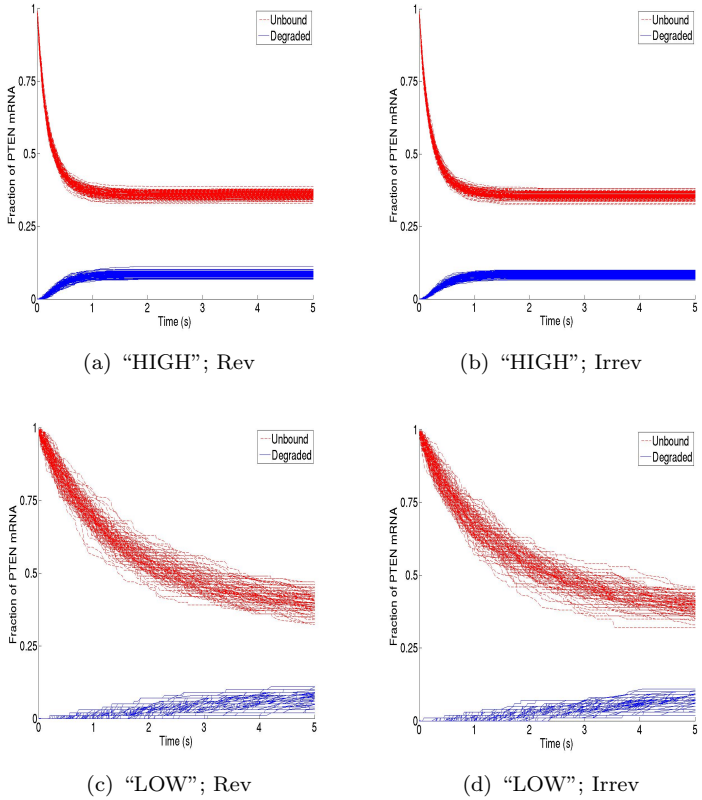


Figure 3.45: Fraction of the PTEN mRNA population degraded comparing SSA and ODE results and diffusion rates ($\theta = 0.3$). ODE results are shown with a thick dashed line, and SSA results are shown with a fine dotted line. $N_b = 2 = N_bC$; miRNAs = "NORMSMALL". agreement between all results regarding the effects of using small, large, normal diffusing and slow diffusing miRNAs on reaction rates.

are less clear when "LOW" numbers of molecules are present however, as we can see from Figures 3.44(c) and 3.44(d), binding events are rare in all cases, so it is difficult to make any definitive conclusions. Figure 3.45 shows the agreement between all results regarding the effects of using small, large, normal diffusing and slow diffusing miRNAs on reaction rates. That is to say, in each case ("NORM/SLOW", "SMALL/BIG"), the miRNA diffusion coefficient has a much greater effect on RNA degradation than size, and that the number of degraded mRNAs increases with the miRNA diffusion coefficient.

3.5.6 SSA Results in Isolation

Figures 3.46-3.48 are presented to show the variation in SSA results (over 100 simulations) for a variety of cases. In each figure, we see that, when the molecule count is "LOW", the spread of simulations is larger than when the molecule count is "HIGH". However, the rate and final value of RNA degraded does not vary much between simulations, suggesting that the system is fairly robust. However, in our Smoldyn results, we sometimes see a divergence from the SSA and ODE results, suggesting that, although the variance between simulations is small, the dynamics themselves may not always reflect what would be seen in a highly crowded spatial environment with multiple (reversible) binding sites, copycats, ceRNAs, and electrostatic interactions between molecules and



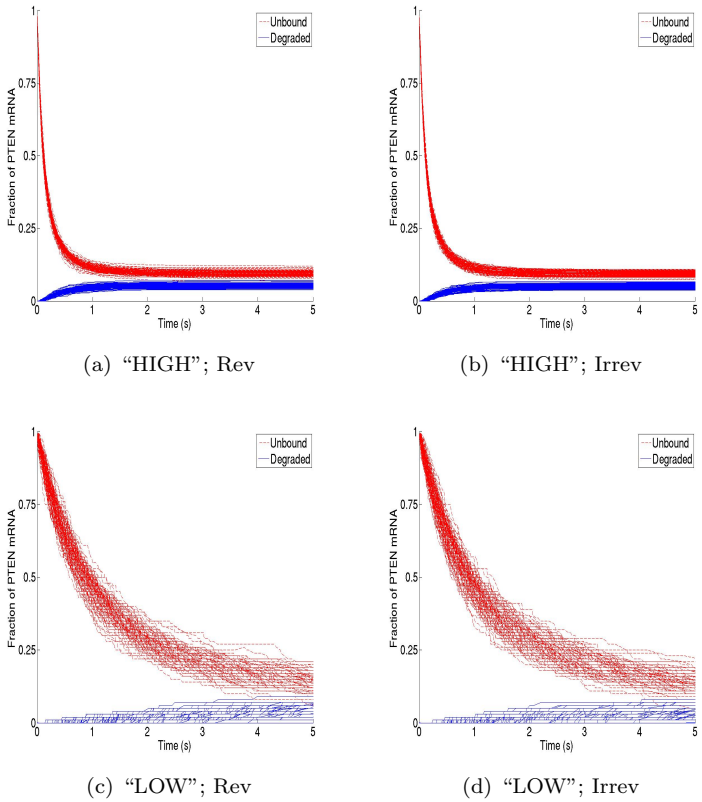
(a) "HIGH"; Rev (b) "HIGH"; Irrev

(c) "LOW"; Rev (d) "LOW"; Irrev

Figure 3.46: Fraction of the PTEN mRNA population degraded (red) and unbound (blue), using the SSA. $N_b = 3$, copycats are absent, and binding is either reversible (as in Figures (a) and (c)) or irreversible (as in Figures (b) and (d)).

obstacles.

In Figure 3.48, we demonstrate the impact of the unbinding rate for reversible binding of miRNAs or copycats. In Figures 3.48(a)-3.48(c), the unbinding rate is taken from (Ala et al., 2013); and in Figures 3.48(d)-3.48(f), we increase that rate 1000-fold. As we can see, in all cases, increasing the unbinding rate leads to a steady increase in the number of degraded RNAs. When $N_b = 1$ and copycats are present (Figures (a) and (d)), the number of unbound PTEN mRNAs decreases to almost zero. We assume that this is a result of copycats unbinding from ceRNAs, resulting in more copycat−PTEN-mRNA binding events. This is because RNA degradation, as a result of miRNA binding, is increasing at a slower rate than overall mRNA binding events are occurring. Conversely, when $N_b = 3$ and copycats are absent (Figures (b) and (e)), the number of unbound PTEN mRNAs increases. This may be because, as ceRNAs are absent, the increased unbinding rate results only in unbinding of miRNAs from PTEN mRNAs. The fraction of PTEN mRNAs unbound



(a) "HIGH"; Rev (b) "HIGH"; Irrev

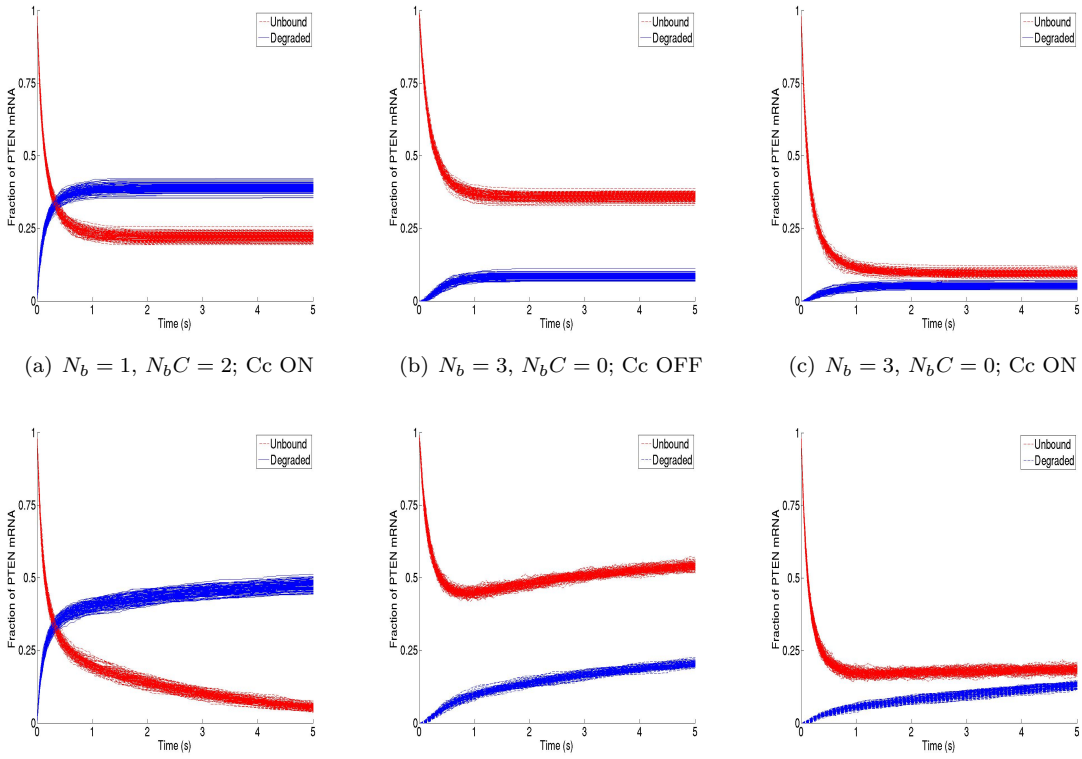
(c) "LOW"; Rev (d) "LOW"; Irrev

Figure 3.47: Fraction of the PTEN mRNA population degraded (red) and unbound (blue), using the SSA. $N_b = 3$, copycats are present, and binding is either reversible (as in Figures (a) and (c)) or irreversible (as in Figures (b) and (d)).

and degraded do not sum up to one as the threshold ($T = 3$) means that there are insufficient miRNAs to degrade all PTEN mRNAs (in the absence of miRNA recycling ($\alpha \neq 0$)). Finally, when $N_b = 3$ and copycats are present (Figures (c) and (f)), we see a slower increase RNA degradation and only a slight increase in the number of unbound mRNAs. This is likely due to the high number of sites blocked by copycats, which must be unbound and occupied by miRNAs, to a threshold of $T = 3$, in order for PTEN mRNA degradation to occur.

3.6 Discussion

In Section 3.4, we declared a number of hypotheses that we wished to test with our models of the PTEN pathway. In this section, we recall each of these hypotheses, and discuss whether we were correct in our predictions.



(a) $N_b = 1$, $N_bC = 2$; Cc ON (b) $N_b = 3$, $N_bC = 0$; Cc OFF (c) $N_b = 3$, $N_bC = 0$; Cc ON

(d) $N_b = 1$, $N_bC = 2$; Cc ON (e) $N_b = 3$, $N_bC = 0$; Cc OFF (f) $N_b = 3$, $N_bC = 0$; Cc ON

Figure 3.48: Fraction of the PTEN mRNA population degraded (red) and unbound (blue), using the SSA. $N_b = 3$, copycats are either present or absent, ceRNAs are either present or absent, molecule count is "HIGH", and binding is reversible. We also compare the effect of changing the reversible reaction rate ($k_b = 0.0003s^{-1}$ in Figures (a)-(c); $k_b = 0.3s^{-1}$ in Figures (d)-(f)).

3.6.1 Hypotheses 1-3: Effect of Macromolecular Crowding

Stationary crowders produce similar dynamics to dilute (non-crowded) conditions with a restricted volume.

In Section 3.5.1, we see that, when PTEN mRNAs diffuse normally, stationary obstacles actually produce similar dynamics to mobile obstacles, although with slightly reduced diffusion in the stationary case. However, immobile obstacles results in greater trapping of large molecules, such as RNAs. This trapping is particularly pronounced when RNA diffusion is increased (PDifc = "FAST"), and apparent Lévy flight motion is observed in a highly crowded domain ($\theta = 0.6$). This Lévy-type motion is more clearly defined with mobile obstacles but is not present for dilute or moderately crowded ($\theta = 0.3$) conditions. In both dilute and crowded conditions, small miRNAs are barely restricted in their diffusion due to their small size ($\geq$ 200-times smaller than obstacles), which means that reaction rates are not obviously affected by macromolecular crowding, as we see in Section 3.5.2.

Reactive mobile obstacles have a bigger impact on diffusion and reaction rates than inert (non-reactive) mobile obstacles.

In Section 3.5.2, we see that adding weak electrostatic interactions between obstacles and RNAs decreases the rate, and amount, of RNA degradation for high and medium trapping probabilities. However, for a highly crowded environment ($\theta = 0.6$), we see that a low trapping probability actually favours RNA degradation, ultimately leading to increased total RNA degradation, although the rate of RNA degradation is reduced. One possible explanation is that molecules which would otherwise be trapped by the highly crowded environment are able to escape trapping by binding to an obstacle and unbinding outside of the trap. This would take time to occur (slow rate), but would ultimately free up more mRNAs to bind to miRNAs that might otherwise have been trapped between obstacles (higher total RNA degradation). This effect is less clear for moderately crowded environments where trapping is less likely to occur.

The size of the obstacles is as important as the total volume excluded.

We do not see any effect of increasing the size of obstacles on the overall amount and rate of RNA degradation, most likely because miRNAs are still capable of diffusing freely between obstacles regardless of their size. The freedom of miRNA movement means that no change in the geometry will affect the reaction rate of miRNA-mRNA binding. For large miRNAs, the size of obstacles has a greater impact on RNA degradation, but ultimately we would need to consider miRNAs to be of the same size as mRNAs in order to see any effect of varying the size of obstacles. Nonetheless, even if the miRNA size is increased, resulting in a reduced rate of RNA degradation, the SSA model can still predict this impact, and the ODE model can provide a good estimate of the average behaviour.

At this stage, Smoldyn, SSA and ODE results agree for the majority of scenarios. The most likely reason for this is that, as the size of miRNAs and RNAs vary over several orders of magnitude, we cannot effectively model macromolecular crowding in a system with obstacles that are of a similar size to miRNAs. This means that miRNAs effectively behave deterministically, implying that an ODE model would describe the system fairly well. However, there are a few cases where the SSA and ODE models diverge from Smoldyn results, which we discuss at the end of this section. We also believe that another possible reason for the similarity in results (across Smoldyn, ODE and SSA models) is that the system is monostable, meaning that there is no effect of small perturbation in the system on RNA degradation. In the final chapter, we look at a system capable of bistability, to see if Smoldyn displays different dynamics compared with SSA and ODE models of the same system.

3.6.2 Hypotheses 4-5: Effect on RNA Degradation

Adding ceRNAs to the system causes a greater upregulation of PTEN mRNAs than the addition of copycat miRNAs or a combination of both ceRNAs and copycats (if copycats can bind to ceRNAs)

We have shown in Section 3.5.2 that adding both ceRNAs and copycats together results in lower RNA degradation levels compared with either ceRNAs or copycats in isolation. All three situations, as we might expect, result in a decrease in RNA degradation. However, we expected that the effects of copycats and ceRNAs to cancel each other out, with copycats blocking ceRNA binding sites,

which we do not observe. This effect is also predicted by SSA simulations. For ODE simulations, the effects of adding ceRNAs and copycats are identical which, again, we would not necessarily expect. These conclusions are true regardless of binding site reversibility.

Reversibility of miRNA or copycat binding to PTEN mRNAs or ceRNAs does not affect PTEN mRNA repression/degradation.

In Figure 3.35, when $N_b = 1$, we see little difference in RNA degradation when ceRNA binding sites are reversible compared with irreversible. The slight differences we do see favour binding site reversibility for reduced RNA degradation. This is not easy to explain as miRNAs and copycats are only unbinding from ceRNAs and there are twice as many binding sites for miRNAs and copycats ($N_bC = N_m(0) + N_d(0)$). When $N_b > 1$, we see a definitive difference between reversibility of binding sites, suggesting that the reversibility has the biggest effect when applied to PTEN MREs. In a dilute environment ($\theta = 0$), when copycats are absent (Cc OFF), binding site irreversibility results in more RNA degradation. Conversely, when copycats are present, binding site reversibility results in more RNA degradation. In highly crowded conditions, the differences between reversible and irreversible binding almost vanish. One possible explanation for this is that unbinding events are merely temporary, with unbound molecules re-binding almost immediately due to limited space through which miRNAs or copycats can diffuse away from their recent binding partners. It is at this point that we see the biggest divergence of the SSA and Smoldyn results, shown in Figures 3.44(a) and 3.44(b). When copycats are absent, the SSA and Smoldyn results agree well. However, when copycats are introduced, we see a large increase in RNA degradation in the SSA for irreversible binding (almost double the value predicted by Smoldyn), and a slight increase for reversible binding. This difference is visible in both the dilute and highly crowded cases, where we also see a reduction in predicted RNA degradation for the case with no copycats. This is surprising as the ODE results match the Smoldyn predictions quite well in this scenario.

In Figure 3.42, when ceRNAs are no longer present but mRNAs have three binding sites ($N_b = 3$) capable of reversible binding to both miRNAs and copycats, we see the biggest divergence of the ODE results from the Smoldyn and SSA results. In particular, the rate and final amount of degraded RNA is vastly overestimated in all cases, particularly when copycats are present.

3.6.3 Summary

In order for ceRNAs and copycats to be used as an effective treatment for preventing miRNA-mediated repression of PTEN, they have to act by reducing RNA degradation in all situations, regardless of PTEN MRE reversibility, the number of MREs, volume fraction, or weak-force interactions with obstacles. Our results suggest that the addition of copycats or ceRNAs (or both) will always reduce the amount of degraded RNAs, which would make them an effective treatment to cancerous cells, without negatively affecting healthy cells. However, in order to make our model predictive, we would require experimental validation, and expansion to include more ceRNA and copycat species, to test these effects on a larger ceRNET. In spite of this, these preliminary results are promising.

As for whether explicit spatial modelling of the system is necessary, we would need to compare our results to experimental data, particularly in the cases when SSA or ODE results diverge from the Smoldyn predictions, to see if our model is a better match to the experimentally observed dynamics. Our model is, however, valuable for obtaining the parameters necessary for use in the SSA and ODE models, using binding radii and diffusion coefficients to determine the effective reaction rates, for example. The parameters we obtain from our simulations do vary from published parameters, such as those from Ala et al. (2013), and performing manual calculations of these rates would be difficult for multiple molecular species and domain conditions. Therefore, it is worthwhile to create the model in Smoldyn to obtain such parameter values, although we would need experimental data to check that these new parameter values result in a better approximation of the system dynamics than currently published parameters.

We have also shown in this chapter that if even one molecular species is able to diffuse freely through the domain, regardless of volume fraction, then modelling macromolecular crowding is largely unnecessary. The results for which we saw a difference in reaction rate between dilute and crowded conditions was in the comparison of binding site reversibility and, of course, when RNAs were capable of weak-force interactions with obstacles. In all other cases, the freedom of movement of miRNAs/copycats was the limiting factor, with all reactions involving miRNAs and copycats depending on the diffusion and size of these small molecules. This is worth noting for future studies involving molecules and obstacles of vastly different sizes. It is possible that modelling obstacles of the same size as miRNAs may have resulted in a change in RNA degradation, but the number of

obstacles required to reach volume exclusion (>5%) would have been too large to simulate using Smoldyn ($\sim 10^{10}$ molecules).

