Intracellular spatial organisation : The effects of non-ideal solutions

Ander Movilla Miangolarra

To cite this version:

Ander Movilla Miangolarra. Intracellular spatial organisation : The effects of non-ideal solutions. Physics [physics]. Université Paris sciences et lettres, 2021. English. NNT : 2021UPSLS076 . tel-03521631

HAL Id: tel-03521631
https://pastel.hal.science/tel-03521631

Submitted on 11 Jan 2022

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Institut Curie

Intracellular spatial organisation: The effects of non-ideal solutions

Organisation spatiale intracellulaire: Les effets des solutions non idéales

Soutenue par
Ander Movilla Miangolarra
Le 10 Décembre 2021

École doctorale n°564
Physique en Île-de-France

Spécialité
Physique

Composition du jury :

David **Lacoste** Dir. de Recherche CNRS, ESPCI Paris	*Président*
John **Palmeri** Dir. de Recherche CNRS, Université de Montpellier	*Rapporteur*
Ulrich **Gerland** Prof., Technical University of Munich	*Rapporteur*
Anđela **Šarić** Dr., University College London	*Examinatrice*
Michele **Castellana** Dr., Institut Curie	*Directeur de thèse*
Jean-François **Joanny** Prof., Collège de France	*Directeur de thèse*

institut Curie | PSL

Abstract

In the cytoplasm of a biological cell there are a myriad of different proteins, lipids and enzymes, each of them performing different tasks. Therefore, the spatial organisation of these chemical species is crucial for the functioning of a cell. Order can emerge in many ways but, in living beings, it often appears as a consequence of free-energy expenditure. Hence, in this thesis, we study how order and organisation can emerge from the interplay between physical interactions and free-energy consumption, both analysing particular examples that arise in cell biology and constructing a more abstract framework for reaction-diffusion systems in the presence of interactions.

In the first part of this thesis, we explore, from a theoretical perspective, two processes where the intracellular medium is patterned and organised, whose common feature is the fact that they both stem from the collective behaviour of a large number of molecules. First, we develop a model for protein aggregation which studies the effect of intracellular obstacles on the coagulation kinetics. Our predictions are then successfully compared with experimental data.

The other example refers to the compaction and location of the bacterial chromosome. We suggest that the chromosome segregates from the rest of the cytoplasm, because of steric interactions between DNA and the intracellular crowders, by means of a mechanism reminiscent of liquid-liquid phase separation. Also, our study indicates that spatial localisation within the cell is dictated by non-equilibrium transcription of mRNAs (which are part of the crowding effect). Our model successfully reproduces much of the phenomenology observed in *Escherichia coli* cells, for example, the positioning of the chromosome at different stages of the cell cycle.

Building on these examples, in the second part of the thesis, we construct a thermodynamically consistent framework to mathematically describe chemical reaction networks in non-ideal solutions. This framework allows us to generalise the results from the classical theory of ideal networks, and aids in elucidating the connection between non-equilibrium chemical reactions and phase separation for a large class of networks, known as complex-balanced networks. Given that complex balance can be fully determined from the topology of the network, we analyse how this topological property of the network can constrain the dynamics of the solution, and what behaviour we can expect when complex balance is broken.

We conclude by discussing how the two approaches used here are related and the contribution each of them can have for the advancement of biophysics.

Résumé

La cellule est l'élément de base composant tous les êtres vivants sur Terre. À l'intérieur de chaque cellule, dans le cytoplasme, il y a une myriade de constituants tels que des protéines, des lipides et des enzymes. Chacune de ces molécules joue un rôle bien spécifique au sein de la cellule, ce qui requiert un contrôle précis de leur position spatiale. L'organisation intracellulaire est donc essentielle.

Pour parvenir à cette organisation spatiale détaillée, les êtres vivants ont recours à stratégies diverses. D'un côté, la mise à profit des interactions entre les constituants de la cellule peut créer de l'ordre sans nécessiter de dépense énergétique, mais cela ne peut créer qu'une organisation statique. D'un autre côté, il existe d'autres stratégies qui sont plus dynamiques mais qui nécessitent une dépense d'énergie (techniquement, une dépense d'énergie libre) qui peut, parfois, être très importante.

L'une des caractéristiques des organismes vivants est la dépense continue d'énergie. Par conséquence, pour comprendre comment l'auto-organisation des organismes vivants émerge, il est nécessaire de comprendre comment cette dépense énergétique modifie les lois physiques que nous connaissons pour les systèmes à l'équilibre thermodynamique.

Il y a donc deux grands axes de recherche dans cette thèse, chacun d'entre eux étant représenté dans une des parties de cette thèse. Dans la première partie, nous proposons une explication théorique pour deux processus où le milieu intracellulaire est organisé spatialement sous forme de motifs. Ces processus découlent tous deux du comportement collectif des molécules. Dans la deuxième partie, nous construisons un cadre thermodynamique pour décrire les réseaux de réactions chimiques dans des solutions non idéales, c'est-à-dire des solutions avec des interactions entre solutés. Cette approche nous permet d'obtenir des résultats très généraux pour les modèles de solutions non idéales avec des réactions chimiques hors équilibre thermodynamique. Cette classe de modèles peut décrire une grande variété de phénomènes dans les cellules biologiques.

Première partie

Cette partie est consacrée à l'analyse théorique des deux processus au cours desquels le milieu intracellulaire s'organise et des motifs spatiaux peuvent apparaître. Dans un premier temps, nous développons un modèle qui a pour but d'étudier l'impact de l'encombrement stérique intracellulaire sur la cinétique d'agrégation des protéines (Chapitre 1). Dans un deuxième temps (Chapitre 2), nous étudions la compaction et la localisation spatiale du chromosome de la bactérie *Escherichia coli* (*E. coli*).

Encombrement stérique et agrégation de protéines

Dans l'environnement intracellulaire, le cytosquelette et d'autres grands obstacles entravent considérablement la diffusion des particules. Cependant, la manière dont ces obstacles affectent les particules dépend de la taille de chacune d'entre elles. Certains obstacles, comme le cytosquelette, n'affectent que les particules de plus de 35-50 nm [Luby-Phelps et al., 1987, Etoc et al., 2018], un seuil parfois appelé "taille des pores du cytoplasme". Pour ce type de particule (taille d'environ 50nm ou plus), le cytosquelette ou d'autres obstacles de grande taille ralentissent fortement la diffusion des particules et peuvent affecter de nombreux processus ; c'est le cas de l'agrégation des protéines.

Nous étudions, d'un point de vue théorique mais complété par des expériences, comment ce type d'encombrement stérique modifie la cinétique d'agrégation des protéines à l'intérieur de la cellule. Nous développons un modèle mathématique pour rendre compte de ces deux coefficients de diffusion très différents et, par comparaison avec les données expérimentales, nous trouvons comment l'encombrement stérique conditionne le nombre et la taille des agrégats des protéines. Enfin, nous calculons approximativement la taille critique des agrégats au-delà de laquelle ils sont piégés dans le cytoplasme cellulaire, et nous constatons qu'elle est proche de la taille du pore du cytoplasme, ce qui est en accord avec notre modèle.

Le nucléoïde chez *E. coli*

L'organisation intracellulaire des bactéries est un cas d'étude très intéressant, car aucune membrane ne sépare les différentes parties de la cellule. Cependant, les bactéries présentent un très haut degré d'organisation intracellulaire. Le cas du chromosome de la bactérie *E. coli*, qui est situé dans une région au centre de la cellule que nous appelons le nucléoïde, est particulièrement spectaculaire. En outre, le nucléoïde contrôle et organise spatialement de nombreux processus cellulaires. Par exemple, il découple spatialement la transcription et la traduction en protéines car les ribosomes, lorsqu'ils sont fixés à l'ARN messager (ARNm), sont exclus du nucléoïde à cause des interactions stériques [Sanamrad et al., 2014]. Finalement, la position du nucléoïde change beaucoup au cours du cycle cellulaire : Quand il n'y a qu'un seul chromosome, il est situé très précisément à la moitié de la cellule, mais juste avant la division cellulaire, les deux chromosomes se positionnent à un quart et trois quarts de la cellule (le long de l'axe principal de la cellule) [Wu et al., 2019].

Pour tenter d'expliquer ces phénomènes, nous avons développé un modèle qui prend en compte les interactions stériques entre le chromosome et les «crowders» de grande taille comme les ribosomes ou les polysomes (ARNm avec un ou plusieurs ribosomes attachés). Même s'il y a d'autres effets affectant potentiellement la taille et l'emplacement du nucléoïde, notre hypothèse de base est que, pour un grand nombre des cas expérimentaux, l'effet dominant est l'encombrement stérique. Notre modèle montre que l'encombrement stérique (couplé à des effets intrinsèquement hors équilibre thermodynamique, comme la transcription et la traduction en protéines) peut expliquer l'apparition du nucléoïde (par un mécanisme similaire à celui de la démixtion de deux liquides). Le modèle peut également expliquer pourquoi la taille de nucléoïde augmente lorsque la transcription d'ARNm est arrêtée et diminue lorsque la traduction en protéines est arrêtée. D'autre part, selon notre modèle, le positionnement central du chromosome, et à 1/4 et 3/4 pour cellules plus grandes, est une conséquence directe de la transcription des ARNm: les ARNm participent à l'encombrement stérique et donc altèrent la pression osmotique dans les différentes zones de la cellule, créant des forces qui pussent le nucléoïde au centre (ou à 1/4 et 3/4, en fonction de la taille du cellule).

Enfin, comme tous les autres modèles, le nôtre présente également des limites. L'une de ces limites est l'hypothèse d'homogénéité dans la direction radiale de la cellule, ce qui implique que le système soit effectivement unidimensionnel et facilite les calculs. Néanmoins, dans certains cas, l'homogénéité radiale est rompue (notamment si le nucléoïde se contracte radialement), ce qui peut avoir des conséquences importantes sur le positionnement du nucléoïde. Par conséquent, nous prédisons, comme conséquence de notre modèle, que si une perturbation contracte le nucléoïde radialement, elle déstabilise également son positionnement.

Deuxième partie

Dans l'analyse précédente, nous avons considéré quelques exemples de phénomènes au sein des cellules qui ont pour point commun de provenir des aspects non idéaux des solutions, notamment l'encombrement stérique et la démixtion des liquides. D'autre part, nous avons vu que des phénomènes hors équilibre thermodynamique (particulièrement des réactions chimiques) contribuent à l'organisation cellulaire. Donc, dans la deuxième partie de cette thèse, nous construisons un cadre thermodynamique général pour décrire les réseaux de réactions chimiques au sein des solutions non idéales.

Cette approche, à la fois générale et thermodynamiquement cohérente, nous permet de généraliser certains résultats de la théorie des réseaux de réactions chimiques idéales tels que les fonctions de Lyapunov (fonctions qui sont minimisées par la dynamique du système) pour les réseaux de type «complex balanced» [Horn and Jackson, 1972]. Ce dernier type de réseau offre un cadre moins restrictif que les réseaux de type «detailed balanced» (qui sont des systèmes à l'équilibre thermodynamique) et, selon la théorie classique de réseaux de réactions chimiques, les réseaux de type «complex balanced» peuvent être caractérisés par leur topologie [Feinberg, 1972]. D'autre part, ce cadre nous permet de formaliser des liens entre les réactions chimiques hors équilibre et les séparations de phase (ou démixtions) pour des réseaux de ce type, pour lesquelles nous pouvons obtenir des diagrammes de phase complets en minimisant leur fonction de Lyapunov. Il est important de noter que cette classe de réseaux peut être identifiée uniquement à partir de la topologie des réseaux, et que l'existence d'une fonction de Lyapunov est indépendante de la cinétique particulière ou de l'énergie libre de chaque système.

Finalement, nous montrons que cette topologie peut contraindre la dynamique des solutions et leur stabilité. En fait, nous montrons que, pour les réseaux de type «complex balanced» avec des interactions locales, l'apparition de «non-equilibrium patterning» (motifs dus à des réactions chimiques hors équilibre) est interdite. À l'inverse, notre théorie prédit que lorsque le réseau n'est plus de type «complex balanced», des dynamiques plus exotiques peuvent apparaître, comme celles de type «Ostwald ripening arrest».

Conclusion

En résumé, cette thèse a présenté deux approches différentes de la même problématique, à savoir l'organisation intracellulaire en présence d'effets hors équilibre thermodynamique. Dans la première partie, nous avons étudié quelques exemples d'un point de vue plutôt effectif, ce qui nous a permis d'analyser de manière simple quelques processus cellulaires et de faire des prédictions qui peuvent être validées expérimentalement. Dans la deuxième partie, en revanche, nous avons utilisé une approche plus générale et abstraite, pour étudier le comportement d'une très grande classe de modèles qui pourraient être utiles pour la modélisation de l'auto-organisation cellulaire. Cependant, cette abstraction a rendu plus difficile la validation de la théorie et nous n'avons pas réussi à décrire certains phénomènes très complexes comme celui du «non-equilibrium patterning». Néanmoins, pour faire progresser la biophysique théorique, une intégration de ces deux approches sera nécessaire, ainsi qu'une synergie réussie entre théorie et expériences.

Acknowledgements

First and foremost, I would like to thank Michele for giving me this opportunity, for your time, and your patience. I would also like to thank Jean-François and Ned for the discussions we have had, always insightful. Also, collaborating with Mathieu and Aléria was one of the great pleasures of the early days of my PhD, for which I am grateful and from whom I learnt a lot. Finally, I would like to thank the *rapporteurs*, John Palmeri and Ulrich Gerland, who have agreed to evaluate my work, as well as the rest of the members of the jury, David Lacoste and Anđela Šarić. *Un grand merci aussi à Romain et Thomas, car si le texte français de cette thèse n'est pas trop mauvais c'est surtout grâce à vous.*

On a more personal note, I would like to thank friends and colleagues from the unit (in fact, the UMR168 as a whole), for a great atmosphere for research but also for your warm welcome and making me feel as one more of the group from the very first day. It has been a pleasure to do research here and it has mostly been so because of you. A special thanks to the people in the theory group (past and present) and the people that have passed through the office, for the always stimulating discussion and company, about science and beyond science.

Y, finalmente, un inmenso gracias a mi familia, por todo su apoyo durante estos años de duro trabajo.

Contents

Preface **1**

I Intracellular organisation from molecular collective processes **5**

1 Protein aggregation in cells **9**

1.1 Physical origin of protein aggregation 10

1.2 Experimental setup 12

1.3 Modelling the dynamical aggregation process 13

1.3.1 Fast-aggregation timescale 14

1.3.2 Slow-aggregation timescale 17

1.4 Comparison with experiments and conclusions 18

2 Intracellular bacterial organisation: the nucleoid **21**

2.1 Characteristics of the nucleoid: Compaction and localisation 22

2.2 Modelling the nucleoid 23

2.2.1 Equilibrium free energy of an interacting inhomogeneous gas 25

2.2.2 DNA free energy 30

2.2.3 Particle currents 31

2.2.4 Model parameters 32

2.2.5 Numerically solving Eqs. (2.1) to (2.3) 33

2.3 Nucleoid formation and size 35

2.3.1 Segregation of DNA and crowders 35

2.3.2 Nucleoid size in the presence of non-equilibrium processes 37

2.3.3 Analytical estimates of the nucleoid size 40

2.4 Consequences of transcription and translation 42

2.4.1 Nucleoid splitting 43

2.4.2 Nucleoid centring 44

2.4.3 Perturbing the nucleoid 48

2.5 Discussion of the results 50

II A theoretical framework for non-ideal chemical reaction networks 57

3 Introduction to the description and thermodynamics of Chemical Reaction Networks 61
3.1 Description of Chemical Reaction Networks . . . 62
3.1.1 Stochastic Chemical Master Equation . . . 63
3.1.2 Deterministic description of a Chemical Reaction Network . . . 64
3.2 Thermodynamical constraints on the dynamics of Chemical Reaction Networks . 64
3.2.1 Stochastic dynamics . . . 64
3.2.2 Deterministic dynamics . . . 66
3.2.3 On the rate constant . . . 67
3.2.4 Rates for non-equilibrium systems . . . 68
3.3 Thermodynamics of Chemical Reaction Networks . . . 69
3.3.1 Thermodynamics of closed Chemical Reaction Networks . . . 70
3.3.2 Non-equilibrium thermodynamics of Chemical Reaction Networks . . . 71

4 The role of complex balance in Reaction-Diffusion networks 73
4.1 Definition of complex and complex-balanced network . . . 73
4.2 Review of results for complex-balanced systems . . . 76
4.3 Complex-balanced steady-state distributions . . . 78
4.3.1 Proof . . . 79
4.3.2 Thermodynamic interpretation of $\tilde{\mu}$ . . . 82
4.4 Lyapunov function for complex-balanced steady states . . . 83
4.5 Lyapunov functionals for spatially heterogeneous solutions . . . 85
4.6 Phenomenology of a complex-balanced mixture . . . 87
4.6.1 Phase diagram of a non-ideal complex-balanced solution . . . 89
4.6.2 General features of the steady state . . . 89

5 Beyond complex balancing? 93
5.1 Patterning when complex balance is broken . . . 93
5.2 Perspectives . . . 96

Conclusion 99

Bibliography 103

Appendices 112

A Supplementary Information for the aggregation model 113

A.1 Experimental Procedures 113
A.1.1 Cell culture 113
A.1.2 Transitory cells transfection by Cry2Olig-mCherry 113
A.1.3 Quantitative estimation of fluorescent protein concentration 113
A.1.4 Optogenetic experiments 114
A.2 Image Analysis 114
A.3 Parameter Fitting 115

B Supplementary Information and Figures for the nucleoid model 117
B.1 Equilibrium free energy of an interacting system: the virial expansion 117
B.2 Auxiliary entropy 119
B.3 Effect of third-order virial terms 120
B.4 Estimate of the diffusion coefficient of DNA 120
B.5 Experimental Methods 122
B.6 Numerical methods 123
B.7 Parameter estimation from experimental data 123
B.8 Nucleoid size at thermodynamic equilibrium 124
B.9 Single chromosome growth condition 124
B.9.1 Scaling of the concentration of chemical species for single-chromosome growth 124
B.9.2 Results for constant RNA concentration 125

C Supplementary Notes on non-ideal Chemical Reaction Networks 127
C.1 Form of the propensity functions for a regular solution theory 127
C.2 Form of the reaction rates in the non-ideal solution theory 128
C.3 Minimisation of the Lagrangian to obtain the phase diagram 129

Bibliography for the Appendices 130

List of Figures

1.1 Experimental system studied and theoretical model 12
1.2 Experimental data and fits to the theoretical model 18

2.1 Sketch of the model for the nucleoid 25
2.2 Steric interactions of DNA segments 28
2.3 Phase diagram for the simplified free energy (2.40) 36
2.4 Steady-state concentration profiles for *E. coli* growing filamentously 38
2.5 Steady-state concentration profiles, for single-chromosome filamentous growth . 39
2.6 mRNA fluxes at steady-state 43
2.7 Transcription splits the nucleoid into two lobes located at 1/4 and 3/4 of the cell length 43
2.8 Non-equilibrium processes centre the nucleoid at midcell 46
2.9 The nucleoid expands in the absence of mRNA synthesis 49
2.10 The nucleoid contracts in the absence of translation initiation 50
2.11 Steady-state profiles for filamentously growing cells 52

3.1 Reaction free energy landscape for a chemical reaction in two different phases . . 68

4.1 Graphical representation of CRNs discussed in Section *A hierarchy of steady states* . 74
4.2 Graphical representation of CRNs discussed in Example 3 77
4.3 Graphical representation of the CRN used in Examples 4 and 5 81
4.4 Steady-state probability distributions for Example 3 82
4.5 The Lyapunov function of a complex-balanced CRN is minimised by its dynamics 88
4.6 Phase diagrams for a chemically reactive non-ideal mixture 90

5.1 LSA for complex-balanced and non-complex-balanced networks 96

B.1 DNA and auxiliary free energies 120
B.2 Effect of third-order virial term 121
B.3 Experimental data and interpolation 124
B.4 Equilibrium concentration profiles for *E. coli* growing filamentously 125

B.5 Nucleoid size for single-chromosome filamentous cells with constant mRNA concentration 126

PREFACE

Science is an essentially anarchic enterprise: theoretical anarchism is more humanitarian and more likely to encourage progress than its law-and-order alternatives.
[...]
And is it not clear that successful participation in a process of this kind is possible only for a ruthless opportunist who is not tied to any particular philosophy and who adopts whatever procedure seems to fit the occasion?

P. K. Feyerabend,
Against Method, 1975

According to the second law of thermodynamics, the entropy of the universe always increases or remains constant [Pippard, 1964]. Informally, entropy is often associated with disorder; fact that, together with the second law of thermodynamics, makes surprising the existence of living beings – creatures known for an almost incredible degree of order and self-organisation, among other characteristics.

This apparent paradox is resolved by noting that, while the entropy of the entire universe must always increase, it may decrease locally. In living organisms, this local decrease in entropy is typically achieved by the consumption of energy (more precisely, free-energy[1]) and the increase of entropy in the surrounding environment, mostly in the form of heat and mass flows from the organism towards the environment. In this process, the main quantity of interest is the free-energy, provided it can be meaningfully defined, because it expresses the maximum amount of energy available to do work [Kondepundi and Prigogine, 2015]. Hence, in an alternative perspective, this free-energy expenditure is used to produce work that can counteract locally the entropic forces that tend towards disorder. Combining theses two ideas, we are led to another seemingly contradictory thought: Free-energy consumption contributes to the increase of entropy but it can also be used to produce work that acts against the entropic forces. Therefore, the crucial concept is that living beings funnel their entropy increase towards their environment, while keeping an ordered internal state[2].

In living beings, the energy expenditure required to maintain homeostasis (the state of steady internal conditions in living beings), can be related to *metabolism*. However, an energy-consuming metabolism is not the only characteristic of the living [Nelson, 2004]. The ability to *reproduce* is also widely considered as a hallmark of life. But energy consumption is also central to reproduction [England, 2013] and, more generally, to the key idea of evolution: natural selection (though the explicit connection, in this case, remains unclear).

In the context of a cell (which is the setting that will occupy us for the rest of this work), free energy is most commonly stored in the form of a chemical potential difference between Adenosine Triphosphate (ATP) and Adenosine Diphosphate (ADP), often referred to as the energy currency of the cell [Phillips et al., 2012]. Indeed, ATP is a high-energy but stable molecule – it typically does not hydrolyse spontaneously in an aqueous solution [Westheimer, 1987] – that can be used to drive chemical reactions or as a source of work.

Therefore, it seems clear now that one of the characteristics of life is its ability to maintain itself away from thermodynamic equilibrium. However, our comprehension of non-equilibrium thermodynamics is far from complete. Mostly due to the pioneering work of L. E. Boltzmann and J. W. Gibbs, for over a century we have had a good understanding of the thermodynamics of systems at equilibrium [Gibbs, 1879]. Slowly, progress was made, by L. Onsager and I. Prigogine among others, regarding the behaviour of systems near equilibrium, within the framework of Linear Irreversible Thermodynamics (LIT) [Onsager, 1931, Glansdorff and Prigogine, 1954]. More recently, a new area of research has emerged that focuses on the thermodynamic properties of

---

[1] Energy, according to the first law of thermodynamics, is conserved if heat is taken into account [Pippard, 1964] and, thus, cannot be "consumed".

[2] In the context of chemical reaction networks, a more rigorous statement of these arguments will be presented in Part II of this thesis.

small (molecular) stochastic systems far away from equilibrium [Jarzynski, 1997, Crooks, 1999, Seifert, 2005, Sekimoto, 2010], which has also enabled some progress in the understanding of chemical reaction networks at the macroscopic scale [Rao and Esposito, 2016]. As a result, systems inherently out of equilibrium, such as living beings, are starting to be understood at the microscopic scale; but there are still plenty of open questions at the macroscopic scale, due to, among other reasons, the spectacular complexity of living beings and the emergent behaviour that arises at larger length-scales [Anderson, 1972].

Nevertheless, order – a central concept to this thesis – can also exist at thermodynamic equilibrium, as is the case, at the atomic scale, in certain metallic alloys [Ashcroft and Mermin, 1976]; or, at the mesoscopic scale, in colloidal suspensions with an interplay between short- and long-range interactions [Campbell et al., 2005, Tarzia and Coniglio, 2006]. In these cases, the system is considered to be in contact with a thermal reservoir, which allows for the energy exchange between the system and the environment, and enables the appearance of order by minimising the free energy of the system. Therefore, a balance between the energy of the system and its entropy appears, and order is produced when the system is capable of minimising its energy and lower its entropy while increasing that of the bath, ensuring that the overall entropy of the universe is still increasing.

Thus, there is not a single universal source of order in nature. While living beings seem to prefer the dissipation of free energy to maintain their ordered structures, there are other mechanisms by which order emerges without the need of dissipative structures. Then, several questions relevant to the topic of this thesis emerge: To which extent does organisation and morphogenesis within living beings require the consumption of free energy? Which tasks could be accomplished without dissipation?

Far from answering these questions, we will content ourselves by describing the interplay that can appear when both sources of order act simultaneously. On the one hand, we consider *passive* mechanisms (no free energy is consumed), which, for the remainder of this work, are the interactions between solutes in a solution. On the other hand, we take into account *active* processes that do require free-energy dissipation. The types of active processes that we consider in this thesis are chemical reactions driven by a free-energy difference between fuel and waste components (e.g., in the context of a biological cell, the chemical potential difference between ATP and ADP).

Much of this thesis is hence devoted to the study of the rich phenomenology that can arise from the crosstalk between these two mechanisms. The system that we bear in mind throughout this work is that of a solution, which is not necessarily dilute, where interactions between the solutes may arise; and fuelled chemical reactions take place, driving the solution out of thermodynamic equilibrium. In contrast, much of the theoretical work in chemical reaction networks has considered only ideal solutions, which assumes that solutes are dilute and interactions among them, negligible. However, order can also emerge when these interactions are taken into account, as explained above, and the interplay between passive and active ordering mechanisms can give rise to novel self-organised structures. We believe that this point of view provides a more realistic, yet still oversimplified, description of the cytoplasm of a cell, given that it is often crowded with macromolecules [Roberts et al., 2002, Höfling and Franosch, 2013] and can even segregate some of its components in droplet-like condensates as a consequence of the interactions among solutes [Brangwynne et al., 2009]. This type of solutions, hereafter termed as non-ideal solutions, are the cornerstone of this work.

Finally and on a more personal note, despite this introduction, which may give the impression of a well-planned research agenda undertaken during these 3 years and 3 months, most of this thesis is the result of *ruthless opportunism*. Rather than following a plan, the work behind the majority of the different chapters in this thesis was undertaken in a fairly *anarchic* way, with chance and curiosity playing a major role. While the topics treated here revolve around the same concepts and I have tried to construct a coherent line of thought throughout the different parts of this thesis, traces of this unplanned opportunism can be observed in the text that follows.

PART I:

INTRACELLULAR ORGANISATION FROM MOLECULAR COLLECTIVE PROCESSES

> *Thermodynamic equilibrium may be characterized by the minimum of the Helmholtz free energy defined usually by:* $F = E - TS$.
> *Are most types of "organisations" around us of this nature? It is enough to ask such a question to see that the answer is negative. Obviously in a town, in a living system, we have a quite different type of functional order.*
>
> I. R. Prigogine,
> *Nobel Lecture*, 1977

The advances in microscopy and quantitative measurements in cell biology has allowed to obtain huge amounts of quantitative data from experiments. This data, while intriguing and often spectacular, is not very informative by itself and it only attains its entire potential when placed within a theoretical framework, when confronted with hypotheses and ideas. Traditionally these frameworks where developed by biologists and were fundamentally qualitative or descriptive. However, as more quantitative data is produced, there seems to be a need for a robust mathematical framework that can make the most of such experimental output. In developing an appropriate quantitative theory, the ideas of theoretical physics, while not always directly applicable, can be of great help.

The first part of this thesis is devoted to the mathematical modelling of two processes that control the spatio-temporal organisation in biological cells, which is established at the cellular scale by the collective behaviour of particles at the molecular scale. First, we describe mathematically a case of protein aggregation process whose dynamics are slowed down for large agglomerates due to the presence of intracellular obstacles. Our theoretical predictions are compared with experiments performed with an optogenetic protein. In the second half of this part, we model the bacterial nucleoid (that is, the part of a bacterial cell in which the chromosome is located) and explain its size and positioning. As mentioned in the Preface, our approach has two main ingredients: we consider the effect of crowders and steric interactions and the consequences of non-equilibrium processes in the cell, such as, transcription and translation. In line with the idea pioneered by I. Prigogine [Prigogine, 1977, Kondepundi and Prigogine, 2015], these *non-equilibrium processes are the source of order* in the system and result in what he called *dissipative structures*.

Ideally, these two examples could be modelled starting from a common theoretical framework that applies to all *dissipative structures*. However, in practice, the particular details of each system make it a very challenging task and often it is best to explore particular models for each of the processes separately.

Nevertheless, the modelling of these two systems has a number of common features. One of them is the fact that their time evolution is derived from physical principles, such as, Fick's law of diffusion or excluded-volume effects. Specifically, in both cases, the effect of intracellular obstacles modulates the dynamics and localisation of the different objects in the system. In the cytoplasm of an eukaryotic cell, there seems to be a well defined mesh size due to the cytoskeleton and other intracellular obstacles [Luby-Phelps et al., 1987, Etoc et al., 2018], which in turn, as we will prove, regulates the size and growth of protein aggregates. In the bacterial nucleoid there is also a characteristic mesh size [Xiang et al., 2021], but in this case formed by the folded chromosome. This mesh size controls, among other things, the spatial location of translation within the cell, as messenger RNAs (mRNAs) with one or several translating ribosomes attached are excluded from the nucleoid to a great extent [Sanamrad et al., 2014], probably due to their size being larger than the mesh size of the nucleoid.

As mentioned above, both systems are out of thermodynamic equilibrium, as is usual in living beings, which are characterised by, among other features, their continuous free-energy consumption. However, these two systems are out of thermodynamic equilibrium in different ways. While the protein aggregation system is out of equilibrium because their initial condition is not an equilibrium state, the bacterial nucleoid is out of equilbrium since there is a continuous expenditure of free energy to maintain a non-equilibrium steady-state. Therefore, in our first example we study the relaxation dynamics towards equilibrium of protein aggregates and, in the second one, the characteristics of the non-equilibrium steady state. Finally, both processes control the spatial organisation on the cell but they do so in different ways: The dynamics of protein aggregation determine, as a function of time, the number and size of the aggregates, while in the bacterial cell the rate of transcription can control the non-equilibrium steady-state and, thus, the positioning of the nucleoid and the localisation of the different components of the transcriptional-translational machinery.

Given the differences between the processes studied in this part of the thesis, the mathematical formalism used in each chapter is markedly different from the ones used in the other chapters. While in the second chapter (regarding the bacterial nucleoid) the analysis takes explicitly into account the types of interactions present in the system and it was mostly limited to its steady state, in the first chapter (about protein aggregation) we have resorted to a simplified kinetic description of the system to study its time evolution in an analytically tractable way. These two approaches can be seen as two faces of the processes that living beings use to control their internal self-organisation: Time dependent processes, like growth or coalescence (e.g. [Wang et al., 2008, Al Jord et al., 2021]), and patterned non-equilibrium steady-states ([Murray and Sourjik, 2017] or [Halatek and Frey, 2018]). Taken together, they enable the spatial organisation of biological cells with an impressive degree of complexity.

Chapter 1

Protein aggregation in cells

Protein aggregation is a process which spans multiple orders of magnitude both in time and space[1]: From nucleation, when a couple of monomers of a given chemical species of interest bind together to initiate the process, to the formation of large clusters containing up to millions of monomers each [Zidar et al., 2018]. Protein aggregation is a widely studied topic in molecular and cell biology due to its connections with numerous diseases and disorders, such as, Alzheimer's and Huntington's disease [Selkoe, 2004]. In these diseases, large micron-sized aggregates appear in cells which are often associated with the loss of function of the proteins involved [Ross and Poirier, 2004]. However, most importantly, protein aggregation has also been linked with the creation of toxic intermediate-size aggregates that interact inappropriately with functional cellular components [Knowles et al., 2014]. Furthermore, protein aggregates have been shown to exclude chromatin and disrupt the expression of certain genes, and the aggregate size appears to play an important role in this process [Li et al., 2016].

Thus, protein aggregation can be seen as a self-assembly process where monomers aggregate to form typically fibrillar structures, with important consequences for cell physiology. However, since toxic intermediates can be created in the process, the final assembly of the system (a single large and inert aggregate, if we are considering irreversible aggregation) is not as important as its time evolution towards this final state. Therefore, here we analyse a process where the kinetics are crucial for its understanding and the final steady state is not as informative as the trajectory followed towards it. More precisely, we consider a process where aggregation is always thermodynamically favoured but can be kinetically arrested, due to the presence of intracellular obstacles.

In this chapter, we first introduce the physico-chemical basis of the process of protein aggregation, describing the different kinetics that may arise in this type of systems and motivating our research. Then we proceed to briefly describe the experiments performed and the mathematical modelling of the process. Finally we compare the output of these two approaches and present the consequences of our research.

---

[1] The work presented in this chapter has been published in *Biophysical Journal*, Ref. [Miangolarra et al., 2021a]. The present chapter is therefore largely based on that reference. The experiments described in this chapter were performed by Aléria Duperray-Susini and Mathieu Coppey.

1.1 Physical origin of protein aggregation

Most proteins in the cell have a stable folded conformation which allows them to appear in a soluble form in the cytoplasm. However, certain proteins do not fold into a globular and soluble structure and, among them, a certain subset is prone to aggregation [Knowles et al., 2014].

Typically, a seed or nucleation event (which gathers a number of proteins to create an oligomer or nucleus of the aggregate) is needed to start the process, after which the protein oligomers grow and become large aggregates [Narayanan et al., 2019]. This is due to the shape of the free energy of the oligomer: If the free energy of the oligomer as a function of its size in monomers (N) has a single maximum at $N = N^*$, then N^* is the critical size of the nucleus and determines the smallest growth-competent unit [Šarić et al., 2016]. Therefore, there is a free-energy barrier to form a nucleus of size N^*, after which the growth of the aggregate is thermodynamically favoured. While the nucleation step can be reversed (small oligomers may be created and afterwards dissolved, especially if the concentration of the protein of interest is not high enough), once large aggregates are formed we can normally consider the aggregation process to be irreversible to a great extent. It is in this later stage of the process that we will focus our efforts.

Given the importance of the kinetics in protein aggregation, a plethora of mathematical models have been developed, depending on the particular features of the aggregating system. For example, for fibrillar assemblies, good agreement was found between the experimental data and a mathematical model considering the elongation and fragmentation of fibrils in only one dimension [Knowles et al., 2009]. Here, however, we will not focus on the shape of the aggregate but on how its overall size and diffusivity affects its growth, as discussed below.

Diffusion-limited irreversible aggregation in the cell

We have already established that, due to energetic considerations, once a large enough protein oligomer has been nucleated, it will continue to grow spontaneously and irreversibly, provided the right conditions are met, i.e. the solution is supersaturated [Narayanan et al., 2019]. However, the kinetics of this growth are still very informative and important from the biological point of view as they determine the speed at which these aggregates will grow, their size and their number. Indeed, if the kinetics are slow enough, aggregation may never happen in a biologically relevant timescale, while if they are fast enough, few of the toxic intermediate-sized oligomers will exist (as they will soon become large and potentially inert [Knowles et al., 2014]). As stressed above, it is not only the final state that matters but also the time evolution of the process.

There are two main factors controlling the timescale over which protein aggregates grow. One of them is the binding affinity, that is, provided two particles meet, how energetically favoured is their binding and, thus, how likely they are to bind. The other factor is the diffusion coefficient, which determines the frequency with which particles bump into each other. If the limiting factor for the aggregation process is the binding affinity of the particles, we talk about reaction-limited aggregation, while if the limiting factor is the diffusivity of the particles then the process is termed diffusion-limited aggregation.

In this chapter, we consider, from the theoretical viewpoint, a diffusion-limited irreversible aggregation process. As a result, the speed at which particles diffuse and meet entirely regulates the speed of the aggregation process. This theoretical approach is complemented by an experimental study (carried out by Aléria Duperray-Susini and Mathieu Coppey) which uses an optogenetic protein that upon blue light exposure oligomerises (CRY2olig). Given that, under frequent blue light exposure, the fragmentation of a cluster is strongly suppressed, one may view this process as an irreversible, out-of-equilibrium aggregation, which leads to the formation of increasingly large clusters. In other words, irreversibility stems from a negligible fragmentation or dissolution rate of the aggregates, at least in the timescale of the experiments [Taslimi et al., 2014]. The role of diffusion is also crucial, because it sets the speed at which the aggregation process unfolds: aggregation processes in the cellular cytoplasm are influenced by the presence of physical obstacles which alter the diffusion dynamics of growing aggregates within the cell [Wojcieszyn et al., 1981, Höfling and Franosch, 2013]. In particular, in Ref. [Etoc et al., 2018], it was shown that, for quasi-spherical nanoparticles in HeLa cells, the diffusivity drops by two or three orders of magnitude as the diameter of the nanoparticle is increased from 50 to 75 nm, due to steric interactions between the particle and the cytosolic meshwork of the cell. Particles above this threshold – which is sometimes referred to as the pore size of the cytoplasm – experience almost no diffusion [Luby-Phelps et al., 1987].

Thus, in view of these facts, we developed a theoretical and analytically tractable model with as few as two free parameters to mathematically describe an aggregation process where agglomerates of different sizes have very different diffusivities. In collaboration with experiments, the model allows to validate the picture of two different timescales controlling large aggregate number and size. Overall, the successful integration and feedback between theory and experiments led to quantitative predictions and their verification, like the critical size above which an aggregate can be thought of as "trapped" in the cytoplasm.

For irreversible aggregation processes, the only possible steady-state is the one where all proteins form a single cluster. However, this is rarely the case in biological cells, as the cytoplasm typically exhibits multiple protein or enzymes clusters scattered all over its volume [An et al., 2008, Narayanan et al., 2019]. Therefore, the physical mechanisms which set the cluster number and size still remain a subject of investigation [Buchner et al., 2013, Castellana et al., 2014], and could have important biological consequences.

In the rest of the chapter, we address the problem of diffusion-limited irreversible aggregation processes in cells, both theoretically and experimentally. We found that two timescales control aggregation processes: One related to the fast diffusion of small clusters, and the other one, slower, potentially related to the hindered diffusion due to the presence of intracellular obstacles. By matching theoretical predictions with experimental results, we estimated the threshold between these two timescales. However, on long timescales a diffusive-like movement can be observed even for large aggregates, which was related to fluctuations that stem from the incoherent effect of a network of active forces in the cell [Guo et al., 2014], such as the rearrangement of the cytoskeleton and endomembranes. Therefore, large aggregates, with a radius comparable or larger than this threshold, can be thought of as being strongly confined, and subject to a dramatic hindrance in the diffusivity which may have an important effect on the aggregation dynamics. Consequently, the slow timescale above is likely to be determined by active fluctuations that affect

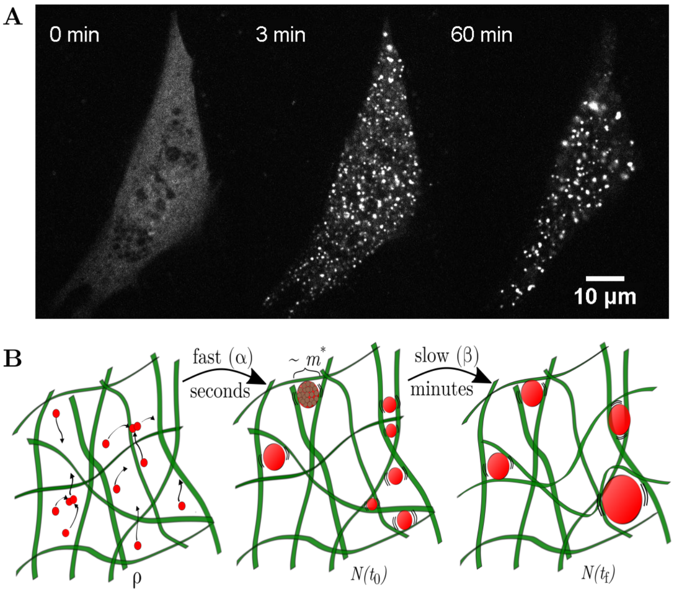


Figure 1.1: Experimental system studied and theoretical model. (A) Fluorescent images of a cell expressing CRY2olig and activated with blue light every 2 min. The cell shows many small clusters by $t = 3$ min, which then mature and coalesce over time. (B) Cartoon depicting the main ingredients of the model. Left: a density ρ of monomers that can freely diffuse and aggregate fast. Middle: $N(t_0)$ larger clusters of size $\sim m^*$ are trapped in the cytoskeleton. Right: $N(t_\mathrm{f})$ large clusters are only able to move and further aggregate when the remodelling of the cytoskeleton sets them free.

the dynamics of intracellular objects larger than the typical pore size of the cytoplasm. Overall, our results shed light into the interplay between aggregation processes, and the dynamics of the crowded environment in the cell cytoplasm.

1.2 Experimental setup

The experimental system under study is an optogenetic protein CRY2olig, which oligomerises upon blue light [Park et al., 2017], fused to the fluorescent tag mCherry, that is transfected into RPE1 cells (retina pigmented epithelium 1, mammalian cells). An important feature of this optogenetic protein is the persistence of its oligomerised state even in the dark, with a half life of around 23 mins [Taslimi et al., 2014].

Twenty-four hours after transfection, cells are exposed to blue light: this blue-light exposure can be regarded as an out-of-equilibrium process, which triggers protein oligomerisation. The period at which cells are exposed to blue light is 120s, which is significantly lower than the half life of the oligomerised state in the absence of light stimulation (23 mins), allowing us to consider the aggregation process as irreversible. The dynamics of these protein clusters are then followed for one hour as shown in Figure 1.1 A: From the images we extracted initial protein concentrations, final concentration of clusters and its size. For details regarding the experimental setup and image analysis, please see Appendix A.

The number of monomers cannot be determined directly from these images as there is a constant relating arbitrary intensity units and the actual concentration in each pixel. An estimate of this constant was obtained by imaging droplets with known concentration of mCherry (the fluorescent tag used in the experiments) and comparing both images.

Qualitatively, we were able to distinguish two different dynamical regimes. The first one is a regime characterised by rapid diffusion and aggregation of small oligomers, see Fig. 1.1 A, which takes place right after the blue light is switched on, and lasts for a time lapse of the order of minutes which is short compared to the imaging time of 1 hr. The second regime is characterised by larger clusters which exhibit slower diffusion or almost no diffusion, resulting in a slower aggregation process. Presumably, the lack of diffusion is produced by the obstacles in the cytoplasm, which trap large aggregates. Hence, any diffusion or movement of these large clusters could be widely attributed to movements and remodelling of the cytoplasmic structure [Guo et al., 2014, Etoc et al., 2018]. These features are summarised in Figure 1.1 B. A similar drop in the diffusivity between large protein aggregates and monomers was also observed in Ref. [Li et al., 2016].

The analysis of these images allowed us to obtain the cluster-size distribution, the cluster concentration and the mean cluster size as functions of the initial protein concentration, which are shown in Fig. 1.2.

1.3 Modelling the dynamical aggregation process

The theoretical basis of irreversible aggregation processes was introduced by von Smoluchowski over a century ago, and it can be summarised into his well-known equation for the coagulation kinetics [Krapivsky et al., 2010]:

$$\frac{dc_i(t)}{dt} = \frac{1}{2} \sum_{j+k=i} k_{j,k} c_j(t) c_k(t) - c_i(t) \sum_k k_{i,k} c_k(t), \tag{1.1}$$

where $c_i(t)$ refers to the concentration of clusters with i monomers, and $k_{i,j}$ is the aggregation rate constant (also known as aggregation kernel) between two clusters of mass i and j. The rate of aggregation $k_{j,k}c_j(t)c_k(t)$ is given by the law of mass action. Upon an appropriate choice of the aggregation kernel $k_{i,j}$, Eq. (1.1) can adequately describe diffusion-limited aggregation processes.

However, the kernel typically does not take account of the effect of obstacles or pores, such as the ones found in the cytoplasm of a cell [Luby-Phelps et al., 1987, Etoc et al., 2018]. To take account of this effect, we leverage the insights from the experiments (previously described in Section 1.2) to build a kernel $k_{i,j}$ based on the separation of the two timescales involved in the aggregation process: On the one hand, there is a fast aggregation timescale (characterised by a rate α), involving monomers and small clusters that diffuse rapidly, and which ultimately leads to the formation of larger agglomerates. On the other hand, there is a slow aggregation timescale (with characteristic rate β) that comprises aggregates larger than the pore size of the cytoplasm. The threshold between these two timescales is the time, t_0, beyond which all clusters are larger than the pore size of the cytoplasm. We denote by m^* the cluster mass at which the agglomerate attains the size of the pore of the cytoplasm and barely diffuses, see Fig. 1.1.

The objective of our model is to examine the effect of a sharp drop in diffusivity with particle size and, therefore, we neglect other hydrodynamic effects, such as size-dependent diffusivity. These assumptions are supported by the conclusions of Ref. [Etoc et al., 2018] where it was found that most of the drop in diffusivity with particle size takes place in a narrow window of size and other variations in diffusivity are small in comparison to this drop. In addition, this simple timescale-separation assumption allows to keep the complexity of the model low, while capturing the essence of the dynamics.

1.3.1 Fast-aggregation timescale

For the fast-aggregation timescale, we choose the following kernel for the Smoluchowski coagulation equation:

$$k_{j,k} = \alpha \left[\theta(m^* - j) + \theta(m^* - k) \right], \tag{1.2}$$

where α is a fast aggregation rate constant, and is assumed to be much larger than the rate constant associated with the slow aggregation timescale β, and $\theta(x)$ is the Heaviside step function. Since, within this timescale, clusters with mass larger than m^* do not diffuse, they do not contribute to the aggregation kernel (1.2). In addition, we will assume that, by the time all clusters are of mass m^* or larger, the contribution of the slow process to the clustering dynamics is negligible, therefore effectively decoupling the timescales involved in the problem.

For the kernel of the type of Eq. (1.2), the following change of variables is known to simplify the Smoluchowski equation [Leyvraz, 2003]:

$$\varphi_i(t) = c_i(t)/N(t), \qquad d\tau = N(t)dt \tag{1.3}$$

where $N(t) = \sum_i c_i(t)$ and $\varphi_i(t)$ represents for the fraction of clusters that is of size i, which yields the following form for the Smoluchowski equations:

$$\frac{d\varphi_i(\tau)}{d\tau} = \alpha \left[\sum_{j+k=i} \varphi_j(\tau)\varphi_k(\tau)\theta(m^* - j) - \varphi_i(\tau)\theta(m^* - i) \right]. \tag{1.4}$$

An important feature of Eq. (1.4) is its recursive structure, i.e., the equation for φ_i only depends on φ_j for $j < i$. One can prove inductively that the solution for the mobile clusters (φ_i for $i \leq m^*$) is given by:

$$\varphi_i(\tau) = \sum_{k=1}^{i} (-1)^{k-1} e^{-k\alpha\tau} \binom{i-1}{k-1}, \tag{1.5}$$

where we have assumed that only monomers are present at $t = 0$, which mathematically it means that $\varphi_i(0) = \delta_{i,1}$ (also known as monodisperse initial condition).

Proof of the solution (1.5)

Given that Eq. (1.4) is a recursive equation for φ_i, in what follows we will attempt an inductive proof of the solution for any φ_i with $i \leq m^*$, for which the Heaviside step function is equal to one. We will show that, for $i \leq m^*$, if the ansatz (1.5) holds for $\varphi_1, \cdots, \varphi_{i-1}$, then it holds for φ_i as well. To achieve this, we insert the ansatz (1.5) in Eq. (1.4), where we evaluate

$$\begin{aligned} \sum_{j=1}^{i-1} \varphi_j(\tau)\varphi_{i-j}(\tau) &= \\ \sum_{j=1}^{i-1} \sum_{n_1=1}^{j} \sum_{n_2=1}^{i-j} (-1)^{n_1+n_2-2} e^{-(n_1+n_2)\tau\alpha} \binom{j-1}{n_1-1} \binom{i-j-1}{n_2-1} &= \\ \sum_{j=1}^{i-1} \sum_{s=2}^{i} \sum_{n_2=1}^{s-1} (-1)^{s-2} e^{-s\tau\alpha} \binom{j-1}{s-n_2-1} \binom{i-j-1}{n_2-1}&, \end{aligned} \tag{1.6}$$

where, in the last equality, we have made the change of variable $s = n_1 + n_2$. Now we can apply Vandermonde's identity:

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k}, \tag{1.7}$$

which yields

$$\begin{aligned} \sum_{j=1}^{i-1} \varphi_j(\tau)\varphi_{i-j}(\tau) =& \sum_{j=1}^{i-1}\sum_{s=2}^{i} (-1)^{s-2} e^{-s\tau\alpha} \binom{i-2}{s-2} \\ =& (i-1)\sum_{s=2}^{i} (-1)^{s-2} e^{-s\tau\alpha} \binom{i-2}{s-2}. \end{aligned} \tag{1.8}$$

Equation (1.4) for φ_i, $i \le m^*$ now reads, assuming the ansatz (1.5) for φ_j, $j < i$,

$$\frac{d\varphi_i(\tau)}{d\tau} = \alpha(i-1)\sum_{s=2}^{i} (-1)^{s-2} e^{-\alpha s\tau} \binom{i-2}{s-2} - \alpha\varphi_i(\tau) \tag{1.9}$$

which can be rewritten as follows

$$\frac{d(\varphi_i(\tau)e^{\alpha\tau})}{d\tau} = \alpha(i-1)\sum_{s=2}^{i} (-1)^{s-2} e^{-\alpha(s-1)\tau} \binom{i-2}{s-2} \tag{1.10}$$

and solved by direct integration along with the monodisperse initial conditions (which make the constant from the integration vanish), yielding

$$\varphi_i(\tau) = \sum_{s=1}^{i} (-1)^{s-1} e^{-s\alpha\tau} \binom{i-1}{s-1}, \tag{1.11}$$

thus proving that if the ansatz (1.5) holds for $\varphi_1, \cdots, \varphi_{i-1}$, then it holds for φ_i as well. It can easily be checked that the ansatz (1.5) is a solution of Eq. (1.4) for $i = 1, 2$ or 3 and, thus, it will hold for all integer i such that $0 < i \le m^*$.

Alternatively, Eq. (1.11) can be recast into the form:

$$\varphi_i(\tau) = \frac{(1-e^{-\alpha\tau})^i}{e^{\alpha\tau}-1} \tag{1.12}$$

by the binomial theorem.

