

# Information Theory in Transmembrane Transport

Présentée le 3 février 2023

Faculté des sciences de base  
Laboratoire de biophysique statistique  
Programme doctoral en physique

pour l'obtention du grade de Docteur ès Sciences

par

**Solange Marie FLATT**

Acceptée sur proposition du jury

Prof. F. Courbin, président du jury  
Prof. P. De Los Rios, directeur de thèse  
Prof. E. Frey, rapporteur  
Prof. K. Kruse, rapporteur  
Prof. S. J. Rahi, rapporteur



# Acknowledgements

A PhD is an enriching four-year journey, punctuated by numerous feelings and shared moments with so many people. A PhD is not merely equations and code, it is before all an incredible human adventure.

First of all, I would like to warmly thank Paolo, for giving me the opportunity to carry out my PhD in his laboratory. For almost five years, he has given me a highly valuable scientific support, guiding me with care towards more and more independence, always being available when needed. Above all, I sincerely appreciated his humanity, his caring and all the values that he places before scientific achievements. I am deeply grateful for this enriching experience.

I would like to thank Stefano and Daniel for all the time they gave me, for explanations, constructive advices and exciting discussions. They patiently supported me in the realisation of the projects and their role was absolutely essential.

Thanks also to all the lab members - Adélaïde, Alexandra, Bruno, Davide, Hugo, Mathieu, Satyam and Shiling - who have been part of the lab for weeks or months and who contributed to broaden my scientific horizons.

I had the honour to share these five years with Adélaïde, my office mate since the beginning. Our common adventure started a long time ago, in 2000, when we met at *La Garanderie*. 15 years later, here are we together at EPFL. We shared joys and sorrows, laughs and moments of discouragement. Every opportunity was taken to escape from science and reshape the world: the academic world, the world of cooking and many others. We spent so much time to discuss, sometimes with serious, often with derision. Definitely, this adventure would have been so less fun without you and I'm extremely thankful for your support.

If I dived into Biophysics a few years ago, it is thanks to Matthieu Wyart and Riccardo. My Master project was an exciting experience and I'm really grateful they made me want to continue the journey in this direction.

C'est dans ma langue maternelle que je suis le plus à l'aise pour exprimer ma reconnaissance envers mes proches qui m'ont accompagnée tout du long, depuis 27 ans, depuis 7 ans et tout spécialement depuis juin 2018.

Papa et Maman, un immense merci pour votre soutien protecteur sans faille, rempli de bienveillance. Chacun à votre manière, avec votre sensibilité propre, vous avez toujours été là pour me soutenir, m'encourager dans mes choix et me conseiller lorsque j'en avais besoin. Je sais que j'ai toujours pu compter sur vous, que je pourrai le faire à chaque instant et c'est le

## Acknowledgements

---

cadeau le plus précieux. Sans vous, je ne serais pas là où je suis maintenant et pour cela ma reconnaissance et ma gratitude sont infinies.

Je souhaite aussi remercier mes grands-parents, Papy et Mamie, qui ont montré un intérêt constant pour mes études, même si je mesure le côté abstrait des réponses que je pouvais leur donner. Leur fierté de grands-parents et leur admiration furent un vrai moteur tout au long de cette aventure.

Emmanuelle, depuis que tu as quitté l'EPFL pour voler vers de nouvelles aventures, tous les endroits du campus brillent à travers les nombreux souvenirs partagés qu'ils réveillent en moi. Les pauses café plus ou moins longues à l'Arcadie, les repas de midi au Vinci (ceux où le temps passe bien trop vite), les salles qui ont hébergé nos sessions de révision et j'en passe. Tout du long, j'ai pu profiter de notre inestimable complicité sororale pour avancer et franchir les obstacles qui se sont posés à moi, grâce à ton écoute et tes conseils toujours précieux et parfaitement à propos. Merci d'avoir rendu ces années si belles et merci de m'apporter tout ce que tu m'apportes chaque jour.

Romain, tu as l'immense mérite d'avoir su me gérer avec justesse et bienveillance au cours des différentes étapes de mon doctorat. Tu as toujours été là comme soutien, pour assurer mon équilibre et me permettre de continuer à avancer. Ton expérience que tu as partagée sans compter fut extrêmement enrichissante et tu as grandement contribué à mon évolution sur ces quelques années. Grâce à toi, me voilà armée pour continuer mon parcours professionnel mais aussi et surtout pour profiter pleinement, à tes côtés, de ce que la vie nous réserve de plus beau.

*Lausanne, December 20, 2022*

S. F.



## Abstract (English)

The past decades have seen the advent of information theory in various fields, from quantum physics to cosmology.

At an intermediary scale between atomic and cosmological scales are biological systems and in particular the cell, as a constitutive element of any living organism. Molecular mechanisms in the cell are tasked to perpetually create and maintain order, against the natural evolution suggested by the laws of thermodynamics. Many agree that these mechanisms are strongly related to the processing of information, but so far most of the established parallels remain at the stage of conceptual analogy. This leaves an obvious gap in a formal and rigorous description of cellular active processes: all the elements are in place at the biochemical level to process information. Through the internal structure of the involved macromolecules, biological systems are able to autonomously evolve, as a result of various information-processing steps. To this end, the present work articulates around two examples of cellular transport, in order to highlight the molecular mechanisms inherent to the processing of information.

In the first part, we establish a model for the transport by ABC Transporters and identify the conditions required for the creation of a concentration gradient across the membrane. The emerging conditions are explicit demonstrations of the different steps in the transport of substrate during which information is processed, induced by the structural and biochemical properties of the transporter. The conclusions extend way beyond a simple conceptual analogy: ABC transporters *are* autonomous Maxwell Demons.

In the second part, our work focuses on the dynamics of Hsp70-driven substrate translocation through membrane. Our model reveals that optimality lies in a balance between the strength of the directionality and the intrinsic diffusion rate of the substrate through the pore. These numerical observations pave the way to a new experimental exploration: the oligomerization of Hsp70 might accelerate the translocation, as an adaptative response to imposed conditions in the surroundings. Finally, we show how kinetic properties of chaperone proteins emerge from the resolution of a simple and minimal model for translocation. Such a model also displays a mathematical structure which brings us to state that the translocation machinery is an autonomous Szilard engine, continuously processing information.

Both projects converge to a bridge between information theory and non-equilibrium thermodynamics applied to active cellular transporters. The analogy between information-processing devices and biological molecular systems is not only conceptual: active cellular mechanisms

## Chapter 0. Abstract (English)

---

evolved to perpetually and autonomously process information. This results in the creation and maintenance of order, that acts as a cornerstone of life.

**Key words:** Information theory | Stochastic Thermodynamics | Kinetic model | Transmembrane transport | ABC transporter | Translocation | Hsp70.

# Résumé

Les dernières décennies ont marqué l'avènement de la théorie de l'information dans de nombreux domaines, de la physique quantique à la cosmologie. Les scientifiques de tous horizons ont montré un intérêt croissant pour cette théorie prometteuse au champ d'application très large.

Entre les échelles atomiques et cosmiques se trouvent les systèmes biologiques et en particulier la cellule, élément constitutif de tout organisme vivant. Les mécanismes moléculaires qui opèrent en son sein ont comme tâche de perpétuellement créer et maintenir de l'ordre, à l'encontre de l'évolution naturelle suggérée par les lois de la thermodynamique. Beaucoup s'accordent sur le fait que ces processus sont liés au traitement de l'information, mais jusqu'à présent, la plupart des parallèles se restreignent à des analogies conceptuelles. Cela laisse un vide évident dans une description rigoureuse des mécanismes cellulaires : au delà de l'analogie, tous les éléments sont présents au niveau biochimique pour un traitement de l'information autonome au sein des systèmes biologiques. C'est dans cette optique que ce travail s'articule autour de deux exemples de transport cellulaire, afin de mettre en exergue les mécanismes moléculaires inhérents au traitement de l'information.

Les deux objets de ce travail sont le transport cellulaire par les transporteurs ABC (*ATP-Binding Cassette*) à travers la membrane plasmique et la translocation de substrats médié par les protéines chaperones Hsp70 (*Heat Shock Protein 70*).

Dans la première partie, nous modélisons le transport cellulaire par les transporteurs ABC et identifions les conditions requises pour efficacement induire un gradient de concentration à travers la membrane. Ces conditions sont les manifestations explicites des différentes étapes du cycle de transport au cours desquelles de l'information est traitée par le système, de par sa structure interne biochimique. Ainsi, les résultats vont bien au-delà d'une simple comparaison conceptuelle : les transporteurs ABC *sont* des Démons de Maxwell autonomes.

La seconde partie se concentre sur la dynamique de translocation d'un substrat à travers une membrane par Hsp70. Notre modèle révèle que l'optimalité se situe dans un équilibre entre la directionnalité imposée à la translocation et le taux de diffusion intrinsèque du substrat à travers le pore. Ces observations ouvrent la voie vers une nouvelle exploration expérimentale : la dimerisation de Hsp70 pour accélérer la translocation, comme une adaptabilité aux conditions imposées par son environnement. Finalement, nous mettons en évidence comment les propriétés cinétiques des protéines chaperones, induites notamment par les interactions

## Chapter 0. Résumé

---

avec d'autres protéines, émergent de la résolution d'un modèle simple de translocation. La structure mathématique ainsi révélée nous amène à formuler que le mécanisme moléculaire de translocation est en réalité un moteur de Szilard revisité qui traite de l'information de manière autonome.

Les deux projets composants cette thèse convergent vers un lien entre la théorie de l'information et la thermodynamique hors-équilibre. L'analogie entre les mécanismes biologiques et les appareils qui traitent de l'information n'est pas que conceptuelle : les structures moléculaires et cellulaires ont évolué pour constamment traiter de l'information au niveau biochimique. Il en résulte la création et le maintien d'ordre au sein des systèmes biologiques, clé de voute essentielle à l'existence de la vie.

**Mots-clés :** Théorie de l'information | Thermodynamique Stochastique | Modèle cinétique | Transport transmembranaire | Transporteur ABC | Translocation | Hsp70.

# Contributions

*Initials:* **PDLR** Paolo De Los Rios / **SZ** Stefano Zamuner / **DMB** Daniel Maria Busiello.

Most of the thesis is written with the subject *we*, for the sake of a more fluent and comfortable readability. In this short paragraph, I want to stress out which were my contributions and the contributions of the different people with whom I collaborated during my PhD.

The grounding and general idea to bridge information theory and cellular transport comes from PDLR in his research plan for the Laboratory of Statistical Biophysics.

The initiation of the model for ABC transporters results from my personal discussions with PDLR and SZ. Following these different inputs, I performed all the analytical and numerical analysis. The part associated to information theory and stochastic thermodynamics results from enriching discussions with DMB, under the supervision of PDLR.

The model for the dynamics of translocation was essentially built on my own, with the advices from PDLR on my request. I performed all the numerical results and oriented the model towards the idea of Hsp70 oligomerisation. I solved analytically the simplified model of translocation to let the logics of transport emerge, with ponctual advices of DMB and PDLR. I drew the analogy with the Szilard Engine, following discussions with PDLR.



# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract (English)</b>	<b>iii</b>
<b>Résumé</b>	<b>v</b>
<b>Contributions</b>	<b>vii</b>
<b>List of Figures</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Towards a fingerprint of Life . . . . .	1
1.2 A challenge to the second law of thermodynamics . . . . .	4
1.3 A first insight in cellular transport . . . . .	12
<b>2 Basic principles on stochastic thermodynamics and kinetic models</b>	<b>15</b>
2.1 From classical to stochastic thermodynamics . . . . .	16
2.2 Time evolution of the system . . . . .	18
2.3 State probability and net flux from kinetic scheme . . . . .	22
2.4 Detailed balance and microscopic reversibility . . . . .	25
2.5 Information entropy . . . . .	26
2.6 Entropy production along a stochastic trajectory . . . . .	27
2.7 Measurement and information . . . . .	29
2.8 Fluctuation theorem and time reversibility . . . . .	31
2.9 ATP as energy source of primary active transporters . . . . .	34
2.10 Thermodynamic forces . . . . .	36
<b>3 ABC transporters as autonomous Maxwell Demons</b>	<b>39</b>
3.1 Biological and structural characterisation of ABC transporters . . . . .	39
3.2 Construction and motivation of the model . . . . .	44
3.3 Characteristic parameters to capture the phenomenology . . . . .	47
3.4 Logics of transport . . . . .	48
	ix

## Contents

---

3.5	Energy cost of information processing . . . . .	53
3.6	Numerical results . . . . .	56
3.7	Reproduction of experimental phenomenology . . . . .	62
3.8	Conclusive discussion . . . . .	70
<b>4</b>	<b>Substrate Translocation induced by Entropic pulling</b>	<b>73</b>
4.1	Biological context for protein translocation . . . . .	73
4.2	Different models to explain the translocation mechanism . . . . .	76
4.3	A kinetic model for protein translocation . . . . .	79
4.4	Translocation as a balance between force and binding frequency . . . . .	87
4.5	Pulling stronger, but less frequently . . . . .	89
4.6	Hsp70 oligomerisation in protein folding . . . . .	92
4.7	Extended model with Hsp70 dimerization . . . . .	94
4.8	Stronger together . . . . .	97
4.9	Analytical study of a simple model of translocation . . . . .	99
4.10	Translocation as an autonomous Szilard Engine . . . . .	106
4.11	Conclusive discussion . . . . .	112
<b>5</b>	<b>Conclusion and outlooks</b>	<b>115</b>
<b>A</b>	<b>Appendix</b>	<b>119</b>
A.1	ABC Transporters . . . . .	119
A.2	Translocation . . . . .	126
	<b>Bibliography</b>	<b>150</b>
	<b>Curriculum Vitae</b>	<b>151</b>



# List of Figures

1.1	Maxwell Demon and Szilard engine thought experiments . . . . .	8
1.2	Experimental realisation of a Szilard-type engine . . . . .	11
2.1	Example of a kinetic model with four states . . . . .	20
2.2	List of spanning trees of a graph . . . . .	22
2.3	List of oriented spanning trees of a graph towards a given vertex . . . . .	23
2.4	Illustration of a stochastic trajectory between states . . . . .	27
2.5	Transition between ATP-bound and ADP-bound states of a substrate . . . . .	35
2.6	Kinetic model for the transport of one substrate molecule through a membrane	36
2.7	Kinetic model for the transport of two substrates as an example of free-energy transduction . . . . .	38
3.1	Protein structure of ABC transporters . . . . .	41
3.2	6-state kinetic model for ABC transporters . . . . .	45
3.3	Analogy between ABC transporters and the Maxwell Demon experiment . . . . .	52
3.4	Measurement and resetting between substrate-bound and -unbound states . . . . .	57
3.5	Energy balance for ABC transporters . . . . .	58
3.6	Concentration gradient as a function of the available energy . . . . .	59
3.7	Concentration gradient as a function of the ratio of equilibrium constants $K_e$ and $K_e^S$ . . . . .	60
3.8	Concentration gradient as a function of the ratio of hydrolysis rates $\eta$ . . . . .	62
3.9	Optimal value of $\eta$ as a function of equilibrium constant $K_e$ . . . . .	63
3.10	Microscopic reversibility and net synthesis rate close to equilibrium . . . . .	64
3.11	Trans-inhibition: decrease of the hydrolysis rate upon addition of substrates on the target side . . . . .	66
3.12	Conserved logic of ABC transporter in a non-linear model . . . . .	68
3.13	Saturation of the performance of the transporter upon increase of substrate on the target side . . . . .	69
4.1	Translocation channel in the ER and the mitochondrion . . . . .	74
4.2	Structure of Hsp70 as mediators for the translocation . . . . .	75

## List of Figures

---

4.3	Three models for the translocation mechanism by Hsp70 . . . . .	77
4.4	Kinetic toy model for the dynamics of translocation . . . . .	81
4.5	Validation of two simulation parameters: the number of sites and the state of incoming binding site upon backward translocation . . . . .	88
4.6	Translocation rate as a function of the available energy . . . . .	89
4.7	Translocation rate as a function of the entropic pulling force $F_1$ . . . . .	90
4.8	Translocation rate as a function of the protein diffusion rate $k_0$ . . . . .	92
4.9	Comparison between the translocation rate and the average pulling force exerted on the protein . . . . .	93
4.10	Oligomerized structure of Hsp70 . . . . .	94
4.11	Extended kinetic model with Hsp70 dimerisation . . . . .	95
4.12	Acceleration of the hydrolysis rate induced by Hsp70 dimerization . . . . .	97
4.13	Average number of Hsp70 on the first binding site . . . . .	99
4.14	Comparison of translocation rate between the two models, for different diffusion rates $k_0$ . . . . .	100
4.15	Contributions to the translocation rate, with one or two Hsp70s on the first bound site . . . . .	101
4.16	Minimal model for an analytical study of translocation . . . . .	102
4.17	Information-driven translocation . . . . .	108

# 1 Introduction

## 1.1 Towards a fingerprint of Life

It is a bewildering, dizzying question I would like to start this introductory chapter with. To set the context, it could not be better asked than through the title of E. Schrödinger's book, "What is Life?" [1]. This question might seem pretentious, all the more if it is raised with the certainty that a clear and non-refutable answer exists. Rightly, in the difficulty to answer it might be the charm and attractiveness of this question. The same attractiveness which has motivated for centuries great thinkers from broad horizons: biologists, physicists, philosophers, ...

As an important preliminary remark, I want to point out that, more than ever, each step towards an answer to the above-raised question instantaneously opens so many other questions. The subjectivity associated to my choice to guide the discussion between non-Life and Life, is fully assumed. Most of all, I dare not pretend to deal with the subject of such a high complexity in a global and exhaustive way. Many works, books, articles have been amazingly well written, each with a different approach, some of them being quoted hereafter.

Among the numerous possible starting point for such a discussion, let's refer to a definition of Life. One among many others. The definition for "Life" given in the Britannica encyclopedia is: *"Matter that shows certain attributes that include responsiveness, growth, metabolism, energy transformation, and reproduction. Although a noun, as with other defined entities, the word life might be better cast as a verb to reflect its essential status as a process"* [2]. In addition to raising the interesting distinction between Life as a state or as a process, this definition essentially stresses different biological criteria for life that stay to be very large-scale and general, but we want and need to go further to grasp the intrinsic property of Life way more fundamentally. Another definition provides smaller-scale criteria, which are of greater interest within the framework of this project: *"An organism is said to be living if it exchanges matter and energy with its surrounding while conserving its autonomy, when it reproduces and evolves by natural*

## Chapter 1. Introduction

---

*selection*"<sup>1</sup> [3]. In this definition are the notions of exchange of energy with its surrounding and autonomy, that will turn out to be at the heart of the different projects in this work. Although these two notions are mostly associated to Life as an existing principle, the counterpart in the above definition is evolution, which rises another approach, maybe the one we should logically start with. To quote an elegant formulation by the biologist Gian-Paolo Dotto: "*What is the driving force that makes an apparently dead seed wake up under the sun, draw water, and germinate into a plant?*" [4]. In other words, what is the founding event during which Life evolved from non-Life?

A partial answer might possibly reside in the association of two concepts, as contradictory and complementary they paradoxically seem: Chance and Necessity<sup>2</sup>, as a reference to the famous eponymous book by Jacques Monod [6], who himself refers to the philosopher from antiquity Democritus<sup>3</sup>. On the one hand is the Chance: it took a combination of circumstances that might be considered to be impossible, but that nevertheless happened. The one which made that, from an ensemble of elementary particles, emerged Life. Some will be tempted to pretend that, at this precise moment, *God did play dice*<sup>4</sup>. On the other hand is the Necessity. The one that makes that, for scores of reasons, that can be physical, biological or chemical, this *had to* happen like that, and it could not be different, so that all the conditions required for the emergence of Life were combined. In this same book, Jacques Monod establishes as a necessary criterion for Life the notion of teleonomy. "*The notion of teleonomy implies the idea of an oriented, coherent and constructive activity. By these standards proteins must be deemed the essential molecular agents of teleonomic performances in living beings*" [6]. We are directly projected to a much smaller scale than where we started from with the initial definition, a scale more instructive to find the fingerprints of constitutive elements of Life: already at the scale of a few hundreds or thousands of atoms, are articulating the first workings of Life. However, this criterion leaves a feeling of incompleteness that might be transcribed as follows: an activity that is *oriented, coherent and constructive* is a starting point to satisfactorily develop the notion of Life, but with respect to what are these three terms defined? Shall they find their base intrinsically to the protein itself? Shouldn't these molecular agents be considered as being part of a more global environment, which would be the only one susceptible to define, through its structure, the criteria and conditions to finally lead to an oriented, coherent and constructive operation? That is one possible hypothesis, of course not the only one, and it is corroborated in the book by Sara Walker and Paul Davies who suggest the existence of a hierarchical organisation, in which "*life may be characterized by context-dependent causal influences, and, in particular, that top-down (or downward) causation where higher levels*

---

<sup>1</sup>Original formulation in French: "*Un organisme est dit vivant lorsqu'il échange de la matière et de l'énergie avec son environnement en conservant son autonomie, lorsqu'il se reproduit et évolue par sélection naturelle.*"

<sup>2</sup>Original title in French: "*Hasard et Nécessité*"[5]

<sup>3</sup>"Everything existing in the universe is the fruit of chance and necessity", Democritus (460-370 BC)

<sup>4</sup>In reference to Einstein's famous quotation "Gott würfelt nicht" [7], although in a totally different context.

*influence and constrain the dynamics of lower levels in organizational hierarchies" [8]. From the fingertip, we touch the heart of the problem and, by extension, the *raison d'être* of this project. Indeed, how is this top-down causality "mediated" from one scale to another, for instance from a biological environment to a protein in it? Information. The word is out and, with it, a *Pandora's box* is opening. Well ahead of a more mathematical and formal approach, a first conceptual overview is necessary to introduce the topic.*

Information might, in some sense, be associated to the "medium" (although immaterial) that supports the top-down causality between different scales of a biological systems, from the environment to the protein. Thanks to this transmission of information, the proper activity of a protein can be described as being oriented and coherent, thus creating order at the molecular level and, by extension, cellular level. Such a time-evolution is far from being the natural dynamics of physical and isolated system, which on the contrary tend to evolve towards the most disordered state. In this respect, Loewenstein argues that *"this information flow, not energy per se, is the prime mover of life that molecular information flowing in circles brings forth the organization we call "organism" and maintains it against the ever-present disorganizing pressures in the physics universe" [9]. Through a more physical and formal approach, Schrödinger goes along with this view and refers to the other extremity of the timeline of any living organism: death. He raises the question to determine "how would we express in terms of statistical theory the marvellous faculty of a living organism, by which it delays the decay into a thermodynamical equilibrium (death) ? It feeds upon negative entropy, attracting, as it were, a stream of negative entropy upon itself, to compensate the entropy increase it produces by itself, to compensate the entropy increase it produces and thus to maintain itself on a stationary and fairly low entropy level" [1].*

The creation of order at the molecular level, associated to operations oriented with a specifically defined goal are at the heart of the emergence of Life and its evolution over time, until it unavoidably reaches its very end. Emergence of life *"may correspond to a physical transition associated with a shift in the causal structure, where information gains direct and context-dependent causal efficacy over the matter in which it is instantiated" [8]. Overall, information associated to this organisation is universal. Indeed, "information even in its functional variant is not limited to humans, but is intrinsic to all forms of life and their devices. Moreover, the goal-directedness of functional information is an objective and universal property of the systems involved" [10].*

The pathway which historically led to understand the notion of information such as briefly introduced in the present paragraph was far from being straightforward. Many steps were necessary to understand that it could be included in a perfectly autonomous and independent manner in the formalism of Physics, and in particular consistently with the law of thermodynamics. Initially, the ability to process information (without being so named) was first attributed to mysterious agents whose properties were puzzlingly anthropomorphic, as if only humans were able to process information. This way to consider the question turned out to

be wrong: in reality, human beings exhibit themselves a fundamental mechanism that takes place at the atomistic and molecular scale. They don't benefit from a general ability that they have to transfer without evidence to molecular systems to explain their proper functioning. This projection into an agent able to process information to such small scales is the object of the next section and, as we will see, was the starting point of a very long way to finally unify information theory and thermodynamics.

This short historical perspective shed light on different elements that aroused my curiosity. My hope was to give a flavour of the merits of studying the molecular mechanisms in biological processes under the angle of their strong interconnection with the processing of information.

As a conclusion, one might ask whether it is possible to answer the question raised by Schrödinger which acts as a starting point of this introduction. I will try to give one through a quotation. An answer whose brevity and its "straight to the point" formulation leave the door open to many other fascinating reflections. "*Life is the universe developing a memory, and our chemical detection system could find it*" [11].

### 1.2 A challenge to the second law of thermodynamics

The second law of thermodynamics is one of the most fundamental laws of physics, governing the evolution of any type of system, from microscopic ones in living organisms, to the evolution of stars and galaxies. It is even believed to hold *the supreme position among the laws of Nature*, according to Arthur Eddington [12]. Coming back to its origins, the second law of thermodynamics was first formulated two centuries ago by Clausius (1850) and Kelvin (1851), inspired by Carnot's work 25 years earlier. Formally, there is no single formulation to the second law, far from it, and there is even less a single interpretation of the various conceptual implications.

This section does not aim, or pretend, to develop in depth the discussion around the second law as well as the similarities and differences between the multiple definitions and various conceptual ways to approach it. However, it is important to present an anchorage point on which the apparent paradox that preceded the emergence of information theory finds its roots. The second law of thermodynamics, according to a widely spread formulation is:

*No device, operating in a cycle, can produce the sole effect of extraction a quantity of heat from a heat reservoir and the performance of an equal quantity of work.*

Kelvin, 1882

In other words, this formulation states that in any system operating cyclically, it is impossible

## 1.2 A challenge to the second law of thermodynamics

---

to produce a work that is larger than the heat extracted from a reservoir. In the middle of the above quotation, the term *sole effect* is more important than what it might seem. It means that the heat flux cannot be induced by any agent or process, external to the device, that necessarily must operate in an independent and isolated manner. By extension, such a formulation unambiguously broke the dream that had likely lived within the hopes and believes of scientists at that time: being able to conceive and build a perpetual machine that nothing would stop, this being achieved in absence of any external operating system. Taking a step back with respect to the notion of cyclic process and simultaneously focusing on the directionality of heat fluxes, an alternative and slightly prior formulation was established by Clausius and is so stated:

*No process is possible for which the sole effect is that heat flows from a reservoir at a given temperature to a reservoir at higher temperature.*

Clausius, 1854

As a corollary, if there is no temperature gradient in a system, there is no heat flow made possible without the action of any external device.

Historically, the notion of information (although not formulated in these words) was first introduced in thermodynamics by Maxwell in the late nineteenth century with a *Gedanken experiment* [13], that was a posteriori named as Maxwell Demon. This opened a new era in the description of physics, that would lead the scientific community to grasp the crucial role played by information and open many questions about how to reconcile this vision with the current tools of physics, in particular thermodynamics. Not only this question remained unsolved for decades, but the debate and discussions were additionally fed with a new thought experiment in 1929, named the Szilard engine. The concept of both thought experiments is developed in the next paragraphs. These revolutionary ideas emerged in a scientific world in which the second law of thermodynamics is or at least seems sacrosanct. In 1935, Arthur Eddington provocatively stated that "*if your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation*" [12]. Such a statement reflects to which extent the second law of thermodynamics is of a vertiginous depth although it is not totally proven at that time. Luckily enough for Science, there are often great thinkers, midway between humility and pretention, who take the risk *to collapse in a deepest humiliation*: a few decades before Eddington, James Clerk Maxwell (1831-1879) came to challenge the scientific community, rightly about the second law of thermodynamics.

In the next paragraph, we take the liberty of a slight deviation from the original formulation of the Maxwell Demon thought experiment, with the aim of introducing it in a slightly more intuitive and pictorial way the original formulation being described at the end of the paragraph.

## Chapter 1. Introduction

---

A box is filled with particles. In the middle of the box is a wall with a trapdoor, on which an idealized agent can operate. This agent was later named according to its creator as the Maxwell Demon. At the very beginning, the system is supposed to be at equilibrium, that is with equal concentrations of particles on the two sides of the box. The aim for the Maxwell Demon is to isolate all the particles on the same side of the box, let's say arbitrarily on the right one, in order to create a concentration gradient. This is done by iteratively operating using a well-defined cyclic process described in figure 1.1A. The different steps are the following. Starting a cycle from an initial situation with particles on both sides of the middle wall (A1), the first step is to detect a particle approaching the door from the left side (A2). The result following the detection is kept in memory (A3), allowing the Demon to operate consecutively. The next operation by the Demon is to open the door to let the particle go from the left to the right side (A4). Once it is done, the door is closed and everything is reset (A5) so that a new cycle can be operated. The Demon is also able to detect particles on the right side, but no opening of the door follows such an observation, keeping the system unchanged for one more cycle. At the end of the iterations, the final outcome is that the concentration of particles is higher on the right side than on the left side (which is theoretically zero).

Coming back to the very original formulation, Maxwell describes a box at uniform temperature in which particles have various velocities. A threshold value for the velocity is defined, so that two types of particles are defined, the fast and slow ones. According to the second law, it is impossible to create a temperature gradient in the box without expenditure of work. By introducing an operable door in the middle of the box, the Demon is able to create a temperature gradient without paying the energy cost that is assumed by the second law. This would be done by operating on the door, coordinately with what is observed about the position and velocities of each particles. After a sequence of such operations, all the slow particles are isolated on one side of the box and the fast one on the other, thus obviously leading to a temperature gradient.

Whatever the version of the Maxwell Demon experiment is chosen, the entropy of the system would thus be decreased, in contradiction with the second law. Behind the experiment is the central question of understanding why a Demon can never operate beyond the apparently fundamental limits imposed by the second law, no matter how intelligent he is. The Maxwell Demon experiment also intrigued because it reopened the utopian perspective on the construction of a perpetual-motion machine. With the hypothesis of existence and feasibility of such a Maxwell Demon, the procedure for perpetual motion, extracting mechanical work from temperature or pressure gradient, is relatively simple and has been discussed by Ehrenberg [14].

During decades, physicists struggled with this apparent conceptual contradiction, without being able to formulate and explain the missing element to the puzzle, the one that would be the link between the Maxwell Demon thought experiment and the second law of thermodynamics. While some try to cripple the legitimacy of the Maxwell Demon by proposing different arguments, others go in the reverse direction, bringing to the table of scientific advances



## 1.2 A challenge to the second law of thermodynamics

---

diverse revisions of the Maxwell's experiment, in which the Demon itself is not even required to implement a similar process. We illustrate it through one example, the automated pressure Demon proposed by Smoluchovski [15, 16]. The setup is very similar to the one illustrated in figure 1.1A. The exception is that there is no Demon operating on the door, but the door itself is built as a spring-loaded trapdoor, acting as a one-way valve, since only particles coming from one side can induce the door opening. The resolution of this automated Maxwell Demon was extensively studied by Skordos et al. in [17]. The numerical modelling of such an experiment showed that the only way, with such a set up, to effectively induce a pressure gradient between the two sides of the door is to systematically remove thermal energy from the door. This de-thermalization prevents random opening that would lead to a flux of particles in the undesired direction. Otherwise, the system stays at equilibrium, without any concentration gradient between the two sides of the door.

Contributions followed one another over the years to build a new formalism able to solve the paradox raised by the Maxwell Demon and thus make it consistent with the second law of thermodynamics. A huge step into this direction was initiated by Leo Szilard (1898 - 1964), who proposed a new thought experiment, the Szilard engine (1929) [18]. More than directly proposing the full solution, this thought experiment established a new framework to guide the reflexion around the apparent contradiction to the second law of thermodynamics, leading many years later to the connexion between thermodynamics and information theory.

In his work, Szilard proposes a new thought experiment, 50 years after Maxwell's one, with a clear objective: to show that the intervention of an intelligent being, apparently allowing the construction of a perpetual machine, is the manifestation of a measurement and memory ability, hence causing the apparent violation of the second law [18]. Before going into the details of this interpretation, let's have a closer look to the experiment itself, known as Szilard engine and schematically illustrated in figure 1.1B.

A cylinder (closed on both ends) is filled with gas at fixed temperature and is associated to a large heat reservoir (not depicted in figure 1.1B). The gas enclosed in the cylinder consists of a single particle and a partition can is inserted in the middle of the cylinder (B1). The Demon detects on which side of the cylinder is the particle, either right or left (B2). The result is stored: this is done by connecting a load through a frictionless pulley on the same side than the one with the particle (B3). As a consequence, the gas expansion driven by the particle is converted into the lift of the load, so that work is performed on it (B4). At the end, the system is reset by disconnecting the load from the system as well as removing the partition and inserting it again in the middle of the cylinder (B5). A new cycle can then be initiated.

This setup can be imagined so that the mechanical displacement of the piston is performed freely. The expression of an external intervention simply consists in the insertion or removal of the partition as well as the connexion between the system and the load so that the partition displacement is converted into a work extraction. If the external agent acts on the system so

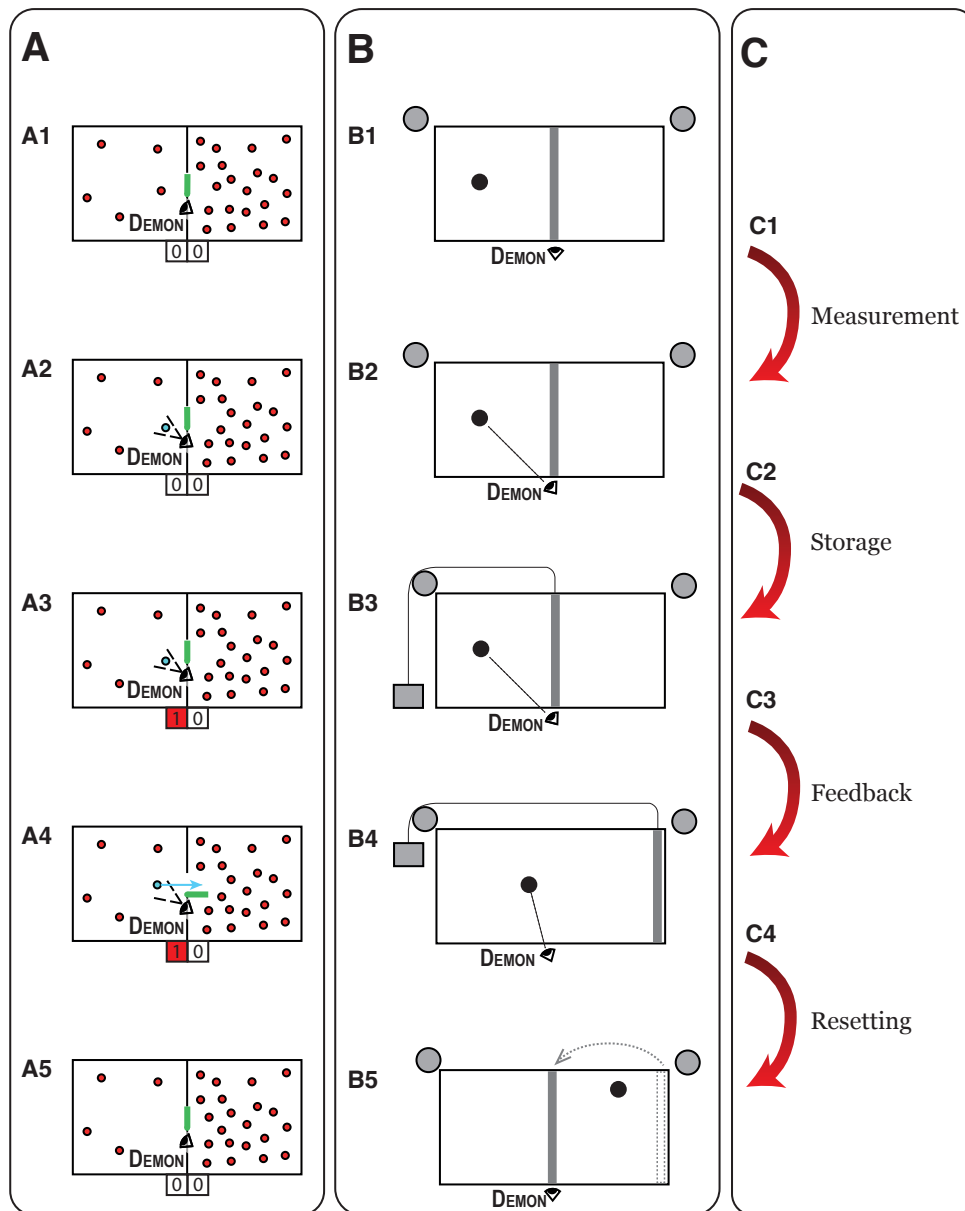


Figure 1.1: Schematic illustration of two major thought experiments: (a slightly simplified version of) the Maxwell Demon (A) and the Szilard engine (B). They are depicted according to the main steps that can be isolated in an information-processing device: measurement, storage, feedback and resetting (C).

## 1.2 A challenge to the second law of thermodynamics

---

that, at every iteration, the gas follows an isothermal expansion (ensured by the heat reservoir), then work is extracted from the system at each iteration of the cyclic process. As for the Maxwell Demon, the apparent contradiction with the second law of thermodynamics is evident, but in this case, and thanks to the relative simplicity of the set up, the formal quantification of the involved thermodynamic quantities has been initiated by Szilard itself, leading to the resolution of the paradox.

The idea that emerged from both experiments the Maxwell Demon and Szilard engine is that there is some cost associated to the measurement process. Brillouin was pioneer in attributing an entropy increase to the measurement [19, 20]. Although it turned out to be incorrect, it paved the way to formally bridging information theory and thermodynamics. 1961, Landauer stated that it is not the measurement itself that increases the entropy of the system, but the erasure at the end of the cyclic operation. This is known as the Landauer's erasure principle [21]. As a cornerstone of the Landauer's erasure principle is logical irreversibility. During the erasure, many different states converge to a single and same state after memory erasure, as presupposed by the idea of resetting the system to its initial configuration. This many-to-one process is associated to a logical irreversibility and thus involves dissipation. The entropy increase that follows the memory erasure must be dissipated in the environment.

In the reverse, the measurement process can be, in principle, performed in a reversible manner [22] and does not necessarily generate entropy dissipation. Bennett also brought significant contributions, in the same direction as the one initiated by Szilard [22]. In particular, he showed that the measurement expands the number of states of the memory, from one to many possible states, which is then compressed during the erasure process. Therefore, whereas the measurement was converted into an entropy decrease of the system by the operation of a Demon, the erasure generates entropy, which offsets the entropy reduction.

In the background of this description are the different steps that are proper to any information-processing cyclic operation: the measurement, storage, feedback and resetting. The correspondence between these four steps with the thought experiments of the Maxwell Demon and Szilard engine is explicitly represented in figure 1.1C.

- **Measurement:** Acquire knowledge on the state of the system, for instance by detecting a particle (A2 and B2).
- **Storage:** The measure is temporarily stored to be converted later in the reduction of the entropy of the system or the extraction of energy from the system. This is done either by storing directly in a memory, whatever the form it has (A3), or by acting "mechanically" so that energy conversion can operate (B3).
- **Feedback:** This is maybe the main step, the one that materialises the previous ones of measurement and storage. This consists in the transfer of one particle through the door

## Chapter 1. Introduction

---

in a fixed direction (A4) or the volumic expansion of the gas (B4).

- **Resetting:** As every cyclic operation must, by definition, end in the same state it started, the resetting includes the erasure of memory and bringing back to its initial physical state. The Maxwell Demon loses trace of the previous measurement and closes the door (A5), whereas the load is disconnected and the piston reinserted in the middle of the cylinder (B5) in the Szilard engine.

The Maxwell Demon and Szilard engine experiments challenge the second law of thermodynamics that was strongly believed to be inviolable. During the last century and as illustrated, many different approaches have been used to explain and bring a physical resolution to this apparent paradox. Whereas important steps were made by Landauer and Bennett to establish a link between the notions of measure and erasure with thermodynamic principles, in the meantime Claude Shannon significantly contributed to advances in the development of information theory in a totally different context. In his work "The Mathematical Theory of Communication", he developed the carrying capacity of communication channels [23].

The central notion that was missing so far is a formal quantification of information, that Shannon introduced as the Information entropy. A very first intuitive characterization of information is a quantification of "what is *a priori* unknown". Far from scientific considerations, if we are told that "the Earth is spherical", it is not very instructive and this claim does not contain a significant information. However, if we are told the result of tomorrow's lottery (and assuming it's correct), the amount of information in this sentence is huge. Similarly, both Maxwell Demon and Szilard engine are the manifestation of external agents able to provide information, *a priori* unknown, on the state of the system. In the first case the information consists in the detection of an approaching particle coming from the "correct" side, whereas in the second case, the information is simply the detection of the particle on one side of the box.

Together with the work by Landauer and Bennett (and many other contributions that we did not necessarily mention), the definition of information entropy by Shannon reconciliated both Maxwell Demon and Szilard engine thought experiments with the law of physics and in particular the second law of thermodynamics. The whole mathematical and physical formalism of information theory that is relevant within the framework of the current work is described in chapter 2.

Although obviously not sufficient, it was necessary to validate the nascent formalism of information theory via experimental realisations, possibly in biological systems but not only [24]. Many authors attempted to implement, by themselves, realisation of tiny demoniac agents at the atomistic level. We will go through one example that perfectly shows how the processing of information can be experimentally realised.

In their work, Toyabe and co-workers implemented a staircase like potential in which a dimeric

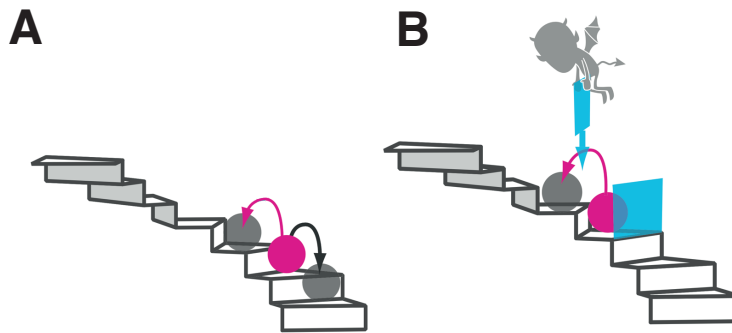


Figure 1.2: Illustration of the experimental setup. **(A)** In its natural evolution, the particle can go in both upward (pink arrow) and downward (grey arrow) direction, the latter occurring with a higher probability. **(B)** Under an external measurement-feedback setup, a "wall" (blue rectangle) is inserted after each upward transition, preventing the particle to go backward. Figure adapted from [25].

particle is embedded [25]. Initially, and without any external intervention, the particle will tend to "fall down" the staircase-like potential (Fig 1.2A). The experiment consists in adding an information-processing device that measures when, due to thermal fluctuations, the particle moves in the wanted (but less frequent) direction corresponding to an upstairs displacement. When such an event takes place, a potential barrier is inserted to prevent the particle to go through the reverse and unwanted transition (Fig 1.2B).

Beyond the set-up itself that largely differs from the Szilard engine, its thermodynamic implications are highly similar. The experimental setup consists in a feedback-based manipulation of a Brownian particle on the basis of information about the location of the particle. These operations result in a conversion of information into energy, precisely like in the Szilard engine experiment. However, although the experimental results obtained in this article confirm the prediction from the theory, there is still one huge discrepancy between such a set-up and biological systems in which life is the manifestation of molecular processing of information. Indeed, in the previously described experiment, the measurement device is external to the system and the resulting information is sent back to the system for a molecular feedback. A huge challenge for experimentalists is to conceive a setup in which all the information processing steps are taking place at the molecular level, exactly like in an information-processing system in the cell. To our knowledge, it has not been successfully achieved yet.

### 1.3 A first insight in cellular transport

The range of biological processes in living organisms is impossible to quantify, many of them being, on their own way a manifestation of the processing of information at the molecular level. The corresponding functions and instructions are encoded into the genome which likely constitutes the most information-loaded biological structure. All the mechanisms that involve substrates recognition and discrimination, transport of any type of particles and molecules, all the ribosome construction machinery are possible illustrations of biological systems in which information processing is at work. Even more specifically, many of them can be reconsidered, beyond the proper biological description of the function, as various manifestations of Maxwell Demons [26].

To set the framework of this thesis, we decided to restrict our study to some cellular transport processes. This choice is motivated on the one hand by the huge diversity of transport mechanisms and the essential role that is played, biologically speaking, in many processes that are absolutely essential for the proper functioning of living organisms. A second argument, possibly even more decisive, is that we considered it was the most natural way to transpose the Maxwell Demon into a biological environment. The transition from a Demon imposing to particles a directionality to cross a door is highly reminiscent of many transporters located in the cellular or organellar membranes.

The proper functioning of living organisms is based, among others, on the essential task of substrates transport through biological membranes, such as the plasma membrane or organelle membranes within the cell (intracellular transport). The range of transportable substrates is very broad, from ions to large macromolecules. Many reasons may highlight the importance of the transport mechanisms and why they are so essential in living organisms. One of these is that, in general the production site of a molecular complex is not spatially close to its destination, where it is intended to operate, to be used.

Fluxes of molecules within the cell are omnipresent and must be very precisely orchestrated: transport of all kind of substrates is performed considering the interplay of various parameters, such as the required transport rate, or a specifically targeted affinity with the different substrate molecules. These parameters that drive the transport to perform such a broad range of tasks suggest also a large diversity of transport processes: biological systems evolved to exhibit diverse mechanisms in order to fulfil the different constrains that are imposed (substrate selectivity, transport rate, directionality).

A proper understanding of these mechanisms is crucial for two main reasons. First, a large number of diseases find their causes in the dysfunction of some transport process at the molecular level [27, 28, 29]: an improved and more detailed understanding might then open great perspectives for treatments or regulations of severe diseases such as cystic fibrosis [30]. Second, in the last few decades, research in drug design showed amazing progresses.

### 1.3 A first insight in cellular transport

---

This was possible only thanks to greater and greater experimental achievements, as well as a simultaneous improved understanding of the different transport mechanisms that are at play: the way the pathogens reach their target and also the way the drug molecules can be transported to their destinations, to prevent or inhibit the manifestation of diseases.

Cellular transport mechanisms are divided into two main classes: passive and active transport [31]. The major distinction between both categories is the requirement or not of an energy source to perform transport.

Passive transport is the simplest of the transport classes, although obviously not less essential. Every type of passive transport is performed down a gradient, either a concentration gradient or an electrochemical gradient. The diffusion of molecules across the membrane either takes place directly through the lipid bilayer constituting the membrane (simple diffusion) or through a transport channel or carrier that aims to facilitate the diffusion from the high to the low concentration region (facilitated diffusion). In the case of passive transport process, the gradient generates the driving force of the transport and no additional energy source is required: the net flow of molecules is always down their gradient.

Contrarily to the passive mechanism, active transport takes place against a concentration or electrochemical gradient (from low to high values). It requires a coupling to an energy source to work against the natural evolution of the system. This coupling can take different forms, one of them being the interaction with an associated primary transport channel, thus indirectly benefiting from the electrochemical gradient of a second solute [32]. Such a transport mechanism is known as a coupled carrier transport, a member of the secondary active transport category. As a famous example, the transport of glucose across the plasmic membrane is coupled to the transport of sodium ions, driven by its electrochemical gradient at the membrane [33].

In the framework of the current work, the discussion will be mainly restricted to primary active transport, in which the energy source is ATP, the energy *currency* driving a huge variety of chemical reactions in the cell.

In particular, we will focus on the description of two active transporters, whose understanding of underlying mechanisms and thermodynamics will be our guideline: transport of substrates by ABC transporters (Chapter 3) and entropic pulling for the translocation of proteins through mitochondrial inner membrane and endoplasmic reticulum (Chapter 4). The two transport mechanisms differ greatly from each other, either shuttling substrates against a concentration gradient or exerting a mechanical force on an incoming substrate protein. However, we show that despite these flagrant differences, striking similar features emerge in the description of both active transporters. Each system displays specific requirements, for instance in the form of kinetic rates or biochemical conditions, necessary to impose a directionality and bring the system out of its equilibrium state, mirroring the information processing nature of the transporters.