Interestingly, we have shown that RNAs diffusing in a highly crowded confined domain undergo apparent Lévy-flight-type motion, but only when RNAs are diffusing fast enough. If RNAs are diffusing normally (representative of their size), then the molecule is merely trapped by the highly crowded environment. This has potential implications for drug delivery. If a large molecule is to be capable of diffusing through a highly crowded environment, it must be diffusing quickly. This could be achieved experimentally by shortening the nucleotide sequence, or by removing fluorescent labels, for example. This could speed up delivery and/or ensure that the molecule reaches its intended destination. Alternatively, the molecule should be as small as possible to avoid trapping altogether, as per miRNA/copycat diffusion. In this regard, copycats may act as a better therapy than ceRNAs as one can be more certain that they will reach the intended destination in the cell, as well as improving their ablility to explore the intracellular space and increase the likelihood of copycat-mRNA interaction.

Overall, we conclude that our model can be used to estimate rate constants and probabilities from molecular weights and sizes, for use in an SSA or ODE model of the proccess. It is our belief that the robustness of this system to stochasticity and spatial heterogeneity is due to the small size of miRNAs, the relatively high molecular count, and the monostability of the system. We look further into the latter case of bistability in the following chapter (Chapter 4), examining the canonical Wnt signalling pathway.

Chapter 4

The Canonical Wnt/βcatenin Signalling Pathway with Cell-Cell Adhesion

As we will describe in this chapter, the Wnt signalling pathway is a good example of a complex biological system that could be better studied using a stochastic computational model; It is a multi-compartmental system, with reactions occurring in the nucleus, cytoplasm and at the cell membrane, and involves many small molecules in low concentrations. Although we did not have the time to finish the complete analysis of the system, in the following chapter we will present the background of the Wnt system, as well as the research we have completed thus far.

4.1 Introduction

In epithelial cells, β-catenin can have a dual role in both Wnt signalling and cell-cell adhesion. We use existing deterministic models of these processes to create spatially-resolved stochastic models, in order to elucidate the impact of low molecule numbers, nucleo-cytoplasmic shuttling of molecules, and spatial heterogeneity, on the level of Wnt target gene transcription and cell-cell adhesion. We hypothesise that these models will more accurately describe the cellular dynamics observed experimentally (such as sustained oscillations of β-catenin expression) than their deterministic counterparts, and may give an insight into the different scenarios that can lead to tumour formation and progression. In this study, we build on the pioneering work by Lee et al., and the more recent work by van Leeuwen et al. and Schmitz et al., by extending their deterministic ODE models of the canonical Wnt pathway and cell-cell adhesion to spatial stochastic simulations of the two processes (Lee et al., 2003; Schmitz et al., 2013; van Leeuwen et al., 2007). We use the stochastic simulator

Smoldyn to create our simulations, and study the impact of Wnt, E-cadherin, spatial heterogeneity, and cytoplasmic retention of β-catenin on β-catenin/TF complex formation (Andrews et al., 2010; Andrews and Bray, 2004).

Schmitz et al. (Schmitz et al., 2013) studied the effects of inter-compartmental shuttling of β-catenin and the destruction complex, (DC), between the nucleus and cytoplasm, on the transcription of Wnt target genes. The measured output of the system is the formation of complexes between β-catenin and its transcription factors (TFs), allowing for ignorance of rates of transcription and translation of target genes. The amount of DC is approximated by the amount of APC, as formation of the DC is restricted by the limited amount of APC in humans (Tan et al., 2012). This depends on the particular cell type which is being considered, but in our study we focus on epithelial cells such as the intestinal epithelial mammalian cell line, Caco-2. This assumption of DC limitation by APC levels is in contrast to the Lee et al. and van Leeuwen et al. models that are based on *Xenopus* cell extracts, in which Axin is the limiting factor of the DC. The schematic is shown in Figure 4.1. A system of coupled ODEs is presented, representing reversible binding of β-catenin to TFs and DC, synthesis, phosphorylation and ubiquitination of β-catenin, shuttling of β-catenin and DC between the nucleus and cytoplasm, and reversible inactivation of DC by Wnt. Initially, a reference model is created whereby Wnt is "off" and DC is restricted to the cytoplasm. Steadily the model is built up to include shuttling of DC, a permanent Wnt signal, and finally a transient Wnt signal. They also report on the influence of inhibition of β-catenin binding to DC, and the inhibition of β-catenin phosphorylation/degradation by DC, in both the nucleus and the cytoplasm. Their results show that shuttling has the greatest effect on β-catenin levels when it results in DC and β-catenin being present in the same compartment, and in particular when they are both present in the cytoplasm, as β-catenin is separated from the TFs.

It has been hypothesised that β-catenin is chaperoned into the nucleus by APC (Henderson and Fagotto, 2002), however, it has recently been suggested that, rather than actively transporting β-catenin into the nucleus, APC acts by retaining β-catenin in the cytoplasm (Schmitz et al., 2013). Once inside the nucleus, β-catenin acts as a cofactor for the T-cell factor/lymphoid enhancing factor (TCF/LEF) family of transcription factors (Giles et al., 2003), responsible for the transcription of Wnt target genes (Takebe et al., 2010). As such, we use β-catenin/TF complex formation as our metric, as transcription of the target genes by this complex in the ultimate step in the canonical

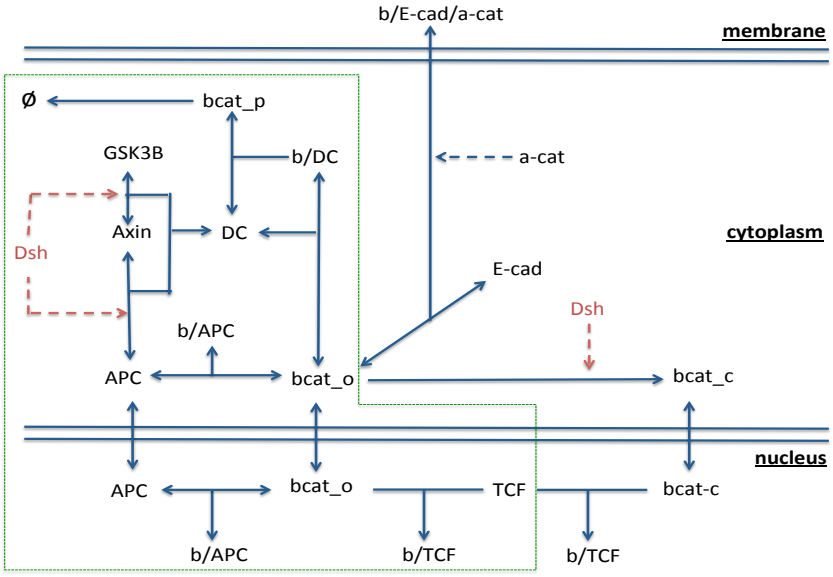


Figure 4.1: Schematic for Schmitz et al. model (inside box) with cell-cell adhesion added.

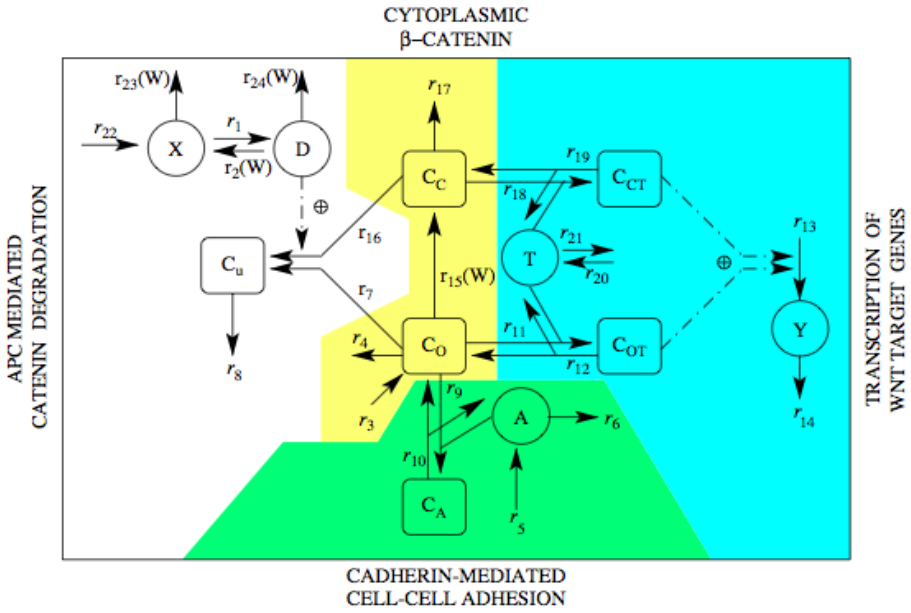


Figure 4.2: Schematic for van Leeuwen et al. model (van Leeuwen et al., 2007).

Wnt pathway, and a direct reflection of the effect of perturbations on the system.

We also examine the model by van Leeuwen et al. (van Leeuwen et al., 2007). They study β-catenin binding to TCF and E-cadherin, as well as regulation by DC, as a system of ODEs. They present two hypotheses for how β-catenin is regulated by its three potential binding partners. Hypothesis 1 (H1)

states that DC, TCF and E-cadherin compete directly for a binding site on β-catenin. Hypothesis two (H2) suggests that Wnt signalling produces a conformational change in β-catenin, resulting in a "closed-form" that binds preferentially to TCF. We model both hypotheses and hope to use our model to test which hypothesis might be most accurate.

4.2 Development of Spatial Stochastic Model

We chose to model the Wnt signalling pathway as, like the PTEN and RNAi pathways, it is a gene regulatory pathway with an involvement in cancer. This, more complex, pathway allows us to expand upon the PTEN and RNAi models, exploring the possibility that spatial modelling of the Wnt pathway would give more divergent results from deterministic and stochastic models. This additional complexity comes in the form of multiple spatial compartments, i.e. the nucleus, cytoplasm, and membrane, and the possibility of bistability. However, we were unable to formulate a simple, biologically feasible, deterministic bistable model of the Wnt pathway that we could also translate to a Smoldyn model. We tried to use the bistable models of Kim et al., Schmitz et al., and McLean et al., but we could not create representative Smoldyn models without sufficient details of parameter values and simple chemical reaction network representations. The Kim et al. model is primarily a qualitatively validated model and includes parameters that remain to be experimentally validated (Schmitz, 2014); For the Schmitz et al. model, whilst it has been proven to be capable of multistationarity by failing the injectivity test, McLean et al. also showed that it is capable of at most two steady states, only one of which is stable, i.e. it is not capable of bistability. Our best chance is with the McLean et al. "shuttle" model which they show to be capable of bistability, although it is specifically designed to be parameter-free, which makes it difficult to convert to Smoldyn. In their paper, they provide a sample parameter set, given by the CNET toolbox, which would result in bistability but, as they themselves state, the parameters are not biologically informed and serve only as a proof of concept parameter set.

Rather than abandon the entire chapter, we decided to present our work so far, building on the Schmitz et al. model and a simple bistable model which has recently been modelled using Smoldyn. This work can be built upon to determine the independent effects of compartmentalisation and bistability in a spatially-detailed model.

4.2.1 Endres Paper - Schlögl Model

In their paper, Endres et al. study the the Schlögl model in a biological setting to show that bistability depends on compartmental volumes and diffusion of molecular species. They extend their study to three dimensions, multi-compartment systems, and particle-based simulations, using the Smoldyn simulation tool described previously. The Schlögl model is often used as the canonical example of a non-equilibrium, bistable chemical reaction system (Vellela and Qian, 2009). In terms of a chemical reaction network, the Schlögl model is given by

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} X \tag{4.1}$$

$$3X \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} 2X + B \tag{4.2}$$

where k_i are rate constants ($i = \pm 1, \pm 2$), X is the copy number of the molecular species of interest, and concentrations A and B are adjustable parameters. For the Schlögl model to work in Smoldyn, the third-order reaction must be split into a second- and a first-order reaction, as the Smoluchowski theory, upon which Smoldyn is based, only works for zero-, first- and second-order reactions. For this, Endres et al. create a second intermediate species, X_2, and two new reaction rates, $k_{\pm 3}$, resulting in a new set of chemical reactions,

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} X \tag{4.3}$$

$$X + X \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} X_2 \tag{4.4}$$

$$X + X_2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} X_2 + B. \tag{4.5}$$

For concentrations $x = X/V$ and $x_2 = X_2/V$, for reaction volume V, the ODE system is given by,

$$\frac{dx}{dt} = -k_{+2}x_2x + (k_{-2}B + k_{-3})x_2 - k_{+3}x^2 - k_{-1}x + k_{+1}A \tag{4.6}$$

$$\frac{dx_2}{dt} = k_{+3}x^2 - k_{-3}x_2 \tag{4.7}$$

where the total concentration of x is given by $x_T = x + 2x_2$, remembering that x_2 is a manufactured intermediary dimer species, introduced for the purpose of compatibility of the Schlögl model with

Table 4.1: Schlögl model coefficients from Endres et al. (Endres, 2015).

Parameter	Value	Units
k_{-1}	1	s^{-1}
k_{+1}A	$5/(6\nu$ [(1)])	nM s^{-1}
k_{-2}	$6\nu/10$	$(nMs)^{-1}$
k_{+2}	B	s^{-1}
k_{-3}	1	s^{-1}
k_{+3}	$6\nu/10$	$(nMs)^{-1}$
V	10ν	μm^3
D_X	3	μm^2 s^{-1}
D_{X_2}	1	μm^2 s^{-1}
B	3.7	nM

Smoldyn. At steady state, the system is equivalent to the original Schlögl model. Endres et al. took the unitless parameter values from the study by Gaspard et al. of bistability in the Schlögl model, and supplemented this with their own parameters for the volume, concentration of B, diffusion coefficients, and the new rate constants $k_{\pm 3}$. A summary of the parameters used in (Endres, 2015) are given in Table 4.1.

Endres et al. tested the Schlögl bistable model in Smoldyn, and proven its ability to maintain the switch points for bistability when compared with the deterministic model. They show that bistability is a fragile state that require fine-tuning of the system towards small cellular/compartmental volumes and fast molecular diffusion, two characteristics associated with a well-mixed system. They hypothesize that bistability may therefore be restricted to eukaryotic cell nuclei (or small bacterial cells), with the bistable switch triggering during the cell cycle when the cellular volume is dynamically changing. They highlight that spatial heterogeneity, compartmentalisation of the cell, and the effects of slow molecular diffusion are often ignored in the study of bistability. They further discuss the negative effects of macromolecular crowding on the speed at which molecules can diffuse in the cell, hindering non-linearity of the system as zero- and first-order reactions are favoured. They compare Smoldyn simulations with Gillespie simulations and conclude that Smoldyn can accurately reproduce the observed bistabilty, in order to use the Smoldyn model to further explore the effects of diffusion and volume on bistability, determined by re-normalising the second-order rate constants for small diffusion coefficients. However, although the paper states that Smoldyn was used to replicate a crowded intracellular environment, there is no evidence that any macromolecular crowding had

[1]Note: ν is a scaling number of order 1.

been simulated, so this is where we could use our model to extend the analysis and further test the impact of spatial detail in modelling bistability.

Mesoscopic cells are often incapable of displaying bistable behaviour, as such complex behaviours can disappear for sufficiently large volumes. The same is true for slow diffusion resulting from macromolecular crowding and compartmentalisation. As such, bistability may be limited to a small range of parameter space which requires fine tuning by both the cell and the modeller of the cell. Bistability may further be limited to bacterial cells or individual eukaryotic cellular compartments, such as the nucleus or the cytoplasm, which is the focus of the study in (Endres, 2015); A loss of bistability is observed for very slow diffusion rates as the low state is favoured, although the specific steady-state values due not change directly. It is likely that Endres et al. simulate crowding by reduction of molecular diffusion coefficients and subsequent renormalisation of the second-order rate constants, however this is not explicitly stated in their paper.

Endres et al. further suggest that molecules up to 110kDa in weight are not impeded by macromolecular crowding or cytoplasmic macrostructures. As β-catenin, Axin and GSK3 are all less than or equal to this weight limit (at 94kDa, 110kDa, and 50kDa, respectively), it is possible that we can ignore the effects of macromolecular crowding in the Wnt model, thereby supporting the assumption that the system is capable of bistability. However, with APC weighing almost triple this limit (at 310kDa), and the DC weighing considerably more with the combined effects of Axin, APC and GSK3, we must at least assume that crowding effects the movement of the DC throughout the domain. What does this suggest if we assume that its binding partner, β-catenin, can move freely throughout the domain? This is something we would wish to explore in future simulations, but we hypothesise that the rate of β-catenin interacting with DC would now depend on the number of each molecular species diffusing throughout the space, the diffusion coefficient of β-catenin, and the mobility of β-catenin's other binding partners, such as TCF in the nucleus, E-cadherin at the membrane, and isolated Axin (not in DC).

4.2.2 Smoldyn Model Formulation

We began by creating a base model for the system in the absence of Wnt ('Wnt off'), validating our results against those obtained by Schmitz et al. and van Leeuwen et al. (Schmitz et al., 2013; van Leeuwen et al., 2007). A cell is simulated as a sphere of radius $r_c\,\mu$m, with an internal sphere

Table 4.2: Diffusion coefficients of molecules.

Parameter	Mass (kDa)	Diffusion coeff. (μm^2)	Ref.
β	90	18.2	(Schmitz et al., 2013)
DC	470	10.5	(Schmitz et al., 2013)
DC-β	560	9.9	(Schmitz et al., 2013)
TF	225	13.44	(Bakalkin et al., 1997)
TF-β	215	12.01	(Bakalkin et al., 1997)

of radius r_n μm representing the nucleus. The volumes of the nucleus and the cytoplasm are equal, as per (Schmitz et al., 2013). This condition is imposed to remove the complications in changing concentrations of molecules that occur when molecules are moving between compartments of varying sizes. DC, β-catenin and TCF are initialised in each compartment according to the concentrations specified in (Schmitz et al., 2013), assuming that the amount of DC is limited by the amount of available Axin. We set the rate constants and dissociation constants for each reaction specified below and assume initially that β-catenin can diffuse freely across the nuclear membrane. Bi-molecular reactions are represented by dissociation constants $K_i = k_{-i}/k_i$, i.e. the ratio of the backward to forward reactions. Dissociation constants are used because the intermediary complexes which are formed in bi-molecular reactions are in rapid equilibrium. In other words, the reversible binding steps are assumed to be very fast, such that the complexes formed are much more likely to split apart into reactants again than to be converted into product (Lee et al., 2003). β-catenin/TF complexes are taken as the metric. As the simulations take a long time to run, it is not feasible to run the system to the steady state equilibrium as seen in Figure 5 of (Schmitz et al., 2013), therefore, we re-calculate the steady state of the ODE model for shorter time scales and use this as a comparison with our model. As such, for a final time of 30 seconds (and anywhere up to 15 minutes), we expect to see the equilibrium concentration of β-catenin/TF to be approximately 14nM, accounting for variability due to stochasticity (Schmitz et al., 2013). We do observe this equilibrium (data not shown).