Evolution of the number of clusters

The number of clusters as a function of our rescaled time τ, during the fast aggregation timescale, can be obtained first by summing Eq. (1.1) over all $i > 0$ with the kernel (1.2):

$$\begin{aligned} &\sum_i \frac{dc_i(t)}{dt} = \\ &\alpha\frac{1}{2}\sum_{i,j+k=i} c_j(t)c_k(t)\left[\theta(m^*-j)+\theta(m^*-k)\right] - \alpha\sum_{i,k} c_i(t)c_k(t)\left[\theta(m^*-i)+\theta(m^*-k)\right] = \\ &\alpha\sum_{i,j+k=i} c_j(t)c_k(t)\theta(m^*-j) - 2\alpha\sum_i c_i(t)\sum_k^{m^*} c_k(t) = \\ &-\alpha N(t)\sum_{j=1}^{m^*} c_j(t). \end{aligned} \tag{1.13}$$

and then by using the change of variables (1.3), which yields the equation for the cluster concentration in the rescaled time τ

$$\frac{dN(\tau)}{d\tau} = -\alpha N(\tau) \sum_{i=1}^{m^*} \varphi_i(\tau). \tag{1.14}$$

Using the so-called *hockey-stick identity*:

$$\binom{m^*}{n} = \sum_{k=0}^{m^*-n} \binom{k+n-1}{n-1}, \tag{1.15}$$

after inserting the solution (1.5) into Eq. (1.14) we obtain

$$\begin{aligned} \frac{dN(\tau)}{d\tau} &= -\,\alpha N(\tau) \sum_{i=1}^{m^*} \sum_{j=1}^{i} (-1)^{j-1} e^{-j\alpha\tau} \binom{i-1}{j-1} \\ &= -\,\alpha N(\tau) \sum_{j=1}^{m^*} \sum_{\hat{\imath}=0}^{m^*-j} (-1)^{j-1} e^{-j\alpha\tau} \binom{\hat{\imath}+j-1}{j-1} \\ &= -\,\alpha N(\tau) \sum_{j=1}^{m^*} (-1)^{j-1} e^{-j\alpha\tau} \binom{m^*}{j}, \end{aligned} \tag{1.16}$$

where, in the second equality, the following change of variable has been performed: $i = \hat{\imath} + j$ to alter the order of summation and, in the third equality, the hockey-stick identity has been used. The solution to Eq. (1.16) is

$$N(\tau) = N(0) \exp\left\{ \left[\sum_{k=1}^{m^*} \frac{(-1)^{k-1}}{k} e^{-k\upsilon\alpha} \binom{m^*}{k} \right]_0^{\tau} \right\}, \tag{1.17}$$

where the brackets denote the difference between their argument evaluated at $\upsilon = \tau$, and at $\upsilon = 0$.

In the limit $\tau \to \infty$, which is equivalent to the concentration of clusters after the fast aggregation timescale has finished, the result is:

$$N(\tau \to \infty) = N(0) \exp\left\{ -\sum_{k=1}^{m^*} \frac{(-1)^{k-1}}{k} \binom{m^*}{k} \right\}, \tag{1.18}$$

which can be rewritten in a way that makes it more explicit the dependency on m^*:

$$N(\tau \to \infty) = \frac{N(0)}{m^*} \exp\left\{ -\sum_{k=1}^{m^*} \frac{(-1)^{k-1}}{k} \binom{m^*}{k} + \log m^* \right\}. \tag{1.19}$$

We can identify

$$\sum_{k=1}^{m^*} \frac{(-1)^{k-1}}{k} \binom{m^*}{k} \tag{1.20}$$

as the m^*-th harmonic number, H_{m^*}, which diverge logarithmically as $m^* \to \infty$. The Euler-Mascheroni constant, γ, is the difference between the n-th harmonic number and the logarithm of n in the limit where $n \to \infty$:

$$\gamma = \lim_{n\to\infty} \left[\sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \binom{n}{k} - \log n \right]. \tag{1.21}$$

Hence, Eq. (1.19), in the large m^* limit, can be recast as

$$N(\tau \to \infty) \overset{m^* \to \infty}{=} \frac{N(0)}{m^*} e^{-\gamma}. \tag{1.22}$$

It is worth noticing that in the large-time limit the results only depend on the critical cluster mass m^*. Thus, Eq. (1.22) is independent of the fast aggregation constant α, which implies that, for the model to be consistent, we only need that the timescale separation hypothesis $\alpha \gg \beta$ is satisfied.

By using Eq. (1.22), we obtain the fraction, $\tilde{N}$, of clusters left in the system after the fast aggregation process has finished

$$\tilde{N}(t_0) = \frac{N(t_0)}{N(0)} = \frac{e^{-\gamma}}{m^*}. \tag{1.23}$$

In Eq. (1.23) we have used that m^* is large, which is justified by the fact that we expect m^* to be of the order of 10^2 or 10^3, and even for $m^* = 10^2$ the error due to this approximation is less than 1%. We will now make use of Eq. (1.23) as the initial condition of the slow-timescale dynamics.

1.3.2 Slow-aggregation timescale

Minutes after the start of the aggregation process, the fast aggregation process is finished (around time $t \sim t_0$). Based on the experiments, we assumed t_0 to be small compared to the final time of the experiment $t_\mathrm{f} = 1\,\mathrm{hr}$, which allows us to neglect it and assume that the slow-aggregation timescale lasts for 1 hr (and not $t_\mathrm{f} - t_0$).

To describe the slow-aggregation regime $t > t_0$, we write a Smoluchowski coagulation equation with $k_{i,j} = \beta$, where β is the slow-aggregation rate. The solution in this case was first given by von Smoluchowski [Krapivsky et al., 2010]. Starting from the Smoluchowski coagulation equation (1.1), with the kernel $k_{j,k} = \beta$:

$$\frac{dc_i(t)}{dt} = \frac{\beta}{2} \sum_{j+k=i} c_j(t)c_k(t) - \beta c_i(t) \sum_k c_k(t), \tag{1.24}$$

we sum over all i and rewrite the constraint on the summation $j + k = i$ with a Kronecker delta, yielding

$$\begin{aligned}\sum_i \frac{dc_i(t)}{dt} =& \frac{\beta}{2} \sum_{i,j,k} \delta_{j+k,i} c_j(t)c_k(t) - \sum_i \beta c_i(t) \sum_k c_k(t) \\ =& \frac{\beta}{2} \sum_{j,k} c_j(t)c_k(t) - \sum_i \beta c_i(t) \sum_k c_k(t),\end{aligned} \tag{1.25}$$

and with the definition $N(t) = \sum_i c_i(t)$ we obtain

$$\frac{dN(t)}{dt} = -\frac{\beta}{2} N(t)^2. \tag{1.26}$$

The solution to this equation is

$$N(t) = \frac{1}{\beta t/2 + C}, \tag{1.27}$$

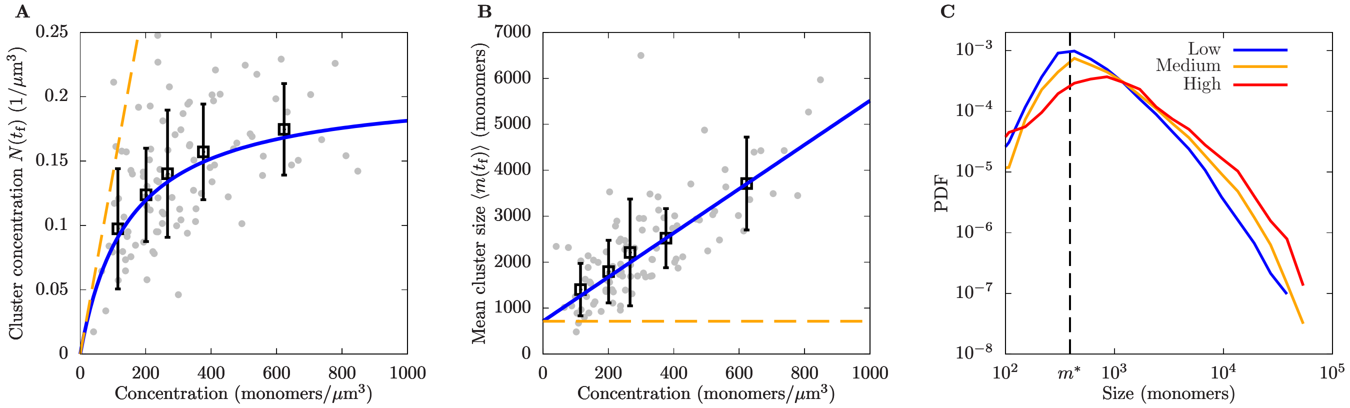


Figure 1.2: Experimental data and fits to the theoretical model. Results obtained for different cells 1 hr after the beginning of the aggregation process. In (A) and (B) we plot the cluster concentration and mean cluster size as functions of the protein density ρ measured in each cell, respectively. Grey dots correspond to results for individual cells, black squares to the average over 20 cells, and error bars to standard deviations. Blue lines correspond to least-square fits of the theoretical expressions for the cluster concentration and cluster size, that is, Eqs. (1.28) and (1.29). The orange dashed lines correspond to the predictions for a passive cytoskeleton, i.e., in the absence of an active dynamics ($\beta = 0$). (C) Cluster-size probability density function (PDF) for different protein concentrations: Low (below 300 monomers/μm^3), medium (between 300 and 600 monomers/μm^3) and high (above 600 monomers/μm^3). The black dashed line corresponds to the estimate of the cluster-mass threshold between timescales, m^*, inferred from (A) and (B).

where, imposing the initial condition $N(0) = \rho\tilde{N}(t_0)$ with ρ is the initial density of monomers, C takes the value $(\rho\tilde{N}(t_0))^{-1}$. Integrating Eq. (1.26) from $t = 0$ to t_f we obtain

$$N(t_\mathrm{f}) = \frac{\rho\tilde{N}(t_0)}{\rho\tilde{N}(t_0)t_\mathrm{f}\beta/2 + 1}. \tag{1.28}$$

Equation (1.28) has two unknown parameters: β and $\tilde{N}(t_0)$ which can be estimated from the experimental data. In particular, $\tilde{N}(t_0)$ can be obtained from relation (1.23), which implies that it is fully determined by the size at which the aggregate attains a size comparable to the pore of the cytoplasm (m^*). Furthermore, using the relationship $\rho = N(t_\mathrm{f})\langle m(t_\mathrm{f})\rangle$ (where $\langle m(t_\mathrm{f})\rangle$ stands for the mean cluster mass at time t_f), one can estimate the mean cluster mass at the end of the experiment as a function of the density of protein:

$$\langle m(t_\mathrm{f})\rangle = \rho\, t_\mathrm{f}\beta/2 + \tilde{N}(t_0)^{-1}. \tag{1.29}$$

In this chapter, equations (1.28) and (1.29), for the slow timescale, and Eq. (1.23) for the fast one, constitute our main theoretical results.

1.4 Comparison with experiments and conclusions

In order to test the model, we fit Eqs. (1.28) and (1.29) to the experimental data for the mean cluster mass and cluster density at the end of the experiment (for details on the fitting procedure, see Appendix A.3). Results are shown in Figure 1.2 A and B. The fit yields $\beta = 9.6\,\mathrm{hr}^{-1}\,\mu\mathrm{m}^3$

and $\tilde{N}(t_0) = 1.4 \times 10^{-3}$. Using Eq. (1.23), one obtains a value of $m^* = 390$ monomers for the mass threshold above which clusters are expected to be trapped in the cytoplasm.

The experimental data also allowed us to quantify the cluster-size distribution, shown in Fig. 1.2 C. Our estimate of m^* is close to the peak of the cluster-size distribution: This result is consistent with the assumptions in the model, where the aggregation process is slowed down for clusters of mass above m^*.

We can also assess the consistency of our result with other experimental data by estimating the pore size of the cytoplasm from the prediction for m^*. To this end, we can utilise the framework of diffusion-limited cluster aggregation (DLCA) [Kolb et al., 1983]. This theory assumes that clusters diffuse freely and bind to each other as soon as they come to contact. If the bonds created by each binding event are rigid and maintain their shape the resulting structure would be a very sparse fractal aggregate. However, in many cases this may not be true and the bonds may rearrange to make a more compact structure.

Taking into account this rearrangement, it has been suggested from numerical simulations that the fractal dimension of the DLCA clusters is approximately $d_f = 2.18$ [Meakin and Jullien, 1988], i.e.,

$$\left(\frac{R_{m^*}}{r_0}\right)^{d_f} \sim m^* \tag{1.30}$$

where R_{m^*} is the radius of an aggregate of mass m^* and $r_0 = 2.5\,\mathrm{nm}$ is the radius of an individual CRY2olig monomer, that is, the average size of a protein containing ~ 500 residues, which allows us to obtain a characteristic radius of the aggregate R_{m^*}. This radius stands for the typical size of the aggregates, and it does not strictly represent the radius of an aggregate, nor implies that the aggregate has spherical shape. Unlike other aggregating proteins that tend to form fibrils [Qiang et al., 2017], it is unclear the geometry of the structures formed by CRY2olig, which is why our estimate does not include any structural information about the aggregate other than the assumption that monomers bind to each other as soon as they come into contact.

Therefore, we obtain for the radius of an aggregate of mass m^*: $R_{m^*} \sim 39\,\mathrm{nm}$. It should be noted that these calculations are correct up to a constant that we cannot determine: Nevertheless, our estimates are coherent with the threshold found in Ref. [Etoc et al., 2018], where they set the threshold between diffusing and non-diffusing particles to be between 25 and 37.5 nm.

Finally, this estimate of the size, together with the good agreement between functional form predicted by the theory and the experimental data, allows us to validate the picture of two different timescales setting the number and size of large protein aggregates, to a great extent. It is possible to refine this model by choosing more complex kernels $k_{i,j}$ for the Smoluchowski coagulation equation in order to explain the cluster-size distribution, but this simple model with only two free parameters (m^* and β) already captures much of the dynamics and the empirical results.

Conclusion and outlook

In this chapter, we have studied diffusion-limited aggregation of CRY2olig protein in mammalian cells, combining an experimental and a theoretical approach.

Our main result is the identification of two different timescales in the aggregation process:

On the one hand, there exists a short timescale where small clusters can freely diffuse and aggregate, leading to the formation of larger agglomerates. On the other hand, later on, large aggregates barely diffuse or do so very slowly. Based on previous work [Luby-Phelps et al., 1987, Guo et al., 2014, Etoc et al., 2018], this effect could be largely due to confinement of the aggregates within the cytoskeleton and other cytosolic obstacles. As a result, large clusters cannot diffuse nor aggregate, unless the confining obstacles move or rearrange on a longer timescale. The predicted threshold between the two timescales corresponds to cluster sizes of ~ 400 monomers, or $\sim 39\,\mathrm{nm}$ of radius, which roughly corresponds to the cytosolic pore size [Etoc et al., 2018].

Our model yields a quantitative estimate of the aggregation rate, β, relative to the long time scale: This rate would characterise the incoherent dynamics of an intracellular network of active forces, such as molecular motors [Guo et al., 2014], which could thus be regarded as an active stirring of the aggregates.

In addition, our analysis demonstrates that clustering of CRY2olig in mammalian cells is markedly different from aggregation in a passive material with a fixed pore size, where the dynamics of the aggregation process would halt as soon as the aggregates' size reaches the pore size. This comparison was made in Fig. 1.2 A and B, where the orange dashed lines represent the predictions for a passive material with the same pore size as that of the cells in our experiment ($\beta = 0$), while solid blue lines represent our model prediction, which includes the active stirring of clusters.

The ideas developed in this study can be generalised to a variety of biological systems that reach a steady state driven by non-equilibrium processes, such as synthesis, degradation, traffic or recycling of proteins [Turner et al., 2005]. In addition, the mechanisms identified here could be extended to the kinetics of other intracellular phenomena, such as liquid-liquid phase separation [Garcia-Jove Navarro et al., 2019]. Indeed, systems under binodal phase separation might exhibit as well two timescales in their coarsening dynamics. The fast timescale rate, α, would represent the diffusion-limited coalescence of droplets in the early kinetics. On the other hand, as droplets grow and diffusion slows down, the main driving force of coarsening would presumably be Ostwald ripening, whose details could be taken into account by a parameter, or function, equivalent to the slow aggregation rate, β. Given that there is a free-energetic cost for a droplet to deform around a network of obstacles [Shin et al., 2018], the effect of obstacles in diffusion would become important only for droplets with a characteristic radius R_{m^*} or larger. Thus, we expect the value of R_{m^*} for this case to be similar to the one predicted by our analysis.

Here, we have focused our analysis on the kinetics of protein aggregation and found that these kinetics too can influence patterns of spatial localisation in cells, such as the size of protein aggregates. Therefore, this mechanism, by which the dynamics are arrested to a large extent, also plays an important role in setting the spatial organisation at the scale of the cell, at least in cases where the processes are slow compared to the timescales of interest in cells and, thus, a steady-state is never reached. In the following chapter we will study the converse case: that of a system that does reach a non-equilibrium steady state which organises the intracellular medium.

Chapter 2

Intracellular bacterial organisation: the nucleoid

Living systems show a high degree of organisation at multiple scales, from the molecular one to the macroscopic scales of organisms and ecosystems[1]. The study of self-organisation and patterning in living beings has attracted much attention from a variety of scientific communities, from molecular and developmental biology to physics, chemistry and applied mathematics. A seminal contribution from the latter field is the morphogenetic mechanism based on reaction-diffusion systems proposed by A. M. Turing in 1952 [Turing, 1952]. Nevertheless, in recent years, many other self-organisation mechanisms have been proposed such as mechano-chemical feedbacks [Boocock et al., 2021], growth-related patterning [Wang et al., 2008] and liquid-liquid phase separation (LLPS) [Brangwynne et al., 2009]. In this chapter we will focus on the spatial organisation of bacteria, which usually lack membrane-enclosed organelles yet achieve a striking degree of organisation.

A notable example of spatial organisation in bacterial cells is that of their DNA: Despite the absence of a nuclear membrane, in many bacteria such as *Escherichia coli* (*E. coli*), the chromosome is not randomly spread throughout the intracellular space, but is markedly localised [Lewis, 2004, Bakshi et al., 2012], and forms a compact structure.

The *E. coli* chromosome is a circular double stranded DNA molecule of around 4.6×10^6 base pairs, $\sim$1.5 mm of contour length, with a persistence length of $\sim$ 50nm. Therefore, if the chromosome behaved as a random coil it would occupy a volume of $\sim 500\,\mu\text{m}^3$ [Verma et al., 2019]. However, the typical *E. coli* cell is a few micrometers in length (a volume of $\sim 2\mu\text{m}^3$), which means that the chromosome has to be highly compacted to fit into the cell. In fact, not only does the chromosome fit into an *E. coli* cell but it does not even fill it completely, as it localises to a particular region of the cell – the nucleoid [Lewis, 2004, Surovtsev and Jacobs-Wagner, 2018].

The localisation and degree of confinement of the nucleoid varies with growth rate and among bacterial species [Gray et al., 2019]. This organisation and localisation of the chromosome has

---

[1] The work presented in this chapter has been published in Proceedings of the National Academy of Sciences of the USA, Ref. [Miangolarra et al., 2021b]. The present chapter is therefore largely based on that reference. The work was done in collaboration with Ned S. Wingreen and Sophia Hsin-Jung Li.

been shown to play an important role in many biological processes, including transcription via the distribution of RNA polymerases [Weng et al., 2019], translation via the localisation of mRNAs and ribosomes [Sanamrad et al., 2014], and the positioning and diffusion of protein aggregates [Coquel et al., 2013].

In this chapter, we aim at theoretically describing the *E. coli* nucleoid, its formation, compaction and localisation at the various stages of the cell cycle. In order to do that, we will construct a mechanism reminiscent of LLPS, based on the steric interactions present in the crowded cytoplasm and including the non-equilibrium effects of transcription and translation, which demonstrates that localisation patterns on the cellular scale emerge spontaneously from microscopic features on a molecular scale. In particular, we show that the segregation of the nucleoid from mRNAs and ribosomes can be understood in terms of excluded-volume effects only, which implies that it is a consequence of equilibrium statistical physics, as in classical phase separations. In addition, our analysis shows that other dynamical features, such as nucleoid positioning, can be driven by the synthesis and degradation of mRNAs, making it a purely non-equilibrium feature. Finally, we compare these results with experimental data obtained from cells growing filamentously [Wu et al., 2019], either with or without chromosome replication, providing important physical and mechanistic insights.

2.1 Characteristics of the nucleoid: Compaction and localisation

Despite the importance of chromosome localisation, the physical causes and regulatory mechanisms of its confinement are still largely unknown. One of the causes of the compaction of the nucleoid could be the fact that the cytoplasm acts as a poor solvent for the chromosome [Xiang et al., 2021], but many other factors could also affect nucleoid compaction, like nucleoid-associated proteins that modify the folding conformation of the chromosome [Dorman, 2013].

The prokaryotic nucleoid is not as structured as its eukaryotic counterpart. Nevertheless, there are still a wide variety of proteins and enzymes that regulate certain aspects of it. One of them is the histone-like nucleoid-structuring protein H-NS that drives the formation of loops between the sites to which it binds [Dame et al., 2020], effectively bridging between different locations in the chromosome. While H-NS primarily bridges between chromosome regions and controls loops, other nucleoid-associated protein bend the DNA, such as HU or Fis (factor for inversion stimulation) and have also been related to compaction of certain regions of the nucleoid [Remesh et al., 2020].

Furthermore, Dps (a nucleoid-associated protein whose expression increases drastically in stationary phase *E. coli* cells) is also known to cause global nucleoid compaction by unespecifically binding to the DNA in a cooperative manner [Almiron et al., 1992, Janissen et al., 2018]. In contrast, the protein MatP binds specifically to a motif called *matS* localised specifically in the Ter region of the chromosome, compacting and organising this part of the genome only [Mercier et al., 2008].

Finally, the supercoiling of the chromosome can be altered by the action of topoisomerases (enzymes that break and rejoin DNA strands) [Dame et al., 2020]. For example, the enzyme gyrase has been shown to control supercoiling of the DNA plectoneme, which has been linked to nucleoid compaction [Stuger et al., 2002].

However, there are also physical forces that compact the nucleoid. It is often suggested that crowding agents in the cytoplasm exert a force against the chromosome which is a crucial factor for the compaction of the nucleoid [Cabrera et al., 2009, Zhang et al., 2009, Yang et al., 2020]. More generally, when an *E. coli* cell is lysed, their intracellular components spread across the solution and, in particular, its chromosome was observed to expand abruptly [Pelletier et al., 2012], implying that it is mostly repulsive forces who hold the together nucleoid within the cell. Overall, this suggest that repulsive interactions like steric or excluded-volume effects are crucial for nucleoid compaction – an insight that we will leverage in the following for the theoretical modelling of the nucleoid.

Regarding the localisation of the nucleoid within the cell, various studies have shown that, while the nucleoid is located at the center of the cell before chromosome replication, during and after replication the daughter chromosomes move out of the centre [Joshi et al., 2011], typically localizing at 1/4 and 3/4 positions on the long cell axis [Wu et al., 2019]. Indeed, this positioning is very robust and occurs for a very wide range of cell lengths [Wu et al., 2019], but it can be perturbed – or even destroyed – by treating the cells with several drugs, such as kasugamycin, due to yet unknown reasons [Xiang et al., 2021]. Moreover, the general mechanism underpinning this localisation pattern remains unclear and has been hypothesised to require an active process [Joyeux, 2019].

2.2 Modelling the nucleoid

Previous theoretical efforts to explain the compaction and localisation of the nucleoid have been mostly based on Monte Carlo simulations [Mondal et al., 2011, Joyeux, 2019]. It was found that excluded-volume effects between DNA and polysomes (mRNAs bound to multiple ribosomes) may account for segregation of the nucleoid from the rest of the cytoplasm [Mondal et al., 2011].

By contrast, in this section, we develop a statistical-physics description of the spatial localisation of the molecular components of the *E. coli* transcriptional-translational machinery (TTM) – composed of DNA, mRNAs, and ribosomes – and identify a set of physical mechanisms underlying their localisation patterns. Unlike previous studies, we leverage semi-analytical methods, i.e., the virial expansion, which allow us to tackle the complexity of the system, and reduce it to a set of computationally tractable reaction-diffusion equations.

We thus describe the dynamics of the *E. coli* TTM by means of a minimal non-equilibrium statistical physics model. The model includes steric interactions only – mathematically described by hard-core interacting potentials – because our aim is to provide a minimal framework that can capture the phenomenology of the bacterial nucleoid. Other interactions may be taken into account, such as weak van der Waals forces among macromolecules, the effects of nucleoid-associated proteins or effective attractive interactions between ribosomes and DNA, due to ribosomes which bind nascent mRNAs linked to DNA. However, there is evidence suggesting that excluded-volume effects are the key players in the formation and morphology of the nucleoid [Yang et al., 2020, Zhang et al., 2009], and that interactions between the chromosome and the cytoplasmic components are essentially repulsive [Pelletier et al., 2012]. Given that interactions are repulsive, to a first approximation, they can be modelled by hard-core potentials [Andersen et al., 1971] and if there were any attractive interactions, provided these are not dominant, they could effectively be taken into account by modifying the steric-interaction parameters (radius or length of particles). Indeed, as shown in the rest of this chapter, steric interactions alone can account for many distinct aspects of nucleoid behavior.

Apart from the inter-particle potentials, we also consider inherently non-equilibrium effects, namely, transcription and translation. In a simplified approach, transcription is modelled by a rate of synthesis of mRNAs linear with the concentration of DNA. Transcription is described by the binding and unbinding of ribosomes to bare mRNAs or polysomes. Finally, we also include the degradation of mRNAs in the model as a linear term in the dynamical equations, see Eqs. (2.1-2.3).

By observing that *E. coli* cells have an approximately cylindrical shape and symmetry, we reduce the three-dimensional cytoplasm to a single dimension along the long cell axis (see Fig. 2.1 A and B) and describe the TTM in terms of the one-dimensional concentrations of DNA segments, mRNAs, and ribosomes. Implicitly, this assumes that the cell is homogeneous in the radial direction, a hypothesis whose consequences will be assessed in the discussion section (Section 2.5). Within this homogeneity approximation, we denote by $c_{\text{DNA}}(x,t)$ the concentration of DNA plectoneme segments at position x along the long cell axis and time t, by $\rho_n(x,t)$ that of polysomes composed of an mRNA and n ribosomes, and by $c_{\text{F}}(x,t)$ that of freely diffusing ribosomes, see Fig. 2.1 A and B. We then consider the reaction-diffusion equations for these concentrations, where we incorporate the diffusion currents and the chemical reactions, i.e., ribosome-mRNA binding and unbinding, mRNA synthesis and degradation:

$$\partial_t c_{\text{DNA}}(x,t) = -\,\partial_x J_{\text{DNA}}(x,t), \tag{2.1}$$

$$\begin{aligned}\partial_t \rho_n(x,t) =& -\,\partial_x J_n(x,t) - k_{\text{on}} c_{\text{F}}(x,t)\rho_n(x,t) - k_{\text{off}}\, n\, \rho_n(x,t) + k_{\text{on}} c_{\text{F}}(x,t)\rho_{n-1}(x,t) \\ &+ k_{\text{off}}(n+1)\rho_{n+1}(x,t) + \alpha\, c_{\text{DNA}}(x,t)\delta_{n,0} - \beta\rho_n(x,t),\end{aligned} \tag{2.2}$$

$$\partial_t c_{\text{F}}(x,t) = -\,\partial_x J_{\text{F}}(x,t) - k_{\text{on}} c_{\text{F}}(x,t)\sum_n \rho_n(x,t) + k_{\text{off}}\sum_n n\rho_n(x,t) + \beta\sum_n n\rho_n(x,t). \tag{2.3}$$

In Eqs. (2.1-2.3), J_{DNA}, J_n, and J_{F} denote the particle currents (derived in Subsections 2.2.1 and 2.2.3), k_{on} and k_{off} the rate constants for ribosome binding and unbinding due to completion of translation, respectively, α the rate at which mRNAs are created locally by transcription, and β the mRNA degradation rate. We assume no-flux (Neumann) boundary conditions for all the components of the TTM at $x = \pm\ell$ where ℓ is the cell half-length.

Regarding the steric interactions, as shown in Fig. 2.1 A and B, we consider ribosomes as spheres of radius R and, because mRNAs and polysomes with n ribosomes are roughly globular polymer coils, we also approximate them as spheres of radius R_0 and R_n, respectively. Because the *E. coli* DNA has a branched, plectonemic structure with a well-defined persistence length and transverse radius [Odijk, 2000], we consider the chromosome as a set of cylindrical segments, where the length of each segment (L, the Kuhn length) corresponds to twice the persistence length [Rubinstein and Colby, 2003]. For the sake of computational tractability, we treat the DNA segments as disconnected, as shown in Fig. 2.1 B.

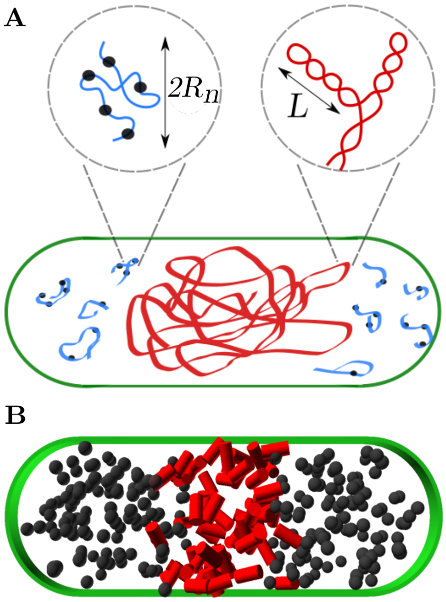


Figure 2.1: Sketch of the model for the nucleoid. (A) Cartoon of an *E. coli* cell and its transcriptional-translational machinery, where the horizontal axis is the single dimension we consider. Blue coils represent mRNAs in polysomes, the red coil denotes the DNA plectoneme and ribosomes are shown in black. Blow-ups: polysome composed of an mRNA and n ribosomes with gyration radius R_n (left) and plectonemic structure of the DNA with persistence length L (right). (B) Schematic of the components underlying the reaction-diffusion model, where the DNA plectoneme is represented by a set of disjoint cylinders, and polysomes and free ribosomes by spheres.

2.2.1 Equilibrium free energy of an interacting inhomogeneous gas

Our model for the nucleoid assumes that all particles interact through hard-core potentials (also known as excluded-volume interactions) and the expression for the particle currents J must be obtained for such system. In order to do that we first need to compute the free energy of the system at thermodynamic equilibrium.

We use the virial expansion, first developed by Onnes [Onnes, 1902] over a century ago, to compute the free energy of a gas of interacting particles. Here we limit ourselves to state the result for hard-sphere potentials to second and third order in the expansion, and refer the interested reader to classical statistical-physics textbooks, e.g., [Huang, 1987, Pathria, 1996], for a complete explanation of the procedure, or to Appendix B.1 for a brief summary.

Free energy of a binary mixture of hard spheres

We first limit ourselves to a binary mixture of hard spheres and later we will include the effect of DNA cylinders. We consider N_{A} and N_{B} particles of species A and B, which are hard spheres with radii R_{A}, R_{B} and diffusion coefficients D_{A}, D_{B}, respectively, confined in a volume $\mathcal{V}$. The hard-sphere potential V_{ij} between the ith and jth particle is zero if the distance between particles is larger than the sum of their radii, and infinity otherwise. Then, the free energy of the system, up to second order in the virial expansion (third viral coefficients) is

$$\begin{aligned} F =& -k_{\mathrm{B}}T \log Z \\ =& k_{\mathrm{B}}T \sum_{a=\mathrm{A,B}} N_a \log \frac{N_a \Lambda_a^3}{\mathcal{V}} - k_{\mathrm{B}}T \log \Bigg[1 - \sum_{a=\mathrm{A,B}} \frac{N_a(N_a-1)}{2\mathcal{V}} B_{aa}^{(2)} - \frac{N_{\mathrm{A}} N_{\mathrm{B}}}{\mathcal{V}} B_{\mathrm{AB}}^{(2)} \\ & - \sum_{a=\mathrm{A,B}} \frac{N_a(N_a-1)(N_a-2)}{6\mathcal{V}^2} B_{aaa}^{(3)} - \sum_{a \neq b} \frac{N_a(N_a-1)N_b}{2\mathcal{V}^2} B_{aab}^{(3)} \Bigg], \end{aligned} \tag{2.4}$$

where Z is the partition function as defined in equilibrium statistical physics, k_B is the Boltzmann constant, Λ_a is the thermal de Broglie wavelength of species a and the ith virial coefficients $B_{ab}^{(i)}$ are given by:

$$B_{ab}^{(2)} = \frac{4\pi}{3}(R_a + R_b)^3, \tag{2.5}$$

$$B_{abc}^{(3)} = \frac{16\pi^2}{9}\left[R_b^3R_c^3 + 3R_aR_b^2R_c^2(R_b + R_c) + R_a^3(R_b + R_c)^3 + 3R_a^2R_bR_c(R_b^2 + 3R_bR_c + R_c^2)\right]. \tag{2.6}$$

By assuming that the volume is large, we can approximate the free energy as

$$\begin{aligned} F \simeq k_\mathrm{B}T\Bigg[&\sum_a N_a \log\frac{N_a\Lambda_a^3}{\mathcal{V}} + \sum_{a=\mathrm{A,B}} \frac{N_a(N_a-1)}{2\mathcal{V}}B_{aa}^{(2)} + \frac{N_\mathrm{A}N_\mathrm{B}}{\mathcal{V}}B_{\mathrm{AB}}^{(2)} \\ &+ \sum_{a=\mathrm{A,B}} \frac{N_a(N_a-1)(N_a-2)}{6\mathcal{V}^2}B_{aaa}^{(3)} + \sum_{a\neq b}\frac{N_a(N_a-1)N_b}{2\mathcal{V}^2}B_{aab}^{(3)}\Bigg], \end{aligned} \tag{2.7}$$

where we have used the Taylor expansion of the logarithm.

Now we consider an infinitesimal distance dx, in which there are dN_a molecules of each species, with one-dimensional concentration $c_a(x) = dN_a/dx$. The volume of each of these infinitesimal slices of the system is $dV = \sigma\, dx$, where σ is the cross section of the system. Then,

$$\begin{aligned} F_0 &= \int_{-\ell}^{\ell} dF_0 \\ &\simeq \int_{-\ell}^{\ell} k_\mathrm{B}T dx \Bigg\{ \sum_{a=\mathrm{A,B}} c_a(x)\left[\log c_a(x) + \log\frac{\Lambda_a^3}{\sigma}\right] + \sum_{a,b=\mathrm{A,B}} B_{ab}^{(2)}\frac{[c_a(x)]^2}{2\sigma} \\ &+ \sum_{a,b,c=\mathrm{A,B}} B_{abc}^{(3)}\frac{c_a(x)c_b(x)c_c(x)}{6\sigma^2}\Bigg\}, \end{aligned} \tag{2.8}$$

where we have approximated $N_a - 1$ with N_a, which yields substantial simplification of the functional form of the free energy. In what follows we drop the term $\log \Lambda_a^3/\sigma$ because it only changes the chemical potential by a constant and will not alter the results. As a result, the logarithms with dimensional arguments remain.

The quantity dF_0/dx in Eq. (2.8) is the free-energy density of the system assuming that the concentrations are uniform along the x axis, where this condition is denoted by the subscript '0'. If we assume that the local free energy density of an inhomogeneous system $f = dF/dx$, is a function of the uniform free-energy density and of the derivatives of the concentration, i.e., $f = f(f_0, \nabla c_a, \nabla^2 c_a ...)$, then we can expand it around f_0, considering the concentration and its derivatives as independent variables, as follows [Cahn and Hilliard, 1958]:

$$f = f_0 + \sum_a \gamma_a \frac{d^2c_a}{dx^2} + \frac{k_\mathrm{B}T}{2\sigma}\sum_{a,b}\kappa_{ab}\frac{dc_a}{dx}\frac{dc_b}{dx}, \tag{2.9}$$

where κ_{ab} are the Cahn-Hilliard coefficients, which account for spatial inhomogeneities in the concentrations and set the free energy cost of concentration gradients in the system. The second term in the right-hand side of Eq. (2.9) does not contribute to the total free energy of the system: when spatially integrated, this term vanishes because of the Neumann (no flux) boundary conditions. Then, the total free energy is

$$\begin{aligned} F &= \int_{-\ell}^{\ell} dF \\ &= \int_{-\ell}^{\ell} dF_0 + \frac{k_{\mathrm{B}}T}{2} \int_{-\ell}^{\ell} \sum_{a,b} \kappa_{ab} \frac{dc_a}{dx} \frac{dc_b}{dx} dx, \end{aligned} \tag{2.10}$$

and the chemical potential for, e.g., species A reads

$$\begin{aligned} \mu_{\mathrm{A}}(x) &= \frac{\delta F}{\delta c_{\mathrm{A}}(x)} \\ &= k_{\mathrm{B}}T \left[1 + \log c_{\mathrm{A}}(x) + \sum_{a=\mathrm{A,B}} c_a(x) B^{(2)}_{a,\mathrm{A}} + \frac{1}{2} \sum_{a,b=\mathrm{A,B}} B^{(3)}_{a,b,\mathrm{A}} c_a(x) c_b(x) - \sum_{a=\mathrm{A,B}} \kappa_{a,\mathrm{A}} \frac{d^2 c_a}{dx^2} \right]. \end{aligned} \tag{2.11}$$

For future convenience, we define ν_a as the non-ideal contribution (which stems from steric interactions) to the chemical potential of each species, e.g., species A:

$$\nu_{\mathrm{A}}(x) = \sum_{a=\mathrm{A,B}} c_a(x) B^{(2)}_{a,\mathrm{A}} + \frac{1}{2} \sum_{a,b=\mathrm{A,B}} B^{(3)}_{a,b,\mathrm{A}} c_a(x) c_b(x) - \sum_{a=\mathrm{A,B}} \kappa_{a,\mathrm{A}} \frac{d^2 c_a}{dx^2} \tag{2.12}$$

Free energy of the full TTM

We will now apply the ideas discussed above to the model for the TTM. In particular, we will use the virial expansion discussed for the binary mixture of hard spheres to obtain an expression for the free energy of the TTM. However, there is an important difference that needs to be taken into account: not all the particles are spheres, because the DNA segments are considered to be cylinders of length L and radius ρ.

In addition, a DNA segment does not interact with another DNA segment in the same way in which it would interact with a polysome or ribosome. In fact, two overlapping DNA plectonemes may be nested into each other, as discussed in Ref. [Mondal et al., 2011]. To model this nesting, while we use the radius ρ to describe overlaps between a DNA cylinder and ribosomes or mRNAs in the virial expansion, we use a smaller, effective radius $\rho' < \rho$ for overlaps between two DNA cylinders. We base the value of ρ' on the hard-sphere model for DNA used for numerical simulations in Ref. [Mondal et al., 2011], where each plectoneme segment is represented as a sequence of four bond beads and two node beads, and all beads have radius ρ. In the simulations, whenever a DNA segment collides with a particle which is not a DNA segment, none of the beads are allowed to overlap with the particle. On the other hand, whenever two DNA segments collide, node beads cannot overlap with each other, but bond beads can, according to the picture above. Given that the two node beads are located at the vertices which connect segments, each node bead contributes half of its volume to each plectoneme segment. The volume that a DNA segment excludes to other DNA segments, which we denote by $\pi\rho'^2 L$, is thus the volume of one node bead,

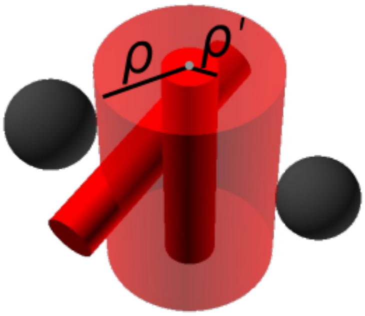


Figure 2.2: Steric interactions of DNA segments. Hard-core interactions between DNA segments (red cylinders) and ribosomes or polysomes (black spheres). DNA segments interact with each other through a cylinder of radius ρ', while they interact with polysomes and ribosomes with a radius $\rho > \rho'$, satisfying Eq. (2.13).

i.e., one fifth of the volume $\pi\rho^2 L$ that the segment excludes to particles other than DNA. As a result, we obtain the relation

$$\rho' = \rho/\sqrt{5} \tag{2.13}$$

between ρ and ρ'. See Fig. 2.2 for a sketch of this interactions.

Moreover, ρ' being an effective parameter, it also includes effects from the nucleoid-associated proteins that control chromosome folding at the molecular scale, as discussed in Section 2.1. Because nucleoid size can be affected by varying the concentration and functionality of these proteins and enzymes, in theory it should be possible to modulate the value of ρ' to effectively account for these biochemical perturbations, at least in a phenomenological manner. Provided repulsive interactions are still dominant, even if there are molecular or enzymatic effects in the compaction of the nucleoid, viewing the interactions within the cell as purely steric may help to obtain a more holistic view of nucleoid organisation, since we can integrate many different phenomena into a single parameter. Moreover, as long as the interactions are mostly repulsive, approximating these by a hard-core potential is theoretically well-grounded [Andersen et al., 1971].

Given the shapes of the particles discussed above, the functional form of the contributions of steric effects to the virial expansion $B_{ab}^{(i)}$ for ribosomes and mRNAs remains the same as for hard spheres, but the interaction of other species with DNA, and of DNA with itself, is different as we are considering DNA to be a set of disjoint cylinders. The second virial coefficient for two cylinders is [Herold et al., 2017]

$$B_{\mathrm{DNA\,DNA}}^{(2)} = 2\pi\rho' \left[L\rho' + \frac{1}{2}(L+\rho')(L+\pi\rho') \right], \tag{2.14}$$

and the virial coefficient between one cylinder and a sphere with radius R_n can be computed by performing integrals of the form (B.4) and yields

$$B_{\mathrm{DNA}\,n}^{(2)} = L\pi(R_n+\rho)^2 + 2\pi R_n \left(\rho^2 + \frac{\pi\rho R_n}{2} + \frac{2}{3}R_n^2 \right), \tag{2.15}$$

where the subscript 'DNA' stands for a DNA cylinder, and n for a sphere of radius R_n.

Before we present the expressions for the third virial coefficients for cylinders, let us define

$$u_{ij} = \mathbb{I}(\mathbf{q}_i \cap \mathbf{q}_j), \tag{2.16}$$

where $\mathbf{q}_i$ denotes the degrees of freedom which specify the position and orientation of a particle, i.e., the position of its center of mass and the Euler angles defining the particle orientation. In addition, $\mathbf{q}_i \cap \mathbf{q}_j$ in Eq. (2.16) stands for the condition that particles i and j overlap, i.e., their hard-core potential is nonzero, and the indicator function $\mathbb{I}$ is equal to one if the condition in its argument is satisfied, and zero otherwise.

The third virial coefficients for interactions that involve cylinders are given by the following integral expressions:

$$B^{(3)}_{\mathrm{D}\,nm} = \int d\mathbf{r}_{n\,\mathrm{D}} d\mathbf{r}_{m\,\mathrm{D}} u^{\mathrm{z}}_{\mathrm{D}\,m} u^{\mathrm{z}}_{\mathrm{D}\,n} u_{nm}, \tag{2.17}$$

$$B^{(3)}_{\mathrm{D\,D'}\,n} = \frac{1}{8\pi^2} \int d\mathbf{r}_{\mathrm{D\,D}} d\mathbf{r}_{\mathrm{D}\,n} d\mathbf{r}_{\mathrm{D'}\,n} d(\cos\theta'_{\mathrm{D}}) d(\cos\theta_{\mathrm{D'}}) d(\cos\theta_n) d\phi_{\mathrm{D}} d\phi_{\mathrm{D'}} d\phi_n u^{\mathrm{z}}_{\mathrm{D\,D'}} u^{\mathrm{z}}_{\mathrm{D}\,n} u_{\mathrm{D'}\,n}, \tag{2.18}$$

$$B^{(3)}_{\mathrm{DD'D''}} = \frac{1}{8\pi^2} \int d\mathbf{r}_{\mathrm{DD'}} d\mathbf{r}_{\mathrm{D'D''}} d\mathbf{r}_{\mathrm{DD''}} d(\cos\theta_{\mathrm{D}}) d(\cos\theta_{\mathrm{D'}}) d(\cos\theta_{\mathrm{D''}}) d\phi_{\mathrm{D}} d\phi_{\mathrm{D'}} d\phi_{\mathrm{D''}} u^{\mathrm{z}}_{\mathrm{DD'}} u^{\mathrm{z}}_{\mathrm{DD''}} u_{\mathrm{D'D''}}, \tag{2.19}$$

where the subscripts label different cylinders, ‘D’ is a shorthand for DNA, n and m label the spheres, and vectors $\mathbf{r}_{ij}$ denote the relative position between the centers of mass of particles i and j. In addition, the superscript z means that the axis of cylinder D is parallel to the z axis, so as to leverage spherical symmetry, and θ, ϕ are the polar and azimuthal angles, respectively. While some simplifications of those integrals are possible, there is no known analytical form for these virial coefficients [Straley, 1973], and we obtained them by numerical integration. Nevertheless, this numerical integration is much less computationally costly than the simulations performed in Refs. [Mondal et al., 2011, Joyeux, 2019], as it only needs to be performed once for fixed parameters.

Then, the total free energy of the system is:

$$\frac{F}{k_{\mathrm{B}}T} = \int_{-\ell}^{\ell} dx \left[\sum_{a=\mathrm{F},n,\mathrm{DNA}} \frac{B^{(2)}_{\mathrm{DNA},a}}{\sigma} c_{\mathrm{DNA}}(x) c_a(x) + \sum_{a,b=\mathrm{DNA,F},n} \frac{B^{(3)}_{\mathrm{DNA},a,b}}{2\sigma^2} c_a(x) c_b(x) c_{\mathrm{DNA}}(x) \right] + \int_{-\ell}^{\ell} dx \left(\frac{dF_0}{k_{\mathrm{B}}T} + \sum_{a,b=\mathrm{DNA,F},n} \frac{\kappa_{ab}}{2} \frac{dc_a}{dx} \frac{dc_b}{dx} \right), \tag{2.20}$$

where the sums run over the chemical species denoted by F, DNA and all polysome species, which we denote by ‘n’ in the sums. Note that F_0 has the same structure as Eq. (2.8): in fact, its form does not change because it involves only species of spherical particles. The difference between F_0 in Eq. (2.20) and F_0 in Eq. (2.8) is in the summation indices, which now span over all polysome species and free ribosomes, which are all modelled as spheres.

The free energy of the system should include an entropic term related to the DNA conformation, but, since this term is small compared with the steric interactions, we neglect it for simplicity (a more detailed explanation is given in Section 2.2.2). The steric interactions have been written explicitly for the DNA segments, and the Cahn-Hilliard terms are analogous to those in Eq. (2.10), except for the numerical values of the coefficients κ_{ab}, which depend on the particle geometry.

In principle, the Cahn-Hilliard coefficients κ_{ab} in Eq. (2.20) can be computed by leveraging hard-core interactions as shown in Ref. [Ilker and Joanny, 2020]. However, for our purposes it is enough to observe that the Cahn-Hilliard terms reflect the cost of concentration gradients related to species a and b. It follows that κ_{ab} is an intrinsic feature of the particles of species a and b: Because κ_{ab} has the dimension of the cube of a length, it must be proportional to a product of the linear sizes of the particles of species a and b. In addition, the relation of the Cahn-Hilliard coefficients to differentials of concentrations over infinitesimal length scales indicates that κ_{ab} is physically related to short rather than long length scales. We thus assume that κ_{ab} equals the minimum between the volume of species a and that of species b.

2.2.2 DNA free energy

When deriving the particle currents, the quantity of interest is the free energy of the particles. For independent spherical particles (ribosomes and polysomes), the entropic term in the free energy is included in the virial expansion. However, for the DNA plectoneme the situation is different: Due to the lack of connectivity between the DNA "cylinders" in the model, we do not obtain the correct entropic term from the virial expansion.

In our model of DNA composed by disconnected cylindrical segments, the energetic part of the free energy is given by the interactions between DNA segments, encoded by the virial-expansion terms, see Subsection 2.2.1. Here, we argue that the entropic part of the real free energy can be neglected, because it is much smaller than the virial terms.

The main contribution is related to the free-energetic cost of confining an ideal polymer. This entropic cost is given by [Edwards and Freed, 1969, De Gennes, 1979]:

$$S \simeq -k_\mathrm{B} \frac{N_\mathrm{DNA} L^2}{L_n^2}, \tag{2.21}$$

where L_n is the typical lengthscale on which the polymer is confined and N_DNA is the number of DNA segments. In the case of the nucleoid, $L_n \sim 1\,\mu\mathrm{m}$ and $N_\mathrm{DNA} = 6 \times 10^3$.

By inserting the numerical values of the parameters we obtain an entropic contribution to the free energy of the order of $10^3\, k_\mathrm{B}T$ and a contribution from the virial terms of order

$$k_\mathrm{B}T \frac{N_\mathrm{DNA}^2 B^{(2)}_\mathrm{DNA,DNA}}{2V_n} \sim 10^4\, k_\mathrm{B}T, \tag{2.22}$$

where $V_n \sim L_n^3$ is the typical volume of a nucleoid. Therefore, there is a difference of one order of magnitude between the entropic and the virial term and, for the sake of simplicity, in this work we neglect the entropic contribution to the free energy, and thus to the current of DNA segments.

Auxiliary entropy

In our analysis of Eqs. (2.1) to (2.3), we will first determine the steady state of the system in the absence of reaction and non-equilibrium terms, by numerically minimising the total free energy (2.20). These profiles are then used as initial conditions to integrate forward in time the reaction-diffusion Eqs. (2.1) to (2.3), which include both reaction and non-equilibrium terms, see Subsection 2.2.5 for details on the numerical methods to solve Eqs. (2.1) to (2.3). At the free-energy minimum, the DNA concentration is nonzero in the nucleoid, while it vanishes outside the nucleoid. Given that these equilibrium profiles are entered as initial conditions in Eqs. (2.1) to (2.3), the vanishing concentration above causes numerical instabilities when these equations are numerically integrated forward in time, and can lead to negative concentrations in the non-equilibrium steady state [Shampine et al., 2005]. To overcome this issue, we included a small, additional entropic term in the free energy (2.20). Therefore, this additional terms is included solely for the stability of the numerical methods used. For details about this term see Appendix B.2, and for a comparison between the magnitude of the real free energy and the auxiliary one see Fig. B.1.

2.2.3 Particle currents

We can now work out the currents by considering Fick's law of diffusion [Crank, 1979], namely,

$$J_a(x) = -\frac{D_a}{k_{\mathrm{B}}T} c_a(x) \partial_x \mu_a, \tag{2.23}$$

where D_a is the diffusion constant, which is given by Einstein-Smoluchowski-Sutherland relation [Dill and Bromberg, 2012]

$$D_a = \zeta_a k_{\mathrm{B}} T, \tag{2.24}$$

and ζ_a is the mobility of species a.

In Eq. (2.23) the chemical potential μ_a is obtained from the derivative of the equilibrium free energy, see Eq. (2.11). We are thus constructing a hydrodynamic theory where we only consider variations in the slow modes, e.g. concentration differences along the cell, implicitly assuming that fluctuations in the fast modes are quickly thermalised and equilibrated, thus locally obeying equilibrium statistical physics with a well defined local temperature, pressure and chemical potential. This is known as the assumption of local equilibrium [Kondepundi and Prigogine, 2015] and is the starting point of the Linear Irreversible Thermodynamics framework used to derive these currents [Groot and Mazur, 1983].