## 2 Basic principles on stochastic thermodynamics and kinetic models

One challenge in the description and study of biological systems is to find an appropriate way to represent them. There is obviously no unique good choice and the answer to this tricky issue depends on many parameters: which level of specificity do we want to include in the model? What are the figure of merits that will serve the research objective? At which scale are the mechanisms that aim to be described? Throughout this research, we are guided by the motivation of modelling the phenomenology of different biological systems while reducing as much as possible the complexity of the model. The ensuing simplified vision of a biological system is almost unavoidably achieved at the expense of biochemical accuracy, but the trade-off is justified to preserve the minimal set of elements that reproduce the phenomenology of the biological process of interest. In that respect, kinetic models (also named kinetic schemes) which are introduced in this chapter are appropriate and convenient representations of biological systems.

We first introduce the concept of stochastic thermodynamics, by showing which are the major advances (and why they were absolutely required) with respect to classical thermodynamics. This discussion is furthermore embedded within the framework of the description of biological systems, to stress why such a formalism is appropriate and widely used in this field.

Second, we focus the discussion around the time-evolution of a system modelled by a kinetic scheme. We introduce the equation that governs the temporal evolution of the system, known as the master equation, and how the steady state can be computed from the different transition rates. An alternative approach based on graph theory is presented, giving interesting insights into the understanding of steady-state solution of kinetic models.

Third, we present different notions of stochastic thermodynamics that are of interest within the framework of this research. In particular, we define the entropy of a system at different scales and show how strongly it is related to the equilibrium or non-equilibrium state of a system. Then, we present the formalism of information theory to show how the fundamental laws of thermodynamics can be rewritten, taking into account the notion of information. Important

results such as the second law of thermodynamics and the Crooks fluctuation theorem are extended to account for the information processing.

Last, we conclude with a simple example of membrane transport to illustrate more concretely some of the basic concepts.

### 2.1 From classical to stochastic thermodynamics

The formalism of classical thermodynamics provides the rigorous framework for the study and description of an open system with its surrounding. In particular, one central question is to understand how a system evolves from a given initial state and relaxes to a final equilibrium state, through the exchange of heat, work and matter with its environment.

Thermodynamics is based on two fundamental laws which are stated (or re-stated) as follows [34, 35]:

1. *First principle:* The first law expresses energy conservation in a closed system. More specifically, in a thermodynamic process, the increment in the internal energy  $\Delta E$  is equal to the difference between the applied work  $W$  and the dissipated heat in the surrounding medium  $Q$ .

$$\Delta E = W - Q \quad (2.1)$$

Equation 2.1 can sometimes be found with different signs, depending on the convention used to report on the directionality of the transfers of heat and work.

2. *Second principle:* The second law reflects the propensity for a system to always tend to a highest disordered state. Considering a system in contact with a heat reservoir at fixed temperature, the total entropy  $S_{tot}$  is defined as the sum of internal entropy  $S$  and the one of the medium  $S_m$ . The variation over time of the total entropy being always positive, a system always moves to states of "higher disorder".

$$\Delta S_{tot} := \Delta S + \Delta S_m \geq 0 \quad (2.2)$$

The interpretation of the entropy is double. First, according to Clausius' work, the change in entropy is related to the heat exchanged  $Q$  with the surrounding environment at temperature  $T$ , that is  $Q = T\Delta S$  [36]. The alternative interpretation is a measure of the disorder of the system. This second approach will be widely used throughout this work.

Usually, a system is described by extensive variables which are macroscopic properties of the system: the energy  $E$ , the volume  $V$  and the number of particles  $N$ . Based on these three parameters, the state variable relevant to describe the state of such a system is the entropy  $S(E, V, N)$ . Correlated to the above second principle is the following definition of an

## 2.1 From classical to stochastic thermodynamics

---

equilibrium state of a system: among all the configurations compatible with the parameters of the system (here  $E, V$  and  $N$ ), the equilibrium state is the one maximising the entropy  $S(E, V, N)$ .

Historically, classical thermodynamics was formulated to report on the equilibrium state of a system. Therefore, the formalism is not appropriate to describe the physics when a system is moved far from its equilibrium state. It was necessary to adapt the level at which the description is made and to move to a less coarse-grained theory. Intermediary between a macroscopic description (the one of classical thermodynamics) and the finest description that can be done at the microscopic scale is the mesoscopic level, rightly the framework for the development of stochastic thermodynamics. The latter is the appropriate support for the physics of irreversible process with entropy dissipation, typical of non-equilibrium physics. To illustrate the importance of such non-equilibrium systems, Prigogine's seminal work states that it is possible for a non-equilibrium system to internally create order, in absence of any external force [37, 38, 39]. In that case, the displacement from equilibrium state is achieved by a chemical driving force. Here is the main difference with respect to classical thermodynamics, in which systems always relax to a state of highest entropy.

In its very first definition, and according to Boltzmann himself, the entropy is necessarily an ensemble property and the reduction of scale to a stochastic trajectory might a priori seem oxymoronic [40]. However, over years, the interest for such a definition applicable for instance to a single event taking place among a sampling of random events increased. At the origin of this question was also the will to extend both first and second laws of thermodynamics to mesoscopic non equilibrium systems such as colloidal particles, RNA, DNA and other biomolecules in interaction with an external experimental set-up (optical tweezers, AFM or micropipets) [41, 42, 43, 44, 45]. These biological systems acted as a framework for the development the stochastic thermodynamics, thanks to the major work of Sekimoto [46]. For an extended development of the whole theory, it is suggested to refer to some articles and reviews [35, 41, 47, 48, 49].

Three conditions have to be met to provide an appropriate basis for a stochastic description. First, and as already mentioned, external mechanical forces or unbalanced chemical potential are driving the non-equilibrium state; second, these systems are in a heat bath, at fixed temperature; last, thermal fluctuations play a significant role in the evolution of the system [35]. The third and last element is the root of the name *stochastic* thermodynamics, i.e. to be able to quantify the stochastic fluctuations of the system. At a macroscopic scale, the system is describe by ensemble and average quantities. The description is moved to a mesoscopic scale in which the definitions of the classical thermodynamics observables (work, heat, entropy) are extended to refer to a single stochastic trajectory [50, 51, 52, 53].

Many biological systems operate out of equilibrium and exchange energy, heat and matter with their environment. A typical example consists in all the processes that are driven by

an ATPase activity. The chemical energy of ATP is used to be converted, for instance, into mechanical work, and more generally to shift the system far from its equilibrium state. The quantification of thermal fluctuations in molecular motors and Brownian ratchet mechanisms is crucial, in particular to understand and quantify the induction of a directionality to a process subject to thermal fluctuations [39, 54, 55, 56].

Leaving aside the mathematical details of the derivation, the laws of thermodynamics (equations 2.1 and 2.2) have their exact counterpart in the formulation of stochastic energetics, in which thermodynamic quantities are defined with respect to single trajectories [35, 49]. The first and second principles are accordingly modified as follows:

$$dE = dw - dq \quad (2.3)$$

$$\langle \Delta s_{tot} \rangle := \langle \Delta s + \Delta s_m \rangle \geq 0 \quad (2.4)$$

where  $\langle \dots \rangle$  denotes the average over all possible trajectories the system can undergo.

These two expressions are obviously not new, since they are just a rewriting of equations 2.1 and 2.2 in which  $dE, dw, dq$  and entropy  $s$  are defined for a single trajectory between different states. In fact, the transition between both formulations is simply to consider much smaller systems than in the macroscopic description. But it is nevertheless assumed that these small parts of the system contain a sufficiently large number of particles, thus the underlying principles still hold [57].

## 2.2 Time evolution of the system

A very broad range of biological systems can be coarse grained and represented as kinetic models with a finite number of discrete states and a set of transitions defined by their kinetic rates. Although a very refined description of biological processes at the level of molecular structure might undoubtedly enrich the understanding of the mechanism, what remains clear is that a very broad range of biochemical aspects can be understood without such a refinement, restricting to a description based on empirical observations such as conformational states, allosteric and kinetic properties of proteins [58]. The model for such a description is called a kinetic model or kinetic scheme.

We consider a kinetic scheme with discrete states  $X_1, X_2, \dots, X_N$ . The system can jump from one state to another with a given probability per unit time, whose proxy is the kinetic rate  $k$  [ $s^{-1}$ ]. From the dynamics of the system results, at each time  $t$ , a probability for the system to be in each state given by  $P(X_1, t), P(X_2, t), \dots, P(X_N, t)$ .

A stochastic process is a generic term to account for all possible evolutions of a system which

evolves probabilistically, subject to random fluctuations. In that case, when considering the specific evolution along a stochastic trajectory, we can generally write

$$P(Y_N, t_N | Y_{N-1}, t_{N-1}; \dots; Y_1, t_1) = \frac{P(Y_N, t_N; Y_{N-1}, t_{N-1}; \dots; Y_1, t_1)}{P(Y_{N-1}, t_{N-1}; \dots; Y_1, t_1)} \quad (2.5)$$

in which time is ordered by convenience:  $t_N > t_{N-1} > \dots > t_2 > t_1$ .

Note that  $\{X_i\}$  refers to the different states of the system, whereas  $\{Y_i\}$  refers to any possible trajectory associated to a time evolution within the set of states  $\{X_i\}$ : each  $Y_i$  can be any of the  $X_i$  states.

A widely studied class of systems are the one following a Markov process, i.e. the system has no memory of past events. The state at time  $t$  is only defined by its state short before at time  $t - dt$ , independently of the whole history from time  $t_0$  to  $t - dt$ . Under this assumption, equation 2.5 gets:

$$P(Y_N, t_N | Y_{N-1}, t_{N-1}; \dots; Y_1, t_1) = P(Y_N, t_N | Y_{N-1}, t_{N-1}) \frac{P(Y_N, t_N; Y_{N-1}, t_{N-1})}{P(Y_{N-1}, t_{N-1})} \quad (2.6)$$

With such an assumption, the Chapman-Kolmogorov equation is written as follows, assuming a discrete set of states  $\{X_i\}$  although it is originally written in a continuous formulation:

$$P(Y_1, t_1 | Y_3, t_3) = \sum_{Y_2 \in \{X_i\}} P(Y_1, t_1 | Y_2, t_2) \cdot P(Y_2, t_2 | Y_3, t_3) \quad (2.7)$$

The formal derivation from the Chapman-Kolmogorov to the master equation is slightly tedious and beyond the scope and aim of this work. Analytical rigorous developments and interpretations can be found in the book by C. Gardiner [59]. In what follows, we will roughly sketch a development, leaving aside all the terms interpreted as drift and diffusion in the stochastic processes and only keeping trace of the discontinuous jump transitions between states.

The transition rate between two states  $Y_i$  and  $Y_j$  can be expressed as the limit of a conditional probability for state occupancy:

$$\lim_{t_j \rightarrow t_i} \frac{1}{t_j - t_i} P(Y_j, t_j | Y_i, t_i) = W(Y_j | Y_i, t_i) \quad \text{for } t_j > t_i \quad (2.8)$$

By differentiating the Chapman-Kolmogorov equation, leading to the so-called differential Chapman-Kolmogorov equation and getting rid of the drift and diffusion terms, the master equation is obtained (here in the discrete state space formulation):

$$\frac{d}{dt} P(Y_2, t_2 | Y_1, t_1) = \sum_{Z \in \{X_i\}} W(Y_2 | Z, t_2) P(Z, t_2 | Y_1, t_1) - W(Z | Y_2, t_2) P(Y_2, t_2 | Y_1, t_1) \quad (2.9)$$

The next and almost last step is to sum over all possible states  $Y_1$  at time  $t_1$ , by apply-

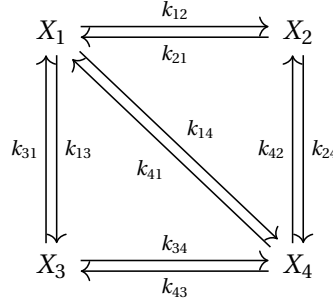


Figure 2.1: Kinetic scheme with 4 different states  $X_1, X_2, X_3$  and  $X_4$ . The allowed transitions in this example (resulting from an arbitrary choice) are shown with arrows between states. The rate corresponding to a transition from state  $X_i$  to  $X_j$  is denoted by  $k_{ij}$ .

ing  $\sum_{Y_i \in \{X_i\}} P(Y_i, t_1) \cdot (\dots)$  on both sides of equation 2.9. Associating the matrix elements  $W(Y_j|Y_i, t)$  to the rates  $k_{ij}$  [ $s^{-1}$ ], the final form of the master equation is given by:

$$\frac{d}{dt}P(Y_j) = \sum_{Y_i \in \{X_i\}} W(Y_j|Y_i, t)P(Y_i, t) - W(Y_i|Y_j, t)P(Y_j, t) \quad (2.10)$$

$$\frac{d}{dt}P(Y_j) = \sum_{Y_i \in \{X_i\}} k_{ij}(t)P(Y_i, t) - k_{ji}(t)P(Y_j, t) \quad (2.11)$$

Turning to a simple example, let  $X_1, X_2, X_3$  and  $X_4$  be the four possible states of a system. A possible set of transitions between the states is represented in Figure 2.1.

The time evolution of the state occupancy probability  $P(\{X_i\})$  is described by the above established master equation (equation 2.11), which holds for each state  $X_i$  of the system,  $i = 1, \dots, 4$  in the example of Figure 2.1. For the sake of brevity, we define the notation  $P_i := P(X_i)$ .

$$\frac{d}{dt}P_i(t) = \underbrace{\sum_j -k_{ij}(t) \cdot P_j(t)}_{\text{out flux}} + \underbrace{\sum_j k_{ji} \cdot P_j(t)}_{\text{in flux}} \quad (2.12)$$

Implicitly, if there is no arrow between two states (for instance between  $X_2$  and  $X_3$ ), then  $k_{2,3} = k_{3,2} = 0$ .

There are many different situations in which the transitions between states of a biological system might be time-dependent and we will mention two of the most common. First, the transition between two states  $X_i$  and  $X_j$  involves another external species, let's say  $Y$  that binds to  $X_i$  to form  $X_j \equiv X_i \cdot Y$ . Then the transition rate  $k_{ij} = k_{ij}^* \cdot [Y](t)$ , where  $k_{ij}^*$  is the binding rate [ $M^{-1}s^{-1}$ ]. However, in this case, the time dependence of the rates can usually be

## 2.2 Time evolution of the system

removed (keeping only time-independent rates such as  $k_{ij}^*$ ) by adding all the possible species that bind and form a complex, thus increasing the number of states to consider. This results also in a non linear formulation of equation 2.12 that involves terms with products of states probabilities and species concentrations. Second, the system might also be governed by an external time-dependent stimulus, for instance to observe its temporal response. The latter case is not the object of the present work.

As subsequently discussed in each specific applications of the master equations, we will restrict our theoretical description to time-independent transition rates. In that case, the system of master equations associated to the study model in figure 2.1 can be linearized and written in the form of a linear system of differential equations  $d\vec{P}(t)/dt = \underline{\underline{M}} \cdot \vec{P}(t)$ :

$$\frac{d}{dt} \begin{pmatrix} P_1(t) \\ P_2(t) \\ P_3(t) \\ P_4(t) \end{pmatrix} = \underbrace{\begin{pmatrix} -(k_{12} + k_{13} + k_{14}) & k_{21} & k_{31} & k_{41} \\ k_{12} & -(k_{21} + k_{24}) & 0 & k_{42} \\ k_{13} & 0 & -(k_{31} + k_{34}) & k_{43} \\ k_{14} & k_{24} & k_{34} & -(k_{41} + k_{42} + k_{43}) \end{pmatrix}}_{\underline{\underline{M}}} \cdot \underbrace{\begin{pmatrix} P_1(t) \\ P_2(t) \\ P_3(t) \\ P_4(t) \end{pmatrix}}_{\vec{P}(t)} \quad (2.13)$$

The steady state of the system corresponds to the case in which all the time derivative cancel out, that is  $\frac{d}{dt} P_i(t) = 0 \forall i$ . Considering an isolated system in which the total number of particles  $\mathcal{N}_{tot}$  is conserved (here  $\sum_i \mathcal{N}(X_i)(t) = \mathcal{N}_{tot}$ , independent of  $t$ ), the equations 2.13 (corresponding to the system depicted in Fig. 2.1) are linearly dependent and the determinant of the matrix is zero. Both properties  $\sum_i M_{i,j} = 0$  and  $M_{i,j} > 0, i \neq j$  have important mathematical consequences. One of them is that there is one and only one eigenvector with a vanishing eigenvalue (with the additional assumption that the graph of the kinetic scheme is strongly connected), which has to be normalised to get a probability vector ( $\sum_i P_i(t) = 1$ ). An alternative but equivalent way is to replace one line of the matrix, let's say the last one, by a normalisation constraint. The following system of equations is solved at steady state (s-s), to obtain the steady-state probabilities  $P_i^{(s-s)}$ :

$$\begin{pmatrix} -(k_{12} + k_{13} + k_{14}) & k_{21} & k_{31} & k_{41} \\ k_{12} & -(k_{21} + k_{24}) & 0 & k_{42} \\ k_{13} & 0 & -(k_{31} + k_{34}) & k_{43} \\ 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} P_1^{(s-s)} \\ P_2^{(s-s)} \\ P_3^{(s-s)} \\ P_4^{(s-s)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad (2.14)$$

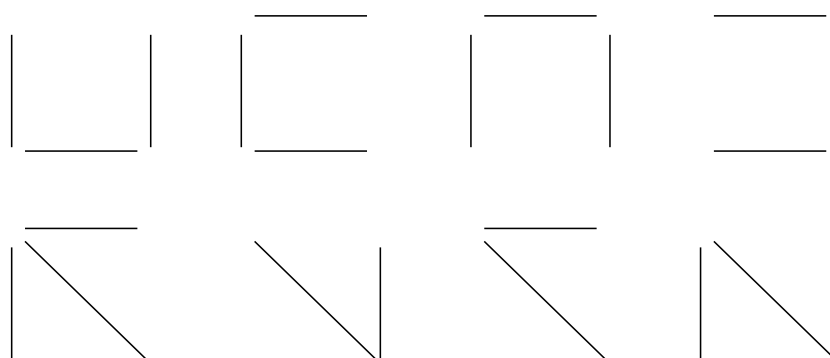


Figure 2.2: Set  $\chi$  of all the undirected spanning trees referring to the kinetic scheme in figure 2.2

### 2.3 State probability and net flux from kinetic scheme

An elegant approach was derived to compute the steady-state probability associated to a kinetic scheme without solving any linear algebra operations such as computing the nullspace or the inverse matrix. This method, named the "diagram method", is used to compute fluxes and state probabilities in terms of the diagram rate constants [60, 75]. The proof is not given in this section, but we will illustrate the mathematical result through the example of the toy kinetic scheme in Figure 2.1, which can be easily extended to any larger kinetic scheme, with the assumption it is strongly connected and without any unidirectional transitions.

In graph theory, a tree is a connected and undirected graph in which any two vertices of the tree are connected by exactly one path. As a corollary, a tree does not contain any loop. A specific type of tree is called *spanning tree*: it is a tree that necessarily contains all the vertices of the graph.

Coming back to the example in Figure 2.1, an exhaustive set of all spanning trees of the system has to be built. In the example, there is a total of 8 (undirected) spanning trees, listed in Figure 2.2. In more complex situations, the total number of spanning trees can be computed using the Kirchhoff's matrix tree theorem [61] and different algorithms to build all of them are reviewed in [62]. As a side remark, the method remains unchanged in case of multiple edges between two vertices. In that case, each spanning tree containing such an edge is duplicated (or eventually more) to contain one of these edges. Of course no spanning tree can contain two of them since they constitute a cycle.

Let  $\chi$  be the set of undirected spanning trees represented in Figure 2.2. To each state  $X_i$ ,  $i = 1, \dots, 4$  is associated the set  $\chi_i$  in which all the edges of the spanning trees are oriented towards  $X_i$  (the example of  $\chi_2$  is given in Figure 2.3).



## 2.3 State probability and net flux from kinetic scheme

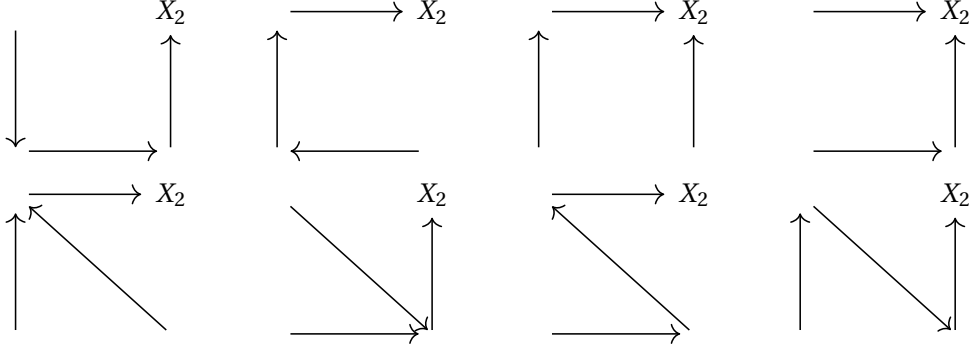


Figure 2.3: Set  $\chi_2$  of all the directed spanning trees with every edge directed toward the vertex  $X_2$

The steady-state probability of being in state  $i$  is given by:

$$P_i = \frac{\sum_{\vec{T}_i \in \chi_i} W_{\vec{T}_i}}{\sum_{j=1}^4 \sum_{\vec{T}_j \in \chi_j} W_{\vec{T}_j}} \quad (2.15)$$

where  $W_{\vec{T}_i}$  is the weight associated to oriented tree  $\vec{T}_i$ , that is

$$W_{\vec{T}_i} = \prod_{\text{edge } \vec{e} \in \vec{T}_i} k_{\vec{e}} \quad (2.16)$$

In other words, the state probability  $P_i$  is proportional (up to a normalization factor) to the sum of the weight of all the trees oriented towards the considered state  $X_i$ .

The probability of observing a given state is important in the description of a biological system, in particular to move from a mesoscopic description to an ensemble description weighted-averaged over all possible states. Another cognate quantity is maybe even more important, especially within the context of cellular transport: the net flux between two states. The quantification of transport in a kinetic scheme will very often, if not always, be related to the net flux between two states.

With the generic expression for the state probability  $P_i$ , we are interested in computing the flux between two states  $i$  and  $j$ , that is:

$$\Phi_{ij} = P_i \cdot k_{ij} - P_j \cdot k_{ji} \quad (2.17)$$

In the set  $\chi_i$ , we can distinguish two subsets of spanning trees, the one that contain the transition  $j \rightarrow i$  (denoted by  $\chi_i^{(+j)}$ ) and the one that do not contain it (denoted by  $\chi_i^{(-j)}$ ). If

## Chapter 2. Basic principles on stochastic thermodynamics and kinetic models

we look more closely at the contribution of one spanning tree of each subset in equation 2.17 to compute  $\Phi_{12}$ , we notice that only one of them contribute in  $P_1 \cdot k_{12} - P_2 \cdot k_{21}$ , as shown in equations 2.18 and 2.19.

$$\text{Example for } \chi^{(+12)} : \begin{array}{c} \begin{array}{ccc} & \xrightarrow{\text{red}} & \\ \uparrow & & \uparrow \\ & \xleftarrow{\text{red}} & \\ & \xleftarrow{\text{black}} & \end{array} & - & \begin{array}{ccc} & \xrightarrow{\text{red}} & \\ \uparrow & & \uparrow \\ & \xleftarrow{\text{red}} & \\ & \xleftarrow{\text{black}} & \end{array} & = 0 \end{array} \quad (2.18)$$

$$\text{Example for } \chi^{(-12)} : \begin{array}{c} \begin{array}{ccc} & \xrightarrow{\text{red}} & \\ \uparrow & \searrow & \downarrow \\ & \xleftarrow{\text{red}} & \end{array} & - & \begin{array}{ccc} & \xleftarrow{\text{red}} & \\ \uparrow & \searrow & \uparrow \\ & \xleftarrow{\text{red}} & \end{array} & \neq 0 \end{array} \quad (2.19)$$

where the red arrows represent the rates  $k_{ij}$  and  $k_{ji}$  in equation 2.17.

It results that only the spanning trees that do not contain the transition  $1 \leftrightarrow 2$  will contribute to the expression for the net flux. The weight of each spanning tree (that enters in the definition of the state probability) has to be multiplied by the transition rate (to compute a flux), and it necessarily closes a cycle, for each considered spanning tree (e.g. red arrows in equation 2.19). Thus, the summation reduces to all the cycles in the graph that contain the edge  $1 \leftrightarrow 2$ . In this simple case, the set of cycles  $C_{12}$  consists only in the two following cycles:

$$C_{12} = \left\{ \begin{array}{ccc} X_1 & \text{---} & X_2 \\ | & & | \\ X_3 & \text{---} & X_4 \end{array} ; \begin{array}{ccc} X_1 & \text{---} & X_2 \\ & \searrow & | \\ & & X_3 & \text{---} & X_4 \end{array} \right\} \quad (2.20)$$

In both terms of equation 2.19, the cycle is crossed in two opposite directions, whereas all the other edges of the spanning tree "converge" to the cycle with a same orientation.

At this point, we need to insert an additional notation that will be denoted by  $w_c$  and which is defined for each  $c \in C_{12}$  and related to the set of arrows flowing into the cycle. If there are  $n_c$  vertices that are not belonging to the cycle  $c$ , the set of arrows flowing into the cycles precisely contains  $n_c$  elements, not necessarily connected.  $w_c$  is sum over all possible configurations of the product of the  $n_c$  directed transitions towards the cycle. If the cycles contains all the vertices of the graph ( $n = 0$ ), then  $w_c = 1$ . Applied to the set  $C_{12}$  defined in 2.20 (with the same order of the list), it follows that

$$w_c|_{c \in C_{12}} = \{1 ; k_{31} + k_{34}\} \quad (2.21)$$

By combining equations 2.19 and 2.21 into eq. 2.17, the latter can be rewritten as follows :

$$\Phi_{12} = \sum_{c \in C_{12}} (W_c^\circ - W_c^\circ) \cdot w_c = \sum_{c \in C_{12}} W_c^\circ \cdot w_c \cdot \left(1 - \frac{W_c^\circ}{W_c^\circ}\right) \quad (2.22)$$

The ratio  $W_c^{\circ}/W_c^{\circ}$  is thermodynamically significant and its interpretation is related to the entropy dissipation along the cycle, as later discussed in section 2.6.

### 2.4 Detailed balance and microscopic reversibility

The extension of classical thermodynamics, in which all the processes are reversible, to a stochastic and non-equilibrium thermodynamics has one important consequence: irreversible phenomena are associated to entropy dissipation. The condition for entropy dissipation cannot be understood without considering microscopic reversibility and detailed balance conditions, which are thermodynamic concepts playing an important role in biological systems and their study through kinetic schemes.

The definition of microscopic reversibility states that *"in a reversible reaction, the mechanism in one direction is exactly the reverse of the mechanism in the other direction"* [63, 64]. In the general framework of molecular machines, the microscopic reversibility imposes that forward and backward motions in a directional process are exactly cancelled. The intrinsic asymmetry is then generated by the "introduction" of energy barriers that actively prevent the backward transition, for instance through a dissociation constant of the considered substrate that depends on the state of the system. [65, 66].

A consequence is the more general principle of detailed balance, which emerged very early in the development of Thermodynamics, starting with a mention by Boltzmann in his work at the end of the 20th century [67]. The formal definition of the principle of detailed balance given by the International Union of Pure and Applied Chemists (IUPAC) is the following: *"When equilibrium is reached in a reaction system (containing an arbitrary number of components and reaction paths), as many atoms, in their respective molecular entities will pass forward, as well as backwards, along each individual path in a given finite time interval. Accordingly, the reaction path in the reverse direction must in every detail be the reverse of the reaction path in the forward direction (provided always that the system is at equilibrium)"* [68].

To illustrate more concretely the detailed balance condition, let us consider  $X_i$  and  $X_j$  as two states of a system, with respective probabilities  $P(X_i)$ ,  $P(X_j)$  and transition rates  $k_{ij}$ ,  $k_{ji}$ . The detailed balance condition is:

$$P(X_i)^{(eq)} k_{ij}^{(eq)} = P(X_j)^{(eq)} k_{ji}^{(eq)} \quad (2.23)$$

By extension, the detailed balance is often written along a cycle of transitions  $X_1 \leftrightarrow X_2 \leftrightarrow \dots \leftrightarrow X_N \leftrightarrow X_1$ . Applying to every transition along the cycle, equation 2.23 can be rewritten

independently of the state probabilities:

$$\left. \frac{k_{1,2} \cdot k_{2,3} \cdot \dots \cdot k_{N-1,N} \cdot k_{N,1}}{k_{1,N} \cdot k_{N,N-1} \cdot \dots \cdot k_{3,2} \cdot k_{2,1}} \right|^{(eq)} = 1 \quad (2.24)$$

When the system is at equilibrium, detailed balance holds by definition. The ensuing constraints, derived at equilibrium, must be satisfied out of equilibrium. Importantly, it is also true that any configuration of the system for which detailed balance is satisfied necessarily coincides with the equilibrium distribution.

## 2.5 Information entropy

To formally define the notion of information entropy, let us start by considering a set of  $N$  states  $X_1, X_2, \dots, X_N$ . It is assumed that every possible transition has its reverse in the kinetic scheme. Each process and its reverse occur at the same flux at equilibrium, as imposed by detailed balance constraints. The discussion of some specific cases with unidirectional transitions is not part of the current introduction [69, 70, 71].

Historically, Shannon derived the statistical entropy within the framework of the theory of communication to quantify the quality of the transmission of a message from one point to another [23]. Whereas the conceptual idea of the current work is quite different, all the formalism remains valid and is introduced directly within the context of a kinetic model. The definition of the statistical entropy for a set of states  $\{X_i\}$  with probabilities  $P(X_i, t)$ ,  $i = 1, \dots, N$  at time  $t$  is given by equation 2.25:

$$S(\{X_i\}, t) = -k_B \sum_{i=1}^N P(X_i, t) \cdot \ln(P(X_i, t)) \quad (2.25)$$

The statistical entropy is also known as information entropy. The rationale for this alternative name is that the entropy  $S(\{X_i\})$  quantifies how much is known on the state of the system, or more rigorously, how much do you gain in the knowledge of the system when measuring it. On the one hand, if the system is fully determined, that is  $P(X_k) = 1$  and  $P(\{X_{i \neq k}\}) = 0$ , then  $S(\{X_i\}) = 0$ , because the state of the system is known without any doubt (in state  $X_k$ ). On the other hand, if all states are equally distributed, that is  $P(X_i) = \frac{1}{N}$ ,  $\forall i = 1, \dots, N$ , then  $S(\{X_i\}) = k_B \ln(N)$ . Furthermore, it can be shown that any intermediary probability distribution between the purely random distribution and the fully deterministic case will give a value  $0 < S(\{X_i\}) < k_B \ln(N)$  [72].

## 2.6 Entropy production along a stochastic trajectory

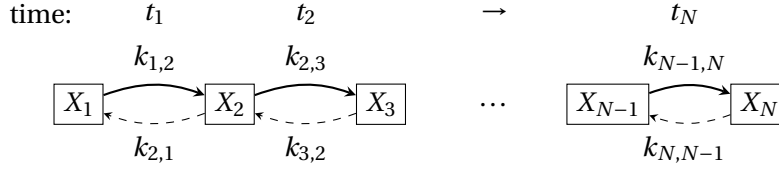


Figure 2.4: Illustration of a stochastic trajectory in a kinetic scheme, from state  $X_1$  to state  $X_N$ . The thick black arrows depicts the forward direction of the trajectory, with corresponding jumps at time  $t_1 < t_2 < \dots < t_N$ . The reverse trajectory is shown with the dashed arrows.  $k_{ij}$  denotes the transition rate from state  $i$  to  $j$ .

## 2.6 Entropy production along a stochastic trajectory

The scope of this paragraph is to present the formalism of entropy production along a single stochastic trajectory, as developed by Seifert in 2005 [41]. This approach is particularly relevant within the framework of our work in which different biological systems are studied as kinetic models with a fixed number of discrete states.

Without losing generality, we consider a stochastic trajectory that goes from state  $X_1$  to state  $X_N$  (assuming the connectivity of the scheme, it is always possible upon an appropriate indexing of the states), as depicted by the thick plain arrows in figure 2.4. Each transition from state  $i$  to state  $i + 1$  takes place at time  $t_i$ , with  $t_1 < t_2 < \dots < t_N$ .

The stochastic entropy of a system defined for a set of states ( $\{X_i\}$ ) suggests a possible expression to quantify the entropy of a single trajectory. Indeed, in an ensemble perspective, at each time  $t$  the precise state of the system is not known but its statistical entropy is defined as a weighted sum of  $\ln(P(X_i))$  (equation 2.25, in which the time dependence is implicit). Restricting to a determined trajectory, along which the state  $X$  is precisely known at each time  $t$ , it is a "reasonable guess" to postulate that the trajectory entropy of the system is given by:

$$s(t) = -\ln[P(X(t))] \quad (2.26)$$

where  $X(t)$  is the state  $X$  of the system at time  $t$ :  $X(t) = X_i$ ,  $t_{i-1} < t \leq t_i$  in Figure 2.4.

The time derivative of the entropy of the system along the trajectory is given by

$$\dot{s}(t) = -\frac{\dot{P}(X)}{P(X)} - \sum_i \delta(t - t_i) \ln\left(\frac{P(X_{i+1})}{P(X_i)}\right) \quad (2.27)$$

where the time dependence  $X = X(t)$  is implicit.

The first term is the consequence of internal variations in the state probability either due to transitions between states or externally-driven time-dependent parameters, whereas the second term ensues from stochastic jumps at fixed discrete times  $t_i$  along the trajectory.

Justified by *a posteriori* arguments in Seifert's article [41], the variation of the total entropy results from two contributions: the change in the entropy of the system itself and the entropy that is dissipated into the environment (medium):

$$\dot{s}_{tot}(t) = \underbrace{\dot{s}(t)}_{\text{system}} + \underbrace{\dot{s}_m(t)}_{\text{medium}} \quad (2.28)$$

where

$$\dot{s}_{tot}(t) = -\frac{\dot{P}(X)}{P(X)} - \sum_i \delta(t-t_i) \ln \left( \frac{P(X_{i+1}) \cdot k_{i+1,j}}{P(X_i) \cdot k_{i,i+1}} \right) \quad (2.29)$$

$$\dot{s}(t) = -\frac{\dot{P}(X)}{P(X)} - \sum_i \delta(t-t_i) \ln \left( \frac{P(X_{i+1})}{P(X_i)} \right) \quad (2.30)$$

$$\dot{s}_m(t) = -\sum_i \delta(t-t_i) \ln \left( \frac{k_{i+1,i}}{k_{i,i+1}} \right) \quad (2.31)$$

Actually, the notion of entropy dissipation into the medium  $s_m$  is somehow unclear and not precisely defined. However, there are some preliminary and intuitive arguments to justify such a proposition. One of them is that the ratio of transition rates  $k_{ij}/k_{ji}$  is related to the difference of free energy between states  $X_i$  and  $X_j$ , that is the energy consumed during the transition  $X_i \rightarrow X_j$  [73, 74].

Let us consider a stochastic trajectory from state  $X_0$  at time  $t_0$  to state  $X_f$  at time  $t_f$  and as previously with a set of jumps  $X_i \rightarrow X_{i+1}$  at time  $t_i$ . The integration over time of the entropy dissipation rate along such a trajectory gives:

$$\Delta s_{tot} = -\ln \left( \frac{P(X_f, t_f)}{P(X_0, t_0)} \right) - \sum_i \ln \left( \frac{k_{i+1,i}(t_i)}{k_{i,i+1}(t_i)} \right) \quad (2.32)$$

$$\Delta s = -\ln \left( \frac{P(X_f, t_f)}{P(X_0, t_0)} \right) \quad (2.33)$$

$$\Delta s_m = -\sum_i \ln \left( \frac{k_{i+1,i}(t_i)}{k_{i,i+1}(t_i)} \right) \quad (2.34)$$

The interpretation of these equations takes on its full meaning when considering a cyclic trajectory. In that case,  $\Delta s = 0$ , and it remains  $\Delta s_{tot} = \Delta s_m$ . At equilibrium, from detailed balance,  $\Delta s_m$  is also equal to zero and thus there is no entropy dissipation at all. Out of equilibrium and over a cyclic trajectory, the whole entropy is dissipated into the environment. As a side remark, this entropy dissipation term is the one that appears in equation 2.22 denoted by  $W_c^\circ / W_c^\circ$ .

To further support the split between the different contributions in equations 2.29 to 2.31, it is interesting to move back from a single trajectory to an ensemble description of the system, by averaging each contribution over the possible trajectories. The averaged form of equations

2.32 to 2.34 is given respectively by equations 2.35 to 2.37:

$$\dot{S}_{tot}(t) = \langle \dot{s}_{tot}(t) \rangle = \sum_{i,j} P(X_i, t) \cdot k_{ij}(t) \cdot \ln \left( \frac{P(X_i, t) \cdot k_{ij}(t)}{P(X_j, t) \cdot k_{ji}(t)} \right) \quad (2.35)$$

$$\dot{S}(t) = \langle \dot{s}(t) \rangle = \sum_{i,j} P(X_i, t) \cdot k_{ij}(t) \cdot \ln \left( \frac{P(X_i, t)}{P(X_j, t)} \right) \quad (2.36)$$

$$\dot{S}_m(t) = \langle \dot{s}_m(t) \rangle = \sum_{i,j} P(X_i, t) \cdot k_{ij}(t) \cdot \ln \left( \frac{k_{ij}(t)}{k_{ji}(t)} \right) \quad (2.37)$$

This result is in accordance with anterior results by Schnakenberg [75] who derived expressions for the rate of entropy production. The present approach, using the formalism of single trajectory, is even more general since the result is not restricted to the description of steady state but its validity holds both at steady state and in transient regime of the time evolution of the system.

## 2.7 Measurement and information

To introduce the formalism of information theory from a thermodynamic perspective, one has to come back to the very first definition of the stochastic entropy given by equation 2.25, rewritten in a slightly different way to introduce a dimensionless equivalent formulation of the Shannon entropy:

$$S(\{X_i\}) = -k_B \sum_{i=1}^N P(X_i) \cdot \ln(P(X_i)) = k_B \cdot H(\{X_i\}) \quad (2.38)$$

where both  $S(\{X_i\})$  and  $H(\{X_i\})$  are known as the Shannon Entropy, simply differing by a unit convention. Note that as  $S(\{X_i\})$  depends on  $\{X_i\}$  only through the probability distribution  $P$  among the different states, it is often written as  $S(P)$ , a convention widely used in a thermodynamic description.

In an equilibrium configuration of the system satisfying the Boltzmann distribution, the well-known relation between entropy, energy and free energy is easily recovered.

$$\begin{aligned} P_{eq}(X_i) &= \frac{\exp(-\beta \mathcal{H}(X_i))}{\mathcal{Z}} \Rightarrow S_{eq}(\{X_i\}) = k_B \langle \beta \mathcal{H}(X_i) \rangle + k_B \ln(\mathcal{Z}) \\ &\Leftrightarrow TS_{eq}(\{X_i\}) = \langle \mathcal{H}(X_i) \rangle - F_{eq}(\{X_i\}) \\ &\Leftrightarrow TS_{eq}(\{X_i\}) = E(\{X_i\}) - F_{eq}(\{X_i\}) \end{aligned} \quad (2.39)$$

## Chapter 2. Basic principles on stochastic thermodynamics and kinetic models

---

where  $\mathcal{H}$  is the Hamiltonian of the system,  $\mathcal{Z}$  the partition function and  $E$  the energy. The non-equilibrium free energy of a system is defined analogously [76, 77, 78], that is :

$$F(P, \mathcal{H}) \equiv \langle \mathcal{H}(X_i) \rangle_P - T \cdot S(P) \quad (2.40)$$

On a set of states  $X = \{X_i\}$ , we define a measurement, that is a process  $M$  that might give different possible outcomes denoted by  $\{m\}$ . Given an outcome  $m \in M$  the probability distribution of  $X$  is modified, from an initial probability  $P(X_i)$  to a new conditional probability  $P(X_i|m)$  (in general  $P(X_i) \neq P(X_i|m)$ ) after measurement which suggests a correlation between the probability distribution of variable  $X$  and the set of possible outcomes  $M$ .

In its very general formulation, the mutual information between two random variables  $X$  and  $Y$ ,  $I(X; Y)$  is defined by the relative entropy between the joint distribution  $P(X, Y)$  and the product of respective distributions  $P(X) \cdot P(Y)$  [79]:

$$\begin{aligned} I(X; Y) &= \sum_{x \in X, y \in Y} P(x, y) \ln \frac{P(x, y)}{P(x)P(y)} = H(X) + H(Y) - H(X, Y) \\ &= H(X) - H(X|Y) \end{aligned} \quad (2.41)$$

Reformulated within the framework of a variable  $X$  and a measurement  $M$  on the state of the system, we have:

$$\begin{aligned} I(X; M) &= H(X) - H(X|M) \\ \Leftrightarrow \Delta S_{meas} &= k_B(H(X|M) - H(X)) = -k_B \cdot I(X; M) \end{aligned} \quad (2.42)$$

whose interpretation is instructive about the conceptual idea of information. The mutual information between a physical system  $X$  (here a set of states  $\{X_i\}$ ) and a measurement process  $M$  operated on it is equal to the difference between the Shannon entropy  $H(X)$  of the system and the Shannon entropy of the system given a measurement process,  $H(X|M)$ . Equivalently, the variation of entropy associated to the measurement is equal, in absolute value, to the mutual information  $I(X; M)$ . The positivity (by construction) of mutual information [80] between two variables  $X$  and  $M$  implies that the measurement always decreases the entropy of the system, by gaining some knowledge about the state of it.

The formalism covering the definition of non-equilibrium free energy, mutual information between a system and measurement, was introduced so far with the aim of getting to a very impactful equation, especially because it provides the resolution of the paradox risen by the thought experiments of Maxwell Demon and Szilard engine. Equation 2.43 is an extension of the second law of thermodynamics that includes the information and its derivation can be



## 2.8 Fluctuation theorem and time reversibility

---

found in the paper by Sagawa et al. [81].

$$W - \Delta F \geq -k_B T I(X; M) \quad (2.43)$$

In particular, in the case of a cyclic process in which the free energy difference  $\Delta F = 0$ , the equation reduces to

$$W \geq -k_B T I(X; M) \quad (2.44)$$

The equation states that the maximum amount of work that can be extracted ( $W < 0$ ) from a cyclic process on a system  $X$  is (in absolute value) at most equal to the information gained on the state of the system during measurement,  $I(X; M)$ .

This equation gives new insights to the Szilard engine. Let us consider a cyclic process consisting in inserting a wall in the middle of the box, moving the wall in the direction opposite to the position of the particle (i.e. perform a volumic expansion of the gas) and finally removing the wall to reset the system. In that case:

$$W = -k_B T \cdot \ln(2) \quad (2.45)$$

$$\left. \begin{array}{l} H(X) = \ln(2) \\ H(X|M) = 0 \end{array} \right\} \Rightarrow I(X; M) = H(X) - H(X|M) = \ln(2) \quad (2.46)$$

The Szilard engine is an example of an optimal setup, in which the whole amount of information issued from the measurement on the system is converted into an extracted work. The question of optimising such a work extraction is beyond the scope of the actual description. It has been shown that the optimal bound of equation 2.44, that is an equality between information and work, can be reached upon reversibility of the feedback process [82, 83]. Although we will not deal with the formal derivation of optimality conditions, tackling the question of the time reversal of stochastic trajectories associated to measurement gives new enlightening insights in the definition and application of information associated to a measurement.

## 2.8 Fluctuation theorem and time reversibility

A system is brought out of equilibrium from a macroscopic state  $A$  to a state  $B$  along a given thermodynamic path  $\gamma$ . The transition is driven by tuning external parameters of the system at a finite rate and the corresponding protocol is not necessarily assumed to be slow. When the experiment is repeated over an ensemble of such trajectories from state  $A$  to state  $B$ , the following inequality from macroscopic thermodynamics holds:

$$\langle W \rangle_\gamma \geq \Delta F := F_B - F_A \quad (2.47)$$

By sampling only one or a very few thermodynamic trajectories from  $A$  to  $B$ , it is clearly

concievable to (apparently) violate the second law of thermodynamics by extracting more work than the free energy difference between the starting and ending states: these occurrences are called the "transient violations of the second principle" [84]. However, in average, the inequality is always satisfied, in accordance with the law of thermodynamics. This results is made even stronger with an equality relation, linking an averaged expression of the work and the free energy difference. The so-called Jarzynski equality [85] is given by:

$$\langle \exp(-\beta W) \rangle_\gamma = \exp(-\beta \Delta F) \quad (2.48)$$

Such an average formulation eliminates the fluctuations that ensue from the stochastic nature of the process. These fluctuations are captured and quantified by the Crooks fluctuation theorem [47]:

$$\frac{P_F(\sigma)}{P_R(-\sigma)} = \exp(\sigma) \quad (2.49)$$

where index  $F$  denotes the forward protocol and index  $R$  refers to the reverse protocol upon time reversal: the operations that drive the system from  $B$  to  $A$  upon time reversal are denoted by  $\lambda_R(t) = \lambda_F(t_{end} - t)$ . During both protocols  $\lambda_F(t)$  and  $\lambda_R(t)$  is an entropy production ( $+\sigma$  and  $-\sigma$  respectively) since the system is driven out of equilibrium. The Crooks fluctuation theorem relates the probabilities  $P_F$  and  $P_R$  of observing a entropy production  $\pm\sigma$  during respectively the forward and backward trajectories. Simply said, the more entropy is produced along a trajectory, the more likely it is (with an exponential growth) to be observed rather than the reverse trajectory. The limit case  $\sigma = 0$  corresponds to an equilibrium process, with no entropy production, in which both forward and backward trajectories are equiprobable, analogously to the detailed balance constrains (equation 2.23).

To mirror the second principle of thermodynamics and the Jarzynski equality that involve the work  $W$  and free energy  $F$ , an alternative formulation of equation 2.49 holds, with the assumption that the initial state  $A$  of the system is at equilibrium:

$$\frac{P_F(\beta W)}{P_R(-\beta W)} = \exp(-\Delta F) \exp(+\beta W) \quad (2.50)$$

In the same way that the second law of thermodynamics was extended to include the information (equation 2.43), an analogous reasoning leads to a fluctuation theorem also involving the information associated to a measurement [86, 82, 87]:

$$\frac{P_F(\gamma_F, m)}{P_R(\gamma_R, m)} = \exp(\beta W(\gamma_F) - \Delta F + \mathcal{I}(\gamma_F, m)) \quad (2.51)$$

where  $\gamma_F$  is a stochastic trajectory associated to a fixed protocol driving from state  $A$  to state  $B$  and  $\gamma_R$  is the reversed trajectory upon time reversal.  $m \in M$  is the outcome of the measurement process and  $\mathcal{I}$  denotes the information at the trajectory level.

Applied to a trajectory among a discrete set of states  $\{X_i\}$ , a trajectory-dependent definition of information associated to a trajectory  $\gamma_F(t) = X_i(t)$  and a measurement outcome  $m \in M$  is given by [80] :

$$\begin{aligned} \mathcal{I}(\gamma_F, m) &= \ln \left( \frac{P_F(m|X_i(t))}{P(m)} \right) \\ &= \ln \left( \frac{P_F(X_i(t)|m)}{P(X_i(t))} \right) \end{aligned} \quad (2.52)$$

The average over all possible trajectories and measurement outcomes defined on set  $X$  and  $M$  brings back to the mutual information (equation 2.41), that is  $\langle \mathcal{I}(\gamma_F, m) \rangle_{\gamma_F \in \{X, M\}} = I(X, M)$

In general, there are many different formulations of the fluctuation theorems, depending on the assumptions on the system itself and with different ranges of validity [88]. In this introduction, we aim to give an overview of the main concept, without going too deep into the formalism. The work of a detailed description with the underlying aim of unifying the different formulations of the fluctuation theorem can be found in numerous reviews and articles [40, 53, 82, 84, 86, 89].