All diffusion coefficients from Table 4.2 were calculated using the formula (Andrews, 2011),

$$D \approx \frac{3270}{4 \times \sqrt[3]{m}}\, \mu m^2 s^{-1} \tag{4.8}$$

where D is the diffusion coefficient to be used with Smoldyn and m is the mass of the molecular

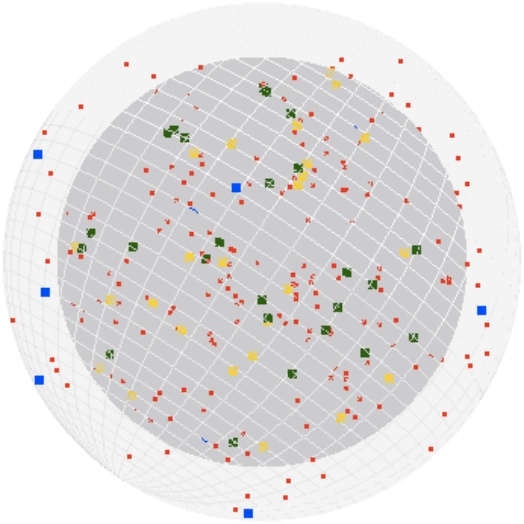


Figure 4.3: Screenshot of the three-dimensional Smoldyn simulation including β-catenin (small red dots in cytoplasm), DC (blue squares in cytoplasm), TF (green dots in nucleus) and β-catenin/TF complexes (yellow squares in nucleus). E-cadherin on the membrane is not shown.

Table 4.3: Smoldyn parameters and initial conditions of Schmitz model.

Parameter	Description	Initial value	Ref.
R_n	Radius of nucleus	3.65μm	(Mazemondet et al., 2009)
R_c	Radius of cell	5.8μm	(Mazemondet et al., 2009)
β	β-catenin	17157 ($\equiv$ 70nM)	(Schmitz et al., 2013)
TF	TCF/Lef	7353 ($\equiv$ 30nM)	(Schmitz et al., 2013)
DC	Destruction complex	10 ($\equiv$ 0.04nM)	(Schmitz et al., 2013)

species in Daltons.

We based our model on the deterministic model of Schmitz et al., and began testing the effects of spatial heterogeneity on the system. We started by including physical obstacles in the cytoplasm to represent volume exclusion and macromolecular crowding. We expected that, initially, the rate of β-catenin-TCF complex formation will be reduced, as there will be restricted access to the nucleus, and increased interaction between β-catenin and the DC in the diminished cytoplasmic volume. However, we expected that, ultimately, the equilibrium concentration of complexes formed will remain the same, with only the time taken to reach the equilibrium changing, as the nuclear concentration of β-catenin gradually increases. Under normal circumstances, there exists an optimal protein density such that the increase in bi-molecular interactions due to crowding is not negated by the negative

Table 4.4: Changes to initial conditions using van Leeuwen model.

Parameter	Description	Initial value	Ref.
β_o	Open-form β-catenin	8504 ($\equiv$ 23nM)	(van Leeuwen et al., 2007)
TF	TCF/Lef	3378 ($\equiv$ 27.5nM)	(van Leeuwen et al., 2007)
DC	Destruction complex	197 ($\equiv$ 0.67nM)	(van Leeuwen et al., 2007)
E-cad	E-cadherin	701 ($\equiv$ 28nM)	(van Leeuwen et al., 2007)

effect on diffusion rates (Schnell and Turner, 2004). However, it seems that there are so few molecules of DC in the cytoplasm, that the interaction between DC and β-catenin is not affected by a reduced cytoplasmic volume (data not shown).

We also tested the impact of a non-uniform initial distribution of molecules, but observed no measurable difference between this and the uniformly distributed case.

In addition to expanding on the model of Schmitz et al. with spatial heterogeneity, we also included E-cadherin on the cell membrane to study the dual role of β-catenin, in Wnt signalling and cell-cell adhesion. We use the deterministic model of van Leeuwen et al. (2007) (van Leeuwen et al., 2007) (and additional information from (Basan et al., 2010) and (Dawes, 2009)) to extend our model in this way. As in (van Leeuwen et al., 2007), we test both the case where the addition of adhesion complexes (AC) serves only to compete with DC and TCF for β-catenin binding (Hypothesis 1), and the case where Wnt signalling produces a different molecular form of β-catenin which binds preferentially to TCF in the nucleus (Hypothesis 2).

We initally represented the adhesion molecules as freely diffusing in the cytoplasm, able to form adhesion complexes with β-catenin which also remained in the cytoplasm (Dawes, 2009). We would extend this model to include a third compartment representing the cell membrane so that we can study β-catenin interacting with E-cadherin on the membrane, with and without the assistance of α-catenin binding in the cytoplasm, and also look at the impact of α-α homodimer formation.

As with the model by Schmitz et al., we wished to use the Smoldyn model to study the impact of macromolecular crowding (van Leeuwen et al., 2007). In this case, we would expect to see an increase in formation of β-catenin/DC as reducing the cytoplasmic volume increases the likelihood of interactions between molecules. Again, there will exist an optimal density of molecules such that the negative effects of crowding on diffusion rates does not interfere with the above enhancement in biomolecular reaction rates. In order to find this optimum value, we experiment with the percentage

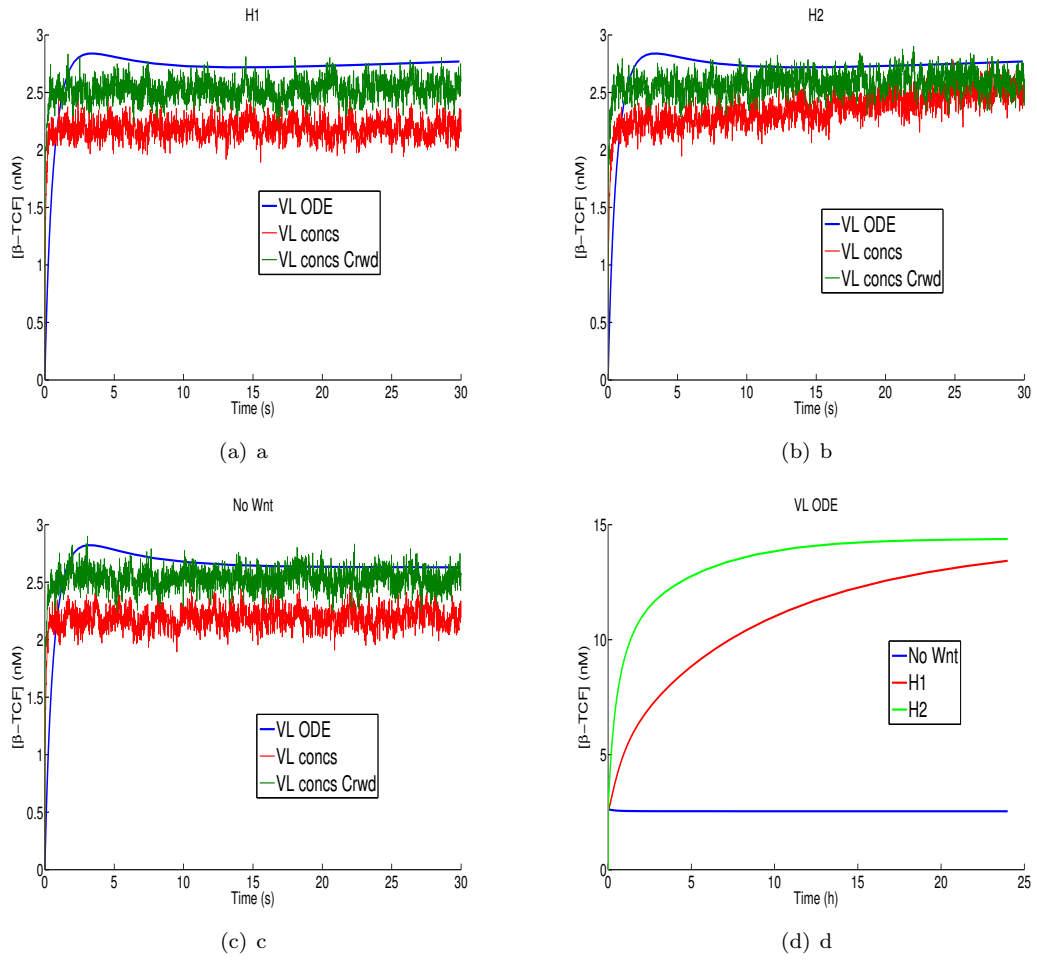
of the cytoplasm that is excluded by obstacles. Our preliminary results are shown in Figures 4.4(a), (b) & (c). We find that there are slight differences initially, and the reduced volume has resulted in a greater probability of DC molecules locating a β-catenin target. The difference in numbers of DC molecules in the Schmitz model vs the van Leeuwen model (increase from 0.04nM to 0.67nM) is clearly impacting on the effect of spatial crowding on the dynamics in the cell. However, as we can see in Figure 4.4(d), the time scale of these results is too short to make any concrete conclusions, so we would have to run longer simulations (on the order of 24 hours simulation time). As stated in Chapter 1, we would have to speed up our simulations or run them on a more powerful computer. To speed things up, we may have to make the time steps longer, or take a hybrid approach, combining the Smoldyn simulations with an SSA model (Klann et al., 2012).

4.3 Results

Results of testing the van Leeuwen ODE model (hypothesis 1, hypothesis 2 and no Wnt, respectively) against our Smoldyn model. These preliminary results (a-c) suggest that the biggest difference between the Smoldyn and ODE models is the initial speed of complex formation in the cell, as well as a slightly lower equilibrium number of complexes formed. (d) Long term solution of van Leeuwen ODE model.

4.4 Discussion

Unfortunately, we did not have sufficient time to complete the Wnt model, however, we have presented preliminary findings and the Endres et al. model could be used with further study to create a multi-compartmental bistable Wnt model. We could further expand on the study of the Wnt pathway to incorporate α-α homodimer formation, competing with adhesion complex formation; and, in the case of cell-cell interactions, possible inclusion of cell-cycle asynchrony between cells. As a next step of research, it would be pertinent to study the difference in dynamics between the *Xenopus* egg extracts, used as a basis for the previous models we have described (Lee et al., 2003; Schmitz et al., 2013; van Leeuwen et al., 2007), and mammalian epithelial cells, using the work of Tan et al. (Tan et al., 2012). There are major differences in protein concentrations between these



(a) a (b) b

(c) c (d) d

Figure 4.4: (a-c) Hypothesis 1, Hypothesis 2 and No Wnt simulations in the short-term. (d) Long term dynamics of van Leeuwen ODE model.

two cell types; most notably, the level of Axin is much higher in mammalian cells. This leads us to the question of whether stochastic modelling of this pathway is necessary for mammalian cells, as we no longer have the rate-limiting step in β-catenin degradation (Tan et al., 2012).

4.4.1 Multi-cellular modelling

Thus far, canonical Wnt signalling has always been modelled as a single cell responding to a Wnt stimulus, released and received by the same cell (autocrine signalling). In reality, however, cells are not isolated entities and may signal to each other (paracrine signalling), affecting the behaviour of their neighbouring cells. Cell-cell communication is particularly important for normal growth, death

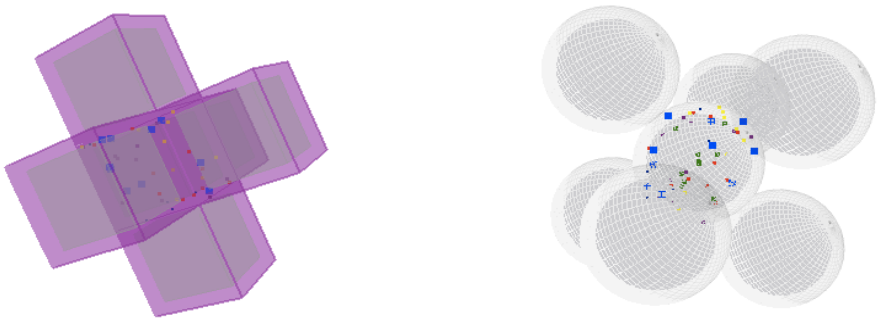
and proliferation of cells, which depend on the level of Wnt stimulus across populations of cells. This has been shown in the mammalian epithelium, in which the gradient of Wnt from the base of the crypt to the tip affects the rate at which cells proliferate up the crypt, differentiate into specialized cells, and eventually die (Lloyd-Lewis et al., 2013). Disruption to this process can easily lead to tumourigenesis, for example if apoptosis (cell death) or proliferation are not properly controlled, this may result in the kind of limitless growth that is synonymous with cancer. Similarly, the response by a population of cells to Wnt stimulus can tell us more about the likelihood of cancers forming than we can gleam from the response by an individual cell, particularly when cells can communicate with each other.

The precise mechanisms by Wnt proteins are secreted in mammalian cells remains to be fully elucidated. However, we can make some hypotheses based on the proposed Wnt secretory pathway in *Drosophila*, given by Gross et al. (Gross and Boutros, 2013). Wnt proteins travel from the endoplasmic reticulum (ER), via the Golgi body, to the plasma membrane, where they are secreted onto exosomes or lipoproteins to be transported through the extracellular space to a receptor on a neighbouring cell. It has been shown that Wnt3a in mammalian cells associates with lipoproteins, specifically high-density lipoproteins (HDL), to travel through the extracellular space, as opposed to vesicles or exosomes (Neumann et al., 2009). Spatial modelling of Wnt secretion may provide more insights into which of the potenial mechanisms best matches observed experimental data, particularly in terms of delays in Wnt release and response.

We would expect to observe a better match to experimental data with the inclusion of Wnt secretion into any model; As Wnt proteins must traverse the extracellular space and locate an appropriate receptor on a neighbouring cell, greater stochasticity and delays will be introduced, dependent on the spatial distribution of the various components of the pathway. Modelling this pathway will also allow us to introduce Wnt antagonists such as Dickkopf 1 (DKK1) and Wnt-inhibitory factor 1 (WIF-1), which can prevent Wnt from binding to its receptors, thereby changing the magnitude of the potential Wnt signal.

Modelling cell-cell interactions is also imperative if one is to fully incorporate the effects of cell-cell adhesion on Wnt target gene expression. Formation of contact between neighbouring cells, via cadherin/catenin interactions, will sequester some of the β-catenin population away from the cytoplasm and nucleus, thereby reducing β-catenin degradation by DC and β-catenin/TCF complex

formation, respectively. At this point, we could suggest two kinds of multicellular models. One model, in which cells are represented as spheres separated by a small amount of extracellular space. We limit the amount of space through which Wnt must travel in order to reach its receptors, as Wnt ligands are insoluble, unstable and strongly associated with cell membranes and the extra-cellular matrix, thereby inhibiting their ability to diffuse well over long distances (Green et al., 2013). One can examine the effects of modelling Wnt secretion and reception explicitly, for example, on potential delays and stochasticity resulting from Wnt's traversal of the extracellular space. We would set the central sphere in the population to be the Wnt secreting cell, which does not possess any Wnt receptors itself, and model Wnt gene expression in each of the surrounding cells (via [β-TCF]). We could later introduce extra-cellular Wnt antagonists into this model. In our second model, cells are represented as cubes which are in direct contact with each other. This would allow for modelling the effects of cell-cell adhesion. This could be a good opportunity to measure any noticable difference in using spheres versus cubes in terms of the level of Wnt reception and cell-cell adhesion.



(a) (b)

Figure 4.5: Different Smoldyn multi-cell models.

Chapter 5

Outlook

In this thesis, we set out to explore the impact of spatial detail and stochasticity when modelling certain complex biological pathways. We aimed to show conclusively that spatial stochastic models perform better than deterministic models in some specific complex systems. In fact, we demonstrated that spatial and stochastic detail only impacts these systems in a small number of specific cases.

Our Markov model with Holcman et al. ((Holcman et al., 2014)) represents the first stochastic models of the PTEN post-transcriptional regulatory pathway, and the first ever model of RNA interference in the nucleus. One of the primary findings in our paper was that unbinding of miRNAs and copycats actually increases the probability of mRNA degradation, resulting in a decrease in the likelihood of mRNA escape from the domain. This effect is not obvious as both copycats and miRNAs can unbind from their target RNA. In (Holcman et al., 2014), we stated that an explicit solution to the Markov model for the reversible binding case was not possible to obtain. However, in this thesis, we expanded upon the original Markov model and provided our solution for the case of reversible binding. We further expanded on this model by building spatial stochastic models of the PTEN and RNAi pathways.

The results we presented show examples of situations in which stochasticity and spatial detail lead to dramatically different modelling results with the same initial conditions. We have also presented many situations which demonstrate the robustness of the systems we elected to model under various conditions. This is somewhat unexpected as we hypothesised from the outset, based on many studies (such as (Burrage et al., 2011), (Schnell and Turner, 2004) and (Mourão et al., 2014b) to name but a few) that spatial detail and stochasticity should greatly impact the dynamics in largely complex systems with low concentrations of reactant molecules. For example, we have shown that

there exist threshold molecular concentrations, spatial volumes, and crowding fractions, above or below which the systems largely follow the dynamics of the ODE models. Similarly, simply adding macromolecular crowders (obstacles) is not sufficient to dramatically effect the dynamics of bi-molecular interactions or escape from a cellular compartment. Rather, these crowders work best when present in certain volume fractions (e.g. $\theta = 0.2$), diffuse with a certain rate (e.g. slowly), are a specific size (e.g. big enough to be impeded by obstacles), and interact weakly with other molecules in the cell.

In addition to looking at the effects of spatial and stochastic detail on the narrow escape problem (NEP) and the mean time to threshold (MTT), we looked at the impact of excluded-volumes on mean squared displacement (MSD). We compare our simulations to the analytical MSD models by Bruna et al. (Bruna, 2012; Bruna and Chapman, 2012a,b). We see some limits to excluded-volume simulations at high volume fractions ($\theta > 0.2$), at which point we begin to observe a kind of anomalous sub-diffusion (similar to a Lévy flight).

The Wnt system was not fully completed as we were unable to find a bistable model that we could reproduce in Smoldyn, and we did not find reliable measurements of the Wnt system in humans to make our spatial model. Instead, we presented preliminary results and proof of concept with the Endres et al. bistable model in Smoldyn. This work can be expanded upon by applying the model to another biological system with an existing bistable ODE model. Continuation of this model for the inter-cellular Wnt pathway could be used to elucidate the delays in signal transduction, and the effect on cell-cell adhesion, resulting from movement of molecules between cellular compartments in a crowded environment. The first two chapters in this thesis show original spatial and stochastic models of two pathways, which proves that this kind of modelling can add value given the right kind of model.