Particle current for hard spheres

By substituting the expression (2.11) for the chemical potential in Eq. (2.23), we obtain the current for hard sphere particles, e.g. free ribosomes:

$$\begin{aligned} J_{\mathrm{F}}(x) = -D_{\mathrm{F}} \Bigg\{ & \partial_x c_{\mathrm{F}}(x) + c_{\mathrm{F}}(x) \sum_{a=\mathrm{F,DNA},n} B^{(2)}_{a,\mathrm{F}} \partial_x c_a(x) \\ & + \frac{c_{\mathrm{F}}(x)}{2} \sum_{a,b=\mathrm{F,DNA},n} B^{(3)}_{a,b,\mathrm{F}} [\partial_x c_a(x) c_b(x) + c_a(x) \partial_x c_b(x)] - c_{\mathrm{F}}(x) \sum_{a=\mathrm{F,DNA},n} \kappa_{a,\mathrm{F}} \frac{d^3 c_a}{dx^3} \Bigg\} \end{aligned} \tag{2.25}$$

Similarly, we obtain the current for the rest of the spherical species (mRNAs and polysomes).

Particle current for the full TTM

Proceeding along the lines of the case for hard spheres, we obtain the DNA current from Eq. (2.23), using the free energy (2.20) and the auxiliary free energy (B.15):

$$J_{\mathrm{DNA}}(x) = -c_{\mathrm{DNA}}(x) D_{\mathrm{DNA}} \left[\partial_x \nu_{\mathrm{DNA}}(x) + \partial_x \mu_{\mathrm{aux}}(x) \right], \tag{2.26}$$

where ν_{DNA} is the excluded-volume term analogous to that in Eq. (2.12), which stems from the fact that the chemical potential of DNA segments does not have entropic contribution. The term μ_{aux} is the derivative of the auxiliary free energy (B.13) with respect to the DNA concentration, $c_{\mathrm{DNA}}(x)$, and its contribution to the current reads

$$\begin{aligned}\partial_x \mu_{\mathrm{aux}}(x) = \; & K_{\mathrm{aux}} e^{-D_{\mathrm{aux}} c_{\mathrm{DNA}}(x)/\langle c_{\mathrm{DNA}}(x)\rangle} \partial_x c_{\mathrm{DNA}}(x) \Bigg\{ \frac{1}{c_{\mathrm{DNA}}(x)} \left(1 - D_{\mathrm{aux}} \frac{c_{\mathrm{DNA}}(x)}{\langle c_{\mathrm{DNA}}(x)\rangle}\right) \\ & + D_{\mathrm{aux}} \log[\sigma c_{\mathrm{DNA}}(x)] - \frac{D_{\mathrm{aux}}}{\langle c_{\mathrm{DNA}}(x)\rangle} \left[1 + \log[2\ell\sigma c_{\mathrm{DNA}}(x)] \left(1 - D_{\mathrm{aux}} \frac{c_{\mathrm{DNA}}(x)}{\langle c_{\mathrm{DNA}}(x)\rangle}\right)\right] \Bigg\}.\end{aligned} \tag{2.27}$$

Combining the results obtained for hard spheres, Eqs. (2.8) and (2.25), with those for the full TTM, Eqs. (2.20) and (2.26), we obtain the currents in Eqs. (2.1) to (2.3). Making use of Eqs. (2.25) and (2.26) and of the virial coefficients previously derived, the currents for DNA, ribosomes and polysomes are fully defined.

For completeness, the currents for ribosomes and polysomes are

$$J_{\mathrm{F}}(x) = -\frac{D_{\mathrm{F}}}{k_{\mathrm{B}}T} c_{\mathrm{F}}(x) \partial_x \mu_{\mathrm{F}}, \tag{2.28}$$

$$J_n(x) = -\frac{D_n}{k_{\mathrm{B}}T} \rho_n(x) \partial_x \mu_n, \tag{2.29}$$

respectively, where the chemical potentials are

$$\mu_{\mathrm{F}}(x) = \log c_{\mathrm{F}}(x) + \sum_{a=\mathrm{F},\mathrm{DNA},n} c_a(x) B^{(2)}_{a,\mathrm{F}} + \frac{1}{2} \sum_{a,b=\mathrm{F},\mathrm{DNA},n} B^{(3)}_{a,b,\mathrm{F}} c_a(x) c_b(x) - \sum_{a=\mathrm{F},\mathrm{DNA},n} \kappa_{a,\mathrm{F}} \frac{d^2 c_a}{dx^2}, \tag{2.30}$$

for free ribosomes, and

$$\mu_n(x) = \log \rho_n(x) + \sum_{a=\mathrm{F},\mathrm{DNA},n} c_a(x) B^{(2)}_{a,n} + \frac{1}{2} \sum_{a,b=\mathrm{F},\mathrm{DNA},n} B^{(3)}_{a,b,n} c_a(x) c_b(x) - \sum_{a=\mathrm{F},\mathrm{DNA},n} \kappa_{a,n} \frac{d^2 c_a}{dx^2}, \tag{2.31}$$

for polysomes.

2.2.4 Model parameters

We fix the model parameters from experiments as follows. First, we consider the parameters on a molecular scale: The radius and length of DNA cylinders are $\rho = 10\,\mathrm{nm}$ and $L = 200\,\mathrm{nm}$ [Odijk, 2000, Mondal et al., 2011], respectively, where $L/2$ is approximately the persistence length of a DNA plectoneme[2] [Cunha et al., 2001, Odijk, 2000].

We take the ribosome radius to be $R = 10\,\mathrm{nm}$ [Mondal et al., 2011], and the radius of a ribosome-free mRNA to be $R_0 = 20\,\mathrm{nm}$ [Kaczanowska and Rydén-Aulin, 2007]. The radius R_n

---

[2] In the literature, there is a large variability in the value for the persistence length of the DNA plectoneme, from the 30-50 nm suggested by [Verma et al., 2019, Xiang et al., 2021] to the 75-100 nm used by [Odijk, 2000, Mondal et al., 2011]. We take 100 nm, because this is the value used in the simulations of Ref. [Mondal et al., 2011].

of an mRNA loaded with n ribosomes is estimated as the sum of the volume of a bare mRNA and n times the volume of a ribosome, i.e., $4/3\pi(R_0^3 + nR^3)$ yielding $R_n = (R_0^3 + nR^3)^{1/3}$.

We estimated the diffusion constant of the different species as follows: $D_{\mathrm{F}} = 0.4\,\mu\mathrm{m}^2/\mathrm{s}$ for ribosomes, and $D_n = 5 \times 10^{-2}\mu\mathrm{m}^2/\mathrm{s}$ for bare mRNAs and polysomes [Bakshi et al., 2012, Sanamrad et al., 2014]. For DNA segments it is harder to obtain an estimate based on experimental data. We therefore estimate their diffusion coefficient as follows: We assume that the drag coefficient on the nucleoid as a whole is low enough for the nucleoid to relax relatively rapidly to mechanical equilibrium (for details, see Appendix B.4) and, given that DNA segments have a linear dimension similar to that of polysomes, we assume that their diffusion coefficients will also be similar. Therefore, we take $D_{\mathrm{DNA}} = 10^{-2}\,\mu\mathrm{m}^2/\mathrm{s}$, a value for the diffusion coefficient of DNA segments which allows the nucleoid to react rapidly to perturbations, but is still not larger than the diffusion coefficient of polysomes.

The parameters relative to the cellular scale are the total number of ribosomes per cell N_{F}, the cell half-length ℓ, both of which will be varied, and the radius of the cellular cross section, which is held constant. Because a central aim is to compare to the experiments in [Wu et al., 2019], we are interested in values for a doubling time of $\sim 2\,\mathrm{hr}$ as in that study. We thus interpolated experimental data points for different growth rates, to obtain the parameter values for the desired growth rate (see Appendix Sections B.5 and B.7) and obtained a total number $N_{\mathrm{F}} \sim 7300$ ribosomes, a cross-sectional radius $R_{\mathrm{cell}} \approx 0.4\,\mu\mathrm{m}$, and a cell half-length $\ell \sim 0.9\,\mu\mathrm{m}$ for a reference cell. In addition, the total mRNA concentration for the reference cell was fixed at $\rho_{\mathrm{tot}} = \sum_n \rho_n = 2400\,\mu\mathrm{m}^{-3}$ [Bartholomäus et al., 2016]. The total number of DNA cylinders for the reference cell was taken to be $N_{\mathrm{DNA}} \sim 6700$ segments [Mondal et al., 2011]. When analysing different situations of biological interest, these parameters may be varied, particularly the overall concentrations of DNA segments, ribosomes, and mRNAs.

Finally, we set the reaction constants for the ribosome binding and unbinding to $k_{\mathrm{on}} = 6 \times 10^{-4}\mu\mathrm{m}/\mathrm{s}$, $k_{\mathrm{off}} = 2.5 \times 10^{-2}/\mathrm{s}$ [Castellana et al., 2016]. The mRNA degradation rate $\beta = 3 \times 10^{-3}/\mathrm{s}$ corresponds to an mRNA half life of $\sim 5\,\mathrm{min}$ [Bernstein et al., 2004], and the mRNA synthesis rate α is estimated from the global steady-state condition of Eq. (2.2), $\alpha N_{\mathrm{DNA}} = \beta N_{\mathrm{mRNA}}$ [Castellana et al., 2016], where N_{mRNA} is the total number of mRNA molecules in the cell, i.e., ρ_{tot} times the cell volume.

2.2.5 Numerically solving Eqs. (2.1) to (2.3)

Due to the complexity and nonlinearity of Eqs. (2.1) to (2.3), in order to solve them we need to resort to numerical methods. We will solve these equations in two stages: First we will minimise the free energy of the system to obtain an equilibrium steady state - which is not yet a solution to Eqs. (2.1) to (2.3) - and then we will integrate forward in time Eqs. (2.1) to (2.3) using the equilibrium steady state as initial condition, in order to obtain a non-equilibrium steady state.

This technique for solving the equations reflects the two types of effects we are considering. On the one hand, we have interactions that do not require any source of external work and, thus, in the absence of any other effect, relax to thermodynamic equilibrium. This implies that their steady state can be found by minimising their free energy [Pippard, 1964]. On the other hand, when integrating forward in time, we are adding intrinsically non-equilibrium effects (since transcription and translation require the cell to consume free energy), which lead the system to a non-equilibrium steady state. Below we argue why a minimisation of the free energy yields a steady state. For further details on the numerical methods used and, in particular, for the integration in time of Eqs. (2.1) to (2.3) from an equilibrium steady state, see Appendix B.6.

Free-energy minimum and equilibrium steady state

The diffusion equations for the mixture of hard spheres discussed in Section 2.2.3 (that is, in the absence of any non-equilibrium contribution) read

$$\partial_t c_{\mathrm{a}}(x,t) = -\partial_x J_{\mathrm{a}}(x,t), \tag{2.32}$$

where a stands for each of the diffusive species involved.

The steady state of Eq. (2.32) combined with no-flux boundary conditions, and with a constraint which fixes the total number of particles to a given value (N_a) for each species yields the following set of relations:

$$-\partial_x J_a(x,t) = 0,\; J_a(\pm\ell,t) = 0, \tag{2.33}$$

$$\int_{-\ell}^{\ell} dx\, c_a(x,t) = N_a. \tag{2.34}$$

We will show that Eqs. (2.33) and (2.34) are tantamount to finding the minimum of the free energy with the constraint (2.34), i.e.,

$$\min_{\{c_a\}} F \tag{2.35}$$

$$\text{subject to Eq. (2.34)}, \tag{2.36}$$

where F is now considered to be a functional of the concentration profiles $c_a(x)$, as discussed in Section 2.2.1, and is minimised with respect to the concentration profiles of all species a.

The Lagrange function of the minimisation problem given by Eqs. (2.35) and (2.36) reads

$$\mathscr{L} = F - \sum_a \lambda_a \left[\int_{-\ell}^{\ell} dx\, c_a(x) - N_a\right], \tag{2.37}$$

where λ_a are the Lagrange multipliers. First, the stationarity condition of $\mathscr{L}$ with respect to $c_a(x)$ is given by

$$\begin{aligned} 0 &= \frac{\delta \mathscr{L}}{\delta c_a(x)} \\ &= \frac{\delta F}{\delta c_a(x)} - \lambda_a \\ &= \mu_a(x) - \lambda_a, \end{aligned} \tag{2.38}$$

where in the third line we used the definition of the chemical potential, see the first line of Eq. (2.11). By taking the derivative of Eq. (2.38) with respect to x and using Eq. (2.23), we obtain $J_{\mathrm{a}}(x) = 0$ for all x, which is equivalent to Eq. (2.33). Second, the stationarity condition of $\mathscr{L}$ with respect to λ_a yields Eq. (2.34). As a result, the minimisation problem (2.35), (2.36) is equivalent to conditions (2.33) and (2.34).

2.3 Nucleoid formation and size

So far we have described the model we will be working with and we now turn to analyse the features of it. The first characteristic of the model we study is whether it gives rise to a nucleoid segregated from the crowders (mRNA and ribosomes) and its size.

A system that relaxes to thermodynamic equilibrium at constant volume and in thermal contact with a constant temperature bath, minimises its free energy $F = U - TS$, where T is the temperature of the bath [Pippard, 1964]. In certain cases, the way a system minimises its free-energy implies segregating some components of the mixture from others. This is the case of the mixture of DNA and crowders (mRNAs and ribosomes) which we are considering here, where, as it will be shown below, the DNA demixes from ribosomes and mRNAs. This phenomenon is known as phase separation and the region of space to which the DNA localises after the segregation can be identified with the bacterial nucleoid. Intuitively, in our model for the nucleoid, phase separation is a consequence of the tendency of the system to reduce its interaction energy (and, thus, its overall free energy), which is achieved by segregating plectonemic DNA from polysomes and ribosomes. However, it can also be viewed as steric interactions giving rise to depletion forces [Asakura and Oosawa, 1954], which drive the phase separation of the nucleoid.

In this section we will give a brief introduction to phase-separation phenomena and analyse the equilibrium behaviour of the model. We will then numerically solve Eqs. (2.1) to (2.3), which include non-equilibrium transcription and translation, and compare these solutions with analytical estimates for the nucleoid size.

2.3.1 Segregation of DNA and crowders

Given a certain set of overall particle number constraints [e.g. the type of constraint introduced by Eq. (2.34)], the fact that the system relaxes to a free-energy minimum implies that, at that minimum, the Hessian matrix of the system

$$\mathbf{H}_{i,j} = \frac{\partial^2 F}{\partial x_i \partial x_j}, \tag{2.39}$$

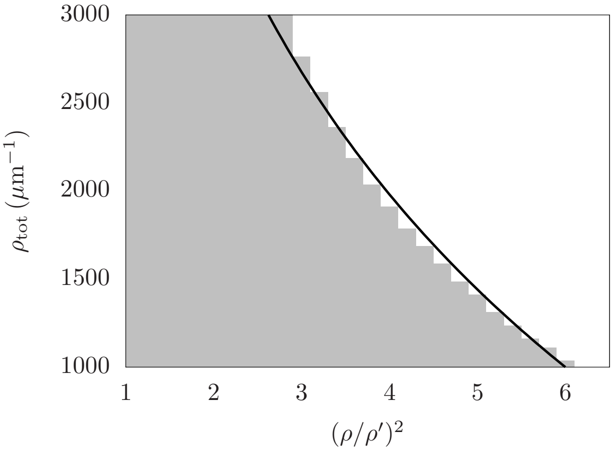
where x_i are the degrees of freedom and F is the free energy of the system, has only positive eigenvalues. If we constrain the system to be homogeneous and find that one or more eigenvalues of the Hessian matrix are negative, then the homogeneous configuration is unstable: focusing on the case of phase separation, this means that the system will demix into two or more distinct phases. This phenomenon is known as spinodal decomposition. In other words, whether spinodal decomposition occurs depends on the curvature of the free-energy function [De Gennes, 1979].

We now consider a simplified version of the free energy density (2.20) of the model:

$$\begin{aligned}\frac{f}{k_\mathrm{B}T} =& c_\mathrm{F}\log c_\mathrm{F} + \sum_n \rho_n \log\rho_n + B^{(2)}_{\mathrm{F,F}}\frac{c_\mathrm{F}^2}{2} + \sum_n B^{(2)}_{n,\mathrm{F}} c_\mathrm{F}\rho_n \\ &+ \frac{1}{2}\sum_{n,n'} B^{(2)}_{n,n'}\rho_{n'}\rho_n + B^{(2)}_{\mathrm{F,DNA}} c_\mathrm{F} c_\mathrm{DNA} + \sum_n B^{(2)}_{n,\mathrm{DNA}} c_\mathrm{DNA}\rho_n + \frac{1}{2}c_\mathrm{DNA}^2 B^{(2)}_{\mathrm{DNA,DNA}},\end{aligned} \tag{2.40}$$

where we have assumed the system is homogeneous and we have only kept the second order virial coefficients for simplicity. Given the total concentrations of ribosomes ($c_{\mathrm{F,tot}}$), mRNA (ρ_tot)

Figure 2.3: Phase diagram for the simplified free energy (2.40). Phase diagram of the free energy (2.40) as a function of the total mRNA concentration ρ_{tot} and the radius by which DNA segments interact with other DNA segments ρ'. The shaded region represents the region of the phase diagram where the homogeneous solution is stable and the white one where it is unstable (region where spinodal decomposition takes place). The approximation of the boundary (black line) was obtained from the best fit to an exponential function.



and DNA segments [$c_{\mathrm{DNA,tot}}$, to distinguish from the position dependent DNA density $c_{\mathrm{DNA}}(x)$], we can obtain the number of free ribosomes and polysomes in a homogeneous system from the kinetics of ribosome binding and unbinding, see Eqs. (2.1) to (2.3):

$$c_{\mathrm{F}} = \frac{c_{\mathrm{F,tot}}}{1 + \rho_{\mathrm{tot}}\frac{k_{\mathrm{on}}}{k_{\mathrm{off}}}}, \qquad \rho_0 = \rho_{\mathrm{tot}} \exp\left(-\frac{c_{\mathrm{F,tot}}}{\frac{k_{\mathrm{off}}}{k_{\mathrm{on}}} + \rho_{\mathrm{tot}}}\right) \quad \text{and} \quad \rho_n = \frac{1}{n!}\rho_0 \left(\frac{k_{\mathrm{on}} c_{\mathrm{F}}}{k_{\mathrm{off}}}\right)^n, \tag{2.41}$$

where, in order to obtain these analytic results, we have assumed that there is no upper bound on the amount of ribosomes that can bind to an mRNA [Castellana et al., 2016], which is justified if the amount of ribosomes that can bind an mRNA is large enough. Substituting these relations in the free energy (2.40), we can obtain the spinodal region of the phase diagram by evaluating the eigenvalues of the Hessian matrix as functions of the concentrations and parameters of the system. If we choose the parameters as Section 2.2.4, but vary the total mRNA concentration ρ_{tot} and the radius by which DNA segments interact with other DNA segments ρ' (which are the two most relevant parameters for the phase behaviour of the system under consideration), we obtain the phase diagram shown in Fig. 2.3.

While the phase diagram Fig. 2.3 provides an estimate for the parameter region where phase separation takes place, the correspondence with the full model is not exact for a number of reasons. First, it is a simplified free energy where only second virial coefficients have been used for simplicity, as opposed to the full model, which includes also third-order virial coefficients. Second, the phase diagram is based on equilibrium thermodynamics, which need not apply to the non-equilibrium model defined by Eqs. (2.1) to (2.3). In this sense, we have obtained the equilibrium phase diagram that is conceptually closest to the non-equilibrium system which we are modeling, because the phase diagram incorporates non-equilibrium effects from ribosome binding and unbinding, as imposed by Eqs. (2.41). Nevertheless, our phase diagram cannot incorporate the effects that may arise from mRNA synthesis and degradation, such as the centering and splitting of the nucleoid that, as it will be shown later, are consequence of transcription. Therefore, the phase diagram Fig. 2.3 is a qualitatively good approximation to the true non-equilibrium dynamics (e.g. it predicts phase separation for the parameter ranges explored numerically later in the text) but it is not exact.

Estimate for the spinodal decomposition within the nucleoid

Here, we will consider a further simplified version of the free energy of the system to estimate the values of the parameters needed to drive spinodal decomposition within the nucleoid, a quantity that will be useful later for the understanding of the splitting of the nucleoid.

We consider the simple case of only two species: DNA segments and bare mRNAs, with the following free energy

$$F = \rho_0 \log \rho_0 + \frac{1}{2} B_{00}^{(2)} \rho_0^2 + B_{0\mathrm{DNA}}^{(2)} c_{\mathrm{DNA}} \rho_0 + \frac{1}{2} c_{\mathrm{DNA}}^2 B_{\mathrm{DNADNA}}^{(2)}. \tag{2.42}$$

Setting all parameters as explained in Section 2.2.4, we can obtain the eigenvalues of the Hessian matrix with respect to these two variables (c_{DNA} and ρ_0) and when any of the two eigenvalues becomes negative, spinodal decomposition occurs. Thus, we find that the concentration of mRNA required at the center of the nucleoid to drive the splitting of the nucleoid is $\rho_0^* \simeq 2500\,\mu\mathrm{m}^{-1}$ (for the typical DNA-segment concentration within the nucleoid of $c_{\mathrm{DNA}} \simeq 6000\,\mu\mathrm{m}^{-1}$, see below and Fig. 2.4 A).

Again, we note that this is not an exact approach and is probably an overestimate of the required mRNA density at the center of the nucleoid. The fact that we only consider bare mRNA and not larger polysomes will cause some of this overestimate. Moreover, considering only second virial coefficients will also cause overestimation of the required mRNA concentration.

2.3.2 Nucleoid size in the presence of non-equilibrium processes

In order to better understand the compaction of the nucleoid, we solved the one-dimensional reaction-diffusion Eqs. (2.1) to (2.3). To compare the predictions of our model with experimental data, in what follows we consider two scenarios for how the concentrations of the molecular species scale with cell length.

Filamentous growth

In filamentous growth, the total number of DNA segments, mRNAs, and ribosomes is proportional to the cell length. Given that one of the primary aims of our analysis is to compare results with the data of Ref. [Wu et al., 2019], which was obtained for slowly growing cells, the cell length is kept fixed, which yields a substantial computational simplification. For each cell length, we first determined the equilibrium steady state of the system by minimising the free energy (2.20), and then numerically integrated the reaction-diffusion Eqs. (2.1) to (2.3) forward in time to reach a non-equilibrium steady state. The results are shown in Fig. 2.4 for different cell lengths, up to the cell length at which the nucleoid spontaneously splits into two lobes, and for different total mRNA densities. The cell length at which the nucleoid spontaneously splits into two lobes is easily identifiable because, when the splitting takes place, the DNA concentration at midcell drops to a value close to zero, comparable to the DNA concentration at the cell poles.

While the non-equilibrium chemical reactions are responsible for the splitting of the nucleoid (as it will be argued in Section 2.4.1), the minimisation of the free energy can account for the existence of a phase-separated nucleoid in the cell due to the steric interactions that give rise to

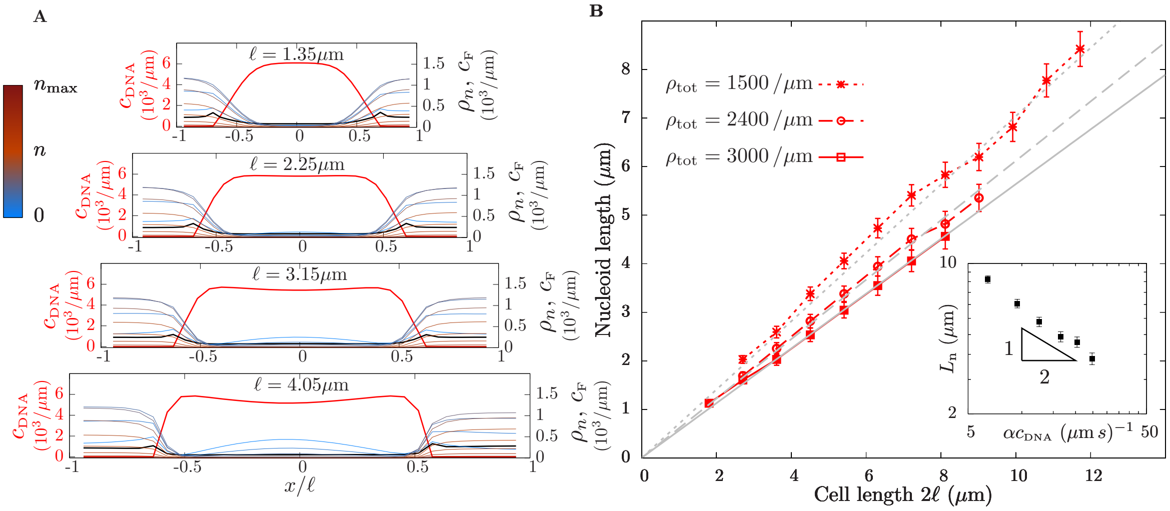


Figure 2.4: Steady-state concentration profiles for *E. coli* growing filamentously. (A) Concentrations along the long axis of the cell of DNA, $c_{\rm DNA}(x)$ (red), and free ribosomes, $c_{\rm F}(x)$ (black), and polysomes $\rho_n(x)$ with the mRNA loading number n indicated by the colour bar. Each panel corresponds to a different cell half-length ℓ, marked on the top of the panel, with total mRNA density $\rho_{\rm tot} = 2400\mu{\rm m}^{-1}$. (B) Nucleoid length versus cell length, 2ℓ, for different values of the total mRNA density, $\rho_{\rm tot}$. For each value of $\rho_{\rm tot}$ the nucleoid length is shown up to the cell length at which the nucleoid splits into two lobes. Red data points were obtained from the numerical solution of the entire model while the grey lines are the analytical estimates. The inset depicts on a log-log plot the power-law relationship between the length at which the nucleoid splits and the rate of mRNA synthesis $\alpha c_{\rm DNA}$, where $c_{\rm DNA}$ is the average DNA concentration along the nucleoid at steady state. The triangle represents the scaling obtained from the simplified model of Eq. (2.49).

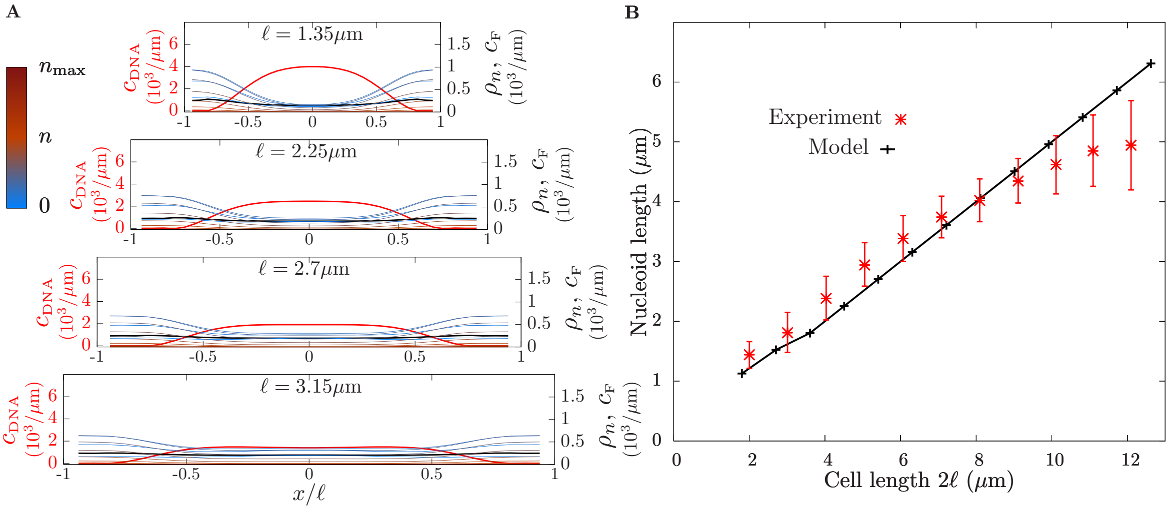


Figure 2.5: Steady-state concentration profiles, for single-chromosome filamentous growth. (A) Concentrations of DNA, $c_{\rm DNA}(x)$, free ribosomes $c_{\rm F}(x)$, and polysomes $\rho_n(x)$, shown as in Fig. 2.4 A. (B) Nucleoid length and standard deviation as a function of cell length from Ref. [Wu et al., 2019] (red) and from the model (black). We found agreement up to cell lengths of around 9μm, after which the prediction of the model deviates from the experimental data.

depletion forces [Asakura and Oosawa, 1954], see Fig B.4 for results obtained in the same way as in Fig. 2.4 but in the absence of non-equilibrium effects. Importantly, neglecting the non-equilibrium effects did not yield a substantial change in nucleoid size.

Furthermore, we observe that the configuration that minimises the free energy excludes ribosomes, mRNAs, and polysomes from the nucleoid to different degrees. For example, the free ribosome concentration is higher at the periphery than within the nucleoid by a factor of ~ 3 while for a polysome with 3 bound ribosomes the ratio rises to ~ 40, which is consistent with experimental observations that polysomes are much more excluded from the nucleoid than are ribosomes [Sanamrad et al., 2014]. As a rule of thumb, we find that the larger the macromolecular complex is, the more excluded it will be from the nucleoid.

The relation between nucleoid and cell length appears to be roughly linear up until the cell length at which the nucleoid begins to split in two, see Fig. 2.4 A and B. Moreover, as shown in Fig. 2.4 B, the higher the total mRNA density, the smaller the nucleoid, implying that a high mRNA density increases the osmotic pressure on the nucleoid, thus making it shrink. In Section 2.3.3 we will make a more formal statement of this observation and we will use it to estimate the nucleoid size analytically.

Single chromosome filamentous growth

So far we have analysed the scaling of nucleoid size with cell size by assuming that the number of DNA segments is proportional to cell length. We now study another case of biological interest, namely, that of a cell with a fixed amount of DNA and varying cell size. This scenario was recently analysed in a dynamic imaging study of the *E. coli* chromosome [Wu et al., 2019], where the initiation of DNA replication and cell division were halted, yielding a single chromosome in a filamentously growing cell. We model this scenario by fixing the number of DNA segments, but allowing the cell size to vary. In addition, the mRNA and ribosome number are no longer proportional to cell length: based on the data in [Kohram, 2021], we assume that the total concentrations of mRNAs and ribosomes decrease linearly with cell length ($\rho_{\text{tot}} \propto a - b\,\ell$, where a and b are constants), approaching zero at $30\,\mu$m – see Appendix B.9 for details.

Results are shown in Fig. 2.5: the model again predicts a roughly linear scaling of the nucleoid size with respect to cell length, while the DNA-segment concentration decreases with cell size. This indicates that the decrease in DNA-segment concentration with cell size is balanced by the decrease of mRNA and ribosome concentrations, so as to keep nucleoid size a linear function of cell size. This can be seen clearly in Fig. 2.5 A where the concentrations of all components of the TTM decrease as the cell size increases. While the model prediction for nucleoid versus cell length agrees reasonably well with experiments [Wu et al., 2019] for cell lengths smaller than $\sim 10\,\mu$m (Fig. 2.5 B), there is a discrepancy for larger cells , see Section 2.5 for further discussion on these results.

Finally, in Fig. B.5, we show an alternative case in which the mRNA and ribosome concentrations are kept constant instead of decreasing with cell size. This yields a predicted nucleoid length very small compared to the experimental data, even for short cells, lending support to the results of Ref. [Kohram, 2021] and the scaling proposed above.

2.3.3 Analytical estimates of the nucleoid size

As seen above, the nucleoid segregates from the rest of the cytoplasm. The nucleoid size at steady state (provided that the nucleoid is single lobed) is set by the mechanical balance of osmotic pressures between the nucleoid and the peripheral cytoplasm. These pressures solely stem from the entropy and steric interactions of the components of the mixture, making the nucleoid size a consequence of equilibrium physics, which is in line with the conclusions of the previous section.

Motivated by the phase separation of the nucleoid and its exclusion of large components of the cytoplasm, in order to estimate the size of this phase-separated nucleoid, we will consider the cell to be divided by two movable walls into three compartments: A central one, the nucleoid, composed exclusively of DNA segments[3], and two lateral ones which include ribosomes and polysomes. To reach mechanical stability, the compartments may expand or contract, moving the walls to an equilibrium position where their osmotic pressures are balanced. The particles interact through steric interactions, described by the virial coefficients. Therefore, we consider the following free energy for the particles within the compartments:

$$F_i = F_{\text{ideal}} + k_{\text{B}}T\frac{N_i^2 B_i}{2V_i}, \tag{2.43}$$

where i denotes the compartment and N_i the particle number, B_i the effective virial coefficient that accounts for the steric interactions among the particles within the compartment, F_{ideal} is the free energy of the ideal gas and V_i the compartment volume. Essesntially, we are coarse graining

---

[3] From the previous section, e.g. Fig 2.4, one can check that this is a good approximation. From the free-energetic point of view, it makes sense to have a compartment enriched in DNA and depleted in crowders, since it decreases the interaction energy due to the form of the inter-particle potentials considered.

all the second virial coefficients of the different particle types that appear in the full free energy (2.20) into a single one for the whole compartment B_i. Then, the osmotic pressure exerted by the compartments is [De Gennes, 1979]

$$\begin{aligned} P_i =& -\frac{\partial F}{\partial V} \\ =& \frac{k_\mathrm{B} T N_i}{V_i}\left(1+\frac{N_i B_i}{2V_i}\right), \end{aligned} \tag{2.44}$$

where the first term stems from the entropic pressure of an ideal gas while the second one, $N_i B_i / V_i$, comes from the steric interactions.

In what follows, we consider Eq. (2.44) in the nucleoid and in the pole compartments, and estimate the respective values of the virial coefficients in the nucleoid, B_n, and at the cell poles, B_p.

In the nucleoid, DNA-DNA interactions dominate, yielding a value of $B_\mathrm{n} \simeq 6.4 \times 10^{-4} \mu\mathrm{m}^3$. For the poles we provide an effective value of the virial coefficient by assuming that all ribosomes are bound to mRNAs, and are equally distributed among them, that is, the poles are occupied by spheres all equal in size. Given that the ratio of ribosomes to mRNAs changes with the amount of mRNA in the cell, the virial coefficients depend on this last parameter. In the cases analysed in Fig. 2.4 B, we obtain the following values for the virial coefficients, using Eq. (2.5), for the corresponding mRNA concentrations:

$$\begin{aligned} B_\mathrm{p}(1500\,\mu\mathrm{m}^{-1}) &= 4.3 \times 10^{-4} \mu\mathrm{m}^3, \\ B_\mathrm{p}(2400\,\mu\mathrm{m}^{-1}) &= 3.7 \times 10^{-4} \mu\mathrm{m}^3, \\ B_\mathrm{p}(3000\,\mu\mathrm{m}^{-1}) &= 3.5 \times 10^{-4} \mu\mathrm{m}^3. \end{aligned} \tag{2.45}$$

By equating the pressures of the compartments, we obtain the equilibrium value for the volumes of each compartment. For filamentous growth, where the number of ribosomes, mRNAs, and DNA segments scales linearly with size ($N_i \propto \ell$), we obtain the solution for the nucleoid size $V_\mathrm{n} = \varphi V$, where φ, the fraction of total volume occupied by the nucleoid, depends on the concentration of mRNAs in the cell:

$$\begin{aligned} \varphi(1500\,\mu\mathrm{m}^{-1}) &= 0.7, \\ \varphi(2400\,\mu\mathrm{m}^{-1}) &= 0.61, \\ \varphi(3000\,\mu\mathrm{m}^{-1}) &= 0.56. \end{aligned} \tag{2.46}$$

As shown in Fig. 2.4 B, these estimates (gray lines) are in good agreement with the numerical solution of the full reaction-diffusion equations (red points).

Nevertheless, the cytoplasm of the cell may not be dilute enough to allow considering only two terms in the virial expansion. The third virial coefficient for the DNA-DNA interaction is $B_\mathrm{n}^{(3)} \simeq 5 \times 10^{-8} \mu\mathrm{m}^6$ and for the polysomes:

$$\begin{aligned} B_\mathrm{p}(1500\,\mu\mathrm{m}^{-1}) &= 9.4 \times 10^{-8} \mu\mathrm{m}^6, \\ B_\mathrm{p}(2400\,\mu\mathrm{m}^{-1}) &= 6.8 \times 10^{-8} \mu\mathrm{m}^6, \\ B_\mathrm{p}(3000\,\mu\mathrm{m}^{-1}) &= 6 \times 10^{-8} \mu\mathrm{m}^6. \end{aligned} \tag{2.47}$$

If we add the third virial coefficient, the fraction of total volume occupied by the nucleoid becomes:

$$\begin{aligned}\varphi(1500\,\mu\mathrm{m}^{-1}) &= 0.69,\\ \varphi(2400\,\mu\mathrm{m}^{-1}) &= 0.60,\\ \varphi(3000\,\mu\mathrm{m}^{-1}) &= 0.55, \end{aligned} \tag{2.48}$$

values that are very close to those obtained with the second virial coefficient alone. Therefore, for simplicity, when comparing with the nucleoid size obtained by numerically solving Eqs. (2.1) to (2.3) we use the value obtained from second order virial coefficients only.

Overall, we find that the inclusion of steric terms in both the nucleoid and mRNA/ribosome compartments makes the nucleoid swell compared to what its size would be with only entropic terms (ideal gas contribution). This is due to the nature of the nucleoid, a long relatively stiff polymer with little entropy per segment compared to ribosomes and mRNAs. Moreover, in Fig. 2.4 B we quantitatively compared the estimates made in this section (grey lines) with the results of the numerical solution to the full model (red data points), which includes non-equilibrium effects, and found good agreement between the two approaches; thus validating the idea of nucleoid size being a consequence of equilibrium physics alone and that the effects of non-equilibrium transcription and translation in nucleoid size are negligible, provided the levels of mRNA and ribosomes are maintained. Therefore, nucleoid size is a consequence of steric interactions alone, which can be parametrised by an effective osmotic pressure to which only large macromolecules contribute (mostly mRNAs and polysomes), since they are the only ones excluded from the nucleoid.

2.4 Consequences of non-equilibrium transcription and translation in intracellular spatial organisation

So far we have analysed the emergence, compaction and size of the nucleoid and concluded that they are a consequence of steric interactions, i.e., mostly driven by processes that do not require free-energy dissipation. In other words, they are *passive* mechanisms. In this section, we will explicitly address the consequences of non-equilibrium transcription and translation in nucleoid positioning, as a result of the particle currents that can be established at steady state.

Currents at steady state

When transcription and translation are taken into account new phenomena occur, particularly at steady state. As seen previously in this chapter, integrating numerically Eqs. (2.1) to (2.3) up to their non-equilibrium steady state yields a profile like the one shown in Fig. 2.6. However, in Fig. 2.6, we stress the fact that there are diffusion currents sustained at steady state due to the transcriptomic activity. Indeed, mRNAs are synthesised within the nucleoid with rate $\alpha c_{\mathrm{DNA}}(x)$, hence their spatial distribution has a local maximum at midcell. Then, mRNAs diffuse towards the poles (see flux arrows in Fig. 2.6), since it is free-energetically favoured. Finally, mRNAs reach the cell poles where, after a typical timescale $\sim \beta^{-1} = 5\,\mathrm{min}$, they are degraded. Note that mRNAs within the nucleoid have no or few ribosomes attached while at the poles they are heavily loaded

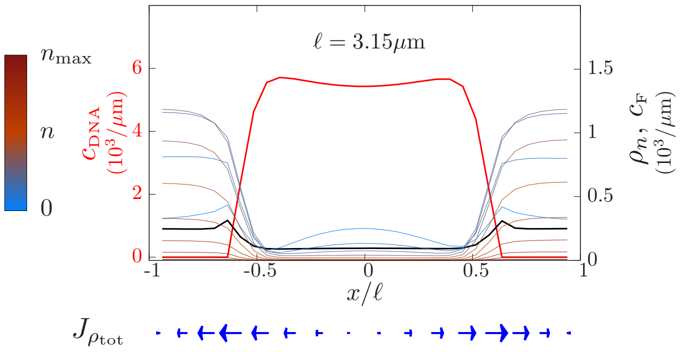


Figure 2.6: mRNA fluxes at steady-state. The plot depicts the concentration profiles at the non-equilibrium steady state. The blue arrows in the bottom depict the steady-state flux of mRNAs, $J_{\rho_{\rm tot}}$ (the arrow length is proportional to the mRNA current). At steady state, mRNAs are synthesised within the nucleoid, wherein they diffuse before escaping to the poles (and are ultimately degraded).

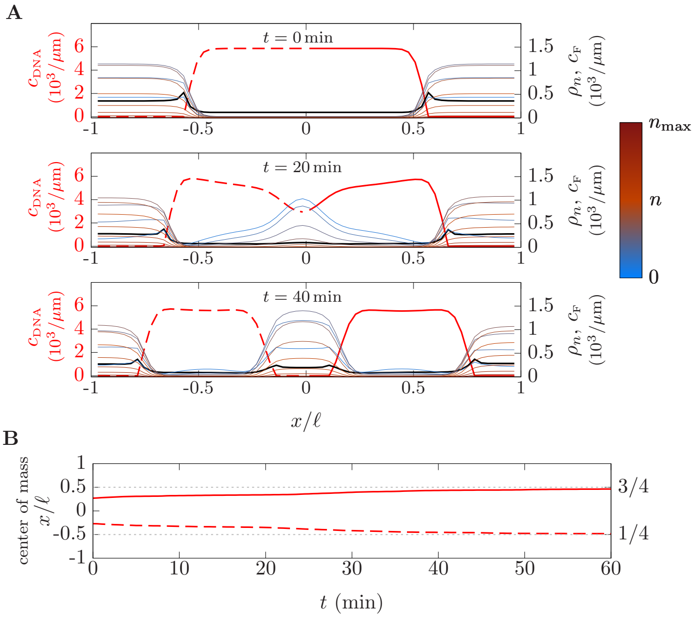


Figure 2.7: Transcription splits the nucleoid into two lobes located at 1/4 and 3/4 of the cell length. (A) Concentration profiles for a filamentous cell, obtained from the equilibrium profile at $t = 0$ by integrating forward in time the reaction-diffusion Eqs. (2.1) to (2.3) in the presence of the non-equilibrium processes until $t = 20\,$min and $t = 40\,$min, for a cell with a half-length $\ell = 4.95\,\mu$m shown as in Fig. 2.4 A. (B) Positions of the center of mass of the left (dashed red curve) and right (solid red curve) halves of the DNA along the long axis of the cell as a fraction of the total cell length, as functions of time.

with them, creating a circulating current for ribosomes too [Castellana et al., 2016]. Altogether, these processes lead to the steady-state distribution shown in Fig. 2.6. In Fig. 2.6 the arrows depict the flux of mRNAs, which is quantified by $J_{\rho_{\rm tot}}(x) = \sum_n J_n(x)$, with $J_n(x)$ given by Eq. (2.29).

These steady-state currents are at the heart of the spatial localisation of the bacterial nucleoid.

2.4.1 Nucleoid splitting

While the linear increase of nucleoid length with cell length is the result of equilibrium osmotic-pressure balance, the splitting of the nucleoid is entirely due to non-equilibrium processes. In fact, for cells with $\rho_{\rm tot} = 2400/\mu$m and a half length of $\sim 4\,\mu$m or larger, the equilibrium steady state used as the initial condition for the reaction-diffusion equations yields a nucleoid with a single lobe. By contrast, the nucleoid splits into two identical lobes positioned at $1/4$ and $3/4$ of the long cell axis when the reaction-diffusion Eqs. (2.1) to (2.3) are integrated forward in time, see Fig. 2.7. Such $1/4$ and $3/4$ positioning of the daughter nucleoids has been ubiquitously observed in experiments [Wu et al., 2019] and is reproduced by the model with a high degree of accuracy, see Fig. 2.7 B.

Scaling for the length at which the nucleoid splits

In what follows, we present a simple argument to explain the dependence of the length at which the nucleoid splits with respect to the underlying parameters, e.g., the mRNA synthesis rate. We take the nucleoid to be a region of length $L_{\rm nucl}$ with homogeneous DNA-segment concentration which extends from $x = -L_{\rm nucl}/2$ to $x = L_{\rm nucl}/2$, with interfaces that are perfectly sharp. The mRNAs synthesised within the nucleoid diffuse until they reach the nucleoid boundaries and, because it is free-energetically favourable, they then escape the nucleoid and not return. As a result, the steady-state concentration of mRNAs within the nucleoid can be modeled by the following diffusion equation with a uniform source term due to mRNA synthesis and absorbing boundary conditions, which represent mRNAs escaping from the nucleoid:

$$D_n \frac{\partial^2 \rho_{\rm tot}(x)}{\partial x^2} + \alpha c_{\rm DNA} = 0, \qquad \rho_{\rm tot}\left(\pm \frac{L_{\rm nucl}}{2}\right) = 0, \tag{2.49}$$

where $\rho_{\rm tot}(x)$ is the total mRNA concentration at position x, D_n the mRNA diffusion constant (as defined in Section 2.2.4), and $\alpha c_{\rm DNA}$ the rate of mRNA synthesis. The solution of the above equation is $\rho_{\rm tot}(x) = (L^2_{\rm nucl}/4 - x^2)\,\alpha\, c_{\rm DNA}/(2D_n)$, whose local maximum within the nucleoid at $x = 0$ takes the value $L^2_{\rm nucl}\,\alpha\, c_{\rm DNA}/(8D_n)$.

We hypothesise that when the mRNA concentration at the center becomes larger than a given threshold, $\rho^*_{\rm tot}$, spinodal decomposition takes place due to steric interactions between mRNAs and DNA, causing the nucleoid to split into two lobes. We thus expect $\rho^*_{\rm tot}$ to roughly correspond to the spinodal line of the phase diagram, but, given the non-equilibrium nature of the system mostly due to mRNA synthesis, it could differ from the equilibrium spinodal boundary. If we assume the threshold $\rho^*_{\rm tot}$ exactly corresponds to the spinodal threshold, then $\rho^*_{\rm tot} \simeq 2.5 \times 10^3\,\mu{\rm m}^{-1}$ (estimated in Section 2.3.1). However, we note that an mRNA concentration of $\sim 10^3\,\mu{\rm m}^{-1}$ to drive the splitting of the nucleoid is in agreement with the observed behavior of the full model (see Fig. 2.7) and, therefore, we can justify *a posteriori* the validity of the approximations.

Whatever value $\rho^*_{\rm tot}$ takes (provided its dependency on α is negligible), this simple model predicts a scaling for the critical length $L^*_{\rm nucl}$ at which the nucleoid starts to divide of the form $L^*_{\rm nucl} \propto (\alpha c_{\rm DNA})^{-1/2}$, obtained from equating the maximum of the mRNA concentration profile to a fixed value $\rho^*_{\rm tot}$. To test the prediction of this simple model, we numerically obtained the length at which the nucleoid divides for different values of α, see the inset in Fig. 2.4 B, and found a good agreement with the proposed scaling.

Nevertheless, our model is too simple to give a mechanistic explanation for nucleoid splitting: no connectivity between DNA segments or modelling of DNA replication is included in the description. However, overall, the model does show that mRNA synthesis can control the localisation of daughter nucleoids, once the nucleoid has been split in two.

2.4.2 Nucleoid centring

As observed in Ref. [Wu et al., 2019], a single bacterial nucleoid has a strong tendency to localise at midcell for all cell sizes. Following the recent suggestion that the central positioning of the nucleoid is regulated by an active process [Joyeux, 2019], we investigated whether the non-equilibrium process of mRNA production, diffusion, ribosome binding, and mRNA degradation can account for nucleoid centring.

We consider the case of a nucleoid that, due to a fluctuation, is not initially at the center of the cell, and test whether the non-equilibrium effects in our model can push the nucleoid back to the cell center. To model this, we use the steady-state profiles obtained for filamentous growth, and shift the concentration profiles towards the right cell pole. The resulting configuration has a nucleoid displaced from the center, and equal mRNA and ribosome concentrations on both sides of the nucleoid. This concentration profile is used as the initial condition for Eqs. (2.1) to (2.3), which we integrate forward in time in the presence of the non-equilibrium terms. As shown in Fig. 2.8, the nucleoid is centred at midcell after $\sim 30\,\mathrm{min}$.

The physical origin of this centring is mRNA synthesis within the nucleoid: The nascent mRNAs diffuse in the nucleoid until they reach one of its boundaries and then escape, with an equal flux to the left and right of the nucleoid. If the nucleoid is not centred, the accumulating mRNAs occupy a greater fraction of the available volume on one side of the nucleoid and thus create a higher osmotic pressure on that side. The resulting pressure difference ultimately drives the nucleoid back to the center of the cell.

The rate at which the nucleoid moves towards the cell center depends on both the pressure difference due to mRNA accumulation, and on the effective viscous drag experienced by the nucleoid as a whole. The effective viscous drag, γ, is given by $F_{\mathrm{nucl}} = \gamma\, v_{\mathrm{nucl}}$, where F_{nucl} is the force applied on the nucleoid (here, the osmotic-pressure difference times the cross-section of the cell), and v_{nucl} is the velocity of the nucleoid. If γ is low, the nucleoid responds fast to any force applied on it and thus quickly reaches a position of mechanical equilibrium, where the osmotic-pressure difference vanishes. In this case, the centring process is only limited by the speed at which mRNAs accumulate on either side of the nucleoid, which sets the pressure differences. By assuming this limit, we can to establish a lower bound for the time needed by the nucleoid to center.

The kinetics obtained in this low-drag limit are shown in Fig. 2.8B (grey line), and they are given by an exponential relaxation with timescale β^{-1}, set by the rate of mRNA degradation (see below). As shown in the figure, the nucleoid centring obtained from the full model lags behind the lower bound, showing that there is a non-negligible contribution from drag on the nucleoid. As shown in the inset, both the lower bound and the result from the full model show an exponential relaxation of the nucleoid position for early times in the centring process.