At this point, it is essential to stress out the following consideration: the above formalism has been derived for an external measurement device that is "observing" a system and performing a measurement on it. In that sense, the measurement device does not directly affect the system itself. However, in the different case studies that will be described in chapters 3 and 4, we will face an internal measurement device in which the measurement process unavoidably affects the state of the system. This modification is crucial in the interpretation that will be made from the measurement process itself.

Formally, equation 2.52 has to be adapted. So far, we were considering a system with states  $\{X_i\}$  and probabilities  $P(X_i)$ , with a measurement device externally operating on it and the trajectory was defined along a stochastic trajectory defined on the different states. Now, let us consider a system split into two subsystems, each of them with  $N$  states. The internal measurement process corresponds to a switch between each subsystem. In absence of measurement, the corresponding subsystem has the probabilities  $P(X_i)$ ,  $i = 1, \dots, N$ . Under the action of an internal measurement device  $M$ , the system is moved the alternative subsystem with a different probability distribution, given by  $P(X_i|m)$ ,  $i = 1, \dots, N$  where  $m$  is the measurement outcome provided by the internal measurement device. Along the stochastic trajectory (defined within the whole  $2N$ -state system, from one subsystem to the other) going from state  $X_i$  to state  $X_i|m$ , the corresponding information  $\mathcal{I}_{\text{measure}}$  inspired by equation 2.52 is thus defined as:

$$\mathcal{I}_{\text{measure}} = \ln \left( \frac{P_F(X_i|m)}{P(X_i)} \right) \quad (2.53)$$

There is no formal definition of such a definition associated to an internal measurement device, However the expression directly comes out from well-established results in the case of external measurement device and there is a strong correspondance between the two cases (internal or external measurement). Therefore, equation 2.53 will be used in the next chapter to quantify the trajectory-dependant information in the case of an internal measurement device.

## 2.9 ATP as energy source of primary active transporters

All active processes in living matter, and in particular active transport of molecules and ions require energy. The major energy source comes from the release of an inorganic phosphate  $P_i$  during the hydrolysis reaction from ATP to ADP, according to the following reaction :



which releases an amount of energy  $\Delta\tilde{G}$ . The reverse reaction that consumes an amount of energy  $\Delta\tilde{G}$  is the synthesis. The notation  $\Delta\tilde{G}$  corresponds to the negative free energy difference of the reaction. At the end,  $\Delta G = -\Delta\tilde{G}$  is defined as a positive value, just a matter of convention. The Gibbs free energy of the reaction is given by:

$$\Delta\tilde{G} = \Delta\tilde{G}_0 + k_B T \ln \frac{[ADP][P_i]}{[ATP]} \quad (2.55)$$

where  $\Delta\tilde{G}_0$  is the standard free energy defined by the equilibrium constant  $K_{eq}$  of the reaction:

$$\Delta\tilde{G}_0 = -k_B T \ln(K_{eq}) = -k_B T \ln \frac{[ADP]_{eq}[P_i]_{eq}}{[ATP]_{eq}} \quad (2.56)$$

Assuming that the inorganic phosphate  $P_i$  is in large excess in the hydrolysis reaction (and thus  $[P_i] \approx [P_i]_{eq}$ ), the Gibbs free energy  $\Delta\tilde{G}$  depends only on the concentrations of ATP and ADP. Introducing the notation

$$\alpha := \frac{[ATP]}{[ADP]} \text{ and } \alpha_{eq} := \frac{[ATP]_{eq}}{[ADP]_{eq}} \quad (2.57)$$

we finally obtain

$$\Delta\tilde{G} = -k_B T \ln \left( \frac{[ATP]}{[ADP]} \cdot \frac{[ADP]_{eq}}{[ATP]_{eq}} \right) = -k_B T \ln \left( \frac{\alpha}{\alpha_{eq}} \right) \quad (2.58)$$

In typical cellular conditions,  $[ATP] = 5 \cdot 10^{-3} M$ ,  $[ADP] = 0.5 \cdot 10^{-3} M$  and  $[P_i] = 10 \cdot 10^{-3} M$ , thus leading to a Gibbs free energy difference in vivo  $-25k_B T \lesssim \Delta\tilde{G} \lesssim -20k_B T$  [90].

With the convention of a positive free energy  $\Delta G$  as a quantification of the energy brought to

## 2.9 ATP as energy source of primary active transporters

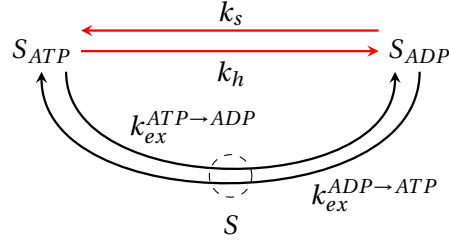


Figure 2.5: The transition between states  $S_{ATP}$  and  $S_{ADP}$  can be done via two routes: hydrolysis and synthesis (red arrows, rates  $k_h$  and  $k_s$  respectively) or exchange (black arrows, rates  $k_{ex}^{ATP \rightarrow ADP}$  and  $k_{ex}^{ADP \rightarrow ATP}$ ). Implicit in the exchange process is an intermediary apo state  $S$  (dashed circle) in which there is no nucleotide bound on the substrate.

the system in the form of  $ATP$ , we have:

$$\Delta G = k_B T \ln \left( \frac{[ATP]}{[ADP]} \cdot \frac{[ADP]_{eq}}{[ATP]_{eq}} \right) = k_B T \ln \left( \frac{\alpha}{\alpha_{eq}} \right) \quad (2.59)$$

Let us consider a substrate protein (denoted by  $S$ ) that can be either bound to  $ATP$  or  $ADP$ . The corresponding states are called the  $ATP$  state  $S_{ATP}$  (respectively  $ADP$  state  $S_{ADP}$ ) of the protein. The transition between  $ATP$  and  $ADP$  states is possible through two different paths (Figure 2.5), either the transitions of hydrolysis/synthesis or exchange. Whereas the first one is a chemical reaction of from  $ATP$  to  $ADP$  or vice versa, the second one consists in the successive release and binding of nucleotide with an transient apo state  $S$  of the substrate.

The expression for the exchange rates is given by equations 2.60 and 2.61

$$k_{ex}^{ATP \rightarrow ADP} = k_{-ATP} \frac{[ADP]k_{+ADP}}{[ADP]k_{+ADP} + [ATP]k_{+ATP}} = k_{-ATP} \frac{k_{+ADP}}{k_{+ADP} + \alpha k_{+ATP}} \quad (2.60)$$

$$k_{ex}^{ADP \rightarrow ATP} = k_{-ADP} \frac{[ATP]k_{+ATP}}{[ADP]k_{+ADP} + [ATP]k_{+ATP}} = k_{-ADP} \frac{\alpha k_{+ATP}}{k_{+ADP} + \alpha k_{+ATP}} \quad (2.61)$$

From the detailed balance condition (equation 2.24), the ratio between hydrolysis and synthesis rates is related to the dissociation constants of  $ATP$  and  $ADP$ :

$$\frac{k_h}{k_s} = \frac{k_{ex}^{ADP \rightarrow ATP}}{k_{ex}^{ATP \rightarrow ADP}} \Big|^{(eq)} = \frac{1}{\alpha_{eq}} \frac{K_{d,ATP}}{K_{d,ADP}} \quad (2.62)$$

where  $K_{d,ADP} = k_{off,ATP}/k_{on,ATP}$  and similarly for  $K_{d,ATP}$ . It follows that the ratio between

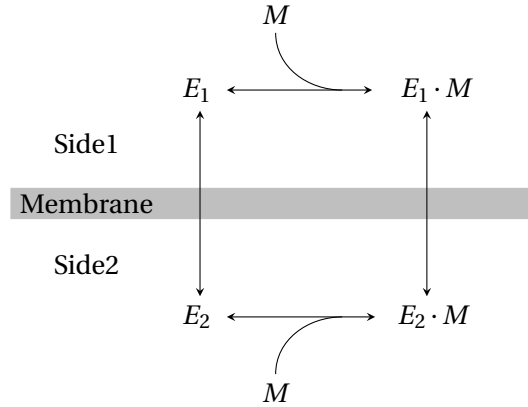


Figure 2.6: Kinetic scheme for the transport of a molecule  $M$  between side 1 and side 2 of the membrane. The transport is mediated by a membrane protein  $E$  that can be in two states  $E_1$  or  $E_2$ .

exchange rates is related to the available Gibbs free energy  $\Delta G$ :

$$\frac{k_{ex}^{ATP \rightarrow ADP}}{k_{ex}^{ADP \rightarrow ATP}} = \frac{1}{\alpha} \frac{K_{d,ATP}}{K_{d,ADP}} = \frac{\alpha_{eq}}{\alpha} \frac{k_h}{k_s} = \exp(-\beta \Delta G) \frac{k_h}{k_s} \quad (2.63)$$

## 2.10 Thermodynamic forces

As a conclusion of this section, we illustrate some of the above-introduced principles through a concrete and simple example of membrane transport [73]. Let's consider a molecule  $M$  that is transported between two sides (side 1 and side 2) of a membrane. The transport is performed by an auxiliary molecule  $E$  which can be seen as a "middleman" in an anthropomorphic picture. The side of the membrane on which the molecule  $M$  binds or detaches is determined by the state of  $E$ . The membrane protein  $E$  is assumed to be in two possible states,  $E_1$  and  $E_2$ , associated to an interaction with side 1 and side 2 respectively (Figure 2.6).

The chemical potentials of molecule  $M$  on each side of the membrane are given by  $\mu_1$  and  $\mu_2$  are defined as:

$$\left. \begin{aligned} \mu_{M,1} &= \mu_M^0 + k_B T \ln([M_1]) \\ \mu_{M,2} &= \mu_M^0 + k_B T \ln([M_2]) \end{aligned} \right\} \Rightarrow X_M := \Delta \mu_M = k_B T \ln \left( \frac{[M_1]}{[M_2]} \right) \quad (2.64)$$

The difference in chemical potential  $\Delta \mu_M$  defined as  $X_M$  corresponds to the driving force that drives the transport of  $M$  from side 1 to side 2.

At equilibrium, the detailed balance holds and the condition on the rates along the cycle is

given by:

$$\frac{\textcircled{O} \text{ product of rates}}{\textcircled{O} \text{ product of rates}} \Big|_{eq} = \frac{k_{E_1 \rightarrow E_1 \cdot M} \cdot k_{E_1 \cdot M \rightarrow E_2 \cdot M} \cdot k_{E_2 \cdot M \rightarrow E_2} \cdot k_{E_2 \rightarrow E_1}}{k_{E_1 \rightarrow E_2} \cdot k_{E_2 \cdot E_2 \cdot M} \cdot k_{E_2 \cdot M \rightarrow E_1 \cdot M} \cdot k_{E_1 \cdot M \rightarrow E_1}} \Big|_{eq} = 1 \quad (2.65)$$

Incorporating the explicit expression for the binding rates of molecule  $M$  on  $E_1, E_2$ , that is  $k_{E_n \rightarrow E_n \cdot M} = [M_n] k_{on, n}$  with  $n = 1, 2$ , the previous equation can be rewritten as:

$$\frac{k_{on, 1} \cdot k_{E_1 \cdot M \rightarrow E_2 \cdot M} \cdot k_{E_2 \cdot M \rightarrow E_2} \cdot k_{E_2 \rightarrow E_1}}{k_{E_1 \rightarrow E_2} \cdot k_{on, 2} \cdot k_{E_2 \cdot M \rightarrow E_1 \cdot M} \cdot k_{E_1 \cdot M \rightarrow E_1}} = \frac{[M_2]_{eq}}{[M_1]_{eq}} \quad (2.66)$$

Importantly, equation 2.66 always holds, the system being at equilibrium or not. It follows that the kinetic parameters are, in general, not independent and always related by the detailed balance constrains. Similarly to equation 2.64, the equilibrium value of  $X_M$  is defined by:

$$X_{M, eq} = k_B T \ln \left( \frac{[M_1]_{eq}}{[M_2]_{eq}} \right) \rightarrow \Delta X_M := X_M - X_{M, eq} \quad (2.67)$$

What changes when the system moves out of equilibrium is that the ratio between the products of rates in one direction of a cycle and in the other is not necessarily equal to 1. In particular:

$$\begin{aligned} \frac{\textcircled{O} \text{ product of rates}}{\textcircled{O} \text{ product of rates}} &= \frac{k_{E_1 \rightarrow E_1 \cdot M} \cdot k_{E_1 \cdot M \rightarrow E_2 \cdot M} \cdot k_{E_2 \cdot M \rightarrow E_2} \cdot k_{E_2 \rightarrow E_1}}{k_{E_1 \rightarrow E_2} \cdot k_{E_2 \cdot E_2 \cdot M} \cdot k_{E_2 \cdot M \rightarrow E_1 \cdot M} \cdot k_{E_1 \cdot M \rightarrow E_1}} \\ &= \frac{[M_1]}{[M_2]} \Big/ \frac{[M_1]_{eq}}{[M_2]_{eq}} = \exp(\Delta X_M) \end{aligned} \quad (2.68)$$

The equilibrium associated to the molecule  $M$  on the two sides of the membrane is defined by a driving force  $\Delta X_M = 0$ , that is  $[M_1]/[M_2] = [M_1]_{eq}/[M_2]_{eq}$ . When the system is moved out of equilibrium, the natural directionality of the transport is governed by the value of  $X_M$  compared to its equilibrium value: if  $\Delta X_M > 0$ , then the molecule moves from side 1 to side 2 and reversely.

In biology, many of the transport mechanisms result from a coupling between different transport of substrates. For instance, as introduced in Section 1.3, various transport mechanisms get advantage of a passive transport through a membrane to actively shuttle another substrate across the membrane. In this context, we extend the previous description of the transport of  $M$  to a second substrate  $S$  to study a simple model of coupled carrier transport. The corresponding scheme is illustrated in Figure 2.7. Similarly to the expression with the binding molecule  $M$ , we can write for  $S$ :  $X_S = \Delta \mu_S = k_B T \ln \left( \frac{[S_1]}{[S_2]} \right)$  and  $\Delta X_S = X_S - X_{S, eq}$ . We assume that the system is such that  $[M_1]/[M_2] > [M_1]_{eq}/[M_2]_{eq}$  and  $[S_1]/[S_2] > [S_1]_{eq}/[S_2]_{eq}$ , that is  $\Delta X_M, \Delta X_S > 0$ , with the additional assumption that  $\Delta X_M > \Delta X_S$ . When considered separately both  $M$  and  $S$  tend to be transported from side 1 to side 2 of the membrane. How can we thermodynamically characterize the global cycle that involves both substrate  $S$  and  $M$ , that is

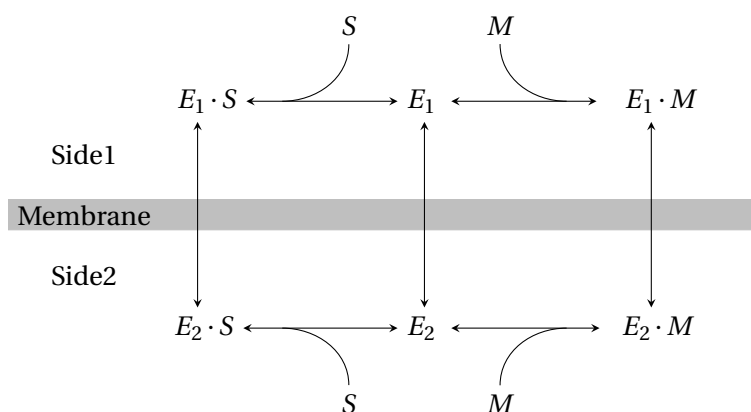


Figure 2.7: The kinetic model in Figure 2.6 is extended with a second substrate molecule  $S$  that can also bind to molecule  $E$ , although not simultaneously with  $M$ .

$E_1 \leftrightarrow E_1 \cdot S \leftrightarrow E_2 \cdot S \leftrightarrow E_2 \leftrightarrow E_2 \cdot M \leftrightarrow E_1 \cdot M \leftrightarrow E_1$ ? This question is crucial to better understand the coupling mechanisms in active transport.

In that case, an equation similar to 2.68 can be written: from the ratio of rates along the global cycle and we find that the resulting driving force of the cycle is given by  $\Delta X_{MS} = \Delta X_M - \Delta X_S$ , positive by hypothesis. Thus the system operates in the direction that consists in moving  $M$  from side 1 to side 2 and moving  $S$  from side 2 to side 1. The free energy of  $M$  is decreased by a value  $\Delta X_M$ , whereas the free energy of  $S$  is increased of  $\Delta X_S$ , the whole system having its overall energy decreased by  $X_M - X_S$ . This is an example of free energy transduction [73] in which the free energy of molecule  $M$  is transferred to the molecule  $S$ , to be used for the transport in the thermodynamically unfavourable direction, that is from side 2 to side 1, whereas the remaining part,  $X_M - X_S$  is dissipated.



## 3 ABC transporters as autonomous Maxwell Demons

The main results presented in this chapter are currently under review and published as a preprint [91]. The majority of the figures are directly reproduced or adapted from this article.

### 3.1 Biological and structural characterisation of ABC transporters

ATP-Binding Cassette transporters (ABC transporters) are one of the most widespread family of active transporters and are present in all kingdoms of life. They play a crucial role in all living organisms, by using energy to drive the transport of several molecules, from scales of ions to large macromolecules. Their abundance and diversity in living systems result from the huge number of tasks that require their ability to perform active transport with an appropriate selectivity, in many different biological organisms [92, 93, 94].

Several nutrients such as lipids, sugars, sterols and vitamins are imported into the cell by ABC transporters and shuttled through the plasma membrane or the organellar membrane (for instance between the cytosol and the endoplasmic reticulum) [95]. Moreover, another crucial role played by ABC transporters is to drive the expel of metabolic waste, toxins or drugs out of the cell.

In addition to being involved in an enormous number of tasks in the cell, ABC transporters are also directly or indirectly implicated in the development of many diseases. On the one hand, the unproper functioning (possibly as a consequence of mutations on the protein sequence) might cause several diseases such as neurodegenerative pathologies (Parkinson, Alzheimer, ...), cystic fibrosis or immune deficiency [96]. On the other hand, their ability to mediate the efflux of drugs (e.g. chemotherapeutic drugs) has been demonstrated, which might promote multidrug resistance [97]. These different mechanisms rise a particular interest for biomedical research on ABC transporters: the precise understanding of the proper functioning is still partially lacking and the potential applications for treatments and regulation of many severe diseases undoubtedly strengthens the interest around the family of ABC transporters [96, 98].

### Chapter 3. ABC transporters as autonomous Maxwell Demons

---

As a preliminary remark which is nonetheless essential, transport by ABC transporters takes place against concentration gradients and thus requires an active process mediated by an energy source, unlike a passive and diffusive transport. Energy consumption is thus required to go against the thermodynamically spontaneous direction of transport: the import or export of selected substrates by ABC transporters is driven by ATP hydrolysis to shuttle molecules against a chemical gradient.

For decades, structural and crystallographic properties of ABC transporters have been extensively characterised, revealing a large spectrum of different conformations that can be observed during the transport cycle [99, 100]. However, there are still gaps in the understanding of the precise energy conversion mechanism: how is the energy released during ATP hydrolysis converted into a chemical potential difference across the membrane? Which are the exact biochemical reactions that induce a conformational change during the transport cycles?

Most scientific advances in the understanding of these mechanisms were possible thanks to the development of experimental techniques to characterize more and more precisely the structure of ABC transporters. These results acted as a starting point to the elaboration of different models that describe the transport mechanism, notably to report on its thermodynamics and energetics.

All ABC transporters share a common molecular structure, a dimer made of two identical monomers (although a few exceptions with distinct monomers have been reported [101]). The first part of the monomer is the Transmembrane Domain (TMD) which spans the plasmic or organellar membrane: it is constituted of multiple  $\alpha$ -helices that are the pathway for the substrate through the lipid bilayer. The second constituent is the nucleotide binding domain (NBD), located inside the cell (that is in the cytoplasm) and is composed of both  $\alpha$ -helices and  $\beta$ -sheets, on which ATP or ADP can bind (or rarely other nucleotides such as GTP in some highly specific cases [102]). Although ABC transporters share a common structure, during billions of years of evolution, members of the ABC transporters family had to adapt to a very broad range of substrates. It results in a large diversity observed in the TMD sequences, whereas the NBD is highly conserved through all ABC transporters reported in the different organisms, likely as a fingerprint of the mechanism itself [103, 104].

Hydrolysis of ATP and nucleotide exchange both drive conformational changes that affect the orientation of the transporter, switching the accessibility of the substrate binding site from one side of the membrane to the other (Figure 3.1 B). When NBD is bound to ADP, substrates bind to or are released from the TMDs on the *in*-side of the membrane. This conformation is usually called open-inside (or inward-facing). Alternatively, when ATP is bound to the NBD, the substrate is caught from or released to the *out*-side of the membrane: this is the

### 3.1 Biological and structural characterisation of ABC transporters

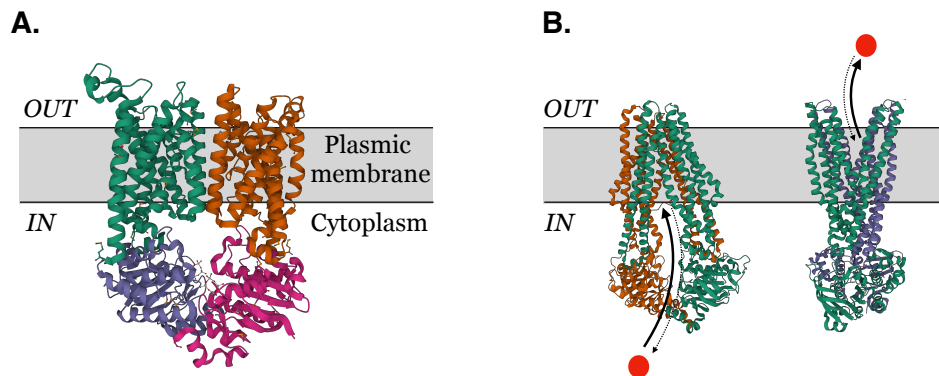


Figure 3.1: Protein structure of ABC transporters from X-Ray experiments. **(A)** Structure of an importer observed in *E. Coli* (*E. Coli* BtuCD structure, PDB 1L7V [105]). One monomer is made of a NBD (in purple for the left monomer, pink for the right monomer) and a TMD (green on the left, orange on the right). The plasmic membrane is schematically drawn in light grey and the two NBDs are located inside the cell. **(B)** Structure of an exporter. The same transporter is observed in two different conformations, inward-facing (left, PDB 3QF4 [106]) and outward-facing (right, PDB 6QV0 [107]). The arrows show the binding and unbinding of substrate (schematic red dot) on the side corresponding to the orientation of the transporter. Plain black arrow is the dominant transition in the case of an export cycle, with respect to the dashed black arrow.

### Chapter 3. ABC transporters as autonomous Maxwell Demons

---

open-outside (or outward-facing) conformation.

At the scale of the transporter in its whole, the global mechanism with a conformational change is relatively well understood presently, but there are still debates and gaps in the understanding of how these specific mechanisms and the underlying intermediary steps are biochemically driven.

Importantly in the working of ABC transporter is the conformational switch of the TMD, depending on the nucleotide bound on the NBD. Each step of the catalytic cycle, which are the ATP binding, ATP hydrolysis as well as ADP and inorganic phosphate release) contribute to the conformational changes in the transporter and, by extension, also input energy into the transport cycle [108].

It is still unclear which precise step in the catalytic cycle induces the conformational change. Some results suggest that the power stroke is rightly generated by ATP hydrolysis, by pushing together and pulling apart the cytoplasmic ends of the TMD [109], possibly through an intermediate conformation of the NBD called *Nucleotide Sandwich Dimer* [110, 99], whereas other results tend to show that the ATP binding generates the power stroke [111, 112].

Mechanistically, the transmission of the signal from the NBD associated to the conformational switch results from the interaction between the helical domain of NBD and the coupling helices of the TMD [93]. The dimeric structure of ABC transporters, and thus the presence of two nucleotide binding domains, raises the question of the role played by the two nucleotides in the transport mechanism. Structural observations show that the nucleotide binding pocket (the precise location of the nucleotide on the NBD) are really close, suggesting a cooperativity of nucleotides in the transport mechanism: it is more reasonable to think of a pair of nucleotides that affect the biochemical properties of the transporter, more than two independent nucleotides, each playing a specific role [113]. Out of a few specific exceptions [114], both ATP binding pockets are always required for the transport mechanism [93].

These conformational transitions, mirroring the interaction from the NBD to the TMD, are obviously essential part of the global mechanism of substrate transport. As a counterpart in the understanding of the transport mechanism is the "reverse" interaction at a distance, that is from the TMD to the NBD. The arising question pertains to the effect of the substrate bound on the TMD. More specifically, it consists in understanding how the structural and kinetic properties of the NBDs are affected by the presence of a substrate on the TMD.

The presence of a substrate strongly affects ATP hydrolysis or nucleotide exchange rates (or both). The induced kinetic asymmetry between states with and without substrates accelerates the ATPase cycle of ABC transporters [93, 115]. Importantly, since the nucleotide binding site is not in direct contact with the substrate, there must necessarily be a long range allosteric transmission pathway, from the substrate binding site to the nucleotide binding domain [116, 117].

### 3.1 Biological and structural characterisation of ABC transporters

---

So far, ABC transporters were described through a general approach, without distinguishing their proper transport function. This choice deliberately lays the foundations for a generic mechanism, but it obviously does not mean that, for an accurate biological description, the structural and functional discrepancies within the ABC transporters family must be totally omitted.

In the past few decades, experimental advances in the characterisation of ABC transporters revealed a large number of structural properties. Very often, they are associated to a specific transport mechanism. The whole family comprises two main families, importers and exporters, that represent the huge majority of transporters. In addition, other less common and more specific transport functions have been reported, such as extractors and mechanotransmitters [104, 118].

Among the importers are two subfamilies, importers of type I and of type II. The essential distinction between both, that likely originates the structural and mechanistic differences, is the task they are assigned to. Type I importers transport diverse small nutrients that are abundant in the cell (sugars, amino acids,...). The orientation switch between the inward-facing and outward-facing conformations is unambiguous and is, as described in the previous section, related to the nucleotide on the NBD. This type of importer has been extensively studied in the case of the maltose transport by E.Coli for instance. Importers of Type II are required to transport larger substrates that are present at low concentration in the cellular environment: a typical member of this family is the protein BtuCD which mediates the import of vitamin B12 [105]. The  $\alpha$ -helices of the TMD form a gate spanning the membrane and the successive rearrangements in their relative structures enclose the substrate in the translocation pathway, progressively shuttling it from the *out*-side to the *in*-side of the membrane.

Exporters are found both in prokaryotes and eukaryotes. In the latter, they are located on the plasma membrane, as for prokaryotes, but also on the organellar membranes [101]. The exact understanding of the mechanism through an experimental description of the structure is much more difficult and still incomplete, because most of the experimental structures do not reveal bound substrate on the TMD [109, 101].

Most of the models (if not all) proposed for ABC transporters, both importers and exporters, find their origins in the alternating access model, an allosteric molecular model that describes a broad range of energy-consuming transporters, such as membrane pumps [120]. Specifically applied to the study of ABC transporters, this description led to various models [93, 109, 121]: the alternating catalytic site model [122] the ATP switch model [113], the reciprocating twin channel model and the constant contact model [123].

Obviously, these models are representative of different transport mechanisms and there is no unified description at the molecular level, so much the required tasks (import ou export) as well as the characteristics of substrates (size, concentration) might be different. However, from a broader perspective, the coupling between substrate- and nucleotide-dependent conformational changes (resulting in different rates between the possible conformations) suggests the

existence of fundamental similarities in the transport by ABC transporters, independently of the substrate and the transport directionality [113, 119, 124].

Therefore, our goal is to develop a model as universal as possible, in which essential features of ABC transporters are captured by a unifying model, as a consequence of a global common mechanism.

## 3.2 Construction and motivation of the model

As a starting point for the elaboration of our model is the work by Douglas Rees et al. [93] in which they built an idealized kinetic model with four states, the minimal number of states that reproduces the working of an ABC transporter: two ATP-bound states and two ADP-bound states, with the corresponding orientation, respectively outward-facing and inward-facing. For each of these conformations, a substrate can be bound on the TMD or not. It is notably shown that three parameters play an important role to optimize the transport performance, illustrated hereafter in the case of an importer (the model omitting all the structural details that distinguish importers from exporters, the analogy with the exporter is direct). First, the affinity of the substrate for the transporter has to be larger in outward-facing than inward-facing conformation. Second, the ATPase activity must be larger when the substrate is bound in the outward-facing conformation. Last, in the absence of substrate, Nucleotide Exchange Factor (NEF, cochaperone proteins that accelerate exchange rates) must favourably drive the transition from an ADP-bound state (inward-facing) to an ATP-bound state (outward-facing) to reset the state of the transporter and thus allow the start of a new transport cycle. An adequate relation between these parameters allows the system, among other, to minimize the number of futile cycles during which energy released by ATP hydrolysis is not effectively involved in the transport of a substrate from one side of the membrane to the other.

Among the three above-mentioned conditions for a performant transport, the second one (a larger ATPase activity in the presence of substrate) is at the heart of our conception of the model, with the aim of improving the fidelity of the model to biochemical and structural arguments. The ATP-bound states must in reality be an ensemble of at least two conformations, characterized by different nucleotide-processing rates, whose relative equilibrium is tuned by substrate binding.

Therefore, we extend the model proposed by Rees and co-workers to report on the allosteric property of the transporter that materializes the interplay between the presence of substrate and its ATPase activity. The initial 4-state model is thus extended to 6 states through the duplication of the ATP-bound states, both in presence and absence of substrates. The four states  $TS$ ,  $DS$ ,  $T$  and  $D$  (independently of the transition rates between each) are the one of the original four-state model [93, 122, 125]. The two additional states  $T * S$  and  $T *$  are added to include the allosteric transition between two states with different nucleotide-processing rates (figure 3.2). Although the model might include few additional transitions, it is also

### 3.2 Construction and motivation of the model

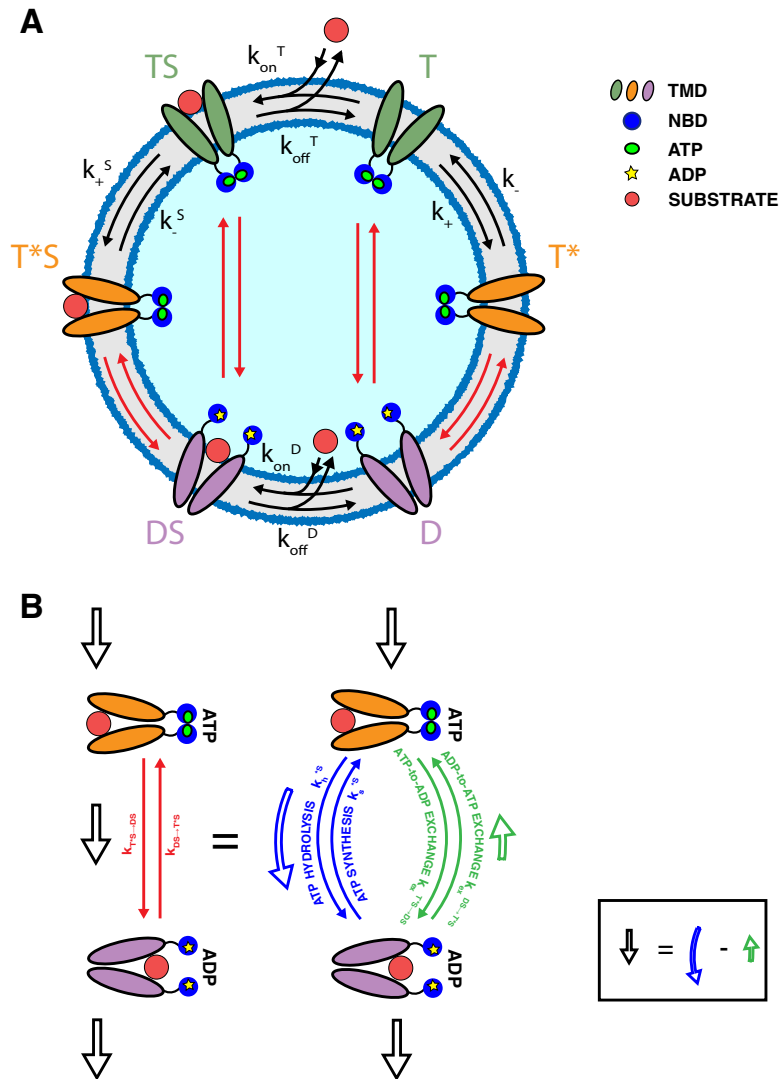


Figure 3.2: **(A)** ABC transporters span the plasma membrane phospholipid bilayer. The ATPase domains (blue circles) are located on the inner side of it and bind ATP (small green circles) or ADP (small yellow circles). When bound to ATP, there are two possible states (orange or green trans-membrane domains), both in an open outside conformation. The relative population of these two states is shifted by the presence or not of a substrate (small red circle) bound on the TMD. The ADP-bound state (open-inside conformation, violet trans-membrane domains) is reached from the ATP-bound conformations through a "two-branch" reaction (red arrows). **(B)** The composite reactions (red arrows in figure A) are implicitly representing the two branches, which are either the ATP hydrolysis/synthesis (blue arrow) or the exchange (green arrows). Overall, the resulting net flux between ATP- and ADP-bound states is depicted with the black hollow arrow.

aimed to stay as simple as possible in order to capture the essential features of the observed phenomenology. Therefore some hypotheses are made to reduce the complexity of the kinetic model.

As a first assumption is the correlation between the transporter orientation and the substrate binding/release ability on the corresponding side of the membrane. In other words, when the transporter is in the ATP-state, respectively ADP-state, the binding and unbinding of substrate are only from/to the *out*- side, respectively the *in*- side, of the membrane. This is qualitatively in line with the description of the transport cycle and also further confirmed by the numerical values of the rates [125] which suggest a difference of four order of magnitude between binding rates on the "good" side with respect to the opposite side (and similarly for the unbinding). Second, the binding/unbinding events between  $T^*$  and  $T^*S$  are not considered. There is obviously no experimental evidence for such an assumption, since the existence of these two additional states  $T^*$  and  $T^*S$  is precisely our main hypothesis, the object of this work. The addition of these two transitions is discussed in section 3.7.

The notation for the different rates corresponding to the transition arrows illustrated in Figures 3.2A and 3.2B is the following:

- $k_h, k_h^S, k_h^*$  and  $k_h^{*S}$  [ $s^{-1}$ ] are the **hydrolysis rates**. They correspond to the hydrolysis of the ATP-bound transporter, respectively in the  $T$  state,  $TS$  state,  $T^*$  state and  $T^*S$  state.
- $k_s, k_s^S, k_s^*$  and  $k_s^{*S}$  [ $s^{-1}$ ] are the **synthesis rates**. They correspond to the synthesis reaction, from an ADP-bound to an ATP-bound state and the same notation for the superscript holds than for the hydrolysis rates.
- $k_{ex}^{T \rightarrow D}$  [ $s^{-1}$ ] is the **exchange rate** from  $T$  to  $D$  state. The other exchange rates are defined with the same notation and the corresponding names of the states.
- $k^{T \rightarrow D}$  and  $k^{D \rightarrow T}$  [ $s^{-1}$ ] (red arrows in Figure 3.2B) are the sum of hydrolysis or synthesis rate and the exchange rate. The same notation holds for the other states, from/to  $T^*$  and in presence of substrate.
- $k_{on}^D$  [ $\mu M^{-1} s^{-1}$ ] and  $k_{off}^D$  [ $s^{-1}$ ] are the **binding and unbinding rates** of substrates, from/to the *in*-side of the membrane. The same is for  $k_{on}^T$  and  $k_{off}^T$ , from/to the *out*-side of the membrane.
- $k_+$  and  $k_-$  [ $s^{-1}$ ] are the **transition rates between  $T$  and  $T^*$** :  $k_+$  from  $T$  to  $T^*$  and  $k_-$  from  $T^*$  to  $T$ .  $k_+^S$  and  $k_-^S$  are similarly defined in presence of substrate.

The two possible transitions between ATP-bound state and ADP-bound state are related to the energy consumption of the system or through exchange. These transitions are unbalanced



### 3.3 Characteristic parameters to capture the phenomenology

out of equilibrium and follow the relation 2.63 introduced in section 2.9 to account for the available energy of the system ( $\Delta G$ ), that is specifically for this model:

$$\frac{k_h}{k_s} = \exp\left(\frac{\Delta G}{k_B T}\right) \frac{k_{ex}^{T \rightarrow D}}{k_{ex}^{D \rightarrow T}} \quad (3.1)$$

$$\frac{k_h^S}{k_s^S} = \exp\left(\frac{\Delta G}{k_B T}\right) \frac{k_{ex}^{TS \rightarrow DS}}{k_{ex}^{DS \rightarrow TS}} \quad (3.2)$$

$$\frac{k_h^*}{k_s^*} = \exp\left(\frac{\Delta G}{k_B T}\right) \frac{k_{ex}^{T^* \rightarrow D}}{k_{ex}^{D \rightarrow T^*}} \quad (3.3)$$

$$\frac{k_h^{*S}}{k_s^{*S}} = \exp\left(\frac{\Delta G}{k_B T}\right) \frac{k_{ex}^{T^*S \rightarrow DS}}{k_{ex}^{DS \rightarrow T^*S}} \quad (3.4)$$

$$(3.5)$$

The detailed balance condition (equation 2.24) must always be satisfied for each of the thermodynamic cycles in Figure 3.2A, not only between exchange and hydrolysis/synthesis, thus restricting the number of independent parameters of the system. Last but not least, we introduce a few parameters to stress the interpretation of some characteristic features of our model.

### 3.3 Characteristic parameters to capture the phenomenology

Based on the different states and rates, we introduce two parameters that conceptually mirror the conception of the model and the different biochemical arguments raised so far.

#### Different ATPase activity between the two ATP-bound states

The states  $T^*$  and  $T^*S$  are characterized by different nucleotide-processing rates compared to states  $T$  and  $TS$  and especially by different hydrolysis rates. To quantify this distinction, we introduce the parameter  $\eta$ :

$$\eta := \frac{k_h^*}{k_h} = \frac{k_h^{*S}}{k_h^S} \quad (3.6)$$

One single value of  $\eta$  is defined on both sides of the system, that is the ratio between the two hydrolysis rates is the same with ( $k_h^{*S}/k_h^S$ ) and without ( $k_h^*/k_h$ ) substrates. The reason is precisely that we aim to point out the existence of two structurally similar but kinetically different ATP-bound states, which exist both in absence and presence of substrates. This is specifically the *raison d'être* of the extension of the model from [93, 125], in which the presence of substrate directly affects the ATPase activity of the ATP-bound (and ADP-bound) states. In our model, the substrates obviously plays a role and the set of three states with substrate ( $TS$ ,  $T^*S$  and  $DS$ ) as a subsystem does not have the same kinetics than the set of three states

without substrate ( $T$ ,  $T^*$  and  $D$ ). This difference is detailed in the next paragraph.

### The presence of substrate shifts the equilibrium towards the one or the other ATP-bound states

In our conception of the model, on the one hand, the binding of the substrate on the TMD tilts the equilibrium in favour of either  $TS$  or  $T^*S$ , each of them with their own ATPase activity. On the other hand, when the substrate is released, the occupation of states  $T^*$  and  $T$  is affected. These two properties are characterised by different equilibrium constants defined in equations 3.7 and 3.8

$$K_e = \frac{k_+}{k_-} \quad (3.7)$$

$$K_e^S = \frac{k_+^S}{k_-^S} \quad (3.8)$$

Combining the imposition of thermodynamic constraints and the introduction of the parameters  $\eta$ ,  $K_e$  and  $K_e^S$ , we report the expression of all the rates in the appendix (Table A.1).

## 3.4 Logics of transport

The time evolution of the occupancy of each state of the system is given by the master equation (equation 2.12) which provides the following set of equations:

$$\frac{dP(TS)}{dt} = -P(TS)(k_+^S + k^{TS \rightarrow DS} + k_{off}^T) + P(T^*S) \cdot k_-^S + P(DS)k_{DS \rightarrow TS} + P(T)[out]k_{on}^T \quad (3.9)$$

$$\frac{dP(T^*S)}{dt} = P(TS)k_+^S - P(T^*S)(k_-^S + k_{T^*S \rightarrow DS}) + P(DS)k_{DS \rightarrow T^*S} \quad (3.10)$$

$$\frac{dP(DS)}{dt} = P(TS)k^{TS \rightarrow DS} + P(T^*S)k_{T^*S \rightarrow DS} - P(DS)(k_{DS \rightarrow TS} + k_{DS \rightarrow T^*S} + k_{off}^D) + P(D)[in]k_{on}^D \quad (3.11)$$

$$\frac{dP(T)}{dt} = -P(T)(k_+ + k^{T \rightarrow D} + [out]k_{on}^T) + P(T^*) \cdot k_- + P(D)k_{D \rightarrow T} + P(T)k_{off}^T \quad (3.12)$$

$$\frac{dP(T^*)}{dt} = P(T)k_+ - P(T^*)(k_- + k_{T^* \rightarrow D}) + P(D)k_{D \rightarrow T^*} \quad (3.13)$$

$$\frac{dP(D)}{dt} = P(T)k^{T \rightarrow D} + P(T^*)k_{T^* \rightarrow D} - P(D)(k_{D \rightarrow T} + k_{D \rightarrow T^*} + [in]k_{on}^D) + P(D)k_{off}^D \quad (3.14)$$

In this problem, we are interested in the non-equilibrium steady-state (NESS):

$$\frac{dP(TS)}{dt} = \frac{dP(T^*S)}{dt} = \frac{dP(DS)}{dt} = \frac{dP(T)}{dt} = \frac{dP(T^*)}{dt} = \frac{dP(D)}{dt} = 0 \quad (3.15)$$

The equations are obviously linearly dependent, from the conservation of the total number of transporters. In order to get a probability distribution for each state, one of the equation

(arbitrarily chosen) can be substituted by a normalisation condition reflecting the conservation of the total number of transporters, that is:

$$P(TS) + P(T^*S) + P(DS) + P(T) + P(T^*) + P(D) = 1 \quad (3.16)$$

The NESS probabilities are associated to a stationary probability occupancy of each state of the transporter. However, it does not imply that the net flux of substrates between the two sides of the membrane equals to zero.

We therefore have to impose another constraint, a zero net flux between the *in*- and *out*- sides of the membrane, which corresponds to a zero net flux between the substrates in solution and substrates bound on the transporter. The two formulations (either relative to the *in* or *out* side of the membrane) are strictly equivalent. In other words, the net flux from the membrane to the *in*- side is necessarily equal (up to a sign convention) to the net flux from the membrane to the *out*- side of the membrane. There is thus a single condition:

$$\begin{aligned} P(TS)k_{off}^T - P(T)[out]k_{on}^T &= 0 \\ \Downarrow \\ P(DS)k_{off}^D - P(D)[in]k_{on}^D &= 0 \end{aligned} \quad (3.17)$$

By substituting the expression of probabilities obtained from the resolution of the master equations (equations 3.9 to 3.14), probabilities that themselves depend on *[in]* and *[out]*, and subsequently solving equation 3.17, an expression for the ratio *[in]/[out]* is found. The sketch of the analytical derivation is described in the Appendix, section A.1.3. The final result is slightly shortened to keep trace only of the emerging mathematical structure of the solution, which gives:

$$\frac{[in]}{[out]} = \frac{[in]_{eq}}{[out]_{eq}} \left[ 1 + \left( \frac{\alpha}{\alpha_{eq}} - 1 \right) \cdot \left( \frac{K_e}{K_e^S} - 1 \right) \cdot \left( 1 - \frac{k_+ k_h^* k_{ex}^{D \rightarrow T}}{k_- k_h k_{ex}^{D \rightarrow T^*}} \right) \cdot \mathcal{F}(\{k\}) \right] \quad (3.18)$$

where  $\mathcal{F}(\{k\})$  is a positive and non-zero function that depends on the whole set of rates. Remarkably, it ensues from the expression that three conditions have necessarily to be satisfied in order to move the system from equilibrium and create a concentration gradient across the membrane, with respect to the equilibrium configuration. As we will see, these conditions can be described as reminiscent of the action of an internal Maxwell Demon. The parallel between the historical experiment of Maxwell Demon (section 1.2) and ABC transporters, which acted as a motivation of our project, is thus a posteriori strengthened by the interpretation of the conditions that emerge from the analytical derivation.

- **Available energy:**  $\alpha \neq \alpha_{eq}$

This first condition is not so surprising: the system requires energy in order to move

from equilibrium. If the concentration ratio between ATP and ADP is equal to the one at equilibrium, then it is thermodynamically impossible for the system to impose and maintain a substrate gradient across the membrane shifted from its equilibrium value.

- **Storage of the measure:**  $K_e \neq K_e^S$

Second, the binding and release of substrate have to affect the kinetics of the system. More precisely the two subsystems  $T$ ,  $T^*$ ,  $D$  and  $TS$ ,  $T^*S$ ,  $DS$  must be driven by different kinetic rates. Following the assumptions that were made to build the model, the asymmetry between the two subsystems is exclusively contained in the equilibrium constants  $K_e$  and  $K_e^S$ . Considering ABC transporters as being driven by an internal and autonomous Maxwell Demon, the condition can be so reformulated: once the substrate is *detected* by the Demon, the measure has to be stored at the molecular level, which consists in tilting the system in one state or another (for instance  $TS$  or  $T^*S$  after binding), in a different way than in absence of substrate.

If  $K_e$  and  $K_e^S$  are equal, the measure is not adequately stored in the system, thus merging the two subsystems as a single one, consequently dissociating the molecular events on the TMD (substrate binding and release) from the kinetics of the nucleotide-processing transitions taking place on the NBD. It is similar to considering a Maxwell Demon that would indeed detect the particle, without being able to record and memorize it for the purpose of operating the door.

The storage of the measurement is the first step of the feedback, whose second step is reflected by the following condition.

- **Directionality of the feedback:**  $k_+^S k_h^{*S} k_{ex}^{DS \rightarrow TS} \neq k_-^S k_h^S k_{ex}^{DS \rightarrow T^*S}$

It is assumed from the two previous paragraphs that the transporter is such that there is available energy in its environment and it is biochemically able to store the measure. Then, the third required condition for transport is related to the directionality of the feedback, which corresponds to the second step after the storage of the measure. More explicitly, it means that, considering the kinetic cycle of import (anti-clockwise in Figure 3.2A) and export (clockwise), their respective total rates along the cycles (i.e. product of the rates) has to be different. The symmetry between both direction must be broken so that the transporter is able to effectively convert the measure into an effective transport of substrate from one side to another.