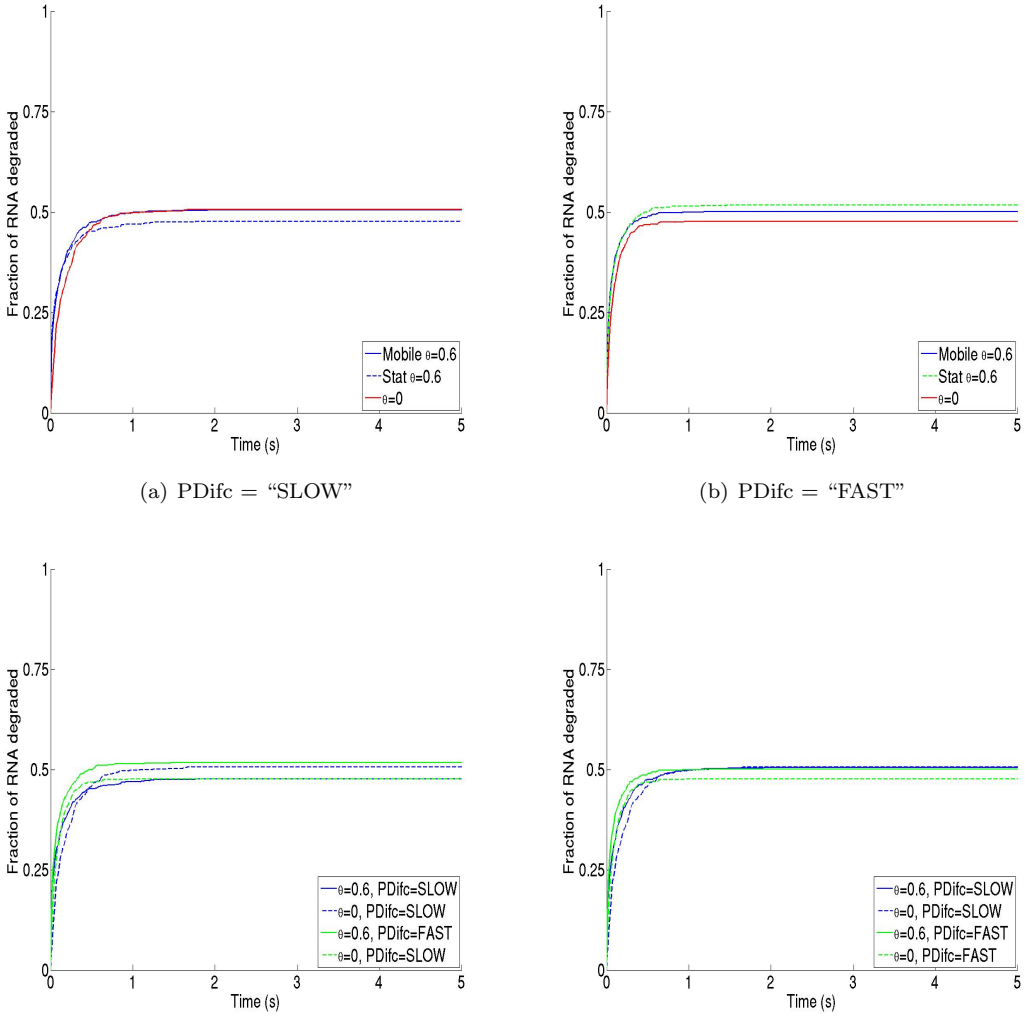
Future work in this field should, in our opinion, be focused on the identification of systems that have not, to date, been well modelled with ODE systems.These studies can be advanced with stochastic and spatial models as we have done, but with the potential to add even more value. A good first extension of the models presented would be to create a hybrid spatial-stochastic model. This would reduce the simulation time compared with a full spatial model, whilst maintaining some degree of spatial detail for interactions that would benefit most greatly from it. For example, the inter-cellular Wnt model could be split into two models; An SSA model for interactions within the cell, with spatial

modelling of the membrane dynamics, such as adhesion, Wnt binding, and Wnt antagonism. Any expansions of the above models should be applied to systems with existing experimental data, as these would aide in model development and have the greatest therapeutic potential.

Appendices

Chapter A

Additional Figures



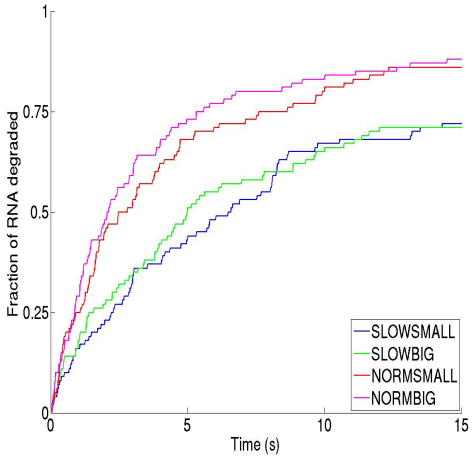
(a) PDifc = "SLOW"

(b) PDifc = "FAST"

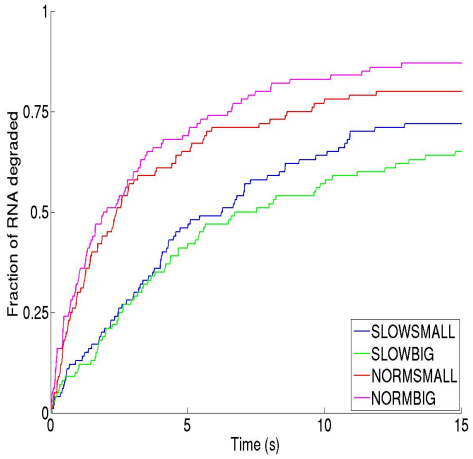
(c) Stat

(d) Mobile

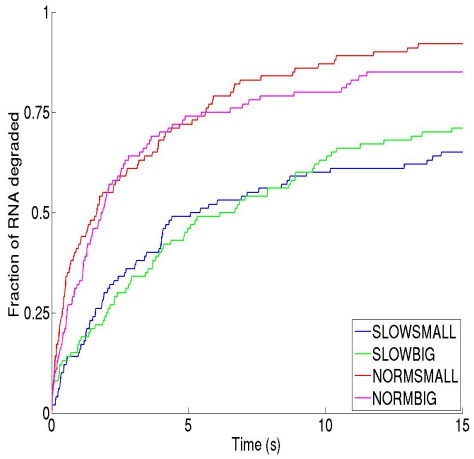
Figure A.1: Fraction of the PTEN mRNA population degraded for big miRNAs moving with unchanged diffusion coefficient ("NORMBIG"; $D_m = 6.67\mu m^2 s^{-1}$). A.1(a)-A.1(b): PTEN mRNA diffusion differs between normal (A.1(a) "SLOW"; $D_p = 0.04\mu m^2 s^{-1}$) and fast (A.1(b) "FAST"; $D_p = 6.67\mu m^2 s^{-1}$). In Figure A.1(c)-A.1(d): We compare stationary obstacles (Figure A.1(c)) to mobile obstacles (Figure A.1(d)). PTEN mRNA diffusion differs between normal ("SLOW"; $D_p = 0.04\mu m^2 s^{-1}$) and fast ("FAST"; $D_p = 6.67\mu m^2 s^{-1}$), and the volume fraction varies between $\theta = \{0, 0.3, 0.6\}$.



(a) $\theta = 0$

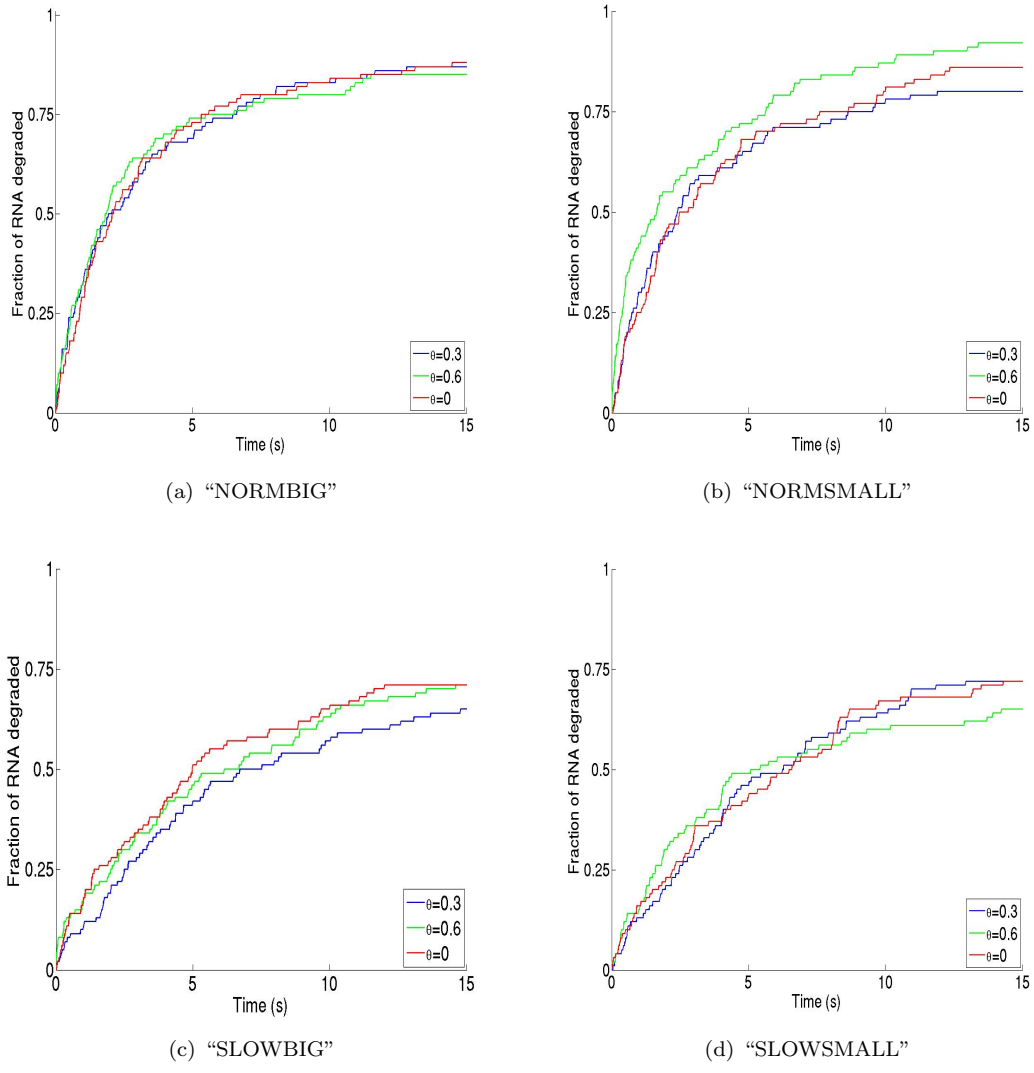


(b) $\theta = 0.3$



(c) $\theta = 0.6$

Figure A.2: Fraction of the PTEN mRNA population degraded for varying miRNA sizes ("*SMALL" and "*BIG") and diffusion coefficents ("NORM*" and "SLOW*") for $\theta = 0, 0.3$ and 0.6. PTEN = miRNA = "LOW".



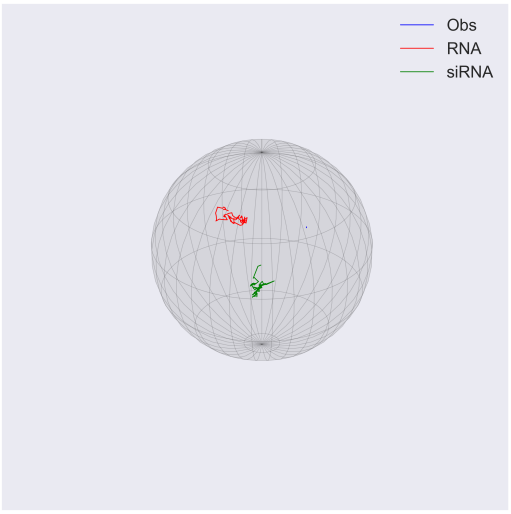
(a) "NORMBIG"

(b) "NORMSMALL"

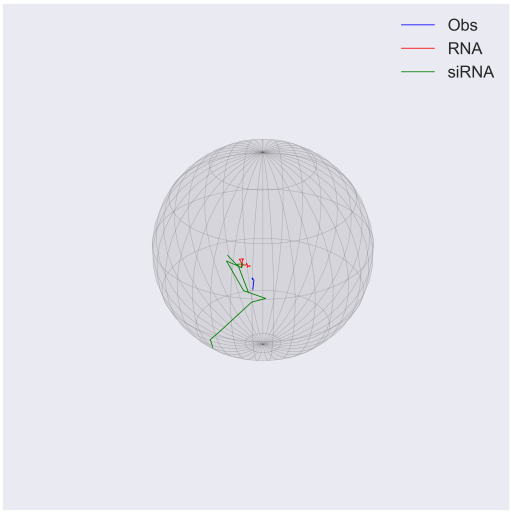
(c) "SLOWBIG"

(d) "SLOWSMALL"

Figure A.3: Fraction of the PTEN mRNA population degraded for each miRNA size ("*SMALL" and "*BIG") and diffusion coefficent ("NORM*" and "SLOW*") for $\theta = 0, 0.3$ and 0.6. PTEN = miRNA = "LOW".

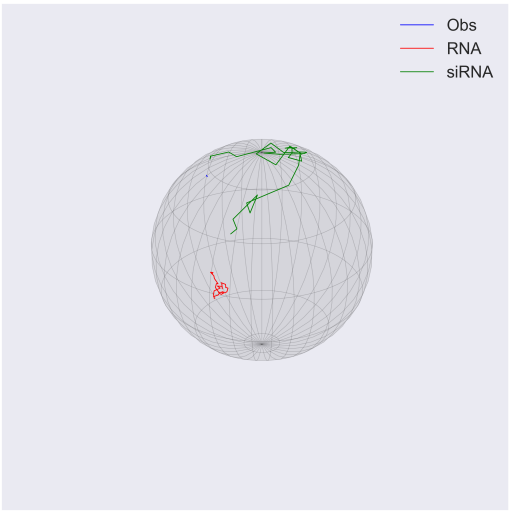


(a) C2: $\theta = 0.1$, $k_f\tau = 10$

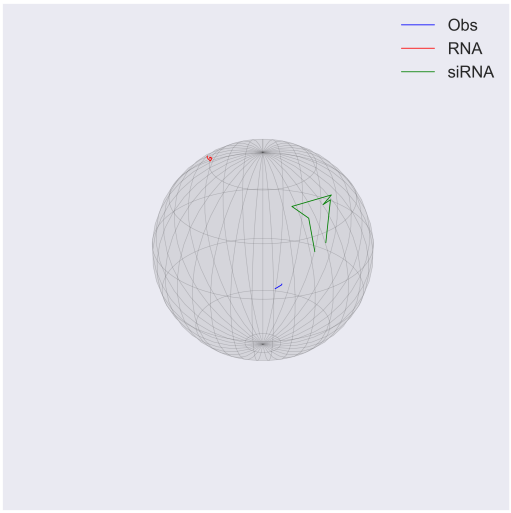


(b) D2: $\theta = 0.3$, $k_f\tau = 10$

Figure A.4: Diffusion of siRNAs, RNAs and obstacles for varying crowding fractions and stationary vs. mobile obstacles. We also show variations in the binding rate, $k_f\tau$.

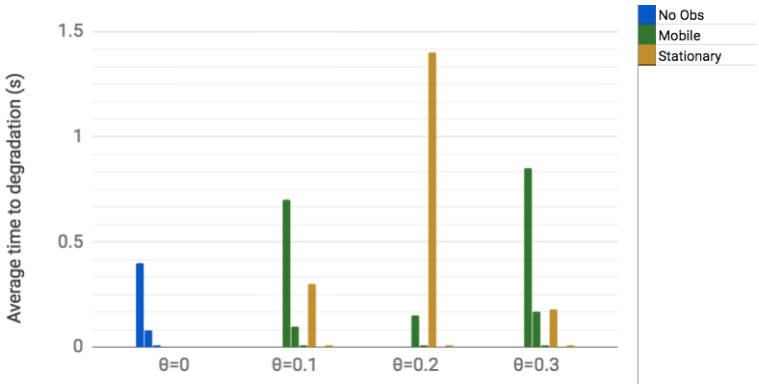


(a) E2: $\theta = 0.2$, $k_f\tau = 10$

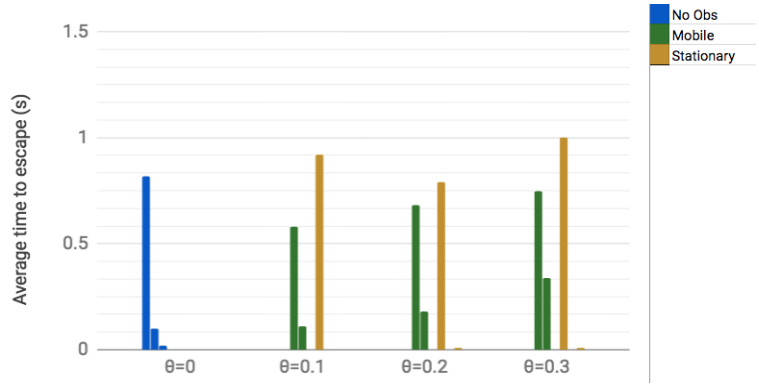


(b) F2: $\theta = 0.3$, $k_f\tau = 10$

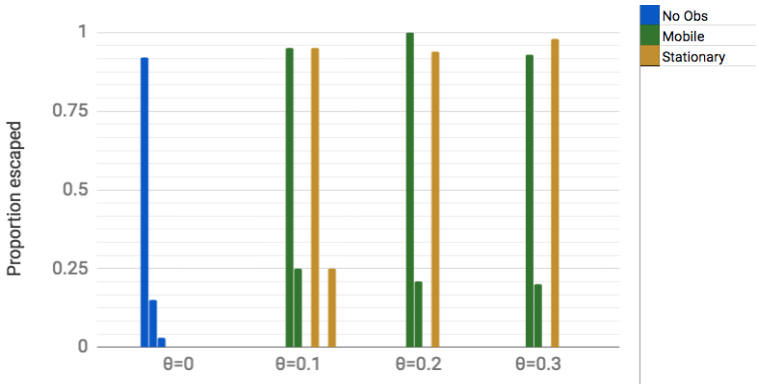
Figure A.5: Diffusion of siRNAs, RNAs and obstacles for varying crowding fractions and stationary obstacles. We also show variations in the binding rate, $k_f\tau$.