Analytical estimates for the centring dynamics in the low-drag limit

The centring of the nucleoid can also be understood in terms of the simplified compartment model we used to estimate the nucleoid size. However, in order to explain the centring dynamics, we need to modify the model and assume that the mRNA synthesised in the nucleoid can diffuse out of the nucleoid to the lateral compartments. If the nucleoid is not centred in the cell and the synthesised mRNAs leave the nucleoid symmetrically to the left and right, then the mRNA density, and thus the osmotic pressure, will increase in the smaller polar compartment, thus pushing the nucleoid towards the center. As a result, the force that we need to consider is the difference in pressure between the poles times the cross section σ of the cell $F = \sigma(P_{\mathrm{L}} - P_{\mathrm{R}})$, where 'L' and

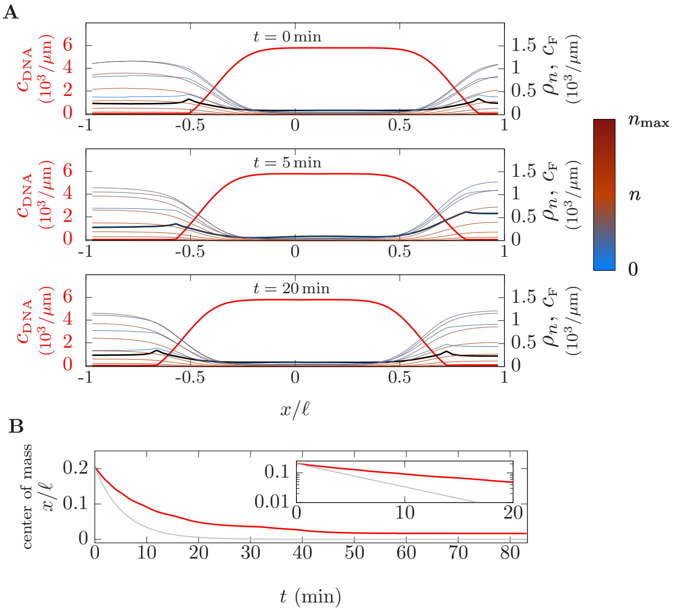


Figure 2.8: Non-equilibrium processes centre the nucleoid at midcell. (A) Concentration profiles obtained by initially shifting the steady-state profiles towards the right cell pole at $t = 0$, and then integrating forward in time the reaction-diffusion Eqs. (2.1) to (2.3) in the presence of non-equilibrium processes to $t = 5\,\mathrm{min}$ and $t = 20\,\mathrm{min}$, for a cell with half-length $\ell = 1.8\,\mu\mathrm{m}$, shown as in Fig. 2.4 A. (B) In red, location of the center of mass of the nucleoid along the long cell axis, as a function of time. In gray, the analytical lower bound obtained by neglecting nucleoid drag. Inset: The quantities depicted are the same as in B, but with the y-axis is in logarithmic scale.

'R' denote the left and right pole, respectively. The dynamical equation for the position of the center of mass of the nucleoid, which we denote by x_n, is:

$$\frac{dx_n}{dt} = \frac{D}{k_{\mathrm{B}}T}\sigma(P_{\mathrm{L}} - P_{\mathrm{R}}), \tag{2.50}$$

where D is a diffusion constant, not necessarily equal to the diffusion constant of DNA segments. In fact, D is an effective diffusion coefficient that includes collective effects of DNA segments diffusing together and potentially other biological effects.

We can provide a lower bound for the time it takes the nucleoid to center by assuming that the drag is small, i.e. the centring of the nucleoid due to a difference in osmotic pressure is only limited by the synthesis of mRNA. In this limit, the nucleoid moves fast enough to prevent a pressure difference between the poles, that is, a quasi-static approximation of Eq. (2.50): $d_t x_{\mathrm{n}} = 0$, which implies $P_{\mathrm{L}} = P_{\mathrm{R}}$. Thus, the centring of the nucleoid is controlled by the rate at which the number of polysomes in the lateral compartments change. In both compartments, the pressure is set by the concentration of polysomes, whose number is set by the following differential equation:

$$\frac{d\rho_{\mathrm{tot_i}}}{dt} = \frac{\alpha c_{\mathrm{DNA}}}{2} - \beta\rho_{\mathrm{tot_i}}, \tag{2.51}$$

where $\rho_{\mathrm{tot_i}}$ is the total mRNA concentration in compartment i and whose solution is

$$\rho_{\mathrm{tot_i}} = \frac{\alpha\, c_{\mathrm{DNA}}}{2\beta} + C_i e^{-\beta t}, \tag{2.52}$$

where C_i is a constant that is set by initial conditions. In the case of the initial condition of Fig. 2.8, C_i takes the value $C_{\mathrm{L}} \simeq +0.2\,\alpha c_{\mathrm{DNA}}/\beta$ and $C_{\mathrm{R}} \simeq -0.2\,\alpha\, c_{\mathrm{DNA}}/\beta$ for the left and right compartment, respectively, as the initial position of the centre of mass of the nucleoid is located at $+0.2/\ell$ and the amount of mRNA is directly proportional to the volume of each compartment. Since $P_{\mathrm{L}} = P_{\mathrm{R}}$ we have $\rho_{\mathrm{tot\,L}}/V_{\mathrm{L}} = \rho_{\mathrm{tot\,R}}/V_{\mathrm{R}}$. Assuming that the nucleoid does not change size during this process, we obtain

$$V_{\mathrm{L}} = \frac{\rho_{\mathrm{tot\,L}}}{\rho_{\mathrm{tot\,L}} + \rho_{\mathrm{tot\,R}}}(V - V_{\mathrm{n}}), \tag{2.53}$$

where V is the total volume of the cell, and V_n the volume of the nucleoid. In the previous relation, the only term that is time-dependent is $\rho_{\mathrm{tot\,L}}$ since $\rho_{\mathrm{tot\,L}} + \rho_{\mathrm{tot\,R}}$ is constant in time. Therefore, the position of the nucleoid is set by V_L, which depends only on $\rho_{\mathrm{tot\,L}}$, which yields the exponential relaxation with timescale β^{-1} mentioned above. This lower bound on the time for centering is depicted in Fig. 2.8 B (grey line).

Validity of the one-dimensional model

In our analysis, we leveraged the cylindrical symmetry of the *E. coli* cell to reduce the number of dimensions of the model to one, which, in the description of the centring dynamics of the nucleoid, is crucial. The real system is three-dimensional and, as opposed to what has been assumed before, it need not be homogeneous in the radial direction. This becomes particularly important at midcell, where the nucleoid may not occupy the whole cellular cross section, potentially allowing the diffusion of polysomes and ribosomes around the nucleoid: such diffusion could have an important effect on the centring dynamics, as it would result in a decrease of the osmotic-pressure difference between the poles and, therefore, a decrease of the centring force. Actually, it would mean that the osmotic pressure difference can vanish without requiring the nucleoid to move, thus disrupting the centring mechanism proposed above. However, the magnitude of this effect depends on how fast polysomes can diffuse around the nucleoid. If such diffusion around the nucleoid is slow (due to, e.g., obstacles in the narrow channel between the membrane and the nucleoid), the osmotic-pressure difference, and thus the centring force, will be sustained.

We now proceed to estimate the diffusive flux around the nucleoid in the presence of a concentration gradient and assess the effect of this flux in the centring dynamics. We assume that the nucleoid is a cylinder of length L_nucl and radius $R - L/2$, where R is the cell radius and L is the length of a DNA segment (we estimate the depletion zone between the membrane and the nucleoid to have a width equal to the persistence length of a DNA plectoneme as this is the lengthscale of the DNA-membrane interaction). The cross section available for diffusion is $\sigma_d = \pi[R^2 - (R - L/2)^2] \simeq \pi R L$ (neglecting terms of order L^2). If we assume that the concentration of mRNAs at one pole is c and that it is zero at the opposite pole, then the concentration gradient is c/L_nucl, yielding for the total flux across J the available cross section

$$J \simeq \frac{D_n c}{L_\mathrm{nucl}} \pi R L, \tag{2.54}$$

where D_n is the diffusion coefficient of mRNAs and polysomes.

Then, the characteristic timescale τ over which these concentration gradients will disappear due to diffusion is

$$\tau \sim \frac{c V_\mathrm{p}}{J} = \frac{R L_\mathrm{p} L_\mathrm{nucl}}{D_n L}, \tag{2.55}$$

where L_p is the length of one of the polar regions ($2\ell = 2L_\mathrm{p} + L_\mathrm{nucl}$) and $V_\mathrm{p} = \pi R^2 L_\mathrm{p}$ is its volume. By substituting the values of these parameters given in Section 2.2.4, and assuming that for our reference cell $L_\mathrm{nucl} \simeq 1\,\mu$m and $L_p \simeq 0.5\,\mu$m, we obtain $\tau \sim 20\,$s.

The value obtained for the diffusion timescale τ is lower than the typical timescale on which synthesis and degradation of mRNAs (β^{-1}) would set the osmotic-pressure difference at the poles. Therefore, most of the osmotic-pressure difference would be removed by diffusion, and the resulting centering force exerted on the nucleoid would be small. However, when estimating τ, we assumed that polysomes diffusing through the narrow channel between the nucleoid and the membrane have same diffusion coefficient as for polysomes diffusing at the poles. This assumption is likely to be inaccurate, because in the channel there are more crowders or obstacles than at the poles, including transertion effects – that is, the simultaneous transcription, translation and insertion of proteins into the membrane that tethers the nucleoid to the membrane [Gorle et al., 2017]. Hence, the validity of the one-dimensional model depends on how fast diffusion of mRNAs and polysomes occurs through the narrow channel between the nucleoid and the membrane. For sufficiently slow diffusion in this region and a sufficiently narrow channel, the corrections to the one-dimensional model (stemming from the nucleoid not occupying the whole cross section) are negligible.

2.4.3 Perturbing the nucleoid

Until now we have analysed how the model behaves naturally, in the absence of stresses or perturbations. However, we can also probe how the model would respond to certain types of perturbations and compare with experimental data. Usually, antibiotics are utilised to modify certain essential functions in *E. coli* and observe the changes in the size and morphology of the nucleoid. Examples of these perturbations are a halt of transcription with rifampicin and an arrest in translation by kasugamycin. Both scenarios are analysed below.

Halt in transcription

It has been shown experimentally that when transcription in *E. coli* is halted, e.g. by treatment with rifampicin, the nucleoid expands [Cabrera et al., 2009, Bakshi et al., 2012]. From a point of view where macromolecular crowding is critical for nucleoid compaction, the halt of mRNA synthesis depletes polysomes, and thus results in a lower osmotic pressure on the nucleoid, allowing its expansion. We tested this scenario with our model by using the non-equilibrium steady state shown in Fig. 2.4 A as the initial condition for Eqs. (2.1) to (2.3), switching off mRNA synthesis ($\alpha = 0$), and integrating forward in time. As shown in Fig. 2.9, the nucleoid expands and spreads over most of the intracellular space. The nucleoid does not take over the entire cell because there are pockets of free ribosomes at both cell poles, which prevent the DNA from occupying these spaces.

The nucleoid reaches its expanded steady state in $\sim 30\,\text{min}$, which is in good agreement with experimental data [Cabrera et al., 2009], see Fig. 2.9 B, despite the fact that parameters like total cell length or growth rate used in our model may not match the (unreported) experimental ones. The agreement is due to the fact that the expansion process is primarily driven by the degradation of mRNAs, hence β is the most relevant parameter for this process. In Fig. 2.9 B the bulk of the expansion happens in the first 10 min – a timescale consistent with the half-life of mRNA (5 min) and with the experimental data. Therefore, the most relevant parameter for the expansion process is the mRNA degradation rate and, given the quantitatively good behaviour of the model, we conjecture that in the experiments by Cabrera and colleagues this parameter takes

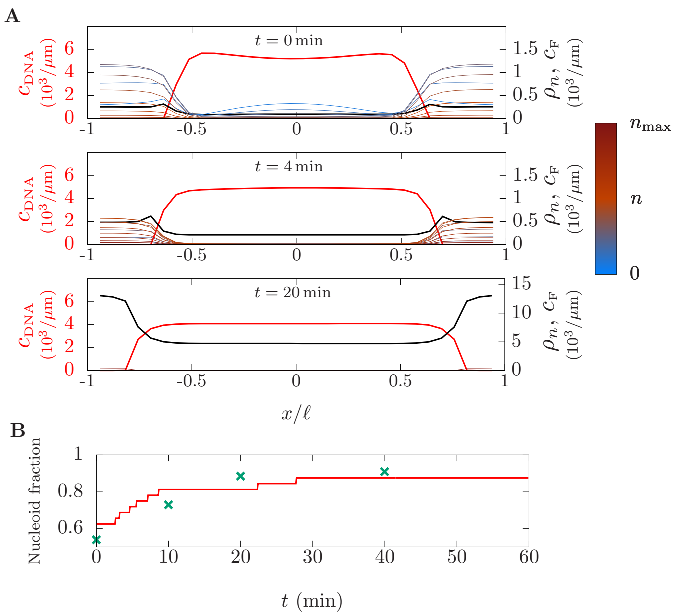


Figure 2.9: The nucleoid expands in the absence of mRNA synthesis. (A) Steady-state profile including mRNA synthesis ($t = 0$) and profiles obtained by integrating forward in time from the steady-state profile at $t = 0$ in the absence of mRNA synthesis ($t = 4\,$mins and $t = 20\,$mins), for a cell with half length $\ell = 3.6\,\mu$m. The concentration profiles are shown as in Fig. 2.4 A. (B) Fraction of the cell volume occupied by the nucleoid, as a function of time (computed as the fraction of length along the axis of the cell with a DNA segment concentration $c_{\mathrm{DNA}} > 1000\ \mu\mathrm{m}^{-1}$). The turquoise data points were obtained from Cabrera et al. [Cabrera et al., 2009], for the case of a cell treated with rifampicin (which blocks mRNA transcription).

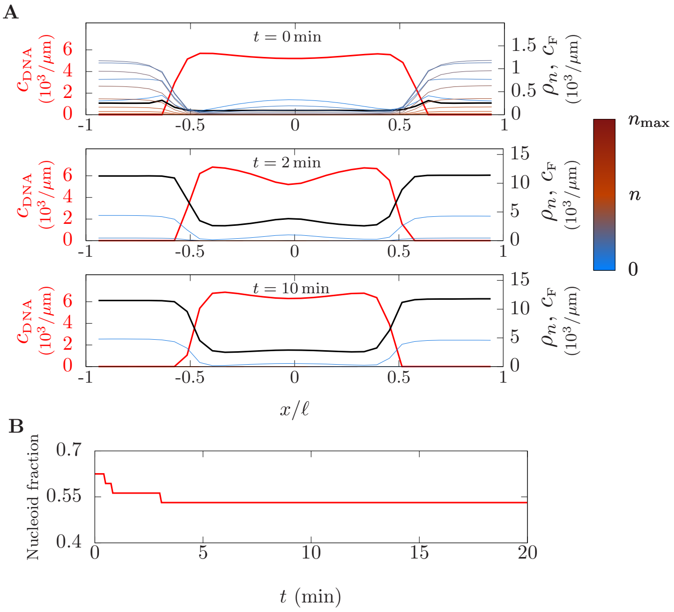
a similar value to the one used here (which was obtained from Ref. [Bernstein et al., 2004]).

Halt in translation

It has been shown that, after inhibition of translation with antibiotics such as kasugamycin or chloramphenicol, the nucleoid contracts [Bakshi et al., 2014]. We therefore tested our model to see if it can explain the contraction of the nucleoid after treatment with kasugamycin. Kasugamycin is known to reduce the abundance of polysomes by inhibiting translation initiation and to substantially compact the nucleoid [Xiang et al., 2021]. Hence, to reproduce the effect of kasugamycin in the model we set the rate of ribosome binding to polysomes to zero ($k_{\mathrm{on}} = 0$) and integrate foward in time from the non-equilibrium steady state shown in Fig. 2.4 A. The results are reported in Fig. 2.10. There is a significant contraction of the nucleoid after $\sim 2\,$min, which is consistent with the timescale on which ribosomes unbind from polysomes $k_{\mathrm{off}}^{-1} = 40\,$s. Nevertheless, the contraction of the nucleoid in the model is smaller than the one reported in Ref. [Xiang et al., 2021], potentially reflecting the effect of transertion on nucleoid size [Bakshi et al., 2014], which is not taken into account in our model. However, the fact that the model predicts a contraction of the nucleoid if ribosomes stop binding to mRNAs suggests that the steric interaction of bare mRNAs with the DNA is enough to compact the nucleoid and that polysomes are not required for nucleoid compaction, which helps explain recent observations [Xiang et al., 2021].

It was also noted that treatment with kasugamicin seems to destabilise the positioning mechanism of daughter nucleoids and cause their coalescence [Xiang et al., 2021]. Given that treatment with kasugamicin decreases the width of the nucleoid [Bakshi et al., 2014], this will increase the diffusion of crowders around the nucleoid, implying that our one-dimensional model may not be valid anymore, in accordance with the findings of the previous section. Therefore, the positioning mechanisms suggested in this chapter may be less effective after treatment with kasugamicin and could eventually cause the coalescence of two daughter nucleoids. As argued below, further

Figure 2.10: The nucleoid contracts in the absence of translation initiation. (A) Steady-state profile including translation initiation ($t = 0$) and profiles obtained by integrating forward in time from the steady-state profile at $t = 0$ in the absence of ribosome binding ($k_{\mathrm{on}} = 0$ for panels $t = 2$ mins and $t = 10$ mins). (B) Time evolution of the fraction of the cell volume occupied by the nucleoid, shown as in Fig. 2.9B.



research should include a more comprehensive description of the radial degrees of freedom in the spatial organisation of *E. coli* to better understand this kind of perturbations and the general mechanisms of self-organisation.

2.5 Discussion of the results

In this chapter, we investigated the physical origins of the intracellular localisation of DNA, messenger RNAs (mRNAs), and ribosomes in bacteria. This is a topic of general interest due to its far-reaching consequences, such as, the spatial organisation of transcription and translation [Gray et al., 2019, Weng et al., 2019], chromosome positioning and segregation [Wu et al., 2019, Joshi et al., 2011], and a wide range of cellular processes regulated by the nucleoid that excludes many macromolecules from the volume which it occupies [Coquel et al., 2013].

Based on steric interactions among DNA, mRNAs and ribosomes, we developed a holistic model for the spatial organisation of the bacterial nucleoid, which is intuitive and can quantitatively describe the behaviour of the nucleoid and account for experimentally tested perturbations. The model predicts the formation of a phase-separated nucleoid, whose size is in agreement with experimental measurements [Wu et al., 2019] for cells smaller than 10 μm (Fig. 2.5). Beyond this cell length, our model is no longer accurate, for reasons that may include the lack of connectivity among modeled DNA segments, uncertainties in the concentration of crowders, and molecular components which have not been explicitly incorporated in the model, such as nucleoid-associated proteins [Dame et al., 2020] or topoisomerases [Stuger et al., 2002] (see Section 2.1).

Moreover, the model highlights the importance of transcription and translation in regulating the size of the nucleoid. On the one hand, it can account for nucleoid expansion as a result of a halt in mRNA synthesis, demonstrating that the progressive degradation of crowders could be the physical cause of the expansion. Indeed, the timescales on which such expansion happens matches the ones observed experimentally [Cabrera et al., 2009], and coincides with the timescales of mRNA turnover. On the other hand, halting translation (by precluding ribosomes from binding to polysomes or mRNAs) further compacts the nucleoid because bare mRNAs are still large enough to sterically compact the nucleoid and the number of free ribosomes is increased.

Our results also underline the role of non-equilibrium effects in the regulation of nucleoid position. The nucleoid is known to localise at midcell [Wu et al., 2019], and we demonstrate that the synthesis of mRNAs and their expulsion from the nucleoid caused by steric effects can give rise to this positioning – see Fig. 2.8. In fact, a perturbation from the central position of the nucleoid induces an osmotic-pressure difference between the two cell poles, which pushes the nucleoid back to midcell. The timescale for this centring depends on both the time it takes to establish an osmotic-pressure difference, which is set by the mRNA turnover time, and the drag experienced by the nucleoid. This drag may be underestimated in our model, because we do not include effects that could slow down nucleoid centering, e.g. the transient attachment of the nucleoid to the membrane by proteins that are simultaneously being transcribed, translated, and inserted into the membrane, also known as transertion [Gorle et al., 2017]. Furthermore, our model shows that non-equilibrium effects are responsible for the ubiquitous nucleoid splitting and localisation at 1/4 and 3/4 positions along the long cell axis. Indeed, the synthesis of mRNAs within the nucleoid, without additional active processes, is a robust mechanism to make the daughter nucleoids localise at 1/4 and 3/4 positions, as observed experimentally [Wu et al., 2019].

Turing patterns at the cellular scale [Wu et al., 2016] also display features, such as, the emergence of a characteristic lengthscale, that could seem similar to the ones produced by our model, in view of Fig. 2.11. These non-equilibrium patterns have been used to investigate many biological features on a cellular and sub-cellular scale, such as the positioning of protein clusters in *E. coli* [Murray and Sourjik, 2017]. However, unlike Turing patterns, our model predicts a phase separation due exclusively to steric interactions, and in the absence of non-equilibrium effects, see Fig. 2.7 A, upper pannel. The patterns reported here are more closely related to the ones produced by other models with an interplay between an equilibrium free energy and non-equilibrium dynamics, e.g., the formation of FtsZ rings in bacteria [Shlomovitz and Gov, 2009], or other non-equilibrium phase-separation models [Li and Cates, 2020], such as models of growing droplets [Zwicker et al., 2017]. Nevertheless, our model provides a conceptually simpler framework to produce these patterns. In fact, unlike a model of physically growing droplets, our analysis involves a conserved order parameter – the total number of DNA segments – and the effect of non-equilibrium terms – mRNA production and degradation – is limited to nucleoid reshaping and repositioning. Despite its simplicity, our model produces a number of experimentally observed patterning effects, such as nucleoid centering at midcell, splitting and positioning of sister lobes during cell division.

In addition, this patterning is not limited to nucleoid splitting into two sister lobes, because our model predicts that the nucleoid can split into more than two lobes, whose size is given by a characteristic length and whose position is tightly controlled, as can be seen in Fig. 2.11. Experimentally, in long filamentously growing cells (where cell division is inhibited but DNA replication is allowed to continue), nucleoids are observed at tightly controlled positions and distances [Wehrens et al., 2018], forming a very similar pattern to the one obtained here, albeit with a shorter characteristic length. In Ref. [Wehrens et al., 2018] a separate nucleoid appears every

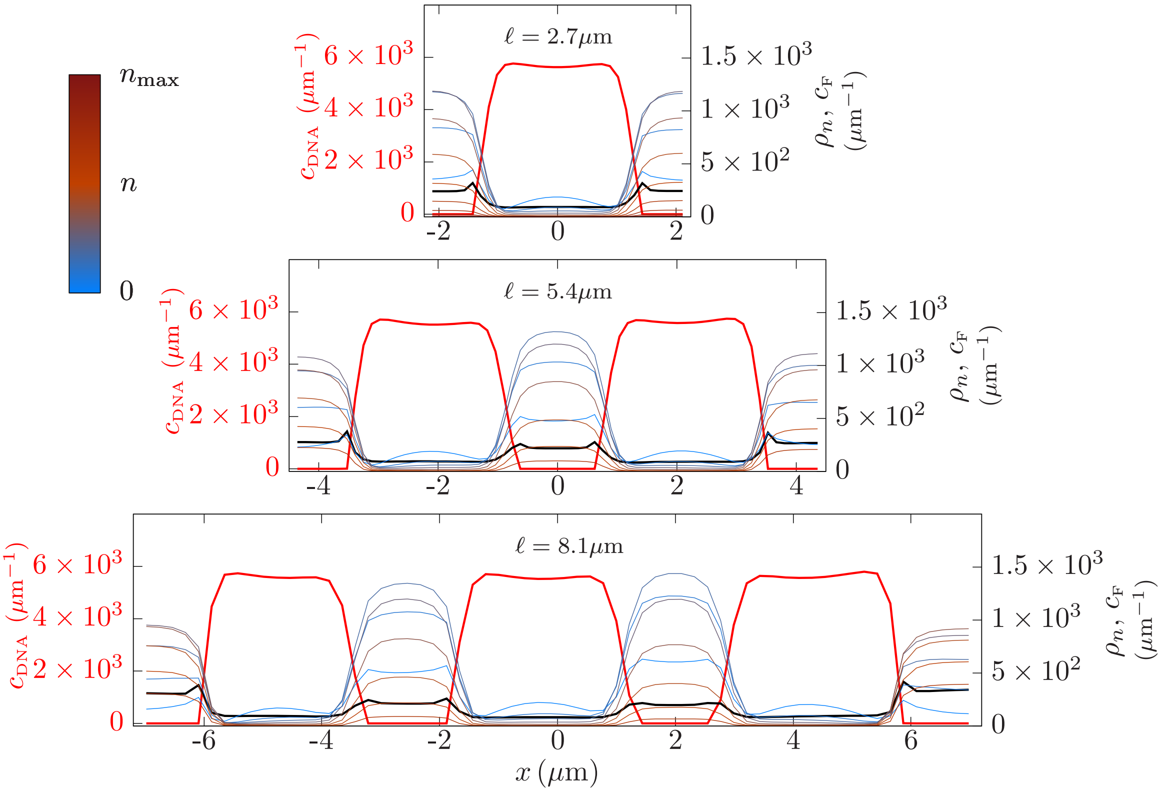


Figure 2.11: Steady-state profiles for filamentously growing cells. Steady-state profiles of the concentrations of the components of the *E. coli* transcriptional-translational machinery (TTM) for filamentously growing cells with mRNA concentration $\rho_{\rm tot} = 2400\,\mu{\rm m}^{-1}$. Colours represent different chemical species of the TTM, as in Fig. 2.4 A, and each of the panels corresponds to a different cell length (as indicated within each panel). To enable comparison between different panels, the lengthscale in all panels was kept constant. As in Fig. 2.7, for cells larger than $\sim 8\,\mu$m the nucleoid splits in two lobes. Furthermore, for cells around $\sim 16\,\mu$m the nucleoid has three distinct lobes, suggesting that this may be a pattern with a characteristic length that exists also for longer cells.

$\sim 2.25\,\mu$m the cell grows in length. This value is roughly half of the one predicted by this model, possibly due to uncertainties in the parameters such as the diffusion coefficient of mRNAs within the nucleoid.

Given its generality, our analysis is not necessarily restricted to the nucleoid of prokaryotic cells [Cohan and Pappu, 2020]. For instance, division of certain phase-separated condensates has been experimentally related to non-equilibrium processes, as is the case for the ParABS partition system, which creates phase-separated condensates of DNA and ParB around *par*S sites, whose division is controlled by the activity of ParB's ATPase activity on ParA [Guilhas et al., 2020]. The activity-driven nucleoid division described in our model may thus constitute a general strategy employed by cells to control the structure and positioning of membraneless compartments.

Future directions: Extensions and experimental tests

Our work implies that steric interactions make the bacterial cytoplasm an effectively poor solvent for the chromosome, as recently indicated by experiments [Xiang et al., 2021]. However, steric interactions may not be the only contribution to the poor-solvent quality of the cytoplasm. Other types of intermolecular interactions [Odijk, 1998] or the effect of nucleoid-associated proteins [Dame et al., 2020] could also affect the solvent quality of the cytoplasm and, therefore, the organisation of the nucleoid in the cell. Notably, certain proteins or enzymes, such as Dps [Janissen et al., 2018] and DNA gyrase [Stuger et al., 2002], have been shown to modify the compaction of the nucleoid. Given that the radii of the steric interaction might be an effective one, phenomenologically, one could account for perturbations where these molecules are involved by varying the parameter ρ' (see Section 2.2.1) and reduce it to account for their presence. Still, for future studies, both theoretical and experimental, research into these other regulators of the nucleoid size could yield a more complete picture of its organisation, and improve the accuracy of the results presented here.

Another of the limitations of this study is the reduction from the three dimensions of an *E. coli* cell to the single one considered here. In reducing the system to one dimension, we tacitly assumed that the nucleoid occupies the whole cross-section. While normally the nucleoid does spread over most of the cross-section, upon treatment by kasugamycin and chloramphenicol the width of the nucleoid decreases [Bakshi et al., 2014] with two important consequences: Since the nucleoid does not occupy the whole cross-section, the osmotic pressures in different regions of the cells can be equilibrated without the need for lateral displacement of the nucleoid, which removes the mechanism that positions the nucleoid at midcell (or 1/4 and 3/4 for daughter nucleoids). In addition, if the nucleoid does not occupy the whole cross-section of the cell, mRNAs might be able to escape radially from the nucleoid, which can induce a deviation from the scaling found in Eq. (2.49) for the length at which the nucleoid splits. Indeed, previously separated daughter nucleoids have been observed to coalesce upon treatment with kasugamycin and chloramphenicol [Schaechter and Laing, 1961, Xiang et al., 2021] and we hypothesise that the decrease in nucleoid width induced by these drugs destabilises the positioning of daughter nucleoids. Then, if the nucleoids come into contact due to fluctuations, they are driven to coalesce by short-range depletion interactions due to ribosomes, mRNAs, and any remaining polysomes. Therefore, we expect that exploring the three-dimensional dynamics of the system might explain several phenomena that for the moment remain unaccounted for.

Moreover, the present work has been devoted to the physical principles governing the global organisation and positioning of the bacterial nucleoid. However, the nucleoid also has a rich internal dynamics [Fisher et al., 2013]. Given that the DNA plectoneme persistence length is close to the cell radius, future studies may investigate whether orientational order emerges, and whether it has implications for intra-nucleoid organisation.

As outlined above, our model (like every model) has certain limitations. Therefore, an important part of building a theory is knowing when it will collapse and will be no longer accurate or able to explain the observations. The grounds for this testing must also be laid.

In the case of the present theory for the organisation of the bacterial nucleoid, the one-dimensional assumption, which turns out to be crucial for nucleoid positioning, is one its limits, as mentioned above. Yet, its consequences are not restricted to the coalescence of daughter nucleoids, as a substantial decrease in the radius of the nucleoid should enable the diffusion of macromolecules around it, destabilising too the centring mechanism. Therefore, if the nucleoid shrinks in the radial dimension, while the average position of the nucleoid may still be the centre[4], there should be larger fluctuations in the position of the nucleoid. These should be measurable by the second moment of the distribution of the nucleoid position, if measurements of enough resolution can be made for a large enough number of cells and long enough times, and treatment by kasugamycin (or erythromycin [Sanamrad et al., 2014]) should result on an increase of the variance of the position.

However, treating the cells with kasugamycin induces a large perturbation in the metabolism of a cell, since not only does it make the nucleoid shrink but it also halts translation, with the widespread consequences this may carry. Thus, alternative approaches that may yield similar outcomes should also be envisaged. In particular, quantitative measurements of the nucleoid in stationary phase *E. coli* could be a good candidate.

When nutrients are scarce and cannot sustain exponential growth, *E. coli* cells enter stationary phase and stop reproducing [Nyström, 2004, Santos et al., 2005]. Upon cessation of growth, the synthesis of Dps is induced [Almiron et al., 1992], which compacts the nucleoid at least in the longitudinal direction [Janissen et al., 2018]. If a radial contraction of the nucleoid was also measured, then, according to our theory, there should be a destabilising effect in the central location of the nucleoid in stationary phase (larger dispersion of the measured positions even if the average is still at midcell). Moreover, stationary phase cells are often associated with a slower transcriptomic activity, which should further amplify the instability of nucleoid location.

Therefore, quantitative assessment of the stationary-phase *E. coli* nucleoid compaction and positioning could also be used to compare with the theory exposed here and explore its limitations. However, care should be taken when extrapolating the results obtained here to the stationary-phase nucleoid, since the highly expressed Dps protein would increase the attractive interactions between DNA segments (in a phenomenon akin to other protein-DNA co-condensations [Quail et al., 2021]) and, eventually, it may outcompete repulsive interactions, thus invalidating the picture based on excluded-volume interactions proposed here [Andersen et al., 1971].

---

[4] Even a completely random positioning within a cell would yield midcell as the average position.

Hence, further work may be needed before applying this framework to certain scenarios like stationary-phase *E. coli*.

PART II:

A THEORETICAL FRAMEWORK FOR NON-IDEAL CHEMICAL REACTION NETWORKS

> *Most of an organism, most of the time, is developing from one pattern into another, rather than from homogeneity into a pattern. One would like to be able to follow this more general process mathematically also. The difficulties are, however, such that one cannot hope to have any very embracing* theory *of such processes, beyond the statement of the equation.*
>
> A. M. Turing,
> *The chemical basis of morphogenesis*, 1952

The examples exposed in the first part of this thesis highlight the importance of interactions between solutes in the cell cytoplasm. These interactions give rise to emergent behaviour like phase-separated organelles (e.g. the nucleoid in certain bacterial cells, see Chapter 2) or control the size of protein aggregates, as found in the first chapter. Nonetheless, Part I was devoted to two particular examples and one would like to have a theory as general as possible for a wide variety of processes, to increase our understanding of them and be able to distinguish between the generic properties of these mechanisms and the details of each one of them.

An umbrella under which many of these processes can be described is that of reaction-diffusion equations. More precisely, we will consider a network of chemical reactions that is driven out of equilibrium to represent the reactions that may happen in the cytoplasm of a cell. In addition, diffusion must be taken into account if we aim at explaining the spatial organisation within a cell. However, unlike classical reaction-diffusion systems, which are assumed to be ideal (i.e. solutions with non-interacting solutes), we need to consider interactions between the components of the solution to better account for the complex dynamics of the crowded cytoplasm and, thus, build more realistic models that describe better the physical reality of a cell. Altogether, the framework we aim at building may be termed non-ideal reaction-diffusion systems.

Beyond the examples of Part I, it is becoming increasingly clear that, in general, the cytoplasm of a cell does not behave like an ideal solution [Zielinski et al., 2017]. There is a plethora of interactions between the solutes that take place within the cytoplasm, mostly involving proteins and other macromolecules, but also ions. Some of the most common interactions that are relevant in the cellular cytoplasm are steric or crowding effects – see Part I and Refs. [Zhou et al., 2008, Mitchison, 2019] – and electrostatic interactions (e.g. [Fall and Keizer, 2001, Wang et al., 2018]). Arguably, the most striking phenomenon caused by these interactions is the emergence of phase-separated condensates within the cytoplasm, also known as membraneless organelles particularly in the cell biology literature, which are now widely studied [Li et al., 2012, Su et al., 2016, Brangwynne et al., 2009]. The composition of these membraneless organelles is different to that of the cytoplasm, as they are typically enriched in a certain type of molecules while they exclude others [Ditlev et al., 2018]. Moreover, it has been hypothesised that they spatially control biochemical reactions [Banani et al., 2017], by modulating their rates and specificity within the condensate. In this sense, the nucleoid described in Chapter 2 shares all of the properties described here for membraneless organelles, since it controls the spatial distribution of transcription and translation, and it excludes large macromolecules from its volume.

Therefore, phase separation phenomena play an important role in the internal spatial organisation of cells and their regulation is crucial for many cellular functions. One of the ways cells can dynamically control the appearance of membraneless organelles (and its composition or function) is through chemical reactions, notably post-translational modifications like phosphorylation [Tsang et al., 2019, Kim et al., 2019] or methylation [Nott et al., 2015]. However, phase separation is also triggered by changes in the environment [Franzmann et al., 2018, Lyon et al., 2021], establishing them as potential sensing and regulatory mechanisms.

From the theoretical perspective, the interplay between interactions within the solution and non-equilibrium chemical reactions has also been widely studied. However, most of these efforts [Huberman, 1976, Glotzer et al., 1995, Wurtz and Lee, 2018, Li and Cates, 2020] were based on effective reaction-diffusion models that, while they can describe patterning and other non-equilibrium phenomena in a simple way, they lack thermodynamic consistency, as they treat chemical reactions as ideal – by modelling their dynamics with mass-action kinetics (MAK) – while the interaction-influenced diffusion that drives phase-separation is not, and there is no explicit non-equilibrium driving. Conversely, a *thermodynamically consistent* description would require that, in the same way in which diffusion is governed by a free energy (that takes into account the interactions), the dynamics of chemical reactions must also reflect this free-energetic dependency. Early progress at addressing non-ideal reaction-diffusion systems in a thermodynamically consistent model was limited to a linear stability analysis for the dynamics of binary solutions [Carati and Lefever, 1997]. More recently, some works have aimed at establishing a deterministic theory for non-ideal chemical reaction networks (CRNs) [Avanzini et al., 2021] and exploring minimal examples for pattern formation [Bazant, 2013, Kirschbaum and Zwicker, 2021] but the link between non-equilibrium CRNs and phase-separation has not yet been elucidated in full generality.

In this part, we aim at building a thermodynamically consistent framework for interacting reaction-diffusion systems that, thus, may exhibit phase-separation at steady state. Therefore, a complete theory of non-ideal CRNs is necessary and here previous efforts are complemented by analysing the behaviour of non-ideal CRNs in the stochastic limit and exploring the consequences of the topology of the network. We do so by first constructing a framework which satisfies that, in the absence of explicit non-equilibrium driving, the system relaxes to thermodynamical equilibrium (Chapter 3). This enables us to naturally adapt and generalise results from the well-established theory of ideal CRNs, explore the connection between non-equilibrium CRNs and phase separation, and obtain Lyapunov functionals for a particular class of CRNs, known as complex-balanced (Chapter 4).

A. M. Turing already stated the difficulty of building a *general* theory for patterning processes and, here, we will fail in doing so, as the applicability of most of our results is limited to complex-balanced networks. However, important consequences can be derived from these types of networks, see Chapter 5, where we also discuss on the implications of considering more general networks.

Chapter 3

Introduction to the description and thermodynamics of Chemical Reaction Networks

We will now take the approach of the *formal kineticist*[1] and define the chemically reacting systems in terms of *elementary reactions*. An elementary reaction is a chemical reaction with no hidden intermediate steps [Kondepundi and Prigogine, 2015], that is, the most basic unit of a chemical transformation[2], which is defined by a set of stoichiometric coefficients and a rule relating reaction rates to composition (as well as other physical variables), which for ideal solutions usually is the law of mass action kinetics [Horn and Jackson, 1972]. But, most importantly, the driving force of elementary reactions (whether in or out of thermodynamic equilibrium) is their free energies of reaction or affinity (no matter how complex these free energies may be), provided a sort of local equilibrium can be defined [Kondepundi and Prigogine, 2015].

Therefore, the thermodynamically consistent approach to chemical reaction networks (CRNs) in non-ideal solutions must be built upon the concept of elementary reactions because it is only by taking into account every process happening in the system that the laws of themodynamics can be recovered for CRNs, even when these are driven out of thermodynamic equilibrium. Once the dynamics of a CRN have been written down in a thermodynamically consistent way, we will be able to properly define thermodynamic quantities, such as entropy production or work; and extend to non-ideal solutions results that are well-known for ideal CRNs.

In order to construct a thermodynamically consistent framework for non-ideal CRNs we will impose two conditions:

- A CRN must relax to thermodynamic equilibrium in the absence of external work.

---

[1] "The formal kineticist, on the other hand, takes a macroscopic viewpoint and his primitive concept is the elementary reaction." F. Horn and R. Jackson, *General Mass Action Kinetics*, 1972.

[2] Accordig to the IUPAC's Gold Book, an elementary reaction is "a reaction for which no reaction intermediates have been detected or need to be postulated in order to describe the chemical reaction on a molecular scale. An elementary reaction is assumed to occur in a single step and to pass through a single transition state." In practice, however, the timescales of relaxation of any additional transition states and any intermediate metastable states must be taken into account for a practical but meaningful definition of elementary reaction.

- The external work done on the CRN must be specified explicitly.

Here we will restrict ourselves to CRNs where a subset of chemical species is chemostatted, that is, kept at a certain chemical potential as a consequence of external work. However, more complex procedures can be devised to keep a CRN out of thermodynamic equilibrium.

Finally, in the rest of this work, we will use the word equilibrium to refer to thermodynamic equilibrium. Conversely, we will call steady state the state of a system with vanishing time derivatives[3].

3.1 Description of Chemical Reaction Networks

We start by describing a CRN comprising N different chemical species and M different reversible reaction pathways. The requirement that every reaction is reversible is a requirement for the thermodynamic consistency of our description of the CRN if all of the reaction affinities are finite. In full generality, a reaction ρ belonging to the CRN can be specified as follows:

$$\sum_i r_i^\rho X_i \leftrightharpoons \sum_i s_i^\rho X_i, \tag{3.1}$$

where X_i denotes the different species labeled with the subindex i. In the rest of the document the indices i and j will be reserved for chemical species. The matrix r_i^ρ denotes the stoichiometry of the forward reaction (specifies number of reactants of type i required by the reaction ρ) and s_i^ρ that of the backwards reaction (products of type i produced by the reaction ρ). Note that, given that the reactions are taken to be reversible, the distinction between reactants and products is arbitrary. Therefore, $v_i^\rho = s_i^\rho - r_i^\rho$ is the net amount of i species created or destroyed along the forward direction of the reaction ρ (denoted as $+\rho$). We also define the vectors $\mathbf{v}^\rho = (v_1^\rho, \cdots, v_N^\rho)$, $\mathbf{r}^\rho = (r_1^\rho, \cdots, r_N^\rho)$ and $\mathbf{s}^\rho = (s_1^\rho, \cdots, s_N^\rho)$ for compactness, which are vectors carrying the information of the type and number of species involved in ρ as net change, reactants or products, respectively. Even more compact, we define the matrices $\mathbf{V} = (\mathbf{v}^1, \cdots, \mathbf{v}^M)$, $\mathbf{R} = (\mathbf{r}^1, \cdots, \mathbf{r}^M)$ and $\mathbf{S} = (\mathbf{s}^1, \cdots, \mathbf{s}^M)$, whose elements are v_α^ρ, r_α^ρ and s_α^ρ, respectively. We finally define the notion of complex $\mathbf{z}$, which is the set (number and type) of particles that take part in a chemical reaction as reactants ($\mathbf{z} = \mathbf{r}^\rho$) or products ($\mathbf{z} = \mathbf{s}^\rho$). A single complex $\mathbf{z}$ may appear in more than one reaction within the network.

Example 1 *Let us consider the following CRN:*

$$\begin{aligned} A + B &\leftrightharpoons C, \\ B &\leftrightharpoons D, \end{aligned} \tag{3.2}$$

whose stoichiometric matrices for the forward and backward reactions are

$$\boldsymbol{R} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \boldsymbol{S} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.3}$$

---

[3] This distinction has been made to avoid confusion with the jargon used in the mathematical literature where an equilibrium point in a dynamical system refers to a point in the system with vanishing time derivatives, independently of any thermodynamic considerations.

where each column denotes a particular species and each row a particular reaction of our CRN. Each of the rows in these two matrices corresponds to the vectors $\mathbf{r}^\rho$ *(for the matrix* $\mathbf{R}$*) and* $\mathbf{S}^\rho$ *(for the matrix* $\mathbf{s}$*). There are, therefore, four complexes in this network, two reactant complexes and two product complexes.* □

3.1.1 Stochastic Chemical Master Equation

If the solutes in a dilute[4] solution diffuse fast (with respect to the typical timescale of chemical reactions) the system may be viewed as well-mixed and can, therefore, be described in terms of a single homogeneous concentration of each of the species across the entire system. In this case a state of the system (number and type of particles) is completely determined by the vector $\mathbf{n} = (n_1, \cdots, n_N)$ where n_i is the number of particles of type i. CRNs where the copy numbers of certain species are low can show large fluctuations and high stochasticity. To describe a stochastic system, each of the states of the system $\mathbf{n}$ will have an associated probability measure $P(\mathbf{n}, t)$ at each instant of time t. The dynamics for the probabilities of states of homogeneous CRNs is given by the Chemical Master Equation (CME) which we write explicitly taking into account the reversibility of all reactions [Gillespie, 1992]:

$$\frac{\partial P(\mathbf{n},t)}{\partial t} = \sum_\rho f_{+\rho}(\mathbf{n}-\mathbf{v}^\rho)P(\mathbf{n}-\mathbf{v}^\rho) + \sum_\rho f_{-\rho}(\mathbf{n}+\mathbf{v}^\rho)P(\mathbf{n}+\mathbf{v}^\rho) - \sum_\rho [f_{+\rho}(\mathbf{n}) + f_{-\rho}(\mathbf{n})]P(\mathbf{n}), \tag{3.4}$$

where the summations over ρ run over all the reactions in the CRN and the rate of the transitions in the network is given by the propensity function f_ρ. In ideal dilute solutions and for elementary reactions, the propensity function f_ρ takes the form

$$f_{+\rho}(\mathbf{n}) = k_{+\rho} \prod_i \frac{n_i!}{(n_i - r_i^\rho)!}, \text{ and } f_{-\rho}(\mathbf{n}) = k_{-\rho} \prod_i \frac{n_i!}{(n_i - s_i^\rho)!}, \tag{3.5}$$

which is known as mass-action kinetics (MAK) for the stochastic CME. In the following section we develop a generalisation of this propensity function for non-ideal systems based on the detailed-balance condition.

In most cases, the dynamics of the CRN will tend to a stationary distribution in the long-time limit. If this is the case, the steady state of the dynamics of the network will be defined by a zero time derivative of $P(\mathbf{n}, t)$ and will therefore satisfy the following relation:

$$\sum_\rho f_{+\rho}(\mathbf{n}-\mathbf{r}^\rho)P(\mathbf{n}-\mathbf{r}^\rho) + \sum_\rho f_{-\rho}(\mathbf{n}-\mathbf{s}^\rho)P(\mathbf{n}-\mathbf{s}^\rho) = \sum_\rho (f_{+\rho}(\mathbf{n}) + f_{-\rho}(\mathbf{n}))P(\mathbf{n}). \tag{3.6}$$

In general, it is not possible to obtain an analytical form for the steady-state distribution of the CME. However, for a broad class of steady states its probability distribution can be constructed analytically, as it will be shown in the following chapter.

---

[4] The diluteness condition has been introduced to ensure that, for the moment, interactions between solutes will not be strong enough to cause inhomogeneities in the solution. Afterwards, we will see that phase separation can occur in the presence of interactions between solutes and, thus, the system cannot be described by a single homogeneous concentration.

3.1.2 Deterministic description of a Chemical Reaction Network

For large particle numbers, by taking the averages of eq. (3.4) and making mean-field assumptions – that is, assuming vanishing correlations – one can derive a set of equations for the concentrations in the macroscopic limit, that is, when $n_i/V = c_i$, for both n_i and V large. T. G. Kurtz [Kurtz, 1972] showed that these two approaches are indeed equivalent in the large volume limit, at least for MAK. In this limit, the state of the system is fully specified by a set of concentrations $(c_1, \cdots, c_N)$ and one obtains the following classical set of equations for the dynamics of the concentrations c_i in a CRN [Schnoerr et al., 2017]:

$$\frac{\partial c_i}{\partial t} = \sum_\rho v_i^\rho (J_{+\rho} - J_{-\rho}), \tag{3.7}$$

where the currents J still need to be determined. For the classical case of MAK, in the deterministic description, the currents take the form

$$J_{+\rho} = k_{+\rho} \prod_i c_i^{r_i^\rho}, \text{ and } J_{-\rho} = k_{-\rho} \prod_i c_i^{s_i^\rho}. \tag{3.8}$$

While both the deterministic and stochastic descriptions refer to the same system, the former one is only accurate for large particle numbers, also known as the thermodynamic limit, where fluctuations are negligible.

3.2 Thermodynamical constraints on the dynamics of Chemical Reaction Networks

In the previous section we introduced the general description of CRNs, both at the stochastic and deterministic level, but, in both cases, a choice for the propensity functions or currents must be made. For ideal solutions, the most common choice is MAK, as outlined above. However, here we are considering solutes that interact among themselves, which are, therefore, not ideal. In this section, our aim is to specify the propensity functions or currents for non-ideal systems by assuming that, if there is no explicit non-equilibrium driving, the system must relax to thermodynamic equilibrium.

3.2.1 Stochastic dynamics

At thermodynamic equilibrium, detailed balance must necessarily hold for every reaction ρ, in order to fulfill the second law of thermodynamics. Detailed balance implies that the probability flux across a reaction ρ in the forward direction must equal the probability flux in the backward direction, thus preserving time reversal symmetry in the system and precluding from systematically extracting work from the fluctuations in the system [Sekimoto, 2010]. Mathematically, for stochastic systems, detailed balance takes the following form

$$P^{\text{eq}}(\mathbf{n}) f_{+\rho}(\mathbf{n}) = f_{-\rho}(\mathbf{n} + \mathbf{v}^\rho) P^{\text{eq}}(\mathbf{n} + \mathbf{v}^\rho), \tag{3.9}$$

where the equilibrium probability distribution $P^{\mathrm{eq}}(\mathbf{n})$ for closed stochastic systems (no energy or matter[5] exchanged with the environment) is given by the canonical Boltzmann distribution:

$$P^{\mathrm{eq}}(\mathbf{n}) = \frac{1}{Z} e^{-\beta F(\mathbf{n})}, \tag{3.10}$$

β being the inverse of temperature times the Boltzmann constant, $F(\mathbf{n})$ the Helmholtz free energy of the system at state $\mathbf{n}$ and Z a normalisation factor, known as the partition function in the statistical physics literature. Importantly, the free energy F that appears in Eq. (3.10) is not necessarily an ideal free energy but can be as complex as needed and include interactions between the different constituents of the system.

Detailed balance, as expressed in Eq. (3.9), yields the following constraint for the propensity functions:

$$\frac{f_{+\rho}(\mathbf{n})}{f_{-\rho}(\mathbf{n}+\mathbf{v}^{\rho})} = \frac{P^{\mathrm{eq}}(\mathbf{n}+\mathbf{v}^{\rho})}{P^{\mathrm{eq}}(\mathbf{n})} = e^{-\beta[F(\mathbf{n}+\mathbf{v}^{\rho})-F(\mathbf{n})]}, \tag{3.11}$$

implying that the reaction rates do depend on the interactions within the system through the changes in free energy provoked by the reaction. Then, we can choose the following functional form for the propensities f_{ρ}:

$$f_{+\rho}(\mathbf{n}) = k_{\rho} e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{r}^{\rho})]}, \qquad f_{-\rho}(\mathbf{n}+\mathbf{v}^{\rho}) = k_{\rho} e^{\beta[F(\mathbf{n}+\mathbf{v}^{\rho})-F(\mathbf{n}+\mathbf{v}^{\rho}-\mathbf{s}^{\rho})]}, \tag{3.12}$$

where k_{ρ} is the reaction constant, which has to be equal for both the forward and the backward reaction[6]. This is a difference with the usual MAK description, where the dependency of the rates on the standard-state chemical potential of the species (the part of the chemical potential that does not depend on the state of the system $\mathbf{n}$) is hidden in the rate constants [Kondepundi and Prigogine, 2015, Rao and Esposito, 2016]. This form for the propensity functions satisfies Eq. (3.11) and means the rate of the reactions is a function of the free energy of the reactant complex.

Note that, while we impose detailed-balance conditions to the rates at equilibrium, this does not mean they are only valid for the equilibrium state, it only ensures that the system relaxes to equilibrium in the absence of external driving or work. By doing so, we found a functional form for the propensity functions (3.12) that we will consider to be valid even if the system is not at thermodynamic equilibrium. This is inspired by the fact that rates of the form of MAK are thought to be an accurate description of reactions both at equilibrium and far away from equilibrium [Groot and Mazur, 1983].

Mass Action Kinetics can be derived from an ideal free energy

The form (3.12) for the propensity functions is not unique but is particularly appealing given that it reduces to MAK for ideal systems. In an ideal system, at least in a lattice model for the solution

---

[5] We have used a canonical Boltzmann distribution for the argument, which implicitly assumes that total number of particles fixed. However, a generalisation can be obtained for a system in equilibrium with a particle reservoir by using the grand-canonical distribution instead.