As a concrete example for an importer, the typical sequence of events that moves the system from equilibrium consists of the detection of the substrate from the *out*-side (binding  $T \rightarrow TS$ ), followed by the storage of the measure and subsequent hydrolysis ( $TS \rightarrow T^*S \rightarrow DS$ ). Then the substrate is released inside ( $DS \rightarrow D$ ). The system is brought back to its initial state through a direct exchange ( $D \rightarrow T$ ), without going through the state  $T^*$ . In fact, only the cycles that go through one hydrolysis/synthesis and one exchange can move the system from equilibrium. Furthermore, the cycles through both

$T^*$  and  $T^*S$  or restricted to the central 4-state cycle are always equilibrated and cannot support the transport of substrate in one preferential direction.

$$\begin{aligned}
 & \frac{T \rightarrow TS \xrightarrow{\text{hydrolysis}} T^*S \xrightarrow{\text{exchange}} DS \rightarrow D \xrightarrow{\text{exchange}} T}{T \xrightarrow{\text{hydrolysis}} D \rightarrow DS \xrightarrow{\text{exchange}} T^*S \rightarrow TS \rightarrow T} : \frac{k_{on}^T k_+^S k_h^{*,S} k_{off}^D k_{ex}^{D \rightarrow T}}{k_h k_{on}^D k_{ex}^{DS \rightarrow T^*S} k_-^S k_{off}^T} \\
 & = \frac{[in]_{eq}}{[out]_{eq}} \frac{k_+^S k_h^{*,S} k_{ex}^{DS \rightarrow TS}}{k_-^S k_h^S k_{ex}^{DS \rightarrow T^*S}} \\
 & \frac{T \rightarrow TS \xrightarrow{\text{hydrolysis}} DS \rightarrow D \xrightarrow{\text{exchange}} T^* \rightarrow T}{T \rightarrow T^* \xrightarrow{\text{hydrolysis}} D \rightarrow DS \xrightarrow{\text{exchange}} TS \rightarrow T} : \frac{k_{on}^T k_+^S k_{off}^D k_{ex}^{D \rightarrow T^*} k_-}{k_+ k_h^* k_{on}^D k_{ex}^{DS \rightarrow TS} k_{off}^T} \\
 & = \frac{[in]_{eq}}{[out]_{eq}} \frac{k_+^S k_h^{*,S} k_{ex}^{DS \rightarrow TS}}{k_-^S k_h^S k_{ex}^{DS \rightarrow T^*S}}
 \end{aligned} \tag{3.19}$$

$$\begin{aligned}
 & \frac{T \rightarrow TS \xrightarrow{\text{hydrolysis}} T^*S \xrightarrow{\text{exchange}} DS \rightarrow D \xrightarrow{\text{exchange}} T^* \rightarrow T}{T \rightarrow T^* \xrightarrow{\text{hydrolysis}} D \rightarrow DS \xrightarrow{\text{exchange}} T^*S \rightarrow TS \rightarrow T} : \frac{k_{on}^T k_+^S k_h^{*,S} k_{off}^D k_{ex}^{D \rightarrow T^*} k_-}{k_+ k_h^* k_{on}^D k_{ex}^{DS \rightarrow T^*S} k_-^S k_{off}^T} \\
 & = \frac{[in]_{eq}}{[out]_{eq}} \\
 & \frac{T \rightarrow TS \xrightarrow{\text{hydrolysis}} DS \rightarrow D \xrightarrow{\text{exchange}} T}{T \xrightarrow{\text{hydrolysis}} D \rightarrow DS \xrightarrow{\text{exchange}} TS \rightarrow T} : \frac{k_{on}^T k_+^S k_{off}^D k_{ex}^{D \rightarrow T}}{k_h k_{on}^D k_{ex}^{DS \rightarrow TS} k_{off}^T} \\
 & = \frac{[in]_{eq}}{[out]_{eq}}
 \end{aligned} \tag{3.20}$$

The same arguments hold for cycles going through synthesis from ADP-bound state to ATP-bound state and exchange for the reverse transition, but this case is obviously thermodynamically much less favourable. Equations 3.19 and 3.20 reflect that the precise path followed by the system is immaterial, in the sense that all paths are thermodynamically equivalent.

The directionality of the transport cycle finds its equivalence into the Maxwell Demon experiment as follows: once the Demon detected and stored the measure, the consecutive action must be dependent on the measure itself. If there is no preferential directionality, i.e. if it opens and closes the door following the same protocol whatever side the particle comes from, then the action of the Demon will not induce any temperature (or equivalently concentration) gradient between both sides of the box.

A pictorial analogy between the Maxwell Demon and ABC transporters is presented in figure 3.3.

These three conditions can be summarized in a sequence of steps that are required for the

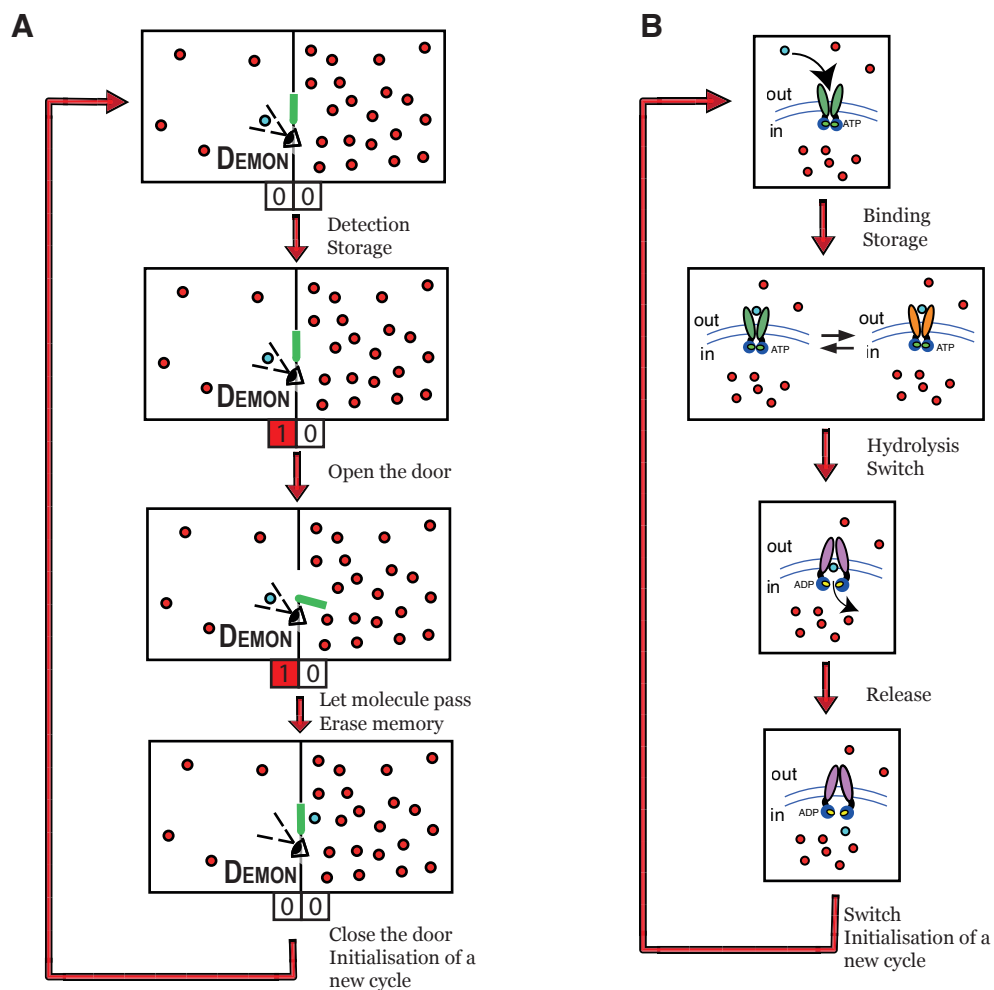


Figure 3.3: **(A)** Illustration of the main constitutive steps associated to the Maxwell Demon experiment. A "double" information-storage device is explicitly represented, with a value of 1 if a particle is detected on one side or the other of the box, a value of 0 otherwise. **(B)** Counterpart of the different steps in the transport cycle by ABC transporter (here importer).

successful action of the autonomous Maxwell Demon that operates at the molecular level the ABC transporter. This reflects a logical AND-like condition that links the three terms in brackets in equation 3.18. First, it must consume energy. Second, it must be able to biochemically record the measure. Third, there must be an asymmetry in the directionality of the transport rates, otherwise the measure and its storage cannot be converted into an effective action, namely the transport of substrate in a preferential direction.

Remarkably, it reveals that the analogy between ABC transporters and the Maxwell Demon is not only conceptual: behind the biochemical properties of ABC transporters is underlying an autonomous Maxwell Demon that operates through a sequence of steps at the molecular level, reproducing each of the information-processing stages that are governing the Maxwell Demon operation in the historical thought experiment.

In section 3.6, we will numerically explore each of these conditions and describe the behaviour of the model more quantitatively. In particular how the concentration gradient  $[in]/[out]$  varies depending on  $\mathcal{F}(\{k\})$ , a term we so far put aside in the discussion of the logical necessary conditions for transport that emerged from the analytical derivation.

### 3.5 Energy cost of information processing

During decades, many physicists struggled to build a formalism of information theory in order to explain (and solve) the paradox of the Maxwell Demon experiment, by considering the thermodynamic cost of the measurement. This opened great perspectives, both fundamentally and through potential applications in various fields. In this section, we *go the other way around* and show how the formalism of information theory applies to the description of ABC transporters as autonomous Maxwell Demon.

Keeping in mind the historical resolution of the Maxwell Demon paradox, we propose an alternative approach that is based on the construction of an energy balance, considering the thermodynamic cost of each information-processing steps during one transport cycle. For the sake of simplicity, the discussion will be focused on the case of a single import cycle. Obviously, the same formalism holds also for an export cycle in the reverse direction (see Appendix A.1.4), thus confirming one more time that import and export processes can be combined in a unified framework.

*Measurement and resetting: A transition between two analogous subsystems*

The only way for the transporter to detect the presence of a substrate, as the Demon would do, is through its binding to the TMD. Considering the three possible states of the NBD as forming one single subsystem, the measurement consists in the transition from the three-state subsystem without substrate ( $T, T^*, D$ ) to the one with substrate ( $TS, T^*S, DS$ ). The complementary process of resetting the system is the reverse transition of substrate release.

Associated to the measurement and resetting is a gain of information about the state of the system: as would the Demon gain information on the state of the system by detecting the approaching particle, the binding of substrate is the only way to internally acquire information at the biochemical level on the state of the system. Similarly, although we refer to it as a *resetting*, the reverse process of unbinding can be described as a gain of information about the state of the system, in the sense that the system gets rid of the degree of freedom on the NBD. In that case the internal measurement takes place between the two subsystems without and with substrate, that is  $\{T, T^*, D\}$   $\{TS, T^*S, DS\}$ . Thus, for both measurement and resetting steps in the transport, the amount of information associated is quantified according to eq 2.53, which transposes to equations 3.21 (measurement from  $T$  to  $TS$ ) and 3.22 (resetting from  $DS$  to  $D$ ) in the case of an import cycle.

$$k_B T I_{\text{measure}} = k_B T \ln \left( \frac{P_3(TS)}{P_3(T)} \right) \quad (3.21)$$

$$k_B T I_{\text{reset}} = k_B T \ln \left( \frac{P_3(D)}{P_3(DS)} \right) \quad (3.22)$$

where  $P_3(T)$  and  $P_3(TS)$  are the steady-state probabilities of the T state conditional to the absence or presence of a bound substrate, respectively. They must be computed on the corresponding 3-state subsystems ( $T-T^*-D$  and  $TS-T^*S-DS$ ) as illustrated in Figure 3.4 A.

#### *Feedback as a source of entropy dissipation*

Although the Maxwell Demon is assumed to be able to operate the door without paying any energy cost, such an eventuality is impossible in a biological system working out of equilibrium. Thus the feedback steps that are at play in the two subsystems  $\{TS, T^*S, DS\}$  and  $\{T, T^*, D\}$  unavoidably cause energy losses in the system, in the form of entropy dissipation into the environment.

More concretely, there are different paths in each feedback loop that might be used to import a substrate. In fact, the choice of the considered path is immaterial, since all dissipate the same amount of entropy, and thus energy. The system is brought to equilibrium by going through one hydrolysis (after the measurement) and one exchange (after resetting). There are formally four combinations that satisfy this constraints. They are listed in equation 3.23 to 3.26 and it is shown that all of them indeed dissipate the same amount of energy. The four additional cycles with one exchange and one synthesis might also be considered, leading exactly to the same entropy dissipation, but they are not mentioned here since thermodynamically almost impossible.

- $TS \rightarrow T^*S \xrightarrow{\text{hydrolysis}} DS$  after measurement and  $D \xrightarrow{\text{exchange}} T$  after resetting

$$\Delta S_{\text{feedback}} = k_B \ln \left( \frac{k_+^S k_h^* k_{ex}^{D \rightarrow T}}{k_-^S k_s^* k_{ex}^{T \rightarrow D}} \right) = k_B \ln \left( \frac{K_e^S k_h^* k_s^* \alpha}{K_e k_s^* k_h^* \alpha_{eq}} \right) \quad (3.23)$$



- $\mathbf{TS} \xrightarrow{\text{hydrolysis}} \mathbf{DS}$  after measurement and  $\mathbf{D} \xrightarrow{\text{exchange}} \mathbf{T}^* \rightarrow \mathbf{T}$  after resetting

$$\Delta S_{\text{feedback}} = k_B \ln \left( \frac{k_h^S k_{ex}^{D \rightarrow T^*} k_-}{k_s^S k_{ex}^{T^* \rightarrow D} k_+} \right) = k_B \ln \left( \frac{K_e^S k_h^* k_s^* \alpha}{K_e k_s^* k_h^* \alpha_{eq}} \right) \quad (3.24)$$

- $\mathbf{TS} \rightarrow \mathbf{T}^* \mathbf{S} \xrightarrow{\text{hydrolysis}} \mathbf{DS}$  after measurement and  $\mathbf{D} \xrightarrow{\text{exchange}} \mathbf{T}^* \rightarrow \mathbf{T}$  after resetting

$$\Delta S_{\text{feedback}} = k_B \ln \left( \frac{k_+^S k_h^* k_{ex}^{D \rightarrow T^*} k_-}{k_-^S k_s^* k_{ex}^{T^* \rightarrow D} k_+} \right) = k_B \ln \left( \frac{K_e^S k_h^* k_s^* \alpha}{K_e k_s^* k_h^* \alpha_{eq}} \right) \quad (3.25)$$

- $\mathbf{TS} \xrightarrow{\text{hydrolysis}} \mathbf{DS}$  after measurement and  $\mathbf{D} \xrightarrow{\text{exchange}} \mathbf{T}$  after resetting

$$\Delta S_{\text{feedback}} = k_B \ln \left( \frac{k_h^S k_{ex}^{D \rightarrow T}}{k_s^S k_{ex}^{T \rightarrow D}} \right) = k_B \ln \left( \frac{K_e^S k_h^* k_s^* \alpha}{K_e k_s^* k_h^* \alpha_{eq}} \right) \quad (3.26)$$

After a slight rearrangement of the expression, we can alternatively show that

$$T \Delta S_{\text{feedback}} = k_B T \ln \left( \frac{[in]_{eq} k_{on}^D k_{off}^T \alpha}{[out]_{eq} k_{off}^D k_{on}^T \alpha_{eq}} \right) \quad (3.27)$$

The definitions (both conceptual and formal) of the measurement, resetting and feedback suggest that the transporter can be, in some sense, described as two three-state systems (Figure 3.4 B), the three states corresponding to the possible states of the NBD. Each system (the two grey triangles) has its own transition rates between the three states, which are (have to be) biochemically modified, depending on the presence or not of substrate. These rates are rightly the one that are comprised in the different steps of feedback (black arrows). The transition between these two systems (red arrows) correspond to the measurement and resetting, through the binding and release of substrate.

Let's take a step back from the equations to consider how energy is expected to be used in the system (Figure 3.4 A). An amount of free energy is brought to the system, in the form of ATP,

with a corresponding free energy  $\Delta G = k_B T \ln([ATP]/[ADP] \cdot [ADP]_{eq}/[ATP]_{eq})$ .

First, a part of it is used to acquire information about the state of the system ( $k_B T I_{\text{measure}}$  and  $k_B T I_{\text{reset}}$ ). This terms numerically appears as a negative contribution because it prevents dissipation in the case  $I_{\text{measure}} + I_{\text{reset}} < 0$ , increasing the final difference in chemical potential. Second, some energy is dissipated by the feedback steps ( $\Delta S_{\text{feedback}}$ ), both to store the measure and go through the hydrolysis/exchange transitions.

At the end of the process, the difference in chemical potential across the membrane

$$\Delta \mathcal{E} := k_B T \ln \left( \frac{[in]}{[out]} \bigg/ \frac{[in]_{eq}}{[out]_{eq}} \right) \quad (3.28)$$

is expected to be equal to the remaining amount of available energy, that is

$$\Delta G - (k_B T I_{\text{measure}} + k_B T I_{\text{reset}}) - \Delta S_{\text{feedback}} \quad (3.29)$$

Combining the definitions and values of these quantities through equations 3.21, 3.22, 3.27, 3.28, A.13, we can indeed show that:

$$\Delta \mathcal{E} = \Delta G - k_B T (I_{\text{measure}} + I_{\text{reset}}) - T \Delta S_{\text{feedback}} \quad (3.30)$$

This alternative approach based on the formalism of information theory is remarkable because it bridges the gap between two conceptions of the same biological system that are apparently very different. On the one hand, we went through the analytical derivation to solve a set of master equations that describe the kinetic of the system and its steady-state, to finally obtain an expression for the concentration ratio of substrates generated across the membrane (equation 3.18). This approach is exclusively based on a biochemical and thermodynamic consideration of the transport process. On the other hand, a more abstract way was used, greatly inspired by the notions and formalism of information theory. We showed that both approaches converge to the same result, supporting our working hypothesis: ABC transporters are autonomous Maxwell Demons whose action is hidden in the sequence of chemical reactions at the molecular level.

### 3.6 Numerical results

The analytical derivation for the steady state concentration gradients highlighted three necessary conditions that must be simultaneously satisfied to move the system out of equilibrium (Equation 3.18). Beyond the logic of these three conditions, the performance of the transporter is defined through the interplay of a whole set of rates and parameters, which cannot be easily captured by the analytical solution. In the first part of this section, we explore how the system

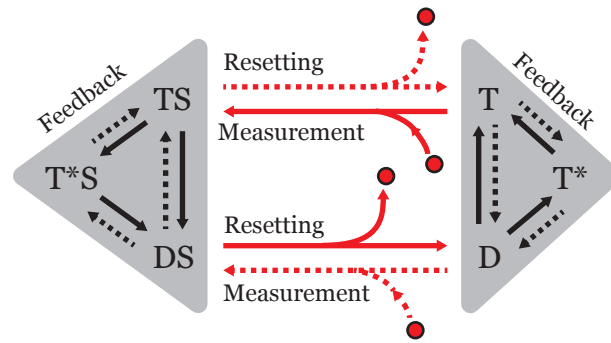


Figure 3.4: Each of the grey triangles corresponds to a three-state system either with (TS-T\*S-DS, left) or without (T-T\*-D, right) bound substrate. Solid and dashed arrows correspond to the typical cycle respectively for the import, respectively export, of substrate. Red arrows are measurement and resetting processes (respectively materialized by binding and unbinding of substrate). Black arrows are the rates realizing the feedback. The composite reactions (hydrolysis/synthesis and exchange) are implicit in the figure.

behaves in non-equilibrium conditions and how its performance depends on the parameters associated to each of the above-mentioned condition. In a second part, we address a few specific examples to show that the model, although very simplistic, is able to qualitatively reproduce experimental results. A possible extension of the model is proposed to give a flavour of possible future results that can be produced with this updated model, opening perspectives to reproduce the phenomenology of other experimental realisations.

If not specifically mentioned, all the results are produced with the rates given in the Appendix, Table A.2.

### 3.6.1 Transport performance far from equilibrium conditions

As a first condition for moving the system out of equilibrium is the presence of an energy source  $\Delta G$  corresponding to the addition of ATP in the biological environment of the transporter. Figure 3.6 shows that similarly for an importer and an exporter, the system reaches relatively fast a plateau upon intake of ATP, around a ratio  $[ATP]/[ADP] \approx 10^{-5}$ . With specific parameters, we observed an intermediary plateau in which the transport performance is slightly lower than the optimal one. However, despite these numerical refinements, the main conclusions always hold and in particular the transport performance is optimal in typical cellular conditions ( $1 \lesssim [ATP]/[ADP] \lesssim 10$  [126, 90], in general 10 in the simulations). The system is always operating at its best, fully harnessing the transport capacity associated to the intrinsic rates of the system. Moreover, the system cannot use the energy source to endlessly

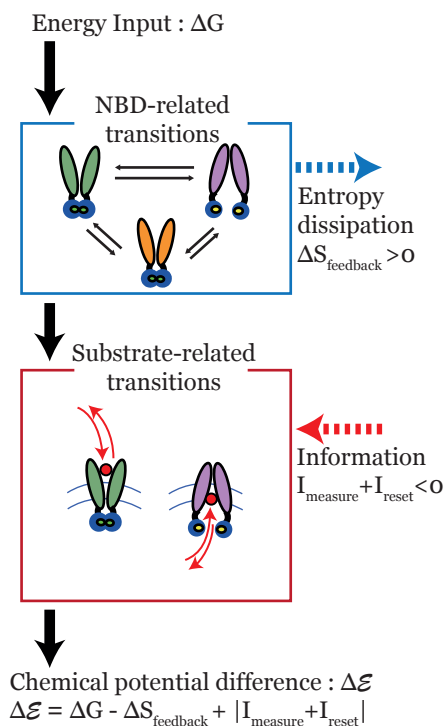


Figure 3.5: Schematic representation of the energy balance for ABC transporters. First, the blue box refers to the transitions in the each 3-state subsystems (either with or without substrate). There is an associated entropy dissipation (blue dashed arrow). Second, the red box refers to the substrate binding and release reactions. The energy associated to the processing of information is depicted with the dashed red arrow. The remaining energy is effectively converted into a net chemical potential difference, through a substrate concentration gradient.

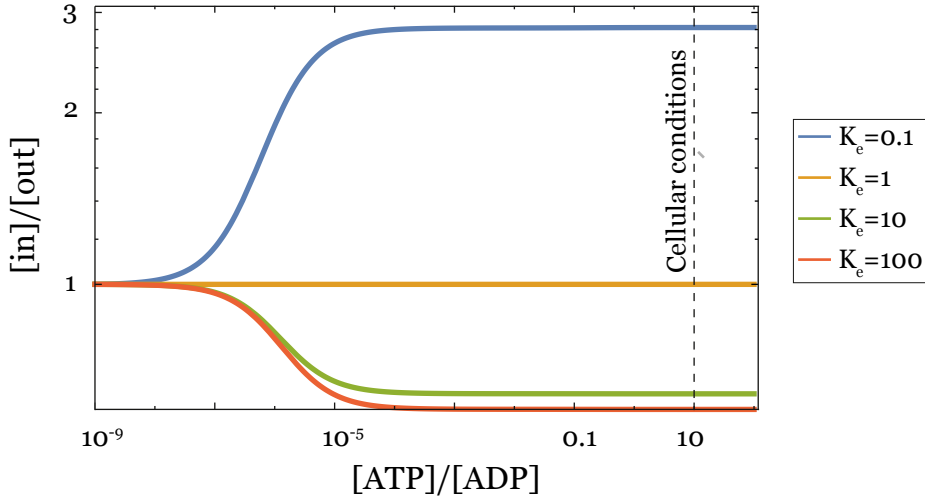


Figure 3.6: Concentration ratio  $[in]/[out]$  as a function of  $\alpha$  for different values of  $K_e$ . In this simulation  $K_e^S = 1, \eta = 10$  and  $[in]_{eq}/[out]_{eq} = 1$ . The vertical dashed line ( $\alpha = 10$ , value used for the future results) corresponds to typical cellular environment (more broadly between 1 and 10 in general).

increase the transport performance: there are asymptotic limits that are imposed by the different rates of the system and their interplay, which induce the presence of the plateau and its corresponding value for the ratio  $[in]/[out]$ .

As a second condition is the storage of the measure associated to the presence or absence of substrate: the equilibrium constant  $K_e$  (eq. 3.7) between states  $T$  and  $T^*$  in absence of substrate has to be different than the one governing the population of states  $TS$  and  $T^*S$  (i.e.  $K_e^S$ , eq. 3.8) when a substrate is bound. In Figure 3.7A, we illustrate this requirement by showing the ratio  $[in]/[out]$  as a function of both equilibrium constants  $K_e$  and  $K_e^S$

As expected, the black dashed line delimits the transition between an importer and an exporter. Over this line, the equality  $K_e = K_e^S$  prevents the system from moving out of equilibrium. In the simulation, the states  $T^*$  and  $T^*S$  are the one with a fast hydrolysis rate ( $\eta = k_h^*/k_h = k_h^{*S}/k_h^S = 10$ ).

For a better understanding of which transport cycle dominates each regime, the fluxes between states are computed when the equilibrium condition on the substrate in solution  $[in] = [in]_{eq}$  and  $[out] = [out]_{eq}$  is imposed. With that given initial condition, the system will evolve towards its non-equilibrium steady-state. This evolution follows one dominant cycle, which is illustrated in Figure 3.7B, both for an importer (top) and an exporter (bottom). The preferential cycle driving the system from equilibrium to a NESS corresponds to what is expected from the construction of the model. In the case of an importer (Figure 3.7B,top), after measurement

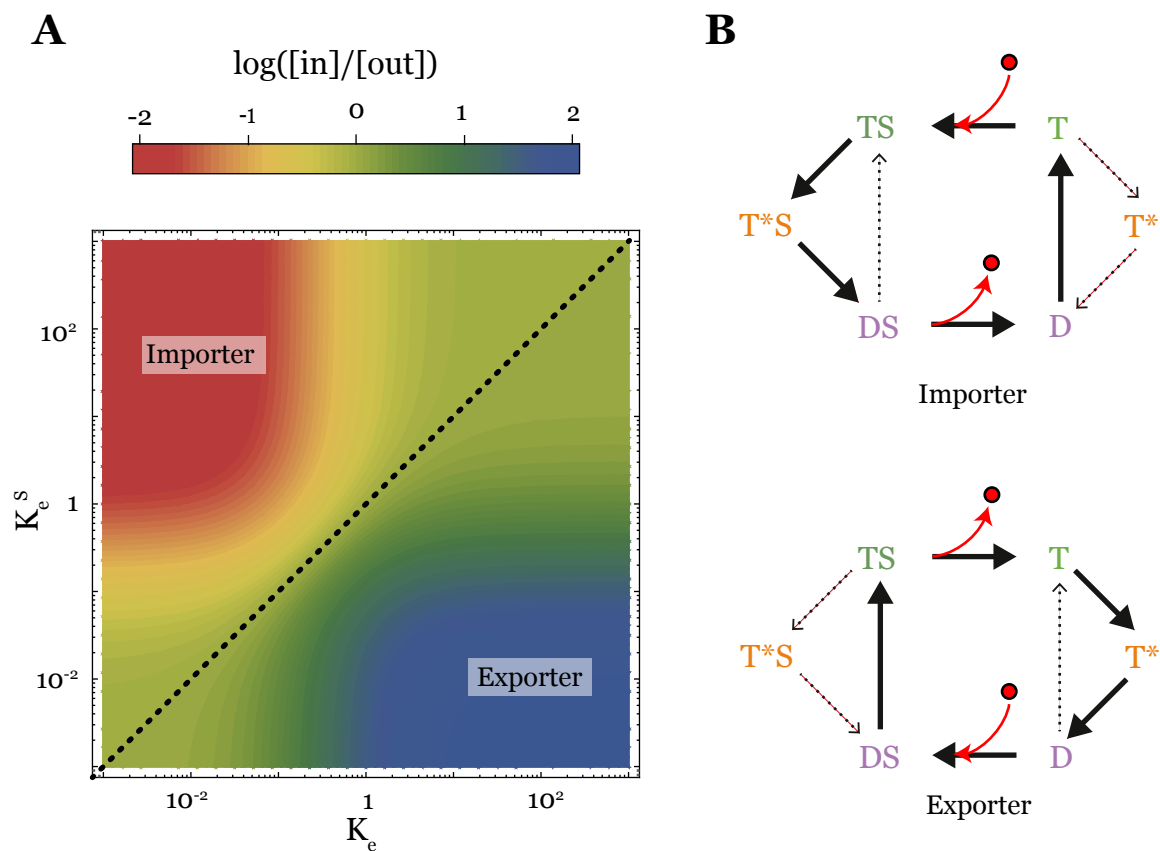


Figure 3.7: **(A)** Concentration ratio  $[in]/[out]$  as a function of  $K_e$  and  $K_e^S$  for different values of  $\eta = 10, \alpha = 10$  and  $[in]_{eq}/[out]_{eq} = 1$ . **(B)** Thick black arrow represent the dominant cycle inducing the transport directionality from the equilibrium concentration conditions ( $[in]_{eq} = [out]_{eq} = 1$ ) towards the steady state of the system). The dashed arrows show the net fluxes between two states which are overall not dominant and oriented in the reverse direction with respect to the global transport direction.

( $T \rightarrow TS$ ), the presence of the substrate is chemically stored by shifting the system towards the state with a fast hydrolysis ( $TS \rightarrow T^*S \rightarrow DS$ ). Then the substrate is released ( $DS \rightarrow D$ ) and nucleotides are exchanged to initiate a new cycle ( $D \rightarrow T$ ).

Alternatively formulated, from the perspective of a Maxwell Demon, the results point out that when the equilibrium constants  $K_e$  and  $K_e^S$  are very different the one from the other, the Demon is able to efficiently acquire the information about the state of the system and detect the presence of a substrate, as it would be able to detect the presence of an approaching particle in the original Maxwell Demon experiment.

The third condition is related to a global directionality that has to be imposed to the system. As shown in equations 3.19, two cycles that both go through one hydrolysis and one exchange, but following the reverse direction must be unequally weighted.

To simplify the analysis and come back to the initial assumption that was used to build the model, a hypothesis is made on the values of the rates: the exchange rates are the same both when it involves the state  $T$  or  $T^*$ , that is  $k_{ex}^{D \rightarrow T} = k_{ex}^{D \rightarrow T^*}$  and  $k_{ex}^{T \rightarrow D} = k_{ex}^{T^* \rightarrow D}$  (and similarly for  $TS$  and  $T^*S$ ). Thus, after imposing the detailed balance constraints, the condition for the directionality can be simplified in the following way:

$$\frac{k_+^S k_h^{*S} k_{ex}^{DS \rightarrow TS}}{k_-^S k_h^S k_{ex}^{DS \rightarrow T^*S}} \neq 1 \implies \eta \neq 1 \quad (3.31)$$

where  $\eta$  is the ratio between hydrolysis rates  $\eta = k_h^*/k_h = k_h^{*S}/k_h^S$  (equation 3.6).

Figure 3.8 shows how the transport performance evolves as a function of  $\eta$ . As expected from the analytical expression, reversing the value of  $\eta$  from  $\eta < 1$  to  $\eta > 1$  is sufficient to reverse the transport directionality from an importer to an exporter. For very small values of  $\eta$ , the hydrolysis and synthesis rates from/to  $T^*$  and  $T^*S$  go to zero, but there is still the exchange rates that are making the transport cycle through  $T^*$  and  $T^*S$  possible, maintaining the system out of equilibrium.

Then, in the regime  $\eta > 1$  there is first an increase in the performance of the transporter that is the consequence of a stronger differentiation between states  $T$  and  $T^*$  ( $TS$  and  $T^*S$  as well). In the language of the Maxwell Demon, it would be associated to the fact that, as a consequence of the detection of an approaching particle, the consecutive action of opening and closing the door isn't the same, depending whether it is detected on the left or on the right. When  $\eta$  is sufficiently increased, the system is largely dominated by the fluxes through the states  $T^*$  and  $T^*S$ , reducing the direct flux between  $T$  and  $D$  as well as between  $TS$  and  $DS$  to zero. Thus the system falls back in an symmetric configuration in which both import and export cycles are equilibrated.

In between, there is an optimal value of  $\eta$ , named  $\eta_{opt}$  that optimizes the transport performance. Interestingly, the value of  $\eta_{opt}$  decreases with  $K_e$  and the evolution of  $\eta_{opt}$  as a

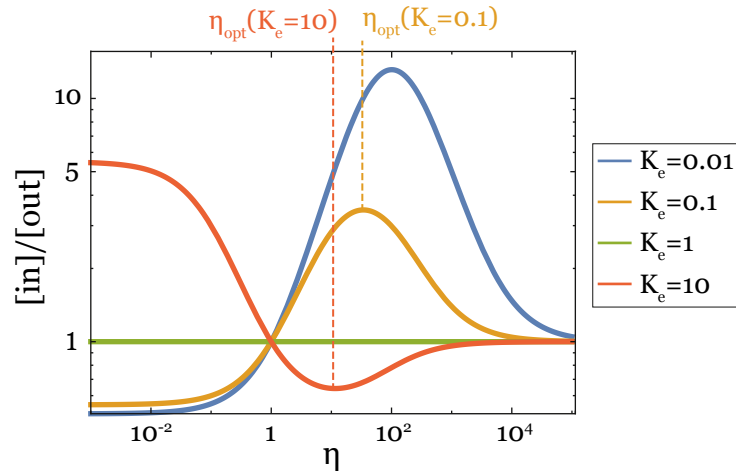


Figure 3.8: Concentration ratio  $[in]/[out]$  as a function of  $\eta$  for different values of  $K_e$ .  $K_e^S = 1$ ,  $\alpha = 10$  and  $[in]_{eq}/[out]_{eq} = 1$ . The vertical coloured dashed lines show in the cases  $K_e = 10$  (red) and  $K_e = 0.1$  (yellow) the value of  $\eta > 1$ , named  $\eta_{opt}$ , for which the concentration ratio  $[in]/[out]$  reaches a local optimum.

function of  $K_e$  for a fixed  $K_e^S$  is shown in Figure 3.9. The explanation is related to the different time scales of the transition rates, and more precisely beyond which value of  $\eta$  the system is effectively dominated by the cycle through both  $TS$  and  $T^*S$  that equilibrates as previously mentioned. The higher  $K_e$ , the smaller has to be  $\eta$  to observe the transition between the two regimes, corresponding to the value  $\eta_{opt}$ .

So far, we quantitatively explored the role played by the main parameters of our model, i.e.  $\alpha$ ,  $K_e$ ,  $K_e^S$  and  $\eta$  in order to explicit their respective effect on the dynamics of the system. The discussion was steered with in mind the autonomous Maxwell Demon that operates in ABC transporter through the different information-processing and energy consuming steps of the transport cycle.

### 3.7 Reproduction of experimental phenomenology

To the best of our knowledge, models that were proposed in the last few decades for transport by ABC transporters often fail to reproduce the very diverse phenomenology observed in experiments. As an example, the mechanism of trans-inhibition was observed with different experimental setups, which is one of the fascinating behaviours exhibited by ABC transporters.

In the next section, we show that the relevance and robustness of our model largely lies in its ability to reproduce the experimental phenomenology observed with ABC transporters. In particular, we point out that we are able to qualitatively reproduce some experimental results,



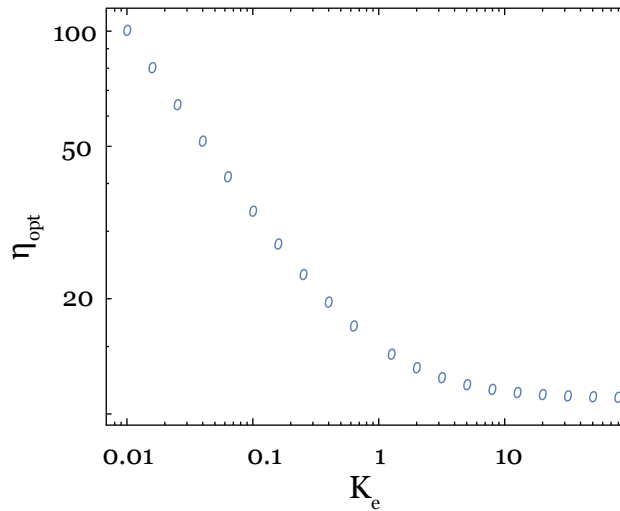


Figure 3.9: Value of  $\eta_{opt}$  for different values of  $K_e$ . The parameters are  $K_e^S = 1$ ,  $\alpha = 10$ . There is no point for  $K_e = K_e^S = 1$ , since the notion of local optimum is not defined in this case, the system being at equilibrium.

despite we have to raise a few assumptions on the numerical values of the rates. We are also facing some limitations that are intrinsic to our conception of the model, which gives us the opportunity to extend it to a slightly more complicated one, upon the addition of possible transitions between states. These results both confirm the robustness of the emerging logic and open exciting perspectives to report on additional experiments.

#### Microscopic reversibility

In section 2.4 was introduced the principle of microscopic reversibility. Specifically applied to our model for ABC transporter, the "energy barrier" that breaks the symmetry and generates a directional process is underlying the feedback steps, both through storage of information and feedback directionality.

Ensuing from microscopic reversibility, two phenomena might be observed, both corresponding to a system working against its "natural" working mechanism. On the one hand, close to equilibrium, the system might use the available energy to produce a net synthesis of ATP against the strong tendency to hydrolyse ATP, which is a thermodynamically much more favourable process. On the other hand, it was also observed experimentally that a system might use the energy from ATP hydrolysis in order to operate according to the reverse motion, which is a priori less favourable [127]. In the following results, we focus on the first described phenomenology and show that, close to equilibrium, our model indeed predicts a net synthesis of ATP.

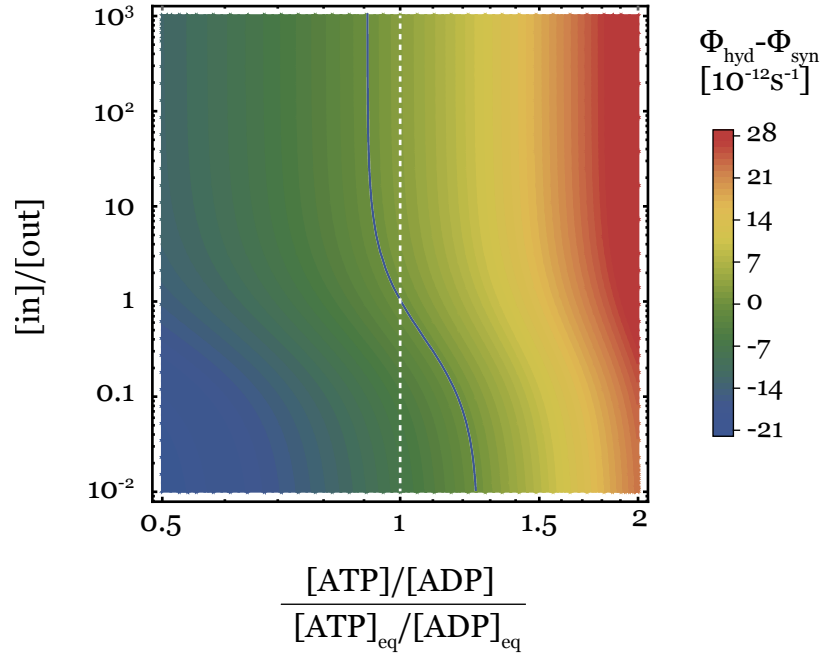


Figure 3.10: Net hydrolysis rate  $\Phi_{hyd} - \Phi_{syn}$ . On the horizontal axis is the ratio  $\alpha/\alpha_{eq}$ , on the vertical axis is the concentration ratio  $[in]/[out]$  (the concentration  $[in]$  is fixed and  $[out]$  varies). The blue curve is the limit  $\Phi_{hyd} - \Phi_{syn} = 0$ : on the left side of the curve, the synthesis is dominant with respect to hydrolysis. The vertical dashed lines corresponds to the equilibrium condition  $\alpha = \alpha_{eq}$ .  $[in] = 1$ ,  $\eta = 10$ ,  $K_e = 100$ ,  $K_e^S = 0.01$ .

We define the net hydrolysis  $\Phi$  [ $s^{-1}$ ] rate as the difference between hydrolysis and synthesis fluxes  $\Phi_{hyd}$  and  $\Phi_{syn}$ :

$$\Phi := \underbrace{[P(T)k_h + P(T^*)k_h^* + P(TS)k_h^S + P(T^*S)k_h^{*S}]}_{\Phi_{hyd}} - \underbrace{[P(D)(k_s + k_s^*)P(DS)(k_s^S + k_s^{*S})]}_{\Phi_{syn}} \quad (3.32)$$

In Figure 3.10, we show that very close to equilibrium and for specific imposed concentrations of substrates, the net hydrolysis flux  $\Phi$  gets negative, corresponding to a dominant synthesis of ATP over hydrolysis.

### Trans-inhibition

In 2008, Gerber et al. performed experimental works highlighting how substrates in solution can inhibit the transport by ABC transporters, a phenomenon called "trans-inhibition" [128]. The inhibitory effect is caused by the increase in the substrate concentration on the target

### 3.7 Reproduction of experimental phenomenology

side after the translocation (the *in*- side for an importer and reversely). In the experiment, Gerber and co-workers studied a transporter called *MaModBC*, specific for molybdate and tungstate substrate. Moreover, its specificity is to have a regulatory domain on the NBD. Its exact function is not well understood but it seems that it plays an essential role to mediate the inhibiting effect of substrates on the working of MaModBC transporter. The authors computed the relative ATP hydrolysis rates in the presence of different substrates. For high concentration of both substrates targeted by the transporter (molybdate and tungstate), the ATP hydrolysis drops really fast, as soon as the substrate concentration is increased (Figure 3.11A). Additionally, the experiment was reproduced with transporters without regulatory domain and the ATPase activity was not affected by the increase in the concentration gradient, confirming that the regulatory domains is necessary for the substrates to inhibit the activity of the transporter.

We used our model to compute the relative ATP hydrolysis rate according to equation 3.33

$$\text{Relative ATPase activity}([in]) = \frac{\Phi_{hyd}([in])}{\Phi_{hyd}([in] = 0)} \quad (3.33)$$

where  $\Phi_{hyd}$  is the hydrolysis net flux as defined in equation 3.32.

Although the ratio  $[in]/[out]$  depends only on the rates, the flux depends also on the absolute value of the concentrations. Thus the concentration  $[in]$  is the parameter tuned in the simulation and  $[out]$  is computed accordingly at steady-state. In Figure 3.11B, we show how our model reproduces this trans-inhibition phenomenon, with a strong decrease of the hydrolysis flux when the concentration on the target (*in*-) side is increased, which is qualitatively in a strong accordance with the experimental results produced by Gerber et al. Unfortunately, this result has to be considered with caution. Indeed, in order to produce this result, the symmetry between the rates, with and without substrates that was so far only broken through the equilibrium constants  $K_e$  and  $K_e^S$  is also internally broken on the exchange rates (through binding rates  $k_{+T}$ ,  $k_{+T}^*$ ,  $k_{+T}^S$  and  $k_{+T}^{*S}$ ). The fundamental reason of it is not so clear, and there are two possible explanations. First, the rates we are using are not compatible with the reproduction of trans-inhibition while keeping the hypotheses of the model. This would require to carefully study the experimental parameters of the experiment to, hopefully, reproduce the results while keeping the symmetry arguments valid. The alternative explanation might be much deeper: in the internal structure of the model, the reproduction of trans-inhibition is incompatible with the hypotheses that were made, and in particular regarding the symmetry breaking between rates in the substrate-bound or -unbound states. We are possibly facing the limitation of our minimal kinetic toy model that we tried to keep as simple as possible.

Interestingly, this limitation arises when we aim to reproduce experimental phenomenology

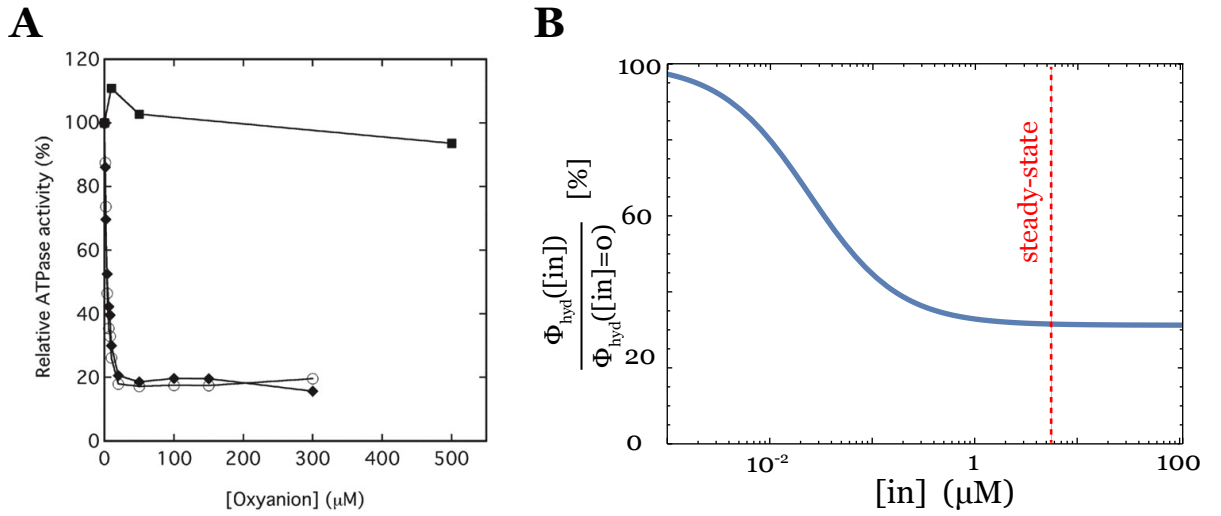


Figure 3.11: **(A)** Relative ATPase activity as a function of the oxyanion concentration. The plot is shown for three different oxyanions: molybdate (open circles), tungstate (solid diamonds) and sulfate (solid squares). Figure from [128]. **(B)** Relative hydrolysis flux as a function of the imposed concentration of substrate on the *in* side of the membrane. The reference value is the hydrolysis flux when [in] goes to zero. The [out] concentration being fixed, the vertical red dotted line corresponds to the steady-state ratio [in]/[out]. Numerical parameters:  $k_{+T} = 0.5\text{s}^{-1}$ ,  $k_{+T}^S = 5\text{s}^{-1}$ ,  $k_{+T}^* = 0.5\text{s}^{-1}$ ,  $k_{+T}^{*,S} = 0.05\text{s}^{-1}$ ,  $\eta = 100$ ,  $K_e = 0.01$ ,  $K_e^S = 0.02$ ,  $\alpha = 10$ , [out] = 1 [ $\mu\text{M}$ )]

that includes also other protein domains, such as the regulatory domain in this specific case, which is on its own a confirmation of the huge biological complexity of the phenomenon.

#### Extension of the model

The previous illustration consisting in the modelling of trans-inhibition shed light on possible limits of our model. To illustrate it with an additional example, we come back to the derivation of the solution (Eq. 3.18). We showed that the steady-state concentration ratio depends only on the rates, and not on the relative concentrations inside and outside the cell. However, this result is not in accordance with experiments found in the literature, showing that there are effects induced by the absolute value of the substrate concentration. In particular, in 2014, Grossmann et al. studied the export of antigens by ABC transporters outside the endoplasmic reticulum lumen, which is essential to the proper working of the immune system [129]. They showed that the increase of the outside concentration of antigens inhibits the transport and thus saturates the internal (lumenal) concentration of antigens (Figure 3.13 B).

In order to break the linearity of our model (at the origin of the independence of  $[in]/[out]$  on the absolute concentrations), we introduce a transition that was neglected so far, namely the possibility for the transporter to bind and release substrates between states  $T^*$  and  $T^*S$ . Before going on with the experiment performed by Grossmann, we first show that all the main principles that were described in the most simple model are robust upon this extension: Figure 3.12 shows that the three conditions detailed in section 3.18 are still necessary conditions to move the system out of equilibrium. More in details, there must be an input of energy (A), a feedback directionality through the parameter  $\eta \neq 1$  (B) and the equilibrium constants  $K_e$  and  $K_e^S$  must be different (C).

Coming back to the experiment by Grossmann, we show in Figure 3.13 that when the outside concentration  $[in]$  is numerically increased, the concentration  $[out]$  is lower than what is observed in the linear case (corresponding to  $k_{on}^{T^*} = 0$ ), thus confirming the existence of an inhibitory effect of the substrate on the activity of ABC transporter. As a side remark, and counter-intuitively, the external peptide concentration in the experiment corresponds in our model to the concentration on the *in*- side of the transporter, relatively to its orientation through the membrane. However, as shown in Figure 3.13A, our result do not exhibit a so strong trans-inhibition that lead to a full saturation of the system as experimentally observed (Figure 3.13 A). It is likely that with other set of parameters, the asymptotic limit can be decreased, but we restricted this work to a more phenomenological reproduction, without paying a specific attention to the quantification of the amplitude.