(a) A: Average Degradation Time

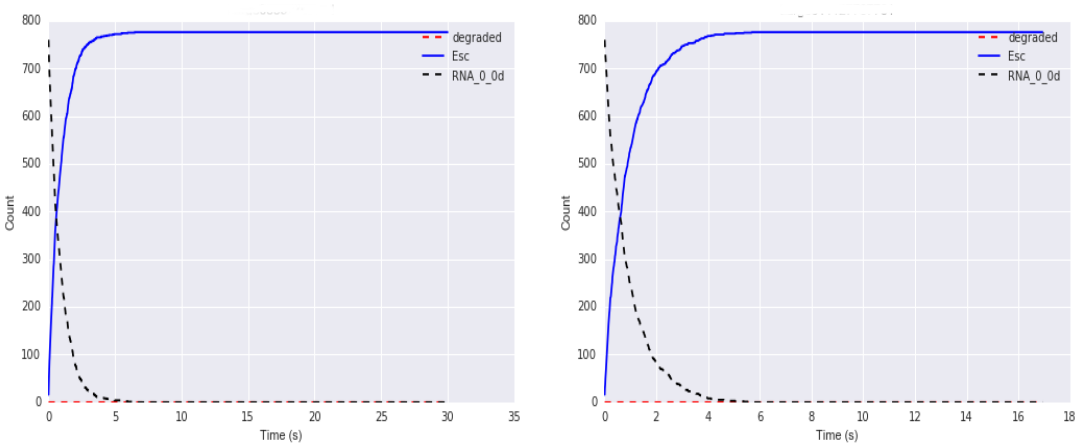


(b) B: Average Escape Time



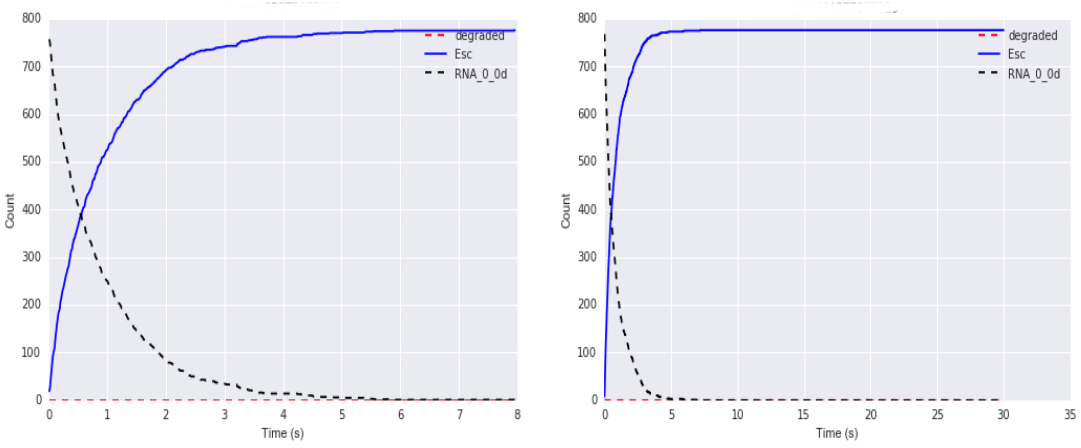
(c) C: Proportion Escaped

Figure A.6: Degradation and escape times for varying numbers of microRNAs. RNAi results for $N_b = 1$, $T = 1$, Mobile Obstacles with radius $0.3\mu m$, $N_p(0) = 1$, and $k_f\tau = 10$. The bars are in groups of three, representing $N_m(0) = 1$, 78, 776 from left to right. Results shown are the average of 55 simulations. For stationary obstacles (yellow bars), we see that volume fraction of $\theta = 0.2$ results in the slowest degradation time. Whereas for mobile obstacles this result is reversed and the degradation time is reduced.



(a) A: $\theta = 0.1$ (b) B: $\theta = 0.2$

Figure A.7: Degradation and escape times for varying crowding fractions. RNAi results for $N_b = 1$, $T = 1$, Mobile Obs with radius $0.3\mu m$, $N_p(0) = 776$, $N_m(0) = 776$, and $k_f\tau = 0.1$. Ave esc times: A: 3.89s, B: 3.88s. Ave deg time for B: 15.69s. Note: No other average degradation times are listed as no other simulations resulted in any RNA degradation.



(a) A: $\theta = 0.1$ (b) B: $\theta = 0.2$

Figure A.8: Degradation and escape times for varying crowding fractions. RNAi results for $N_b = 20, T = 10$, Mobile Obs with radius $0.3\mu m$, $N_p(0) = 776$, $N_m(0) = 776$, and $k_f\tau = 0.1$. Ave esc times: A: 3.64s, B: 3.89s.

Chapter B

Smoldyn Code

```
# set up 3D smoldyn environment
dim 3
species siRNA copycat RNA_degraded Holes Esc Obs
species RNA_0_0d
species RNA_0_1d

# set diffusion coefficients
difc siRNA 6.67
difc copycat 6.67
difc RNA_degraded 0
difc Holes 0
difc Esc 0
difc Obs 0.08756 # made from radius using formula
difc RNA_0_0d 6.67
difc RNA_0_1d 6.67

# set colour for snapshots of simulations
color siRNA 1 0 0
color copycat 0 1 0
color RNA_degraded 0 0 0
color Holes 0 0 0
```

```
color Esc 0 0 0
color Obs 1 1 0
color RNA_0_0d 0.1 0.9 0.2
color RNA_0_1d 0.2 0.2 0.1

# set size of particles for snapshots of simulations
display_size siRNA 2
display_size copycat 2
display_size RNA_degraded 0.1
display_size Holes 10
display_size Esc 0.1
display_size Obs 0.1
display_size RNA_0_0d 5
display_size RNA_0_1d 5

# set start time, end time and time step
time_start 0
time_stop 30
time_step 0.0001

# set size of boundaries
boundaries 0 0 18  # micrometers
boundaries 1 0 18
boundaries 2 0 18
frame_thickness 0

# set cytoplasm/cell surface
start_surface cyt
color both 0.4  0.4  0.4  0.07
action both all reflect
polygon front edge
```

```
thickness 1
panel sph 9 9 9 5.67 50 50 n1
end_surface

# set nucleus surface
start_surface nuc
action both all reflect
polygon front edge
color both  0.8 0.8 0.81 1
panel sph 9 9 9 3.57 80 80 n2
thickness 1
end_surface

# set nucleus compartment
start_compartment nucleus
surface nuc
point 9 9 9
end_compartment

# set cell compartment
start_compartment cell
surface cyt
point 9 9 9
end_compartment

# set cytoplasm compartment as whole cell excluding nucleus
start_compartment cytoplasm
compartment equal cell
compartment andnot nucleus
end_compartment
```

```
# set initial numbers of molecules (and cell surface holes if using)
compartment_mol 776 siRNA nucleus
compartment_mol 0 copycat nucleus
compartment_mol 776 RNA_0_0d nucleus
surface_mol 3000 Holes(back) nuc sphere n1
compartment_mol 836 Obs nucleus

# molecule creation and degradation
reaction_cmpt nucleus k_Rdeg1 RNA_0_0d  -> 0
reaction_rate k_Rdeg1 0.0003
reaction_cmpt nucleus k_Rdeg2 RNA_0_1d  -> 0
reaction_rate k_Rdeg2 0.0003
reaction_cmpt nucleus k_Rprod0 0 -> RNA_0_0d
reaction_rate k_Rprod0 0.2
reaction_cmpt nucleus k_siRprod0 0 -> siRNA
reaction_rate k_siRprod0 0.2
reaction_cmpt nucleus k_siRdeg0 siRNA -> 0
reaction_rate k_siRdeg0 0.0003

# (reflective) interactions between molecules
reaction_cmpt nucleus rxnc0 RNA_0_0d + RNA_0_0d -> RNA_0_0d + RNA_0_0d
binding_radius rxnc0 0.003
product_placement rxnc0 bounce 0.013
reaction_cmpt nucleus rxnc1 RNA_0_0d + RNA_0_1d -> RNA_0_0d + RNA_0_1d
binding_radius rxnc1 0.003
product_placement rxnc1 bounce 0.013
reaction_cmpt nucleus rxnc2 RNA_0_1d + RNA_0_0d -> RNA_0_1d + RNA_0_0d
binding_radius rxnc2 0.003
product_placement rxnc2 bounce 0.013
reaction_cmpt nucleus rxnc3 RNA_0_1d + RNA_0_1d -> RNA_0_1d + RNA_0_1d
binding_radius rxnc3 0.003
```

```
product_placement rxnc3 bounce 0.013

# more (reflective) interactions between molecules
reaction_cmpt nucleus rxnss0 siRNA + siRNA -> siRNA + siRNA
binding_radius rxnss0 0.003
product_placement rxnss0 bounce 0.013
reaction_cmpt nucleus rxnsd0 siRNA + copycat -> siRNA + copycat
binding_radius rxnsd0 0.003
product_placement rxnsd0 bounce 0.013
reaction_cmpt nucleus rxndd0 copycat + copycat -> copycat + copycat
binding_radius rxndd0 0.003
product_placement rxndd0 bounce 0.013

# (reflective) interactions between molecules and obstacles
reaction_cmpt nucleus RNAExc0 RNA_0_0d + Obs -> RNA_0_0d + Obs
binding_radius RNAExc0 0.3215
product_placement RNAExc0 bounce 0.3315
reaction_cmpt nucleus RNAExc1 RNA_0_1d + Obs -> RNA_0_1d + Obs
binding_radius RNAExc1 0.3215
product_placement RNAExc1 bounce 0.3315

# more (reflective) interactions between molecules and obstacles
reaction_cmpt nucleus siRExc0 siRNA + Obs -> siRNA + Obs
binding_radius siRExc0 0.3215
product_placement siRExc0 bounce 0.3315
reaction_cmpt nucleus decExc0 copycat + Obs -> copycat + Obs
binding_radius decExc0 0.3215
product_placement decExc0 bounce 0.3315

# (reflective) interactions between obstacles
reaction_cmpt nucleus ObsExc Obs + Obs -> Obs + Obs
```

```
binding_radius ObsExc 0.64
product_placement ObsExc bounce 0.65

# reaction between RNA and hole, used to model escape
reaction rxn1 Holes(back) + RNA_0_0d(bsoln) -> Holes(back) + Esc
binding_radius rxn1 0.0045

# bi-molecular reactions (reaction network)
reaction_cmpt nucleus k_fs1 siRNA + RNA_0_0d -> RNA_degraded
binding_radius k_fs1 0.003
reaction_probability k_fs1 0.5
reaction_cmpt nucleus k_fd1 copycat + RNA_0_0d -> RNA_0_1d
binding_radius k_fd1 0.003
reaction_probability k_fd1 0.5

output_files outputfile.txt

cmd b molcountheader outputfile.txt
cmd n 100 molcount outputfile.txt

end_file
```

Chapter C

Maria Bruna Mean Squared Displacement Code

```
# Maria Bruna Formula

# 8.177 for sphere. This is L^d=vol(domain) or (4/3)*pi*L^3 for sphere
L = 13.81

# input diffusion coefficient of (blue) moving particles of interest
# e.g. RNA_0_0d
D_b0 = 0.17

# input diffusion coefficient of Obs
D_r0 = 0.17

# non-dimensional diffusion coefficient of (blue) moving particles of interest
# (D_b0*[(1/L^2)/(D_b0/L^2)] = D_b0/D_b0 = 1)
D_b = 1

# non-dimensional diffusion coefficient of (red) obstacles
D_r = 1

# non-dimensional diameter of moving particles (point particle = 0)
```

```
Eps_b = 0

# diameter of obstacles (point particle = 0) (divide by L)
Eps_r = 0.6/L

# number of moving molecules
N_b = 1

# number of obstacles
N_r = 4050

# dimension of system
d = 3

# non-dimensional effective diffusion coefficient of moving particles
D11 = D_b*(1- (N_b - 1)*(Eps_b**d)*(np.pi / d)
      - N_r*( ( (Eps_b + Eps_r) / 2)**d) * (2*np.pi / d)*(D_b / (D_b + D_r) ) )

# re-dimensionalised diffusion coefficient
D_effMB = D11 * D_b0

# Notes:
# If everything is diffusing normally, using <x^2> = 6Dt,
# should find that D_eff = M/6 = D_effMB
# where M = <x^2>/t is the slope of the MSD graph
# Should find that increasing volume fraction decreases diffusion coefficient
# For moving vs stationary obstacles, should find that diffusing obstacles
# boosts moving particle diffusion
```

Chapter D

Maria Bruna Mean Squared Displacement Smoldyn Model

```
# Maria Bruna volume exclusion model:


# This model has only one molecule (of RNA_0_0d) which is a point particle
# Reactions with obstacles (Obs) are volume excluded
# D_obs = 0, Phi_Obs = 10%, Phi_RNA = 0, D_RNA = 0.044 mu^2 s^-1
# Boundaries are periodic to avoid problems of reflection periodicity
# eps_obs = diameter = 0.3, eps_RNA = 0.0001
# Aim is to try to represent moving particles as a point particle but still
# have it exluded from the obstacles.
# No nucleus. Instead, replaced by smaller compartment using the
# fact that
# vol_cyt = (9^3)*4*pi/3 - ((5.67^3)*4*pi/3) = (8.18^3)*4*pi/3



dim 3
species RNA_0_0d Obs

difc RNA_0_0d 0.17
difc Obs 0
```

```
color RNA_0_0d 0.5 0.4 0.5
color Obs 0.6 0.8 1

display_size RNA_0_0d 5
display_size Obs 1

time_start 0
time_stop 30
time_step 0.0001

boundaries 0 0 18 a #micrometers
boundaries 1 0 18 a
boundaries 2 0 18 a
frame_thickness 0

start_surface cyt
color both 0.4  0.4  0.4  0.07
polygon front edge
thickness 1
panel sph 9 9 9 8.18 50 50 c1
action both all transmit
end_surface

start_compartment cytoplasm
surface cyt
point 9 9 9
end_compartment

compartment_mol 1 RNA_0_0d cytoplasm
compartment_mol 0 Obs cytoplasm
```

```
reaction_cmpt cytoplasm RNAExc0 RNA_0_0d + Obs -> RNA_0_0d + Obs
binding_radius RNAExc0 0.3
product_placement RNAExc0 bounce 0.31

output_files outputfile.txt msdfile.txt
cmd b molcountheader outputfile.txt
cmd n 100 molcountincmpt cytoplasm outputfile.txt
cmd n 10 meansqrdisp RNA_0_0d all msdfile.txt
end_file
```