[6] Note that this is not the most general form of the propensity functions. A more general expression will be presented later in Section 3.2.3.

(see Appendix C.1) with total volume $V = \sum_i n_i$, the free energy is

$$F_{\text{id}} = \sum_i n_i \mu_i^0 + \beta^{-1}\left(\sum_i \log(n_i!) - \log(V!)\right), \tag{3.13}$$

where μ_i^0 is the standard-state chemical potential of species i. Then, the rates take the following form:

$$f_{+\rho}(\mathbf{n}) = k_\rho e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{r}^\rho)]} = k_\rho e^{\beta\sum_i r_i^\rho \mu_i^0}\frac{(V-\sum_i r_i^\rho)!}{V!}\prod_i \frac{n_i!}{(n_i - r_i^\rho)!}, \tag{3.14}$$

where $(V - \sum_i r_i^\rho)!/V!$ can be approximated by $V^{-\sum_i r_i^\rho}$. After redefining

$$k_{+\rho} = k_\rho \frac{\exp(\beta\sum_i r_i^\rho \mu_i^0)}{V^{\sum_i r_i^\rho}}, \tag{3.15}$$

we find that the propensity functions (3.12) for an ideal free energy matches the propensity function for stochastic MAK, Eq. (3.5). Note that, after redefining the reaction constant $k_{+\rho}$, the forward and backward reaction constants are no longer equal.

Example 2 *Let us consider the second reaction in Eq. (3.2):*

$$B \rightleftharpoons D \tag{3.16}$$

Then the propensity functions, according to Eq. (3.12), are

$$f_{+\rho}(\boldsymbol{n}) = k_\rho \exp(\beta[F(\boldsymbol{n}) - F(\boldsymbol{n} - \mathbf{r}^\rho)]) \quad f_{-\rho}(\mathbf{n}) = k_\rho \exp(\beta[F(\boldsymbol{n}) - F(\boldsymbol{n} - \mathbf{s}^\rho)]). \tag{3.17}$$

Let us now define the chemical potential of the i-th species for stochastic (small) systems as follows:

$$\mu_i(\boldsymbol{n}) = F(\boldsymbol{n}) - F(\boldsymbol{n} - \boldsymbol{e}_i), \tag{3.18}$$

where $\boldsymbol{e}_i$ is the unit vector with all entries zero except for the i-th entry that is 1. Then, the propensity functions of unimolecular reactions (like $B \rightleftharpoons D$) can be written as

$$f_{+\rho}(\boldsymbol{n}) = k_\rho \exp[\beta\mu_B(\boldsymbol{n})] \qquad f_{-\rho}(\boldsymbol{n}) = k_\rho \exp[\beta\mu_D(\boldsymbol{n})], \tag{3.19}$$

which, again, reduces to mass action kinetics for unimolecular reactions if we consider an ideal system. □

3.2.2 Deterministic dynamics

In systems with large particle numbers, where a deterministic description, as the one presented in Section 3.1.2, is accurate, one can adapt the rates (3.12) obtained for the stochastic system to this macroscopic limit. When the particle numbers $\mathbf{n}$ are large the differences in free energies that appear in the rates (3.12) can be expanded in Taylor series and rewritten as

$$F(\mathbf{n}) - F(\mathbf{n} - \mathbf{r}^\rho) \simeq \sum_i r_i^\rho \mu_i, \tag{3.20}$$

where μ_i is the standard chemical potential of species i, $\mu_i = \partial f(\mathbf{c})/\partial c_i$, and $f(\mathbf{c})$ is the free energy of the system per unit volume in the deterministic notation. Eq. (3.20) becomes exact in the limit when $\mathbf{n} \to \infty$, while $\mathbf{r}^\rho$ remains finite.

Therefore, the currents in a deterministic and non-ideal CRN, whose dynamics are given by Eq. (3.7), can be written as

$$J_{+\rho} = k_\rho e^{\beta \sum_i r_i^\rho \mu_i}, \qquad J_{-\rho} = k_\rho e^{\beta \sum_i s_i^\rho \mu_i}, \tag{3.21}$$

which is an expression conceptually similar to those given by other approaches to build thermodynamically consistent descriptions for deterministic non-ideal CRNs [Avanzini et al., 2021, Bazant, 2013].

Once again, the currents (3.21) match their ideal MAK counterpart (3.8) if the chemical potentials used in the rates are those of an ideal solution, i.e., $\mu_i = \beta^{-1} \log c_i + \mu_i^0$. In this case, we obtain, e.g. for the forward reaction,

$$J_{+\rho} = k_\rho e^{\beta \sum_i r_i^\rho \mu_i} = k_{+\rho} \prod_i c_i^{r_i^\rho}, \tag{3.22}$$

which matches the deterministic MAK, Eq. (3.8), and where $k_{+\rho} = k_\rho \exp(\beta \sum_i r_i^\rho \mu_i^0)$.

In Sections 3.2.1 and 3.2.2, we have written the chemical reaction rates for both stochastic and deterministic systems in terms of their free energies. Given that one of the aims of the work presented here is to describe CRNs in phase-separating systems, in Appendix C.1 we briefly review the Flory-Huggings or regular solution theory with interactions, widely used to model phase separation, in order to obtain free energies and chemical reaction rates (stochastic and deterministic) for phase-separating solutions, results that will be used in the following chapter. In Appendix C.2, we show how to obtain the non-ideal chemical reaction rates from activity coefficients, more common in the chemistry literature.

3.2.3 On the rate constant

Until now we have assumed k_ρ to be a constant of the reaction and completely insensitive to any change of state in the system. However, in full generality, this constant could depend on the state of the system, as the system is part of the environment the reaction takes place in. These effects can be disregarded for most cases in ideal or dilute solutions, but in non-ideal ones these effects may not be negligible, particularly in the case of phase separation, where there could be two (or more) very different environments where the chemical reaction takes place, thereby accelerating it or slowing it down. Nevertheless, detailed balance, Eq. (3.11), must continue to hold. This means that the forward reaction constant for a state $\mathbf{n}$ must be equal to the backward reaction constant for a state $\mathbf{n}+\mathbf{v}^\rho$. One way to ensure this equality while keeping the state-dependency of the reaction constants is to make k_ρ a function of the state after subtracting the reactant complex, that is $\mathbf{n}-\mathbf{r}^\rho$ for the forward case and $\mathbf{n}+\mathbf{v}^\rho-\mathbf{s}^\rho$ for the backward one. Noting that the detailed-balance equality is still satisfied by virtue of $\mathbf{n}-\mathbf{r}^\rho = \mathbf{n}+\mathbf{v}^\rho-\mathbf{s}^\rho$, we can write down the following more general rates:

$$f_{+\rho}(\mathbf{n}) = k_\rho(\mathbf{n}-\mathbf{r}^\rho) e^{\beta(F(\mathbf{n})-F(\mathbf{n}-\mathbf{r}^\rho))}, \qquad f_{-\rho}(\mathbf{n}+\mathbf{v}^\rho) = k_\rho(\mathbf{n}+\mathbf{v}^\rho-\mathbf{s}^\rho) e^{\beta(F(\mathbf{n}+\mathbf{v}^\rho)-F(\mathbf{n}+\mathbf{v}^\rho-\mathbf{s}^\rho))}. \tag{3.23}$$

From a physical perspective, this means that the rate constant can only be dynamically affected by the entire system except for the complex that takes part in the reaction.

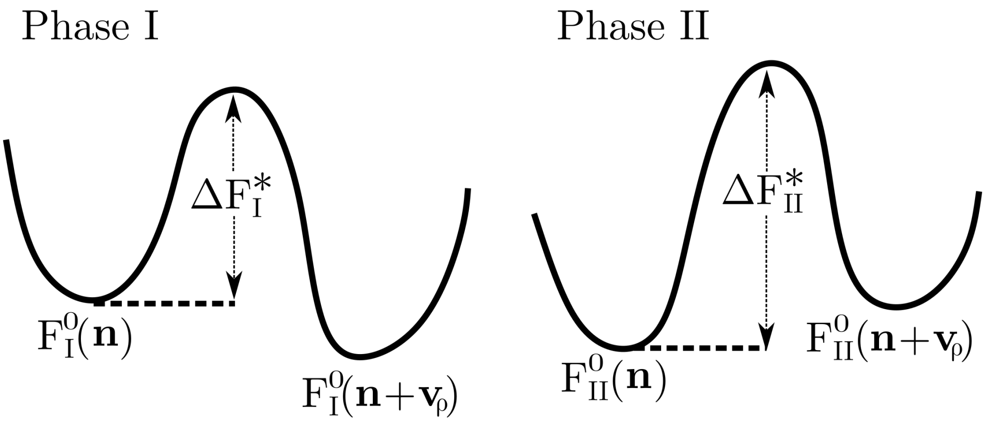


Figure 3.1: Reaction free energy landscape for a chemical reaction in two different phases. The horizontal dimension is the reaction coordinate and the vertical one specifies the height of the reaction free energy landscape F^0 (which is equal to F except for the fact that it does not include the entropic term $\prod_i \log n_i!$, as is customary [Kondepundi and Prigogine, 2015]). For each phase there are two minima in the free energy landscape corresponding to whether the reaction has occurred (right-hand minimum) or not (left-hand one). The free energy of the system (F) specifies the free energy of the system before the reaction happens [$F^0(\mathbf{n})$] and after [$F^0(\mathbf{n}+\mathbf{v}_\rho)$] but it does not determine the height of the reaction free energy barrier ΔF^*. Thus, how ΔF^* depends on the environement can be crucial for the behaviour of the CRN. The subscripts specify to which of the phases the value of the variable refers to.

In analogy with the classical transition state theory, we can think of the microscopic mechanism of a reaction as a diffusion process in a reaction free-energy landscape in the reaction coordinates [Kondepundi and Prigogine, 2015, Hänggi et al., 1990], see Fig. 3.1. Then, the value of the rate constant depends on the height of the free energy barrier ΔF^* of the reaction. While the free energies of the reactants and products (the stable states in the reaction landscape) have free energies defined by F, this is no longer the case for the height of the barrier ΔF^*. For systems where more than one phase coexist, how the height of the barrier (and, thus, the value of k_ρ) depends on which phase the reaction takes place in can be crucial for the behaviour of the CRN. Fig. 3.1 graphically summarises the arguments given in this section and it highlights the effect a different phase or environment can have in the reaction free-energy landscapes and, thus, on the reaction constants k_ρ. Note that, setting a value of the barrier imposes a relation between the forward and backward reactions and is therefore equivalent to requiring detailed balance to hold at equilibrium.

3.2.4 Rates for non-equilibrium systems

Until now we have considered the rates or propensity functions of chemical reactions that, even if they are not at thermodynamic equilibrium, they will eventually relax to such a state as a consequence of the functional form of these rates. However, we are not interested in systems that relax to equilibrium but in systems that, like living beings, dissipate free-energy to maintain themselves away from thermodynamic equilibrium.

Let us consider that a certain subset of the N species of the system are connected to different particle reservoirs (chemostats) that will keep their concentration constant. Then, in general, the system will not relax to equilibrium due to the work done on the system by the chemostats. In this case, the space of possible states of the system is reduced to N' dimensions (as the chemostatted species have their concentrations fixed) and the thermodynamic constraint on the propensity functions, Eq. (3.9), takes the following form:

$$\frac{f_{+\rho}(\mathbf{n})}{f_{-\rho}(\mathbf{n}+\mathbf{v}^{\rho})} = e^{-\beta(F(\mathbf{n}+\mathbf{v}^{\rho})-F(\mathbf{n})+\sum_j v_j^{\rho}\mu_j)} \tag{3.24}$$

where, now, $\mathbf{n}$ is a vector that only encodes the particle number of the N' non-chemostatted species, F in the first term in the exponential also refers to the free energy of non-chemostatted species and the summation over j runs only over the chemostatted species. The second term ($\sum_j v_j^{\rho}\mu_j$) specifies the chemical work done by the chemostats (with chemical potentials fixed at μ_j) when a reaction ρ occurs, which pushes the system out of equilibrium. Here, an approximation has been made by which the chemostatted species are ideal (no interactions with the rest of the chemical species) and abundant, hence, their free energy change can be replaced by the sum of chemical potentials, as is done for deterministic CRNs, see Eq. (3.20), and we are able to neglect fluctuations around their chemostatted values. Therefore, the system is now fully specified by a vector $\mathbf{n}$ that includes the particle numbers of the non-chemostatted species only. Finally, as discussed in Section 3.2.3, the rate constants may have state dependencies and the rates including the contributions from the chemostats can take the following form

$$\begin{aligned} f_{+\rho}(\mathbf{n}) &= \tilde{k}_{\rho} g_{\rho}(\mathbf{n}-\mathbf{r}^{\rho}) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{r}^{\rho})+\sum_j r_j^{\rho}\mu_j]}, \\ f_{-\rho}(\mathbf{n}+\mathbf{v}^{\rho}) &= \tilde{k}_{\rho} g_{\rho}(\mathbf{n}+\mathbf{v}^{\rho}-\mathbf{s}^{\rho}) e^{\beta[F(\mathbf{n}+\mathbf{v}^{\rho})-F(\mathbf{n}+\mathbf{v}^{\rho}-\mathbf{s}^{\rho})+\sum_j s_j^{\rho}\mu_j]}, \end{aligned} \tag{3.25}$$

where $\sum_j r_j^{\rho}\mu_j$ and $\sum_j s_j^{\rho}\mu_j$ include the contribution of the chemostats for the forward and backward reactions, respectively. For the sake of clarity, in the remainder of the text, we will reserve the index j to denote the chemostatted species and the index i to denote the rest of chemical species (non-chemostatted). Finally, $\tilde{k}_{\rho}$ is constant and any state dependency k_{ρ} may have is absorbed into the function g_{ρ}. We note that a similar, and equally general, form for the kinetics has been used in other contexts, such as the modelling of molecular motors [Jülicher et al., 1997].

Analogously, in the deterministic limit, the above propensity functions become the following currents

$$\begin{aligned} J_{+\rho}(\mathbf{c}) &= \tilde{k}_{\rho} g_{\rho}(\mathbf{c}) e^{\beta[\sum_i r_i^{\rho}\mu_i+\sum_j r_j^{\rho}\mu_j]}, \\ J_{-\rho}(\mathbf{c}) &= \tilde{k}_{\rho} g_{\rho}(\mathbf{c}) e^{\beta[\sum_i s_i^{\rho}\mu_i+\sum_j s_j^{\rho}\mu_j]}, \end{aligned} \tag{3.26}$$

where, μ_i is the chemical potential of the non-chemostatted species and is thus a dynamic variable, while μ_j corresponds to the chemical potential of chemostatted species and is therefore fixed.

3.3 Thermodynamics of Chemical Reaction Networks

Until now we have mostly considered the dynamics of non-ideal CRNs, but for the rest of the chapter we will focus on their energetics and review certain thermodynamic relationships valid for stochastic systems far away from equilibrium. These relationships will be useful in the following chapter, as they will shed light into the thermodynamic interpretations of the results obtained there.

3.3.1 Thermodynamics of closed Chemical Reaction Networks

In this section we consider a closed stochastic non-ideal CRN. This implies that the system, if prepared in an arbitrary state, will always relax to its equilibrium state, since it is not connected to any chemostat or any other source of work. Therefore, we can define a thermodynamic potential [Rao and Esposito, 2018]

$$G(\mathbf{n},t) = k_B T \log P(\mathbf{n},t) + F(\mathbf{n}) \tag{3.27}$$

whose average value, taken over the probabilities $P(\mathbf{n},t)$, acts as a Lyapunov function of the system and takes its minimum value at thermodynamical equilibrium. Thus, G can be called the stochastic free energy of the system, as it includes a contribution from the probability P of the system being at state $\mathbf{n}$ and time t.

Then, entropy of the system is [Schmiedl and Seifert, 2007]

$$\begin{aligned} S(\mathbf{n},t) =& -\frac{\partial G(\mathbf{n},t)}{\partial T} \\ =& -k_B \log P(\mathbf{n},t) + s(\mathbf{n}), \end{aligned} \tag{3.28}$$

where $s(\mathbf{n}) = -\partial F(\mathbf{n})/\partial T$. The enthalpy is thus $H(\mathbf{n}) = G(\mathbf{n},t) + TS(\mathbf{n},t)$.

In closed solutions, the change in enthalpy across a reaction ρ – i.e. the difference in enthalpy after reaction ρ occurs – is the heat flow from (or to) the thermal reservoir the system is in contact with:

$$\Delta_\rho H(\mathbf{n}) = H(\mathbf{n}+\mathbf{v}_\rho) - H(\mathbf{n}) = -Q_\rho^{\text{env}}, \tag{3.29}$$

which can be seen as a statement of the first law of thermodynamics for closed systems at constant volume [Pippard, 1964].

Likewise, the change in the entropy of the system across a reaction ρ in a closed system takes the following form:

$$\begin{aligned} \Delta_\rho S(\mathbf{n},t) &= S(\mathbf{n}+\mathbf{v}_\rho,t) - S(\mathbf{n},t) \\ &= -k_B \log \frac{P(\mathbf{n}+\mathbf{v}_\rho,t)}{P(\mathbf{n},t)} + \frac{\Delta_\rho H(\mathbf{n}) - \Delta_\rho F(\mathbf{n})}{T} \\ &= -\frac{\Delta_\rho G(\mathbf{n},t) + Q_\rho^{\text{env}}}{T}, \end{aligned} \tag{3.30}$$

where, in all cases, Δ_ρ stands for the difference between the state functions after and before the reaction ρ. Finally, the changes in the entropy of the thermal reservoir due to a reaction ρ occurring in a closed system are given by the exchanged heat [Kondepundi and Prigogine, 2015, Rao and Esposito, 2016]:

$$\Delta_\rho S^{\text{env}} = \frac{Q_\rho^{\text{env}}}{T}. \tag{3.31}$$

Therefore, taking into account both the entropy of the system and that of the environment, the total entropy production across a reaction is

$$\Delta_\rho S^{\text{tot}} = \Delta_\rho S(\mathbf{n},t) + \Delta_\rho S^{\text{env}} = -\frac{\Delta_\rho G(\mathbf{n},t)}{T}. \tag{3.32}$$

A full stochastic trajectory is completely determined by the occurrence of a set of chemical reactions $\{\rho_k\}$ at times $\{t_k\}$. The trajectory can then be encoded in the instantaneous current for each of the reactions during the dynamics [Rao and Esposito, 2018]:

$$j_\rho(\mathbf{n},\tau) = \sum_k \delta_{\rho,\rho_k}\delta_{\mathbf{n},\mathbf{n}_{t_k}}\delta(\tau - t_k), \tag{3.33}$$

whose probability can be obtained from the CME (3.4).

For a trajectory described by Eq. (3.33) – which follows the dynamics of the CME (3.4) – entropy production can be written as [Seifert, 2005, Rao and Esposito, 2018]

$$\Delta S^{\text{tot}}[\mathbf{n}_\mathbf{t}] = \int_0^t d\tau \left[\frac{\partial P(\mathbf{n},\tau)}{\partial \tau} + \sum_{\mathbf{n},\rho}(\Delta_\rho S(\mathbf{n}) + \Delta_\rho S^{\text{env}})j_\rho(\mathbf{n},\tau)\right] = -\frac{\Delta G[\mathbf{n}_\mathbf{t}]}{T}, \tag{3.34}$$

where

$$\Delta G[\mathbf{n}_\mathbf{t}] = \int_0^t d\tau \left[\frac{\partial P(\mathbf{n},\tau)}{\partial \tau} + \sum_{\mathbf{n},\rho}\Delta_\rho G(\mathbf{n},t)j_\rho(\mathbf{n},\tau)\right] \tag{3.35}$$

is the change in free energy along the stochastic trajectory and we use $[\mathbf{n}_\mathbf{t}]$ to denote a full stochastic trajectory from $\tau = 0$ to time $\tau = t$. The two previous equations are differences of a state variable along a trajectory that depends on the stochastic occurrences of chemical reactions and the changes in probabilities, these last ones obtained from solving the CME. We emphasise that the previous relations hold only for closed systems, where the entropy production is, indeed, given by the change in free energy, and a generalisation to open non-equilibrium systems will be exposed later.

Finally, it can be proved that the average total entropy difference along a trajectory is always positive [Seifert, 2005]

$$\langle \Delta S^{\text{tot}} \rangle \geq 0, \tag{3.36}$$

which is a statement of the second law of thermodynamics. This also implies that the average stochastic free energy G decreases during a trajectory, as expected in closed systems where no work is being done.

3.3.2 Non-equilibrium thermodynamics of Chemical Reaction Networks

Here we consider a CRN as in the previous section but that has been pushed out of equilibrium by placing it in contact with different particle reservoirs, at given chemical potentials. In this case, there are energy fluxes flowing through the network that require a free-energy expenditure to be maintained, which, as noted in the *Preface*, is a hallmark of living beings.

Therefore, the reservoirs are exerting a work on the system and a restatement of the first law of thermodynamics, Eq. (3.29), is needed to account for it. Across a reaction ρ there will be now two different terms that contribute to the change in enthalpy [Rao and Esposito, 2018]

$$\Delta_\rho H(\mathbf{n}) = -Q_\rho^{\text{env}} - Q_\rho^{\text{chem}} + W_\rho^{\text{chem}}, \tag{3.37}$$

where $W_\rho^{\text{chem}} = \sum_j \mu_j \mathbf{v}_j^\rho$ is the work made by the chemostats across reaction ρ, and $Q_\rho^{\text{chem}} = -T\sum_j s_j \mathbf{v}_j^\rho$ is the heat exchanged with the chemostats, with $h_j - Ts_j = \mu_j$ being the enthalpic

(h_j) and entropic (s_j) part of the chemostatted chemical potentials μ_j, respectively. This is an expression of the first law of thermodynamics for a system at constant volume on which work is being done by placing it in contact with chemostats at different chemical potentials. The decomposition of the heat the system exchanges follows from the identification of work performed in the system as $W_\rho^{\text{chem}} = \sum_j \mu_j \mathbf{v}_j^\rho$, which is a meaningful and intuitive definition, particularly if we keep in mind the canonical view of how a biological cell stores energy: in the form of a chemical potential difference between ATP and ADP.

As before, let us now consider the changes in the entropy of the system across a reaction ρ, which, unlike before, may be driven by the work performed by the chemostats:

$$\begin{aligned}\Delta_\rho S(\mathbf{n},t) &= -k_B \log \frac{P(\mathbf{n}+\mathbf{v}_\rho,t)}{P(\mathbf{n},t)} + \frac{\Delta_\rho H(\mathbf{n}) - \Delta_\rho F(\mathbf{n})}{T} \\ &= -\frac{\Delta_\rho G(\mathbf{n},t) + Q_\rho^{\text{env}} + Q_\rho^{\text{chem}} - W_\rho^{\text{chem}}}{T}.\end{aligned} \tag{3.38}$$

Similarly, the entropy change of the thermal bath is

$$\Delta_\rho S^{\text{env}} = \frac{Q_\rho^{\text{env}} + Q_\rho^{\text{chem}}}{T}. \tag{3.39}$$

These two relations now include contributions from the chemostats, both in the form of heat and work.

Then, the total entropy production along a stochastic trajectory for a CRN on which chemical work is being exerted is given by [Rao and Esposito, 2018]

$$\Delta S^{\text{tot}}[\mathbf{n}_\mathbf{t}] = \int_0^t d\tau \left[\frac{\partial P(\mathbf{n},\tau)}{\partial \tau} + \sum_{\mathbf{n},\rho} (\Delta_\rho S(\mathbf{n}) + \Delta_\rho S^{\text{env}}) j_\rho(\mathbf{n},\tau)\right] = -\frac{\Delta G[\mathbf{n}_\mathbf{t}] - W^{\text{chem}}[\mathbf{n}_\mathbf{t}]}{T} \tag{3.40}$$

where, unless specified, differences in state variables (S^{tot} and G) are taken along a full trajectory and not only across a single reaction and $W^{\text{chem}}[\mathbf{n}_\mathbf{t}]$ is the total chemical work performed along the trajectory. Again, one can prove that the average entropy production is positive [Seifert, 2005, Rao and Esposito, 2018],

$$\langle \Delta S^{\text{tot}} \rangle \geq 0, \tag{3.41}$$

even in the presence of external work. This yields a bound for the work required to transform a system from one state to another, which reminiscent of the inequalities obtained from equilibrium thermodynamics:

$$\langle W^{\text{chem}} \rangle \geq \langle \Delta G \rangle. \tag{3.42}$$

However, $\langle \Delta G \rangle$ need not be evaluated at thermodynamic equilibrium but can correspond to any non-equilibrium steady state which requires free-energy dissipation in order to be maintained.

A refinement of these ideas can be found in [Rao and Esposito, 2018], obtained by separating reversible from irreversible work sources. However, for our purposes, the bounds stated in this section are enough and will be applied in the next chapter.

Chapter 4

The role of complex balance in Reaction-Diffusion networks

For over a century it has been known that if for each reaction ρ the backward and forward rates are equal, then the steady state of a CRN with MAK satisfies the thermodynamic equilibrium condition [Lewis, 1925]. However, the steady state of a CRN with MAK need not be the point of thermodynamic equilibrium, as pointed out first by R. Wegscheider in 1902 and later, in a more general way, by L. Onsager in 1931 [Onsager, 1931].

Nevertheless, a formalism akin to that of equilibrium thermodynamics can emerge, for MAK but also for more general kinetics, in the sense of being able to define a function whose value is minimised by the kinetics of the system and relating its minimum value with the steady state of the system. This is definitely the case for complex-balanced systems, as it will be shown below.

4.1 Definition of complex and complex-balanced network

In the previous chapter, we defined a *complex* as the set of chemical species (with their respective particle numbers) that take part in reaction, either as a reactant or as a product. Its most general expression is the vector

$$\mathbf{z} = (z_1, \cdots, z_i, \cdots) \tag{4.1}$$

where the index i runs over all chemical species and z_i is an integer coefficient that specifies the number of molecules of the species X_i that appear in the complex $\mathbf{z}$. Then, any CRN (as described in Section 3.1) can be represented as a graph whose nodes represent the complexes that take part in the reactions and there is an edge between two complexes if and only if the reaction $\mathbf{z}_m \leftrightharpoons \mathbf{z}_n$ exists, where $\mathbf{z}_m$ and $\mathbf{z}_n$ denote two different complexes (see e.g. Fig 4.1).

In a deterministic CRN, whose kinetics are given by Eq. (3.7), the *rate of creation of the complex* $\mathbf{z}$ is defined as

$$J_{+\mathbf{z}} = \sum_{\rho:\mathbf{s}^\rho=\mathbf{z}} J_{+\rho} + \sum_{\rho:\mathbf{r}^\rho=\mathbf{z}} J_{-\rho} \tag{4.2}$$

where the subscript $\rho : \mathbf{s}^\rho = \mathbf{z}$ indicates that the sum is taken over the reactions ρ whose product complex equals the complex $\mathbf{z}$ and similarly for $\rho : \mathbf{r}^\rho = \mathbf{z}$, which denotes that that the sum is

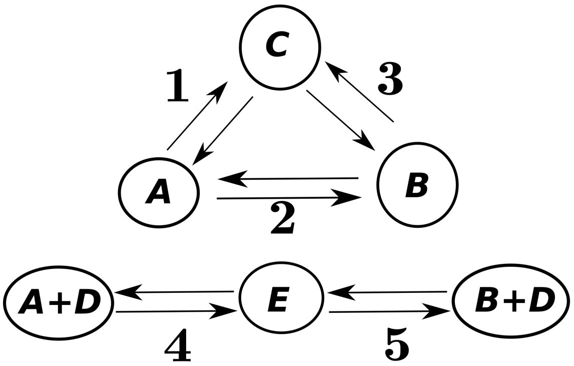


Figure 4.1: Graphical representation of CRNs discussed in Section *A hierarchy of steady states*. The CRN has 6 complexes each of them represented in one of the nodes of the network: A, B, C, $A + D$, E and $B + D$. The five reactions present in the CRN are numbered.

taken over those reactions whose reactant complex equals the complex $\mathbf{z}$. Analogously, we can define the rate of annihilation of the complex $\mathbf{z}$ as

$$J_{-\mathbf{z}} = \sum_{\rho:\mathbf{s}^{\rho}=\mathbf{z}} J_{-\rho} + \sum_{\rho:\mathbf{r}^{\rho}=\mathbf{z}} J_{+\rho}. \tag{4.3}$$

Then, a deterministic network is said to have a *complex-balanced steady state* if its steady state satisfies the condition that the creation rate and the annihilation rate of each complex are equal [Horn and Jackson, 1972]. Mathematically, this means that

$$J_{+\mathbf{z}} = J_{-\mathbf{z}} \quad \forall\, \mathbf{z}. \tag{4.4}$$

We have defined complex-balancing only for a deterministic network but below we will see how this condition will also affect the stochastic dynamics.

A hierarchy of steady-states

Given the definitions above, a complex-balanced steady state is only a subset of all the steady states a general CRN can have. However, complex-balanced steady states are still a broad class of steady states which includes, but is not restricted to, detailed-balance steady states.

Therefore, we can order these types of steady states in terms of their generality:

1. The most restrictive condition we can impose to a steady state is that of detailed balance:

   $$J_{+\rho} = J_{-\rho} \quad \forall\, \rho, \tag{4.5}$$

   condition that corresponds to a system at thermodynamic equilibrium and implies that the rate of the forward reaction equals the rate of the backward reaction, for every reaction ρ.

2. More general than detailed-balanced steady states are complex-balanced steady states, which satisfy that

   $$\sum_{\rho:\mathbf{s}_{\rho}=\mathbf{z}} J_{+\rho} + \sum_{\rho:\mathbf{r}_{\rho}=\mathbf{z}} J_{-\rho} = \sum_{\rho:\mathbf{s}_{\rho}=\mathbf{z}} J_{-\rho} + \sum_{\rho:\mathbf{r}_{\rho}=\mathbf{z}} J_{+\rho} \quad \forall\, \mathbf{z}, \tag{4.6}$$

   which implies that the creation rate and the annihilation rate of each complex are equal. In the example of Fig. 4.1, it corresponds to having the net rate $(J_{+\rho} - J_{-\rho})$ in reaction 1 equal to those of 2 and 3, and, thus, reactions 4 and 5 have to be detailed balanced (zero net rate for each reaction). While it is not the case for the CRN in Fig. 4.1, there are CRN topologies that only accept complex-balanced steady states, as we will see in the Section 4.2.

3. Finally, we have the most general class of steady states, which is just defined by vanishing time derivatives of the dynamical equation (3.7); condition that, by splitting it into the contributions of each complex $\mathbf{z}$, can be recast into

$$\sum_m z_i^m \left[\sum_{\rho:\mathbf{s}_\rho=\mathbf{z}_m} J_{+\rho} + \sum_{\rho:\mathbf{r}_\rho=\mathbf{z}_m} J_{-\rho} \right] = \sum_m z_i^m \left[\sum_{\rho:\mathbf{r}_\rho=\mathbf{z}_m} J_{+\rho} + \sum_{\rho:\mathbf{s}_\rho=\mathbf{z}_m} J_{-\rho} \right] \quad \forall\, i, \quad (4.7)$$

   where m is an index that labels each of the complexes in the network and the integer z_i^m represents its components. Then, there are no longer constraints between the net rates of each reaction [other than those imposed by Eq. (3.7)]. For Fig. 4.1, this implies that there can appear cycles at steady state where, for example, the species A is created by reaction 1 but annihilated by reaction 4, which breaks complex balance. A generic steady state can allow cycles that cannot be visualised directly from the network representation of the CRN in terms of complexes (such as the graphical representation in Fig. 4.1), while a complex balanced steady state only allows cycles that can be directly visualised from the network representation of a CRN in terms of complexes.

From this hierarchical classification, it can be clearly seen that detailed balancing (4.5) implies complex balancing (4.6) and that complex balancing (4.6) implies the system is at steady state (4.7). However, the converse is not true. Indeed, a steady state (4.7) does not necessarily have to be complex balanced (4.6) and a complex-balanced steady state (4.6) is more general than a detailed-balanced one (4.5). Therefore, complex balance is a less restrictive constrain than detailed balance, but it is still less general than a fully generic steady state. Moreover, a complex-balanced network need not be at thermodynamic equilibrium and can sustain net currents at steady state.

Complex balance in networks with mass action kinetics

As a particular case of especial importance, we give the explicit condition for complex-balancing in a network with MAK. The dynamics of a deterministic and ideal CRN [Eqs. (3.7) and (3.8)], with state independent rate constants, can be written as

$$\frac{\partial c_i}{\partial t} = \sum_\rho v_i^\rho \left(k_{+\rho} \prod_i c_i^{r_i^\rho} - k_{-\rho} \prod_i c_i^{s_i^\rho} \right), \quad (4.8)$$

where the non-equilibrium contribution of the chemostats has been absorbed into the rate constants $k_{+\rho}$ and $k_{-\rho}$. A deterministic and ideal CRN has a complex balanced steady state if for each complex $\mathbf{z}$ [Anderson et al., 2010]:

$$\sum_{\rho:\mathbf{s}^\rho=\mathbf{z}} \left(k_{+\rho} \prod_i (c_i^{\mathrm{SS}})^{r_i^\rho} - k_{-\rho} \prod_i (c_i^{\mathrm{SS}})^{s_i^\rho} \right) = \sum_{\rho:\mathbf{r}^\rho=\mathbf{z}} \left(k_{+\rho} \prod_i (c_i^{\mathrm{SS}})^{r_i^\rho} - k_{-\rho} \prod_i (c_i^{\mathrm{SS}})^{s_i^\rho} \right), \quad (4.9)$$

where c_i^{SS} is the steady-state concentration of species i[1].

---

[1] This expression differs from the one given in Ref. [Anderson et al., 2010] due to the explicit consideration that every reaction can happen in both senses, to better illustrate the alignment with thermodynamic principles.

4.2 Review of results for complex-balanced systems

In this section we review well-known results from the literature of ideal CRNs, which we will later generalise to non-ideal networks.

It was first realised half a century ago that CRNs that relax to complex-balanced steady states have a certain number of useful properties. In the seminal work by F. Horn and R. Jackson in 1972, they showed that if a deterministic CRN with MAK accepts a complex-balanced steady state (i.e. Eq. (4.9) is fulfilled) then the system is *quasi-thermodynamic*, which they defined as a system that minimises a *pseudo-Helmholtz function* [Horn and Jackson, 1972]:

$$\mathscr{L} = \sum_i c_i \left(\log c_i - \log c_i^{\mathrm{SS}} - 1 \right), \tag{4.10}$$

where c_i^{SS} is the steady-state concentration of species i. The choice of terminology was due to the fact that Horn and Jackson were looking for CRNs that satisfied a formalism similar to that of equilibrium thermodynamics, where the minimum of a free energy determines the equilibrium point of a system. Indeed, they found the function $\mathscr{L}$, whose minimum specifies the steady state of the system, but this function is not necessarily equal to the equilibrium free energy. This is due to the fact that, as pointed out by Wegscheider and Onsager among others, for a CRN with MAK to relax to equilibrium there are some constraints the rate constants must satisfy, which relate them to the standard-state chemical potentials of the species, as in Eq. (3.15). However, to minimise a *pseudo-Helmholtz function* like Eq. (4.10), these constraints are not needed and, therefore, it may correspond to a MAK system out of thermodynamic equilibrium.

Shortly thereafter, M. Feinberg showed that, just by examining the topology of the network, one can know if the steady state of the network must be complex balanced [Feinberg, 1972]. Given a CRN, one can define the deficiency of the network, δ, which is given by

$$\delta = C - L - S \tag{4.11}$$

where C is the number of complexes in the network, L is the number of linkage classes or connected components in the network and S is the dimension of the stoichiometric subspace, which amounts to the number of degrees of freedom in the dynamics. If $\delta = 0$, then the system necessarily has a complex-balanced steady state. Therefore, we can assess whether a MAK system will have a pseudo-Helmholtz function as a Lyapunov function only from looking at the topology of the network. Moreover, the deficiency of a network has an interpretation in terms of cycles: it gives the number of cycles that can exist at steady state but cannot be directly visualised from looking at the network graph in terms of the complexes. In other words, it gives the number of *emergent* or *hidden* cycles of a CRN [Polettini et al., 2015].

Example 3 *To illustrate the concept of deficiency we discuss the CRNs depicted in Fig. 4.2.*

First, we compute the deficiency δ of the CRN on the left of Fig. 4.2, which corresponds to a simple enzyme kinetics model with a substrate S, a product P, an enzyme E and an enzyme-substrate complex ES. This CRN has 6 complexes C (including the empty complex $\emptyset$ that represents exchanges with particle reservoirs) and two connected components L. Finally, the dimension of the stoichiometric subspace S is 3,

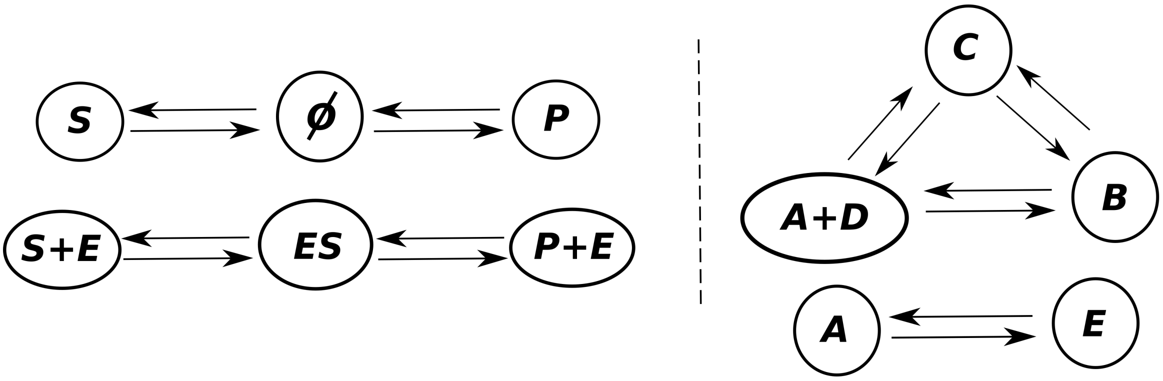


Figure 4.2: Graphical representation of CRNs discussed in Example 3. The left CRN has deficiency $\delta = 1$ and the right one is zero-deficient $\delta = 0$. Therefore, the network on the right will necessarily have a complex-balanced steady state but this is not the case for the left-hand one.

since the number of enzymes E and enzyme-substrate complexes ES is always conserved. Then deficiency $\delta = C - L - S = 1$, which means that:

1. *The steady state of this network need not be complex balanced.*
2. *There is one emergent cycle that is not represented graphically in the network representation of the complexes (see Fig. 4.2, left CRN). This emergent cycle is: the extraction of a substrate particle S from the particle reservoir, its binding to the enzyme to form a ES complex, conversion to a product P and final withdrawal of the product P from the system to the particle reservoir.*

The second CRN (right CRN of Fig. 4.2) is zero-deficient. It has 5 complexes C, 2 linkage classes L and a stoichiometric subspace, S, of 3 (since there are two conservation laws: $D + B + C$ is constant and the total mass of the system is also constant since there is no exchange with particle reservoirs). Then $\delta = C - L - S = 0$ which means that the steady state of this network is necessarily complex balanced and thus, if the kinetics are MAK, it has a Lyapunov function of the form (4.10). □.

More recently, it was found by D. F. Anderson, G. Craciun and T. D. Kurtz that, in the stochastic description, the steady-state probability distribution of a complex-balanced CRN with MAK can be written as a product of independent Poisson distributions [Anderson et al., 2010]. Under this constraint, the steady-state distribution π_{CB} of the network with dynamics (3.4) and propensity functions (3.5) is given by

$$\pi_{\mathrm{CB}}(\mathbf{n}) = \prod_{i=1}^{M} \frac{(c_i^{\mathrm{SS}})^{n_i}}{n_i!} e^{-c_i^{\mathrm{SS}}}, \tag{4.12}$$

where c_i^{SS} are the steady-state solutions of the complex-balanced networks in the deterministic limit [this is, Eqs. (3.7) and (3.8)]. Note that the expression (4.12) for the steady-state probability distribution is valid only for vectors $\mathbf{n}$ satisfying all the conservation laws the CRN may have, otherwise, its value is zero. This distribution includes the equilibrium distribution of the CRN (the Boltzmann distribution for an ideal system), but is not limited by thermodynamic equilibrium and may be applied to cases far from equilibrium. Furthermore, they also proved a more general result for propensity functions of the form

$$f_{+\rho}(\mathbf{n}) = k_{+\rho}\frac{\theta(\mathbf{n})}{\theta(\mathbf{n}-\mathbf{r}^{\rho})}, \tag{4.13}$$

which, in complex-balanced systems, yields a steady-state distribution with the following form [Anderson et al., 2010]:

$$\pi_{\mathrm{CB}}(\mathbf{n}) = \frac{M}{\theta(\mathbf{n})}\prod_{i=1}(c_i^{\mathrm{SS}})^{n_i}, \tag{4.14}$$

where M is a normalisation constant and θ is a function that maps the vector of integer numbers $\mathbf{n}$ into a single real-valued positive number.

To conclude with this review of the literature, recently D. F. Anderson and T. D. Nguyen genralised the Lyapunov function of complex-balanced MAK systems [Eq. (4.10)] to networks with product-form stationary distributions [Anderson and Nguyen, 2019]. In the following, we will further generalise some of these results and will give an interpretation of them in terms of thermodynamic quantities.

4.3 Complex-balanced steady-state distributions

Closely inspecting Eqs. (4.13) and (4.14) and identifying $\theta(\mathbf{n})$ with the Boltzmann factor $e^{-\beta F(\mathbf{n})}$, one can already get a hint of the connection with the thermodynamical constraints imposed on the propensity functions f_ρ in the previous chapter. In this section, we will look at a slight generalisation of the probability distribution (4.14), by considering a complex-balanced CRN, with propensity functions given by Eq. (3.25). However, in the following, we will assume that the function g_ρ in Eq. (3.25) is the same for all reactions, that is, $g_\rho = g$. Therefore, we do not require that the function g is a constant, but we do need the fact that the rate constants of all reactions depend on the state $\mathbf{n}$ in the same way, which implies that g for all reaction must be the same function. This simplification allows us to prove the result and the consequences of relaxing this hypothesis will be examined in the following Chapter 5.

Under these conditions, we claim that, for complex-balanced CRNs the steady-state solution of the stochastic dynamics (3.4) with the propensity functions (3.25) takes the form:

$$\pi_{\mathrm{CB}}(\mathbf{n}) = \frac{e^{-\beta[F(\mathbf{n})+\sum_i \tilde{\mu}_i n_i]}}{Z}, \tag{4.15}$$

where $\tilde{\mu}_i$ are parameters that will depend on the chemostats to which the system is connected (and that push the system out of equilibrium) and the reaction constants of the CRN. In fact, the parameter $\tilde{\mu}_i$ is related to c_i^{SS} in the previous section, but we have changed its form to better illustrate its thermodynamic meaning. Importantly, these $\tilde{\mu}_i$ parameters can be obtained from the ideal and deterministic CRN, greatly simplifying the task of obtaining analytically the steady-state distribution of the system. By analogy with equilibrium statistical physics, we denote by Z the normalisation factor of the probability distribution.

The strategy we will follow to prove the result is similar to that used by Anderson, Craciun and Kurtz in Ref. [Anderson et al., 2010], where the key point is to simplify the complex-balanced steady-state relation until we obtain that of a deterministic CRN modelled with MAK, given in Eq. (4.9). Then, we will be in a position to compare our parameters $\tilde{\mu}_i$ to the steady-state concentrations of the chemical species in the deterministic MAK case and write down the full probability distribution at steady state for complex-balanced non-ideal networks.

4.3.1 Proof

The CME (3.4), with rates of the form (3.25) and $g_\rho = g$ for all reactions ρ, at steady state becomes

$$\begin{aligned}
&\sum_\rho \tilde{k}_\rho g(\mathbf{n}-\mathbf{s}_\rho) e^{\beta[F(\mathbf{n}-\mathbf{v}_\rho)-F(\mathbf{n}-\mathbf{s}_\rho)+\sum_j r_j^\rho \mu_j]} P(\mathbf{n}-\mathbf{v}_\rho) + \\
&\sum_\rho \tilde{k}_\rho g(\mathbf{n}-\mathbf{r}_\rho) e^{\beta[F(\mathbf{n}+\mathbf{v}_\rho)-F(\mathbf{n}-\mathbf{r}_\rho)+\sum_j s_j^\rho \mu_j]} P(\mathbf{n}+\mathbf{v}_\rho) = \\
&\quad \sum_\rho \left(\tilde{k}_\rho g(\mathbf{n}-\mathbf{r}_\rho) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{r}_\rho)+\sum_j r_j^\rho \mu_j]} + \tilde{k}_\rho g(\mathbf{n}-\mathbf{s}_\rho) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{s}_\rho)+\sum_j s_j^\rho \mu_j]} \right) P(\mathbf{n}).
\end{aligned} \tag{4.16}$$

Dividing the previous expression by $P(\mathbf{n})$ and inserting an ansatz for the solution of the form (4.15) we obtain:

$$\begin{aligned}
&\sum_\rho \tilde{k}_\rho \left\{ g(\mathbf{n}-\mathbf{s}_\rho) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{s}_\rho)+\sum_i v_i^\rho \tilde{\mu}_i+\sum_j r_j^\rho \mu_j]} + g(\mathbf{n}-\mathbf{r}_\rho) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{r}_\rho)-\sum_i v_i^\rho \tilde{\mu}_i+\sum_j s_j^\rho \mu_j]} \right\} \\
&= \sum_\rho \tilde{k}_\rho \left\{ g(\mathbf{n}-\mathbf{r}_\rho) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{r}_\rho)+\sum_j r_j^\rho \mu_j]} + g(\mathbf{n}-\mathbf{s}_\rho) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{s}_\rho)+\sum_j s_j^\rho \mu_j]} \right\}.
\end{aligned} \tag{4.17}$$

We can write this relation in terms of a summation over each of the complexes $\mathbf{z}$ separately

$$\begin{aligned}
&\sum_{\mathbf{z}} \sum_{\rho:\mathbf{s}_\rho=\mathbf{z}} \tilde{k}_\rho g(\mathbf{n}-\mathbf{s}_\rho) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{s}_\rho)+\sum_i v_i^\rho \tilde{\mu}_i+\sum_j r_j^\rho \mu_j]} + \\
&\sum_{\mathbf{z}} \sum_{\rho:\mathbf{r}_\rho=\mathbf{z}} \tilde{k}_\rho g(\mathbf{n}-\mathbf{r}_\rho) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{r}_\rho)-\sum_i v_i^\rho \tilde{\mu}_i+\sum_j s_j^\rho \mu_j]} = \\
&\sum_{\mathbf{z}} \left\{ \sum_{\rho:\mathbf{r}_\rho=\mathbf{z}} \tilde{k}_\rho g(\mathbf{n}-\mathbf{r}_\rho) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{r}_\rho)+\sum_j r_j^\rho \mu_j]} + \sum_{\rho:\mathbf{s}_\rho=\mathbf{z}} \tilde{k}_\rho g(\mathbf{n}-\mathbf{s}_\rho) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{s}_\rho)+\sum_j s_j^\rho \mu_j]} \right\},
\end{aligned} \tag{4.18}$$

where the subscript $\rho : \mathbf{s}_\rho = \mathbf{z}$ denotes that the sum runs only over all reactions ρ whose product complex $\mathbf{s}_\rho$ equals the complex $\mathbf{z}$. Until now we are considering an arbitrary steady state. Introducing the complex balance constraint, we can impose that, separately, the fluxes across all complexes vanish (that is $J_{+\mathbf{z}} = J_{-\mathbf{z}} \,\forall\, \mathbf{z}$), which, can be rewritten in the following form, given that the complex $\mathbf{z}$ is fixed:

$$\begin{aligned}
&g(\mathbf{n}-\mathbf{z}) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{z})]} \left\{ \sum_{\rho:\mathbf{s}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta[\sum_i v_i^\rho \tilde{\mu}_i+\sum_j r_j^\rho \mu_j]} + \sum_{\rho:\mathbf{r}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta[-\sum_i v_i^\rho \tilde{\mu}_i+\sum_j s_j^\rho \mu_j]} \right\} = \\
&\qquad g(\mathbf{n}-\mathbf{z}) e^{\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{z})]} \left\{ \sum_{\rho:\mathbf{r}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta \sum_j r_j^\rho \mu_j} + \sum_{\rho:\mathbf{s}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta \sum_j s_j^\rho \mu_j} \right\}.
\end{aligned} \tag{4.19}$$

Then, Eq. (4.18) will be satisfied if for each complex, $\mathbf{z}$, Eq. (4.19) is satified. We can now divide both sides by $g(\mathbf{n}-\mathbf{z})\exp(\beta[F(\mathbf{n})-F(\mathbf{n}-\mathbf{z})])$ to obtain

$$\begin{aligned}\sum_{\rho:\mathbf{s}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta[\sum_i (z_i - r_i^\rho)\tilde{\mu}_i + \sum_j r_j^\rho \mu_j]} + \sum_{\rho:\mathbf{r}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta[-\sum_i (s_i^\rho - z_i)\tilde{\mu}_i + \sum_j s_j^\rho \mu_j]} \\ = \sum_{\rho:\mathbf{r}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta\sum_j r_j^\rho \mu_j} + \sum_{\rho:\mathbf{s}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta\sum_j s_j^\rho \mu_j}, \quad (4.20)\end{aligned}$$

where we have substituted $\mathbf{v}_\rho = \mathbf{s}_\rho - \mathbf{r}_\rho$ and, according to each of the sums, one of this terms can be replaced by the complex $\mathbf{z}$. Finally, given that $\mathbf{z}$ is fixed in the previous equation, we can divide the whole expression by $\exp(\beta\sum_i z_i\tilde{\mu}_i)$, yielding

$$\begin{aligned}\sum_{\rho:\mathbf{s}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta[-\sum_i r_i^\rho\tilde{\mu}_i + \sum_j r_j^\rho \mu_j]} + \sum_{\rho:\mathbf{r}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta[-\sum_i s_i^\rho\tilde{\mu}_i + \sum_j s_j^\rho \mu_j]} = \\ \sum_{\rho:\mathbf{r}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta[\sum_j r_j^\rho \mu_j - \sum_i z_i^\rho\tilde{\mu}_i]} + \sum_{\rho:\mathbf{s}_\rho=\mathbf{z}} \tilde{k}_\rho e^{\beta[\sum_j s_j^\rho \mu_j - \sum_i z_i^\rho\tilde{\mu}_i]}. \quad (4.21)\end{aligned}$$

At this point, if we make the following definitions:

$$c_i^{\mathrm{SS}} = \exp(-\beta(\tilde{\mu}_i + \mu_i^0)), \quad (4.22)$$

$$k_{+\rho} = \tilde{k}_\rho \exp(\beta[\sum_i r_i^\rho \mu_i^0 + \sum_j r_j^\rho \mu_j]), \quad (4.23)$$

$$k_{-\rho} = \tilde{k}_\rho \exp(\beta[\sum_i s_i^\rho \mu_i^0 + \sum_j s_j^\rho \mu_j]), \quad (4.24)$$

then we recover the complex balance condition (4.9) for an ideal and deterministic network. Hence, a CRN for which the deterministic steady-state is complex balanced, accepts a steady-state solution of the form (4.15) for its stochastic and non-ideal version. The parameters $\tilde{\mu}_i$ that appear in the steady-state distribution can be obtained using Eq. (4.22) and solving the ideal and deterministic CRN for the steady state concentrations (c_i^{SS}), with rates constants as specified in Eqs. (4.23) and (4.24), where the standard-state chemical potentials and non-equilibrium contributions from the chemostats have been absorbed into the rate constants, as is customary in MAK.