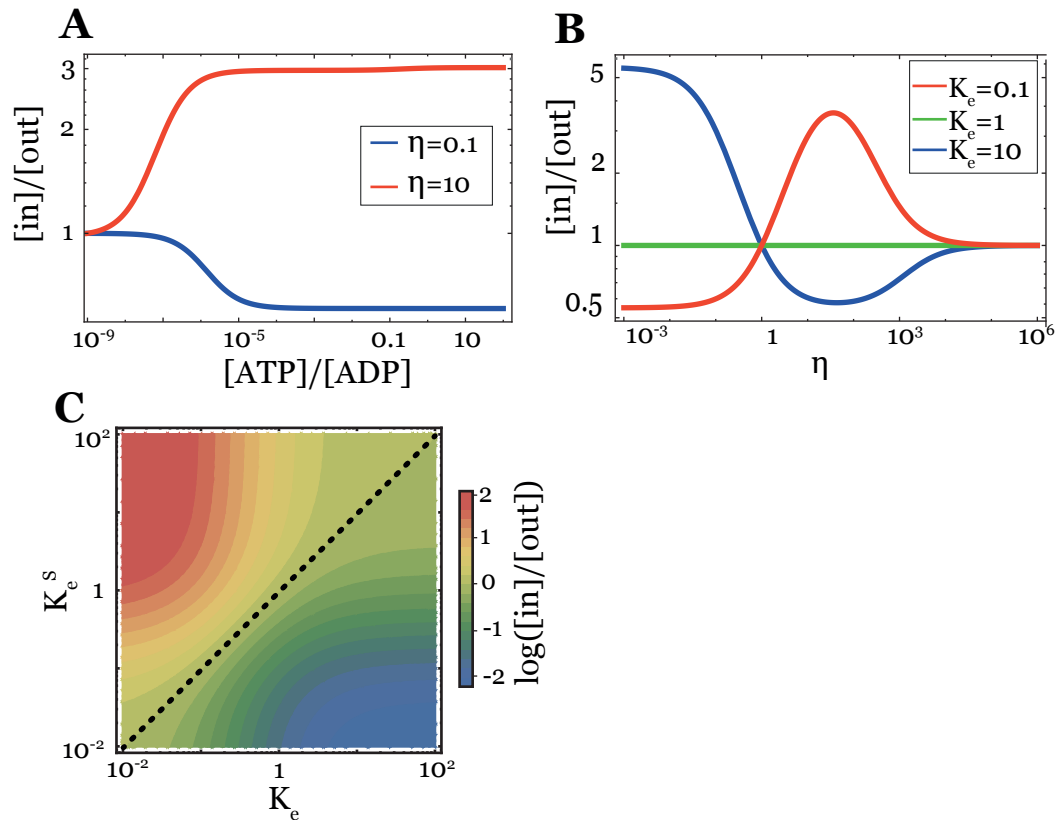


Figure 3.12: The logic of ABC transporter remains unchanged upon addition of the binding (unbinding) transition between  $T^*$  and  $T^*S$ . **(A)** Concentration ratio  $[in]/[out]$  as a function of  $[ATP]/[ADP]$ , for  $\eta = 0.1$  (blue) and  $\eta = 10$  (red).  $K_e = 10$  and  $K_e^S = 1$ . **(B)** Concentration ratio  $[in]/[out]$  as a function of  $\eta$ .  $K_e^S = 1$  and  $K_e = 0.1$  (red), 1 (green), 10 (blue).  $[ATP]/[ADP] = 10$ . **(C)** Concentration ratio  $[in]/[out]$  as a function of  $K_e$  and  $K_e^S$ .  $[ATP]/[ADP] = 10$  and  $\eta = 10$ .

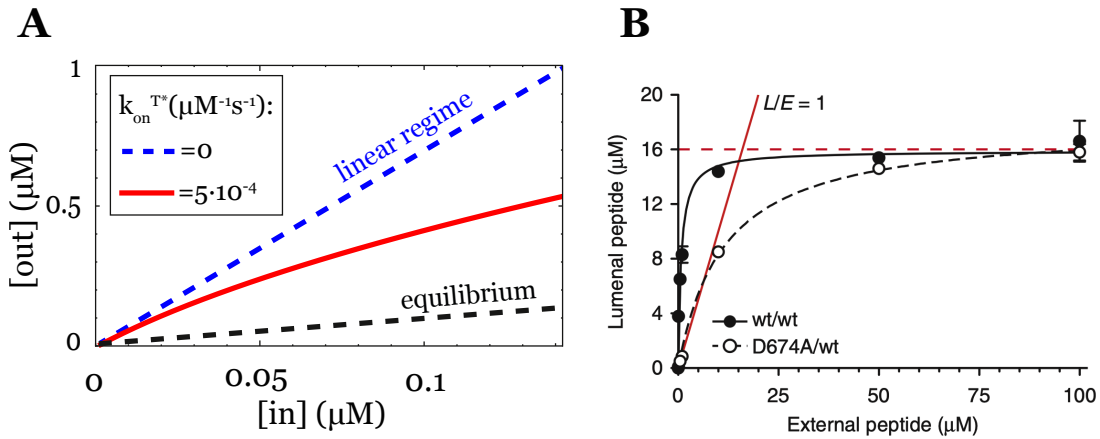


Figure 3.13: **(A)** Steady-state [out] concentration as a function of an imposed [in] concentration, for different binding rates from state  $T^*$  to  $T^*S$  ( $k_{on}^{T^*} = 0$  (dashed blue, corresponding to the initial model),  $k_{on}^{T^*} = 10^{-4}$  (red). The equilibrium state of the system is shown with the black dashed line.  $k_+ = k_+^S = 10^{-3}$  [ $\text{s}^{-1}$ ],  $k_{+T} = 5$  [ $\text{s}^{-1}$ ],  $k_{+T}^* = 0.05$  [ $\text{s}^{-1}$ ],  $K_e = 10$ ,  $K_e^S = 0.1$ ,  $\eta = 10$ ,  $\alpha = 10$ . **(B)** Internal (Luminal) concentration of peptide as a function of the external concentration. Experimental results are shown for a wild-type transporter (wt/wt, black dot) and a mutant complex (D674A/wt, open circles). The red dashed horizontal line is the saturation threshold. The red line (L/E=1) shows the linear regime in which both concentrations are equal. Figure from [129]

### 3.8 Conclusive discussion

In this chapter, we built a minimal toy model for ABC transporters whose building blocks are exclusively imported from structural and biochemical results that have been developed for thirty years.

The picture that has emerged from our model shows that there are striking similarities between the transport by ABC transporters and the action of the Maxwell Demon in the historical thought experiment. The similarities are not only conceptual, but it amazingly turns out that they arise from the mathematical solution of the model. All the information-processing steps associated to the operation of an autonomous Maxwell Demon find their counterpart both in the different steps of the transport cycle and also, more strikingly, as logical necessary conditions to shift the system from its equilibrium state. These conditions are associated to a symmetry breaking, either between substrate-bound and -unbound states ( $K_e \neq K_e^S$ ) or more globally in the transport directionality ( $\eta \neq 1$ ). As a consequence of these logical conditions is the existence of different routes to turn an importer to an exporter.

These results are further supported by an energy balance that drives the transport by ABC transporters, bridging all the thermodynamic quantities that describe the system: the available energy, entropy dissipation, the amount of internally processed information and the resulting chemical potential across the membrane.

In the second part of this work, we showed that the model is able to reproduce the phenomenology of different experimental results, thus giving a flavour of the potentiality of such a model, despite its apparent simplicity. On the other hand, we also faced the limitations of the model, induced by the same above-mentioned simplicity. Thus the model was extended in order to reproduce non-linear behaviours and additional experimental results. This evolution showed that the logic of ABC transporters is conserved and not restricted to a model in its greatest simplicity. This obviously strengthens the robustness of our model and opens perspectives on promising developments that such a model can originate to push further the fundamental understanding of transport mechanism by ABC transporters.

In the continuity, one natural extension of the model of great interest would be to tackle the question of substrate selectivity in the transport process, which turns out to play a crucial role in many biological natural processes, as well as in the development of new drugs that need to target specific substrate molecules. It would be very interesting to highlight how this selection mechanism processes information, possibly through the implementation of a *demonic kinetic proofreading* model aiming at optimising the required selectivity of the transport process. These results might possibly exhibit the existence of competition and inhibition between cognate and non-cognate substrates, as already described in the literature [130].



Taking a step back, all these results implies that ABC transporter not only behaves like Maxwell Demon, but indeed is an autonomous Maxwell Demon, in the sense that each of the information-processing step of the Maxwell Demon naturally takes place in the form of biochemical reactions at the molecular level. The formalism of information theory is not only "used to" describe in a excessively simple way how active biological systems such as ABC transporters work, but indeed, the internal information processing devices are essential constitutive part of the transporter itself.

As a final word, this work opens many doors to explore other cellular transporters, always from the angle of Information Theory. In this context, the next section is dedicated to the study of a different type of cellular transport, the translocation of substrate proteins by chaperone proteins Hsp70, through the membrane of mitochondria and endoplasmic reticulum.



## 4 Substrate Translocation induced by Entropic pulling

### 4.1 Biological context for protein translocation

Translocation is the biological process during which proteins move between cellular compartments through aqueous and gated channels. It takes place in various cellular organisms and organelles: translocation has been observed among others in the Endoplasmic Reticulum (ER) of eukaryotes, the plasma membrane, mitochondria, chloroplasts and peroxisomes [131]. Several principles are shared by these translocation mechanisms such as the requirement of molecular chaperones in the cytosol and inside the organelles as well as the consumption of ATP (or GTP) to drive the hydrolysis step [132], supporting a unified framework for their study and description, although each of them has its own specificities, as a consequence of the biological environment and constraints.

In this work, we focus on the translocation driven by a specific molecular chaperone, Hsp70, which mostly takes place in two cellular organelles, mitochondria and the ER [93].

An evolutionary hypothesis to explain the origin of mitochondria is endosymbiosis [133]. It is very likely that billions of years ago, some bacteria survived endocytosis by other cells and evolved towards symbiosis with the hosting cell becoming mitochondria. The initial bacterial DNA was drastically reduced and simplified during evolution, to keep only a few specific functions. The remaining functions are part of the symbiotic interaction between mitochondria and the cell, that is essentially metabolic and to be the power supply of the cell. Many other functions are not encoded in the mitochondrial DNA: the import of proteins into the mitochondria is thus required to guarantee its proper functioning.

All the cellular proteins resulting from mRNA transcription by ribosomes are divided into two categories: the ones that are used within the cell and the ones that are synthesised to be exported outside the cell. Whereas, in the first case, the translation is performed by free polysomes, in the second case, mRNA strand is translated by individual ribosomes on the rough ER and then translocated into the ER Lumen. After being translocated into the lumen,

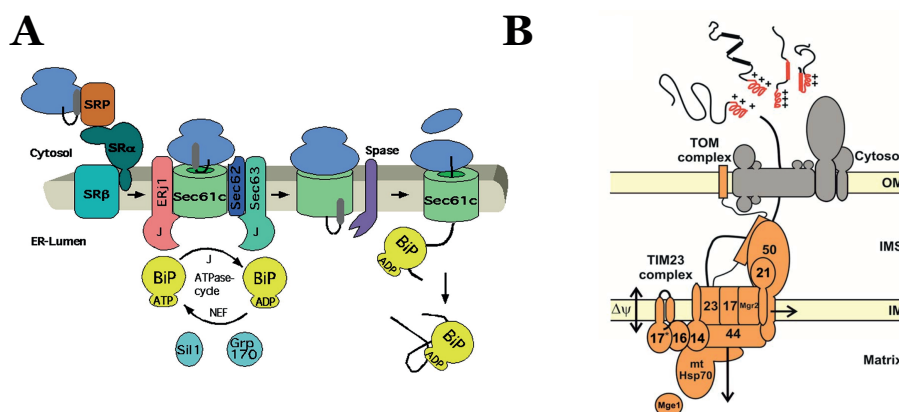


Figure 4.1: Schematic representation of the translocation channel: **(A)** in the ER. Figure from [134] **(B)** in the mitochondrion. Figure from [135].

the proteins are packed into small vesicles that are packaged to be delivered outside the ER. Finally, the vesicles fuse with the plasmic membrane to, after intermediary steps, discharge their proteic content outside the cell or on the extracellular side of the plasma membrane.

The structure of the channel in the ER and mitochondria as well as the translocation mechanism itself turn out to be very similar, as illustrated with the schematic representation of the translocation channel in Figures 4.1A and 4.1B.

An extended description of the structures are available in the literature for mitochondria and for the ER [134, 136, 135]. In the next paragraph, we focus on the main elements of the translocation machinery, in particular the ones that are part of our model.

The central subunit of the translocation channel, the one through which protein is translocated, is the complex Sec61c in the ER (referred by (ER)) and Tim23 in mitochondria (referred by (mt)) and other subunits are structured all around it to help translocation.

The Hsp70 chaperone protein (BiP (ER), mtHsp70 (mt)) is the motor for the translocation, through successive binding to the translocating substrate protein after its insertion into the pore. It stimulates the translocation of precursor proteins [137] and drives the translocation through a mechanism that cannot be understood without having a closer look to the structural properties of Hsp70.

Hsp70 is made of two main domains: first the Nucleotide Binding Domain (NBD) that hosts the binding of either ATP or ADP. The NBD itself is composed of different lobes surrounding the nucleotide denoted by Lobe I and Lobe II (Figures 4.2A,B in light and dark blue, respectively). Second is the Substrate Binding Domain (SBD) composed of two subdomains, SBD $\alpha$  and

## 4.1 Biological context for protein translocation

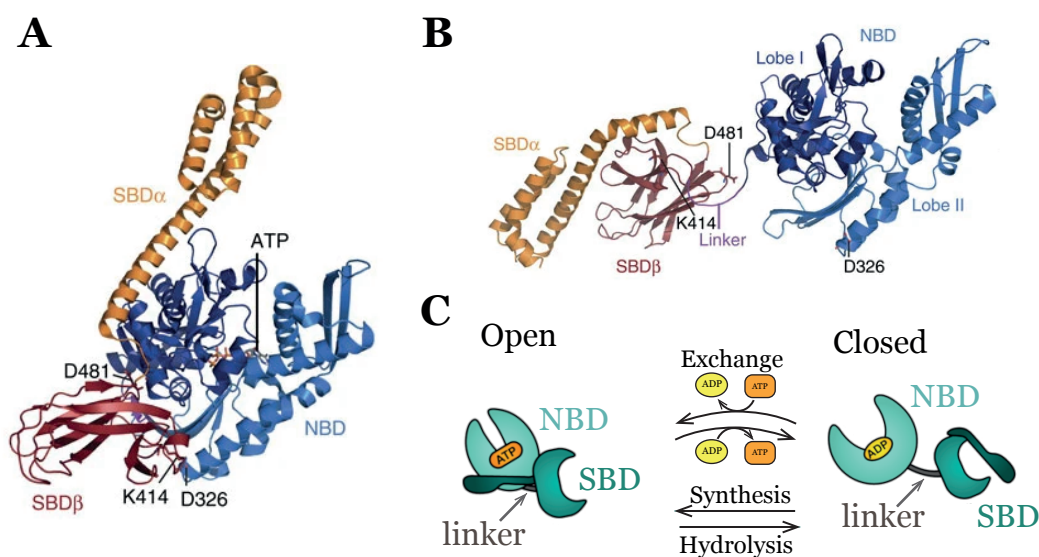


Figure 4.2: **(A)** and **(B)**: Protein structure of Hsp70 with its different subunits in the open **(A)** and closed **(B)** conformations. The SBD subunits are in orange and pink. The NBD subunits are in light and dark blue. Visible and exposed in the closed conformation **(B)** is the linker (purple). Figures from [138]. **(C)** Two kinetic paths between the open and closed conformations: exchange or hydrolysis/synthesis. Figure adapted from [139].

SBD $\beta$ , which are also known as Helicoidal Lid SubDomain (HLSD) and Substrate Binding SubDomain (SBSD) (Figures 4.2A,B in orange and red, respectively). The conformation of the SBD is greatly affected by the nucleotide on the NBD, through an allosteric pathway between the NBD and the SBD. The subdomain SBD $\alpha$  acts as a lid closing upon SBD $\beta$  and "capturing" the substrate when bound (Figure 4.2B)[138].

The two domains NBD and SBD are distant over the protein sequence, with a long linker in between. The conformational rearrangement of the protein upon nucleotide binding also affects the structure of the linker: the linker is rigid and hidden when bound to ATP since the different domains are folding around it. When ADP is bound on the NBD, the linker is getting more flexible and exposed to the environment, enabling the possible binding of other proteins on the linker itself.

The schematic representation of the allosteric conformational change between open and closed conformations is shown in figure 4.2C. There are two possible routes to switch from one to the other, which are either hydrolysis/synthesis or exchange of nucleotides ATP $\leftrightarrow$ ADP.

In the structure of both translocation channels, the presence of J-proteins anchored to the membrane stands out: Sec63 and ERj1 (ER, [134]) and Tim14 (mt, [140]) that stimulate the

ATPase activity of Hsp70. Although the precise interactions are still unclear, it is believed that their function during the translocation process is improved and/or stabilized thanks to their interaction with J-like proteins: Sec62 (ER, [141]) and Tim16, Tim17 (mt, [142]).

In addition to J-proteins, other co-chaperones play a role in the translocation. The Nucleotide Exchange Factors (NEFs) increase the exchange rates from ADP-bound to ATP-bound states, to indirectly accelerate the release of Hsp70 from its substrate. The NEF are Sil1, Grp170 (mt) and Mge1(ER) [136, 143].

An important part of the translocation machinery in the mitochondria is the Tim44 subunit that acts as a organizational hub between all the different subunits [135]. Interestingly, Tim44 seems to have no counterpart in the ER.

### 4.2 Different models to explain the translocation mechanism

The structure and the role of each component is increasingly understood, as well as the interactions between each of them. Yet the precise mechanism by which chaperone protein Hsp70 drives translocation is still debated. Historically, three main mechanisms have been proposed in the last decades: the power stroke model, the Brownian ratchet model and the entropic pulling model. Here, we give a brief description of the first two, whereas the third one is the cornerstone of the current project.

#### 4.2.1 Brownian ratchet model

The first model proposed in 1994 is known as the Brownian ratchet model (Figure 4.3A). The dynamics, in the supposed absence of Hsp70, is governed by Brownian motion due to thermal fluctuations without any preferential direction of motion. The model suggests a cooperativity between the protein complex Tim44 and Hsp70: Tim 44 first binds to the emerging portion of the translocating chain, before transferring it to Hsp70. It follows that the binding of Hsp70 close to the pore prevents retro-translocation, thus imposing the direction of the subsequent thermal motion. Once a new portion of the chain is emerging, the process is repeated, opening the possibility of binding another Hsp70 [143]. In this model, there is no net force that is strictly applied on the translocating polypeptide, but only a directionality that results from obvious steric arguments.

#### 4.2.2 Power stroke model

Two years later, Horst et al. proposed an alternative model known as the power stroke model (Figure 4.3B) [146]. In this description, Hsp70 binds to the substrate protein close to the pore,

## 4.2 Different models to explain the translocation mechanism

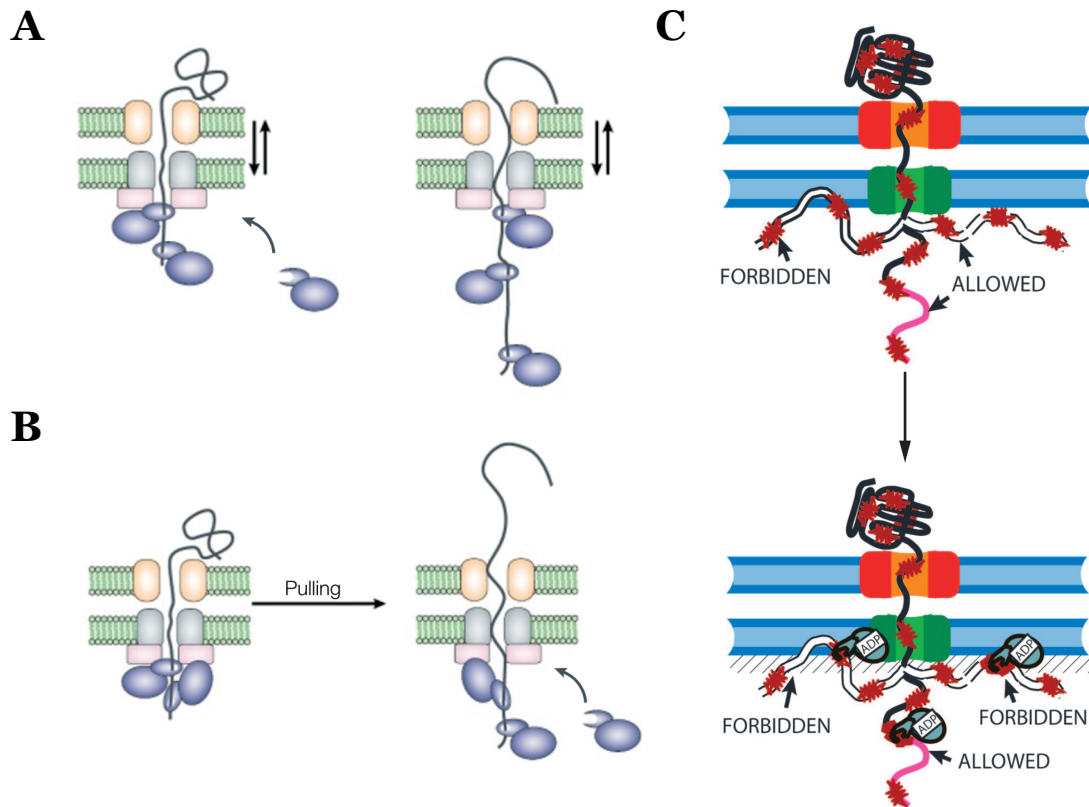


Figure 4.3: Schematic illustration of the different existing translocation mechanisms: **(A)** the Brownian ratchet model, Figure adapted from [144]; **(B)** the power stroke model, Figure from [144]; **(C)** the entropic pulling model, Figure from [145].

also through a cooperative mechanism with the Tim44 complex. Once the incoming protein is captured by Hsp70, a conformational switch of the chaperone protein follows the hydrolysis of ATP and induces a power stroke on the protein. Contrarily to the Brownian ratchet, there is an effective mechanical force exerted on the protein, whereas thermal fluctuations are not considered to explain the translocation.

These two models are based on very different arguments but they are not mutually incompatible in the description of biological processes, in the sense that both mechanisms might simultaneously play a role to drive the translocation [144].

### 4.2.3 Entropic pulling model

The third model can be seen as a reconciliation of both previous leading ideas: the entropic pulling model [145, 147] conserves the idea of an effective pulling force exerted on the substrate (as in the power stroke model, although it is not a purely mechanical force), combining it with the idea of a motion induced by thermal fluctuations (Brownian ratchet model).

During translocation, the portion of the protein in the mitochondrial matrix or in the ER lumen can move freely, always with the physical constraint that it cannot move beyond the membrane. It results in a set of allowed and forbidden conformations (Figure 4.3C, top). The binding of Hsp70 leads to possible steric clash between Hsp70 and the membrane surrounding the pore, thus reducing the number of reachable configurations (Figure 4.3C, bottom). Thermodynamically, it corresponds to a reduction of the entropy of the system induced by the binding of Hsp70 which is associated to a force exerted on the substrate that drives the translocation inside the mitochondria or the ER.

More precisely, the entropy variation associated to the reduction of the number of reachable configurations of the protein has been analytically solved [145]. The free energy profile  $f(n_K)$  is given by:

$$f(n_K) = \frac{3}{2} k_B T \frac{R^2}{b^2 n_K} \quad (4.1)$$

where  $R$  is the radius of the particle (here Hsp70) that binds and causes the entropy decrease,  $b$  is the length of a Kuhn segment of the polymer and  $n_K$  is the number of Kuhn segments between the pore and the binding site.

Beyond the proportionality constants, the essential point of this formula is that the free energy is inversely proportional to  $n_K$  and thus the distance along the protein between the pore and the binding site. The way this formula, and more generally the entropic pulling mechanism, enters into our model is presented in section 4.3.

Although the entropic pulling itself is based on mechanical and statistical arguments, the way it is involved in protein translocation in biological systems reveals the translocation as an active process, in which the nucleotide-dependant and allosteric nature of Hsp70 plays



### 4.3 A kinetic model for protein translocation

---

a key role. The associated mechanism of ultra-affinity significantly enhances the affinity of the chaperone for its substrate [148, 149]. When bound to ATP, the SBD of Hsp70 is in an open conformation with large binding and unbinding rates. In contrast, when bound to ADP, the binding and unbinding rates are low since the SBD is in a closed conformation. Out of equilibrium, there is a mix of the different timescales relative to ATP and ADP bound states leading to an effective affinity, precisely known as ultra-affinity, that is not reachable in any equilibrium configuration and that might be order of magnitudes larger than the equilibrium one with ATP or ADP. Ultra-affinity furthermore results in an acceleration of the translocation rate, by increasing the time Hsp70s are bound on the substrate protein and consequently the pulling force exerted by Hsp70 on the translocating protein by entropic pulling. So far, most of the developments on the entropic pulling mechanism have been focused on a description at the level of a single Hsp70, on how the binding is converted into a force, but there is a lack in a more global and dynamical perspective of the translocation process.

In the next section, we aim at building a minimal model based on the entropic pulling model to reproduce the phenomenology of the translocation in order to reveal different inherent mechanisms. First we present the kinetic model that we built to study the dynamics of the protein translocation. Then, we show how different parameters affect the overall dynamics of the system and which are the possible consequences on the dynamical evolution of biological systems, also opening hypotheses on possible structural rearrangements of Hsp70s.

### 4.3 A kinetic model for protein translocation

We model a protein that is translocating through the pore of a membrane, either toward the ER lumen or the matrix of the mitochondrion. The model is sufficiently generic so that specific details relative to the distinction between mitochondria and ER are not taken into consideration.

The translocated proteins are of finite size and the insertion of the terminal at the beginning of the process is an important part to understand the translocation in its whole. However, our model focuses on a transient regime (leaving aside the possible "boundary effects") that operates at steady-state, that is when the translocation rate of the protein is constant over time. To that aim, we define a simulation box that contains a fixed number of binding sites (Figure 4.4A). In solution, chaperone proteins Hsp70s can be either in the ATP state or in the ADP state and can bind to an empty binding site. The possible transitions that can occur at the level of a single site ( $S_n$ , the  $n^{th}$  site from the pore) on the protein sequence are the following (Figure 4.4B) :

- $k_{on,H_T}^{S_n}$  and  $k_{on,H_D}^{S_n}$  [ $\mu M^{-1} s^{-1}$ ]: **Binding rates** of Hsp70 in the ATP and ADP state respectively;

## Chapter 4. Substrate Translocation induced by Entropic pulling

---

- $k_{off,H_T}^{S_n}$  and  $k_{off,H_D}^{S_n}$  [ $s^{-1}$ ]: **Unbinding rates** of Hsp70 in the ATP and ADP state respectively;
- $k_{ex,H_T \rightarrow H_D}^{S_n}$  and  $k_{ex,H_D \rightarrow H_T}^{S_n}$  [ $s^{-1}$ ]: **Exchange rates** (respectively ATP  $\rightarrow$  ADP and ADP  $\rightarrow$  ATP) on the  $n^{th}$  site of the protein sequence;
- $k_h^{S_n}$  and  $k_s^{S_n}$  [ $s^{-1}$ ]: **Hydrolysis and synthesis rates** on the  $n^{th}$  site of the protein sequence.

In addition, there are two rates that do not involve a single Hsp70 on a given site but the whole protein. For each non-empty sequence in which  $n_1$  denotes the index of the first occupied binding site on the protein sequence, the system can undergo either a forward or a backward translocation, whose rates are given by:

- $k_f^{(n_1)}$  [ $s^{-1}$ ]: **Forward translocation rate**. During the forward translocation, the index of the first non-empty binding site shifts from  $n_1$  to  $n_1 + 1$
- $k_b^{(n_1)}$  [ $s^{-1}$ ]: **Backward translocation rate** (retro-translocation). During the backward translocation, the index of the first non-empty binding site shifts from  $n_1$  to  $n_1 - 1$

The numerical values of the rates as well as experimental results that provide orders of magnitude for the different parameters introduced hereafter are presented and discussed in the Appendix (Table A.6, equations A.43 and A.44).

To build a more detailed approach of the model, we detail each of the main parameters and their transition rates. In particular, we expose how they are defined in accordance with the entropic pulling free energy (equation 4.1). The hypotheses considered to model the different rates are also discussed. Here, we first present the expression of all the rates, but for thermodynamic consistency the system must always satisfy detailed balance. It follows that all the rates cannot be chosen independently and their expression as a function of independent parameters is shown in the Appendix (Tables A.3 to A.5).

**Entropic pulling force:** To the entropic pulling free energy (equation 4.1) is associated a pulling force exerted on the translocating substrate. To explicitly report on this pulling force, the expression for the entropic pulling free energy is alternatively rewritten in the following way:

$$f(n_1) = F_1 \cdot \frac{x_1^2}{x_{n_1}} \quad (4.2)$$

where  $x_1$  is the position of the first binding site, the closest to the pore and  $x_{n_1}$  is the position of the  $n_1^{th}$  binding site (the first with a Hsp70 along the sequence). An expression for  $x_{n_1}$  cannot be considered without tackling the question of the distribution of the binding sites along

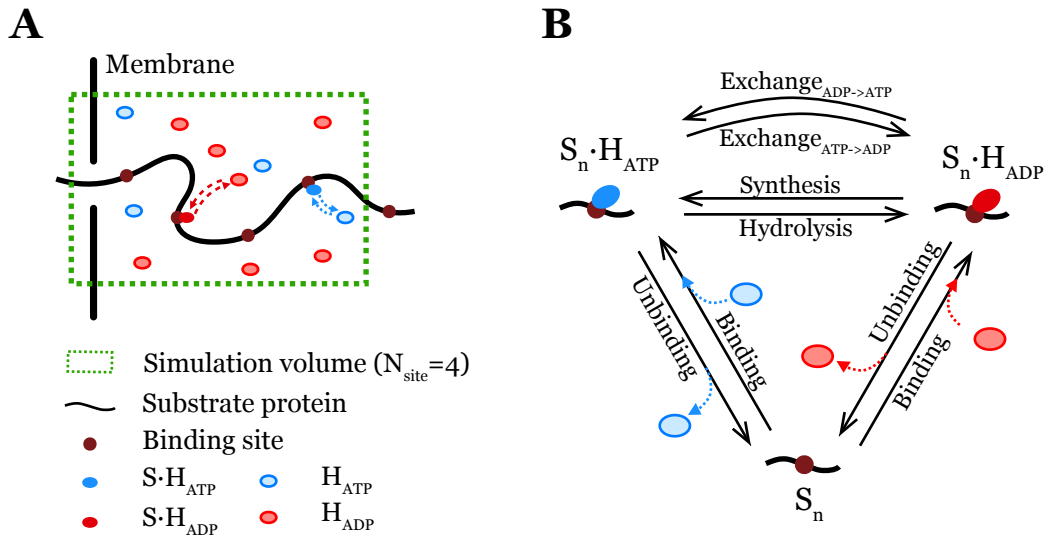


Figure 4.4: **(A)** A simulation volume (green dashed box) contains the pore and a portion of the translocating protein with exactly  $N_s \equiv N_{\text{site}}$  binding sites (brown circle; here  $N_s = 4$ ). The simulation volume contains Hsp70 in solution that can be either in an ADP or ATP states (respectively denoted by  $H_{\text{ADP}}$ , light red oval or  $H_{\text{ATP}}$ , light blue oval). These chaperone proteins can bind or unbind (dashed arrows) to any empty site  $S_n$  of the protein, leading to the states  $S_n \cdot H_{\text{ATP}}$  with dark red oval or  $S_n \cdot H_{\text{ATP}}$  with dark blue oval. **(B)** Possible transitions on a site  $S_n$  of the protein sequence.

## Chapter 4. Substrate Translocation induced by Entropic pulling

---

the protein sequence, which is far from being trivial. As a very first and rough description of the translocation, it can be considered that the binding can occur according to a quasi-continuous distribution, that is on every amino-acid of the protein sequence. However, experimental and structural studies of interactions between substrate proteins and Hsp70 reveal that the average distance statistically corresponds to 36 amino-acids (abbreviated by aa) [150, 151]. Our model neglects possible fluctuations around this average value and we thus fixed  $\Delta x = 35 \text{ aa} = 10.5 \text{ nm}$ . The position of the first binding site after entering the mitochondrion or ER lumen was found to be between 8 and 15 aa [145] and we fixed  $x_1 = 8 \text{ aa} = 2.4 \text{ nm}$ . The distance between the binding site  $n_1$  and the pore is thus defined by:

$$x_{n_1} = x_1 + (n_1 - 1) \cdot \Delta x \quad (4.3)$$

Going back to equation 4.2 and its interpretation, the parameter  $F_1$  is a measure of the pulling force exerted on the protein upon Hsp70 binding on the first site. For this reason, we often refer to  $F_1$  as the entropic pulling force parameter. The comparison between expressions 4.1 and 4.2 gives an expression for  $F_1$  as a function of the parameters of the system:

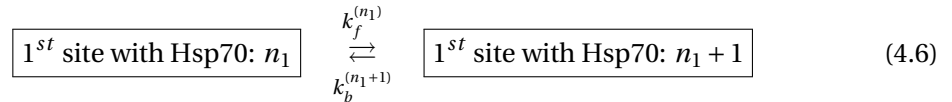
$$F_1 = \frac{3}{2} k_B T \frac{R^2}{bx_1^2} \quad (4.4)$$

The larger the parameter  $F_1$ , the stronger is the pulling force exerted on the translocating protein, on the first site but also by extension on the further ones. Its numerical value is expressed in the Appendix, equation A.44.

**Forward and backward translocation:** The entropic pulling force parameter  $F_1$  tunes the forward and backward translocation rates by favouring the forward translocation event with respect to the backward one. The ratio between both is related to the free energy difference between two consecutive states:

$$\frac{k_f^{(n_1)}}{k_b^{(n_1+1)}} = \exp\left(\frac{f(n_1) - f(n_1 + 1)}{k_B T}\right) \quad (4.5)$$

where  $k_f^{(n_1)}$  and  $k_b^{(n_1+1)}$  are the rates for two reverse transitions, as illustrated in equation 4.6.



Whereas  $F_1$  tunes the asymmetry between the rates, and by extension the strength of the

### 4.3 A kinetic model for protein translocation

imposed directionality, there is for now no parameter that mirrors the speed of the translocation process, irrespective of its directionality. A parameter  $k_0$  is introduced to give an explicit expression for both rates  $k_f$  and  $k_b$ :

$$k_f^{(n_1)} = k_0 \exp\left(\frac{f(n_1) - f(n_1 + 1)}{k_B T}\right) \quad (4.7)$$

$$k_b^{(n_1+1)} = k_0 \quad (4.8)$$

As a preliminary remark, the choice to include the whole contribution of the entropic pulling into the forward translocation rate is arbitrary and many other choices would have been thermodynamically consistent. Our choice was motivated by an alternative but equivalent formulation of the entropic pulling mechanism: having in mind the kinetic theory of gases, the successive collisions of the chaperone with the membrane and pore can be seen as generating a repulsive force exerted on the membrane[152]. Within this perspective, it was decided to mirror it in an increase of the forward translocation rate, rather than a decrease of the backward translocation rate. The latter is constant and interpreted as the consequence of thermal fluctuations, independent on the position of the bound chaperone on the substrate. The parameter  $k_0$  is referred to be a diffusion rate, a proxy for the diffusion constant of the protein through the pore. It reflects the speed of the translocation but does not directly impact the directionality of the translocation by favouring one direction or the other. Its expected numerical value is discussed in the appendix (equation A.44).

**Binding and unbinding rates:** The free energy gradient associated to the pulling force exerted on the substrate unavoidably has a thermodynamic cost that has to be "paid" on other transitions in the system. Considering the cycle including backward and forward translocation and crossed in two reverse directions (4.9 and 4.10), detailed balance condition in the form of equation 4.11 has to be satisfied.

$$\text{Binding on site } n_1 \longrightarrow \text{forward translocation} \longrightarrow \text{Unbinding from site } n_1 + 1 \quad (4.9)$$

$$\text{Unbinding from site } n_1 \longleftarrow \text{backward translocation} \longleftarrow \text{Binding on site } n_1 + 1 \quad (4.10)$$

$$\frac{k_{on,H_X}^{S_{n_1}} \cdot k_f^{S_{n_1}} \cdot k_{off,H_X}^{S_{n_1+1}}}{k_{off,H_X}^{S_{n_1}} \cdot k_b^{S_{n_1+1}} \cdot k_{on,H_X}^{S_{n_1+1}}} = 1 \quad (4.11)$$

where  $H_X$  can be either  $H_T$  or  $H_D$ .

Here again, for similar arguments than the one described in the case of forward and backward translocation, we assume that the whole contribution of the energy cost associated to entropic

## Chapter 4. Substrate Translocation induced by Entropic pulling

---

pulling free energy is paid on the binding, whereas the unbinding rates  $k_{off,H_X}^{S_{n1}}$  only depend on chemical affinities of either  $H_T$  or  $H_D$  for the substrate. Therefore, the binding and unbinding rates are defined as follows:

$$k_{on,H_X}^{S_n} = k_{on,H_X}^{(0)} \exp\left(-\frac{f(n)}{k_B T}\right) \quad (4.12)$$

$$k_{off,H_X}^{S_n} = k_{off,H_X}^{(0)} \quad (4.13)$$

**Hydrolysis and synthesis:** The model does not involve explicitly the J-proteins and their interactions with chaperone proteins (mtHsp70 or BiP). In itself, the whole mechanism associated to the acceleration of ATPase activity is very complex, involving the interplay between subunits Tim14 (Pam18) and Tim16 (Pam16) of the TIM23 complex [153] in the mitochondria. In the ER, a similar acceleration is observed, mediated by the Sec63 subunit of the translocation channel [154].

We only keep its essential feature in our model, that is the resulting acceleration of the ATPase activity. It is made in a very simplistic way, by introducing a numerical factor  $\lambda$  quantifying the acceleration of both hydrolysis and synthesis rates when bound close to the membrane, on the first binding site after its insertion in the mitochondrial matrix or ER lumen. The hydrolysis ( $k_h$ ) and synthesis ( $k_s$ ) rates are defined accordingly:

$$k_h^{S_1} = \lambda k_h^S \quad k_h^{S_{n>1}} = k_h^S \quad (4.14)$$

$$k_s^{(1)} = \lambda k_s^S \quad k_s^{S_{n>1}} = k_s^S \quad (4.15)$$

**Exchange:** The exchange rates only depend on the affinity of ATP and ADP to chaperone proteins Hsp70 when bound to the substrate. In particular, they do not depend on the binding position. Exchange rates are given by:

$$k_{ex}^{S_n \cdot H_D \rightarrow S_n \cdot H_T} = k_{-D}^S \frac{\alpha k_{+T}^S}{k_{+D}^S + \alpha k_{+T}^S} \quad (4.16)$$

$$k_{ex}^{S_n \cdot H_T \rightarrow S_n \cdot H_D} = k_{-T}^S \frac{k_{+D}^S}{k_{+D}^S + \alpha k_{+T}^S} \quad (4.17)$$

where, as a reminder,  $\alpha$  denotes the ratio  $[ATP]/[ADP]$ .

**Boundary of the simulation volume:** When the protein undergoes a forward translocation, the site  $N_s$  escapes from the simulation volume and the possibly bound Hsp70 is no longer considered. On the reverse, when the protein is backward translocated, the question of the incoming binding sites (i.e. site  $N_s$  after the translocation) arises. We hypothesise that the

### 4.3 A kinetic model for protein translocation

site entering the simulation volume is always empty. Alternatively, it could be in the state  $S_{N_s} \cdot H_T$  or  $S_{N_s} \cdot H_D$ , with a distribution that follows the one of a protein in solution (without considering any translocation). In the results part, it is shown that this assumption does not affect the overall dynamics of the system.

**Hsp70 in solution:** It is assumed that Hsp70 in solution are at fixed concentrations, either thanks to a chemostat or by hypothetically reaching equilibrium with the surrounding Hsp70 in solution sufficiently fast. Consequently, the binding and unbinding events do not effectively affect the concentrations of free chaperones bound to ATP and ADP, respectively  $[H_T]$  and  $[H_D]$ .

Their concentrations are expressed by considering a simple two-state system with  $H_T$  and  $H_D$  at steady state, leading to equation 4.18 and 4.19.

$$[H_T] = [H]_{tot} \frac{k_{ex}^{H_D \rightarrow H_T} + k_s}{(k_{ex}^{H_D \rightarrow H_T} + k_s) + (k_{ex}^{H_T \rightarrow H_D} + k_h)} \quad (4.18)$$

$$[H_D] = [H]_{tot} \frac{k_{ex}^{H_T \rightarrow H_D} + k_h}{(k_{ex}^{H_D \rightarrow H_T} + k_s) + (k_{ex}^{H_T \rightarrow H_D} + k_h)} \quad (4.19)$$

where, similarly to the substrate-bound rates,

$$k_{ex}^{H_D \rightarrow H_T} = k_{-D} \frac{\alpha k_{+T}}{k_{+D} + \alpha k_{+T}} \quad (4.20)$$

$$k_{ex}^{H_T \rightarrow H_D} = k_{-T} \frac{k_{+D}}{k_{+D} + \alpha k_{+T}} \quad (4.21)$$

Again, all the parameters are not independent the one from the other and their expressions after imposing detailed balance conditions are given in the Appendix, Tables A.3 to A.5.

#### 4.3.1 Mathematical formulation of the model

So far, we have presented in details the different transitions involving a single chaperone protein, as well as the forward and backward translocations the substrate protein can undergo. In the present section, we introduce the formalism used to solve the system at steady state and compute the main quantity of interest to describe the dynamics of translocation, the net translocation rate of the protein through the pore.

The portion of the protein in the simulation volume (figure 4.4A) is defined by its sequence of  $N_s \equiv N_{\text{site}}$  binding sites, as a vector  $\vec{s}$  of size  $N_s$ :  $\vec{s} := (\sigma_1, \dots, \sigma_{N_s})$ . Each  $\sigma_i$  is the state of the binding site  $i$ , that is  $\sigma_i \in \{S_i \cdot H_T, S_i \cdot H_D, S_i\}$ . Therefore, there are  $3^{N_s}$  sequences which are combined in a single vector  $\vec{\Psi} := (\vec{s}_1, \dots, \vec{s}_{3^{N_s}})$  that contains all the possible configurations of

the protein in the simulation volume.

For each sequence  $\vec{s}_j \in \vec{\Psi}$ , the possible transitions are either localised on one site (transitions from Fig 4.4B) or involve the whole protein with either the backward or forward translocation. The dynamics of the system can be written as a set of linear differential equations, in the form

$$\frac{d}{dt} \vec{P}(\vec{\Psi}) = \underline{\underline{M_k}} \cdot \vec{P}(\vec{\Psi}) \stackrel{\text{steady-state}}{=} 0 \quad (4.22)$$

where  $\underline{\underline{M_k}}$  is a  $3^{N_s} \cdot 3^{N_s}$  matrix, with elements corresponding to the total transition rates between two sequences (equation 4.23). As an example, the full matrix is written in the appendix (equations A.46 and A.47) for the simple case of a system with two sites, corresponding to a  $9 \times 9$  matrix.

$$M_k^{(i,j)} = \sum_{n=1}^{N_s} \left[ \underbrace{k_{h/s}^{(s_j^{(n)} \rightarrow s_i^{(n)})}}_{\text{Hydrolysis or synthesis}} + \underbrace{k_{\text{ex}}^{(s_j^{(n)} \rightarrow s_i^{(n)})}}_{\text{Exchange}} + \underbrace{\tilde{k}_{\text{off/on}}^{(s_j^{(n)} \rightarrow s_i^{(n)})}}_{(\text{Un-})\text{Binding}} \right] + \underbrace{k_{f/b}^{(\vec{s}_j \rightarrow \vec{s}_i)}}_{\text{Translocation}} \quad (4.23)$$

where the notation  $\tilde{k}$  for the (un-)binding rates aims to avoid the confusion between the binding rates  $k_{on,H_X} [\mu\text{M}^{-1}\text{s}^{-1}]$  and the effective binding rate  $\tilde{k}_{on,H_X} [\text{s}^{-1}] = [H_X] \cdot k_{on,H_X}$ . Of course,  $\tilde{k}_{off,H_X} = k_{off,H_X}$ .

The time derivative  $d\vec{\Psi}/dt$  is imposed to be zero at steady-state (s-s). Steady-state means that the probability of each sequence is constant in time, even if the protein is translocating at a constant rate. Equation 4.22 is then solved by computing the kernel of the matrix  $\underline{\underline{M_k}}$ , giving after appropriate normalisation the steady-state probability of each sequence ( $P_{s-s}(\vec{\Psi})$ ).

The main quantity of interest that will be used to describe the system and in particular its dynamics is the net translocation rate,  $R$  (equation 4.24). This quantity directly follows from the sequence probability distribution  $P_{s-s}(\vec{\Psi})$ :

$$R = \sum_{\vec{s}_i \in \vec{\Psi}} P_{s-s}(\vec{s}_i) \cdot \left[ k_f^{(n_1(\vec{s}_i))} - k_b^{(n_1(\vec{s}_i))} \right] \quad (4.24)$$

where  $k_f^{(n_1(\vec{s}_i))}$  and  $k_b^{(n_1(\vec{s}_i))}$  refer to the equations 4.7 and 4.8 applied to the first bound site associated to protein sequence  $\vec{s}_i$ , that is  $n_1(\vec{s}_i)$ .

### 4.3.2 An alternative stochastic approach to solve the system

At the very beginning of the project, we started to simulate the system using a Kinetic Monte Carlo simulation (Gillepsie algorithm), before moving to the above-presented method consisting in solving a linear system of equations. A few simulations were performed with the exactly same definition of the possible rates and the same hypotheses than the one presented so far.



## 4.4 Translocation as a balance between force and binding frequency

---

For one figure, we show in the appendix that indeed the two methods provide the same results (Figure 4.8 in the main text reproduced in the Appendix, Figure A.2). However, this method was not used for a long time and thus is not extensively described.

### 4.4 Translocation as a balance between force and binding frequency

In the first part of this project, we focus our study on three parameters:  $\alpha$  and the two main parameters that characterise the entropic pulling,  $k_0$  and  $F_1$ . We first show how the translocation performance is increased upon uptake of energy into the system. Then we discuss how the "strength" of the entropic pulling force represented by  $F_1$  and its "velocity" via the parameter  $k_0$  (a proxy for the diffusion constant of the protein through the pore) modify the observed net translocation rate  $R$ .

We highlight an optimal regime, suggesting a balance between on the one hand the imposition of a directionality to the translocation process and on the other hand the binding frequency of Hsp70 on the substrate, indeed the key-element necessary for translocation.

In a second part, motivated by the interpretation of this directionality-affinity balance, we study the impact of an hypothetical Hsp70 dimerization, in which two chaperones consecutively bind the one on the other. We show that the effect of a possible dimerisation is significant and might be of high relevance in some specific biological conditions. These results reflect an adaptability of the system to external conditions, for instance the presence of an obstacle which possibly reduces the intrinsic diffusion rate of the substrate through the channel.

Before turning to the numerical results of the model and the description of how the different parameters affect the translocation rate of the protein, we briefly describe a few preliminary results that are not directly related to the physics of the model but that deal with the choices and hypotheses that are made in the implementation and interpretation of it.