Chapter E

PTEN SSA Python Model

```
import numpy as np
import pandas as pd
from scipy import integrate
import matplotlib.pyplot as plt
import scipy

fig = plt.figure()
plt.hold(True)
plt.grid('on')

for number_runs in range(100):
    alphas = np.zeros((1, 31))
    t_final = 5000
    tau = np.zeros((1,1))
    t = 0

    Nb = 4
    T = 2
    numsir = 776
    numdec = 776
    numrna = 776
```

```
alpha = 1


k_fm = 0

k_bm = 0

k_fd = 0

k_bd = 0

k_fp = 0

k_bp = 0

k_fpm = 0.00005

k_fpd = 0.00005

k_bpm = 0.0005

k_bpd = 0.0005


R_0 = 776*np.ones((1,1)) #miR

R_1 = 776*np.ones((1,1)) #dec

R_2 = 776*np.ones((1,1)) #RNA_0_0d

R_3 = 0*np.ones((1,1)) #RNA_0_1d

R_4 = 0*np.ones((1,1)) #RNA_0_2d

R_5 = 0*np.ones((1,1)) #RNA_0_3d

R_6 = 0*np.ones((1,1)) #RNA_0_4d

R_7 = 0*np.ones((1,1)) #RNA_1_0d

R_8 = 0*np.ones((1,1)) #RNA_1_1d

R_9 = 0*np.ones((1,1)) #RNA_1_2d

R_10 = 0*np.ones((1,1)) #RNA_1_3d

R_11 = 0*np.ones((1,1)) #RNA_2_0d

R_12 = 0*np.ones((1,1)) #RNA_2_1d

R_13 = 0*np.ones((1,1)) #RNA_2_2d


alphas[0,0] = k_fm #miR trans

alphas[0,1] = k_bm*R_0[0,t] #miR deg

alphas[0,2] = k_fd #copycat trans
```

```
alphas[0,3] = k_bd*R_1[0,t] #copycat deg
alphas[0,4] = k_fp #mRNA trans
alphas[0,5] = k_bp*R_2[0,t] #mRNA deg
alphas[0,6] =  k_bpd*R_3[0,t] #RNA_0_1d  -> RNA_0_0d + d
alphas[0,7] =  k_bpd*R_4[0,t] #RNA_0_2d  -> RNA_0_1d + d
alphas[0,8] =  k_bpd*R_5[0,t] #RNA_0_3d  -> RNA_0_2d + d
alphas[0,9] =  k_bpd*R_6[0,t] #RNA_0_4d  -> RNA_0_3d + d
alphas[0,10] =  k_bpd*R_8[0,t] #RNA_1_1d  -> RNA_1_0d + d
alphas[0,11] =  k_bpd*R_9[0,t] #RNA_1_2d  -> RNA_1_1d + d
alphas[0,12] =  k_bpd*R_10[0,t] #RNA_1_3d  -> RNA_1_2d + d
alphas[0,13] =  k_bpm*R_7[0,t] #RNA_1_0d  -> RNA_0_0d + m
alphas[0,14] =  k_bpm*R_8[0,t] #RNA_1_1d  -> RNA_0_1d + m
alphas[0,15] =  k_bpm*R_9[0,t] #RNA_1_2d  -> RNA_0_2d + m
alphas[0,16] =  k_bpm*R_10[0,t] #RNA_1_3d  -> RNA_0_3d + m
alphas[0,17] =  k_fpd*R_2[0,t]*R_1[0,t] #RNA_0_0d + d  -> RNA_0_1d
alphas[0,18] =  k_fpd*R_3[0,t]*R_1[0,t] #RNA_0_1d + d  -> RNA_0_2d
alphas[0,19] =  k_fpd*R_4[0,t]*R_1[0,t] #RNA_0_2d + d  -> RNA_0_3d
alphas[0,20] =  k_fpd*R_5[0,t]*R_1[0,t] #RNA_0_3d + d  -> RNA_0_4d
alphas[0,21] =  k_fpd*R_7[0,t]*R_1[0,t] #RNA_1_0d + d  -> RNA_1_1d
alphas[0,22] =  k_fpd*R_8[0,t]*R_1[0,t] #RNA_1_1d + d  -> RNA_1_2d
alphas[0,23] =  k_fpd*R_9[0,t]*R_1[0,t] #RNA_1_2d + d  -> RNA_1_3d
alphas[0,24] =  k_fpm*R_2[0,t]*R_0[0,t] #RNA_0_0d + m  -> RNA_1_0d
alphas[0,25] =  k_fpm*R_3[0,t]*R_0[0,t] #RNA_0_1d + m  -> RNA_1_1d
alphas[0,26] =  k_fpm*R_4[0,t]*R_0[0,t] #RNA_0_2d + m  -> RNA_1_2d
alphas[0,27] =  k_fpm*R_5[0,t]*R_0[0,t] #RNA_0_3d + m  -> RNA_1_3d
alphas[0,28] =  k_fpm*R_7[0,t]*R_0[0,t] #RNA_1_0d + m  -> RNA_2_0d
alphas[0,29] =  k_fpm*R_8[0,t]*R_0[0,t] #RNA_1_1d + m  -> RNA_2_1d
alphas[0,30] =  k_fpm*R_9[0,t]*R_0[0,t] #RNA_1_2d + m  -> RNA_2_2d


TAU = 0
while TAU < t_final:
```

```
r1 = np.random.uniform(0,1)
while r1 == 0:
    r1 = np.random.uniform(0,1)

r2 = np.random.uniform(0,1)
alpha0 = sum(alphas[0,:])
tau1 = (1.0/alpha0) * np.math.log(1/r1)
if alpha0 == 0:
    tau1 = t_final - TAU + 10
    alpha0 = 1

TAU = TAU +tau1
t = t+1

R_0_temp = np.zeros((1,t+1))
R_1_temp = np.zeros((1,t+1))
R_2_temp = np.zeros((1,t+1))
R_3_temp = np.zeros((1,t+1))
R_4_temp = np.zeros((1,t+1))
R_5_temp = np.zeros((1,t+1))
R_6_temp = np.zeros((1,t+1))
R_7_temp = np.zeros((1,t+1))
R_8_temp = np.zeros((1,t+1))
R_9_temp = np.zeros((1,t+1))
R_10_temp = np.zeros((1,t+1))
R_11_temp = np.zeros((1,t+1))
R_12_temp = np.zeros((1,t+1))
R_13_temp = np.zeros((1,t+1))

R_0_temp[0,0:(t)] = R_0
R_1_temp[0,0:(t)] = R_1
```

```
R_2_temp[0,0:(t)] = R_2

R_3_temp[0,0:(t)] = R_3

R_4_temp[0,0:(t)] = R_4

R_5_temp[0,0:(t)] = R_5

R_6_temp[0,0:(t)] = R_6

R_7_temp[0,0:(t)] = R_7

R_8_temp[0,0:(t)] = R_8

R_9_temp[0,0:(t)] = R_9

R_10_temp[0,0:(t)] = R_10

R_11_temp[0,0:(t)] = R_11

R_12_temp[0,0:(t)] = R_12

R_13_temp[0,0:(t)] = R_13


R_0 = R_0_temp

R_1 = R_1_temp

R_2 = R_2_temp

R_3 = R_3_temp

R_4 = R_4_temp

R_5 = R_5_temp

R_6 = R_6_temp

R_7 = R_7_temp

R_8 = R_8_temp

R_9 = R_9_temp

R_10 = R_10_temp

R_11 = R_11_temp

R_12 = R_12_temp

R_13 = R_13_temp


tau_temp = np.zeros((1,t+1))

tau_temp[0,0:t] = tau

tau = tau_temp
```

```
tau[0,t] = TAU


if r2 >= 0 and r2 < sum(alphas[0,0:1])/alpha0:
    R_0[0,t] = R_0[0,t-1] + 1
    R_1[0,t] = R_1[0,t-1]
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]


elif r2 >= sum(alphas[0,0:1])/alpha0
 and r2 < sum(alphas[0,0:2])/alpha0 and R_0[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1] - 1
    R_1[0,t] = R_1[0,t-1]
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
```

```
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:2])/alpha0
 and r2 < sum(alphas[0,0:3])/alpha0:
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:3])/alpha0
 and r2 < sum(alphas[0,0:4])/alpha0  and R_1[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] - 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
```

```
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:4])/alpha0
 and r2 < sum(alphas[0,0:5])/alpha0:
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1]
    R_2[0,t] = R_2[0,t-1] + 1
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]
elif r2 >= sum(alphas[0,0:5])/alpha0
 and r2 < sum(alphas[0,0:6])/alpha0 and R_2[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1]
    R_2[0,t] = R_2[0,t-1] - 1
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
```

```
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:6])/alpha0
 and r2 < sum(alphas[0,0:7])/alpha0 and R_3[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1] + 1
    R_3[0,t] = R_3[0,t-1] - 1
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:7])/alpha0
 and r2 < sum(alphas[0,0:8])/alpha0 and R_4[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] + 1
    R_3[0,t] = R_3[0,t-1] + 1
```

```
    R_4[0,t] = R_4[0,t-1] - 1
    R_2[0,t] = R_2[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:8])/alpha0
 and r2 < sum(alphas[0,0:9])/alpha0 and R_5[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] + 1
    R_4[0,t] = R_4[0,t-1] + 1
    R_5[0,t] = R_5[0,t-1] - 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:9])/alpha0
 and r2 < sum(alphas[0,0:10])/alpha0 and R_6[0,t-1] > 0:
```

```
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] + 1
    R_5[0,t] = R_5[0,t-1] + 1
    R_6[0,t] = R_6[0,t-1] - 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:10])/alpha0
 and r2 < sum(alphas[0,0:11])/alpha0 and R_8[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] + 1
    R_7[0,t] = R_7[0,t-1] + 1
    R_8[0,t] = R_8[0,t-1] - 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]
```

```
elif r2 >= sum(alphas[0,0:11])/alpha0
 and r2 < sum(alphas[0,0:12])/alpha0 and R_9[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] + 1
    R_8[0,t] = R_8[0,t-1] + 1
    R_9[0,t] = R_9[0,t-1] - 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]


elif r2 >= sum(alphas[0,0:12])/alpha0
 and r2 < sum(alphas[0,0:13])/alpha0 and R_10[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] + 1
    R_9[0,t] = R_9[0,t-1] + 1
    R_10[0,t] = R_10[0,t-1] - 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
```

```
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:13])/alpha0
 and r2 < sum(alphas[0,0:14])/alpha0 and R_7[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1] + 1
    R_1[0,t] = R_1[0,t-1]
    R_7[0,t] = R_7[0,t-1] - 1
    R_2[0,t] = R_2[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:14])/alpha0
 and r2 < sum(alphas[0,0:15])/alpha0 and R_8[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1] + 1
    R_1[0,t] = R_1[0,t-1]
    R_8[0,t] = R_8[0,t-1] - 1
    R_3[0,t] = R_3[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
```

```
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:15])/alpha0
 and r2 < sum(alphas[0,0:16])/alpha0 and R_9[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1] + 1
    R_1[0,t] = R_1[0,t-1]
    R_9[0,t] = R_9[0,t-1] - 1
    R_4[0,t] = R_4[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:16])/alpha0
 and r2 < sum(alphas[0,0:17])/alpha0 and R_10[0,t-1] > 0:
    R_0[0,t] = R_0[0,t-1] + 1
    R_1[0,t] = R_1[0,t-1]
    R_10[0,t] = R_10[0,t-1] - 1
    R_5[0,t] = R_5[0,t-1] + 1
```

```
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:17])/alpha0
 and r2 < sum(alphas[0,0:18])/alpha0 and R_2[0,t-1] > 0 and R_1[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] - 1
    R_2[0,t] = R_2[0,t-1] - 1
    R_3[0,t] = R_3[0,t-1] + 1
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:18])/alpha0
 and r2 < sum(alphas[0,0:19])/alpha0 and R_3[0,t-1] > 0 and R_1[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1]
```

```
        R_1[0,t] = R_1[0,t-1] - 1
        R_3[0,t] = R_3[0,t-1] - 1
        R_4[0,t] = R_4[0,t-1] + 1
        R_2[0,t] = R_2[0,t-1]
        R_5[0,t] = R_5[0,t-1]
        R_6[0,t] = R_6[0,t-1]
        R_7[0,t] = R_7[0,t-1]
        R_8[0,t] = R_8[0,t-1]
        R_9[0,t] = R_9[0,t-1]
        R_10[0,t] = R_10[0,t-1]
        R_11[0,t] = R_11[0,t-1]
        R_12[0,t] = R_12[0,t-1]
        R_13[0,t] = R_13[0,t-1]

    elif r2 >= sum(alphas[0,0:19])/alpha0
     and r2 < sum(alphas[0,0:20])/alpha0 and R_4[0,t-1] > 0 and R_1[0,t-1] > 0 :
        R_0[0,t] = R_0[0,t-1]
        R_1[0,t] = R_1[0,t-1] - 1
        R_4[0,t] = R_4[0,t-1] - 1
        R_5[0,t] = R_5[0,t-1] + 1
        R_2[0,t] = R_2[0,t-1]
        R_3[0,t] = R_3[0,t-1]
        R_6[0,t] = R_6[0,t-1]
        R_7[0,t] = R_7[0,t-1]
        R_8[0,t] = R_8[0,t-1]
        R_9[0,t] = R_9[0,t-1]
        R_10[0,t] = R_10[0,t-1]
        R_11[0,t] = R_11[0,t-1]
        R_12[0,t] = R_12[0,t-1]
        R_13[0,t] = R_13[0,t-1]
```

```
elif r2 >= sum(alphas[0,0:20])/alpha0
 and r2 < sum(alphas[0,0:21])/alpha0 and R_5[0,t-1] > 0 and R_1[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] - 1
    R_5[0,t] = R_5[0,t-1] - 1
    R_6[0,t] = R_6[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]


elif r2 >= sum(alphas[0,0:21])/alpha0
 and r2 < sum(alphas[0,0:22])/alpha0 and R_7[0,t-1] > 0 and R_1[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] - 1
    R_7[0,t] = R_7[0,t-1] - 1
    R_8[0,t] = R_8[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
```

```
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:22])/alpha0
 and r2 < sum(alphas[0,0:23])/alpha0 and R_8[0,t-1] > 0 and R_1[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] - 1
    R_8[0,t] = R_8[0,t-1] - 1
    R_9[0,t] = R_9[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:23])/alpha0
 and r2 < sum(alphas[0,0:24])/alpha0 and R_9[0,t-1] > 0 and R_1[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1] - 1
    R_9[0,t] = R_9[0,t-1] - 1
    R_10[0,t] = R_10[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
```

```
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:24])/alpha0
 and r2 < sum(alphas[0,0:25])/alpha0 and R_2[0,t-1] > 0 and R_0[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1] - 1
    R_1[0,t] = R_1[0,t-1]
    R_2[0,t] = R_2[0,t-1] - 1
    R_7[0,t] = R_7[0,t-1] + 1
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:25])/alpha0
 and r2 < sum(alphas[0,0:26])/alpha0 and R_3[0,t-1] > 0 and R_0[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1] - 1
    R_1[0,t] = R_1[0,t-1]
    R_3[0,t] = R_3[0,t-1] - 1
    R_8[0,t] = R_8[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_4[0,t] = R_4[0,t-1]
```

```
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:26])/alpha0
 and r2 < sum(alphas[0,0:27])/alpha0 and R_4[0,t-1] > 0 and R_0[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1] - 1
    R_1[0,t] = R_1[0,t-1]
    R_4[0,t] = R_4[0,t-1] - 1
    R_9[0,t] = R_9[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:27])/alpha0
 and r2 < sum(alphas[0,0:28])/alpha0 and R_5[0,t-1] > 0 and R_0[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1] - 1
    R_1[0,t] = R_1[0,t-1]
    R_5[0,t] = R_5[0,t-1] - 1
```

```
    R_10[0,t] = R_10[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:28])/alpha0
 and r2 < sum(alphas[0,0:29])/alpha0 and R_7[0,t-1] > 0 and R_0[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1] - 1
    R_1[0,t] = R_1[0,t-1]
    R_7[0,t] = R_7[0,t-1] - 1
    R_11[0,t] = R_11[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:29])/alpha0
 and r2 < sum(alphas[0,0:30])/alpha0 and R_8[0,t-1] > 0 and R_0[0,t-1] > 0 :
```

```
    R_0[0,t] = R_0[0,t-1] - 1
    R_1[0,t] = R_1[0,t-1]
    R_8[0,t] = R_8[0,t-1] - 1
    R_12[0,t] = R_12[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_13[0,t] = R_13[0,t-1]

elif r2 >= sum(alphas[0,0:30])/alpha0
 and r2 < sum(alphas[0,0:31])/alpha0 and R_9[0,t-1] > 0 and R_0[0,t-1] > 0 :
    R_0[0,t] = R_0[0,t-1] - 1
    R_1[0,t] = R_1[0,t-1]
    R_9[0,t] = R_9[0,t-1] - 1
    R_13[0,t] = R_13[0,t-1] + 1
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
```

```
else:
    R_0[0,t] = R_0[0,t-1]
    R_1[0,t] = R_1[0,t-1]
    R_2[0,t] = R_2[0,t-1]
    R_3[0,t] = R_3[0,t-1]
    R_4[0,t] = R_4[0,t-1]
    R_5[0,t] = R_5[0,t-1]
    R_6[0,t] = R_6[0,t-1]
    R_7[0,t] = R_7[0,t-1]
    R_8[0,t] = R_8[0,t-1]
    R_9[0,t] = R_9[0,t-1]
    R_10[0,t] = R_10[0,t-1]
    R_11[0,t] = R_11[0,t-1]
    R_12[0,t] = R_12[0,t-1]
    R_13[0,t] = R_13[0,t-1]

alphas[0,0] = k_fm #miR trans
alphas[0,1] = k_bm*R_0[0,t] #miR deg
alphas[0,2] = k_fd #copycat trans
alphas[0,3] = k_bd*R_1[0,t] #copycat deg
alphas[0,4] = k_fp #mRNA trans
alphas[0,5] = k_bp*R_2[0,t] #mRNA deg
alphas[0,6] =  k_bpd*R_3[0,t] #RNA_0_1d  -> RNA_0_0d + d
alphas[0,7] =  k_bpd*R_4[0,t] #RNA_0_2d  -> RNA_0_1d + d
alphas[0,8] =  k_bpd*R_5[0,t] #RNA_0_3d  -> RNA_0_2d + d
alphas[0,9] =  k_bpd*R_6[0,t] #RNA_0_4d  -> RNA_0_3d + d
alphas[0,10] =  k_bpd*R_8[0,t] #RNA_1_1d  -> RNA_1_0d + d
alphas[0,11] =  k_bpd*R_9[0,t] #RNA_1_2d  -> RNA_1_1d + d
alphas[0,12] =  k_bpd*R_10[0,t] #RNA_1_3d  -> RNA_1_2d + d
alphas[0,13] =  k_bpm*R_7[0,t] #RNA_1_0d  -> RNA_0_0d + m
```

```
        alphas[0,14] =  k_bpm*R_8[0,t] #RNA_1_1d  -> RNA_0_1d + m
        alphas[0,15] =  k_bpm*R_9[0,t] #RNA_1_2d  -> RNA_0_2d + m
        alphas[0,16] =  k_bpm*R_10[0,t] #RNA_1_3d  -> RNA_0_3d + m
        alphas[0,17] =  k_fpd*R_2[0,t]*R_1[0,t] #RNA_0_0d + d  -> RNA_0_1d
        alphas[0,18] =  k_fpd*R_3[0,t]*R_1[0,t] #RNA_0_1d + d  -> RNA_0_2d
        alphas[0,19] =  k_fpd*R_4[0,t]*R_1[0,t] #RNA_0_2d + d  -> RNA_0_3d
        alphas[0,20] =  k_fpd*R_5[0,t]*R_1[0,t] #RNA_0_3d + d  -> RNA_0_4d
        alphas[0,21] =  k_fpd*R_7[0,t]*R_1[0,t] #RNA_1_0d + d  -> RNA_1_1d
        alphas[0,22] =  k_fpd*R_8[0,t]*R_1[0,t] #RNA_1_1d + d  -> RNA_1_2d
        alphas[0,23] =  k_fpd*R_9[0,t]*R_1[0,t] #RNA_1_2d + d  -> RNA_1_3d
        alphas[0,24] =  k_fpm*R_2[0,t]*R_0[0,t] #RNA_0_0d + m  -> RNA_1_0d
        alphas[0,25] =  k_fpm*R_3[0,t]*R_0[0,t] #RNA_0_1d + m  -> RNA_1_1d
        alphas[0,26] =  k_fpm*R_4[0,t]*R_0[0,t] #RNA_0_2d + m  -> RNA_1_2d
        alphas[0,27] =  k_fpm*R_5[0,t]*R_0[0,t] #RNA_0_3d + m  -> RNA_1_3d
        alphas[0,28] =  k_fpm*R_7[0,t]*R_0[0,t] #RNA_1_0d + m  -> RNA_2_0d
        alphas[0,29] =  k_fpm*R_8[0,t]*R_0[0,t] #RNA_1_1d + m  -> RNA_2_1d
        alphas[0,30] =  k_fpm*R_9[0,t]*R_0[0,t] #RNA_1_2d + m  -> RNA_2_2d

    plt.plot(tau[0,:]/1000, R_2[0,:]/numrna, 'r--')
    RNAs = R_11[0,:] + R_12[0,:] + R_13[0,:]
    plt.plot(tau[0,:]/1000, (RNAs)/numrna, 'b--')
    plt.hold(True)

plt.plot(tau[0,:]/1000, R_2[0,:]/numrna, 'r--', label='RNA')
RNAs = R_11[0,:] + R_12[0,:] + R_13[0,:]
plt.plot(tau[0,:]/1000, (RNAs)/numrna, 'b--', label='RNA_deg')
plt.title('SSA siRs/Decs = 776')
plt.xlabel('Time (s)')
plt.ylabel('Fraction of PTEN mRNA')
plt.xlim(0, 5)
plt.ylim(0, 1)
```

```
plt.legend(loc='right')

plt.show()

plt.savefig('figurename.png')
```

Chapter F

PTEN ODE Python Model

```
import numpy as np
from scipy import integrate
import matplotlib.pyplot as plt
import scipy
import seaborn as sns

param = 6.7321

def AlaODEForThesis2016(t, y, param):

    alpha = 1
    K = param

    k_fm = 0
    k_bm = 0
    k_fd = 0
    k_bd = 0
    k_fp = 0
    k_bp = 0
    k_fs = 0
    k_bs = 0
```

```
    k_fpm = K
    k_0pm = 0
    k_fpd = K
    k_bpd = 0
    k_0pd = 0
    k_fsm = 0
    k_bsm = 0
    k_fsd = 0
    k_bsd = 0

    dydt = np.zeros((8, 1))
    dydt[0] = k_fm - y[0]*k_bm - y[0]*(y[2]*k_fpm + y[3]*k_fsm)
              + y[6]*k_bsm + (1-alpha)*y[4]*k_0pm
    dydt[1] = k_fd - y[1]*k_bd - y[1]*(y[2]*k_fpd + y[3]*k_fsd)
              + y[5]*k_bpd + y[7]*k_bsd
    dydt[2] = k_fp - y[2]*k_bp - y[2]*(y[0]*k_fpm + y[1]*k_fpd)
              + y[5]*k_bpd
    dydt[3] = k_fs - y[3]*k_bs - y[3]*(y[0]*k_fsm + y[1]*k_fsd)
              + y[6]*k_bsm + y[7]*k_bsd
    dydt[4] = y[2]*y[0]*k_fpm - y[4]*k_0pm
    dydt[5] = -y[5]*k_0pd + y[2]*y[1]*k_fpd - y[5]*k_bpd
    dydt[6] = y[3]*y[0]*k_fsm - y[6]*k_bsm
    dydt[7] = y[3]*y[1]*k_fsd - y[7]*k_bsd
    return dydt

# Integrating the ODE(s):
if __name__ == '__main__':

    # Integrator: use ``vode'' with ``backward differentiation formula''
    r = integrate.ode(AlaODEForThesis2016).set_integrator('vode', method='bdf')
```

```
# Time range
t_start = 0.0
t_final = 5.0
delta_t = 0.01

# Number of time steps: 1 extra for initial condition
num_steps = np.floor((t_final - t_start)/delta_t) + 1



# Set initial condition(s):
y_init = np.array([0.563, 0.563, 0.563, 0, 0, 0, 0, 0])
r.set_initial_value(y_init, t_start)

r.set_f_params(par)
# Create vectors to store trajectories

t = np.zeros((num_steps, 1))
X = np.zeros((num_steps, 8))
t[0] = t_start
X[0] = y_init
pars = np.zeros((num_steps, 8))
pars[0] = y_init

# Integrate the ODE(s) across timestep, delta_t
k = 0
while r.successful() and k < num_steps:
    r.integrate(r.t + delta_t)

    # Store the results to plot later
    t[k] = r.t
    X[k, :] = r.y
```

```
    k += 1


# Convert from concentrations to molecule number
a = 6.022*10**(23)*2290*10**(-6)*10**(-18)
X *= a


# Convert to proportion of initial concentration
X /= 776  # (y_init[0]*a)


# Plot trajectories:
plt.hold(True)
sns.set_style("darkgrid")
plt.grid('on')
plt.plot(t, X[:, 2], 'r--', label='Unbound')
plt.plot(t, X[:, 4], 'b', label='Degraded')
plt.title('PTEN_ODECoopObs30Stat')
plt.xlabel('Time (s)')
plt.ylabel('Molecule count')
plt.xlim(0, 5)
plt.ylim(0, 1)
plt.legend(3loc='right')
plt.figure(figsize=(12, 14))
```

References

Ala, U., Karreth, F. A., Bosia, C., Pagnani, A., Taulli, R., Léopold, V., Tay, Y., Provero, P., Zecchina, R., and Pandolfi, P. P. (2013). Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *PNAS*, 110(18):7154–7159.