Therefore, the steady-state distribution of a non-ideal complex-balanced CRN has the form of an effective Boltzmann distribution, where the standard-state chemical potentials μ_i^0 are shifted by an amount $\tilde{\mu}_i$. Physically, it is interesting the fact that, for complex-balanced steady states, the free-energetic contribution decouples from the non-equilibrium terms, embodied in the shifted chemical potentials $\tilde{\mu}$. This may not be true for more general steady-states.

This result is similar to Theorem 6.6 of Ref. [Anderson et al., 2010], whose main result is the probability distribution (4.14), but here we have generalised it slightly to include rates of the form (3.25), which includes the function g that could be of interest in phase-separated systems as it modulates the rates depending on the environment. Moreover, our approach clarifies the thermodynamic origin of the rates (3.25) and the steady-state distribution (4.15), as is shown in the next subsection.

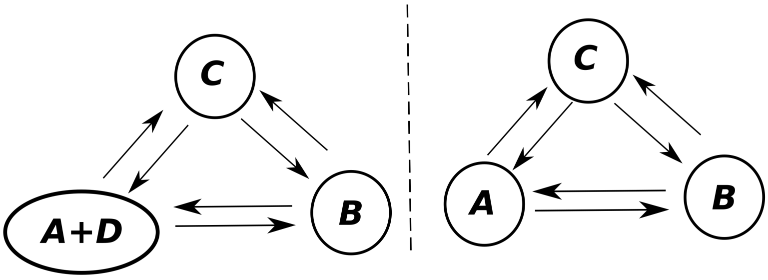


Figure 4.3: Graphical representation of the CRN used in Examples 4 and 5. These two CRNs are necessarily complex balanced ($\delta = 0$ in both cases). Unimolecular networks (as the one on the right) are always complex balanced. In the case of the left-hand side network, despite the presence of a bimolecular complex $A+D$, the network is still complex balanced, since all the A or D particles are created or destroyed through the complex $A + D$, hence, in this case, the steady-state condition implies the complex-balance condition (4.9). The left CRN corresponds to the one used in Example 4; the right one corresponds to Example 5.

Example 4 *Let us consider the following CRN (see Fig. 4.3 for a graphical representation):*

$$A + D \rightleftharpoons B \quad B \rightleftharpoons C \quad C \rightleftharpoons A + D, \tag{4.25}$$

with a free energy taken from a regular solution theory (see Appendix C.1). Given that the number of particles is not conserved, for simplicity, we assume that the solvent volume is conserved and allow the total volume to vary (alternatively, we could have chosen A and D particles as occupying half the volume as C or B). Thus, $V = N + n_{sol} + n_D$, where N is the number of conserved particles $N = n_A + n_B + n_C$ and n_{sol} is the number of solvent particles. Moreover, given any initial condition, the variable n_D will always be ensalved to the variable n_A as they are both created and annihilated together, keeping constant the initial particle difference between the two. For simplicity, we assume that the system is driven out of equilibrium solely by imposing a non-equilibrium chemical potential difference $\Delta\mu_{neq} = \log(5/2)\beta^{-1}$ in the transition from C to $A + D$, which is equivalent to assuming that the reaction $C \rightleftharpoons A + D$ actually is $C + E \rightleftharpoons A + D + F$, where E and F are chemostatted species with chemical potentials satisfying $\mu_E - \mu_F = \log(5/2)\beta^{-1}$.

We assume all the reaction constants are equal to each other and note that the network is necessarily complex-balanced: unimolecular reactions are always complex balanced and here particles A and D are only created or destroyed via the complex $A + D$, therefore the steady-state condition yields that the creation and annihilation rates of the complex $A + D$ must be equal. Equivalently, one could compute the deficiency of the network and check that $\delta = 0$, hence the steady state must be complex balanced. Then, the steady-state of the CRN (4.25), modelled stochastically with dynamics (3.4) and propensity functions (3.25), can be obtained from the following deterministic rate equations (that include entropic terms and the non-equilibrium driving, but not any other term of the free energy):

$$\begin{aligned}
\frac{dn_A}{dt} &= n_B + 2.5n_C - 2n_A n_D \\
\frac{dn_B}{dt} &= n_C + n_A n_D - 2n_B \\
\frac{dn_C}{dt} &= n_B + n_A n_D - 3.5n_C.
\end{aligned}$$

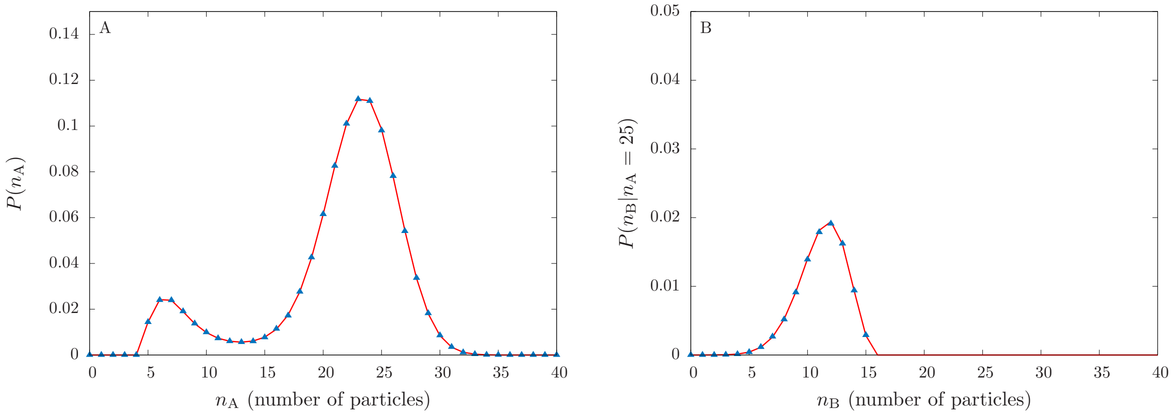


Figure 4.4: Steady-state probability distributions for Example 3. Marginal probability distribution (A) and conditional probability distribution (B) for Example 3 at steady state. The blue triangles are the probabilities obtained from a Gillespie simulation and the red line is the analytical solution given by Eq. (4.26).

Taking as initial conditions $N = 40$ *and* $n_A - n_D = 5$*, the solution of the above system at steady-state is* $n_A^{SS} \simeq 8.1$, $n_B^{SS} \simeq 19.1$, $n_C^{SS} \simeq 12.7$ *and* $n_D^{SS} \simeq 3.1$*. Identifying,* $n_i^{SS} = e^{-\beta\tilde{\mu}_i}$*, we get the values of shifted non-equilibrium chemical potentials and we can obtain a full steady-state probability distribution*

$$\pi_{CB}(\boldsymbol{n}) = \frac{e^{-\beta[F(\mathbf{n})+\sum_i \tilde{\mu}_i n_i]}}{Z}, \tag{4.26}$$

once the free energy F *has been specified and the normalisation constant* Z *computed. Note that there are two conservation laws (*$N = 40$ *and* $n_A - n_D = 5$*), and four chemical species, hence, in practice,* $\pi_{neq}(\boldsymbol{n})$ *is a distribution with only two independent variables.*

Let us consider the following regular solution free energy, that comprises standard-state chemical potentials (taken with respect to that of species B), interactions (parametrised by χ*) and entropic terms:*

$$F(n_A, n_B, n_C, n_D) = \beta^{-1}\left[\log(n_A! n_B! n_C! n_D! n_{sol}!) - \log V!\right] \\ + n_A\mu_A^0 + n_C\mu_C^0 + n_D\mu_D^0 + \chi\frac{n_A n_C}{V}, \tag{4.27}$$

where $V = n_A + n_B + n_C + n_D + n_{sol}$ *and* n_{sol} *is the number of solvent molecules. For the parameters* $\chi = 10\beta^{-1}$ *and* $\mu_A^0 = \mu_C^0 = \mu_D^0 = \log(2)\beta^{-1}$*, we obtain the steady-state probability depicted in Fig. 4.4, that matches closely the one obtained from a simulation of the same CRN using the Gillespie algorithm [Gillespie, 1977]. Simulations were started in parallel from random poissonian initial conditions (provided they satisfy the constraints above) and the samples were obtained after allowing the simulations to relax to the steady state.* □

4.3.2 Thermodynamic interpretation of $\tilde{\mu}$

The direct consequence of the non-equilibrium driving of the complex-balanced CRN is the appearance of a term $\tilde{\mu}$ in the stationary distribution (4.15). Effectively, this term shifts the standard-state chemical potential μ^0 by a constant. However, the thermodynamic implications of this shift in the standard-state chemical potentials can be further clarified.

As shown in Section 3.3.2, one can bound the work required to make a non-equilibrium transformation of the system by the difference in free energies between these non-equilibrium states. Here, we consider a simpler case of obtaining a non-equilibrium complex-balanced steady state from an initial equilibrium state, by connecting the system to several chemostats. Then, in the stochastic description, the average work required to do that is bounded by

$$\langle W^{\text{chem}} \rangle \geq \langle G_f \rangle - \langle G_{\text{eq}} \rangle, \tag{4.28}$$

where G is the stochastic free energy defined by Eq. (3.27).

At equilibrium, the probability distribution is just the canonical distribution (or grand-canonical distribution in case the particle numbers are not conserved), which is, up to a normalisation constant (the partition function Z), the exponential of the free energy F. Thus, from Eq. (3.27), $G_{\text{eq}} = -k_{\text{B}} T \log Z$, as expected from equilibrium statistical physics and where Z is the equilibrium partition function. However, for a non-equilibrium complex-balanced state,

$$\begin{aligned}\langle G_{\text{CB}} \rangle &= \sum_{\mathbf{n}} \pi_{\text{CB}}(\mathbf{n}) \sum_i \tilde{\mu}_i n_i - k_{\text{B}} T \log Z_{\text{CB}} \\ &= \sum_i \tilde{\mu}_i \langle n_i \rangle - k_{\text{B}} T \log Z_{\text{CB}},\end{aligned} \tag{4.29}$$

where Z_{CB} is the complex-balanced normalisation constant, which is not necessarily equal to the equilibrium partition function Z, and $\langle n_i \rangle$ is the average number of i particles, according to the complex-balanced distribution π_{CB}.

Therefore, the work required for such a transformation is bounded by

$$\langle W^{\text{chem}} \rangle \geq \sum_i \tilde{\mu}_i \langle n_i \rangle_f + k_{\text{B}} T \log \frac{Z}{Z_{\text{CB}}}, \tag{4.30}$$

where $\langle n_i \rangle_f$ is the average number of particles i at the final non-equilibrium complex-balanced steady state. In the previous relation, the ratio of partition functions is reminiscent of the equilibrium bound for a process, which is only attained if the transformation is fully reversible ($\Delta S^{\text{tot}} = 0$) during the transformation. However, the appearance of the $\sum_i \tilde{\mu}_i \langle n_i \rangle_f$ is exclusively due to the non-equilibrium work performed on the system, since $\tilde{\mu}_i = 0$ in the absence of such work.

Thus, this inequality suggests a thermodynamic interpretation of the shifted chemical potentials $\tilde{\mu}$, since the bound depends crucially on them and is the only inherently non-equilibrium contribution to the bound.

4.4 Lyapunov function for complex-balanced steady states

Under fairly general conditions, the logarithm of the steady-state probability distribution in the stochastic model is a Lyapunov function of the deterministic one, provided an appropriate scaling is used to bridge between the different volume and particle-number scales [Ge and Qian, 2016]. In particular, this has been shown for ideal and complex-balanced CRNs [Anderson et al., 2015]. Here, we show that in non-ideal complex-balanced CRNs the following function is minimised by the dynamics

$$\mathscr{L}(\mathbf{c}) = -\lim_{V \to \infty} \frac{1}{V} \log(\pi_{\text{CB}}(\mathbf{n})) = \beta \left(f(\mathbf{c}) + \sum_i \tilde{\mu}_i c_i \right) + \log(Z)/V, \tag{4.31}$$

where $c_i = n_i/V$ and $f(\mathbf{c}) = F(\mathbf{n})/V$. Indeed, this approach generalises to a certain extent the results of Anderson and Nguyen [Anderson and Nguyen, 2019] for product-form stationary states of CRNs. We abuse slightly the terminology by calling this function a Lyapunov function as we do not prove it is strictly positive or bounded from below, we will only prove that it is minimised by the dynamics. Nevertheless, the fact that this function is minimised is sufficient for our purposes.

Since the normalisation constant Z does not depend on time (it only depends on the non-equilibrium steady state), the time derivative of the Lyapunov function can be written as

$$\begin{aligned}\frac{d\mathscr{L}}{dt} =& \sum_{i'} \frac{\partial \mathscr{L}}{\partial c_{i'}} \frac{\partial c_{i'}}{\partial t} \\ =& \beta \sum_{i'} (\mu_{i'} + \tilde{\mu}_{i'}) \left[\sum_{\rho} v^{\rho}_{i'} k_\rho g(\mathbf{c}) (e^{\beta[\sum_i r^\rho_i \mu_i + \sum_j r^\rho_j \mu_j]} - e^{\beta[\sum_i s^\rho_i \mu_i + \sum_j s^\rho_j \mu_j]}) \right], \end{aligned} \tag{4.32}$$

where we have used Eq. (3.7) with currents given by Eq. (3.26). Note we have used two different dummy indices, i and i', to make it clear to which sum each variable corresponds to. We can recast the previous expression into

$$\begin{aligned}\frac{d\mathscr{L}}{dt} =& \sum_{\rho} \sum_{i'} k_\rho g(\mathbf{c}) (\mu_{i'} + \tilde{\mu}_{i'}) (s^\rho_{i'} - r^\rho_{i'}) e^{\beta[\sum_i r^\rho_i (\mu_i + \tilde{\mu}_i) - \sum_i r^\rho_i \tilde{\mu}_i + \sum_j r^\rho_j \mu_j]} \\ &+ \sum_{\rho} \sum_{i'} k_\rho g(\mathbf{c}) (\mu_{i'} + \tilde{\mu}_{i'}) (r^\rho_{i'} - s^\rho_{i'}) e^{\beta[\sum_i s^\rho_i (\mu_i + \tilde{\mu}_i) - \sum_i s^\rho_i \tilde{\mu}_i + \sum_j s^\rho_j \mu_j]}, \end{aligned} \tag{4.33}$$

by adding and subtracting the $\tilde{\mu}$-terms in the arguments of the exponential functions. We now can repeatedly apply the inequality[2] $e^a(b-a) \leq e^b - e^a$ to the sums of chemical potentials to arrive at:

$$\begin{aligned}\frac{d\mathscr{L}}{dt} \leq& \sum_{\rho} k_\rho g(\mathbf{c}) e^{\beta[\sum_j r^\rho_j \mu_j - \sum_i r^\rho_i \tilde{\mu}_i]} \left[e^{\beta \sum_i (\mu_i + \tilde{\mu}_i) s^\rho_i} - e^{\beta \sum_i r^\rho_i (\mu_i + \tilde{\mu}_i)} \right] \\ &+ \sum_{\rho} k_\rho g(\mathbf{c}) e^{\beta[\sum_j s^\rho_j \mu_j - \sum_i s^\rho_i \tilde{\mu}_i]} \left[e^{\beta \sum_i (\mu_i + \tilde{\mu}_i) r^\rho_i} - e^{\beta \sum_i s^\rho_i (\mu_i + \tilde{\mu}_i)} \right]. \end{aligned} \tag{4.34}$$

This previous expression can now be separated in terms of the different complexes in the system:

$$\begin{aligned}&\frac{d\mathscr{L}}{dt} \leq \\ &\sum_{\mathbf{z}} g(\mathbf{c}) \Bigg\{ \sum_{\rho:\mathbf{s}^\rho = \mathbf{z}} k_\rho e^{\beta[\sum_j r^\rho_j \mu_j - \sum_i r^\rho_i \tilde{\mu}_i + \sum_i (\mu_i + \tilde{\mu}_i) s^\rho_i]} - \sum_{\rho:\mathbf{r}^\rho = \mathbf{z}} k_\rho e^{\beta[\sum_j r^\rho_j \mu_j - \sum_i r^\rho_i \tilde{\mu}_i + \sum_i r^\rho_i (\mu_i + \tilde{\mu}_i)]} \\ &\quad + \sum_{\rho:\mathbf{r}^\rho = \mathbf{z}} k_\rho e^{\beta[\sum_j s^\rho_j \mu_j - \sum_i s^\rho_i \tilde{\mu}_i + \sum_i (\mu_i + \tilde{\mu}_i) r^\rho_i]} - \sum_{\rho:\mathbf{s}^\rho = \mathbf{z}} k_\rho e^{\beta[\sum_j s^\rho_j \mu_j - \sum_i s^\rho_i \tilde{\mu}_i + \sum_i s^\rho_i (\mu_i + \tilde{\mu}_i)]} \Bigg\}, \end{aligned} \tag{4.35}$$

where, invoking complex balancing as in the previous section and dividing by $\exp[\beta \sum_i z_i(\mu_i +$

---

[2] The inequality stems from $1 + x \leq e^x$, $\forall x \in \mathbb{R}$ and setting $x = b - a$. We use it by identifying (for the first term) $a = \sum_i (\mu_i + \tilde{\mu}_i) r^\rho_i$ and $b = \sum_i (\mu_i + \tilde{\mu}_i) s^\rho_i$ (and conversely for the second).

$\tilde{\mu}_i)]$, we have an equality for each complex $\mathbf{z}$

$$\sum_{\rho:\mathbf{s}^\rho=\mathbf{z}} k_\rho e^{\beta[\sum_j r_j^\rho \mu_j - \sum_i r_i^\rho \tilde{\mu}_i]} - \sum_{\rho:\mathbf{r}^\rho=\mathbf{z}} k_\rho e^{\beta[\sum_j r_j^\rho \mu_j - \sum_i r_i^\rho \tilde{\mu}_i]} + \sum_{\rho:\mathbf{r}^\rho=\mathbf{z}} k_\rho e^{\beta[\sum_j s_j^\rho \mu_j - \sum_i s_i^\rho \tilde{\mu}_i]} - \sum_{\rho:\mathbf{s}^\rho=\mathbf{z}} k_\rho e^{\beta[\sum_j s_j^\rho \mu_j - \sum_i s_i^\rho \tilde{\mu}_i]} = 0, \quad (4.36)$$

equality that is given by the complex-balance condition for deterministic MAK networks [Eq. (4.9)] together with Eqs. (4.22) to (4.24). Summing over all complexes yields the desired inequality:

$$\frac{d\mathscr{L}}{dt} \leq 0, \quad (4.37)$$

and, thus, $\mathscr{L}$ is minimised (or remains unchanged) along a trajectory for a deterministic complex-balanced non-ideal CRN.

Therefore, unlike in Section 2.2.5 where we minimised the free energy F since we were looking for the equilibrium state, here we minimise the function $\mathscr{L}$ that includes the free energy density f and the standard-state chemical potentials μ_i^0 have been shifted by $\tilde{\mu}_i$ due to the work performed on the system by the chemostats. Consequently, the CRN does not relax to thermodynamic equilibrium and, at steady state, is able to sustain chemical currents.

4.5 Lyapunov functionals for spatially heterogeneous solutions

To be able to describe phase-separating systems we need to take into account the spatial degrees of freedom. In the deterministic description, concentrations are now a function of space, $c_i(\mathbf{x})$, within a volume Ω and the free energy is a functional of these concentrations, $F[\mathbf{c}(\mathbf{x})]$.

The time derivative of the concentration fields $c_i(\mathbf{x})$ is given by the following reaction-diffusion equation

$$\frac{dc_i(\mathbf{x})}{dt} = -\nabla \cdot \mathbf{J}_i + \sum_\rho v_i^\rho (J_{+\rho} - J_{-\rho}), \quad (4.38)$$

where the first term in the right-hand side (RHS) of the equation corresponds to diffusion and the second one to the chemical reactions. As in the linear irreversible thermodynamics framework [Groot and Mazur, 1983, Kondepundi and Prigogine, 2015], the driving force of the diffusion current $\mathbf{J}_i$ is the gradient of chemical potentials $\nabla\mu_i$. More precisely, the diffusion currents take the following form $\mathbf{J}_i = -\sum_j \mathscr{M}_{ij} \nabla \mu_j$, where $\mathscr{M}_{ij}$ is the mobility matrix. We assume no-flux boundary conditions $\mathbf{J}_i|_{\mathbf{x}=\partial\Omega} = 0$ for the non-chemostatted species, where $\partial\Omega$ denotes the boundaries of the volume Ω.

We now proceed to show that the Lyapunov function obtained above for a complex-balanced CRN is also a Lyapunov functional if we consider the diffusion of solutes and solvent within the solution. Then, the time evolution of the functional $\int \mathscr{L}(\mathbf{x}) d\mathbf{x}$ becomes

$$\begin{aligned} \int \frac{d\mathscr{L}}{dt} d\mathbf{x} &= \sum_i \int d\mathbf{x} \frac{d\mathscr{L}}{dc_i} \frac{dc_i}{dt} \\ &= \sum_i \int d\mathbf{x}\, \beta(\mu_i(\mathbf{x}) + \tilde{\mu}_i) \left[-\nabla \cdot \mathbf{J}_i + \sum_\rho v_i^\rho (J_{+\rho} - J_{-\rho}) \right], \quad (4.39) \end{aligned}$$

where $\mu_i(\mathbf{x}) = \delta F[\mathbf{c}(\mathbf{x})]/\delta c_i(\mathbf{x})$ is the local chemical potential and, in the second equality, Eq. (4.38) has been used. From the results in the previous section, we know that the contribution of the second term in the square brackets of the RHS of the equation to the time derivative of $\mathscr{L}$ is negative, since it corresponds to the chemical reactions. Therefore, to prove that $\int \mathscr{L}(\mathbf{x})d\mathbf{x}$ is minimised by the dynamics is sufficient to show that the contribution of the first term in the RHS is also negative.

We note that the contribution of diffusion to the time evolution of $\mathscr{L}$ can be written as

$$\sum_i \int d\mathbf{x}\,(\mu_i(\mathbf{x}) + \tilde{\mu}_i)\nabla \cdot \mathbf{J}_i = \sum_i \int d\mathbf{x}\,[\nabla \cdot \{(\mu_i(\mathbf{x}) + \tilde{\mu}_i)\mathbf{J}_i\} - \nabla(\mu_i(\mathbf{x}) + \tilde{\mu}_i) \cdot \mathbf{J}_i]\,. \quad (4.40)$$

For the first term, we have that

$$\int d\mathbf{x}\nabla \cdot \{(\mu_i(\mathbf{x}) + \tilde{\mu}_i)\mathbf{J}_i\} = \int_{\partial\Omega} dS(\mu_i(\mathbf{x}) + \tilde{\mu}_i)\mathbf{J}_i \cdot \hat{\mathbf{n}} = 0, \quad (4.41)$$

where $\hat{\mathbf{n}}$ is the verctor normal to the surface $\partial\Omega$. In the first equality we have used the divergence theorem and, in the second one, the fact that we are considering no-flux boundary conditions, although it may be generalised to other appropriate boundary conditions. Regarding the second term in Eq. (4.40), we have that

$$-\nabla(\mu_i(\mathbf{x}) + \tilde{\mu}_i) \cdot \mathbf{J}_i = -\nabla\mu_i(\mathbf{x}) \cdot \mathbf{J}_i = \dot{S}_{\text{diff}} \geq 0, \quad (4.42)$$

where $\dot{S}_{\text{diff}}$ is the entropy production rate due to diffusion. In the first equality we have used the fact that $\tilde{\mu}_i$ are constants (hence $\nabla\tilde{\mu}_i = 0$). Then, we can identify the remaining terms as the entropy production due to the diffusion process, which necessarily has to be greater or equal to zero, provided the Onsager reciprocal relations for the mobility matrix $\mathscr{M}_{ij}$ are met [Groot and Mazur, 1983, Kondepundi and Prigogine, 2015].

We finally obtain that

$$\int \frac{d\mathscr{L}}{dt} d\mathbf{x} \leq 0, \quad (4.43)$$

as both the contribution from chemical reactions and from diffusion are negative. Thus, a non-ideal complex-balanced reaction-diffusion system will minimise the functional $\int \mathscr{L} d\mathbf{x}$, which we can now use to obtain information about the steady state, as done with the free energy F for systems that relax to thermodynamic equilibrium.

Example 5 *Let us now consider the following CRN (see Fig. 4.3 for a graphical representation):*

$$A \leftrightharpoons B \quad B \leftrightharpoons C \quad C \leftrightharpoons A, \quad (4.44)$$

with a free energy taken from a regular solution theory, as before. Again, for simplicity, we assume that the system is driven out of equilibrium solely by imposing a non-equilibrium chemical potential difference $\Delta\mu$ for the transition from C to A.

We take all the reaction constants equal to each other and note that the network is necessarily complex-balanced because all chemical reactions are unimolecular. Then, the steady-state of the CRN (4.44), modelled stochastically with dynamics (3.4) and propensity functions (3.25), can be obtained from the following *deterministic rate equations (that include entropic terms and the non-equilibrium driving, but not any other term of the free energy):*

$$\begin{aligned}\frac{dc_A}{dt} &= c_B + e^{-\beta\Delta\mu}c_C - 2c_A\\ \frac{dc_B}{dt} &= c_C + c_A - 2c_B\\ \frac{dc_C}{dt} &= c_B + c_A - (1+e^{-\beta\Delta\mu})c_C.\end{aligned}$$

The solution of the above system at steady state is

$$\begin{aligned}c_A^{SS} &= c_C^{SS}\left(\frac{1}{2}+e^{-\beta\Delta\mu}\right)\frac{2}{3}\\ c_B^{SS} &= c_C^{SS}\left[\left(\frac{1}{2}+e^{-\beta\Delta\mu}\right)\frac{1}{3}+\frac{1}{2}\right].\end{aligned}$$

Identifying, $c_i^{SS} = e^{-\beta\tilde{\mu}_i}$, and noting that we can express them with respect to that of species C, we get the values of shifted non-equilibrium chemical potentials $\tilde{\mu}_i$ and we can obtain the Lyapunov function of the system

$$\mathscr{L}(\boldsymbol{c}) = \beta f(\boldsymbol{c}) - c_A \log\left[\left(1+2e^{-\beta\Delta\mu}\right)\frac{1}{3}\right] - c_B\log\left[\frac{1}{2}+\left(\frac{1}{2}+e^{-\beta\Delta\mu}\right)\frac{1}{3}\right] - \log Z, \tag{4.45}$$

where the non-equilibrium partition function has not been explicitly computed but, since it is a constant term along the dynamics, it will not alter the location of the minima of $\mathscr{L}$ in the space of concentrations **c**.

Let us consider the following regular solution free energy in the deterministic limit, that comprises interaction and entropic terms and includes Cahn-Hilliard terms in order to model smoothly the potentially non-uniform system:

$$f(c_A, c_B, c_C) = \beta^{-1}\sum_{\alpha=A,B,C} c_\alpha \log c_\alpha + \chi_{22}c_A^2 + \chi_{12}c_Ac_B + \sum_{i,i'}\kappa_{i,i'}(\nabla c_i \nabla c_{i'}), \tag{4.46}$$

where $\chi_{22} = -2$, $\chi_{12} = -7$ and $\kappa_{i,i'}$ are the Cahn Hilliard coefficients [Cahn and Hilliard, 1958], which, as introduced in Section 2.2.1, set the free-energetic cost of interfaces and their width. This free energy and the reaction-diffusion equation (4.38) can describe the dynamics of the system, which minimise its Lyapunov functional (4.45), as seen in the Figure 4.5 for a one dimensional system. In particular, the upper panel shows that the Lyapunov functional (4.45) is minimised by the numerical time evolution of the reaction-diffusion system, the middle panel shows the concentration profiles at steady state and the lower panel the net chemical current from species C to A at steady state as a function of the position x. □

4.6 Phenomenology of a complex-balanced mixture

Since the Lyapunov functional for complex-balanced systems found in the previous section is minimised by the dynamics, it carries plenty of information about the steady-state in which the macroscopic system will settle. The phase diagram of a mixture, in the case of phase separation, states whether for a certain parameter set the steady state of the system is homogeneous or there are multiple different phases. Thus, the phase diagram follows from the Lyapunov functional

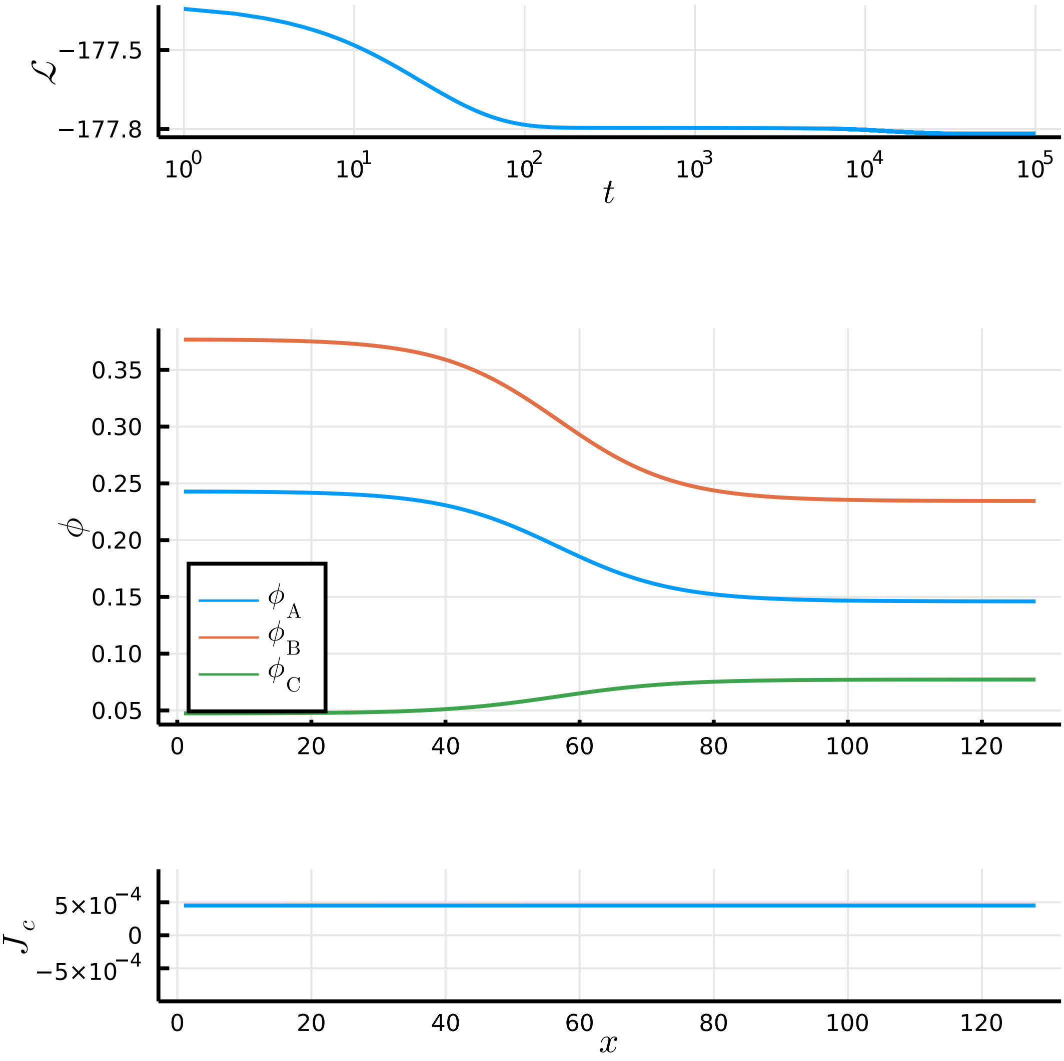


Figure 4.5: The Lyapunov function of a complex-balanced CRN is minimised by its dynamics. Numerical results for the CRN (4.44) obtained from perturbing a homogeneous solution and integrating Eq. (4.38) forward in time until it reaches a steady state. *Top panel*: evolution of the value of the Lyapunov functional with time. This functional is minimised with the dynamics but it does not attain the value 0 for two reasons: First, the constant term $-\log(Z)$ has been neglected. Second, the dynamics may not have reached the global minimum of $\mathscr{L}$ but only a local one. *Middle panel*: Steady-state configuration of the system, displaying the volume fractions of each species as a function of the spatial coordinate $\phi(x)$. In this case, the system reaches at a steady state where there is phase separation. *Lower panel*: Net reaction flux at steady state for the third reaction in the CRN (4.44) as a function of the spatial coordinate, showing that the system is not detailed-balanced. Note that, while in the upper panel, the horizontal dimension refers to time, it is no longer the case for the lower ones, where it represents the space in a one-dimensional system.

(4.31) in the same way it would do for a classical free energy: the phase coexistence lines are obtained from the minimisation of $\mathscr{L}$ with the appropriate constraints (particle conservation, in most of the cases we consider).

4.6.1 Phase diagram of a non-ideal complex-balanced solution

In this section we will consider a non-ideal solution with the CRN (4.44) in the deterministic limit, and obtain its phase diagram. Therefore, we need to minimise the Lyapunov functional found in Example 5 subject to the particle conservation constraint $V\phi_N = \int d\mathbf{x}[\phi_A(\mathbf{x}) + \phi_B(\mathbf{x}) + \phi_C(\mathbf{x})]$, where $\phi_i(\mathbf{x})$ are the continuous volume fraction fields of each of the species at point $\mathbf{x}$ in space (as is customary when discussing phase separation) and ϕ_N is the overall volume fraction of solutes. Then, the Lagrangian that needs to be minimised is

$$L = \int d\mathbf{x}\, \mathscr{L}(\phi) - \lambda \left(V\phi_N - \int d\mathbf{x}\, [\phi_A(\mathbf{x}) + \phi_B(\mathbf{x}) + \phi_C(\mathbf{x})] \right), \tag{4.47}$$

where λ is the Lagrange multiplier associated with the conservation of solute particles. If we assume the bulk free energy of the homogeneous phases is large compared to the free-energetic cost of the interfaces, then we can drop the continuity of the concentration fields in space and have a simpler description of the system only in terms of two homogeneous but different phases. Considering a finite interface would, in general, yield a small correction to the phase diagram that we find below. Then, we need to minimise the following Lagrangian:

$$\begin{aligned} L = {} & V^{(1)}\mathscr{L}(\phi^{(1)}) + (V - V^{(1)})\mathscr{L}(\phi^{(2)}) \\ & - \lambda \left[V\phi_N - V^{(1)}(\phi_A^{(1)} + \phi_B^{(1)} + \phi_C^{(1)}) - (V - V^{(1)})(\phi_A^{(2)} + \phi_B^{(2)} + \phi_C^{(2)}) \right], \end{aligned} \tag{4.48}$$

where the superscripts (1) and (2) stand for the different phases, $V^{(1)}$ is the volume of one of the phases, and we recall that $\mathscr{L}$ is the one obtained in Example 5. For simplicity, we consider the following free energy

$$f(\phi) = \sum_{\alpha} \phi_\alpha \log \phi_\alpha - \chi \phi_A^2. \tag{4.49}$$

Minimisation of the Lagrangian (4.48), yields the phase diagram in Fig. 4.6 (for details on the minimisation, see Appendix C.3). Regions **2** and **3** in the phase diagrams correspond to parts of the parameter space where phase separation occurs, as it can be seen from the concentration profiles of the upper panels, while in regions **1** and **4** the homogeneous configuration is stable. These results are reminiscent of equilibrium phase diagrams but we recall that here there are net chemical fluxes being maintained at steady state due to the non-equilibrium driving. In fact, from the form of the free energy (4.49), we see that species A drives phase separation. Thus, since the overall concentrations are modulated by the non-equilibrium work $\Delta\mu$, whether the steady state displays one phase or a coexistence of phases also depends on the value of $\Delta\mu$.

4.6.2 General features of the steady state

As seen above, the phenomenology of a complex-balanced system does not change much with respect to a non-ideal solution in equilibrium, but the non-equilibrium terms can move the system

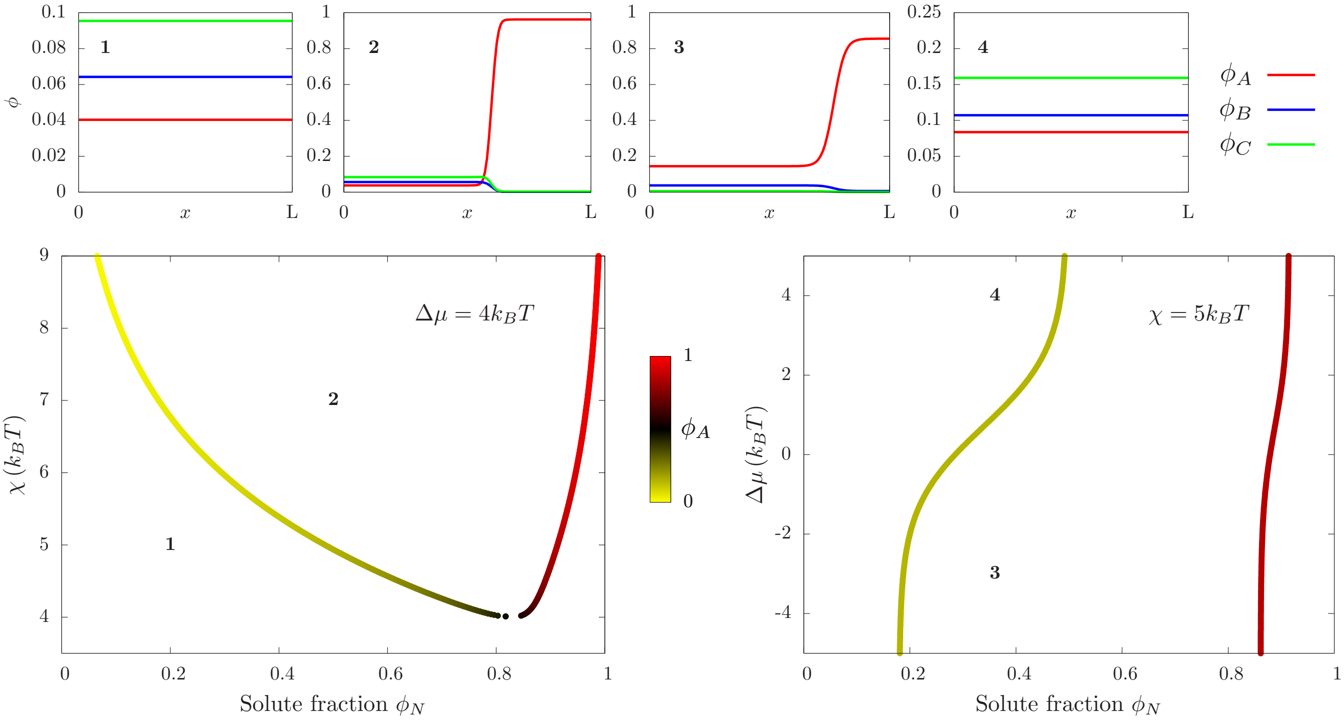


Figure 4.6: Phase diagrams for a chemically reactive non-ideal mixture. *Lower panels*: Phase diagrams obtained by minimisation of Eq. (4.48) as a function of total solute fraction ϕ_N and interaction parameter χ or non-equilibrium driving $\Delta\mu$ (as defined in Example 5). The color code indicates the amount of A particles along the phase coexistence lines (volume fraction). *Upper panels*: Each of the panels has associated a number which states their corresponding point in the phase diagrams. These panels depict the steady-state volume fraction profiles of the system in a one dimensional space, x, from 0 to L. These results were obtained by numerically integrating in time the reaction-diffusion equations, as in Fig. 4.5, with the parameters set by their point in the phase diagram. Note that, in the upper panels, the width of the interface is finite to ease the numerical integration, unlike in the lower panels, where the phase diagram was obtained for vanishing interfaces.

across the phase diagram, enabling the switching on and off of phase separation. This can be clearly seen from the minimisation of L [Eq. (4.47)] that involves taking functional derivatives with respect to the volume fractions $\phi_i(\mathbf{x})$:

$$\frac{\delta L}{\delta \phi_i(\mathbf{x})} = \mu_i(\mathbf{x}) + \tilde{\mu}_i - \lambda = 0, \tag{4.50}$$

which implies that the chemical potential is homogeneous (since $\tilde{\mu}_i$ and λ are constants) as it would be the case for a mixture at thermal equilibrium [*cf.* Section 2.2.5, and, in particular Eq. (2.38)]. Hence, if we define non-equilibrium patterning as the type of steady-state patterning characterised by displaying diffusion currents, it is excluded for solutions driven out of equilibrium by a complex-balanced CRN since there cannot be any diffusion currents at steady state (the driving force of the diffusion currents $\nabla \mu_i = 0$). Nevertheless, as shown in Example 5, the system can sustain local chemical currents at steady state, due to the chemical work supplied by the chemostats. However, these chemical currents do not couple to the diffusion process to produce diffusion currents at steady state.

One of the consequences of our analysis is the fact that breaking complex-balance is a minimal requirement for non-equilibrium pattern formation in non-ideal solutions with local interactions, at least when modelled in a thermodynamically consistent way, which generalises previous considerations [Carati and Lefever, 1997, Kirschbaum and Zwicker, 2021]. Nevertheless, patterning in non-ideal mixtures can be achieved by the addition of surfactants or long-range interactions, which can yield states of suppressed Ostwald ripening [Tarzia and Coniglio, 2006, Pham et al., 2010].

Our analysis highlights the fact that in non-ideal media complex balance (as defined in the previous chapters) can be broken in two different ways. Breaking complex balance can be achieved by a suitable CRN topology, e.g. [Carati and Lefever, 1997], but, in a system that exhibits coexistence between two different phases, one can also break complex balance by allowing the rates of the reactions to depend on their local environment in different ways (for example, in Ref. [Kirschbaum and Zwicker, 2021] they found a patterned steady-state by allowing one reaction to depend on the concentration of an enzyme that localises preferentially at one of the phases). Here, to prove the form of the steady-state distribution and the Lyapunov function for complex-balanced systems, we assumed that, even if the reaction were allowed to depend on their environment via a function g_ρ, this function had to be the same for all reactions. Hence, if different reactions have different g_ρ-functions, our results may not hold and, as demonstrated in Ref. [Kirschbaum and Zwicker, 2021], patterning may occur.

In biological cells, phase separation has been hypothesised to perform many functions, such as, accelerating biochemical reactions within the condensate irrespective of the rate of the reaction in the dilute phase [Lyon et al., 2021]. Our work also implies that, in order to accelerate a reaction only in one of the phases, breaking complex balance is necessary. Indeed, in a complex-balanced system, the chemical potential of every species is the same in both phases (in the condensate and outside it and, in fact, according to Eq. (4.50), it is perfectly homogeneous across space). Then, given that the force driving the chemical reactions are the chemical potentials, the reaction rates in both phases are related, making it impossible to regulate the rates of chemical reactions in each phase independently and suggesting that breaking complex balance in one of the two ways outlined above is crucial. However, there are other functions that phase separation in biological cells can perform with complex-balanced CRNs or even at thermodynamic equilibrium, like sequestration of molecules or noise buffering [Klosin et al., 2020].

Complex balance is known to be a key feature of the CRN that determines not only its behaviour [Anderson et al., 2010] but also its thermodynamic properties [Polettini et al., 2015]. Here, we have further stressed the connection between the topology of the CRN and the thermodynamically consistent structure of the physical system to generalise results from ideal CRNs and explore non-equilibrium thermodynamics of complex-balanced non-ideal CRNs. However, for non-complex-balanced systems little is known still and, given our results, further research regarding the behaviour of this type of networks will be of the utmost importance both from the physical and the biological point of view. Some steps in this direction will be taken in the chapter that follows.

Chapter 5

Beyond complex balancing?

In this work we have shown that for chemically-reactive non-ideal solutions we can obtain results analogous to those of ideal CRNs, provided the system is modelled in a thermodynamically consistent way. This implies that the rates of the chemical reactions will be affected by the interactions between the species in the system and, therefore, MAK is no longer a valid description for the dynamics of the CRN. In the previous chapter, generalising MAK for a non-ideal solution, we obtained the steady-state probability distribution for a stochastic complex-balanced CRN and the Lyapunov function of its deterministic counterpart, which specifies the phase diagram of the system.

Our results are of particular importance for non-equilibrium phase-separating systems, where a wide variety of non-equilibrium behaviour can arise [Weber et al., 2019]. However, they were limited to a class of CRNs known as complex-balanced and, as discussed above, many of the inherently non-equilibrium phenomena can only occur for systems that are not complex balanced. Examples of these are: the modulations of reaction rates in one phase (e.g. a phase-separated condensate in a cell) independently of the rest of the system [Lyon et al., 2021] or non-equilibrium patterning [Carati and Lefever, 1997]. Thus, going beyond complex-balanced networks seems imperative to understand the phenomenology non-ideal solutions may present in biological cells.

This is a very challenging task and, in the present chapter, we will limit ourselves to analysing a particular example and provide directions towards which future research should be directed.

5.1 Patterning when complex balance is broken

As an example of behaviour that can arise when complex balance is broken, we consider a simplified version of the model analysed in Ref. [Kirschbaum and Zwicker, 2021]: A ternary mixture (two chemical species and a solvent) where the two chemical species are exchanged via chemical reactions of the form $A \leftrightharpoons B$.

First we will consider the equilibrium behaviour of the system. Afterwards, we will compare it to its behaviour out of thermodynamic equilibrium while restricted to complex-balanced networks. Finally, we will explore the effects of breaking complex balance out of equilibrium.

Equilibrium behaviour of the mixture

Let us assume that the system is not chemically reactive and is well described by the following equilibrium free energy density

$$\frac{f}{k_{\mathrm{B}}T} = c_A \log(c_A) - c_A + c_B \log(c_B) - c_B - \chi c_A^2 + \frac{\kappa}{2}\left(|\nabla c_A|^2 + |\nabla c_B|^2\right), \tag{5.1}$$

where κ are the Cahn-Hilliard coefficients that penalise the creation of interfaces. The eigenvalues of the Hessian matrix of the free energy density yield the stability of the homogeneous solution and can be used to obtain the parameter region for spinodal decomposition (see Chapter 2, Section 2.3.1). Considering a homogeneous system (where $\nabla c_i = 0$), we obtain that the smallest (most negative) eigenvalue of the Hessian matrix is given by

$$\lambda_- = \frac{1}{c_A} - 2\chi, \tag{5.2}$$

which controls the stability of the system and, hence, whether spinodal decomposition occurs. Therefore, when $2\chi > \frac{1}{c_A}$, the homogeneous configuration of the system is unstable and spinodal decomposition will occur.

Complex-balanced behaviour

Let us now consider the case of the exchange $A \leftrightharpoons B$ that can happen via two different reaction pathways: one with a non-equilibrium contribution from the chemostats $\Delta\mu$ and the other one without any external work. Then, one can write the resulting reaction diffusion-equations for the system:

$$\frac{\partial c_A}{\partial t} = D_A \nabla^2 \mu_A + k_1(e^{\beta(\mu_B + \Delta\mu)} - e^{\beta\mu_A}) + k_2(e^{\beta\mu_B} - e^{\beta\mu_A}) \tag{5.3}$$

$$\frac{\partial c_B}{\partial t} = D_B \nabla^2 \mu_B - k_1(e^{\beta(\mu_B + \Delta\mu)} - e^{\beta\mu_A}) - k_2(e^{\beta\mu_B} - e^{\beta\mu_A}), \tag{5.4}$$

where $\mu_A = \delta f / \delta c_A$ and $\mu_B = \delta f / \delta c_B$ are the chemical potentials of species A and B, respectively. A concentration homogeneous profile is a steady-state solution to Eqs. (5.3) and (5.4), provided the chemical potentials satisfy

$$e^{\beta\mu_A} = \frac{k_1 e^{\beta\Delta\mu} + k_2}{k_1 + k_2} e^{\beta\mu_B}. \tag{5.5}$$

In order to analyse the stability of this solution we perform a Linear Stability Analysis (LSA) (for an introduction to LSA see e.g. [Kondepundi and Prigogine, 2015]). Briefly, we consider small deviations from this homogeneous solution (δc_A) and analyse them in Fourier space to find if these perturbations decay back to the homogeneous solution or they are amplified and the system becomes unstable. To linear order in δc_A and δc_B, in Fourier space we have

$$\frac{\partial \delta c_A}{\partial t} = -k^2 D_A \delta\mu_A + k_1(e^{\beta(\mu_B + \Delta\mu)}\delta\mu_B - e^{\beta\mu_A}\delta\mu_A) + k_2(e^{\beta\mu_B}\delta\mu_B - e^{\beta\mu_A}\delta\mu_A) \tag{5.6}$$

$$\frac{\partial \delta c_B}{\partial t} = -k^2 D_B \delta\mu_B - k_1(e^{\beta(\mu_B + \Delta\mu)}\delta\mu_B - e^{\beta\mu_A}\delta\mu_A) - k_2(e^{\beta\mu_B}\delta\mu_B - e^{\beta\mu_A}\delta\mu_A), \tag{5.7}$$

where k is the Fourier wave number and $\delta\mu_A = \delta c_A \left[\frac{1}{c_A} - 2\chi + \kappa k^2\right]$, $\delta\mu_B = \delta c_B \left[\frac{1}{c_B} + \kappa k^2\right]$. Solving this linear system, amounts to finding the eigenvalues and eigenvectors of the Jacobian matrix; the stability of the homogeneous solution being ensured if, for every k, both eigenvalues are negative.

The sign of the determinant encodes information about the sign of the eigenvalues, since it is equal to their product. Thus, in order to assess the stability of the homogeneous system, we analyse the sign of the determinant of the Jacobian matrix (J) of the system defined by Eqs. (5.6) and (5.7). To uncover the behaviour of the system, we simplify the superfluous parameters as follows: $D_A = D_B = 1$. Then, for small wave numbers ($k \to 0$, long-wavelength perturbation) the determinant takes the form

$$\det(J) = \frac{2(1-2\chi c_A^0)(k_1 + k_2 e^{\beta\Delta\mu})}{c_A^0} k^2 + \mathcal{O}(k^4), \tag{5.8}$$

where the superscript '0' indicates the value of the homogeneous concentrations. By inspection, one can see that the change of sign in the determinant corresponds to the spinodal condition for equilibrium systems (5.2), because k_1 and k_2 are strictly non-negative. When the determinant changes sign, one of the eigenvalues of J also changes sign, marking the threshold for stability, which matches the stability condition for the mixture at equilibrium. This is in line with the findings of Chapter 4 where we saw that the Lyapunov functional of complex-balanced solutions is close to that of a system at thermodynamic equilibrium and cannot sustain diffusion currents at steady state.