As illustrated in figure 4.4 A, the simulated box is of finite size, with a fixed number of binding sites. Based on the actual understanding of the translocation mechanism by entropic pulling, it is expected that the most important contribution to the pulling force comes from chemical transitions taking place close to the membrane, suggesting that it should not be necessary to simulate a too large number of sites. We simulated the evolution of the translocation rate as a function of  $[ATP]/[ADP]$ , for  $N_s = 2, 4, 6$ , and an example of the results is shown in figure 4.5A, which is representative of all the observations that were made with different set of parameters. Close to equilibrium, there is a visible difference between  $N_s = 2$  and  $N_s \geq 4$ , whereas moving far from equilibrium, the difference gets negligible and all the three curves overlap. Based on these observations, from now on, all the simulations are performed with  $N_s = 4$ . Beyond technical considerations, this first observation interestingly provides a length scale of the portion of the protein on which the entropic pulling mechanism is effectively taking place and leading to the translocation of the protein, either inside the mitochondria or ER: the order of

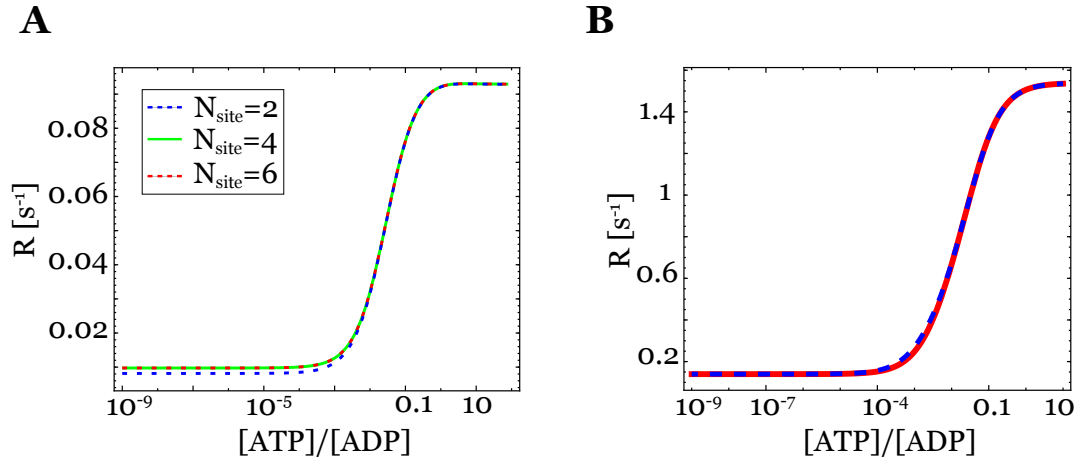


Figure 4.5: **(A)** Translocation rate as a function of  $[ATP]/[ADP]$ . The simulation is made with different number of binding sites in the simulation volume:  $N_s = 2$  (dashed blue),  $N_s = 4$  (green) and  $N_s = 6$  (dashed red).  $F_1 = 6\text{pN}$ ,  $k_0 = 10^{-2}\text{s}^{-1}$ . **(B)** Translocation rate as a function of  $[ATP]/[ADP]$ . The red curve is computed with the assumption of an empty incoming binding site into the simulation volume during a retro translocation event. Alternatively, for the dashed blue curve, it is assumed that the distribution of incoming binding site is representative of the steady-state distribution on a protein in solution, without translocation mechanism involved.  $F_1 = 6\text{pN}$ ,  $k_0 = 10^{-2}\text{s}^{-1}$ .

magnitude is around a hundred amino acids.

Second, we hypothesize in the model that when there is a retro-translocation event, the binding site that enters into the simulation volume is empty. It might seem more realistic to consider a probability distribution that follows the distribution that would be observed for a protein in solution, independently of any translocation event. The results associated to both possible implementations are shown in figure 4.5B. Clearly, although there is a tiny difference between both curves around  $[ATP]/[ADP] \approx 10^{-3}$ , it is reasonable to consider that our assumption to model an empty incoming site when retro translocating is correct, in the sense that it does not affect at all the effective translocation rate  $R$ . Moreover, it confirms also the conclusion from figure 4.5A, according to which the translocation is mediated by chemical transitions taking place essentially on binding sites closer to the membrane than the fourth one, thus validating the choice of parameter  $N_s = 4$ .

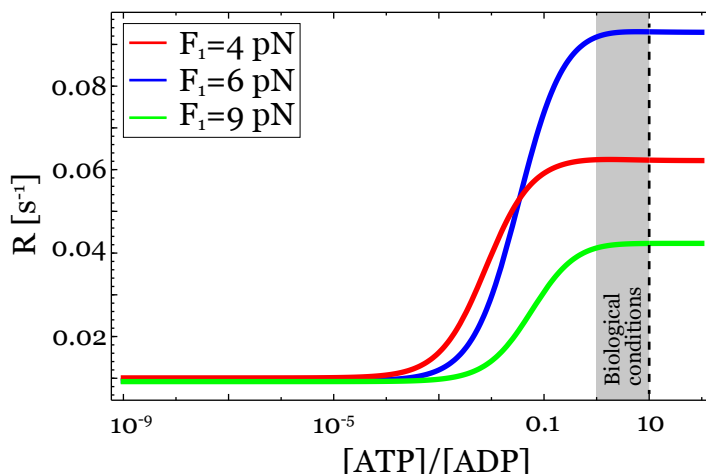


Figure 4.6: Translocation rate as a function of  $[ATP]/[ADP]$  for different values  $F_1 = 4, 6$  and  $9$  pN. The region corresponding to typical biological conditions is represented with a grey rectangle. The vertical dashed line illustrates the value  $[ATP]/[ADP] = 10$  that is used in further simulations.  $k_0 = 10^{-2} \text{s}^{-1}$ .

## 4.5 Pulling stronger, but less frequently

As for all active transport processes, energy is required to efficiently drive the translocation of proteins, to accelerate the transport with respect to what it would be in equilibrium conditions. We start by showing how the translocation rate  $R$  depends on the available energy, through the ratio  $[ATP]/[ADP]$ . Figure 4.6 shows how the translocation rate increases when energy is brought to the system. At equilibrium, the dynamics is driven by the successive binding and translocation events (as in the Brownian ratchet model), the binding of another Hsp70 on the first site preventing the retro-translocation. The uptake of energy favours the cycle that involves the hydrolysis of  $H_{ATP}$  close to the membrane before the exchange  $H_{ADP} \rightarrow ADP$  after translocation. For a given set of parameters, a plateau is reached around  $[ATP]/[ADP] \approx 0.1$ , which is slightly below the typical biological conditions,  $1 \lesssim [ATP]/[ADP] \lesssim 10$  [155, 156]. No precise experimental result was found for the ratio  $[ATP]/[ADP]$  in the ER lumen, but we expect it to be in order of magnitude equal to the one in the cytosol and mitochondria.

To study the dynamics of translocation, we focus on the two main parameters that tune the translocation rates  $k_f$  and  $k_b$ . At a first stage, the diffusion rate  $k_0$  is discussed, before turning to the entropic pulling force parameter  $F_1$  at a second stage.

First of all, as the translocation rate  $R$  is proportional to  $k_0$ , it is more convenient to define the relative translocation rate  $R/k_0$  in order to get rid of this explicit dependence on  $k_0$  and

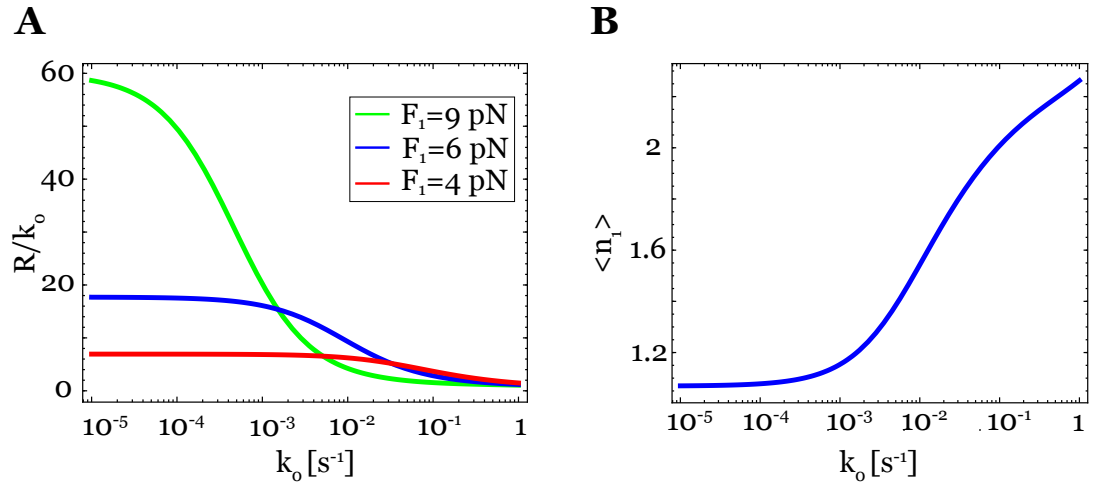


Figure 4.7: **(A)** Translocation rate as a function of  $k_0$  for  $F_1 = 4, 6, 9$  pN and  $[ATP]/[ADP] = 10$ . **(B)** The position of the Hsp70 chaperone the closer to the membrane (here the position refers to the discrete index  $n_1$  of the binding site  $1, 2, \dots, N_S$ ) is averaged according to the steady-state distribution of all sequences. The average  $\langle n_1 \rangle$  is plotted as a function of  $k_0$ , with  $F_1 = 6$  pN and  $[ATP]/[ADP] = 10$ .

thus isolate the hidden interplay between  $k_0$  and all the other transition rates. We show in Figure 4.7A how the relative translocation rate  $R/k_0$  depends on the diffusion rate for different entropic pulling forces  $F_1$ . The decreasing behaviour can be easily understood as a competition between on the one hand the diffusion rate and on the other hand the binding rate. When the diffusion rate  $k_0$  is large, the translocation associated to the presence of a chaperone protein on the second (or further) binding site is more probable to happen, well before the binding of another Hsp70 on the first binding site. As a consequence, the system does not perform optimally since the closer to the membrane is the Hsp70, the stronger the entropic pulling. The interpretation of the decrease is supported by figure 4.7B, which illustrates the average position of the first bound site (i.e. where is the closer chaperone with respect to the pore), as a function of  $k_0$ . Whereas it is predominantly on the first site for low  $k_0$  (average index  $\langle n_1 \rangle < 1.2$ ), it gets further when  $k_0$  is increased (average index  $\langle n_1 \rangle > 2$ ) because the translocation occurs faster, reducing the time in which sequences with  $n_1 = 1$  are observed, and thus explaining the decrease of the relative translocation rate  $R/k_0$ .

In addition to the diffusion rate  $k_0$  is the parameter  $F_1$  that characterises the exerted pulling force on the incoming protein.

Figure 4.8A shows how the relative translocation rate  $R$  depends on the parameter  $F_1$ . Independently of the value of  $k_0$ , it is observed that the translocation rate  $R$  has a non monotonic

behaviour as a function of  $F_1$  and reaches a local maximum. In the first regime, that is for low values of  $F_1$ , increasing the entropic pulling force improves the translocation, by pulling stronger on the incoming protein, through a stronger asymmetry between inward and outward translocation rates. However, as illustrated by equation 4.12 and the associated discussion, the symmetry breaking has a cost that is paid on the binding rates of the chaperone proteins in solutions to the incoming polypeptide. Thus when  $F_1$  is too large, the system moves to a second regime in which it is not favourable to increase  $F_1$ , because the cost on the binding is too high and is not compensated by the gain on the translocation directionality. The transition value of  $F_1$  associated to local maximum is denoted by  $F_1^*$  and an illustration is given with the vertical blue dashed line in figure 4.8A for the parameter  $k_0 = 10^{-3}$ . Obviously, there is no unique value of  $F_1^*$  but it depends on  $k_0$  and the dependence is illustrated in figure 4.8B. The exhibited decrease of  $F_1^*$  as a function of  $k_0$  can be interpreted in the following way: when the diffusion rate is decreased, the translocation events (in both directions, forward and backward) are less and less frequent. In that case, it is optimal for the system to translocate in the good direction, by imposing a stronger asymmetry between inward and outward translocation rates, respectively  $k_f$  and  $k_b$  or, alternatively formulated, to pull stronger on the incoming substrate. This holds despite the associated energy cost on the binding. In the reverse, for large  $k_0$  and a high translocation velocity, the systems works optimally with a smaller value of  $F_1$ : binding events are more frequent on the sites left empty after translocation, although the fluctuations between forward and backward are more important (because of the weaker asymmetry between  $k_f$  and  $k_b$ ).

Clearly, there is a strong interconnection between on the one hand the translocation rate  $R$  that depends on  $F_1$  and the force exerted on the substrate, in average, during the steady-state of the translocation mechanism. The average force  $\langle F \rangle$  is defined as:

$$\langle F \rangle = \sum_{\{\sigma\}} P(\sigma) \cdot F(\sigma) \quad (4.25)$$

where  $F(\sigma)$  is the entropic pulling force, depending on the position of the first Hsp70 in the corresponding protein sequence  $\sigma$ .

$R$  and  $\langle F \rangle$  are two ways of quantifying the mechanism of translocation, it is worth to compare both, and in particular with in mind the perspective of the optimal regime discussed in figure 4.8, with respect to  $R$ . On the same plot are shown  $R$  and  $\langle F \rangle$ , for two values of  $k_0$  (figure 4.9). There is a non-negligible shift in the optimal value of  $F_1$  associated to  $\langle F \rangle$  and  $F_1^*$  defined for the translocation rate  $R$ , the latter being slightly larger. However, these are quantitative differences that are expected not to drastically change the qualitative description and understanding of the translocation mechanism. Therefore, we will always be using the translocation rate  $R$  to quantify the ability of the system to translocate a substrate protein, mostly leaving aside the average force  $\langle F \rangle$ .

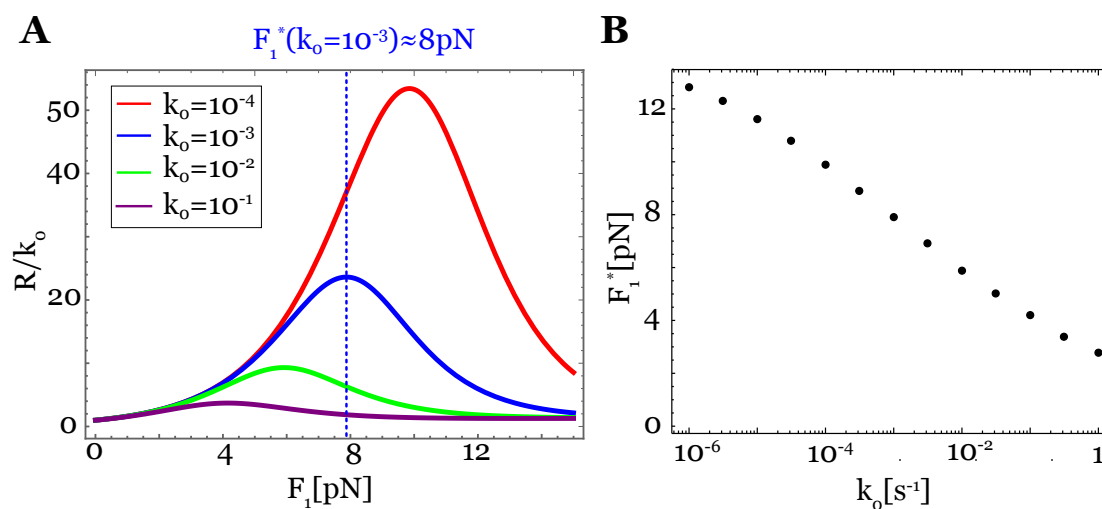


Figure 4.8: (A) Relative translocation rate  $R/k_0$  as a function of  $F_1$  for  $k_0 = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$  and  $[ATP]/[ADP] = 10$ . The vertical dashed blue line represents the optimal value of  $F_1$ , denoted by  $F_1^*$  in the case  $k_0 = 10^{-3} s^{-1}$ . (B) Optimal value  $F_1^*$  plotted for a set of parameters  $k_0$ .  $[ATP]/[ADP] = 10$ .

Following the discussion and interpretation of figure 4.8B, a natural question arises: if the system works optimally by "pulling stronger" on the protein when the intrinsic diffusion rate  $k_0$  is small, what if the system is able to bind a second Hsp70 on a same site, exerting an even stronger pulling force on the incoming protein? The answer to this question is developed in section 4.8 after a presentation of the Hsp70 oligomerization, both from physical and biological perspectives

## 4.6 Hsp70 oligomerisation in protein folding

The configuration in which a Hsp70 is bound to another has been extensively observed and described, both *in vivo* and *in vitro*, and for various Hsp70 family members [157, 158]. These works all tend to show that dimerization and higher order assemblies might be a general property of Hsp70. These observations have been mostly performed in experiments where Hsp70 assist the folding of non-native proteins. In this case, it has been observed that the ADP-bound form of Hsp70 (as well as the free form, essentially neglected here) are prone to oligomerize. However, the specific case of Hsp70 dimerization within the framework of protein import into either the mitochondria or the ER has not been reported in the literature. There is no experimental evidence for such a conformation in that precise process but, importantly, there is also no evidence for its impossibility, for any biological, biochemical or physical

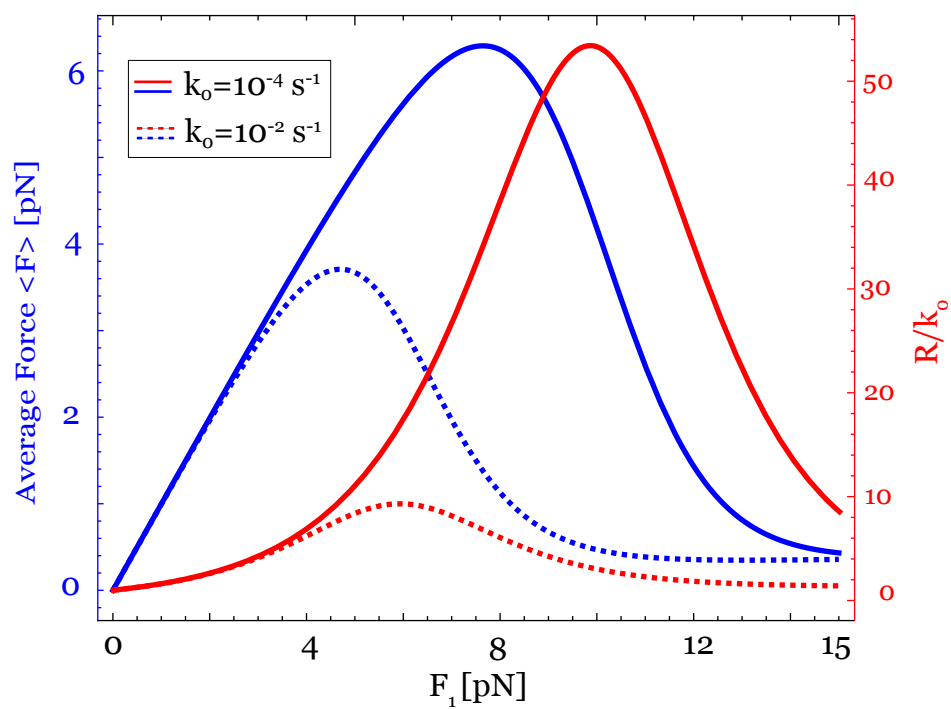


Figure 4.9: Red: Relative translocation rate  $R/k_0$  as a function of  $F_1$ . Blue: Average force  $\langle F \rangle$  as a function of  $F_1$ . The comparison between these two quantities is made for  $k_0 = 10^{-4} \text{ s}^{-1}$  (continuous lines) and  $k_0 = 10^{-2} \text{ s}^{-1}$  (dashed lines).  $[ATP]/[ADP] = 10$ .

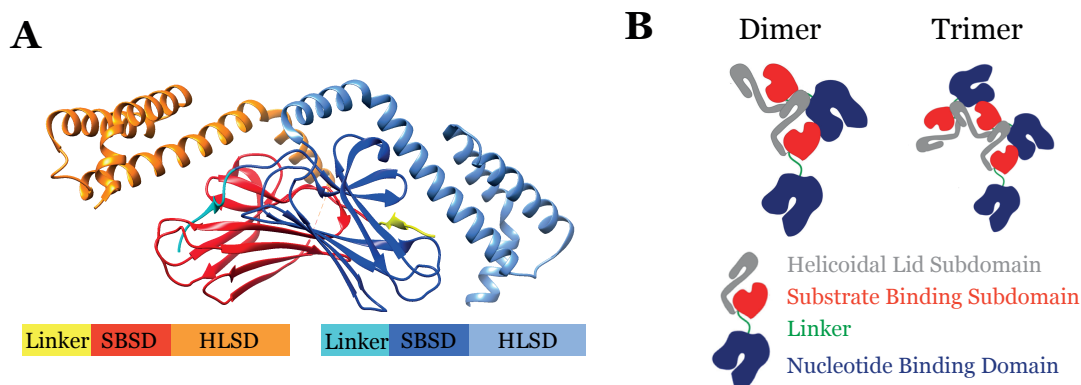


Figure 4.10: Dimer structure of the Substrate Binding Domain (A) PDB structure 4R5J from Leu et al. [161]. The structure shows two dimerized SBDs of Hsp70 (yellow-red-orange and shades of blue) with their different subdomains (Linker, Substrate Binding Subdomain (SBS) and Helicoidal Lid Subdomain (HLS)). (B) Schematic representation of two Hsp70 oligomers: dimer (left) and trimer (right). Figure from [157].

arguments. Thus, we extend the model to a possible dimeric state, in which a second Hsp70 is bound on the initially bound chaperone protein.

The allosteric transformation between ADP and ATP bound states has a strong effect on the different possible dimer states. On the one hand, the oligomerization is mediated by the linker between the NBD and SBD and, on the other hand, the linker is exposed in the ADP state but "inserted" in the NBD in the ATP bound state [159]. These are simple arguments that are in favour of a strong dominance of monomers in the ATP-bound state whereas higher order oligomers are observed in the ADP state (schematically illustrated in figure 4.10B) [157, 158, 160].

## 4.7 Extended model with Hsp70 dimerization

Two additional states are added into the initial model (figure 4.11, dashed grey rectangle) which are the  $S_n \cdot H_{ADP} \cdot H_{ADP}$  and  $S_n \cdot H_{ADP} \cdot H_{ATP}$ . Theoretically, there are few other possible transitions that might be included into the updated model, but we always try to keep it as simple as possible despite the addition of two states. The hypotheses that were made to follow this approach are double. First, we neglect the binding (and unbinding) of a Hsp70 dimer that could have been formed in solution, which also means that the fixed concentrations  $[H_{ATP}]$  and  $[H_{ADP}]$  remain unchanged with respect to the initial formulation of the model (Equations



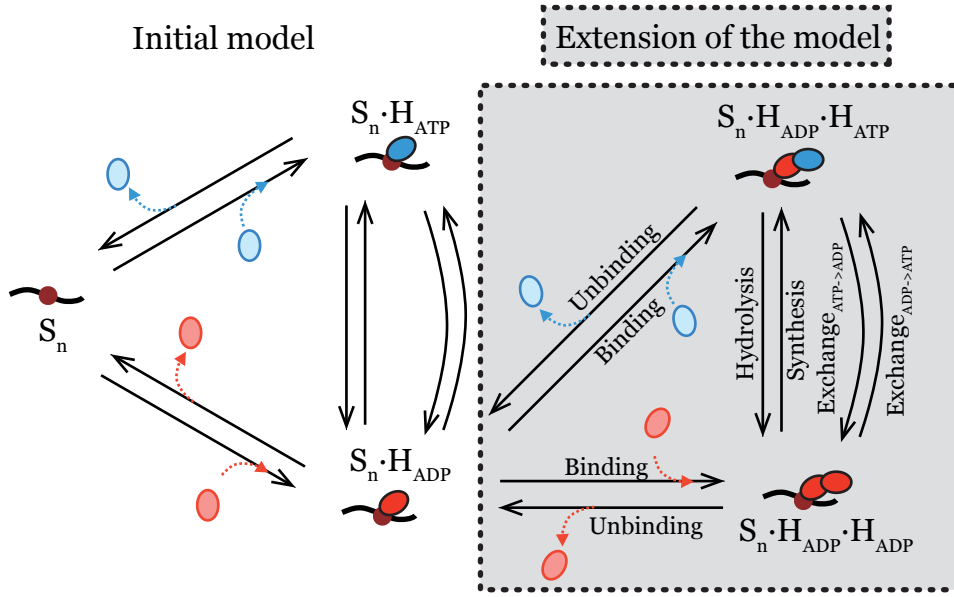


Figure 4.11: Extension of the model (from Figure 4.4B). Two states are added (grey rectangle) which correspond to a dimer structure of Hsp70: On the state  $S_n \cdot H_{ADP}$  can bind a second Hsp70 either in ATP state or ADP state, leading respectively to  $S_n \cdot H_{ADP} \cdot H_{ATP}$  (right top) and  $S_n \cdot H_{ADP} \cdot H_{ADP}$  (right bottom) states.

4.18 and 4.19). Second, we neglect the possibility for the first bound chaperone (in the ADP state) to go through synthesis or exchange. Biologically, it is not clear whether this can be considered and with which kinetics, but since it would lead to states that are by assumption not part of the kinetic model ( $S_n \cdot H_{ATP} \cdot H_{ADP}$  and  $S_n \cdot H_{ATP} \cdot H_{ATP}$ ), this possibility is left aside. Indeed it is likely that the presence of the second Hsp70 drastically reduce the exchange rates  $ADP \rightarrow ATP$  on the first chaperone because the second one constitutes a steric obstacle to such an exchange process.

The free energy associated to a dimer binding has to be adapted, given that the size of the bound particle increases (eq. 4.1). To account for this increase, but keeping the definition of  $F_1$  unchanged for sake of results comparability, we introduce a parameter  $\gamma$  in equation 4.2 :

$$f_2(x) := \gamma F_1 \cdot \frac{x_1^2}{x} \quad (4.26)$$

The appropriate value of the parameter  $\gamma$  is not clearly defined and it will depend, among other

factors, on the spatial arrangement of the dimer. Since the free energy depends quadratically of the size of the bound chaperone(s) [145], as a first approximation, it is reasonable to assume that  $\gamma \in [2^{2/3}, 2^2]$  that is an intermediate value between purely linear and volumetric rearrangements. We found out that varying  $\gamma$  within this range of values does not affect the results qualitatively, in particular the evolution of the translocation rate, and thus all the results are performed with  $\gamma = 2$ . This choice has its part of arbitrariness and for more realistic results, it is possible that a parameter  $\gamma \approx 1.8$  sticks more with existing numerical simulations performed in the recent past [unpublished work]. Figure A.1 shows that in the case of fast translocation ( $k_0 = 10^{-2} \text{ s}^{-1}$ ), the difference between  $\gamma = 1.8$  and  $\gamma = 2$  is very weak. By slowing the translocation ( $k_0 = 10^{-5} \text{ s}^{-1}$ ), there is a more important difference between  $\gamma = 1.8$  and  $\gamma = 2$ , around approximatively 10%, but that does not change the qualitative interpretation of our simulations.

The rates related to the entropic pulling free energy profile (which are the forward translocation rates and the binding rates) have to be modified accordingly, including the factor  $\gamma$ .

The binding rate of the second Hsp70 depends on  $\gamma$  in the following way, in order to satisfy detailed balance :

$$k_{on,H_T}^{(n)} = \begin{cases} k_{on,H_T}^{(0)} \cdot \exp[-\beta f(x_n)] & \text{monomer} \\ k_{on,H_T}^{(0)} \cdot \exp[-\beta(f_2(x_n) - f(x_n))] & \text{dimer} \end{cases} \quad (4.27)$$

$$k_{off,H_T}^{(n)} = k_{off,H_T}^{(0)} \quad (4.28)$$

$$k_{on,H_D}^{(n)} = \begin{cases} k_{on,H_D}^{(0)} \cdot \exp[-\beta f(x_n)] & \text{monomer} \\ k_{on,H_T}^{(0)} \cdot \exp[-\beta(f_2(x_n) - f(x_n))] & \text{dimer} \end{cases} \quad (4.29)$$

$$k_{off,H_D}^{(n)} = k_{off,H_D}^{(0)} \quad (4.30)$$

Similarly, the inward and backward translocation rates are modified accordingly:

$$k_f^{(n)} = \begin{cases} k_0 \exp[-\beta(f(x_{n+1}) - f(x_n))] & \text{monomer} \\ k_0 \exp[-\beta(f_2(x_{n+1}) - f_2(x_n))] & \text{dimer} \end{cases} \quad (4.31)$$

$$k_b^{(n)} = k_0 \quad (4.32)$$

Finally, the exact same procedure is used to solve the steady state, through the resolution of a  $5^{N_s}$ -dimensional system of linear equations (similar to eq. 4.22). The new net translocation rate denoted by  $R^{(2)}$  is defined similarly to equation 4.24, that is:

$$R^{(2)} = \sum_{\vec{s}_i \in \tilde{\Psi}} P_{s-s}(\vec{s}_i) \cdot [k_f(\vec{s}_i) - k_b(\vec{s}_i)] \quad (4.33)$$

where  $k_f(\vec{s}_i)$  and  $k_b(\vec{s}_i)$  depend on the position of the first bound site of protein sequence  $\vec{s}_i$ , according to equations 4.31 and 4.32. A summary of the expression for all the rates is given in

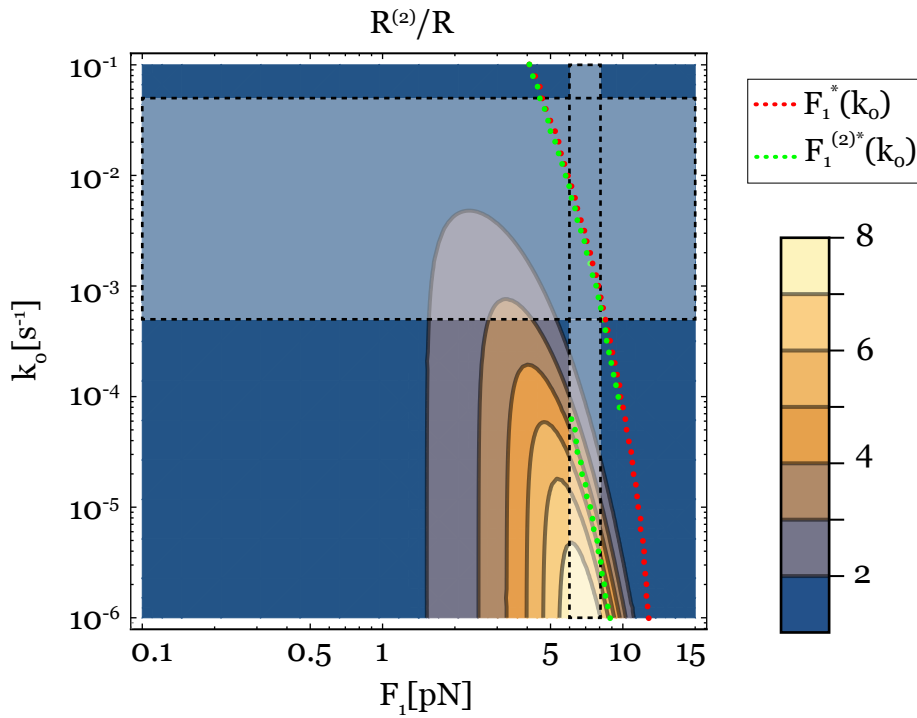


Figure 4.12: Ratio of the translocation rates  $R^{(2)}$  associated to a possible double binding of Hsp70 and  $R$ , with at most one Hsp70 per site. The ratio is plotted as a function of both parameters  $k_0$  and  $F_1$ . The colour points show the optimal value  $F_1^*$  (red) and  $F_1^{(2)*}$  (green) for a discrete set of values  $k_0$ . The two light dashed rectangles highlight the region compatible with typical biological conditions, that is  $k_0 = 5 \cdot 10^{-4} - 5 \cdot 10^{-2} \text{ s}^{-1}$  and  $F_1 = 6 - 8 \text{ pN}$ .

the Appendix, Tables A.3 to A.5.

## 4.8 Stronger together

To quantify the effect of the possible binding of a second Hsp70, we compare the updated translocation rate  $R^{(2)}$  given by equation 4.33 with the "basal" one computed in the first part of this chapter,  $R$  (equation 4.24). We compute the ratio  $R^{(2)}/R$ , which will be often referred (by misuse of language) to the acceleration or enhancement of the translocation, to give an appreciation of how much faster is the translocation if there is a possible dimerization of Hsp70 on the substrate binding sites. The result is shown as a function of the two parameters  $F_1$  and  $k_0$  in Figure 4.12.

As a first observation, we notice the presence of a orange-yellow region, corresponding to a

ratio  $R^{(2)}/R \approx 4-8$ , reflecting a significant enhancement of the translocation. This acceleration takes place at low values of  $k_0$ , i.e. for slow diffusive translocating proteins. By slow, it is meant approximately 2-3 orders of magnitudes slower than what is expected in typical biological conditions, that is  $k_0 \approx 5 \cdot 10^{-4} - 5 \cdot 10^{-2} \text{ s}^{-1}$  (light dashed rectangle). The red points show the optimal value  $F_1^*$  as a function of  $k_0$ , computed with the model with only a single binding per site (see figure 4.8B). Similarly, the same optimum value of  $F_1$  is computed when an additional binding is allowed in the extended model, thus defining  $F_1^{(2)*}$  and corresponding to the green dots.

The discontinuity of  $F_1^{(2)*}$  reveals a phase transition between two regimes.

- On the upper part of Figure 4.12 is the regime driven by the translocation with one single Hsp70 (red and green line almost overlap, because dimeric states form very rarely). The range of parameters for which the translocation is optimal (that is the the red and green curves) falls within the expected biological conditions (see Appendix A.2.2), that is  $k_0 \approx 5 \cdot 10^{-4} - 5 \cdot 10^{-2} \text{ s}^{-1}$  and  $F_1 \approx 6 - 8 \text{ pN}$ . It is highly remarkable because, according to our model, biological systems evolved so that a mechanism such as protein translocation takes place in conditions that are optimal or quasi optimal, thus enabling an efficient translocation of substrate proteins into the ER lumen and the mitochondria.
- Let us consider the situation in which  $k_0$  is decreased while assuming a fixed value  $F_1 \approx 6 - 8 \text{ pN}$  (illustrated by the thin light rectangle). On the one hand, we notice that we fall again on the optimality condition associated to multiple binding (that is the lower part of the green dashed line). On the other hand, it also coincides with the range of parameters for which the translocation is strongly enhanced upon a possible second binding (illustrated by the yellow-orange region). Both observations together are very strong: if for some reasons the diffusion of the substrate along the pore is not favourable, there is an emerging adaptation of the system, in which two Hsp70 will simultaneously bind to accelerate the translocation and perform it almost optimally.

To strengthen the idea that the system adapts to external conditions through a second binding (and not simply a new steady state involving essentially binding sites with a single Hsp70), we computed the average number of Hsp70 that are bound at steady state on the first binding site.

Figure 4.13 supports the emergence of the dimeric state, by showing that for  $k_0 < 10^{-3}$  the average number of chaperone proteins bound on the first binding site (which can also be empty) increases to a value larger than 1 ( $\sim 1.3$ ).

The discontinuity observed on the green dotted line in figure 4.12 comes from the presence of two distinct local maxima when plotting the translocation rate  $R^{(2)}$  as a function of  $F_1$  (figures 4.14A to 4.14D). The one associated to a lower value of  $F_1$  progressively increases, until getting

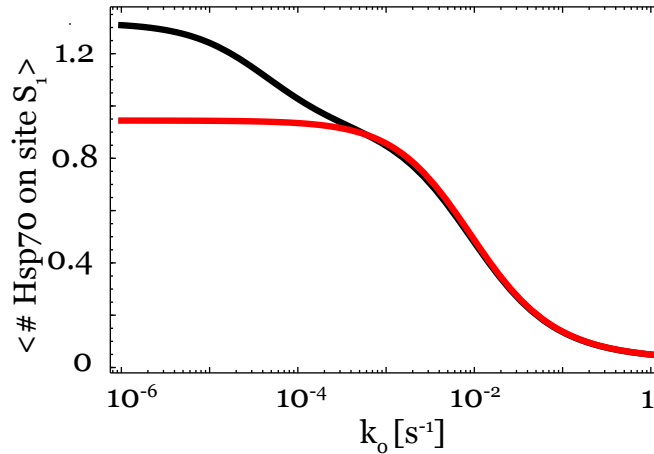


Figure 4.13: Average number of chaperone on the first binding site  $S_1$ , the one the closest to the pore. The result is shown with the initial model (single binding, red) and with the extended model (possible double binding, black).  $F_1 = 6$  pN and  $[ATP]/[ADP] = 10$ .

dominant with respect to the one on the left which corresponds to the red dotted line in figure 4.12.

To push further the interpretation of the two peaks, we consider more closely the case  $k_0 = 10^{-5} \text{ s}^{-1}$  (Figure 4.14C). We show in Figure 4.15 both translocation rates  $R$  and  $R^{(2)}$  (respectively the red and black lines) and additionally split  $R^{(2)}$  into two contributions: the one with one Hsp70 on the first bound site and the one with two Hsp70 on the first bound site. These two contributions are illustrated with the black dashed line (one Hsp70) and black dotted line (two Hsp70s).

Concretely, these observations are possibly reflecting the adaptation of the system to the external environment. A decrease of the diffusion rate  $k_0$  is the consequence of the presence of a misfolded protein at the pore, whose conformation makes the translocation mechanistically more difficult. As a consequence, and to adapt to such a situation, there is a combined action of multiple Hsp70s to exert a stronger force. The mechano-adaptation highlighted by these results reminds of other biological systems that exhibit such similar behaviours [162, 163].

## 4.9 Analytical study of a simple model of translocation

The results of the numerical simulations of both the first model and its extension reveal new interesting features of the translocation mechanism and possible implications on structural rearrangement of chaperone proteins within the framework of protein translocation. This

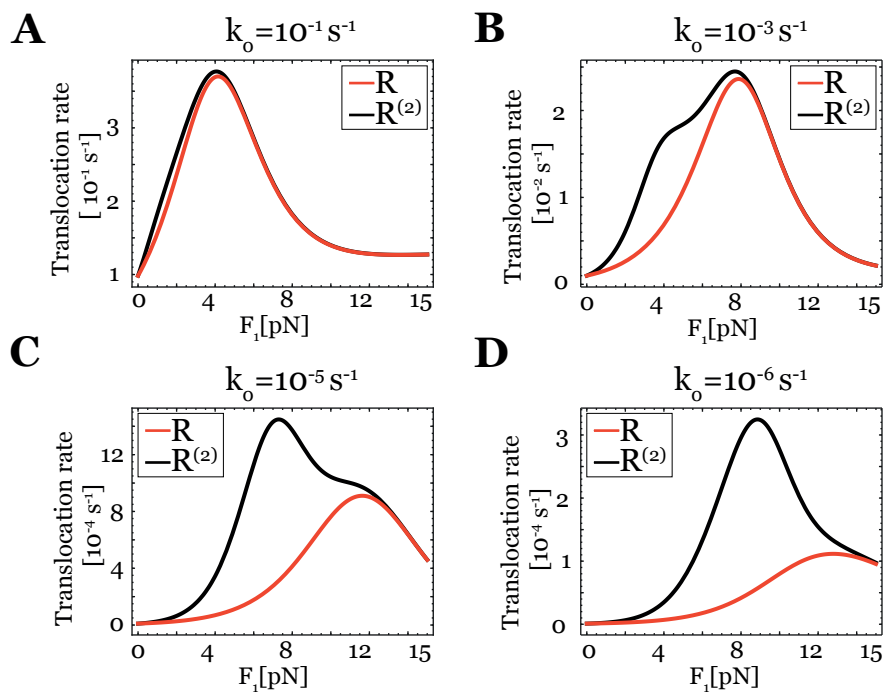


Figure 4.14: Translocation rates  $R$  (red) and  $R^{(2)}$  (black) as a function of  $F_1$ , for different values of the diffusion rate: (A)  $k_0 = 10^{-1} \text{ s}^{-1}$ ; (B)  $k_0 = 10^{-3} \text{ s}^{-1}$ ; (C)  $k_0 = 10^{-5} \text{ s}^{-1}$ ; (D)  $k_0 = 10^{-6} \text{ s}^{-1}$ .  $[ATP]/[ADP] = 10$ .

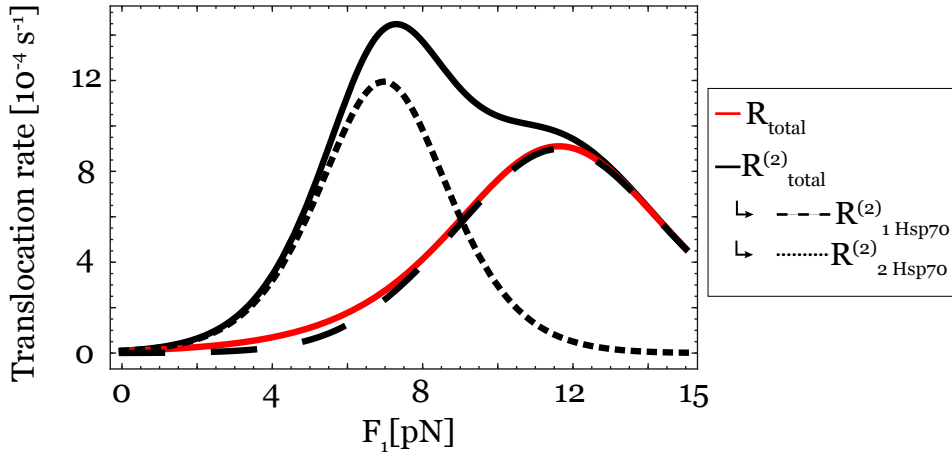


Figure 4.15: Translocation rates  $R$  (red) and  $R^{(2)}$  (black) as a function of  $F_1$  in the case  $k_0 = 10^{-5} \text{ s}^{-1}$ . The black curve is split as a sum of two contributions: a translocation rate induced by a single Hsp70 on the first bound site ( $R_{1 \text{ Hsp70}}^{(2)}$ , black dashed line) and a translocation rate induced by two Hsp70 on the first bound site ( $R_{2 \text{ Hsp70}}^{(2)}$ , black dotted line).  $[ATP]/[ADP] = 10$ .

model obviously contains some limitations: numerical simulations provide very instructive results about the dynamics of translocation but the analytical solution, although formally possible, gets very quickly unfeasible as soon as the size of the simulation volume increases (system of  $3^{N_s}$  or  $5^{N_s}$  linear equations). Our next goal is thus to develop a simpler description highlighting the intrinsic logic of translocation, even at the expense of biological accuracy.

With that aim, the model is further simplified with the following assumptions:

- There are only two binding sites on the protein;
- At most one Hsp can be bound on the protein (not only on a given site);
- When a Hsp70 is on the second site, the protein can undergo a backward translocation, not a forward one;
- As in the original model, the concentration in solution is chemostated to its steady-state value in solution.

Considering also the possible state of Hsp70 either with  $ATP$  or  $ADP$ , there are 5 possible configurations for the 2-site protein sequence, which are illustrated in figure 4.16 with the possible associated transitions. We address the question of an analytical expression for a translocation rate in this simplified model.

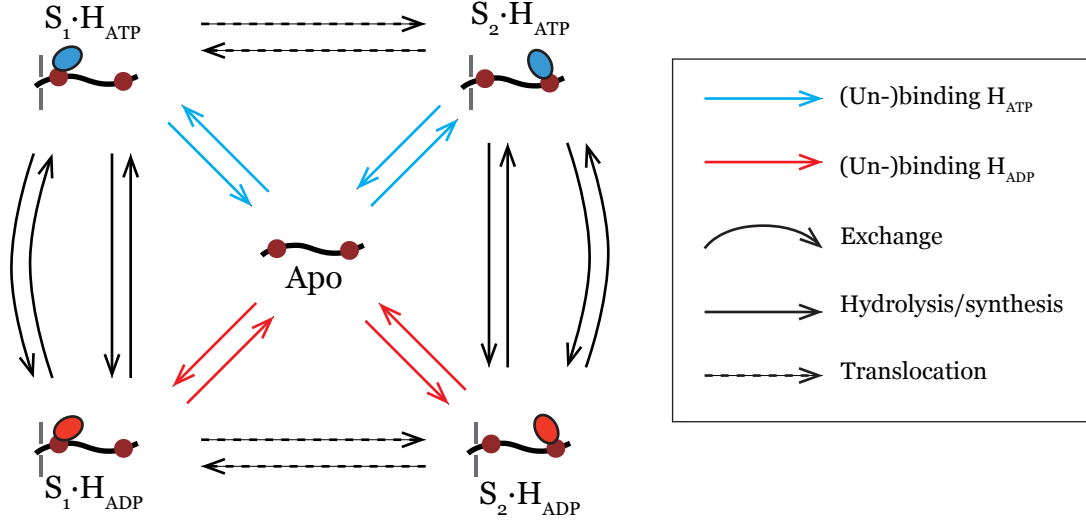


Figure 4.16: Simple kinetic model of translocation. The protein sequence with two sites can be in only 5 configurations. The membrane is represented with a grey rectangle on the left of the protein sequence. The notation  $H_T$  (resp.  $H_D$ ) is used in the text to refer to the states  $H_{ATP}$  (resp.  $H_{ADP}$ ).

Hereafter, we present only the starting and ending points of the analytical development in the main text. A more detailed derivation is reported in the Appendix (A.2.6). The translocation rate  $R$  is given by:

$$R = \underbrace{P(S_1 \cdot H_T) \cdot k_{f,H_T}^{(1)} - P(S_2 \cdot H_T) \cdot k_{b,H_T}^{(2)}}_{\phi_T} + \underbrace{P(S_1 \cdot H_D) \cdot k_{f,H_D}^{(1)} - P(S_2 \cdot H_D) \cdot k_{b,H_D}^{(2)}}_{\phi_D} \quad (4.34)$$

where we explicitly considered the state  $H_X$  of the chaperone in the forward and backward transition rate (e.g.  $k_{f,H_X}^{(1)}$ ), although it is assumed to be invariant.

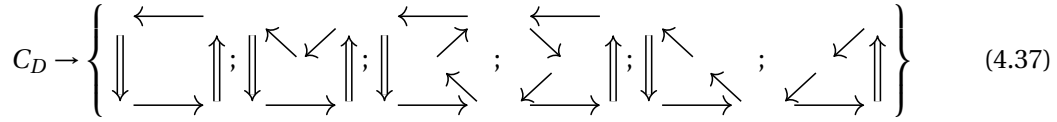
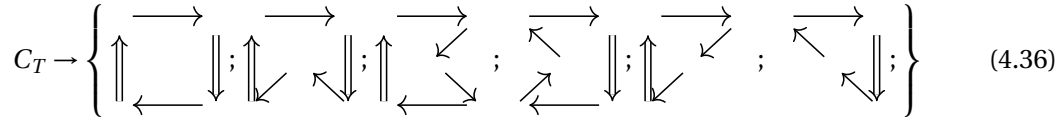
The translocation rate is the sum of two net fluxes ( $\phi_T$  and  $\phi_D$ ) for which equation 2.22 (remined below) relative to the previously introduced formalism of kinetic models is applicable. As a reminder, the equation to compute the flux between two states  $X_1$  and  $X_2$  in a kinetic model is:

$$\Phi_{12} = \sum_{c \in C_{12}} (W_c^\circ - W_c^\circ) \cdot w_c = \sum_{c \in C_{12}} W_c^\circ \cdot w_c \cdot \left(1 - \frac{W_c^\circ}{W_c^\circ}\right) \quad (4.35)$$

where  $C_{12}$  is the list of cycles that contain the transition between  $X_1$  and  $X_2$ . The corresponding sets in the translocation model are  $C_T$  (between  $S_1 \cdot H_T$  and  $S_2 \cdot H_T$ ), and similarly for  $C_D$ .



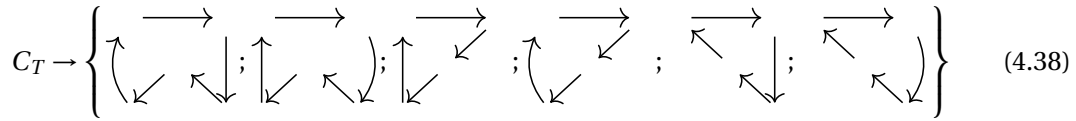
#### 4.9 Analytical study of a simple model of translocation



where the double arrows stand for the combination of exchange and hydrolysis/synthesis. The cycle with one (respectively two) double arrows is the concise representation of two (respectively four) distinct cycles in the summation.