Alberts, B., Johnson, A., Lewis, J., Raff, M., and Roberts, K. (2002). *Molecular Biology of the Cell 4th edition*. Garland Science.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2008). *Molecular Biology of the Cell*. Garland Science, 5 edition.

Alder, B. J. and Wainwright, T. E. (1967). Velocity autocorrelations for hard spheres. *Physical review letters*, 18(23):988.

Alder, B. J. and Wainwright, T. E. (1970). Decay of the velocity autocorrelation function. *Physical review A*, 1(1):18.

Ambs, S., Prueitt, R. L., Yi, M., Hudson, R. S., Howe, T. M., Petrocca, F., Wallace, T. A., Liu, C. G., Volinia, S., Calin, G. A., Yfantis, H. G., Stephens, R. M., and Croce, C. M. (2008). Genomic Profiling of MicroRNA and Messenger RNA Reveals Deregulated MicroRNA Expression in Prostate Cancer. *Cancer research*, 68(15):6162–6170.

Ameres, S. L., Martinez, J., and Schroeder, R. (2007). Molecular Basis for Target RNA Recognition and Cleavage by Human RISC. *Cell*, 130(1):101–112.

Andrews, S. S. (2011). Spatial and Stochastic Cellular Modeling with the Smoldyn Simulator. pages 519–542. Springer New York, New York, NY.

Andrews, S. S., Addy, N. J., Brent, R., and Arkin, A. P. (2010). Detailed simulations of cell biology with Smoldyn 2.1. *PLoS Computational Biology*, 6(3):e1000705.

Andrews, S. S. and Bray, D. (2004). Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Physical Biology*, 1(3):137–151.

Arvey, A., Larsson, E., Sander, C., Leslie, C. S., and Marks, D. S. (2010). Target mRNA abundance dilutes microRNA and siRNA activity. *Molecular Systems Biology*, 6:1–7.

Baccarini, A., Chauhan, H., Gardner, T. J., Jayaprakash, A. D., Sachidanandam, R., and Brown, B. D. (2011). Kinetic Analysis Reveals the Fate of a MicroRNA following Target Regulation in Mammalian Cells. *Current Biology*, 21(5):369–376.

Bakalkin, G., Yakovleva, T., and Terenius, L. (1997). The Leu-enkephalin-encoding sequence DNA-binding factor (LEF) is the transcription factor YY1. *Biochemical and biophysical research communications*, 231(1):135–139.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–297.

Basan, M., Idema, T., Lenz, M., Joanny, J.-F., and Risler, T. (2010). A Reaction-Diffusion Model of the Cadherin-Catenin System: A Possible Mechanism for Contact Inhibition and Implications for Tumorigenesis. *Biophysical Journal*, 98(12):2770–2779.

Berezhkovskii, A. M., Dagdug, L., and Bezrukov, S. M. (2014). Discriminating between Anomalous Diffusion and Transient Behavior inMicroheterogeneous Environments. *Biophysical Journal*, 106(2):L09–L11.

Berezhna, S. Y., Supekova, L., Supek, F., Schultz, P. G., and Deniz, A. A. (2006). siRNA in human cells selectively localizes to target RNA sites. *PNAS*, 103(20):7682–7687.

Berger, A. H., Knudson, A. G., and Pandolfi, P. P. (2011). A continuum model for tumour suppression. *Nature*, 476(7359):163–169.

Berry, H. and Chaté, H. (2014). Anomalous diffusion due to hindering by mobile obstacles undergoing Brownian motion or Orstein-Ulhenbeck processes. *Physical Review E*, 89(2):022708.

Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I., and Filipowicz, W. (2006). Re-

lief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress. *Cell*, 125(6):1111–1124.

Bienz, M. and Clevers, H. (2000). Linking colorectal cancer to Wnt signaling. *Cell*, 103(2):311–320.

Bosia, C., Pagnani, A., and Zecchina, R. (2013). Modelling Competing Endogenous RNA Networks. *PLoS ONE*, 8(6):e66609.

Brodersen, P. and Voinnet, O. (2009). Revisiting the principles of microRNA target recognition and mode of action. *Nature Reviews Molecular Cell Biology*, 10(2):141–148.

Bruna, M. (2012). *Excluded-volume Effects in Stochastic Models of Diffusion*. PhD thesis.

Bruna, M. and Chapman, S. J. (2012a). Diffusion of multiple species with excluded-volume effects. *The Journal of Chemical Physics*, 137(20):204116.

Bruna, M. and Chapman, S. J. (2012b). Excluded-volume effects in the diffusion of hard spheres. *Physical Review E*, 85(1):011103.

Bruna, M. and Chapman, S. J. (2013). Diffusion of Finite-Size Particles in Confined Geometries. *Bulletin of mathematical biology*, 76(4):947–982.

Burnett, J. C., Rossi, J. J., and Tiemann, K. (2011). Current progress of siRNA/shRNA therapeutics in clinical trials. *Biotechnology Journal*, 6(9):1130–1146.

Burrage, K., Burrage, P. M., Leier, A., Marquez-Lago, T., and Nicolau, Jr, D. V. (2011). Stochastic simulation for spatial modelling of dynamic processes in a living cell. In *Design and Analysis of Biomolecular Circuits*, pages 43–62. Springer.

Cai, J., Guan, H., Fang, L., Yang, Y., Zhu, X., Yuan, J., Wu, J., and Li, M. (2013). MicroRNA-374a activates Wnt/β-catenin signaling to promote breast cancer metastasis. *Journal of Clinical Investigation*.

Calin, G. A., Dumitru, C. D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., and Rai, K. (2002). Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *PNAS*, 99(24):15524–15529.

Carthew, R. W. and Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 136(4):642–655.

Castel, S. E. and Martienssen, R. A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nature Reviews Genetics*, 14(2):100–112.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA. *Cell*, 147(2):358–369.

Chatterjee, S., Fasler, M., Büssing, I., and Großhans, H. (2011). Short Article. *Developmental Cell*, 20(3):388–396.

Cheviakov, A. F., Ward, M. J., and Straube, R. (2010). An asymptotic analysis of the mean first passage time for narrow escape problems: Part ii: The sphere. *Multiscale Modeling & Simulation*, 8(3):836–870.

Cheviakov, A. F. and Zawada, D. (2013). Narrow-escape problem for the unit sphere: Homogenization limit, optimal arrangements of large numbers of traps, and the N^2 conjecture. *Physical Review E*, 87(4-1):042118.

Cho, K.-H., Baek, S., and Sung, M.-H. (2006). Wnt pathway mutations selected by optimal β-catenin signaling for tumorigenesis. *FEBS Letters*, 580(15):3665–3670.

Chu, E. C. and Tarnawski, A. S. (2004). PTEN regulatory functions in tumor suppression and cell biology. *Medical science monitor : international medical journal of experimental and clinical research*, 10(10):RA235–41.

Cuccato, G., Polynikis, A., Siciliano, V., Graziano, M., di Bernardo, M., and di Bernardo, D. (2011). Modeling RNA interference in mammalian cells. *BMC Systems Biology*, 5(1):19.

Dawes, A. T. (2009). A mathematical model of alpha-catenin dimerization at adherens junctions in polarized epithelial cells. *Journal of Theoretical Biology*, 257(3):480–488.

de Giorgio, A., Krell, J., Harding, V., Stebbing, J., and Castellano, L. (2013). Emerging Roles of Competing Endogenous RNAs in Cancer: Insights from the Regulation of PTEN. *Molecular and Cellular Biology*, 33(20):3976–3982.

Di Cristofano, A. and Pandolfi, P. P. (2000). The multiple roles of PTEN in tumor suppression. *Cell*, 100(4):387–390.

Ebert, M. S. (2010). *Molecular Titration by MicroRNAs and Target Mimic Inhibitors*. PhD thesis, MIT.

Ebert, M. S., Neilson, J. R., and Sharp, P. A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature Methods*, 4(9):721–726.

Eiring, A. M., Harb, J. G., Neviani, P., Garton, C., Oaks, J. J., Spizzo, R., Liu, S., Schwind, S., Santhanam, R., Hickey, C. J., Becker, H., Chandler, J. C., Andino, R., Cortes, J., Hokland, P., Huettner, C. S., Bhatia, R., Roy, D. C., Liebhaber, S. A., Caligiuri, M. A., Marcucci, G., Garzon, R., Croce, C. M., Calin, G. A., and Perrotti, D. (2010). miR-328 Functions as an RNA Decoy to Modulate hnRNP E2 Regulation of mRNA Translation in Leukemic Blasts. *Cell*, 140(5):652–665.

El-Samad, H., Khammash, M., Petzold, L., and Gillespie, D. (2005). Stochastic modelling of gene regulatory networks. *International Journal of Robust and Nonlinear Control*, 15(15):691–711.

Elcock, A. H. (2010). Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Current Opinion in Structural Biology*, 20(2):196–206.

Endres, R. G. (2015). Bistability: Requirements on Cell-Volume, Protein Diffusion, and Thermodynamics. *PLoS ONE*, 10(4):e0121681.

Erban, R. (2014). From molecular dynamics to Brownian dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 470(2167):20140036–20140036.

Erban, R., Chapman, J., and Maini, P. (2007). A practical guide to stochastic simulations of reaction-diffusion processes. *arXiv.org*.

Erban, R. and Chapman, S. J. (2009). Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions. *Physical Biology*, 6(4):046001.

Espenson, J. (2002). *Chemical Kinetics and Reaction Mechanisms*. McGraw-Hill Science/Engineering/Math.

Fabian, M. R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA Translation and Stability by microRNAs. *Annual Review of Biochemistry*, 79(1):351–379.

Feng, J., Sun, G., Yan, J., Noltner, K., Li, W., Buzin, C. H., Longmate, J., Heston, L. L., Rossi, J., and Sommer, S. S. (2009). Evidence for X-Chromosomal Schizophrenia Associated with microRNA Alterations. *PLoS ONE*, 4(7):e6121.

Fletcher, A. G., Breward, C. J. W., and Chapman, S. J. (2012). Mathematical modelling of monoclonal conversion in the colonic crypt. *Journal of Theoretical Biology*, 300(c):118–133.

Franosch, T., Höfling, F., Bauer, T., and Frey, E. (2010). Chemical Physics. *Chemical Physics*, 375(2-3):540–547.

Fulton, A. B. (1982). How crowded is the cytoplasm? *Cell*, 30(2):345–347.

Fusco, D., Accornero, N., Lavoie, B., Shenoy, S. M., Blanchard, J. M., Singer, R. H., and Bertrand, E. (2003). Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Current biology : CB*, 13(2):161–167.

Gagnon, K. T. and Corey, D. R. (2012). Argonaute and the nuclear RNAs: new pathways for RNA-mediated control of gene expression. *Nucleic acid therapeutics*, 22(1):3–16.

Gagnon, K. T., Li, L., Chu, Y., Janowski, B. A., and Corey, D. R. (2014). RNAi Factors Are Presentand Active in Human Cell Nuclei. *CellReports*, 6(1):211–221.

Georgescu, M. M. (2011). PTEN Tumor Suppressor Network in PI3K-Akt Pathway Control. *Genes & Cancer*, 1(12):1170–1177.

Gershon, N. D., Porter, K. R., and Trus, B. L. (1985). The cytoplasmic matrix: its volume and surface area and the diffusion of molecules through it. *PNAS*, 82(15):5030–5034.

Giles, R. H., van Es, J. H., and Clevers, H. (2003). Caught up in a Wnt storm: Wnt signaling in cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1653(1):1–24.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361.

Gillespie, D. T. (2000). The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306.

Gillespie, D. T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733.

Giza, D. E., Vasilescu, C., and Calin, G. A. (2014). MicroRNAs and ceRNAs: therapeutic implications of RNA networks. *Expert Opinion on Biological Therapy*, 14(9):000–000.

Goentoro, L. and Kirschner, M. W. (2009). Evidence that Fold-Change, and Not Absolute Level, of β-Catenin Dictates Wnt Signaling. *Molecular Cell*, 36(5):872–884.

Green, J. L., La, J., Yum, K. W., Desai, P., Rodewald, L.-W., Zhang, X., Leblanc, M., Nusse, R., Lewis, M. T., and Wahl, G. M. (2013). Paracrine Wnt signaling both promotes and inhibits human breast tumor growth. *PNAS*, 110(17):6991–6996.

Griffiths-Jones, S. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, 34(90001):D140–D144.

Grima, R. and Schnell, S. (2006). A systematic investigation of the rate laws valid in intracellular environments. *Biophysical chemistry*, 124(1):1–10.

Gross, J. C. and Boutros, M. (2013). Secretion and extracellular space travel of Wnt proteins. *Current Opinion in Genetics & Development*, 23(4):385–390.

Grünwald, D. and Singer, R. H. (2010). In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport. *Nature*, 467(7315):604–607.

Guan, Y. and Krone, S. M. (2004). WinSSS: Stochastic Spatial Simulator. *Bulletin of the Ecological Society of America*, 85(3):102–104.

Guo, H., Ingolia, N. T., Weissman, J. S., and Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308):835–840.

Hansen, T. B., Kjems, J., and Damgaard, C. K. (2013). Circular RNA and miR-7 in Cancer. *Cancer research*, 73(18):5609–5612.

Hansen, T. B., Wiklund, E. D., Bramsen, J. B., Villadsen, S. B., Statham, A. L., Clark, S. J., and Kjems, J. o. r. (2011). miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *The EMBO Journal*, 30(21):4414–4422.

Henderson, B. R. and Fagotto, F. (2002). The ins and outs of APC and beta-catenin nuclear transport. *EMBO reports*, 3(9):834–839.

Höfling, F. and Franosch, T. (2013). Anomalous transport in the crowded world of biological cells. *Reports on Progress in Physics*, 76(4):046602.

Höfling, F., Munk, T., Frey, E., and Franosch, T. (2008). Critical dynamics of ballistic and Brownian particles in a heterogeneous environment. *The Journal of Chemical Physics*, 128(16):164517.

Holcman, D., Dao Duc, K., Jones, A., Byrne, H., and Burrage, K. (2014). Post-transcriptional regulation in the nucleus and cytoplasm: study of mean time to threshold (MTT) and narrow escape problem. *Journal of Mathematical Biology*, 70(4):805–828.

Hollander, M. C., Blumenthal, G. M., and Dennis, P. A. (2011). PTEN loss in the continuum of common cancers, rare syndromes and mouse models. *Nature Reviews Cancer*, 11(4):289–301.

Hon, L. S. and Zhang, Z. (2007). The roles of binding site arrangement and combinatorial targeting in microRNA repression of gene expression. *Genome Biology*, 8(8):R166.

Hudson, R. S., Yi, M., Esposito, D., Glynn, S. A., Starks, A. M., Yang, Y., Schetter, A. J., Watkins, S. K., Hurwitz, A. A., Dorsey, T. H., Stephens, R. M., Croce, C. M., and Ambs, S. (2012). MicroRNA-106b-25 cluster expression is associated with early disease recurrence and targets caspase-7 and focal adhesion in human prostate cancer. *Oncogene*, 32(35):4139–4147.

Huse, J. T., Brennan, C., Hambardzumyan, D., Wee, B., Pena, J., Rouhanifard, S. H., Sohn-Lee, C., le Sage, C., Agami, R., Tuschl, T., and Holland, E. C. (2009). The PTEN-regulating microRNA miR-26a is amplified in high-grade glioma and facilitates gliomagenesis in vivo. *Genes & Development*, 23(11):1327–1337.

Hutvagner, G. (2002). A microRNA in a Multiple-Turnover RNAi Enzyme Complex. *Science*, 297(5589):2056–2060.

Hwang, H. W., Wentzel, E. A., and Mendell, J. T. (2007). A Hexanucleotide Element Directs MicroRNA Nuclear Import. *Science*, 315(5808):97–100.

Iorio, M. V., Ferracin, M., Liu, C. G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M., Menard, S., Palazzo, J. P., Rosenberg, A., Musiani, P., Volinia, S., Nenci, I., Calin, G. A., Querzoli, P., Negrini, M., and Croce, C. M. (2005). MicroRNA Gene Expression Deregulation in Human Breast Cancer. *Cancer research*, 65(16):7065–7070.

Janowski, B. A., Huffman, K. E., Schwartz, J. C., Ram, R., Nordsell, R., Shames, D. S., Minna, J. D., and Corey, D. R. (2006). Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nature Structural & Molecular Biology*, 13(9):787–792.

Jeffries, C. D., Fried, H. M., and Perkins, D. O. (2011). Nuclear and cytoplasmic localization of neural stem cell microRNAs. *RNA*, 17(4):675–686.

Jha, S. K. and Langmead, C. J. (2012). Exploring behaviors of stochastic differentialequation models of biological systems usingchange of measures. *BMC Bioinformatics*, 13(Suppl 5):S8.

Kammerer, A., Höfling, F., and Franosch, T. (2008). Cluster-resolved dynamic scaling theory and universal corrections for transport on percolating systems. *EPL (Europhysics Letters)*, 84(6):66002.

Karreth, F. A., Tay, Y., Perna, D., Ala, U., Tan, S. M., Rust, A. G., DeNicola, G., Webster, K. A., Weiss, D., Perez-Mancera, P. A., Krauthammer, M., Halaban, R., Provero, P., Adams, D. J., Tuveson, D. A., and Pandolfi, P. P. (2011). In Vivo Identification of Tumor- Suppressive PTEN ceRNAs in an Oncogenic BRAF-Induced Mouse Model of Melanoma. *Cell*, 147(2):382–395.

Kartha, R. V. and Subramanian, S. (2014). Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. *Frontiers in genetics*, 5:8.

Kemler, R. (1993). From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion. *Trends in genetics : TIG*, 9(9):317–321.

Khan, S., Zou, Y., Amjad, A., Gardezi, A., Smith, C. L., Winters, C., and Reese, T. S. (2011). Sequestration of CaMKII in dendritic spines in silico. *Journal of Computational Neuroscience*, 31(3):581–594.

Khanin, R. and Higham, D. J. (2007). Chemical Master Equation and Langevin regimes for a gene transcription model. *Computational Methods in Systems Biology*, pages 1–14.

Khanin, R. and Higham, D. J. (2009). Mathematical and computational modelling of post-transcriptional gene regulation by microRNAs. *Handbook of Statistical Systems Biology*.

Kim, W., Kim, M., and Jho, E. h. (2013). Wnt/β-catenin signalling: from plasma membrane to nucleus. *Biochemical Journal*, 450(1):9–21.

Klann, M., Ganguly, A., and Koeppl, H. (2012). Hybrid spatial Gillespie and particle tracking simulation. *Bioinformatics*, 28(18):i549–i555.