Non-complex-balanced behaviour

Finally, if we assume one of the reaction pathways considered above is sensitive to changes in the environment (as discussed in Section 3.2.3) complex balance is broken. We can describe this situation by the following version of the previous reaction-diffusion equations

$$\frac{\partial c_A}{\partial t} = D_A \nabla^2 \mu_A + k_1 g(c_A)[e^{\beta(\mu_B + \Delta\mu)} - e^{\beta\mu_A}] + k_2(e^{\beta\mu_B} - e^{\beta\mu_A}) \tag{5.9}$$

$$\frac{\partial c_B}{\partial t} = D_B \nabla^2 \mu_B - k_1 g(c_A)[e^{\beta(\mu_B + \Delta\mu)} - e^{\beta\mu_A}] - k_2(e^{\beta\mu_B} - e^{\beta\mu_A}), \tag{5.10}$$

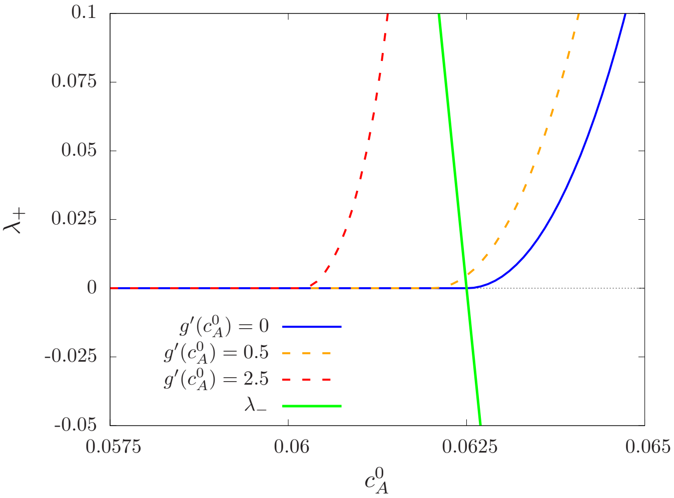
where we have included the function $g(c_A)$ to represent the variation of the rate constant of the first reaction with the environment of the reaction and, in particular, with the concentration of species A.

Carrying out a similar LSA with further simplifications to neatly observe the effect of breaking complex balance ($k_1 = k_2 = 1$), one finds, for the value of the determinant of the Jacobian matrix for small k

$$\det(J) = \frac{(1 + e^{\beta\Delta\mu})(4 - 8\chi c_A^0) + c_A^0 g'(c_A^0)(1 - e^{\beta\Delta\mu})}{2c_A^0} k^2 + \mathcal{O}(k^4), \tag{5.11}$$

where $g'(c_A^0)$ is the derivative of g with respect to c_A evaluated at $c_A = c_A^0$. Thus, the sign of the determinant for small k is no longer set by the equilibrium spinodal decomposition as there is a contribution from the derivative of g, the function that describes how the rate of a chemical

Figure 5.1: LSA for complex-balanced and non-complex-balanced networks. Largest eigenvalue (evaluated at the wave number k that maximises it) for each of the values of a homogeneous c_A^0. The eigenvalues λ_+ are compared with the equilibrium spinodal condition λ_- (solid green line). λ_- crosses 0 at the same point where the $g'(c_a) = 0$ line becomes positive (solid blue line), proving that for complex-balanced systems, despite being out of equilibrium, their stability is still given by the equilibrium spinodal condition. The parameters are set as in Section 5.1, with $\Delta\mu = 5k_\mathrm{B}T$, $c_B^0 = 1$ and $g(c_A^0) = 1$.



reaction changes with the environment. It should be noted that, in order to have the stability modified by non-equilibrium effects as found here, two conditions are necessary: $\Delta\mu \neq 0$ and $g'(c_A^0) \neq 0$ (and not equal for every reaction in the network). The first one implies that the system is not at thermodynamic equilibrium and the second one gives a local measure of how the rate of one of the chemical reactions changes due to the environment it take place in. If any of these conditions is not respected, then the stability of the system would be determined by the equilibrium spinodal condition, as for the complex-balanced or equilibrium case.

Note that here we have only analysed the stability of the system for small k or, in other words, long wavelength perturbations. In order to examine it with more generality, numerically, the value of the largest eigenvalue (with any wavelength) for each parameter set was obtained, see Fig. 5.1. As it can be seen from the figure, the complex-balanced ($g'(c_A^0) = 0$, solid blue line) eigenvalues yield the same condition for the instability as for spinodal decomposition at equilibrium λ_-, Eq. (5.2). This suggest that more exotic behaviour could appear for non-complex-balanced solutions, such as Ostwald ripening arrest for non-complex-balanced systems, as found in Ref. [Kirschbaum and Zwicker, 2021].

Therefore, we have given minimum conditions for non-equilibrium patterning to appear – complex balance must be broken – but not a sufficient condition. Further efforts will be required in order to give general and sufficient conditions for non-equilibrium patterning to appear and characterise the phenomenology that may arise.

5.2 Perspectives

It is therefore necessary to go beyond complex-balancing in order to be able to explain non-equilibrium phenomena such as patterning, which are ubiquitous in living beings.

A small step in this direction could be the analysis of networks that we will call weakly non-complex-balanced. This is a network whose topology is not limited to a complex-balanced network (that is, its deficiency δ is not 0). However, if the reactions that break the complex-balance constraint are slow when compared to the rest, then one would expect that the complex-balanced behaviour still dominates the dynamics but, arguably, with a small deviation due to these slow reactions. However, how large could this deviation be and what would its functional form be remains unclear.

Hence, we suggest a series expansion around the complex-balanced steady state to address this problem. Plenty of series expansions have been devised to analyse the steady-state behaviour of systems out of equilibrium (recently, e.g. [Freitas et al., 2021]). Nevertheless, most of this expansions take as a starting point the state of the system at thermodynamic equilibrium which means that, in principle, they cannot account for phenomena happening far away from equilibrium. In contrast, in this case, the expansion would not be based in the equilibrium steady state but in the complex-balanced one, which can be arbitrarily far away from equilibrium. Thus, we would be limited by the topology of the network and the timescales of the reactions rather than by the *distance* to thermodynamic equilibrium.

In practice, this would mean to obtain the steady-state distribution for the closest complex-balanced network and then add perturbatively the effects of the slow reactions that break complex balance, where the expansion parameter would be the ratio between the reaction constant of the fast reactions and the slow ones. This could be done in an analogous way to other expansions (e.g. [Proesmans et al., 2016, Freitas et al., 2021]) but taking as an ansatz the complex-balanced steady state plus a correction, that is, an exponential with argument $\beta[F(\mathbf{n}) + \sum_i \tilde{\mu}_i n_i + g(\mathbf{n})]$, where $g(\mathbf{n})$ quantifies the deviation from complex balance and could be obtained to different orders in the ratio of reaction constants.

Apart from increasing our understanding of general reaction networks, this expansion would also have consequences from the practical point of view. In real biological systems, it is seldom the case that the full reaction networks are known with great precision. Therefore, it might occur that a CRN which, to the best of our knowledge, is complex balanced, in reality it is not so, because there could be slow or rare reactions taking place that are hard to measure or identify. Then, this expansion would provide a means by which the results presented here could be robustly applied to complex biological settings, where deviations from complex-balancing, even if unidentified, may occur.

Finally, it would still be a challenge to find Lyapunov functionals even for the weakly non-complex-balanced case, if they exist [Ge and Qian, 2016]. This is because arguments like the ones given in Section 4.5 are unlikely to hold for more complex steady states. In fact, any deviation from a Boltzmann distribution other than a linear one at the level of the exponential of particle numbers (which is the case of complex-balanced networks) would invalidate the proof of the Lyapunov functional for the diffusion process, at least as performed here. Thus, in the future, new arguments to find Lyapunov functionals will need to be devised, in order to obtain a more general understanding of non-equilibrium patterning events. But, for the moment, to the best of my knowledge, these more general Lyapunov functionals remain unknown.

CONCLUSION

For a theory that has been well corroborated can only be superseeded by one of a higher level of universality; that is, by a theory that is better testable and which, in addition, contains the old, well corroborated theory [...].

K. R. Popper
The Logic of Scientific Discovery, 1959

The present work has been devoted to the study of theoretical biophysics and, in particular, patterning and self-organisation processes in non-equilibrium solutions. This topic has been analysed from two different perspectives, embodied by each of the two major parts of this manuscript.

In Part I, we theoretically addressed two examples of processes that organise the intracellular space in different ways: one that sets the size and number of large protein aggregates and another one which explains the appearance of a nucleoid in bacteria and its localisation within the cell.

To a great extent, the scientific knowledge is built upon confrontation between hypotheses and facts, or, in this case, between theoretical predictions and quantitative measurements. Proceeding in accordance with these ideas, analysis of microscopy images were performed to confront the aggregation model. We found good agreement between the model and the experimental data and hypothesised that such a timescale separation could influence many other intracellular phenomena that involve large agglomerates or structures, such as phase-separated droplets. Regarding the modelling of the bacterial nucleoid, our results were compared with the existing empirical evidence and additional measurements were suggested for several perturbations of the bacterial nucleoid, which should further validate the theory or force us to modify it, or even fully discard it.

In contrast, the objective of Part II was markedly different. In a purely theoretical endeavour, its aim was to understand the behaviour of reaction-diffusion models in the most general way possible, both from the dynamical-system perspective (for example, what kind of behaviour can we expect from the reaction-diffusion equations at steady state) and from the thermodynamical point of view (e.g. how does thermodynamics constrain the form of the reaction-diffusion equations). We achieved a detailed characterisation of a large class of CRNs known as complex-balanced, for which we ruled out the presence of diffusion currents at steady state (and, hence, non-equilibrium patterning cannot exist for these networks) and obtained a Lyapunov functional for these systems with a clear thermodynamic interpretation. In doing so, results that had been obtained for specific networks (e.g. [Bazant, 2013, Kirschbaum and Zwicker, 2021]) were generalised, but the resulting theory is not general enough to explain how non-equilibrium patterning arises in non-ideal solutions.

However, it remains unclear what are the repercussions of the results obtained in Part II for complex biological systems. Clearly, this is the price one has to pay for constructing such an abstract and general theory: it is still far away from making concrete falsifiable hypothesis, especially in relation to biological systems. The applicability of these results is also limited by their heavy dependency on the topology of the CRN, which is well defined from a mathematical point of view but in real living beings it might be hard to obtain in full detail. In particular, let us assume that, for a certain process in a given organism, molecular biologists have managed to discover the main players and their interactions. However, one cannot expect that this network is completely isolated from the rest of the organism nor that all chemical reactions have been discovered, no matter how weak or slow they are. Then, if the known network were complex-balanced but there were still some hidden nodes that break this complex-balancing, in view of Part II, what can we say about the system? Chapter 5 hints at how progress in this direction may be achieved, which may enable the robust application of these results in biological systems.

Nevertheless, the utility of this theory does not lie as much in its predictive power as in the *order* that it introduces in the plethora of results that have been obtained for particular models: this abstract approach enables the identification of a large class of networks that, even if driven out of equilibrium, still have many of the features associated to systems relaxing to thermodynamic equilibrium. From the theoretical point of view, it aids in relating network topologies to the phenomenology of non-ideal solutions out of thermodynamic equilibrium and has the potential to help in the construction of future models because it has increased, in an intuitive way, our understanding of the behaviour that these models may exhibit.

The original motivation for the work undertaken in Part II is the existence of *effective* reaction-diffusion models for non-ideal solutions that, while they are not thermodynamically consistent in the sense of Part II (chemical reactions and diffusion do not stem from a single free energy), have achieved a remarkable accuracy in describing experimental observations (e.g. Ref. [Zwicker et al., 2014]). Even the model for the nucleoid presented in Chapter 2 lacks thermodynamic consistency according to Part II, yet it is a simple model that can recapitulate plenty of the experimentally observed phenomena and produce new hypotheses, which may or may not be confirmed. Both are simple but effective models that could be thought of as coarse-grained models, where the microscopic details have been integrated out or absorbed into certain parameters, and that display this paradoxical behaviour under certain approximations (in the sense of an apparent lack of thermodynamic consistency in the dynamics). To the best of my knowledge, there is no systematic way to obtain these effective models from a coarse-graining procedure that starts from microscopic and thermodynamically consistent models. This class of *active matter* models are not limited to reaction-diffusion systems [Li and Cates, 2020] but also encompasses systems with non-reciprocal interactions [Saha et al., 2020] or self-propelled particles [Tailleur and Cates, 2008]. In some cases, attempting a coarse-graining procedure may not even make sense (e.g. for self-propelled particles it seems an overwhelming task) but, in others, I expect it to greatly increase our understanding of systems out of thermodynamic equilibrium, if such a mechanism could be devised. From our results, it seems clear that inherently active patterning phenomena as the one described in Ref. [Li and Cates, 2020], Ref. [Zwicker et al., 2014], or Ref. [Saha et al., 2020] cannot arise from the coarse graining of a complex-balanced CRN[1], but, by no means this implies that these effective models cannot be obtained starting from more general networks.

Each of the parts of this thesis represents two orthogonal approaches to theoretical biophysics. A modelling approach to theoretical biophysics utilises *effective* representations of processes while retaining their physical interpretation and eases the obtention of clear falsifiable predictions. This enables rapid rejection of the model if the hypotheses are proven wrong and advancement in the theoretical and empirical understanding of the system under consideration. Conversely, we have also undertaken a more abstract approach, aiming at deciphering the generalities of a certain class of reaction-diffusion systems, which enables the extraction of general features of a wide variety of models in an intuitive way. However, as we have seen above, the drawback of this type of general intuitive understanding is the difficulty in testing such a theory in real systems and the limitations when trying to explain complex emergent phenomena like non-equilibrium patterning.

---

[1] Note that, since patterning can also emerge at thermodynamic equilibrium, we are only referring to patterning that arises as a consequence free-energy dissipation and that it would not exist in the absence of this dissipation.

Therefore, to make further progress in the understanding of the physics underpinning biological systems, both approaches will be needed: general results that set the main lines of the behaviour of non-equilibrium systems and particular models that yield concrete refutable hypotheses regarding the functioning of biological cells or organisms. Ideally, eventually these two approaches would be linked by a rigorous coarse-graining procedure that, in an understandable way, explains the functioning and self-organisation of the living from physical principles. Tackling the complexity of the living and comprehending the emergent phenomena in biological systems will require theoretical efforts at every scale, and a successful integration of theoretical predictions and experimental results that enables the empirical testing of general and abstract theories.

Bibliography

[Al Jord et al., 2021] Al Jord, A., Letort, G., Eichmuller, A., Chanet, S., Huynh, J.-R., Gov, N. S., Voituriez, R., Terret, M.-É., and Verlhac, M.-H. (2021). Cytoplasmic forces functionally reorganize nuclear condensates in oocytes. *bioRxiv*.

[Almiron et al., 1992] Almiron, M., Link, A. J., Furlong, D., and Kolter, R. (1992). A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev.*, 6(12b):2646.

[An et al., 2008] An, S., Kumar, R., Sheets, E. D., and Benkovic, S. J. (2008). Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. *Science*, 320(5872):103.

[Andersen et al., 1971] Andersen, H. C., Weeks, J. D., and Chandler, D. (1971). Relationship between the hard-sphere fluid and fluids with realistic repulsive forces. *Phys. Rev. A*, 4:1597.

[Anderson et al., 2015] Anderson, D. F., Craciun, G., Gopalkrishnan, M., and Wiuf, C. (2015). Lyapunov functions, stationary distributions, and non-equilibrium potential for reaction networks. *Bull. Math. Biol.*, 77(9):1744.

[Anderson et al., 2010] Anderson, D. F., Craciun, G., and Kurtz, T. G. (2010). Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Biol.*, 72(8):1947.

[Anderson and Nguyen, 2019] Anderson, D. F. and Nguyen, T. D. (2019). Results on stochastic reaction networks with non-mass action kinetics. *Math. Biosc. Eng.*, 16(4):2118.

[Anderson, 1972] Anderson, P. W. (1972). More is different. *Science*, 177(4047):393.

[Asakura and Oosawa, 1954] Asakura, S. and Oosawa, F. (1954). On interaction between two bodies immersed in a solution of macromolecules. *J. Chem. Phys.*, 22(7):1255.

[Ashcroft and Mermin, 1976] Ashcroft, N. W. and Mermin, N. D. (1976). *Solid state physics*. Holt, Rinehart and Winston, New York, London.

[Avanzini et al., 2021] Avanzini, F., Penocchio, E., Falasco, G., and Esposito, M. (2021). Nonequilibrium thermodynamics of non-ideal chemical reaction networks. *J. Chem. Phys.*, 154(9):094114.

[Bakshi et al., 2014] Bakshi, S., Choi, H., Mondal, J., and Weisshaar, J. C. (2014). Time-dependent effects of transcription-and translation-halting drugs on the spatial distributions of the *Escherichia coli* chromosome and ribosomes. *Mol. Microbiol.*, 94(4):871.

[Bakshi et al., 2012] Bakshi, S., Siryaporn, A., Goulian, M., and Weisshaar, J. C. (2012). Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol. Microbiol.*, 85(1):21.

[Banani et al., 2017] Banani, S. F., Lee, H. O., Hyman, A. A., and Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.*, 18(5):285.

[Bartholomäus et al., 2016] Bartholomäus, A., Fedyunin, I., Feist, P., Sin, C., Zhang, G., Valleriani, A., and Ignatova, Z. (2016). Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philos. Trans. R. Soc. A*, 374(2063):20150069.

[Bazant, 2013] Bazant, M. Z. (2013). Theory of chemical kinetics and charge transfer based on nonequilibrium thermodynamics. *Acc. Chem. Res.*, 46(5):1144.

[Bernstein et al., 2004] Bernstein, J. A., Lin, P.-H., Cohen, S. N., and Lin-Chao, S. (2004). Global analysis of *Escherichia coli* RNA degradosome function using DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.*, 101(9):2758.

[Boocock et al., 2021] Boocock, D., Hino, N., Ruzickova, N., Hirashima, T., and Hannezo, E. (2021). Theory of mechanochemical patterning and optimal migration in cell monolayers. *Nat. Phys.*, 17(2):267.

[Brangwynne et al., 2009] Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A. A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, 324(5935):1729.

[Buchner et al., 2013] Buchner, A., Tostevin, F., and Gerland, U. (2013). Clustering and optimal arrangement of enzymes in reaction-diffusion systems. *Phys. Rev. Lett.*, 110(20):208104.

[Cabrera et al., 2009] Cabrera, J. E., Cagliero, C., Quan, S., Squires, C. L., and Jin, D. J. (2009). Active transcription of rRNA operons condenses the nucleoid in *Escherichia coli*: Examining the effect of transcription on nucleoid structure in the absence of transertion. *J. Bacteriol.*, 191(13):4180.

[Cahn and Hilliard, 1958] Cahn, J. W. and Hilliard, J. E. (1958). Free energy of a nonuniform system. I. Interfacial free energy. *J. Chem. Phys.*, 28(2):258.

[Campbell et al., 2005] Campbell, A. I., Anderson, V. J., van Duijneveldt, J. S., and Bartlett, P. (2005). Dynamical arrest in attractive colloids: The effect of long-range repulsion. *Phys. Rev. Lett.*, 94(20):208301.

[Carati and Lefever, 1997] Carati, D. and Lefever, R. (1997). Chemical freezing of phase separation in immiscible binary mixtures. *Phys. Rev. E*, 56:3127.

[Castellana et al., 2016] Castellana, M., Hsin-Jung Li, S., and Wingreen, N. S. (2016). Spatial organization of bacterial transcription and translation. *Proc. Natl. Acad. Sci. U.S.A.*, 113(33):9286.

[Castellana et al., 2014] Castellana, M., Wilson, M. Z., Xu, Y., Joshi, P., Cristea, I. M., Rabinowitz, J. D., Gitai, Z., and Wingreen, N. S. (2014). Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.*, 32(10):1011.

[Cohan and Pappu, 2020] Cohan, M. C. and Pappu, R. V. (2020). Making the case for disordered proteins and biomolecular condensates in bacteria. *Trends Biochem. Sci.*, 45(8):668.

[Coquel et al., 2013] Coquel, A.-S., Jacob, J.-P., Primet, M., Demarez, A., Dimiccoli, M., Julou, T., Moisan, L., Lindner, A. B., and Berry, H. (2013). Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect. *PLOS Comput. Biol.*, 9(4):1.

[Crank, 1979] Crank, J. (1979). *The mathematics of diffusion*. Oxford university press.

[Crooks, 1999] Crooks, G. E. (1999). Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60(3):2721.

[Cunha et al., 2001] Cunha, S., Woldringh, C. L., and Odijk, T. (2001). Polymer-mediated compaction and internal dynamics of isolated *Escherichia coli* nucleoids. *J. Struct. Biol.*, 136(1):53.

[Dame et al., 2020] Dame, R. T., Rashid, F.-Z. M., and Grainger, D. C. (2020). Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.*, 21(4):227.

[De Gennes, 1979] De Gennes, P.-G. (1979). *Scaling concepts in polymer physics*. Cornell University press.

[Dill and Bromberg, 2012] Dill, K. and Bromberg, S. (2012). *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience*. Garland Science.

[Ditlev et al., 2018] Ditlev, J. A., Case, L. B., and Rosen, M. K. (2018). Who's in and who's out—compositional control of biomolecular condensates. *J. Mol. Biol.*, 430(23):4666. Phase Separation in Biology and Disease.

[Dorman, 2013] Dorman, C. J. (2013). Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat. Rev. Microbiol.*, 11(5):349.

[Edwards and Freed, 1969] Edwards, S. F. and Freed, K. F. (1969). The entropy of a confined polymer. I. *J. Phys. A*, 2(2):145.

[England, 2013] England, J. L. (2013). Statistical physics of self-replication. *J. Chem. Phys.*, 139(12):09B623_1.

[Etoc et al., 2018] Etoc, F., Balloul, E., Vicario, C., Normanno, D., Liße, D., Sittner, A., Piehler, J., Dahan, M., and Coppey, M. (2018). Non-specific interactions govern cytosolic diffusion of nanosized objects in mammalian cells. *Nat. Mater.*, 17(8):740.

[Fall and Keizer, 2001] Fall, C. P. and Keizer, J. E. (2001). Mitochondrial modulation of intracellular Ca2+ signaling. *J. Theor. Biol.*, 210(2):151.

[Feinberg, 1972] Feinberg, M. (1972). Complex balancing in general kinetic systems. *Arch. Ration. Mech. Anal.*, 49(3):187.

[Fisher et al., 2013] Fisher, J. K., Bourniquel, A., Witz, G., Weiner, B., Prentiss, M., and Kleckner, N. (2013). Four-dimensional imaging of *E. coli* nucleoid organization and dynamics in living cells. *Cell*, 153(4):882.

[Franzmann et al., 2018] Franzmann, T. M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A. S., Nüske, E., Richter, D., Baumeister, W., Grill, S. W., Pappu, R. V., Hyman, A. A., and Alberti, S. (2018). Phase separation of a yeast prion protein promotes cellular fitness. *Science*, 359(6371).

[Freitas et al., 2021] Freitas, N., Falasco, G., and Esposito, M. (2021). Linear response in large deviations theory: a method to compute non-equilibrium distributions. *New J. Phys.*, 23(9):093003.

[Garcia-Jove Navarro et al., 2019] Garcia-Jove Navarro, M., Kashida, S., Chouaib, R., Souquere, S., Pierron, G., Weil, D., and Gueroui, Z. (2019). RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates. *Nat. Commun.*, 10(1):3230.

[Ge and Qian, 2016] Ge, H. and Qian, H. (2016). Mesoscopic kinetic basis of macroscopic chemical thermodynamics: A mathematical theory. *Phys. Rev. E*, 94:052150.

[Gibbs, 1879] Gibbs, J. W. (1879). On the equilibrium of heterogeneous substances. *Trans. Conn. Acad. Arts Sci.*, 3:108.

[Gillespie, 1977] Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340.

[Gillespie, 1992] Gillespie, D. T. (1992). A rigorous derivation of the chemical master equation. *Phys. A: Stat. Mech. Appl.*, 188(1):404.

[Glansdorff and Prigogine, 1954] Glansdorff, P. and Prigogine, I. (1954). Sur les propriétés différentielles de la production d'entropie. *Physica*, 20(7):773.

[Glotzer et al., 1995] Glotzer, S. C., Di Marzio, E. A., and Muthukumar, M. (1995). Reaction-controlled morphology of phase-separating mixtures. *Phys. Rev. Lett.*, 74:2034.

[Gorle et al., 2017] Gorle, A. K., Bottomley, A. L., Harry, E. J., Collins, J. G., Keene, F. R., and Woodward, C. E. (2017). DNA condensation in live *E. coli* provides evidence for transertion. *Mol. BioSyst.*, 13:677.

[Gray et al., 2019] Gray, W. T., Govers, S. K., Xiang, Y., Parry, B. R., Campos, M., Kim, S., and Jacobs-Wagner, C. (2019). Nucleoid size scaling and intracellular organization of translation across bacteria. *Cell*, 177(6):1632.

[Groot and Mazur, 1983] Groot, S. R. and Mazur, P. (1983). *Non-equilibrium Thermodynamics*. Dover.

[Guilhas et al., 2020] Guilhas, B., Walter, J.-C., Rech, J., David, G., Walliser, N. O., Palmeri, J., Mathieu-Demaziere, C., Parmeggiani, A., Bouet, J.-Y., Le Gall, A., and Nollmann, M. (2020). ATP-driven separation of liquid phase condensates in bacteria. *Mol. Cell*, 79(2):293.

[Guo et al., 2014] Guo, M., Ehrlicher, A., Jensen, M., Renz, M., Moore, J., Goldman, R., Lippincott-Schwartz, J., Mackintosh, F., and Weitz, D. (2014). Probing the stochastic, motor-driven properties of the cytoplasm using force spectrum microscopy. *Cell*, 158(4):822.

[Halatek and Frey, 2018] Halatek, J. and Frey, E. (2018). Rethinking pattern formation in reaction–diffusion systems. *Nat. Phys.*, 14(5):507.

[Hänggi et al., 1990] Hänggi, P., Talkner, P., and Borkovec, M. (1990). Reaction-rate theory: Fifty years after Kramers. *Rev. Mod. Phys.*, 62:251.

[Herold et al., 2017] Herold, E., Hellmann, R., and Wagner, J. (2017). Virial coefficients of anisotropic hard solids of revolution: The detailed influence of the particle geometry. *J. Chem. Phys.*, 147(20):204102.

[Höfling and Franosch, 2013] Höfling, F. and Franosch, T. (2013). Anomalous transport in the crowded world of biological cells. *Rep. Prog. Phys.*, 76:046602.

[Horn and Jackson, 1972] Horn, F. and Jackson, R. (1972). General mass action kinetics. *Arch. Ration. Mech. Anal.*, 47(2):81.

[Huang, 1987] Huang, K. (1987). *Statistical Mechanics, 2nd Edition*. Wiley, 2 edition.

[Huberman, 1976] Huberman, B. A. (1976). Striations in chemical reactions. *J. Chem. Phys.*, 65(5):2013.

[Ilker and Joanny, 2020] Ilker, E. and Joanny, J.-F. (2020). Phase separation and nucleation in mixtures of particles with different temperatures. *Phys. Rev. Research*, 2:023200.

[Janissen et al., 2018] Janissen, R., Arens, M. M., Vtyurina, N. N., Rivai, Z., Sunday, N. D., Eslami-Mossallam, B., Gritsenko, A. A., Laan, L., de Ridder, D., Artsimovitch, I., Dekker, N. H., Abbondanzieri, E. A., and Meyer, A. S. (2018). Global DNA compaction in stationary-phase bacteria does not affect transcription. *Cell*, 174(5):1188.

[Jarzynski, 1997] Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78(14):2690.

[Joshi et al., 2011] Joshi, M. C., Bourniquel, A., Fisher, J., Ho, B. T., Magnan, D., Kleckner, N., and Bates, D. (2011). *Escherichia coli* sister chromosome separation includes an abrupt global transition with concomitant release of late-splitting intersister snaps. *Proc. Natl. Acad. Sci. U.S.A.*, 108(7):2765.

[Joyeux, 2019] Joyeux, M. (2019). Preferential localization of the bacterial nucleoid. *Microorganisms*, 7(7):204.

[Jülicher et al., 1997] Jülicher, F., Ajdari, A., and Prost, J. (1997). Modeling molecular motors. *Rev. Mod. Phys.*, 69(4):1269.

[Kaczanowska and Rydén-Aulin, 2007] Kaczanowska, M. and Rydén-Aulin, M. (2007). Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiol. Mol. Biol. R.*, 71(3):477.

[Kim et al., 2019] Kim, T. H., Tsang, B., Vernon, R. M., Sonenberg, N., Kay, L. E., and Forman-Kay, J. D. (2019). Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. *Science*, 365(6455):825.

[Kirschbaum and Zwicker, 2021] Kirschbaum, J. and Zwicker, D. (2021). Controlling biomolecular condensates via chemical reactions. *J. R. Soc. Interface*, 18(179):20210255.

[Klosin et al., 2020] Klosin, A., Oltsch, F., Harmon, T., Honigmann, A., Jülicher, F., Hyman, A. A., and Zechner, C. (2020). Phase separation provides a mechanism to reduce noise in cells. *Science*, 367(6476):464.

[Knowles et al., 2014] Knowles, T. P. J., Vendruscolo, M., and Dobson, C. M. (2014). The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.*, 15(6):384.

[Knowles et al., 2009] Knowles, T. P. J., Waudby, C. A., Devlin, G. L., Cohen, S. I. A., Aguzzi, A., Vendruscolo, M., Terentjev, E. M., Welland, M. E., and Dobson, C. M. (2009). An analytical solution to the kinetics of breakable filament assembly. *Science*, 326(5959):1533.

[Kohram, 2021] Kohram, M. (2021). *Bacterial growth mechanisms and their role in cell size homeostasis and senescence*. PhD thesis, University of Pittsburgh.

[Kolb et al., 1983] Kolb, M., Botet, R., and Jullien, R. (1983). Scaling of kinetically growing clusters. *Phys. Rev. Lett.*, 51:1123.

[Kondepundi and Prigogine, 2015] Kondepundi, D. and Prigogine, I. (2015). *Modern Thermodynamics*. John Wiley & Sons, Ltd.

[Krapivsky et al., 2010] Krapivsky, P. L., Redner, S., and Ben-Naim, E. (2010). *A kinetic view of statistical physics*. Cambridge University Press.

[Kurtz, 1972] Kurtz, T. G. (1972). The relationship between stochastic and deterministic models for chemical reactions. *J. Chem. Phys.*, 57(7):2976.

[Lewis, 1925] Lewis, G. N. (1925). A new principle of equilibrium. *Proc. Natl. Acad. Sci. U.S.A.*, 11(3):179.

[Lewis, 2004] Lewis, P. J. (2004). Bacterial subcellular architecture: recent advances and future prospects. *Mol. Microbiol.*, 54(5):1135.

[Leyvraz, 2003] Leyvraz, F. (2003). Scaling theory and exactly solved models in the kinetics of irreversible aggregation. *Phys. Rep.*, 383(2):95.

[Li et al., 2016] Li, L., Liu, H., Dong, P., Li, D., Legant, W. R., Grimm, J. B., Lavis, L. D., Betzig, E., Tjian, R., and Liu, Z. (2016). Real-time imaging of Huntingtin aggregates diverting target search and gene transcription. *eLife*, 5:e17056.

[Li et al., 2012] Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D. S., Banani, S. F., Russo, P. S., Jiang, Q.-X., Nixon, B. T., and Rosen, M. K. (2012). Phase transitions in the assembly of multivalent signalling proteins. *Nature*, 483(7389):336.

[Li and Cates, 2020] Li, Y. I. and Cates, M. E. (2020). Non-equilibrium phase separation with reactions: a canonical model and its behaviour. *J. Stat. Mech.: Theory Exp.*, 2020(5):053206.

[Luby-Phelps et al., 1987] Luby-Phelps, K., Castle, P. E., Taylor, D. L., and Lanni, F. (1987). Hindered diffusion of inert tracer particles in the cytoplasm of mouse 3T3 cells. *Proc. Natl. Acad. Sci. U.S.A.*, 84(14):4910.

[Lyon et al., 2021] Lyon, A. S., Peeples, W. B., and Rosen, M. K. (2021). A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.*, 22(3):215.

[MATLAB, 2018] MATLAB (2018). The MathWorks Inc., Natick, Massachusetts, United States.

[Meakin and Jullien, 1988] Meakin, P. and Jullien, R. (1988). The effects of restructuring on the geometry of clusters formed by diffusion-limited, ballistic, and reaction-limited cluster-cluster aggregation. *J. Chem. Phys.*, 89(1):246.

[Mercier et al., 2008] Mercier, R., Petit, M.-A., Schbath, S., Robin, S., El Karoui, M., Boccard, F., and Espéli, O. (2008). The matp/*matS* site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. *Cell*, 135(3):475.

[Miangolarra et al., 2021a] Miangolarra, A. M., Duperray-Susini, A., Coppey, M., and Castellana, M. (2021a). Two timescales control the creation of large protein aggregates in cells. *Biophys. J.*, 120(12):2394.

[Miangolarra et al., 2021b] Miangolarra, A. M., Li, S. H.-J., Joanny, J.-F., Wingreen, N. S., and Castellana, M. (2021b). Steric interactions and out-of-equilibrium processes control the internal organization of bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, 118(43).

[Mitchison, 2019] Mitchison, T. J. (2019). Colloid osmotic parameterization and measurement of subcellular crowding. *Molecular Biology of the Cell*, 30(2):173. PMID: 30640588.

[Mondal et al., 2011] Mondal, J., Bratton, B. P., Li, Y., Yethiraj, A., and Weisshaar, J. C. (2011). Entropy-based mechanism of ribosome-nucleoid segregation in *E. coli* cells. *Biophys. J.*, 100(11):2605.

[Murray and Sourjik, 2017] Murray, S. M. and Sourjik, V. (2017). Self-organization and positioning of bacterial protein clusters. *Nat. Phys.*, 13(10):1006.

[Narayanan et al., 2019] Narayanan, A., Meriin, A., Andrews, J. O., Spille, J.-H., Sherman, M. Y., and Cisse, I. I. (2019). A first order phase transition mechanism underlies protein aggregation in mammalian cells. *eLife*, 8:e39695.

[Nelson, 2004] Nelson, P. (2004). *Biological physics*. WH Freeman New York.

[Nott et al., 2015] Nott, T., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T. D., Bazett-Jones, D., Pawson, T., Forman-Kay, J., and Baldwin, A. (2015). Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell*, 57(5):936.

[Nyström, 2004] Nyström, T. (2004). Stationary-phase physiology. *Annu. Rev. Microbiol.*, 58:161.

[Odijk, 1998] Odijk, T. (1998). Osmotic compaction of supercoiled DNA into a bacterial nucleoid. *Biophys. Chem.*, 73(1):23.

[Odijk, 2000] Odijk, T. (2000). Dynamics of the expanding DNA nucleoid released from a bacterial cell. *Phys. A: Stat. Mech. Appl.*, 277(1):62.

[Onnes, 1902] Onnes, H. K. (1902). Expression of the equation of state of gases and liquids by means of series. *KNAW, Proceedings*, 4:125.

[Onsager, 1931] Onsager, L. (1931). Reciprocal relations in irreversible processes. I. *Phys. Rev.*, 37(4):405.

[Park et al., 2017] Park, H., Kim, N. Y., Lee, S., Kim, N., Kim, J., and Heo, W. D. (2017). Optogenetic protein clustering through fluorescent protein tagging and extension of CRY2. *Nat. Commun.*, 8(1):30.

[Pathria, 1996] Pathria, R. K. (1996). *Statistical Mechanics*. Butterworth-Heinemann, Oxford, 2 edition.

[Pelletier et al., 2012] Pelletier, J., Halvorsen, K., Ha, B.-Y., Paparcone, R., Sandler, S. J., Woldringh, C. L., Wong, W. P., and Jun, S. (2012). Physical manipulation of the *Escherichia coli* chromosome reveals its soft nature. *Proc. Natl. Acad. Sci. U.S.A.*, 109(40):E2649.

[Pham et al., 2010] Pham, B. T., Zondanos, H., Such, C. H., Warr, G. G., and Hawkett, B. S. (2010). Miniemulsion polymerization with arrested ostwald ripening stabilized by amphiphilic raft copolymers. *Macromolecules*, 43(19):7950.

[Phillips et al., 2012] Phillips, R., Kondev, J., Theriot, J., Garcia, H. G., and Orme, N. (2012). *Physical biology of the cell*. Garland Science.

[Pippard, 1964] Pippard, A. B. (1964). *Elements of Classical Thermodynamics*. Cambridge University Press.

[Polettini et al., 2015] Polettini, M., Wachtel, A., and Esposito, M. (2015). Dissipation in noisy chemical networks: The role of deficiency. *J. Chem. Phys.*, 143(18):184103.

[Prigogine, 1977] Prigogine, I. (1977). Nobel lecture: Time, structure and fluctuations.

[Proesmans et al., 2016] Proesmans, K., Cleuren, B., and den Broeck, C. V. (2016). Linear stochastic thermodynamics for periodically driven systems. *J. Stat. Mech.: Theory Exp.*, 2016(2):023202.

[Qiang et al., 2017] Qiang, W., Yau, W.-M., Lu, J.-X., Collinge, J., and Tycko, R. (2017). Structural variation in amyloid-β fibrils from Alzheimer's disease clinical subtypes. *Nature*, 541(7636):217.

[Quail et al., 2021] Quail, T., Golfier, S., Elsner, M., Ishihara, K., Murugesan, V., Renger, R., Jülicher, F., and Brugués, J. (2021). Force generation by protein–DNA co-condensation. *Nat. Phys.*, page 1.

[Rao and Esposito, 2016] Rao, R. and Esposito, M. (2016). Nonequilibrium thermodynamics of chemical reaction networks: Wisdom from stochastic thermodynamics. *Phys. Rev. X*, 6:041064.

[Rao and Esposito, 2018] Rao, R. and Esposito, M. (2018). Conservation laws and work fluctuation relations in chemical reaction networks. *J. Chem. Phys.*, 149(24):245101.

[Remesh et al., 2020] Remesh, S. G., Verma, S. C., Chen, J.-H., Ekman, A. A., Larabell, C. A., Adhya, S., and Hammel, M. (2020). Nucleoid remodeling during environmental adaptation is regulated by HU-dependent DNA bundling. *Nat. Commun.*, 11(1):2905.

[Roberts et al., 2002] Roberts, K., Alberts, B., Johnson, A., Walter, P., and Hunt, T. (2002). Molecular biology of the cell. *New York: Garland Science*, 32(2).

[Ross and Poirier, 2004] Ross, C. A. and Poirier, M. A. (2004). Protein aggregation and neurodegenerative disease. *Nat. Med.*, 10(7):S10.

[Rubinstein and Colby, 2003] Rubinstein, M. and Colby, R. H. (2003). *Polymer physics*, volume 23. Oxford university press New York.

[Saha et al., 2020] Saha, S., Agudo-Canalejo, J., and Golestanian, R. (2020). Scalar active mixtures: The nonreciprocal Cahn-Hilliard model. *Physical Review X*, 10(4):041009.

[Sanamrad et al., 2014] Sanamrad, A., Persson, F., Lundius, E. G., Fange, D., Gynnå, A. H., and Elf, J. (2014). Single-particle tracking reveals that free ribosomal subunits are not excluded from the *Escherichia coli* nucleoid. *Proc. Natl. Acad. Sci. U.S.A.*, 111(31):11413.

[Santos et al., 2005] Santos, J. R., Ferrat, G. C., and Eichelmann, M. G. (2005). La fase estacionaria en la bacteria *Escherichia coli*. *Rev. Latinoam. Microbiol.*, 47(3-4):92–101.

[Schaechter and Laing, 1961] Schaechter, M. and Laing, V. O. (1961). Direct observation of fusion of bacterial nuclei. *J. Bacteriol.*, 81(4):667.

[Schmiedl and Seifert, 2007] Schmiedl, T. and Seifert, U. (2007). Stochastic thermodynamics of chemical reaction networks. *J. Chem. Phys.*, 126(4):044101.

[Schnoerr et al., 2017] Schnoerr, D., Sanguinetti, G., and Grima, R. (2017). Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J. Phys. A*, 50(9):093001.

[Seifert, 2005] Seifert, U. (2005). Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Phys. Rev. Lett.*, 95:040602.

[Sekimoto, 2010] Sekimoto, K. (2010). *Stochastic Energetics*. Springer-Verlag Berlin Heidelberg.

[Selkoe, 2004] Selkoe, D. J. (2004). Cell biology of protein misfolding: The examples of Alzheimer's and Parkinson's diseases. *Nature Cell Biology*, 6(11):1054.

[Shampine et al., 2005] Shampine, L. F., Thompson, S., Kierzenka, J. A., and Byrne, G. D. (2005). Non-negative solutions of ODEs. *Appl. Math. Comput.*, 170(1):556.

[Shin et al., 2018] Shin, Y., Chang, Y.-C., Lee, D. S., Berry, J., Sanders, D. W., Ronceray, P., Wingreen, N. S., Haataja, M., and Brangwynne, C. P. (2018). Liquid nuclear condensates mechanically sense and restructure the genome. *Cell*, 175(6):1481.

[Shlomovitz and Gov, 2009] Shlomovitz, R. and Gov, N. S. (2009). Membrane-mediated interactions drive the condensation and coalescence of FtsZ rings. *Phys. Biol.*, 6(4):046017.

[Straley, 1973] Straley, J. P. (1973). Third virial coefficient for the gas of long rods. *Mol. Cryst. Liq. Cryst.*, 24(1-2):7.

[Stuger et al., 2002] Stuger, R., Woldringh, C. L., van der Weijden, C. C., Vischer, N. O. E., Bakker, B. M., van Spanning, R. J., Snoep, J. L., and Weterhoff, H. V. (2002). DNA supercoiling by gyrase is linked to nucleoid compaction. *Mol. Biol. Rep.*, 29(1):79.

[Su et al., 2016] Su, X., Ditlev, J. A., Hui, E., Xing, W., Banjade, S., Okrut, J., King, D. S., Taunton, J., Rosen, M. K., and Vale, R. D. (2016). Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science*, 352(6285):595.

[Surovtsev and Jacobs-Wagner, 2018] Surovtsev, I. V. and Jacobs-Wagner, C. (2018). Subcellular organization: A critical feature of bacterial cell replication. *Cell*, 172(6):1271.

[Tailleur and Cates, 2008] Tailleur, J. and Cates, M. (2008). Statistical mechanics of interacting run-and-tumble bacteria. *Phys. Rev. Lett.*, 100(21):218103.

[Tarzia and Coniglio, 2006] Tarzia, M. and Coniglio, A. (2006). Pattern formation and glassy phase in the ϕ^4 theory with a screened electrostatic repulsion. *Phys. Rev. Lett.*, 96(7):075702.

[Taslimi et al., 2014] Taslimi, A., Vrana, J. D., Chen, D., Borinskaya, S., Mayer, B. J., Kennedy, M. J., and Tucker, C. L. (2014). An optimized optogenetic clustering tool for probing protein interaction and function. *Nat. Commun.*, 5(1):4925.

[Tsang et al., 2019] Tsang, B., Arsenault, J., Vernon, R. M., Lin, H., Sonenberg, N., Wang, L.-Y., Bah, A., and Forman-Kay, J. D. (2019). Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. *Proc. Natl. Acad. Sci. U.S.A.*, 116(10):4218.

[Turing, 1952] Turing, A. M. (1952). The chemical basis of morphogenesis. *Philos. Trans. R. Soc. B*, 237(641):37.

[Turner et al., 2005] Turner, M. S., Sens, P., and Socci, N. D. (2005). Nonequilibrium raftlike membrane domains under continuous recycling. *Phys. Rev. Lett.*, 95:168301.

[Verma et al., 2019] Verma, S. C., Qian, Z., and Adhya, S. L. (2019). Architecture of the *Escherichia coli* nucleoid. *PLOS Genet.*, 15(12):1.

[Wang et al., 2008] Wang, H., Wingreen, N. S., and Mukhopadhyay, R. (2008). Self-organized periodicity of protein clusters in growing bacteria. *Phys. Rev. Lett.*, 101:218101.

[Wang et al., 2018] Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., Poser, I., Pappu, R. V., Alberti, S., and Hyman, A. A. (2018). A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell*, 174(3):688.

[Weber et al., 2019] Weber, C. A., Zwicker, D., Jülicher, F., and Lee, C. F. (2019). Physics of active emulsions. *Rep. Prog. Phys.*, 82(6):064601.

[Wehrens et al., 2018] Wehrens, M., Ershov, D., Rozendaal, R., Walker, N., Schultz, D., Kishony, R., Levin, P. A., and Tans, S. J. (2018). Size laws and division ring dynamics in filamentous *Escherichia coli* cells. *Curr. Biol.*, 28(6):972.

[Weng et al., 2019] Weng, X., Bohrer, C. H., Bettridge, K., Lagda, A. C., Cagliero, C., Jin, D. J., and Xiao, J. (2019). Spatial organization of RNA polymerase and its relationship with transcription in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, 116(40):20115.

[Westheimer, 1987] Westheimer, F. H. (1987). Why nature chose phosphates. *Science*, 235(4793):1173.

[Wojcieszyn et al., 1981] Wojcieszyn, J. W., Schlegel, R. A., Wu, E. S., and Jacobson, K. A. (1981). Diffusion of injected macromolecules within the cytoplasm of living cells. *Proc. Natl. Acad. Sci. U.S.A.*, 78(7):4407.

[Wu et al., 2016] Wu, F., Halatek, J., Reiter, M., Kingma, E., Frey, E., and Dekker, C. (2016). Multistability and dynamic transitions of intracellular Min protein patterns. *Mol. Syst. Biol.*, 12(6):873.

[Wu et al., 2019] Wu, F., Swain, P., Kuijpers, L., Zheng, X., Felter, K., Guurink, M., Solari, J., Jun, S., Shimizu, T. S., Chaudhuri, D., Mulder, B., and Dekker, C. (2019). Cell boundary confinement sets the size and position of the *E. coli* chromosome. *Curr. Biol.*, 29(13):2131.

[Wurtz and Lee, 2018] Wurtz, J. D. and Lee, C. F. (2018). Chemical-reaction-controlled phase separated drops: Formation, size selection, and coarsening. *Phys. Rev. Lett.*, 120:078102.

[Xiang et al., 2021] Xiang, Y., Surovtsev, I. V., Chang, Y., Govers, S. K., Parry, B. R., Liu, J., and Jacobs-Wagner, C. (2021). Interconnecting solvent quality, transcription, and chromosome folding in *Escherichia coli*. *Cell*, 184(14):3626.

[Yang et al., 2020] Yang, D., Männik, J., Retterer, S. T., and Männik, J. (2020). The effects of polydisperse crowders on the compaction of the *Escherichia coli* nucleoid. *Mol. Microbiol.*, 113(5):1022.

[Zhang et al., 2009] Zhang, C., Shao, P. G., van Kan, J. A., and van der Maarel, J. R. C. (2009). Macromolecular crowding induced elongation and compaction of single DNA molecules confined in a nanochannel. *Proc. Natl. Acad. Sci. U.S.A.*, 106(39):16651.

[Zhou et al., 2008] Zhou, H.-X., Rivas, G., and Minton, A. P. (2008). Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.*, 37(1):375. PMID: 18573087.

[Zidar et al., 2018] Zidar, M., Kuzman, D., and Ravnik, M. (2018). Characterisation of protein aggregation with the Smoluchowski coagulation approach for use in biopharmaceuticals. *Soft Matter*, 14:6001.

[Zielinski et al., 2017] Zielinski, M. W., McGann, L. E., Nychka, J. A., and Elliott, J. A. W. (2017). Non-ideal solute chemical potential equation and the validity of the grouped solute approach for intracellular solution thermodynamics. *J. Phys. Chem. B*, 121(46):10443.

[Zwicker et al., 2014] Zwicker, D., Decker, M., Jaensch, S., Hyman, A. A., and Jülicher, F. (2014). Centrosomes are autocatalytic droplets of pericentriolar material organized by centrioles. *Proc. Natl. Acad. Sci. U.S.A.*, 111(26):E2636.

[Zwicker et al., 2017] Zwicker, D., Seyboldt, R., Weber, C. A., Hyman, A. A., and Jülicher, F. (2017). Growth and division of active droplets provides a model for protocells. *Nat. Phys.*, 13(4):408.

[Šarić et al., 2016] Šarić, A., Michaels, T. C. T., Zaccone, A., Knowles, T. P. J., and Frenkel, D. (2016). Kinetics of spontaneous filament nucleation via oligomers: Insights from theory and simulation. *J. Chem. Phys.*, 145(21):211926.

Appendix A

Supplementary Information for the aggregation model

A.1 Experimental Procedures

I include this section of the Appendix for completeness yet it is not my own work but that of my collaborators, as the experiments were carried out exclusively by Aléria Duperray-Susini and Mathieu Coppey.

A.1.1 Cell culture

The immortalized hTERT RPE1 cells (Human Retinal Pigmented Epithelium) were cultivated in DMEM F12 without Phenol Red (Gibco, Life Technologies) supplemented with 10% Fetal Bovine Serum (FBS) without antibiotic, hereafter called the growth medium. They were maintained at 37°C in humidified atmosphere with 5% CO_2, tested and certified as mycoplasma free.

A.1.2 Transitory cells transfection by Cry2Olig-mCherry

RPE1 cells were detached by trypsin and centrifuged for 3 min, 100 g at room temperature to eliminate it. The pellet was kept and resuspended on growth medium. They were transfected on suspension by jetPrime (Polyplus transfection), with 1 μg of DNA plasmid vector Cry2Olig-mCherry (purchased from Addgene, number 60032), and then platted on fluorodishes. According to the recommendation of manufactory, the medium was replaced after four hours by a fresh one. From there, the manipulation of cells was done in the complete dark.

A.1.3 Quantitative estimation of fluorescent protein concentration

To estimate the concentration of proteins in cells using the fluorescent signal, we calibrated the intensity on the camera using mCh-6His protein purified at 4.19 mg/ml (a gift from El Marjou. A, Platform of Curie Institute). We performed serial dilutions of the stock solution (1, 1:2, 1:4, 1:8, 1:10, 1:16, 1:32, 1:64, 1:100, 1:128, 1:1000) in the cell growth medium, and the medium alone was used for background estimation. For each dilution, we put a drop of 10 μl into a fluorodish and we imaged the drop using the exact same parameters as for the cell imaging experiments. Two images were acquired at a focus right above the coverslip, as for cell imaging. We then quantified the average fluorescent intensity using Fiji. The total intensity of the image was background subtracted and averaged over the size of the whole image. Data were plotted and gave rise to a linear relationship between raw intensities of the images and concentrations of recombinant fluorescent proteins. We fitted data with a line and used the value of the slope to convert intensities into concentrations.