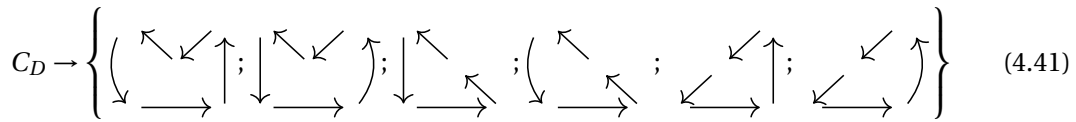
From the combined summation on both sets  $C_T$  and  $C_D$ , two classes of cycles lead to vanishing terms. First, the cycles  $c$  for which  $W_c^\circ/W_c^\circ = 1$ . Such cycles are the one that go through either only exchange or only hydrolysis/synthesis. Second the cycles that go through two translocation transitions, one in the forward and one in the backward direction. These terms vanish because the contribution in the summation over  $C_T$  is the exact opposite than the one in the summation over  $C_D$ , thus summing to zero in equation 4.34.

It remains six terms in both sums (over  $C_T$  and  $C_D$ ). The corresponding set of cycles, with the associated values of  $W_c^\circ/W_c^\circ$  (or the inverse) and  $w_c$  are the following:



$$\frac{W_c^\circ}{W_c^\circ} \rightarrow \left\{ \frac{\alpha}{\alpha_{eq}}; \frac{\alpha_{eq}}{\alpha}; \frac{P_D P_T^{eq}}{P_T P_D^{eq}}; \frac{\alpha P_D P_T^{eq}}{\alpha_{eq} P_T P_D^{eq}}; \frac{P_T P_D^{eq}}{P_D P_T^{eq}}; \frac{\alpha_{eq} P_T P_D^{eq}}{\alpha P_D P_T^{eq}} \right\} \quad (4.39)$$

$$w_c \rightarrow \left\{ 1; 1; k_{b,H_D}^{(2)} + k_s^{S_2} + k_{ex}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} + k_{off,H_D}^{S_2}; k_{b,H_D}^{S_1} + k_s^{S_2} + k_{ex}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} + k_{off,H_D}^{S_2}; k_{f,H_D}^{(1)} + k_s^{S_1} + k_{ex}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} + k_{off,H_D}^{S_1}; k_{f,H_D}^{(1)} + k_s^{S_1} + k_{ex}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} + k_{off,H_D}^{S_1} \right\} \quad (4.40)$$



$$\frac{W_c^\circ}{W_c^\circ} \rightarrow \left\{ \frac{\alpha_{eq}}{\alpha}; \frac{\alpha}{\alpha_{eq}}; \frac{P_T P_D^{eq}}{P_D P_T^{eq}}; \frac{\alpha_{eq} P_T P_D^{eq}}{\alpha P_D P_T^{eq}}; \frac{P_D P_T^{eq}}{P_T P_D^{eq}}; \frac{\alpha P_D P_T^{eq}}{\alpha_{eq} P_T P_D^{eq}} \right\} \quad (4.42)$$

$$w_c \rightarrow \left\{ 1; k_{b,H_T}^{(1)} + k_h^{S_2} + k_{ex}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} + k_{off,H_T}^{S_2}; k_{b,H_T}^{(2)} + k_h^{S_2} + k_{ex}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} + k_{off,H_T}^{S_2}; \right. \\ \left. k_{f,H_T}^{(1)} + k_h^{S_1} + k_{ex}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} + k_{off,H_T}^{S_1}; k_{f,H_T}^{(1)} + k_h^{S_1} + k_{ex}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} + k_{off,H_T}^{S_1} \right\} \quad (4.43)$$

In what comes next, the notation assumes that the exchange rates do not depend on the site, that is

$$k_{ex}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} = k_{ex}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} =: k_{ex}^{S \cdot H_D \rightarrow S \cdot H_T} \quad (4.44)$$

$$k_{ex}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} = k_{ex}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} =: k_{ex}^{S \cdot H_T \rightarrow S \cdot H_D} \quad (4.45)$$

The kinetic asymmetry between transitions on site 1 ( $S_1 \cdot H_T \leftrightarrow S_1 \cdot H_D$ ) and site 2 ( $S_2 \cdot H_T \leftrightarrow S_2 \cdot H_D$ ) is only modelled through the hydrolysis and synthesis rates. In addition, we kept the analytical trace of the distinction between  $k_{0,H_T}$  and  $k_{0,H_D}$ , although they are assumed to be the same.

After an analytical development of the summation (sketched in the Appendix, section A.2.6), we obtain for the net translocation rate:

$$R = \overset{\rightarrow}{\underset{\nwarrow}{\kappa}} \cdot \underbrace{\left(1 - \frac{\alpha_{eq}}{\alpha}\right)}_{(1)} \cdot \underbrace{\left(\frac{k_{off,H_T}^{(0)} k_0^D}{k_{off,H_D}^{(0)} k_0^T} - 1\right)}_{(2)} \cdot \underbrace{(\lambda - 1)}_{(3)} \cdot \mathcal{F}(\{k\}) \quad (4.46)$$

where (1), (2) and (3) refer to the below discussion and

$$\mathcal{F}(\{k\}) = \left[ k_{off,H_D}^{(0)} \cdot \frac{\frac{\alpha}{\alpha_{eq}} k_{ex}^{H_T \rightarrow H_D}}{k_h + \frac{\alpha}{\alpha_{eq}} k_{ex}^{H_T \rightarrow H_D}} + k_{ex}^{S \cdot H_D \rightarrow S \cdot H_T} \left( \frac{P_D k_{on,H_D}^{(0)}}{P_T k_{on,H_T}^{(0)}} + 1 \right) \right] \quad (4.47)$$

Equation 4.46 sheds light on three conditions that have to be satisfied to move the system from its equilibrium dynamics, namely the absence of translocation ( $R = 0$ ). These conditions remarkably mirror some main biological features of Hsp70 and the translocation machinery. The enumeration (1, 2 and 3) refers to the annotation in equation 4.46.

1. *The system is moved far from equilibrium conditions:*

The term (1) vanishes if and only if  $\alpha = \alpha_{eq}$ . As a fingerprint of the non-equilibrium nature of translocation, the uptake of energy in the form of ATP is necessary to observe a net translocation of the protein.

2. *Hydrolysis is accelerated close to the membrane:*

The condition  $\lambda = k_h^{S_1} / k_h^{S_2} \neq 1$  formally breaks the symmetry between the two sites

## 4.9 Analytical study of a simple model of translocation

1 and 2 of the protein. If both are governed by a same dynamics, the translocation is not effective. This condition finds its counterparts in the biological description of the translocation: the role of J-protein cochaperones is rightly to accelerate the ATP-ase activity of Hsp70 close to the membrane

3. *The unbinding rate of the substrate is larger in the  $H_T$  state than in the  $H_D$  state:*

As a preliminary remark, the discussion assumes that  $k_{0,H_T} = k_{0,H_D} = k_0$ . The rationale is that the diffusion constant of the protein through the pore is not affected by the state of the bound chaperones  $H_T$  or  $H_D$ . We kept trace of the distinction for a more accurate analytical result, but it is expected that the ratio  $k_0^T / k_0^D$  is unitary.

This being said, the condition restricts to  $k_{off,H_D}^{(0)} \neq k_{off,H_T}^{(0)}$ . The structure of the Hsp70 SBD illustrated in figure 4.2 with the closing of the lid, capturing the substrate, is the biological manifestation of the present condition. In the closed conformation of the SBD (with  $H_D$ ), the unbinding rate is much smaller than in the open state (with  $H_T$ ).

More abstractly, the condition acts as an additional symmetry breaking requirement. In the previous point, we discussed the symmetry breaking between sites 1 and 2. Here, the other symmetry of the kinetic scheme has to be broken, that is between states of the protein with bound  $H_T$  or  $H_D$ . This symmetry breaking is achieved via the condition  $k_{off,H_D}^{(0)} \neq k_{off,H_T}^{(0)}$ .

Interestingly the double symmetry breaking acts as an overall requirement for a cyclic directionality. If at least one of them is not broken, the dynamics of the system is stuck in a stationary regime in which no net translocation is achieved.

It is remarkable how such a simple model is able to grasp the logic of the protein translocation. The different conditions mirror key biological aspects, structural and functional, that are essential to an efficient translocation. However, this has to be moderated, by keeping in mind that these results do not accurately model the dynamics of the translocation.

The main limitation of the current model, that makes it importantly differ from the numerical simulations, is the restriction to a single Hsp70 on the protein sequence. In real biological instances, the simultaneous action of Hsp70, binding to the incoming site "as soon as possible" is at the heart of the performance of the translocation for multiple reasons: it prevents retro-translocation, it leads to a stronger exerted force, even before the previous Hsp70 has detached from the substrate and thus decreases the time scales in game during the translocation.

The three conditions reveal the emerging logic of the translocation machinery. In the next section, we go further in this direction and show that the translocation is in fact the result of an information-processing cycle of operations, mediated by chaperone protein Hsp70.

## 4.10 Translocation as an autonomous Szilard Engine

In this section, we go deeper into the identification of the information processing steps achieved during the translocation of proteins. The analytical development is guided by the interpretation of the different biochemical transitions between states and how they are involved in translocation. We start by presenting the derivation that ensue from the definition of the translocation rate and show in which way the information emerges from these equations. The result is then discussed by representing the translocation machinery as a stochastic autonomous Szilard engine, as suggested by the analytical expression.

First of all, one hypothesis is made with respect to the kinetic scheme shown in figure 4.16 to further simplify the analytical development and its interpretation. We believe this assumption does not change the conceptual implications of our results, its validity being discussed at the end of this section.

We hypothesise that the *apo* state (empty protein) is a transient state in the consecutive unbinding and binding events of Hsp70 in the same state, either ATP-bound or ADP-bound. That is we neglect the possible "cross-interaction", consisting of releasing a Hsp70 in an ATP-state and binding a Hsp70 in an ADP-state, or vice versa. As a consequence, there are two possible transitions between states  $S_1 \cdot H_T \leftrightarrow S_2 \cdot H_T$  and similarly between  $S_1 \cdot H_D \leftrightarrow S_2 \cdot H_D$ : i) the translocation, as defined in the previous models; ii) the sliding of Hsp70 along the substrate protein, through consecutive unbinding and binding on two neighbour sites.

The representation of the different transitions is illustrated in figure 4.17A and we briefly introduce or remind the notation used for each of them.

- **Free energy along the substrate protein:**  $f_i := f(x_i) = F_1 x_1^2 / x_i$ ,  $i = \{1, 2\}$  (equation 4.2).
- **Translocation:**  $k_{f,H_X}^{(1)} := k_{0,H_X} \exp(\beta(f_1 - f_2))$  and  $k_{b,H_X}^{(2)} = k_{0,H_X}$  for the forward and backward translocation rates ( $X = \{D, T\}$ ).
- **Binding-release:** The "sliding" rates (terminology used to reflect the consecutive unbinding and binding on two neighbour sites) are defined by:

$$k_{slide}^{S_1 \cdot H_T \rightarrow S_2 \cdot H_T} := k_{slide}^T \exp(-\beta f_2) \quad (4.48)$$

$$k_{slide}^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} := k_{slide}^T \exp(-\beta f_1) \quad (4.49)$$

$$k_{slide}^{S_1 \cdot H_D \rightarrow S_2 \cdot H_D} := k_{slide}^D \exp(-\beta f_2) \quad (4.50)$$

$$k_{slide}^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D} := k_{slide}^D \exp(-\beta f_1) \quad (4.51)$$

so that detailed balance is satisfied with the translocation rates.

- **Hydrolysis-synthesis:**  $k_h^{S_1} = \lambda k_h^S$  and  $k_h^{S_2} = k_h^S$ . Similarly is defined synthesis, to satisfy detailed balance with exchange.

## 4.10 Translocation as an autonomous Szilard Engine

- **Exchange:**  $k_{ex}^{S \cdot H_T \rightarrow S \cdot H_D}$  and  $k_{ex}^{S \cdot H_D \rightarrow S \cdot H_T}$  are the exchange rates and do not depend on the considered binding site 1 or 2.
- **Total transition rate:** we define the total transition rate between two states on the same binding site as  $k_{tot}^{S_i \cdot H_T \rightarrow S_i \cdot H_D} := k_h^{S_i} + k_{ex}^{S \cdot H_T \rightarrow S \cdot H_D}$  and  $k_{tot}^{S_i \cdot H_D \rightarrow S_i \cdot H_T} := k_s^{S_i} + k_{ex}^{S \cdot H_D \rightarrow S \cdot H_T}$ .

Coming back to the very first definition of translocation rate,

$$R = P(S_1 \cdot H_T) \cdot k_{f,H_T}^{(1)} - P(S_2 \cdot H_T) \cdot k_{b,H_T}^{(2)} + P(S_1 \cdot H_D) \cdot k_{f,H_D}^{(1)} - P(S_2 \cdot H_D) \cdot k_{b,H_D}^{(2)} \quad (4.52)$$

we can develop the expression by introducing the formalism with the spanning trees explained in section 2.3 and already used in the previous analytical development (section 4.9). In that case, the structure of the kinetic scheme is more simple without the central state, reducing the length of the expressions. All the steps are written in the Appendix, section A.2.7.

$$\begin{aligned}
 R = \frac{1}{\Sigma_{TOT}} & \left[ k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_{slide}^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_{f,H_D}^{(1)} \right] \\
 & \cdot \left( 1 - \frac{P_{S_1}(S_1 \cdot H_T) P_{S_2}(S_2 \cdot H_D)}{P_{S_1}(S_1 \cdot H_D) P_{S_2}(S_2 \cdot H_T)} \cdot \frac{k_{b,H_D}^{(2)} k_{slide}^{S_1 \cdot H_T \rightarrow S_2 \cdot H_T}}{k_{f,H_D}^{(1)} k_{slide}^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T}} \right) \\
 + \frac{1}{\Sigma_{TOT}} & \left[ k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_{slide}^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D} k_{tot}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} k_{f,H_T}^{(1)} \right] \\
 & \cdot \left( 1 - \frac{P_{S_1}(S_1 \cdot H_D) P_{S_2}(S_2 \cdot H_T)}{P_{S_1}(S_1 \cdot H_T) P_{S_2}(S_2 \cdot H_D)} \cdot \frac{k_{b,H_T}^{(2)} k_{slide}^{S_1 \cdot H_D \rightarrow S_2 \cdot H_D}}{k_{f,H_T}^{(1)} k_{slide}^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D}} \right)
 \end{aligned} \quad (4.53)$$

where  $\Sigma_{TOT}$  is the sum of all the spanning trees over the four-state kinetic model.

At this point, we need to push further the interpretation of the two types of fractions (\*) and (\*\*) in brackets.

### 1. (\*) Ratio of conditional probabilities:

The force exerted by entropic pulling on the substrate protein depends on the relative position of Hsp70 with respect to the pore, not on the nucleotide bound on it. Thus, in general, the systems works optimally if it favours successive binding of Hsp70 on the first site, more than on the second one.

Taking a closer look to the terms in red in Equation 4.53, we see that there are two cycles

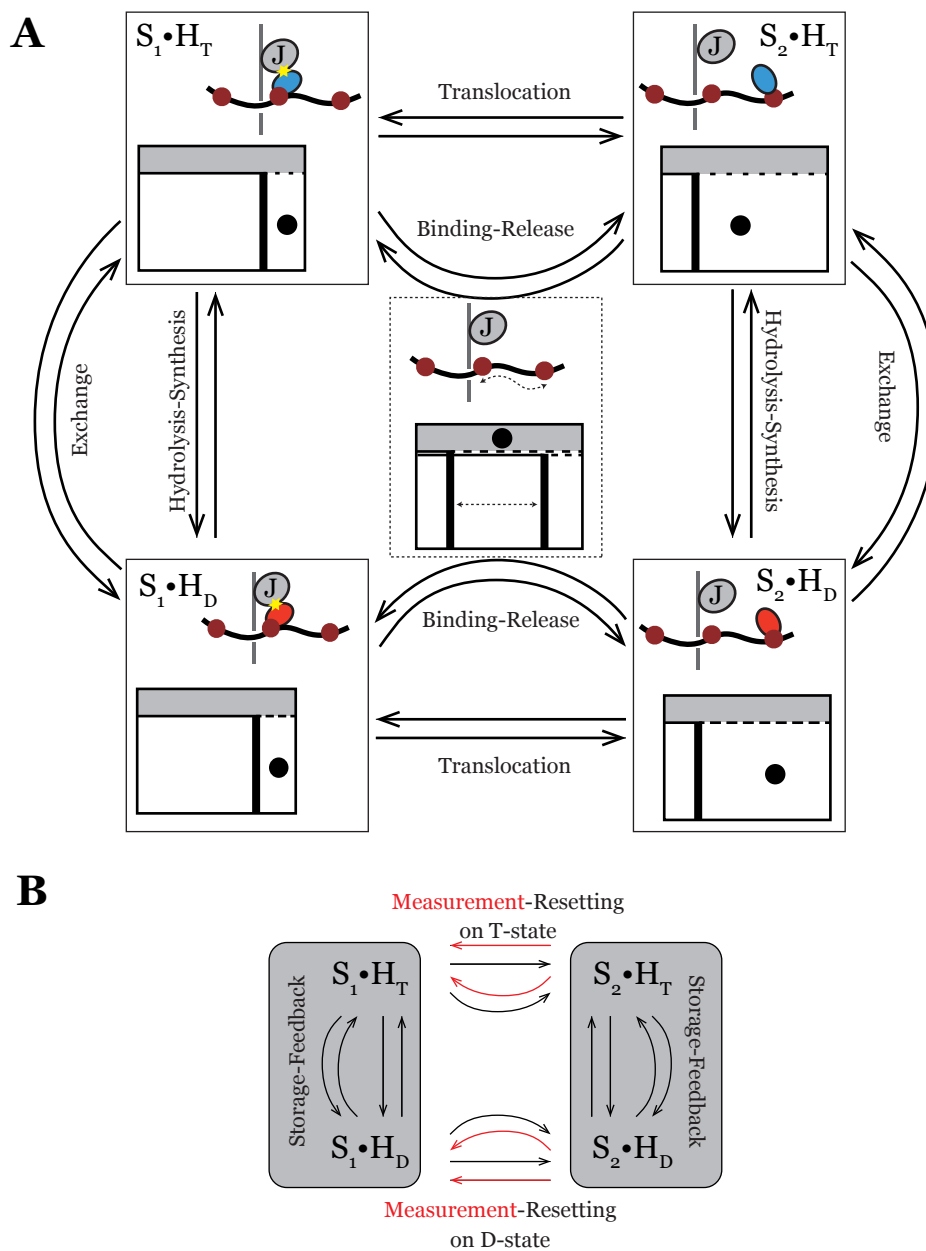
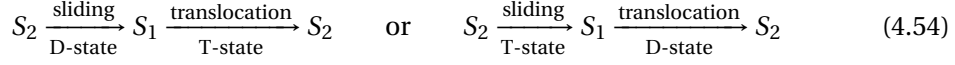


Figure 4.17: **(A)** Model for the translocation in which the empty protein sequence is a transient step in the sliding of Hsp70 along the substrate, leading to a 4-state model with the different transitions described in the figure. To each state in the translocation model (boxes (I) to (V), top) is associated an equivalent picture inspired from a revisited Szilard Engine (boxes (I) to (V), bottom). **(B)** Illustration of the information-processing steps involved in the translocation, with respect to the model illustrated in (A).

## 4.10 Translocation as an autonomous Szilard Engine

that are contributing to a net forward translocation. These two cycles are either



It is thus reasonable to postulate that the measurement process is defined by the transition from site  $S_2$  to site  $S_1$  and the resetting that follows the measurement is the transition from site  $S_1$  to site  $S_2$ .

Formally speaking, the paradigm of the measurement and resetting is very general, by identifying the measurement to any transition from  $S_2$  to site  $S_1$  and the resetting to any transition from  $S_1$  to  $S_2$ . The information associated to both steps is defined as a trajectory-dependant information (Section 2.8) and the ideas behind these definitions are strongly similar to the one used to describe ABC Transporters (Section 3.5).

Importantly, in the most general description, measurement is any transition  $S_2 \rightarrow S_1$ . However, if we refer more specifically to the two cycles that "carry" the forward translocation, the measurement consists in both cases in the sliding transition, associated to the unbinding from  $S_2$  and consecutive binding on  $S_1$ . Similarly, the resetting from  $S_1$  to  $S_2$  is always in the form of a forward translocation.

$$\text{Measurement } S_2 \cdot H_T \xrightarrow{\text{sliding}} S_1 \cdot H_T : \quad I_{meas,T} := \ln \left( \frac{P_{S_1}(S_1 \cdot H_T)}{P_{S_2}(S_2 \cdot H_T)} \right) \quad (4.55)$$

$$\text{Measurement } S_2 \cdot H_D \xrightarrow{\text{sliding}} S_1 \cdot H_D : \quad I_{meas,D} := \ln \left( \frac{P_{S_1}(S_1 \cdot H_D)}{P_{S_2}(S_2 \cdot H_D)} \right) \quad (4.56)$$

$$\text{Resetting } S_1 \cdot H_T \xrightarrow{\text{translocation}} S_2 \cdot H_T : \quad I_{reset,T} := \ln \left( \frac{P_{S_2}(S_2 \cdot H_T)}{P_{S_1}(S_1 \cdot H_T)} \right) \quad (4.57)$$

$$\text{Resetting } S_1 \cdot H_D \xrightarrow{\text{translocation}} S_2 \cdot H_D : \quad I_{reset,D} := \ln \left( \frac{P_{S_2}(S_2 \cdot H_D)}{P_{S_1}(S_1 \cdot H_D)} \right) \quad (4.58)$$

where  $I_{meas,T} = -I_{reset,T} =: I_T$  and  $I_{meas,D} = -I_{reset,D} =: I_D$ .

In the end, we obtain:

$$\frac{P_{S_1}(S_1 \cdot H_T) P_{S_2}(S_2 \cdot H_D)}{P_{S_1}(S_1 \cdot H_D) P_{S_2}(S_2 \cdot H_T)} = \exp(I_{meas,T} + I_{reset,D}) \quad (4.59)$$

$$\frac{P_{S_1}(S_1 \cdot H_D) P_{S_2}(S_2 \cdot H_T)}{P_{S_1}(S_1 \cdot H_T) P_{S_2}(S_2 \cdot H_D)} = \exp(I_{meas,D} + I_{reset,T}) \quad (4.60)$$

### 2. (\*\*) Possible ways to accelerate the translocation:

The two terms (\*\*) are the two following fractions:

$$\exp(\beta\chi_T) := \frac{k_{f,H_T}^{(1)} k_{slide}^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D}}{k_{b,H_T}^{(2)} k_{slide}^{S_1 \cdot H_D \rightarrow S_2 \cdot H_D}} \quad \text{and} \quad \exp(\beta\chi_D) := \frac{k_{f,H_D}^{(1)} k_{slide}^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T}}{k_{b,H_D}^{(2)} k_{slide}^{S_1 \cdot H_T \rightarrow S_2 \cdot H_T}} \quad (4.61)$$

## Chapter 4. Substrate Translocation induced by Entropic pulling

Following the detailed balance constrains, they can be rewritten only as a function of exchange rates:

$$\exp(\beta\chi) := \frac{k_{ex}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_{ex}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T}}{k_{ex}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} k_{ex}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D}} = \exp(\beta\chi_T) = \exp(-\beta\chi_D) \quad (4.62)$$

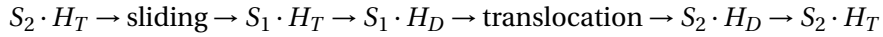
Putting all the ingredients together, we obtain the final expression for the net translocation rate:

$$\begin{aligned} R = & \frac{1}{\Sigma_{TOT}} \left[ k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_{slide}^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_{f, H_D}^{(1)} \right] \cdot (1 - \exp(I_{meas, T} + I_{reset, D}) \exp(\beta\chi_T)) \\ & + \frac{1}{\Sigma_{TOT}} \left[ k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_{slide}^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D} k_{tot}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} k_{f, H_T}^{(1)} \right] \cdot (1 - \exp(I_{meas, D} + I_{reset, T}) \exp(\beta\chi_D)) \end{aligned} \quad (4.63)$$

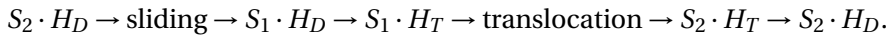
or equivalently with a slightly condensed notation:

$$\begin{aligned} R = & \frac{1}{\Sigma_{TOT}} \left[ k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_{slide}^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_{f, H_D}^{(1)} \right] \cdot (1 - \exp(I_T - I_D + \beta\chi)) \\ & + \frac{1}{\Sigma_{TOT}} \left[ k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_{slide}^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D} k_{tot}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} k_{f, H_T}^{(1)} \right] \cdot (1 - \exp(-I_T + I_D - \beta\chi)) \end{aligned} \quad (4.64)$$

The structure of the last equation provides many insights to better understand the translocation as an information processing mechanism. First of all, the mathematical structure highlights that there are exactly two possible trajectories that drive the forward translocation, either



or



Each of the trajectories has a weight, in the form  $(1 - [\dots])$ , where the two bracketed terms  $[\dots]$  are the inverse of each other: necessarily, one cycle will contribute positively to the forward translocation and the other will contribute negatively (i.e induce a backward translocation).

In accordance with the definition of exchange rates (independent on the binding site),  $\chi = 0$ . However, this term might, from a broader perspective, play an important role. Indeed, different hypotheses go in the direction of  $\chi \neq 0$ , that could lead to the acceleration of Hsp70. A first one is to include in our model that binding a NEF far from the pore is more favourable, as for Hsp70. It would result in a symmetry breaking in the ratio of exchange rates, with



#### 4.10 Translocation as an autonomous Szilard Engine

$k_{ex}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} / k_{ex}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} > k_{ex}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} / k_{ex}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D}$ . Another hypothesis comes from informal discussions: it might be that the ATPase activity of the Hsp110 NEF is essential to the protein disaggregation. Similarly during the translocation, the energy consumed from the ATPase activity could thus lead to a term  $\chi \neq 0$  that would accelerate the translocation. Obviously these are hypotheses that still have to be studied and confirmed, but our model paves the way to understand the different biochemical mechanisms that could contribute to accelerate the translocation through a membrane.

Overall, when considering all the elements and putting them into perspective with the biological description of the translocation, a global picture of the mechanism is captured by our model. The interplay between Hsp70 and cochaperones induces a directionality in the two subsystems that are either on site 1 ( $S_1 = \{S_1 \cdot H_T, S_1 \cdot H_D\}$ ) or on site 2 ( $S_2 = \{S_2 \cdot H_T, S_2 \cdot H_D\}$ ). First, the activity of J-protein is targeted around its anchorage point, on Tim44, and results in an acceleration of the hydrolysis on the first binding site. Second, the NEFs favor the exchange from ADP-bound to ATP-bound Hsp70 on both sites  $S_1$  and  $S_2$  but the effect of their interaction with Hsp70 dominates after the translocation, that is on  $S_2$ . As a consequence, when considering the two subsystems (site 1 and site 2) as non interacting, and comparing the relative occupancy of  $H_T$ - or  $H_D$ - bound sites, we get  $P_{S_1}(S_1 \cdot H_T) < P_{S_2}(S_2 \cdot H_T)$  and  $P_{S_1}(S_1 \cdot H_D) > P_{S_2}(S_2 \cdot H_D)$ . Therefore,  $I_T - I_D < 0$ , which means that the translocation cycle that go through  $S_2 \cdot H_T \rightarrow \text{sliding} \rightarrow S_1 \cdot H_T \rightarrow S_1 \cdot H_D \rightarrow \text{translocation} \rightarrow S_2 \cdot H_D \rightarrow S_2 \cdot H_T$  is amplified and dominates the overall translocation mechanism.

At equilibrium, there is no information that is driven between the two sites and thus in both cases,  $I_{meas} + I_{reset} = 0$ , leading unavoidably to  $R = 0$ . The preponderance of this cycle is especially true since the product of rates along the dominant cycle

$$k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_s^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_{f, H_D}^{(1)}$$

is expected to be way larger than the product of rates on the alternative cycle from the following equalities and inequalities:

$$k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} > k_{tot}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} \quad (4.65)$$

$$k_s^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} > k_s^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D} \quad (4.66)$$

$$k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} > k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} \quad (4.67)$$

$$k_{f, H_D}^{(1)} = k_{f, H_T}^{(1)} \quad (4.68)$$

As a conclusion to this chapter on the translocation as an information processing mechanism, we build a revisited Szilard Engine that shares all the different stages of the translocation machinery. The correspondence is illustrated in figure 4.17A, bottom illustration of boxes  $I$  to  $V$ .

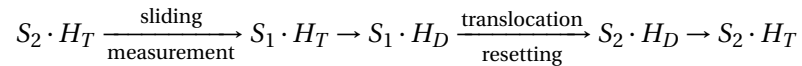
## Chapter 4. Substrate Translocation induced by Entropic pulling

---

In order to reproduce the two possible binding sites in the simple model of translocation, the setup consists of a box with a wall that can be in two different positions. The particle is assumed to be always located on the right of the wall when it is inside the box. When the wall is on the right, it moves in a small volume (A.I and A.III). Conversely, when the wall is on the left, the particle moves in a larger volume (A.II and A.IV). When a Hsp70 is bound on the first site, it exerts a pulling force to move to a more favourable state, a state of higher entropy. Similarly, the favourable evolution of the system when a particle is in a small volume is to undergo a volume expansion, by pushing on the wall and thus increasing the available volume (A.I  $\rightarrow$  A.II and A.III  $\rightarrow$  A.IV).

Also, it is possible for Hsp70 to detach and bind from the substrate protein. This has to be part of the Szilard device: the particle can escape from the box and move to a side reservoir (A.V), pictured as a grey rectangle on the top of the box. The bound nucleotide does not directly affect the translocation rates but strongly modifies the affinity of the chaperone for the substrate, and thus the sliding rates. In the perspective of the Szilard engine, it corresponds to a membrane between the box and the reservoir that can have different permeability properties. On the one hand, it can let the particle go to the reservoir and come back into the box relatively freely, as would Hsp70 do in the ATP-bound state by frequently binding and unbinding. On the other hand, the membrane can be almost impermeable, in the same way the dissociation constant of ADP-bound Hsp70 for the substrate is very low.

Importantly the stochasticity associated to this description makes it differ conceptually from the original Szilard Engine, in which every step is preformed externally and in a totally deterministic way, without any fluctuations. Indeed, among all the possible cycles that are included in our model, the only one that is iteratively repeated in the original Szilard Engine can be assimilated to:



in which even the internal transitions  $S_1 \cdot H_T \rightarrow S_1 \cdot H_D$  and  $S_2 \cdot H_D \rightarrow S_2 \cdot H_T$  are not explicit part of the description.

The exact transposition of the translocation machinery into an autonomous stochastic Szilard Engine supports the analytical development that led us to understand the translocation as a result of an autonomously information processing molecular device.

### 4.11 Conclusive discussion

Throughout this project about the translocation of substrate proteins across the membrane, we studied the dynamics of the translocation based on the entropic pulling model. The aim was to highlight the fundamental biological properties of the system that drive the translocation process, and how they affect the performance of the translocation, quantified by the net

translocation rate.

As a basis of our work is the entropic pulling model, one possible model for the translocation mechanism across the membrane. Our model includes all the physical and biological features of the entropic pulling that were so far reported in the literature, essentially characterized by considering a single Hsp70 interacting with a substrate protein. The interest of our study lies in the description at a larger scale, that is a substrate protein interacting with multiple Hsp70 chaperone proteins in solution. The model provides a precise description of dynamics, from which exciting features emerge to better understand the underlying mechanism.

The results of this project essentially articulate around two directions.

First, we showed from numerical simulations that the translocation is optimally achieved when the interplay between two counterbalancing effects is adequate: on the one hand, the entropic pulling force must be large enough to impose a directionality to the translocation. On the other hand, this same parameter has to be sufficiently small, so that Hsp70s bind frequently enough to the protein to generate a pulling force. Resulting from these observations, we also showed the existence of a range of biological conditions in which the dimerization of Hsp70 strongly accelerates the translocation rate of substrate protein, as an adaptative process with respect to the environment. This might occur for instance in the presence of an obstacle to the translocation, such as a misfolded substrate protein, that would unavoidably slow down the translocation.

Second, we went through the analytical solution of a simpler model of translocation to reveal the logics of transport and its fundamental mathematical structure. Interestingly, all the biological conditions emerge as logical requirements, such as the acceleration of ATP-ase activity and different affinity for the substrate in the ATP- and ADP- states, all this being possible only upon uptake of energy with respect to equilibrium conditions. As a final step of this work, we showed that all the steps of the translocation mechanism find their counterpart in a *revisited* autonomous Szilard Engine, in which an information flow between different states of the system drives the translocation.

Our results are very far-reaching and complementary. The numerical simulations are meant to be realistic and representative of biological systems. Despite evident assumptions and simplifications in the model, we obtain highly relevant results, especially the predicted dimerization of Hsp70 to drive the translocation. These numerical observations open interesting perspectives for new experimental set-ups to answer the following question: are Hsp70s observed in a dimeric form during the translocation when the diffusion rate of the protein is reduced, for instance if it is in a misfolded state? The experimental challenge behind is significant, but nonetheless exciting.

Importantly, all these results are characteristic of the steady-state of the translocation, that is when the boundary effects (insertion of the protein through the membrane, for instance)

#### Chapter 4. Substrate Translocation induced by Entropic pulling

---

are neglected. Moreover the protein is of finite size and is probably relatively fast translocated in its whole. Future numerical simulations using Kinetic Monte Carlo algorithm would be adequate to include this effect into the model and quantify how does it affect the translocation rate, and to compare it with the steady-state transient regime we focused on in this work.

The theoretical part of the model is much more fundamental and contains important assumptions and limitations. The main one being that only one Hsp70 can be bound simultaneously on the whole two-site protein. In spite of this, and here is the strength of the results, logics of transport is amazingly grasped and reveals the information-processing nature of the translocation machinery. A future challenging analytical and numerical research would be the extension of these results by raising some of the main hypothesis, to capture more precisely the *real* energy consumption balance that operates during the protein translocation, in particular the one associated to the processing of information.

## 5 Conclusion and outlooks

Throughout this thesis, we described biological systems as information-processing devices and showed that this characterisation finds its roots in the biochemical interactions at the molecular level. During our research, we aspired to describe cellular transporters with models as simple as possible, still grasping the essential features of the inherent mechanisms suggested by the biological description of the systems. Finding the appropriate balance between simplification and the right amount of complexity was a real challenge, which we faced using kinetic models and applying the formalism of stochastic thermodynamics to study and characterise the transport processes. We shed light on the logical mathematical structure, remarkably mirroring the information-processing stages that are intrinsically part of these processes, letting us state that active and energy consuming biological systems are internally and autonomously perpetually processing information to achieve their assigned task. This approach was used to tackle two types of membrane transporters: ABC transporters and the translocation machinery involving chaperone proteins Hsp70s.

In the first part, we built a kinetic model for ABC transporters. As a cornerstone is the experimentally observed dependence of the ATPase activity on substrate binding. The associated allosteric property was reproduced by introducing two distinct ATP-bound states with different hydrolysis rates, consecutively modelling the effect of the substrate as a population shift between these two states. We quantified the steady-state concentration gradient that results from such a characterisation.

With this innovative description, we revealed the intrinsic logic of the transport mechanism through the simultaneous necessity of different conditions. These three requirements are strongly related to symmetry breakings: the symmetry between kinetic properties of ATP-bound states with and without substrate; the symmetry between the direction of transport (import or export), and finally the symmetry of time-reversal, specifically by breaking the detailed balance and moving the system out of equilibrium. These symmetry breakings perfectly mirror the information-processing steps of measurement, storage, feedback and resetting that

we identified as integral parts of the transport cycle, exactly as in the Maxwell Demon thought experiment. Last, we built an energy balance for the transport cycle by ABC transporters, thus quantifying the energy and entropy variation associated to each information-processing step. Moreover, we were able to qualitatively reproduce the phenomenology of various experimental data, thus supporting the validity and robustness of our model.

In a nutshell, we showed, starting from a biochemical description, that ABC transporters are autonomous Maxwell Demons, in which all the information processing is constantly taking the form of successive molecular interactions.

In the second part, we moved to the study of a completely different transport mechanism, the translocation of a protein through a membrane by entropic pulling induced by Hsp70s.

First, the numerical simulations reveal that the translocation optimality is the consequence of a balance between, on the one hand, the affinity of Hsp70 for the substrate and, on the other hand, the exerted pulling force. This result led us to postulate the existence of a so far unobserved biological state of Hsp70 in the framework of translocation. Indeed, we showed numerically that the translocation is strongly accelerated under specific conditions if two chaperones bind to a same site to form a dimer, strengthening the pulling force. This is particularly so when the external conditions are not favourable to the translocation, for instance when the translocating substrate is misfolded.

Second, we studied the translocation through a simpler analytical model, explicating the different conditions which have to be satisfied to impose a directionality to the translocation. In a two site-model, we showed that the symmetry between kinetic properties of the two sites has to be broken, as well as the symmetry between ATP-state and ADP-state of Hsp70 on binding sites. Even more fundamentally, energy has to be brought to the system to move the system from equilibrium and induce a translocation directionality. In the last part, we showed that the translocation results from acquisition of information about the relative position of the membrane with respect to Hsp70, which is exactly the characteristic features of a (slightly revisited) Szilard Engine. The key message that arises from this work within the framework of information theory is the description of the translocation mechanism as an autonomous Szilard Engine. It is something new that was, as far as we know, not described in the literature.

All the results we produced in the two projects open exciting perspectives for future work, both in the direction of a more precise description for each of them, as well as with the aim of building a unified framework for cellular transporters and, by extension, many other active biological systems.

In particular, one interesting perspective for ABC Transporters would be to explore more deeply the non-linear effects that arise from the complexification of our model. Another is to study the reactivity to changes in external conditions, extending the observations from steady-state to the time-evolution during the transitional regime.

The main challenge in the translocation would be to extend the formalism of information

---

theory that we built in the second part of the project to the first and more accurate description of translocation. We believe that the associated mechanism is in fact a collection of "intricate" Szilard engines that act simultaneously to generate a more complex mechanism than the one analytically described in this work.

As a whole, we believe that all the conclusions drawn throughout the present work can be unified in a more global formalism. With that aim, a future work direction would be to establish a generic model to grasp through the formalism of information theory the characteristics of wider class of transport mechanisms, taking steps back from specific examples. In that sense, we aim to generalize kinetic transitions such as binding and unbinding, hydrolysis and synthesis as carrying and converting information inside the biological systems. As a corollary, it would also reveal the crucial nature of symmetry breaking as a necessary condition to perform any autonomous information-processing task at the molecular level.

In this work, we built a bridge between biological description of transport and information theory. Whereas many articles artificially introduce elements from the latter into the model, we showed that the biochemical characterisation naturally leads to the proper nature of biological systems, as information-processing devices. Thought experiments that acted as a starting point to the development of formalism of information theory arise at the end of this process. ABC transporters, the translocation channel and hypothetically most of active biological processes evolved for billions of years to autonomously process information at the molecular level, without interference of small external intelligent and *demoniac* beings.

This likely does not fully answer the question "*What is Life?*", but we believe to have provided in this work the evidence that information is one touchstone among others, that has been essential to the emergence and development of Life since its origin billions of years ago.





# **A Appendix**

## **A.1 ABC Transporters**

### **A.1.1 Analytical expression of all the rates**

All the transition rates in the model are expressed as a function of a set of independent parameters. The expression written in the right column of Table A.1 satisfy detailed balance constraints.

### **A.1.2 Numerical values of the rates**

If not stated otherwise, the numerical values used for the simulations are shown in table A.2.

## Appendix A. Appendix

Table A.1: Expression of the rates as a function of independent parameters, after introducing detailed balance constrains as well as  $\eta$ ,  $K_e$  and  $K_e^S$ . All the terms in the right column "Expression" can be defined independently.

Rate	Expression
$k_{on}^T$	$k_{on}^T$
$k_{on,H_D}^{(0)}$	$k_{on,H_D}^{(0)}$
$k_+$	$k_+$
$k_-$	$\frac{k_+}{K_e}$
$k_h$	$k_h$
$k_{ex}^{T \rightarrow D}$	$\frac{k_{-T}k_{+D}}{k_{+D} + \alpha k_{+T}}$
$k_h^*$	$\eta k_h$
$k_{ex}^{T^* \rightarrow D}$	$\frac{k_{-T}^*k_{+D}^*}{k_{+D}^* + \alpha k_{+T}^*}$
$k_h^S$	$k_h^S$
$k_{ex}^{TS \rightarrow DS}$	$\frac{k_{-T}^S k_{+D}^S}{k_{+D}^S + \alpha k_{+T}^S}$
$k_h^{*S}$	$\eta k_h^S$
$k_{ex}^{T^*S \rightarrow DS}$	$\frac{k_{-T}^{*S} k_{+D}^{*S}}{k_{+D}^{*S} + \alpha k_{+T}^{*S}}$

Rate	Expression
$k_{off}^T$	$k_{off}^T$
$k_{off}^D$	$k_{off}^D$
$k_+^S$	$k_+^S$
$k_-^S$	$\frac{k_+^S}{K_e^S}$
$k_s$	$\frac{[in]_{eq}}{[out]_{eq}} \frac{k_h k_{off}^T k_{on}^D k_s^S}{k_h^S k_{off}^D k_{on}^T}$
$k_{ex}^{D \rightarrow T}$	$\frac{[in]_{eq}}{[out]_{eq}} \frac{\alpha k_{-T} k_{off}^T k_{on}^D k_{+D} k_s^S}{\alpha_{eq} k_h^S k_{off}^D k_{on}^T (k_{+D} + \alpha k_{+T})}$
$k_s^*$	$\frac{[in]_{eq}}{[out]_{eq}} \frac{\eta K_e k_h k_{off}^T k_{on}^D k_s^S}{k_h^S k_{off}^D k_{on}^T}$
$k_{ex}^{D \rightarrow T^*}$	$\frac{[in]_{eq}}{[out]_{eq}} \frac{\alpha K_e k_{-T}^* k_{off}^T k_{on}^D k_{+D}^* k_s^S}{\alpha_{eq} k_h^S k_{off}^D k_{on}^T (k_{+D}^* + \alpha k_{+T}^*)}$
$k_s^S$	$k_s^S$
$k_{ex}^{DS \rightarrow TS}$	$\frac{\alpha k_{-T}^S k_{+D}^S k_s^S}{\alpha_{eq} k_h^S (k_{+D}^S + \alpha k_{+T}^S)}$
$k_s^{*S}$	$\eta K_e^S k_s^S$
$k_{ex}^{DS \rightarrow T^*S}$	$\frac{\alpha K_e^S k_{-T}^{*S} k_{+D}^{*S} k_s^S}{\alpha_{eq} k_h^S (k_{+D}^{*S} + \alpha k_{+T}^{*S})}$

Rate	Value	Rate	Value
$k_{on}^T$	$0.5 \mu M^{-1} s^{-1}$	$k_{off}^T$	$0.01 s^{-1}$
$k_{on}^D$	$0.5 \mu M^{-1} s^{-1}$	$k_{off}^D$	$0.01 s^{-1}$
$k_+$	$1 s^{-1}$	$k_+^S$	$1 s^{-1}$
$k_h$	$0.02 s^{-1}$	$k_h^S$	$0.02 s^{-1}$
		$k_s^S$	$2 \cdot 10^{-9} s^{-1}$
$k_{-T}$	$10^{-4} s^{-1}$	$k_{-T}^S$	$10^{-4} s^{-1}$
$k_{-T}^*$	$10^{-4} s^{-1}$	$k_{-T}^{*S}$	$10^{-4} s^{-1}$
$k_{+T}$	$0.5 s^{-1}$	$k_{+T}^S$	$0.5 s^{-1}$
$k_{+T}^*$	$0.5 s^{-1}$	$k_{+T}^{*S}$	$0.5 s^{-1}$
$k_{+D}$	$0.1 s^{-1}$	$k_{+D}^S$	$0.1 s^{-1}$
$k_{+D}^*$	$0.1 s^{-1}$	$k_{+D}^{*S}$	$0.1 s^{-1}$

Table A.2: Numerical values of the rates. The rates that are not defined in this table ensue from detailed balance conditions and thermodynamic constraints (see S.I., Table A.1).

### A.1.3 Sketch of the analytical derivation

We solve the equation 3.17 referring to the state probabilities  $P(TS)$  and  $P(T)$  but an analog derivation obviously holds also for  $P(DS)$  and  $P(D)$  leading exactly to the same final result.

Let's start by decomposing the state probabilities  $P(TS)$  and  $P(T)$  relatively to the set of spanning trees converging respectively to  $TS$  and  $T$ . The decomposition can be pushed further by splitting the system in two three-state systems and possible transitions in between.

In what follows,  $\Gamma^{(T)}$  refers to the set of oriented spanning trees toward the vertex  $T$  of the graph, similarly for all other states.  $\Gamma_{3,S}^{(DS)}$  is the restriction to the three-state system with substrate  $\{TS, T^*S, DS\}$  and denotes the set of spanning trees towards  $DS$  in this subsystem.

$\prod_{l \in \gamma} k_l(T)$  is the product of all the rates in the oriented spanning tree  $\gamma$ .

$$P(T) = \sum_{\gamma \in \Gamma^{(T)}} \prod_{l \in \gamma} k_l(T) \quad (\text{A.1})$$

$$= \sum_{\gamma \in \Gamma_{3,S}^{(TS)}} \prod_{l \in \gamma} k_l \cdot \sum_{\gamma \in \Gamma_3^{(T)}} \prod_{l \in \gamma} k_l \cdot k_{off}^T \quad (\text{A.2})$$

$$+ \sum_{\gamma \in \Gamma_{3,S}^{(DS)}} \prod_{l \in \gamma} k_l \cdot \sum_{\gamma \in \Gamma_3^{(T)}} \prod_{l \in \gamma} k_l \cdot k_{off}^D \quad (\text{A.3})$$

$$+ \sum_{\gamma \in \Gamma_{3,S}^{(TS)}} \prod_{l \in \gamma} k_l \cdot k_{off}^T [in] k_{on}^D (k^{T* \rightarrow D} + k_-) \quad (\text{A.4})$$

$$+ \sum_{\gamma \in \Gamma_3^{(T)}} \prod_{l \in \gamma} k_l \cdot k_{off}^T k_{off}^D (k_-^S + k^{T*S \rightarrow DS}) \quad (\text{A.5})$$

and similarly

$$P(TS) = \sum_{\gamma \in \Gamma^{(TS)}} \prod_{l \in \gamma} k_l(TS) \quad (\text{A.6})$$

$$= \sum_{\gamma \in \Gamma_3^{(T)}} \prod_{l \in \gamma} k_l \cdot \sum_{\gamma \in \Gamma_{3,S}^{(TS)}} \prod_{l \in \gamma} k_l \cdot [out] k_{on}^T \quad (\text{A.7})$$

$$+ \sum_{\gamma \in \Gamma_3^{(D)}} \prod_{l \in \gamma} k_l \cdot \sum_{\gamma \in \Gamma_{3,S}^{(TS)}} \prod_{l \in \gamma} k_l \cdot [in] k_{on}^D \quad (\text{A.8})$$

$$+ \sum_{\gamma \in \Gamma_3^{(T)}} \prod_{l \in \gamma} k_l \cdot [out] k_{on}^T k_{off}^D (k^{T*S \rightarrow DS} + k_-^S) \quad (\text{A.9})$$

$$+ \sum_{\gamma \in \Gamma_{3,S}^{(TS)}} \prod_{l \in \gamma} k_l \cdot [out] k_{on}^T [in] k_{on}^D (k_- + k^{T* \rightarrow D}) \quad (\text{A.10})$$

Substituting these two expressions in

$$P(TS)k_{off}^T - P(T)[in]k_{on}^T = 0$$

most of the terms cancel out and it remains

$$P(TS)k_{off}^T - P(T)[in]k_{on}^T \quad (\text{A.11})$$

$$= \sum_{\gamma \in \Gamma_{3,S}^{(DS)}} \prod_{l \in \gamma} k_l \cdot \sum_{\gamma \in \Gamma_3^{(T)}} \prod_{l \in \gamma} k_l \cdot k_{off}^D [out] k_{on}^T - \sum_{\gamma \in \Gamma_{3,S}^{(TS)}} \prod_{l \in \gamma} k_l \cdot \sum_{\gamma \in \Gamma_3^{(D)}} \prod_{l \in \gamma} k_l \cdot k_{off}^T [in] k_{on}^D \quad (\text{A.12})$$

Finally, imposing a zero flux between states  $T$  and  $TS$ , it leads to:

$$\frac{[in]}{[out]} = \frac{\sum_{\gamma \in \Gamma_{3,S}^{(DS)}} \prod_{l \in \gamma} k_l \cdot \sum_{\gamma \in \Gamma_3^{(T)}} \prod_{l \in \gamma} k_l \cdot k_{off}^D k_{on}^T}{\sum_{\gamma \in \Gamma_{3,S}^{(TS)}} \prod_{l \in \gamma} k_l \cdot \sum_{\gamma \in \Gamma_3^{(D)}} \prod_{l \in \gamma} k_l \cdot k_{off}^T k_{on}^D} = \frac{k_{off}^D k_{on}^T P_3(T) P_3(DS)}{k_{off}^T k_{on}^D P_3(TS) P_3(D)} \quad (\text{A.13})$$

This formulation is the most general formulation that is only built on the topology of the model. The next step is to push the analytical development further, by introducing the definition of the rates (table), thus taking into account the biochemical considerations that were introduced into the physical and biochemical conception of our model, through the explicit definition of the transition rates.