Kogan, Y., Halevi Tobias, K. E., Hochman, G., Baczmanska, A. K., Leyns, L., and Agur, Z. (2012). A

new validated mathematical model of the Wnt signalling pathway predicts effective combinational therapy by sFRP and Dkk. *Biochemical Journal*, 444(1):115–125.

Kopelman, R. (1986). Rate processes on fractals: theory, simulations, and experiments. *Journal of Statistical Physics*, 42(1-2):185–200.

Kopelman, R. (1988). Fractal Reaction-Kinetics. *Science*, 241(4873):1620–1626.

Krichevsky, A. M., King, K. S., Donahue, C. P., Khrapko, K., and Kosik, K. S. (2003). A microRNA array reveals extensive regulation of microRNAs during brain development. *RNA*, 9(10):1274–1281.

Krieghoff, E. (2006). Nucleo-cytoplasmic distribution of beta-catenin is regulated by retention. *Journal of Cell Science*, 119(7):1453–1463.

Krone, S. M. (2004). Spatial models: stochastic and deterministic. *Mathematical and Computer Modelling*, 40(3-4):393–409.

Langlois, M. A. (2005). Cytoplasmic and Nuclear Retained DMPK mRNAs Are Targets for RNA Interference in Myotonic Dystrophy Cells. *Journal of Biological Chemistry*, 280(17):16949–16954.

Lee, E., Salic, A., Krüger, R., Heinrich, R., and Kirschner, M. W. (2003). The Roles of APC and Axin Derived from Experimental and Theoretical Analysis of the Wnt Pathway. *PLoS Biology*, 1(1):e10.

Lee, E. J., Baek, M., Gusev, Y., Brackett, D. J., Nuovo, G. J., and Schmittgen, T. D. (2007). Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. *RNA*, 14(1):35–42.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854.

Leslie, N. R. and Foti, M. (2011). Non-genomic loss of PTEN function in cancer: not in my genes. *Trends in Pharmacological Sciences*, 32(3):131–140.

Li, D. M. and Sun, H. (1997). TEP1, encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor beta. *Cancer research*, 57(11):2124–2129.

Li, Z. and Rana, T. M. (2012). Molecular Mechanisms of RNA-Triggered Gene Silencing Machineries. *Accounts of Chemical Research*, 45(7):1122–1131.

Liang, H., Zhang, J., Zen, K., Zhang, C.-Y., and Chen, X. (2013). Nuclear microRNAs and their unconventional role in regulating non-coding RNAs. *Protein & Cell*, 4(5):325–330.

Liao, J.-Y., Ma, L.-M., Guo, Y.-H., Zhang, Y.-C., Zhou, H., Shao, P., Chen, Y.-Q., and Qu, L.-H. (2010). Deep Sequencing of Human Nuclear and Cytoplasmic Small RNAs Reveals an Unexpectedly Complex Subcellular Distribution of miRNAs and tRNA 3 Trailers. *PLoS ONE*, 5(5):e10563.

Lipkow, K. and Odde, D. J. (2008). Model for Protein Concentration Gradients in the Cytoplasm. *Cellular and Molecular Bioengineering*, 1(1):84–92.

Liu, J. (2004). Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science*, 305(5689):1437–1441.

Lloyd-Lewis, B., Fletcher, A. G., Dale, T. C., and Byrne, H. M. (2013). Toward a quantitative understanding of the Wnt/ β-catenin pathway through simulation and experiment. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, pages n/a–n/a.

Lorenz, C. D. and Ziff, R. M. (2001). Precise determination of the critical percolation threshold for the three-dimensional "Swiss cheese" model using a growth algorithm. *The Journal of Chemical Physics*, 114(8):3659.

Luby-Phelps, K., Taylor, D. L., and Lanni, F. (1986). Probing the structure of cytoplasm. *The Journal of cell biology*, 102(6):2015–2022.

MacDonald, B. T., Tamai, K., and He, X. (2009). Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. *Developmental Cell*, 17(1):9–26.

Macfarlane, L.-A. and Murphy, P. R. (2010). MicroRNA: Biogenesis, Function and Role in Cancer. *Current genomics*, 11(7):537–561.

MacLean, A. L., Rosen, Z., Byrne, H. M., and Harrington, H. A. (2015). Parameter-free methods distinguish Wnt pathway models and guide design of experiments. *PNAS U.S.A*, q-bio.QM(9):2652–2657.

Marquez-Lago, T. T., Leier, A., and Burrage, K. (2012). Anomalous diffusion and multifractional

Brownian motion: simulating molecular crowding and physical obstacles in systems biology. *IET systems biology*, 6(4):134–142.

Matsuda, H., Putzel, G. G., Backman, V., and Szleifer, I. (2014). Macromolecular Crowding as a Regulator of Gene Transcription. *Biophysical Journal*, 106(8):1801–1810.

Matzke, M. A. and Birchler, J. A. (2005). RNAi-mediated pathways in the nucleus. *Nature Reviews Genetics*, 6(1):24–35.

Mazemondet, O., John, M., Leye, S., Rolfs, A., and Uhrmacher, A. M. (2012). Elucidating the Sources of β-Catenin Dynamics in Human Neural Progenitor Cells. *PLoS ONE*, 7(8):e42792.

Mazemondet, O., John, M., Maus, C., Uhrmacher, A. M., and Rolfs, A. (2009). Integrating diverse reaction types into stochastic models: a signaling pathway case study in the imperative π-calculus. In Rossetti, M. D., Hill, R. R., Johansson, B., Dunkin, A., and Ingalls, R. G., editors, *Winter Simulation Conference*, pages 932–943. Winter Simulation Conference.

Mazo, R. M. (2008). *Brownian Motion*. Fluctuations, Dynamics, and Applications. OUP Oxford.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell*, 15(2):185–197.

Meng, F., Henson, R., Lang, M., Wehbe, H., Maheshwari, S., Mendell, J. T., Jiang, J., Schmittgen, T. D., and Patel, T. (2006). Involvement of Human Micro-RNA in Growth and Response to Chemotherapy in Human Cholangiocarcinoma Cell Lines. *Gastroenterology*, 130(7):2113–2129.

Mettetal, J. T. and van Oudenaarden, A. (2007). Necessary Noise. *Science*, 317(5837):463–464.

Minton, A. P. (1983). The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Molecular and cellular biochemistry*, 55(2):119–140.

Minton, A. P. (2001). The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. *Journal of Biological Chemistry*, 276(14):10577–10580.

Mirams, G. R., Byrne, H. M., and King, J. R. (2009). A multiple timescale analysis of a mathematical model of the Wnt/β-catenin signalling pathway. *Journal of Mathematical Biology*, 60(1):131–160.

Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., Leith, J., and Kosmrlj, A. (2009). How a protein searches for its site on dna: the mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical*, 42(43):434013.

Mourão, M., Kreitman, D., and Schnell, S. (2014a). Unravelling the impact of obstacles in diffusion and kinetics of an enzyme catalysed reaction. *Physical Chemistry Chemical Physics*, 16(10):4492.

Mourão, M. A., Hakim, J. B., and Schnell, S. (2014b). Biophysical Review. *Biophysical Journal*, 107(12):2761–2766.

Mukherji, S., Ebert, M. S., Zheng, G. X. Y., Tsang, J. S., Sharp, P. A., and van Oudenaarden, A. (2011). MicroRNAs can generate thresholds in target gene expression. *Nature Genetics*, 43(9):854–859.

Neumann, S., Coudreuse, D. Y. M., van der Westhuyzen, D. R., Eckhardt, E. R. M., Korswagen, H. C., Schmitz, G., and Sprong, H. (2009). Mammalian Wnt3a is Released on Lipoprotein Particles. *Traffic*, 10(3):334–343.

Neves, S. R. and Iyengar, R. (2009). Models of Spatially Restricted Biochemical Reaction Systems. *Journal of Biological Chemistry*, 284(9):5445–5449.

Nicolau, Jr, D. V. and Burrage, K. (2008). Stochastic simulation of chemical reactions in spatially complex media. *Computers & Mathematics with Applications*, 55(5):1007–1018.

Nicolau, Jr, D. V., Hancock, J. F., and Burrage, K. (2007). Sources of Anomalous Diffusion on Cell Membranes: A Monte Carlo Study. *Biophysical Journal*, 92(6):1975–1987.

Nusse, R. and Varmus, H. E. (1982). Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell*, 31(1):99–109.

Nüsslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature*, 287(5785):795–801.

Orsulic, S., Huber, O., Aberle, H., Arnold, S., and Kemler, R. (1999). E-cadherin binding prevents beta-catenin nuclear localization and beta-catenin/LEF-1-mediated transactivation. *Journal of Cell Science*, 112 (Pt 8):1237–1245.

Paul, C. P., Good, P. D., Winer, I., and Engelke, D. R. (2002). Effective expression of small interfering RNA in human cells. *Nature biotechnology*, 20(5):505–508.

Pedrero, J. M. G., Carracedo, D. G., Pinto, C. M. o., Zapatero, A. n. H., Rodrigo, J. P., Nieto, C. S., and Gonzalez, M. V. (2005). Frequent genetic and biochemical alterations of the PI3-K/AKT/PTEN pathway in head and neck squamous cell carcinoma. *International Journal of Cancer*, 114(2):242–248.

Perren, A., Komminoth, P., Saremaslani, P., Matter, C., Feurer, S., Lees, J. A., Heitz, P. U., and Eng, C. (2000). Mutation and expression analyses reveal differential subcellular compartmentalization of PTEN in endocrine pancreatic tumors compared to normal islet cells. *The American Journal of Pathology*, 157(4):1097–1103.

Petersen, C. P., Bordeleau, M.-E., Pelletier, J., and Sharp, P. A. (2006). Short RNAs Repress Translation after Initiation in Mammalian Cells. *Molecular Cell*, 21(4):533–542.

Poliseno, L., Salmena, L., Riccardi, L., Fornari, A., Song, M. S., Hobbs, R. M., Sportoletti, P., Varmeh, S., Egia, A., Fedele, G., Rameh, L., Loda, M., and Pandolfi, P. P. (2010a). Identification of the miR-106b 25 MicroRNA Cluster as a Proto-Oncogenic PTEN-Targeting Intron That Cooperates with Its Host Gene MCM7 in Transformation. *Science Signaling*, 3(117):ra29–ra29.

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., and Pandolfi, P. P. (2010b). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, 465(7301):1033–1038.

Politz, J. C. R., Zhang, F., and Pederson, T. (2006). MicroRNA-206 colocalizes with ribosome-rich regions in both the nucleolus and cytoplasm of rat myogenic cells. *PNAS*, 103(50):18957–18962.

Ragan, C., Zuker, M., and Ragan, M. A. (2011). Quantitative Prediction of miRNA-mRNA Interaction Based on Equilibrium Concentrations. *PLoS Computational Biology*, 7(2):e1001090.

Reingruber, J., Abad, E., and Holcman, D. (2009). Narrow escape time to a structured target located on the boundary of a microdomain. *The Journal of Chemical Physics*, 130(9):094909.

Rintoul, M. D. and Torquato, S. (1997). Precise determination of the critical threshold and exponents in a three-dimensional continuum percolation model. *Journal of Physics A: Mathematical and General*, 30(16):L585.

Robb, G. B., Brown, K. M., Khurana, J., and Rana, T. M. (2005). Specific and potent RNAi in the nucleus of human cells. *Nature Structural & Molecular Biology*, 12(2):133–137.

Roberts, T. C. (2014). The MicroRNA Biology of the Mammalian Nucleus. *Molecular Therapy—Nucleic Acids*, 3(8):e188.

Rowan, A. J., Lamlum, H., Ilyas, M., Wheeler, J., Straub, J., Papadopoulou, A., Bicknell, D., Bodmer, W. F., and Tomlinson, I. P. M. (2000). APC Mutations in Sporadic Colorectal Tumors: A Mutational "Hotspot" and Interdependence of the "Two Hits". *PNAS U.S.A*, 97(7):3352–3357.

Salmena, L., Carracedo, A., and Pandolfi, P. P. (2008). Tenets of PTEN Tumor Suppression. *Cell*, 133(3):403–414.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell*, 146(3):353–358.

Saxton, M. J. (1994). Anomalous diffusion due to obstacles: A Monte Carlo study. *Biophysical Journal*, 66(2):394–401.

Saxton, M. J. (1996). Anomalous diffusion due to binding: A Monte Carlo study. *Biophysical Journal*, 70(3):1250–1262.

Saxton, M. J. (2007a). A Biological Interpretation of Transient Anomalous Subdiffusion. I. Qualitative Model. *Biophysical Journal*, 92(4):1178–1191.

Saxton, M. J. (2007b). Modeling 2D and 3D diffusion. *Methods in molecular biology (Clifton, N.J.)*, 400:295–321.

Schmitz, Y. (2014). Compartmental Modelling of the Wnt Pathway: Elucidating the role of nucleocytoplasmic shuttling of β-catenin and its antagonists. *Cell Signalling*, pages 1–143.

Schmitz, Y., Rateitschak, K., and Wolkenhauer, O. (2013). Analysing the impact of nucleocytoplasmic shuttling of β-catenin and its antagonists APC, Axin and GSK3 on Wnt/β-catenin signalling. *Cellular Signalling*, 25(11):2210–2221.

Schnell, S. and Turner, T. E. (2004). Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Progress in Biophysics and Molecular Biology*, 85(2-3):235–260.

Schuss, Z., Singer, A., and Holcman, D. (2007). The narrow escape problem for diffusion in cellular microdomains. *PNAS*, 104(41):16098–16103.

Seitz, H. (2009). Redefining MicroRNA Targets. *Current Biology*, 19(10):870–873.

Sharma, R. P. and Chopra, V. L. (1976). Effect of the Wingless (wg1) mutation on wing and haltere development in Drosophila melanogaster. *Developmental Biology*, 48(2):461–465.

Song, M. S., Salmena, L., and Pandolfi, P. P. (2012). The functions and regulation of the PTEN tumour suppressor. *Nature Reviews Molecular Cell Biology*, 13(5):283–296.

Steck, P. A., Pershouse, M. A., Jasser, S. A., Yung, W. K., Lin, H., Ligon, A. H., Langford, L. A., Baumgard, M. L., Hattier, T., Davis, T., Frye, C., Hu, R., Swedlund, B., Teng, D. H., and Tavtigian, S. V. (1997). Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nature Genetics*, 15(4):356–362.

Sumazin, P., Yang, X., Chiu, H.-S., Chung, W.-J., Iyer, A., Llobet-Navas, D., Rajbhandari, P., Bansal, M., Guarnieri, P., Silva, J., and Califano, A. (2011). An Extensive MicroRNA-Mediated Network of RNA-RNA Interactions Regulates Established Oncogenic Pathways in Glioblastoma. *Cell*, 147(2):370–381.

Takebe, N., Harris, P. J., Warren, R. Q., and Ivy, S. P. (2010). Targeting cancer stem cells by inhibiting Wnt, Notch, and Hedgehog pathways. *Nature Publishing Group*, 8(2):97–106.

Tamguney, T. and Stokoe, D. (2007). New insights into PTEN. *Journal of Cell Science*, 120(23):4071–4079.

Tan, C. W., Gardiner, B. S., Hirokawa, Y., Layton, M. J., Smith, D. W., and Burgess, A. W. (2012). Wnt Signalling Pathway Parameters for Mammalian Cells. *PLoS ONE*, 7(2):e31882.

Tang, R., Li, L., Zhu, D., Hou, D., Cao, T., Gu, H., Zhang, J., Chen, J., Zhang, C.-Y., and Zen, K. (2011). Mouse miRNA-709 directly regulates miRNA-15a/16-1 biogenesis at the post-transcriptional level in the nucleus: evidence for a microRNA hierarchy system. *Cell Research*, 22(3):504–515.

Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S. M., Ala, U., Karreth, F., Poliseno, L., Provero, P., Di Cunto, F., Lieberman, J., Rigoutsos, I., and Pandolfi, P. P. (2011). Coding-Independent Regulation of the Tumor Suppressor PTEN by Competing Endogenous mRNAs. *Cell*, 147(2):344–357.

Tay, Y., Rinn, J., and Pandolfi, P. P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature*, 505(7483):344–352.

Tian, L., Fang, Y.-x., Xue, J.-l., and Chen, J.-z. (2013). Four microRNAs promote prostate cell proliferation with regulation of PTEN and its downstream signals in vitro. *PLoS ONE*, 8(9):e75885.

Tian, T. and Burrage, K. (2004). Binomial leap methods for simulating stochastic chemical kinetics. *The Journal of Chemical Physics*, 121(21):10356.

Valencia-Sanchez, M. A., Liu, J., Hannon, G. J., and Parker, R. (2006). Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & Development*, 20(5):515–524.

van Leeuwen, I. M. M., Byrne, H. M., Jensen, O. E., and King, J. R. (2007). Elucidating the interactions between the adhesive and transcriptional functions of -catenin in normal and cancerous cells. *Journal of Theoretical Biology*, 247(1):77–102.

Vargas, D. Y., Raj, A., Marras, S. A. E., Kramer, F. R., and Tyagi, S. (2005). Mechanism of mRNA transport in the nucleus. *PNAS*, 102(47):17008–17013.

Vellela, M. and Qian, H. (2009). Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited. *Journal of The Royal Society Interface*, 6(39):925–940.

Ventura, A. and Jacks, T. (2009). MicroRNAs and Cancer: Short RNAs Go a Long Way. *Cell*, 136(4):586–591.

Vilaseca, E., Isvoran, A., Madurga, S., Pastor, I., Garcés, J. L., and Mas, F. (2011). New insights into diffusion in 3D crowded media by Monte Carlo simulations: effect of size, mobility and spatial distribution of obstacles. *Physical Chemistry Chemical Physics*, 13(16):7396.

Wachsmuth, M., Waldeck, W., and Langowski, J. (2000). Anomalous diffusion of fluorescent probes inside living cell nuclei investigated by spatially-resolved fluorescence correlation spectroscopy. *Journal of Molecular Biology*, 298(4):677–689.

Wang, T.-Y., Chen, K.-C., Hsu, D. F., and Kao, C.-Y. (2009). Combining Agent-Based Models with Stochastic Differential Equations for Gene Regulatory Networks. In *2009 Ninth IEEE International Conference on Bioinformatics and BioEngineering (BIBE)*, pages 405–409. IEEE.

Wang, X., Li, Y., Xu, X., and Wang, Y.-h. (2010). Toward a system-level understanding of microRNA pathway via mathematical modeling. *BioSystems*, 100(1):31–38.

Wawra, C., Kühl, M., and Kestler, H. A. (2007). Extended analyses of the Wnt/b-catenin pathway: Robustness and oscillatory behaviour. *FEBS Letters*, 581(21):4043–4048.

Zhang, J.-g., Wang, J.-j., Zhao, F., Liu, Q., Jiang, K., and Yang, G.-h. (2010). MicroRNA-21 (miR-21) represses tumor suppressor PTEN and promotes growth and invasion in non-small cell lung cancer (NSCLC). *Clinica chimica acta; international journal of clinical chemistry*, 411(11-12):846–852.

Zhou, H.-X., Rivas, G., and Minton, A. P. (2008). Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences *. *Annual Review of Biophysics*, 37(1):375–397.

Zimmerman, S. B. and Minton, A. P. (1993). Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annual Review of Biophysics and Biomolecular Structure*, 22:27–65.