A.1.4 Optogenetic experiments

All experiments were performed using 100x objectives (oil immersion, numerical aperture 1.4) by Inverted Spinning Disk Confocal Roper/Nikon, EMCCD 512x512 evolve (pixel size: 16 μm) photometrics come from to Imaging Nikon Center (PICT-LM) in Curie Institute. Live imaging was on normal growth condition and preserved by Life Imaging Service Yokogawa head: CSU-X1 integrated in Metamorph software by Gataca Systems. Twenty-four hours after transfection, cells were kept at 37°C and were imaged before any activation with blue light over 17 z-stack (0.5 μm) at 561 nm (0.134 mW). The same cells were imaged at the end of the activation routine using the same 17 z-stacks while keeping the same focus. Optogenetic activations were performed every two minutes for a total duration of one hour, using the laser blue light at 491 nm (0.506 mW). We selected cells for further image quantification based on their visible viability, on their presence in the field of view at the end of the experiment (some cells escaped the field of view after one hour), and on the absence of pixels saturation (very bright, saturated clusters could appear over the time course of the experiment). All laser settings and parameters of the camera (time of exposition, gain) were kept constant for all experiments and calibration of the concentration.

A.2 Image Analysis

The initial concentration of the protein is obtained from the cell image at the initial time, $t = 0$. The cell is separated from the background and the intensity is computed as the average of the intensity in the cell after subtracting the background intensity, using Matlab [MATLAB, 2018]. We estimated the volume of the cells by measuring the area of the cell just above the coverslide and assuming an effective height such that the total intensity of the 3D final image equals the total intensity of this 2D initial image times this effective height. This effective height parameter varies from cell to cell and has a mean value of 1.1μm and a standard deviation of 0.4μm.

In order to quantify the size and frequency of the cluster at $t = 1$ hr, we smoothed the image with a gaussian filter, substracted the mean background intensity, located the local maxima of intensity in the image, and performed a watershed transform to estimate the spatial extent of each cluster [Meyer, 1994]. The size of the clusters is then determined by considering that the cluster is composed of the pixels that have at least one fifth of the intensity of the maximum of such cluster. In addition, we considered a bright spot to be a cluster only if the intensity of its peak is at least 2000 arbitrary units above the background intensity—which corresponds to peaks with at least $\sim$ 20 monomers. Once the clusters are located and their boundaries defined, we add up the total intensity of each of them, separately, to obtain an estimate of the mass of each cluster, i.e., the total number of monomers in each of the clusters. This number might be slightly underestimated due to a potential self-quenching effect of the fluorescent tag upon aggregation.

A.3 Parameter Fitting

The two datasets that we want to fit with Eqs. (1.28) and (1.29), i.e., cluster density and mean cluster mass, have different units and numerical values. In what follows, we will introduce a least-square minimization such that, when minimising the squares to find the best fitting parameters, both datasets are equally taken into account. To achieve this, we introduce

$$\sum_i \left\{ \frac{1}{\mu_1} [f_1(x_i) - y_i^{(1)}] \right\}^2 + \sum_i \left\{ \frac{1}{\mu_2} [f_2(x_i) - y_i^{(2)}] \right\}^2 , \tag{A.1}$$

where $f_{1,2}$ are defined by Eqs. (1.28) and (1.29), the 2-tuples $(x_i, y_i)^{(1,2)}$ denote each of the datapoints i of each dataset (1 or 2, cluster concentration or cluster size), and $\mu_{1,2}$ are the mean values of the datapoints of each dataset: $\mu_{1,2} = {M_{1,2}}^{-1} \sum_{i=1}^{M_{1,2}} y_i^{(1,2)}$, $M_{1,2}$ being the number of datapoints.

Appendix B

Supplementary Information and Figures for the nucleoid model

B.1 Equilibrium free energy of an interacting system: the virial expansion

In this Appendix we briefly explain how to obtain the equilibrium free energy of a binary gas of hard spheres, as the one presented in Section 2.2.1 that yields Eq. (2.4). The model is composed of N_A and N_B particles of species A and B, which are hard spheres with radii R_A, R_B, respectively, confined in a volume $\mathcal{V}$. We assume that the system has cylindrical symmetry, and thus $\mathcal{V}$ is effectively one dimensional, so as to apply our findings to the model defined by Eqs. (2.1) to (2.3).

Without taking into account reaction processes, we work out the free energy from the partition function

$$Z = \int \Big(\prod_i d\mathbf{x}_i d\mathbf{p}_i\Big) e^{-\beta H(\mathbf{x},\mathbf{p})} = Z_\text{ideal}\frac{1}{\mathcal{V}^N}\int \Big(\prod_i d\mathbf{x}_i\Big) e^{-\beta V_\text{int}(\mathbf{x})}, \tag{B.1}$$

where $\mathbf{x}$ and $\mathbf{p}$ are the positions and momenta of the particles, Z_ideal is the partition function of the system in the absence of excluded-volume interactions, i.e., an ideal gas [Huang, 1987]. In order to compute this partition function we will resort to the virial expansion [Onnes, 1902]. We introduce the Mayer function f:

$$f(\mathbf{x}_i, \mathbf{x}_j) = e^{-\beta V_{ij}(\mathbf{x}_i,\mathbf{x}_j)} - 1, \tag{B.2}$$

where V_{ij} is the interaction potential between the ith and jth particle. Hence, the interaction term can be expanded as follows

$$e^{-\beta\sum_{i<j} V_{ij}(\mathbf{x}_i,\mathbf{x}_j)} = \prod_{i<j} e^{-\beta V_{ij}(\mathbf{x}_i,\mathbf{x}_j)} = \prod_{i<j}[1 + f(\mathbf{x}_i, \mathbf{x}_j)], \tag{B.3}$$

where, assuming low density, we expect that, on average, $f(x_i, x_j)$ is small compared to one. The leading contribution in Eq. (B.3) is $\mathcal{O}(1)$—the ideal gas—and the subleading one is $\mathcal{O}(f(x_i, x_j))$, which yields the second term in the virial expansion. Here we will also compute the third virial term, $\mathcal{O}(f(x_i, x_j)^2)$, because the density in our system is not low enough to allow us to safely neglect three-particle interactions. As shown in Fig. B.2, in some cases the second virial term does not yield the phase separation in the presence of reaction and out-of-equilibrium terms, while the addition of the third virial term provides such phase separation.

The second virial term consists of a summation of terms like the following

$$\frac{1}{\mathcal{V}^N}\int\prod_k d\mathbf{x}_k f(\mathbf{x}_i,\mathbf{x}_j)=\frac{1}{\mathcal{V}}\int dr_{ij}e^{-\beta V_{ij}(\mathbf{r}_{ij})}, \tag{B.4}$$

where we have used the fact that the system is one dimensional and integrated out all spatial variables except one, $\mathbf{r}_{ij}=\mathbf{x}_i-\mathbf{x}_j$. The potential V_{ij} depends on the relative position of species i and j considered, and here is taken to be a hard-sphere potential, which is non-zero only when the distance between two particles is smaller than the sum of their respective radii (R_i+R_j). Mathematically, it takes the following form:

$$V(\mathbf{r}_1,\mathbf{r}_2)=\begin{cases}0 & \text{if} \quad |\mathbf{r}_1-\mathbf{r}_2|\geq R_i+R_j\\ \infty & \text{if} \quad |\mathbf{r}_1-\mathbf{r}_2|<R_i+R_j\end{cases}. \tag{B.5}$$

For a given species i there are $N_i(N_i-1)/2$ combinations of the Mayer f function: as a result

$$\frac{N_a(N_a-1)}{\mathcal{V}}\int d\mathbf{r}_{ij}e^{-\beta V_{ij}(\mathbf{r}_{ij})}=-\frac{N_a(N_a-1)}{2\mathcal{V}}\frac{4\pi}{3}(2R_a)^3. \tag{B.6}$$

Conversely, as far as the hard-sphere interaction between species A and B is concerned, we have $N_\mathrm{A}N_\mathrm{B}$ terms, yielding

$$\frac{N_\mathrm{A}N_\mathrm{B}}{\mathcal{V}}\int d\mathbf{r}_{ij}e^{-\beta V_{ij}(\mathbf{r}_{ij})}=-\frac{N_\mathrm{A}N_\mathrm{B}}{\mathcal{V}}\frac{4\pi}{3}(R_\mathrm{A}+R_\mathrm{B})^3\equiv-\frac{N_aN_b}{\mathcal{V}}B^{(2)}_{\mathrm{AB}}. \tag{B.7}$$

In general, the term $B^{(n)}_{ab}$, which enters the n-th order of the virial expansion for species a and b, is known as virial coefficient. With this new notation, the partition function to order $\mathcal{O}(f(x_i,x_j))$ is

$$Z=Z_\mathrm{ideal}\left\{1-\left[\sum_{a=\mathrm{A,B}}\frac{N_a(N_a-1)}{2\mathcal{V}}B^{(2)}_{aa}+\frac{N_\mathrm{A}N_\mathrm{B}}{\mathcal{V}}B^{(2)}_{\mathrm{AB}}\right]\right\}, \tag{B.8}$$

where the partition function of an ideal gas of indistinguishable particles is

$$Z_\mathrm{ideal}=\frac{\mathcal{V}^N}{\prod_{a=\mathrm{A,B}}N_a!\Lambda_a^{3N_a}}, \tag{B.9}$$

and Λ_a is the thermal de Broglie wavelength of species a.

The third virial coefficient is obtained by adding up the contributions of the three-particle interactions in Eq. (B.1), i.e., terms of the form

$$\int d\mathbf{x}_id\mathbf{x}_jd\mathbf{x}_kf(\mathbf{x}_i,\mathbf{x}_j)f(\mathbf{x}_j,\mathbf{x}_k)f(\mathbf{x}_k,\mathbf{x}_i). \tag{B.10}$$

Proceeding along the same lines as for the first virial term, we count the number of times when such a term appears in Z, which depends on the chemical species of the particles involved. For clusters of three particles all of the same chemical species we have the following contribution:

$$-\sum_{a=\mathrm{A,B}}\frac{N_a(N_a-1)(N_a-2)}{6}\frac{160\pi^2R_a^6}{3\mathcal{V}^2}=-\sum_{a=\mathrm{A,B}}\frac{N_a(N_a-1)(N_a-2)}{6\mathcal{V}^2}B^{(3)}_{aaa}, \tag{B.11}$$

where the summation over a involves the different chemical species A and B. If two particles belong to the same species and a third particle to a different species, we have

$$\sum_{a\neq b} -\frac{N_a(N_b-1)N_b}{2}\frac{16\pi^2}{9\mathscr{V}^2}(8R_a^3R_b^3+15R_a^2R_b^4+6R_aR_b^5+R_b^6) = \sum_{a\neq b} -\frac{N_a(N_b-1)N_b}{2\mathscr{V}^2}B^{(3)}_{abb}. \tag{B.12}$$

These contributions are added up to the free energy of the system, to which higher order corrections can be introduced by computing the integrals involving products of 4 or more Mayer f functions.

Taking the logarithm of the partition function (B.8), once the third virial correction has been added, yields the free energy (2.4) in the main text.

B.2 Auxiliary entropy

In our analysis, we first determine the steady state of the system in the absence of reaction and out-of-equilibrium terms, by minimising the total free energy (2.20). These steady-state profiles will then be used as initial conditions to integrate forward in time the reaction-diffusion Eqs. (2.1) to (2.3), which include both reaction and out-of-equilibrium terms. At the free-energy minimum, the DNA concentration is nonzero in the nucleoid, while it vanishes outside the nucleoid. Given that these equilibrium profiles are entered as initial conditions in Eqs. (2.1) to (2.3), the vanishing concentration above causes numerical instabilities when these equations are numerically integrated forward in time, and can lead to negative concentrations in the out-of-equilibrium steady state [Shampine et al., 2005]. To overcome this issue, we included a small, additional entropic term in the free energy (2.20):

$$F_{\text{aux}} = K_{\text{aux}}\, k_{\text{B}}T \int_{-\ell}^{\ell} dx\, e^{-D_{\text{aux}} c_{\text{DNA}}(x)/\langle c_{\text{DNA}}(x)\rangle} \frac{c_{\text{DNA}}(x)}{N_{\text{DNA}}} \log\left[2\ell c_{\text{DNA}}(x)\right], \tag{B.13}$$

where $\langle c_{\text{DNA}}(x)\rangle = N_{\text{DNA}}/(2\ell)$ is the average DNA concentration across the cell, and K_{aux} and D_{aux} are constants that we set to $0.2\, N_{\text{DNA}}$ and 10, respectively, in order to obtain a negligible auxiliary free energy in the nucleoid and a non-negligible one at the poles. The exponential term in Eq. (B.13) is such that, if the coordinate x lies in the nucleoid bulk, then $c_{\text{DNA}}(x)$ is of the same order of magnitude as $\langle c_{\text{DNA}}(x)\rangle$ and thus, by choosing D_{aux} sufficiently large, the contribution to F_{aux} vanishes. On the other hand, the exponential is approximately equal to one outside the nucleoid, where $c_{\text{DNA}}(x) \ll \langle c_{\text{DNA}}(x)\rangle$, and Eq. (B.13) reproduces a term that resembles the standard entropy of mixing of a polymer in the mean-field approximation

$$\frac{c_{\text{DNA}}(x)}{N_{\text{DNA}}} \log\left[2\ell c_{\text{DNA}}(x)\right], \tag{B.14}$$

which dominates the integral in F_{aux}.

In what follows, we will add F_{aux} to the total free energy Eq. (2.20), by setting

$$F \to F + F_{\text{aux}}. \tag{B.15}$$

As a result, the minimization of Eq. (2.20) tends to also maximize the entropy (B.14), thus spreading out a fraction of DNA segments outside the nucleoid bulk. This procedure alters only slightly

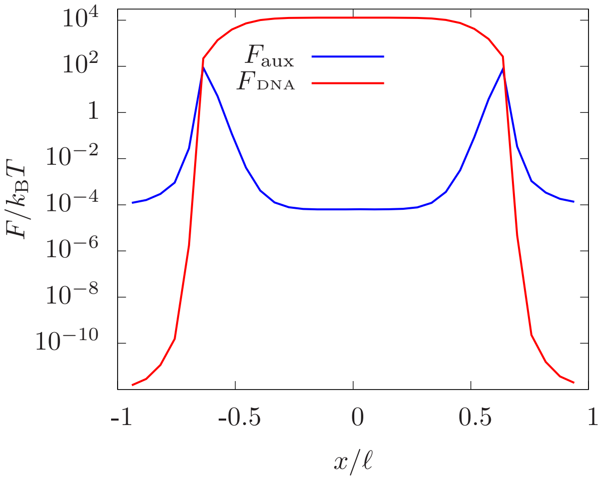


Figure B.1: DNA and auxiliary free energies. DNA and auxiliary free energies (F_{DNA} and F_{aux}, respectively)for a cell of size 3.6 μm at the out-of-equilibrium steady state. The DNA free energy only takes into account the self-interaction terms of DNA cylinders, because the mean-field entropic term of the DNA chain is neglected, see Section 2.2.2.

the free-energy minimum: As shown in Fig. B.1, the auxiliary free energy does not vanish outside of the nucleoid, but it is orders of magnitude smaller than the original free energy within the nucleoid. Notwithstanding this, such a small free energy prevents numerical instabilities in the integration of the reaction-diffusion equations.

B.3 Effect of third-order virial terms

Finally, we can evaluate the effect of adding a third virial coefficient in our model. In Fig. B.2, we depict the equilibrium minimum of the free-energy and the out-of-equilibrium steady state, to both second and third order in the virial expansion. There is a marked difference, particularly for the out-of-equilibrium steady state, where, without a third order term, the phase-separated nucleoid disappears due to the non-equilibrium synthesis of mRNA.

B.4 Estimate of the diffusion coefficient of DNA

In order to assess the validity of the value for the diffusion coefficient of DNA segments used in the model, we estimated the drag force exerted on the nucleoid by the viscous fluid surrounding it (which in the cell can be identified with the cytoplasm).

We consider a cylindrical nucleoid with radius $R - L/2$, assuming that there is a gap of width $L/2$ (the persistence length of DNA segments) in the radial direction between the nucleoid and the plasma membrane. For a nucleoid moving as a whole with velocity v_{nucl}, we assume the radial velocity profile of the viscous fluid surrounding the nucleoid that decreases linearly from velocity v_{nucl} at the nucleoid to zero at the membrane. We thus have a velocity gradient in the radial

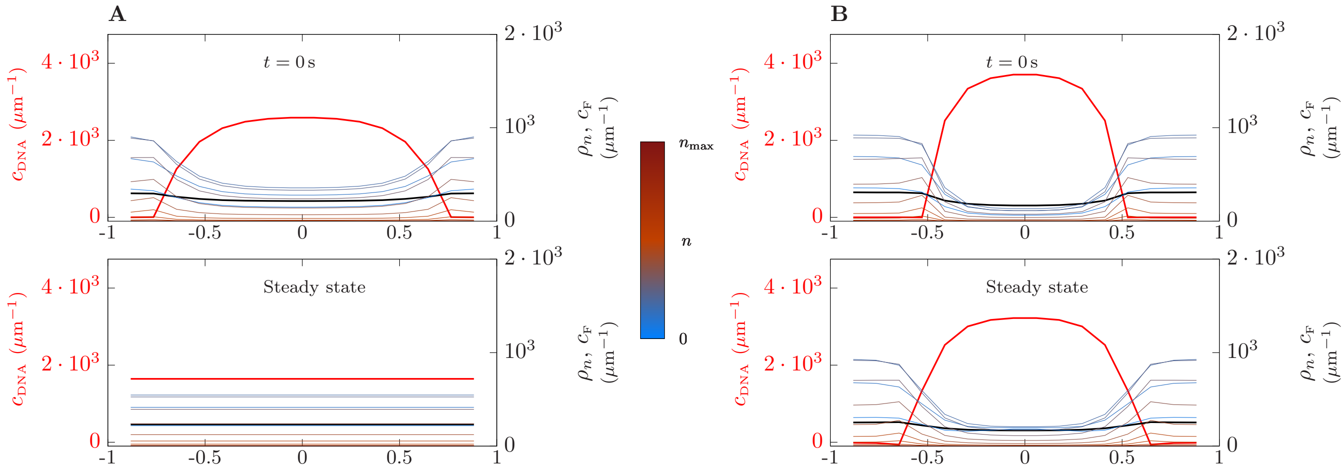


Figure B.2: Effect of third-order virial term. (*A*) Profiles corresponding to the free-energy minimization (top) and out-of-equilibrium steady state (bottom), obtained with the second virial coefficients only, for a cell length of 3.6 μm for single-chromosome growth. As in the main text, concentrations of DNA, $c_{\mathrm{DNA}}(x)$, and free ribosomes, $c_{\mathrm{F}}(x)$ are shown in red and black respectively and polysome concentrations $\rho_n(x)$ are also shown, where the mRNA loading number n is specified by the color box. (*B*) Same quantities as in A, with the third virial coefficients.

direction

$$\frac{\Delta v}{L} = \frac{2v_{\mathrm{nucl}}}{L} \tag{B.16}$$

from the nucleoid to the membrane. Then, the shear force F_{nucl} exerted on the nucleoid is [Lautrup, 2011]

$$F_{\mathrm{nucl}} = A_{\mathrm{nucl}}\,\nu\,\Delta v, \tag{B.17}$$

where ν and A_{nucl} are the viscosity of the fluid and the area of the nucleoid, respectively. For the reference cell of length 1.8 μm, the nucleoid will be around 1μm long and have a radius of $R - L/2$, which gives an area for the nucleoid $A_{\mathrm{nucl}} = 1.9\,\mu\mathrm{m}^2$. We use the viscosity of water $\nu \sim 1\,\mathrm{mPa}\cdot\mathrm{s}$, although the viscosity of the cytoplasm might be larger. The ratio between the force F_{nucl} and the velocity of the nucleoid is defined as the drag coefficient [Kubo et al., 2012]

$$\begin{aligned}\gamma &= \frac{F_{\mathrm{nucl}}}{v_{\mathrm{nucl}}}\\ &= \frac{2A_{\mathrm{nucl}}\nu}{L}\\ &\simeq 4.7 k_{\mathrm{B}}T\,\mathrm{s}/\mu\mathrm{m}^2.\end{aligned} \tag{B.18}$$

Given that the chromosome of the reference cell is composed of N_{DNA} segments, if one naively ignores the hydrodynamic coupling between segments, one could infer that the drag coefficient per segment is $\gamma_{\mathrm{s}} = \gamma/N_{\mathrm{DNA}}$. Then, by the fluctuation-dissipation relation, the diffusion coefficient of each DNA segment would be

$$D = \frac{k_{\mathrm{B}}T}{\gamma_{\mathrm{s}}} \simeq 10^3\,\mu\mathrm{m}^2/\mathrm{s}, \tag{B.19}$$

which is significantly higher than the value used in the model. While this estimate is clearly an oversimplification, the overall low drag of the nucleoid suggests that processes where the diffusion coefficient of segments could matter, i.e., nucleoid centering, are not limited by viscous drag (and, hence, by the diffusion coefficient), but by the osmotic-pressure difference.

Our actual choice of the diffusion coefficient for the DNA segments is significantly lower than the value (B.19) obtained with this estimate to be consistent with the measured diffusion coefficient of mRNAs that have a similar linear dimension while agreeing with the result of the estimate (B.18), namely, that viscous drag should not limit the dynamics of the nucleoid. Both in the centering dynamics and in the expansion of the nucleoid after halting transcription, the timescale is set by the synthesis and degradation of mRNA (β^{-1}), and not by the diffusion coefficient.

B.5 Experimental Methods

I include this section of the Appendix for completeness yet it is not my own work but that of my collaborators, as the experiments were carried out exclusively by Sophia Hsin-Jung Li.

In this study, we used *E. coli* wild type strain NCM3722. To achieve different growth rates, cells were cultured at 37°C in chemostats and in batch. For slow growth rates (0.1 and 0.6 h^{-1}), carbon-limiting chemostats with corresponding dilution rates were used, whereas for faster growth, batch cultures with glucose minimal media (0.9 h^{-1}) and defined rich media (1.7 h^{-1}) were employed. The chemostat (Sixfors, HT) volume was 300 mL with oxygen and pH probes to monitor the culture. The pH was maintained at 7.2 ± 0.1 and the aeration rate was set at $4.5\,l\,h^{-1}$. 40 mM MOPS media (M2120, Teknova) was used with glucose (0.4 %, Sigma G8270), ammonia (9.5 mM NH4Cl, Sigma A9434), and phosphate (1.32 mM K2HPO4, Sigma P3786) added separately. For the defined rich media, additional Supplement EZ 5X and 10X ACGU Solution (Teknova) were added. In carbon-limiting chemostats, glucose concentration was reduced to 0.08 %. All the sample collection happened after chemostat cultures reached steady state or when batch culture reached OD600 0.3.

To measure cell size, 750 μL of culture was fixed with 250 μL 20% paraformaldehyde at room temperature for 15 min, washed with PBS twice, and stored at 4°C until imaging. Then, 1 μL of cells were placed on 1% low-melting agar pad (Calbiochem) made with PBS and imaged with inverted Nikon90i epifluorescent microscope equipped with a 100 × 1.4 NA objective (Nikon) and Hamamazu Orca R2 CCD camera. NIS Elements software (Nikon) was used to automate image acquisition for phase contrast images. Segmentation, quantification of fluorescence intensity, and cell-length measurements were further analyzed in MATLAB [MATLAB, 2018] using customized programs.

To infer ribosome number per cell, cell number per OD600 and total RNA per OD600 were measured separately. Cell number per OD600 was calculated by serial dilution and plating. To measure total RNA, 1.5 mL of culture was pelleted by centrifugation for 1 min at 1.3×10^4 X g. The pellet was frozen on dry ice and the supernatant was taken to measure absorbance at 600 nm for cell loss. The pellet was then washed twice with 0.6 M HClO4 and digested with 0.3 M KOH for 1 hour at 37° C. The solution was then precipitated with 3 M HClO4 and the supernatant was collected. The pellet was re-extracted again with 0.5 M HClO4. The supernatants were combined and absorbance measured at 260 nm using Tecan Infinite 200 Pro (Tecan Trading AG, Switzerland). Total RNA concentration was determined by multiplying the A260 absorbance with 31

(μg RNA/mL) as the extinction coefficient.

B.6 Numerical methods

For both the minimisation of the free energy (2.20) and the time integration of the reaction-diffusion Eqs. (2.1) to (2.3), numerical methods are required, as the equations are too complex to be solved analytically. We discretise the spatial degrees of freedom of the system into a mesh that satisfies:

$$2\ell = \Delta x N_{\mathrm{d}}, \tag{B.20}$$

where Δx is the distance between two neighboring points and N_{d} is the number of points taken to describe the concentration profile of each of the species in the system, which was set to $N_{\mathrm{d}} = 32,\ 64$, depending on the desired accuracy.

Using this discretisation of space, we evaluated the spatial derivatives in Eq. (2.20), and obtained a minimum of the free energy by using an algorithm for constrained gradient-based optimization [Kraft, 1994]. We used the C implementation of the NLopt library [Johnson, 2007].

For the time integration with the non-equilibrium effects, we wrote down a set of ordinary differential equations, where to each chemical species and each point in the mesh (defined by Eq. B.20) corresponds a function of time, and such functions are coupled to neighboring points in space through the discretized spatial derivatives, and to other chemical species through the local chemical reactions. To solve this system we used an implementation of the backward differentiation formula (BDF) method in Mathematica [Wolfram Research, Inc., 2018].

B.7 Parameter estimation from experimental data

Cell length, cross-sectional radius and number of ribosomes were inferred from experimental measurements of *E. coli* colonies growing in different chemostatted conditions (as explained in Section B.5). The data yields the values of these parameters for different growth rates. The estimate of the number of ribosomes is derived from the total amount of 23S and 16S ribosomal RNA (rRNA), considering that two-thirds of the total mass of a ribosome comes from rRNA [Berg et al., 2002].

We analyzed twelve different nutrient limitations, which correspond to four groups of similar growth rates: the parameters—cell length, cross-sectional radius, number of ribosomes, and growth rate—were averaged across data points belonging to the same growth rate.

Finally, the values of the parameters above as functions of the growth rate were fitted with an exponential by using the least-square method. The parameter values for a growth rate of $\log(2)/2/\mathrm{hr}$, which corresponds to a doubling time of $2\,\mathrm{hr}$, were obtained via interpolation, by evaluating the exponential fit at the reference growth rate $\log(2)/2\,\mathrm{hr}^{-1}$, see Section 2.2.4. The interpolations are shown in Fig. B.3, and the values of the corresponding parameters are given in Section 2.2.4.

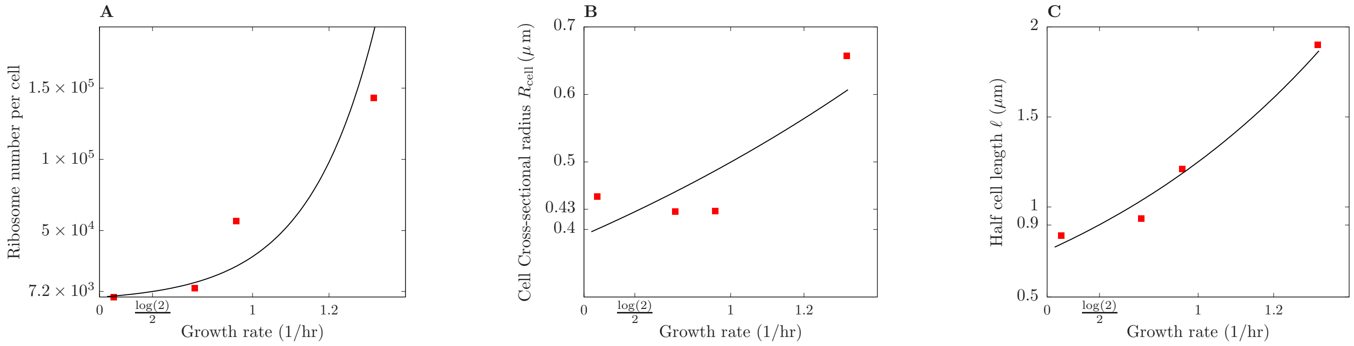


Figure B.3: Experimental data and interpolation. (*A*) Ribosome number per cell as a function of growth rate. Points correspond to experimental data, and solid curves to the exponential fit. The inferred parameters for the desired growth rate ($\log(2)/2\mathrm{hr}^{-1}$) are marked on each axis. (*B*) Same as *A*, for cell cross-sectional radius. (*C*) Same as *A*, for half cell length.

B.8 Nucleoid size at thermodynamic equilibrium

In this section we depict the results obtained for the filamentous growth scaling as in the main text, Fig. 2.4, but in the absence of the out-of-equilibrium chemical reactions, i.e., for a passive system (see Fig. B.4). The format of the Figure is the same as that of Fig. 2.4.

B.9 Single chromosome growth condition

B.9.1 Scaling of the concentration of chemical species for single-chromosome growth

In the single-chromosome case we assume that the concentration of mRNA and ribosomes scales linearly with the growth rate and, based on the data of Ref. [Kohram, 2021], that the growth rate of *E.coli* decreases linearly with cell length until it reaches zero at $\sim 20\,\mu$m. In order to compare the model predictions with the experimental data in Ref. [Wu et al., 2019], we assume the same linear law, but with a slope such that the growth rate, g, reaches zero at 30 μm, because in the data of Ref. [Wu et al., 2019] cells appear to grow up to that length. In addition, we assume that the mRNA degradation rate, β, decreases linearly in the same way the growth rate does, motivated by the expectation that for slow growth it would be inefficient to turn over mRNA quickly.

With the above considerations, we can write the following relations:

$$g(\ell) = g_0 - a\,\ell, \tag{B.21}$$

$$N_{\mathrm{mRNA}}(\ell) = 2\ell N_{\mathrm{mRNA}_0} g(\ell), \tag{B.22}$$

$$\beta(\ell) = \beta_0(g_0 - a\,\ell) \tag{B.23}$$

where, for a cell that grows up to 30 μm, $a = g_0/(15\,\mu\mathrm{m})$ and the constants g_0, N_{mRNA_0}, and β_0 are set by observing that the values of their respective functions must match those of the reference

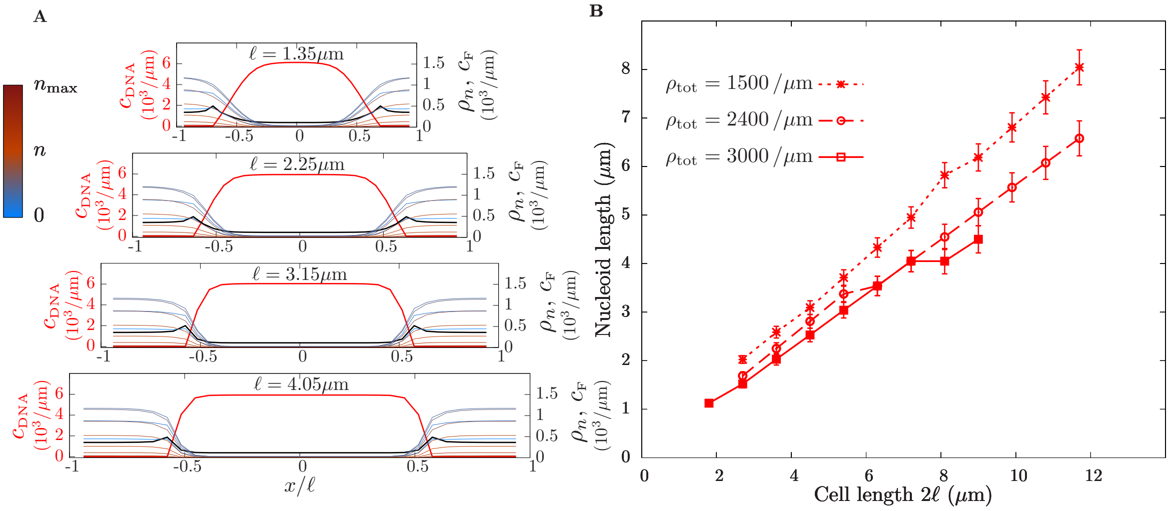


Figure B.4: Equilibrium concentration profiles for *E. coli* growing filamentously. This figure shows the concentration profiles and nucleoid lengths exactly as in Fig. 2.4 except that the non-equilibrium terms in the equations have not been added. Therefore, this figure represents the results of the model at thermodynamnic equilibrium.

cell of the main text, see Section 2.2.4. Those are:

$$g(\ell = 0.9\,\mu\mathrm{m}) = \frac{\log(2)}{2}\mathrm{hr}^{-1}, \tag{B.24}$$

$$\frac{N_{\mathrm{mRNA}}(\ell = 0.9\,\mu\mathrm{m})}{2\ell\sigma} = 2400\,\mu\mathrm{m}^{-3}, \tag{B.25}$$

$$\beta(\ell = 0.9\,\mu\mathrm{m}) = 3 \times 10^{-3}/\mathrm{s}. \tag{B.26}$$

Then, the mRNA synthesis rate, $\alpha(\ell)$, is fixed by the relation $\alpha(\ell)N_{\mathrm{DNA}} = \beta(\ell)N_{\mathrm{mRNA}}(\ell)$ see the main text, Section 2.2.4, where $N_{\mathrm{mRNA}}(\ell)$ and $\beta(\ell)$ are given by Eqs. (B.22) and (B.23).

B.9.2 Results for constant RNA concentration

This figure depicts the predicted nucleoid length for single-chromosome growth, as in Section 2.3.2 *Single-chromosome growth*, but keeping mRNA and ribosome concentration constant. As it can be seen in Fig. B.5, the predicted nucleoid length is much smaller than the measured one, which led us to conclude that mRNA and ribosome concentrations will fall with cell length in single-chromosome filamentously growing cells, which is consistent with the observed reduction in growth rate in Ref. [Kohram, 2021].

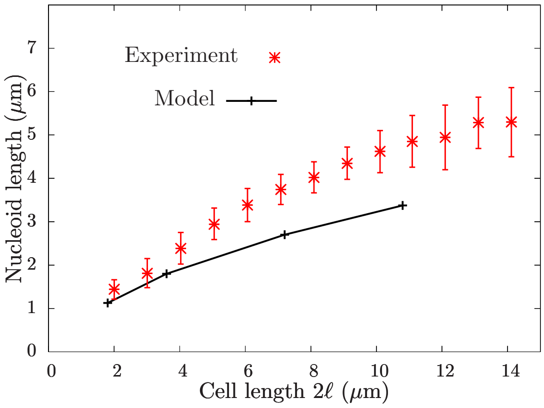


Figure B.5: Nucleoid size for single-chromosome filamentous cells with constant mRNA concentration. This Figure is analogous to Fig. 2.5 B in the main text. However, unlike the scaling of Section 2.3.2 *Single-chromosome filamentous growth*, here we consider cells which grow filamentously with a single chromosome, with constant mRNA and ribosome concentrations, independent of cell length. The predicted nucleoid length is small compared to the experimentally measured length, even for small cells.

Appendix C

Supplementary Notes on non-ideal Chemical Reaction Networks

C.1 Form of the propensity functions for a regular solution theory

We now consider a model of a solution based on a lattice where each chemical species (including the solvent) occupies one lattice site, effectively neglecting any differences in molecular volumes. Throughout most of this manuscript we will consider a free energy with virial coefficients to correct for interactions among particles, where the non-ideality of the systems stems from.

In a lattice of size V (number of lattice sites) occupied by N different species, provided that $\sum_{i=1}^{N} n_i = V$, the configurational entropy is given by

$$S = k_B \log \Omega = k_B \log \frac{V!}{\prod_i n_i!}, \tag{C.1}$$

where Ω is the number of microstates and k_B is the Boltzmann constant. The internal free energy of each species is given, as before, by the standard-state chemical potential μ_α^0. Moreover, in this regular solution model, we take into account interactions among neighbouring sites, whose energy, in a mean-field approximation with respect to space, is

$$\sum_{\alpha,\beta} \frac{\chi_{\alpha,\beta}}{2V} \left[n_\alpha n_\beta (1-\delta_{\alpha,\beta}) + n_\alpha (n_\alpha - 1)\delta_{\alpha,\beta} \right] = \sum_{\alpha,\beta} \frac{\chi_{\alpha,\beta}}{2V} \left[n_\alpha n_\beta - n_\alpha \delta_{\alpha,\beta} \right] \tag{C.2}$$

where $\chi_{\alpha,\beta}$ is the energy of the interaction between a particle α and a particle β but can also be interpreted as a matrix of virial coefficients.

Taken into account the previous considerations, the free energy for a homogeneous mixture of chemical species in the regular solution model reads

$$\begin{aligned} F =& U - TS \\ =& \beta^{-1} \left[\sum_\alpha \log(n_\alpha!) - \log(V!) \right] + \sum_\alpha \mu_\alpha^0 n_\alpha + \sum_{\alpha,\beta} \frac{\chi_{\alpha,\beta}}{2V} \left[n_\alpha n_\beta - n_\alpha \delta_{\alpha,\beta} \right]. \end{aligned} \tag{C.3}$$

With this expression of the free energy we can now derive an expression for the propensity functions, as given in eq. (3.12). The forward (or backward) rates are a function of the free energy difference of the complex:

$$\begin{aligned} &F(\mathbf{n}) - F(\mathbf{n} - \mathbf{r}^\rho) = \\ &\Delta F_{\text{id}} + \sum_\alpha r_\alpha^\rho \sum_\beta \frac{\chi_{\alpha,\beta}}{V} n_\beta - \sum_{\alpha,\beta} \frac{\chi_{\alpha,\beta}}{2V} r_\beta^\rho r_\alpha^\rho + \sum_\alpha \frac{\chi_{\alpha,\alpha}}{2V} r_\alpha^\rho \end{aligned} \tag{C.4}$$

where ΔF_{id} is the ideal part of the free energy difference of the complex, as given by eq. (3.14). In the last line of eq. (C.4), only the first term in $\chi_{\alpha,\beta}$ scales with the particle number, hence, in the thermodynamic limit, the rest of the interacting terms are negligible. In general, the propensity function in the stochastic system for an interacting regular solution is complicated, as seen in Eq. (C.4), but for a unimolecular reaction it simplifies to:

$$f_{+\rho}(\mathbf{n}) = k_\rho e^{\beta(F(\mathbf{n}) - F(\mathbf{n} - \mathbf{r}^\rho))} = k_\rho e^{\beta \mu_i} = k_\rho \frac{n_i}{V} e^{\beta(\mu_i^0 + \sum_j \frac{\chi_{i,j}}{V} n_\beta)}, \tag{C.5}$$

where i is the reactant of the reaction $+\rho$ and the definition of chemical potential (3.18) has been used.

In the thermodynamic limit, the deterministic rates can be written, according to eqs. (3.12) and (C.4) as follows

$$J_{+\rho} = k_{+\rho} \prod_\alpha c_\alpha^{r_\alpha^\rho} \exp\left(\sum_{\alpha,\beta} r_\alpha^\rho \chi_{\alpha,\beta} c_\beta \right), \tag{C.6}$$

where the discrete particle numbers have been replaced with continuous concentrations and the internal energy part of the chemical potential has been absorbed by the rate constant $k_{+\rho}$. However, the rates (C.6) can be written in the more general form

$$J_{+\rho} = k_\rho \exp\left(\sum_\alpha r_\alpha^\rho \mu_\alpha \right) \tag{C.7}$$

by identifying

$$\mu_\alpha = \log c_\alpha + \mu_\alpha^0 + \sum_\beta \chi_{\alpha,\beta} c_\beta. \tag{C.8}$$

We note that reaction rates of this form have already been suggested in this deterministic limit in the context of phase-separating battery materials [Bazant, 2013] and that a similar result for the reaction rates can be derived from activity coefficients (see Appendix C.2), which are common in the chemistry literature.

C.2 Form of the reaction rates in the non-ideal solution theory

Chemists have been dealing for long with non-ideal solutions, such as the one we are attempting at describing, and its chemical potential its typically written as [Kondepundi and Prigogine, 2015]

$$\mu_\alpha = \mu_\alpha^0 + \log(\gamma_\alpha(\mathbf{c}) c_\alpha), \tag{C.9}$$

where $\gamma_\alpha(\mathbf{c})$ is the activity coefficient which, itself, can depend on the vector of concentrations of the mixture $\mathbf{c}$. Note that, if we identify $\gamma_\alpha(\mathbf{c}) = \exp\left(\sum_\beta \chi_{\alpha,\beta} c_\beta\right)$ we recover the regular solution chemical potential in the thermodynamic limit eq. (C.8). Then, the reaction rate becomes

$$R_{+\rho} = k_{+\rho} \prod_\alpha (c_\alpha \gamma_\alpha)^{r_\alpha^\rho} e^{-\beta \mu^*}, \tag{C.10}$$

where we have made explicit the height of the barrier, μ^*, and implicit the standard chemical potentials of the reactants. μ^* can be interpreted as the free energy of the activated complex (*cf.* Fig. 3.1) and can be identified with $\mu_0^* + \log \gamma^*$, that, after absorbing μ_0^* into the reaction constant yields the rate [Madon and Iglesia, 2000]

$$R_{+\rho} = \frac{\tilde{k}_{+\rho}}{\gamma^*} \prod_\alpha (c_\alpha \gamma_\alpha)^{r^\rho_\alpha}. \tag{C.11}$$

This expression of the reaction rates from the point of view of non-ideal solutions has the advantage of providing an intuitive explanation for the effect of changes in concentrations on the reaction rates. Changes in concentrations will affect the rates not only through changes in the chemical potential of reactants and products but also, according to eq. (C.11), through the degree of solvation of the activated complex $*$ [Madon and Iglesia, 2000], parametrised by γ^*. This interpretation of the reaction rate gives an expression with a physical meaning for how the reaction rate should vary with the environment in which the reaction takes place.

C.3 Minimisation of the Lagrangian to obtain the phase diagram

In order to find the steady state of the solution we need to minimise the Lyapunov functional (or the Lagrangian, once particle conservation constraints are taken into account), Eq. (4.47). Substantial simplification is made by neglecting the contribution of the interfaces and considering the system as two homogeneous phases, which implies that the actual function that needs to be minimised is the Lagrangian (4.48).

First, we reduce the dimensionality of the problem by equating the derivatives of the Lagrangian with respect to the concentrations of the species:

$$\frac{\partial L}{\partial \phi_i^{(k)}} = \frac{\partial L}{\partial \phi_{i'}^{(k)}} = 0, \tag{C.12}$$

where i and i' are any two given chemical species in the system, and (k) refers to the different phases in the system. In practice, this yields relationships of the analogous to the equality of chemical potentials at equilibrium (which here include the shifted chemical potential term $\tilde{\mu}$ due to the non-equilibrium complex-balancing). For simple free energies like Eq. (4.49), we have

$$\phi_B = \phi_A e^{2\chi\phi_A + \tilde{\mu}_A - \tilde{\mu}_B}, \tag{C.13}$$

$$\phi_C = \phi_A e^{2\chi\phi_A + \tilde{\mu}_A - \tilde{\mu}_C}, \tag{C.14}$$

which reduces the problem to just three variables: $\phi_A^{(1)}$, $\phi_A^{(2)}$ and $V^{(1)}$.

We finally need to enforce stationarity condition of the Lagrangian with respect to volume:

$$\frac{\partial L}{\partial V^{(1)}} = 0, \tag{C.15}$$

which, together with

$$\frac{\partial L}{\partial \phi_A^{(1)}} = 0, \tag{C.16}$$

$$\frac{\partial L}{\partial \phi_A^{(2)}} = 0, \tag{C.17}$$

is a fully determined system for the unknown variables $\phi_A^{(1)}$, $\phi_A^{(2)}$ and $V^{(1)}$. However, the resulting conditions are transcendental equations which, in general, have no explicit analytical solution. Therefore, they need to be solved numerically. Even numerically, it is still a hard problem for parameter sets near the critical point, which is why in Fig. 4.6 the density of data-points around the critical point decreases.

Bibliography for the Appendices

[Bazant, 2013] Bazant, M. Z. (2013). Theory of chemical kinetics and charge transfer based on nonequilibrium thermodynamics. *Acc. Chem. Res.*, 46(5):1144.

[Berg et al., 2002] Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002). *Biochemistry*. New York: W H Freeman.

[Huang, 1987] Huang, K. (1987). *Statistical Mechanics, 2nd Edition*. Wiley, 2 edition.

[Johnson, 2007] Johnson, S. G. (2007). The NLopt nonlinear-optimization package.

[Kohram, 2021] Kohram, M. (2021). *Bacterial growth mechanisms and their role in cell size homeostasis and senescence*. PhD thesis, University of Pittsburgh.

[Kondepundi and Prigogine, 2015] Kondepundi, D. and Prigogine, I. (2015). *Modern Thermodynamics*. John Wiley & Sons, Ltd.

[Kraft, 1994] Kraft, D. (1994). Algorithm 733: TOMP–Fortran modules for optimal control calculations. *ACM Trans. Math. Softw.*, 20(3):262.

[Kubo et al., 2012] Kubo, R., Toda, M., and Hashitsume, N. (2012). *Statistical physics II: nonequilibrium statistical mechanics*, volume 31. Springer Science & Business Media.

[Lautrup, 2011] Lautrup, B. (2011). *Physics of continuous matter: exotic and everyday phenomena in the macroscopic world*. CRC press.

[Madon and Iglesia, 2000] Madon, R. J. and Iglesia, E. (2000). Catalytic reaction rates in thermodynamically non-ideal systems. *J. Mol. Catal. A Chem.*, 163(1):189.

[MATLAB, 2018] MATLAB (2018). The MathWorks Inc., Natick, Massachusetts, United States.

[Meyer, 1994] Meyer, F. (1994). Topographic distance and watershed lines. *Signal Process.*, 38(1):113.

[Onnes, 1902] Onnes, H. K. (1902). Expression of the equation of state of gases and liquids by means of series. *KNAW, Proceedings*, 4:125.

[Shampine et al., 2005] Shampine, L. F., Thompson, S., Kierzenka, J. A., and Byrne, G. D. (2005). Nonnegative solutions of ODEs. *Appl. Math. Comput.*, 170(1):556.

[Wolfram Research, Inc., 2018] Wolfram Research, Inc. (2018). *Mathematica, version 10.0*. Wolfram Research, Inc., Champaign, Illinois.

[Wu et al., 2019] Wu, F., Swain, P., Kuijpers, L., Zheng, X., Felter, K., Guurink, M., Solari, J., Jun, S., Shimizu, T. S., Chaudhuri, D., Mulder, B., and Dekker, C. (2019). Cell boundary confinement sets the size and position of the *E. coli* chromosome. *Curr. Biol.*, 29(13):2131.

RÉSUMÉ

Le cytoplasme d'une cellule est constitué d'une myriade de macromolécules comme des protéines des lipides et des enzymes. Chacune de ces molécules jouent un rôle bien spécifique au sein de la cellule, ce qui requiert un contrôle précis de leurs position spatiale. Dans la première partie de cette thèse nous proposons une explication théorique à deux processus où le milieu intracellulaire est organisé spatialement sous forme de motifs. Ces processus découlent tous deux du comportement collectif de ces molécules. Nous développons dans un premier temps un modèle qui a pour but d'étudier l'impact de l'encombrement stérique intracellulaire sur la cinétique d'agrégation de ces protéines. Nous montrons ensuite que les prédictions de ce modèle corrèlent de manière satisfaisant avec les données expérimentales de nos collaborateurs. Dans un deuxième temps, nous étudions la compaction et la localisation spatiale du chromosome de la bactérie *Escherichia coli* (*E. coli*). Nous expliquons la condensation du chromosome par les répulsion stériques entre l'ADN et d'autres macromolécules comme les ARNs messagers ou les ribosomes. Cette transition de phase s'assimile d'un point de vue théorique à la démixtion entre deux liquides. La localisation particulière du genome d'*E. coli* est quand-à-elle expliquée par l'activité transcriptomique des ARNs messagers. Cette hypothèse nous permet de reproduire la signature spécifique du chromosome d'*E. coli* in vivo, situé à la moitié de la cellule avant la division du chromosome et au 1/4 et au 3/4 après sa division. Dans la deuxième partie de cette thèse, nous construisons un cadre thermodynamique pour décrire les réseaux de réactions chimiques au sein de solutions non idéales. Cette approche nous permet de généraliser certains résultats de la théorie des réseaux de réactions chimiques idéales. De plus, ce cadre aide à formaliser les connections entre les réactions chimiques hors-équilibre et les séparations de phase pour les réseaux de type «complex-balanced». Ces derniers offrent un cadre moins restrictif à la théorie des réseaux de type «detailed-balanced». Nous caractérisons les réseaux de type «complex-balanced» par leur topologie. Nous montrons alors que cette topologie peut contraindre la dynamique des solutions. À l'inverse, notre théorie prédit que lorsque le réseau n'est plus de type «complex-balanced» des dynamiques plus exotiques peuvent apparaître.

MOTS CLÉS

Physique statistique hors-équilibre, biologie cellulaire, organisation spatiale, démixtion, solutions non-idéales, réseaux de réaction chimiques.

ABSTRACT

In the cytoplasm of a biological cell there are a myriad of different proteins, lipids and enzymes, each of them performing different tasks. The spatial organisation of these chemical species is crucial for the correct functioning of a cell. In the first part of this thesis we will explore, from a theoretical perspective, two processes where the intracellular medium is patterned and organised, whose common feature is the fact that they both stem from the collective behaviour of a large number of molecules. First, we develop a model for protein aggregation which studies the effect of intracellular obstacles on the coagulation kinetics. Our predictions are then successfully compared with experimental data obtained by our collaborators. The other example refers to the compaction and location of the bacterial chromosome. We suggest that the chromosome segregates from the rest of the cytoplasm, because of steric interactions between DNA and the intracellular crowders, by means of a mechanism reminiscent of liquid-liquid phase separation. Moreover, our study indicates that spatial localisation of the chromosome within the cell is dictated by non-equilibrium transcription of mRNAs (which are part of the crowding effect). Our model successfully reproduces the localisation pattern of the *Escherichia coli* chromosome, which is positioned at the center of the cell before the division of the chromosome and at 1/4 and 3/4 of the cell after division. Building on these examples, in the second part of the thesis we construct a thermodynamically consistent framework to mathematically describe chemical reaction networks in non-ideal solutions. This framework allows us to generalise the results from the classical theory of ideal networks, and aids in elucidating the connection between non-equilibrium chemical reactions and phase separation for a large class of networks, known as complex-balanced networks. Complex-balanced networks are a class of chemical reaction networks that is less restrictive than detailed-balanced ones. Given that complex balancing can be fully determined from the topology of the network, we discuss how this topological property of the network can constrain the dynamics of the solution and what behaviour one can expect when complex balancing is broken.

KEYWORDS

Non-equilibrium statistical physics, cell biology, spatial organisation, phase separation, non-ideal solutions, chemical reaction networks.