As a preliminary step, we will determine how the presence or absence of substrate affects each of the term in eq (), that is between  $P_3(T)$  and  $P_{3,S}(TS)$  as well as between  $P_3(D)$  and  $P_{3,S}(DS)$ .

$$k^{T^*S \rightarrow DS} k^{TS \rightarrow DS} = k^{T^* \rightarrow D} k^{T \rightarrow D} \quad (\text{A.14})$$

$$k_-^S k^{TS \rightarrow DS} = \frac{K_e}{K_e^S} k_- k^{T \rightarrow D} \quad (\text{A.15})$$

$$k^{T^*S \rightarrow DS} k_+^S = k^{T^* \rightarrow D} k_+ \quad (\text{A.16})$$

$$k^{T^*S \rightarrow DS} k^{DS \rightarrow TS} = \frac{[out]_{eq} k_{on}^T k_{off}^D}{[in]_{eq} k_{off}^T k_{on}^D} k^{T^* \rightarrow D} k^{D \rightarrow T} \quad (\text{A.17})$$

$$k_-^S k^{DS \rightarrow TS} = \frac{[out]_{eq} k_{on}^T k_{off}^D}{[in]_{eq} k_{off}^T k_{on}^D} \frac{K_e}{K_e^S} k_- k^{D \rightarrow T} \quad (\text{A.18})$$

$$k^{DS \rightarrow T^*S} k_+^S = \frac{[out]_{eq} k_{on}^T k_{off}^D}{[in]_{eq} k_{off}^T k_{on}^D} k^{D \rightarrow T^*} k_+ \quad (\text{A.19})$$

It follows two relations between  $P_3(T)$  and  $P_{3,S}(TS)$  as well as between  $P_3(D)$  and  $P_{3,S}(DS)$ .

$$P_3(DS) = P_3(D) + \left( \frac{K_e}{K_e^S} - 1 \right) k_- k^{T \rightarrow D} \quad (\text{A.20})$$

$$P_3(TS) = \frac{[out]_{eq} k_{on}^T k_{off}^D}{[in]_{eq} k_{off}^T k_{on}^D} \left( P_3(T) + \left( \frac{K_e}{K_e^S} - 1 \right) k_- k^{D \rightarrow T} \right) \quad (\text{A.21})$$

## Appendix A. Appendix

---

Combining equations A.20 and A.21 into equation A.13, we obtain:

$$\frac{[in]}{[out]} = \frac{[in]_{eq} \left[ P_3(D) + \left( \frac{K_e}{K_s^e} - 1 \right) k_- k^{T \rightarrow D} \right] P_3(T)}{[out]_{eq} \left[ P_3(T) + \left( \frac{K_e}{K_s^e} - 1 \right) k_- k^{D \rightarrow T} \right] P_3(D)} \quad (\text{A.22})$$

$$\frac{[in]/[out]}{[in]_{eq}/[out]_{eq}} - 1 = \frac{\left( \frac{K_e}{K_s^e} - 1 \right)}{\left[ P_3(T) + \left( \frac{K_e}{K_s^e} - 1 \right) k_- k^{D \rightarrow T} \right] P_3(D)} \{ k_- k^{T \rightarrow D} P_3(T) - k_- k^{D \rightarrow T} P_3(D) \} \quad (\text{A.23})$$

Let's now focus on the term between {...} in the previous equation

$$[k_- k^{T \rightarrow D} P_3(T) - k_- k^{D \rightarrow T} P_3(D)] \quad (\text{A.24})$$

$$= k_- k^{T \rightarrow D} (k_- k^{D \rightarrow T} + k_- k^{D \rightarrow T^*} + k^{T^* \rightarrow D} k^{D \rightarrow T^*}) \quad (\text{A.25})$$

$$- k_- k^{D \rightarrow T} (k_- k^{T \rightarrow D} + k_+ k^{T^* \rightarrow D} + k^{T^* \rightarrow D} k^{T \rightarrow D}) \quad (\text{A.26})$$

$$= k_- \cdot [k_- k^{T \rightarrow D} k^{D \rightarrow T^*} - k_+ k^{T^* \rightarrow D} k^{D \rightarrow T}] \quad (\text{A.27})$$

The transitions between  $T$  and  $D$  as well as between  $T^*$  and  $D$  have to be decomposed into the hydrolysis/synthesis and exchange. As a consequence from detailed balance, only the cycles that go through one hydrolysis/synthesis and one exchange contribute to the term, otherwise the product of rates are equal in both directions of the cycle. Thus the last expression can be rewritten as:

$$k_- \cdot [k_- k^{T \rightarrow D} k^{D \rightarrow T^*} - k_+ k^{T^* \rightarrow D} k^{D \rightarrow T}] \quad (\text{A.28})$$

$$= k_- \cdot [(k_- k_{ex}^{T \rightarrow D} k_s^* - k_+ k_h^* k_{ex}^{D \rightarrow T}) + (k_- k_h k_{ex}^{D \rightarrow T^*} - k_+ k_{ex}^{T^* \rightarrow D} k_s)] \quad (\text{A.29})$$

$$= k_- \cdot \left[ \left( \frac{\alpha_{eq}}{\alpha} - 1 \right) k_+ k_h^* k_{ex}^{D \rightarrow T} + \left( \frac{\alpha}{\alpha_{eq}} - 1 \right) k_+ k_s k_{ex}^{T^* \rightarrow D} \right] \quad (\text{A.30})$$

$$= k_- \left( \frac{\alpha}{\alpha_{eq}} - 1 \right) \left( k_+ k_s k_{ex}^{T^* \rightarrow D} - \frac{\alpha_{eq}}{\alpha} k_+ k_h^* k_{ex}^{D \rightarrow T} \right) \quad (\text{A.31})$$

The final step of the derivation consists in introducing from table A.1 the explicit form of the

rates. After a few calculations, we obtain:

$$k_- \left( \frac{\alpha}{\alpha_{eq}} - 1 \right) \left[ k_+ k_s k_{ex}^{T^* \rightarrow D} - \frac{\alpha_{eq}}{\alpha} k_+ k_h^* k_{ex}^{D \rightarrow T} \right] \quad (A.32)$$

$$= k_- \left( \frac{\alpha}{\alpha_{eq}} - 1 \right) \left[ k_+ k_s^S k_{ex}^{T^* \rightarrow D} \frac{[in]_{eq} k_{off}^T k_{on}^D}{[out]_{eq} k_{on}^T k_{off}^D} \left( 1 - \frac{k_+ k_h^* k_{ex}^{D \rightarrow T}}{k_- k_h k_{ex}^{D \rightarrow T^*}} \right) \right] \quad (A.33)$$

Finally, plugging back equation A.33 into eq A.23, we obtain the final expression:

$$\frac{[in]}{[out]} = \frac{[in]_{eq}}{[out]_{eq}} \left[ 1 + \left( \frac{\alpha}{\alpha_{eq}} - 1 \right) \cdot \left( \frac{K_e}{K_e^S} - 1 \right) \cdot \left( 1 - \frac{k_+ k_h^* k_{ex}^{D \rightarrow T}}{k_- k_h k_{ex}^{D \rightarrow T^*}} \right) \cdot \frac{k_+ k_- k_s^S k_{ex}^{T^* \rightarrow D}}{P_3(D) P_3(TS)} \right] \quad (A.34)$$

#### A.1.4 Energy balance for an exporter

As stated in the main text, the energy balance reported for an importer also holds for an exporter and all the physical quantities are defined analogously.

The available energy  $\Delta G$  is obviously the same, that is

$$\Delta G = k_B T \log \left( \frac{\alpha}{\alpha_{eq}} \right) \quad (A.35)$$

Along the export cycle, the substrate is bound on the in- side of the membrane (from  $D$  to  $DS$ , "measurement") and released on the out- side (from  $TS$  to  $T$ , resetting). The corresponding information associated to each of the two processes is:

$$k_B T I_{\text{measure}} = k_B T \ln \left( \frac{P_3(DS)}{P_3(D)} \right) \quad (A.36)$$

$$k_B T I_{\text{reset}} = k_B T \ln \left( \frac{P_3(T)}{P_3(TS)} \right) \quad (A.37)$$

The entropy dissipation is defined on a cycle going through hydrolysis in absence of substrate and exchange when a substrate is bound. The different routes for such a cycle are equivalent and we illustrate it through a single example:

- $DS \xrightarrow{\text{exchange}} TS$  after measurement and  $T \rightarrow T^* \xrightarrow{\text{hydrolysis}} D$  after resetting

$$\Delta S_{\text{feedback}} = k_B T \ln \left( \frac{k_+^S k_h^{*,S} k_{ex}^{D \rightarrow T}}{k_-^S k_s^{*,S} k_{ex}^{T \rightarrow D}} \right) = k_B T \ln \left( \frac{K_e k_s^{*,S} k_h^* \alpha}{K_e^S k_h^{*,S} k_s^* \alpha_{eq}} \right) \quad (A.38)$$

## Appendix A. Appendix

---

which is equal to:

$$T\Delta S_{\text{feedback}} = k_B T \ln \left( \frac{[\text{out}]_{eq} k_{off}^D k_{on}^T \alpha}{[\text{in}]_{eq} k_{on}^D k_{off}^T \alpha_{eq}} \right) \quad (\text{A.39})$$

Finally, the difference in chemical potential across the membrane at the end of the process is:

$$\Delta \mathcal{E} := k_B T \ln \left( \frac{[\text{out}]}{[\text{in}]} \bigg/ \frac{[\text{out}]_{eq}}{[\text{in}]_{eq}} \right) \quad (\text{A.40})$$

Combining the equations A.36, A.37, A.39, A.40, A.13, we can indeed show that the exact same energy balance relation holds for importers and exporters, that is:

$$\Delta \mathcal{E} = \Delta G - k_B T (I_{\text{measure}} + I_{\text{reset}}) - T\Delta S_{\text{feedback}} \quad (\text{A.41})$$

## A.2 Translocation

### A.2.1 Analytical expressions of the rates and numerical values of the parameters

The detailed balance constrains is imposed on the system so that all the rates are not independant the one from the other. In Table A.3 to A.5 are listed the expressions for the rates as a function of independant parameters.

### A.2.2 Numerical estimations of rates and parameters

#### Numerical estimation of $F_1$ :

The numerical estimation of the entropic pulling force  $F_1$  comes from equation 4.4, which is as a reminder

$$F_1 = \frac{3}{2} k_B T \frac{R^2}{bx_1^2} \quad (\text{A.42})$$

As a numerical estimation, we took  $k_B T = 4\text{pN} \cdot \text{nm}$ ,  $x_1 \approx 8$  residues  $\approx 2.5 - 3\text{nm}$ ,  $b \approx 1\text{nm}$  and  $R_{70} \approx 3\text{nm}$ . Hence



Hydrolysis-synthesis and exchange rates in substrate-bound states	
$k_h^{S_n}$	$\begin{cases} k_h^S \cdot \lambda & n = 1 \\ k_h^S & n > 1 \end{cases}$
$k_s^{S_n}$	$\begin{cases} \lambda \alpha_{eq} k_h^S \frac{k_{-D} k_{+T} k_{on,0}^{T-TS} k_{off,0}^{DS-D}}{k_{-T} k_{+D} k_{on,0}^{D-DS} k_{off,0}^{TS-T}} & n = 1 \\ \alpha_{eq} k_h^S \frac{k_{-D} k_{+T} k_{on,0}^{T-TS} k_{off,0}^{DS-D}}{k_{-T} k_{+D} k_{on,0}^{D-DS} k_{off,0}^{TS-T}} & n > 1 \end{cases}$
$k_{ex}^{S_n \cdot H_T \rightarrow S_n \cdot H_D}$	$\frac{k_{on,H_D}^{(0)} k_{off,H_T}^{(0)} k_{-T} k_{+D}}{k_{on,H_T}^{(0)} k_{off,H_D}^{(0)} k_{-D} k_{+T}} k_{-D} k_{+T} \frac{k_{+T}^{S_n}}{k_{+D}^{S_n} + \alpha k_{+T}^{S_n}}$
$k_{ex}^{S_n \cdot H_D \rightarrow S_n \cdot H_T}$	$k_{-D} \frac{\alpha k_{+T}^{S_n}}{k_{+D}^{S_n} + \alpha k_{+T}^{S_n}}$

Table A.3: Analytical expressions for the transition rates of Hsp70s bound on the substrate, as a function of independent parameters.

Hydrolysis-synthesis and exchange rates in solution	
$k_h$	$k_h$
$k_s$	$\alpha_{eq} k_h \frac{k_{-D} k_{+T}}{k_{-T} k_{+D}}$
$k_{ex}^{H_T \rightarrow H_D}$	$k_{-T} \frac{k_{+D}}{k_{+D} + \alpha k_{+T}}$
$k_{ex}^{H_D \rightarrow H_T}$	$k_{-D} \frac{\alpha k_{+T}}{k_{+D} + \alpha k_{+T}}$

Table A.4: Analytical expressions for the rates between  $H_T$  and  $H_D$  in solution, as a function of independent parameters.

**Appendix A. Appendix**

---

Binding, unbinding and translocation rates	
$k_{on,H_T}^{S_n}$	$\begin{cases} k_{on,H_T}^{(0)} \cdot \exp[-\beta f(x_n)] & \text{monomer} \\ k_{on,H_T}^{(0)} \cdot \exp[-\beta(f_2(x_n) - f(x_n))] & \text{dimer} \end{cases}$
$k_{off,H_T}^{S_n}$	$k_{off,H_T}^{(0)}$
$k_{on,H_D}^{S_n}$	$\begin{cases} k_{on,H_D}^{(0)} \cdot \exp[-\beta f(x_n)] & \text{monomer} \\ k_{on,H_D}^{(0)} \cdot \exp[-\beta(f_2(x_n) - f(x_n))] & \text{dimer} \end{cases}$
$k_{off,H_D}^{S_n}$	$k_{off,H_D}^{(0)}$
$k_{in}^{(n)}$	$\begin{cases} k_0 \exp[-\beta(f(x_{n+1}) - f(x_n))] & \text{monomer} \\ k_0 \exp[-\beta(f_2(x_{n+1}) - f_2(x_n))] & \text{dimer} \end{cases}$
$k_{out}^{(n)}$	$\begin{cases} 0 & n=1, \text{ monomer \& dimer} \\ k_0 & n>1, \text{ monomer \& dimer} \end{cases}$

Table A.5: Analytical expressions for the binding, unbinding and translocation rates as a function of independent parameters.

$$F_1 \approx 6 - 8 \text{pN} \quad (\text{A.43})$$

### Numerical estimation of $k_0$ :

There is no experimental data that provides a precise estimation for the value of the parameter  $k_0$ . Nevertheless, there are numerical models [164] corroborating experimental results on protein translocation [165] which provide a broad range of potentially acceptable values for the parameter  $k_0$  in our model. However, we have to point out that the assumptions of the model are very different from our model, so these values have to be considered even more carefully.

The set of inward and backward translocation rates is in the range  $2 - 20 \text{min}^{-1}$ , which corresponds to  $0.03 - 0.3 \text{s}^{-1}$ . In particular, far from the pore, an equal translocation rate in both directions is given by  $12 \text{min}^{-1} = 0.2 \text{s}^{-1}$ .

Second, we compared the difference between inward and backward translocation rate associated to chaperone binding. In Liebermeister2001, the net effect is an inward translocation with a rate of  $21 \text{min}^{-1}$ . In our model, taking  $F_1 \approx 6 - 8 \text{pN}$  leads to a net effect associated to the binding on the first site equal to  $\approx 20 - 45 \cdot k_0$ . Thus, we get  $k_0 \approx 0.4 - 1 \text{min}^{-1} \approx 7 \cdot 10^{-3} - 2 \cdot 10^{-2} \text{s}^{-1}$ .

Hence, we keep the broad range

$$k_0 \approx 5 \cdot 10^{-3} - 5 \cdot 10^{-1} \text{s}^{-1} \quad (\text{A.44})$$

### Numerical values of the parameters

Besides  $F_1$  and  $k_0$ , all the other independent parameters that are part of the expressions in tables A.3 to A.5 are listed in table A.6.

### A.2.3 Full Matrix in the simple case $N_{site} = 2$

Let's consider only two sites on the protein in the simulation volume. The basis of protein sequences is the following:

$$\begin{aligned} \vec{\Psi} = \{ & (S_1 \cdot H_T, S_2 \cdot H_T), (S_1 \cdot H_T, S_2 \cdot H_D), (S_1 \cdot H_T, S_2), (S_1 \cdot H_D, S_2 \cdot H_T), \\ & (S_1 \cdot H_D, S_2 \cdot H_D), (S_1 \cdot H_D, S_2), (S_1, S_2 \cdot H_T), (S_1, S_2 \cdot H_D), (S_1, S_2) \} \end{aligned} \quad (\text{A.45})$$

The matrix  $\underline{\underline{M}}_k$  with elements  $M_k^{(i,j)}$  defined according to expression 4.23 is given by equations

**Appendix A. Appendix**

---

Numerical value of independent parameters	
Parameter	Num. value
$\alpha$	10
$\alpha_{eq}$	$10^{-9}$
$k_{-T}$	$1.33 \cdot 10^{-4} \text{s}^{-1}$
$k_{-D}$	$0.022 \text{s}^{-1}$
$k_{+T}$	$0.13 \mu\text{M}^{-1} \text{s}^{-1}$
$k_{+D}$	$0.267 \mu\text{M}^{-1} \text{s}^{-1}$
$k_{-D}^S$	$0.022 \text{s}^{-1}$
$k_{+T}^S$	$0.13 \mu\text{M}^{-1} \text{s}^{-1}$
$k_{+D}^S$	$0.267 \mu\text{M}^{-1} \text{s}^{-1}$
$k_h$	$6 \cdot 10^{-4} \text{s}^{-1}$
$k_h^S$	$10^{-3} \text{s}^{-1}$
$\lambda$	1800
$k_{off,H_D}^{(0)}$	$4.7 \cdot 10^{-4} \text{s}^{-1}$
$k_{off,H_T}^I$	$2 \text{s}^{-1}$
$k_{on,H_D}^{(0)}$	$10^{-3} \mu\text{M}^{-1} \text{s}^{-1}$
$k_{on,H_T}^{(0)}$	$0.45 \mu\text{M}^{-1} \text{s}^{-1}$
$x_1$	$8 \cdot 0.3 \text{nm}$
$\Delta x$	$35 \cdot 0.3 \text{nm}$

Table A.6: The numerical values are from different references in the literature [126, 166, 167, 168].

A.46 (non-diagonal elements) and A.47 (diagonal elements). To shorten the notation in the matrix, we define  $S_i \cdot H_X =: X_i$ .

$$\left( \begin{array}{cccccccccc} M_{1,1} & k_{tot}^{D_2 \rightarrow T_2} & [H_T]k_{on,H_T}^{S_2} & k_{tot}^{D_1 \rightarrow T_1} & 0 & 0 & [H_T]k_{on,H_T}^{S_1} & 0 & 0 \\ k_{tot}^{T_2 \rightarrow D_2} & M_{2,2} & [H_D]k_{on,H_D}^{S_2} & 0 & k_{tot}^{D_1 \rightarrow T_1} & 0 & 0 & [H_T]k_{on,H_T}^{S_1} & 0 \\ k_{off,H_T}^{S_2} & k_{off,H_D}^{S_2} & M_{3,3} & 0 & 0 & k_{tot}^{D_1 \rightarrow T_1} & k_{b,H_T}^{(2)} & 0 & [H_T]k_{on,H_T}^{S_1} \\ k_{tot}^{T_1 \rightarrow D_1} & 0 & 0 & M_{4,4} & k_{tot}^{D_2 \rightarrow T_2} & [H_T]k_{on,H_T}^{S_2} & [H_D]k_{on,H_D}^{S_1} & 0 & 0 \\ 0 & k_{tot}^{T_1 \rightarrow D_1} & 0 & k_{tot}^{T_2 \rightarrow D_2} & M_{5,5} & [H_D]k_{on,H_D}^{S_2} & 0 & [H_D]k_{on,H_D}^{S_1} & 0 \\ 0 & 0 & k_{tot}^{T_1 \rightarrow D_1} & k_{off,H_T}^{S_2} & k_{off,H_D}^{S_2} & M_{6,6} & 0 & k_{H_D}^{(2)} & [H_D]k_{on,H_D}^{S_1} \\ k_{off,H_T}^{S_1} + k_{f,H_T}^{(1)} & k_{f,H_T}^{(1)} & k_{f,H_T}^{(1)} & k_{off,H_D}^{S_1} & 0 & 0 & M_{7,7} & k_{tot}^{D_2 \rightarrow T_2} & [H_T]k_{on,H_T}^{S_2} \\ 0 & k_{off,H_T}^{S_1} & 0 & k_{f,H_D}^{(1)} & k_{off,H_D}^{S_1} + k_{f,H_D}^{(1)} & k_{f,H_D}^{(1)} & T_2 \rightarrow D_2 \\ 0 & 0 & k_{off,H_T}^{S_1} & 0 & 0 & k_{off,H_D}^{S_1} & k_{off,H_T}^{S_2} + k_{f,H_T}^{(2)} & M_{8,8} & [H_D]k_{on,H_D}^{S_2} \\ & & & & & & & & M_{9,9} \end{array} \right) \quad (\text{A.46})$$

Where the diagonal elements are

$$\begin{aligned} M_{1,1} &= -(k_{tot}^{T_2 \rightarrow D_2} + k_{off,H_T}^{S_2} + k_{tot}^{T_1 \rightarrow D_1} + k_{off,H_T}^{S_1} + k_{f,H_T}^{(1)}) \\ M_{2,2} &= -(k_{tot}^{D_2 \rightarrow T_2} + k_{off,H_D}^{S_2} + k_{tot}^{T_1 \rightarrow D_1} + k_{f,H_T}^{(1)} + k_{off,H_T}^{(1)}) \\ M_{3,3} &= -([H_T]k_{on,H_T}^{S_2} + [H_D]k_{on,H_D}^{S_2} + k_{tot}^{T_1 \rightarrow D_1} + k_{f,H_T}^{(1)} + k_{off,H_T}^{S_1}) \\ M_{4,4} &= -(k_{tot}^{D_1 \rightarrow T_1} + k_{tot}^{T_2 \rightarrow D_2} + k_{off,H_T}^{S_2} + k_{off,H_D}^{S_1} + k_{f,H_D}^{S_1}) \\ M_{5,5} &= -(k_{tot}^{D_1 \rightarrow T_1} + k_{tot}^{D_2 \rightarrow T_2} + k_{off,H_D}^{S_2} + k_{off,H_D}^{S_1} + k_{f,H_D}^{S_1}) \\ M_{6,6} &= -(k_{tot}^{D_1 \rightarrow T_1} + [H_T]k_{on,H_T}^{S_2} + [H_D]k_{on,H_D}^{S_2} + k_{f,H_D}^{(1)} + k_{off,H_D}^{S_1}) \\ M_{7,7} &= -([H_T]k_{on,H_T}^{S_1} + k_{b,H_T}^{(2)} + [H_D]k_{on,H_D}^{S_1} + k_{tot}^{T_2 \rightarrow D_2} + k_{off,H_T}^{S_2} + k_{b,H_T}^{(2)}) \\ M_{8,8} &= -([H_T]k_{on,H_T}^{S_1} + [H_D]k_{on,H_D}^{S_1} + k_{b,H_D}^{(2)} + k_{tot}^{D_2 \rightarrow T_2} + k_{off,H_D}^{S_2} + k_{f,H_D}^{(2)}) \\ M_{9,9} &= -([H_T]k_{on,H_T}^{S_1} + [H_D]k_{on,H_D}^{S_1} + [H_T]k_{on,H_T}^{S_2} + [H_D]k_{on,H_D}^{S_2}) \end{aligned} \quad (\text{A.47})$$

#### A.2.4 Dependence of the system on $\gamma$

The choice of the numerical value of parameter  $\gamma$  is not so trivial. We show in figure A.1A and A.1B how the translocation rate  $R^{(2)}$  depends on  $\gamma$ . Obviously, the rate  $R$  does not depend on  $\gamma$ , since it refers to the initial model with only a single possible binding on each site. The horizontal dashed lines serve as reference value. For high diffusion rates  $k_0$  ( $k_0 = 10^{-2}\text{s}^{-1}$ , figure A.1A), as the dimerization does not play a significant role, the dependence of  $R^{(2)}$  on  $\gamma$  is negligible. For slower diffusive process ( $k_0 = 10^{-5}\text{s}^{-1}$ , figure A.1B), the dependence is much stronger. However, we expect that varying  $\gamma$  in a range between 1.8 and 2.2 does not significantly modify the qualitative interpretation of the results. Therefore, all the simulations are performed with a unique value of  $\gamma = 2$ .

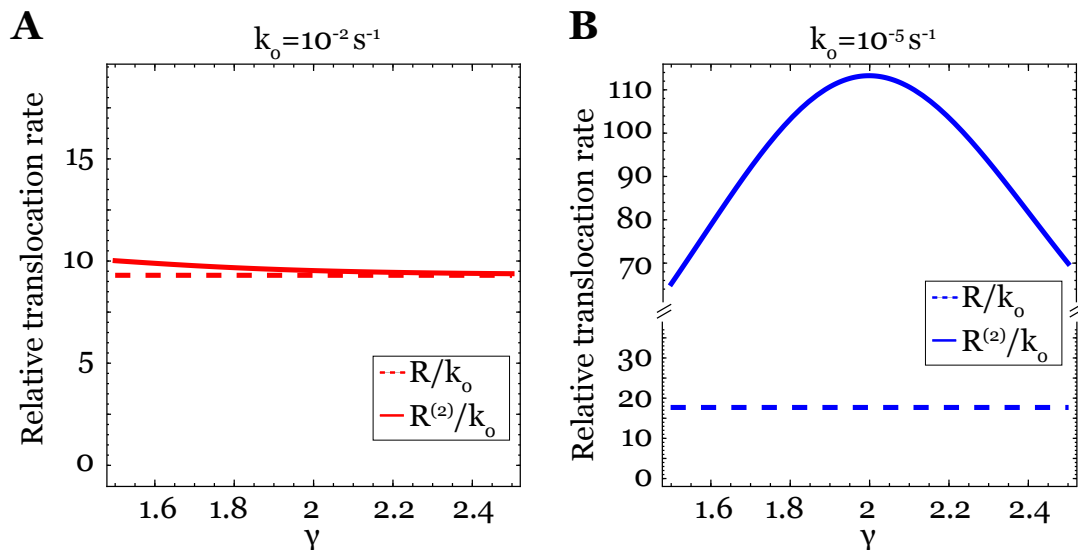


Figure A.1: Relative translocation rate as a function of  $\gamma$ . The reference value ( $R/k_0$ , independent of  $\gamma$ ) is represented with dashed lines. The relative translocation rate  $R^{(2)}/k_0$  that depends on  $\gamma$  is a continuous line.  $F_1 = 6\text{pN}$ ,  $[ATP]/[ADP] = 10$ . and (A)  $k_0 = 10^{-2}\text{s}^{-1}$ , (B)  $k_0 = 10^{-5}\text{s}^{-1}$ .

### A.2.5 Implementation of a Kinetic Monte Carlo algorithm

As mentioned in the main text, the model was first solved by using with a Kinetic Monte Carlo algorithm, before turning to the resolution of a linear system of equations to get the translocation rate at steady-state. Although the switch from one method to the other was made early along the study of our model, we show in figure A.2 that both methods provide the same result.

With our hypotheses and quantity of interest that we defined, it was not useful to go on with the KMC algorithm. However, clearly, this is a much more powerful (although way more computationally expensive) technique if we aim to extend the model with other objectives. For instance, it might be interesting to study the transient regime in which the system evolves towards the steady-state of the translocation, or, even more realistic, to include in the KMC algorithm the modelisation of the boundary effects to study the insertion of the protein at the beginning of the translocation.

### A.2.6 Analytical development for the translocation rate in the simple model

Applying formula 2.22 to the set of parameters 4.38 to 4.40 (translocation with  $H_T$ ) and 4.41 to 4.43 (translocation with  $H_D$ ), there are in total 12 contributing terms that will be split in two

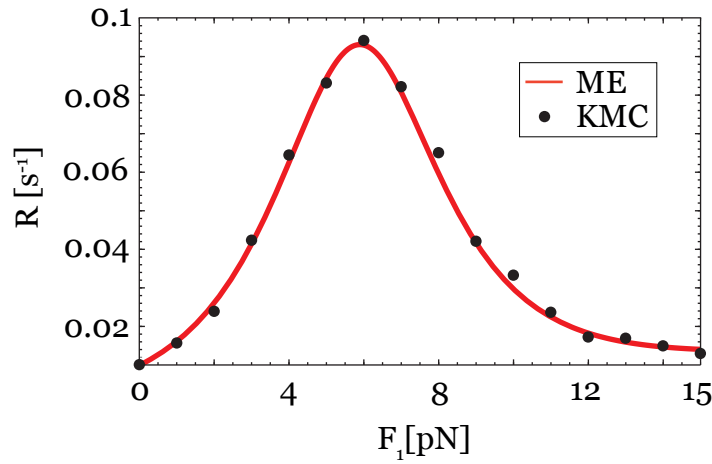


Figure A.2: Comparison between the Kinetic Monte Carlo algorithm ("KMC", black dots) and the solution obtained by solving the Master equation ("ME", red line). The parameters of the simulation are the same than in Figure 4.8 (green curve,  $k_0 = 10^{-2} s^{-1}$ ).

categories for a first analytical development. There are the two first terms in each set that go through the five states and for which  $w_c = 1$ . We will deal with them later (equations A.63 and A.64). We start to focus on the other terms, the one represented by a triangular-shaped cycle.

### Preliminary remark - notation in this development

- Exchange rates on the substrate, independant on the binding site:  $k_{ex,S}^{T \rightarrow D}$  and  $k_{ex,S}^{D \rightarrow T}$ ;
- Diffusion rates  $k_0^T$  (respectively  $k_0^D$ ) when Hsp70 in ATP-state (resp ADP-state) is bound on the substrate;

### Step 1: Combine contributions of similar cycles with hydrolysis/synthesis and exchange

Let's first introduce a notation to shorten the following expressions. The notation  $\Sigma_{X_i}$  denotes the term in  $w_c$  associated to the converging flux from state  $S_i \cdot H_X$  to the "triangular" cycle that does not contain  $S_i \cdot H_X$ . With this notation, let's combine the terms three and four in the

## Appendix A. Appendix

summation over cycles illustrated in 4.38.

$$\begin{array}{c} \begin{array}{c} \rightarrow \\ \uparrow \leftarrow \\ \leftarrow \end{array} \Sigma_{D_2} \left( 1 - \frac{P_T P_D^{eq}}{P_D P_T^{eq}} \right) + \begin{array}{c} \rightarrow \\ \leftarrow \uparrow \\ \leftarrow \end{array} \Sigma_{D_2} \left( 1 - \frac{P_T P_D^{eq} \alpha_{eq}}{P_D P_T^{eq} \alpha} \right) \end{array} \quad (A.48)$$

$$= \begin{array}{c} \rightarrow \\ \uparrow \leftarrow \\ \leftarrow \end{array} \Sigma_{D_2} \cdot \left[ \left( 1 - \frac{P_T P_D^{eq}}{P_D P_T^{eq}} \right) + \frac{k_{ex}^{D-T}}{\gamma \lambda k_h^S} \left( 1 - \frac{P_T P_D^{eq} \alpha_{eq}}{P_D P_T^{eq} \alpha} \right) \right] \quad (A.49)$$

$$= \begin{array}{c} \rightarrow \\ \uparrow \leftarrow \\ \leftarrow \end{array} \Sigma_{D_2} \cdot \left[ \left( 1 + \frac{k_{ex}^{D-T}}{\gamma \lambda k_h^S} \right) - \frac{P_T P_D^{eq}}{P_D P_T^{eq}} \left( 1 + \frac{\alpha_{eq} k_{ex}^{D-T}}{\alpha \gamma \lambda k_h^S} \right) \right] \quad (A.50)$$

Similarly, we get three other terms for the other pair of cycles, thus giving the four terms in total:

$$\begin{array}{c} \rightarrow \\ \leftarrow \uparrow \\ \leftarrow \end{array} \Sigma_{D_2} \cdot \left[ \left( 1 + \frac{k_{ex}^{D-T}}{\gamma \lambda k_h^S} \right) - \frac{P_T P_D^{eq}}{P_D P_T^{eq}} \left( 1 + \frac{\alpha_{eq} k_{ex}^{D-T}}{\alpha \gamma \lambda k_h^S} \right) \right] \quad (A.51)$$

$$\begin{array}{c} \rightarrow \\ \leftarrow \downarrow \\ \leftarrow \end{array} \Sigma_{D_1} \cdot \left[ \left( 1 + \frac{\alpha_{eq} k_{ex}^{D-T}}{\alpha \gamma k_h^S} \right) - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \left( 1 + \frac{k_{ex}^{D-T}}{\gamma k_h^S} \right) \right] \quad (A.52)$$

$$\begin{array}{c} \leftarrow \\ \downarrow \rightarrow \\ \leftarrow \end{array} \Sigma_{T_2} \cdot \left[ \left( 1 + \frac{\alpha_{eq} k_{ex}^{D-T}}{\alpha \gamma \lambda k_h^S} \right) - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \left( 1 + \frac{k_{ex}^{D-T}}{\gamma \lambda k_h^S} \right) \right] \quad (A.53)$$

$$\begin{array}{c} \leftarrow \\ \leftarrow \uparrow \\ \leftarrow \end{array} \Sigma_{T_1} \cdot \left[ \left( 1 + \frac{k_{ex}^{D-T}}{\gamma k_h^S} \right) - \frac{P_T P_D^{eq}}{P_D P_T^{eq}} \left( 1 + \frac{\alpha_{eq} k_{ex}^{D-T}}{\alpha \gamma k_h^S} \right) \right] \quad (A.54)$$

### Step 2: Summing the flux terms associated to a translocation in the $H_T$ state (and similarly with $H_D$ )

We now sum the equations associated to the net flux when  $H_T$  (respectively  $H_D$ ) is bound on the protein, that is equations (A.51)+(A.52) (resp. (A.53)+(A.54)), using the relations between the ratio of cycles:

$$\begin{array}{c} \rightarrow \\ \uparrow \leftarrow \\ \leftarrow \end{array} / \begin{array}{c} \rightarrow \\ \leftarrow \downarrow \\ \leftarrow \end{array} = \lambda \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \quad \text{and} \quad \begin{array}{c} \leftarrow \\ \leftarrow \uparrow \\ \leftarrow \end{array} / \begin{array}{c} \leftarrow \\ \downarrow \rightarrow \\ \leftarrow \end{array} = \frac{1}{\lambda} \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \quad (A.55)$$

$$(A.51) + (A.52) \rightarrow \begin{array}{c} \rightarrow \\ \leftarrow \downarrow \\ \leftarrow \end{array} \Sigma_{D_1} \left\{ \left[ \left( 1 - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) \left( 1 - \lambda \frac{\Sigma_{D_2}}{\Sigma_{D_1}} \right) \right] + \frac{k_{ex}^{D-T}}{\gamma k_h^S} \left[ \left( \frac{\alpha_{eq}}{\alpha} - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) \left( 1 - \frac{\Sigma_{D_2}}{\Sigma_{D_1}} \right) \right] \right\} \quad (A.56)$$



$$(A.53) + (A.54) \rightarrow \downarrow \begin{array}{c} \nearrow \\ \leftarrow \end{array} \Sigma_{T_2} \left\{ \left[ \left( 1 - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) \left( 1 - \frac{1}{\lambda} \frac{\Sigma_{T_1}}{\Sigma_{T_2}} \right) \right] + \frac{k_{ex}^{D \rightarrow T}}{\gamma k_h^S} \frac{1}{\lambda} \left[ \left( \frac{\alpha_{eq}}{\alpha} - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) \left( 1 - \frac{\Sigma_{T_1}}{\Sigma_{T_2}} \right) \right] \right\} \quad (A.57)$$

**Step 3: Compute the total flux, either with  $H_D$  or  $H_T$**

The same method is used, that is computing the sum using the ratio between the two cycles. Precisely, the ratio of interest here is:

$$\downarrow \begin{array}{c} \nearrow \\ \leftarrow \end{array} / \begin{array}{c} \nearrow \\ \leftarrow \\ \downarrow \end{array} = \lambda \frac{k_0^D}{k_0^T} \quad (A.58)$$

Giving:

$$(A.56) + (A.57) \rightarrow \begin{array}{c} \nearrow \\ \leftarrow \end{array} \left\{ \left( 1 - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) \left[ \Sigma_{D_1} - \lambda \Sigma_{D_2} + \frac{k_0^D}{k_0^T} \lambda \Sigma_{T_2} - \frac{k_0^D}{k_0^T} \Sigma_{T_1} \right] + \frac{k_{ex}^{D \rightarrow T}}{\gamma k_h^S} \left( \frac{\alpha_{eq}}{\alpha} - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) \left[ \Sigma_{D_1} - \Sigma_{D_2} + \frac{k_0^D}{k_0^T} \Sigma_{T_2} - \frac{k_0^D}{k_0^T} \Sigma_{T_1} \right] \right\} \quad (A.59)$$

We go one step further in the development of the expression, by focusing on the terms between [...]. They can be rewritten as follows :

$$\left[ \Sigma_{D_1} - \lambda \Sigma_{D_2} + \frac{k_0^D}{k_0^T} \lambda \Sigma_{T_2} - \frac{k_0^D}{k_0^T} \Sigma_{T_1} \right] = (-1 + \lambda) \left( k_{ex}^{D \rightarrow T} \left( \frac{\alpha_{eq} k_0^D}{\alpha k_0^T \gamma} - 1 \right) + k_{off, H_D}^{(0)} \left( \frac{k_{off, H_T}^{(0)} k_0^D}{k_{off, H_D}^{(0)} k_0^T} - 1 \right) \right) \quad (A.60)$$

$$\left[ \Sigma_{D_1} - \Sigma_{D_2} + \frac{k_0^D}{k_0^T} \Sigma_{T_2} - \frac{k_0^D}{k_0^T} \Sigma_{T_1} \right] = (-1 + \lambda) k_h^S \left( \gamma - \frac{k_0^D}{k_0^T} \right) \quad (A.61)$$

Substituting into equation A.59, we get:

$$(A.59) \rightarrow \begin{array}{c} \nearrow \\ \leftarrow \end{array} (\lambda - 1) \left\{ \left( 1 - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) \left( k_{ex}^{D \rightarrow T} \left( \frac{\alpha_{eq} k_0^D}{\alpha k_0^T \gamma} - 1 \right) + k_{off, H_D}^{(0)} \left( \frac{k_{off, H_T}^{(0)} k_0^D}{k_{off, H_D}^{(0)} k_0^T} - 1 \right) \right) + \frac{k_{ex}^{D \rightarrow T}}{\gamma k_h^S} \left( \frac{\alpha_{eq}}{\alpha} - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) k_h^S \left( \gamma - \frac{k_0^D}{k_0^T} \right) \right\} \quad (A.62)$$

**Step 4: Contribution of the "pentagonal" cycles through the five states**

We now turn to the cycles that go through all the states, which are the two first in the lists (4.41)

## Appendix A. Appendix

and (4.38) to compute the expression:

$$\left( \begin{array}{c} \longrightarrow \\ \swarrow \quad \searrow \\ \downarrow \end{array} \right) (1 - \frac{\alpha_{eq}}{\alpha}) + \left( \begin{array}{c} \longrightarrow \\ \swarrow \quad \searrow \\ \downarrow \end{array} \right) (1 - \frac{\alpha}{\alpha_{eq}}) + \left( \begin{array}{c} \swarrow \quad \searrow \\ \longrightarrow \\ \downarrow \end{array} \right) (1 - \frac{\alpha}{\alpha_{eq}}) + \left( \begin{array}{c} \swarrow \quad \searrow \\ \longrightarrow \\ \downarrow \end{array} \right) (1 - \frac{\alpha_{eq}}{\alpha}) \quad (A.63)$$

We rewrite each cycle using its ratio with the cycle that was arbitrarily used as a "reference" so far ( $\begin{array}{c} \longrightarrow \\ \swarrow \quad \searrow \\ \downarrow \end{array}$ ), so that the summation with equation A.62 will be easier. We get:

$$(A.63) \rightarrow \begin{array}{c} \longrightarrow \\ \swarrow \quad \searrow \\ \downarrow \end{array} (1 - \frac{\alpha_{eq}}{\alpha}) (1 - \lambda) k_{ex}^{D-T} \left[ \frac{P_D k_{on,H_D}^{(0)}}{P_T k_{on,H_T}^{(0)}} - \frac{k_0^D k_{off,H_T}^{(0)}}{k_0^T k_{off,H_D}^{(0)}} \right] \quad (A.64)$$

### Step 5: Sum of the two contributions to get the final expression

The last step of the derivation is to sum equations A.62 and A.64 to get the net translocation rate  $R$  defined in equation 4.34. This results from the sum of all the 12 terms in 4.41 and 4.38 with corresponding weights. The last few steps are:

$$R = \begin{array}{c} \longrightarrow \\ \swarrow \quad \searrow \\ \downarrow \end{array} (\lambda - 1) \left\{ \left( 1 - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) \left( k_{ex}^{D-T} \left( \frac{\alpha_{eq} k_0^D}{\alpha k_0^T \gamma} - 1 \right) + k_{off,H_D}^{(0)} \left( \frac{k_{off,H_T}^{(0)} k_0^D}{k_{off,H_D}^{(0)} k_0^T} - 1 \right) \right) \right. \\ \left. + \frac{k_{ex}^{D-T}}{\gamma k_h^S} \left( \frac{\alpha_{eq}}{\alpha} - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) k_h^S \left( \gamma - \frac{k_0^D}{k_0^T} \right) \right. \\ \left. + (-1) \cdot \left( 1 - \frac{\alpha_{eq}}{\alpha} \right) k_{ex}^{D-T} \left( \frac{P_D k_{on,H_D}^{(0)}}{P_T k_{on,H_T}^{(0)}} - \frac{k_0^D k_{off,H_T}^{(0)}}{k_0^T k_{off,H_D}^{(0)}} \right) \right\} \quad (A.65)$$

$$= \begin{array}{c} \longrightarrow \\ \swarrow \quad \searrow \\ \downarrow \end{array} (\lambda - 1) \left\{ \left( 1 - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) k_{off,H_D}^{(0)} \left( \frac{k_{off,H_T}^{(0)} k_0^D}{k_{off,H_D}^{(0)} k_0^T} - 1 \right) \right. \\ \left. + k_{ex}^{D-T} \cdot \left( \frac{k_{off,H_T}^{(0)} k_0^D}{k_{off,H_D}^{(0)} k_0^T} - 1 \right) \left( 1 - \frac{\alpha_{eq}}{\alpha} \right) \left( 1 + \frac{P_D k_{on,H_D}^{(0)}}{P_T k_{on,H_T}^{(0)}} \right) \right\} \quad (A.66)$$

If we furthermore note that

$$\left( 1 - \frac{P_D P_T^{eq}}{P_T P_D^{eq}} \right) = \left( 1 - \frac{\alpha_{eq}}{\alpha} \right) \cdot \frac{\frac{\alpha}{\alpha_{eq}} k_{ex}^{T-D}}{k_h + \frac{\alpha}{\alpha_{eq}} k_{ex}^{T-D}} \quad (A.67)$$

we finally obtain the following expression for the translocation rate:

$$R = \begin{array}{c} \longrightarrow \\ \swarrow \quad \searrow \\ \downarrow \end{array} (\lambda - 1) \cdot \left( \frac{k_{off,H_T}^{(0)} k_0^D}{k_{off,H_D}^{(0)} k_0^T} - 1 \right) \cdot \left( 1 - \frac{\alpha_{eq}}{\alpha} \right) \cdot \left[ k_{off,H_D}^{(0)} \cdot \frac{\frac{\alpha}{\alpha_{eq}} k_{ex}^{T-D}}{k_h + \frac{\alpha}{\alpha_{eq}} k_{ex}^{T-D}} + k_{ex}^{D_i \rightarrow T_i} \left( \frac{P_D k_{on,H_D}^{(0)}}{P_T k_{on,H_T}^{(0)}} + 1 \right) \right] \quad (A.68)$$

### A.2.7 Translocation machinery as an information processing device

#### Derivation of equation 4.53

$$R = P(S_1 \cdot H_T) \cdot k_{f,H_T}^{(1)} - P(S_2 \cdot H_T) \cdot k_{b,H_T}^{(2)} + P(S_1 \cdot H_D) \cdot k_{f,H_D}^{(1)} - P(S_2 \cdot H_D) \cdot k_{b,H_D}^{(2)} \quad (\text{A.69})$$

$$\begin{aligned}
R = \frac{1}{\Sigma_{TOT}} & \left[ k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_s^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_{f,H_D}^{(1)} \right. \\
& - k_{tot}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} k_s^{S_1 \cdot H_T \rightarrow S_2 \cdot H_T} k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_{b,H_D}^{(2)} \\
& + k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_{b,H_T}^{(2)} k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_{f,H_D}^{(1)} \\
& - k_{tot}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} k_{f,H_T}^{(1)} k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_{b,H_D}^{(2)} \\
& + k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_s^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D} k_{tot}^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{f,H_T}^{(1)} \\
& - k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_s^{S_1 \cdot H_D \rightarrow S_2 \cdot H_D} k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_{b,H_T}^{(2)} \\
& + k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_{b,H_D}^{(2)} k_{tot}^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{f,H_T}^{(1)} \\
& - k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_s^{S_1 \cdot H_D \rightarrow S_2 \cdot H_D} k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_{b,H_T}^{(2)} \\
& + (k_s^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D} k_{f,H_D}^{(1)} - k_s^{S_1 \cdot H_D \rightarrow S_2 \cdot H_D} k_{b,H_D}^{(2)}) \cdot (\dots) \\
& \left. + (k_s^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{f,H_T}^{(1)} - k_s^{S_1 \cdot H_T \rightarrow S_2 \cdot H_T} k_{b,H_T}^{(2)}) \cdot (\dots) \right] \quad (\text{A.70})
\end{aligned}$$

where  $\Sigma_{TOT}$  is the sum of all the spanning trees over the four-state kinetic model.

$$\begin{aligned}
R = \frac{1}{\Sigma_{TOT}} & \left[ k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_s^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_{f,H_D}^{(1)} \right. \\
& - k_{tot}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} k_s^{S_1 \cdot H_T \rightarrow S_2 \cdot H_T} k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_{b,H_D}^{(2)} \\
& + k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_s^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D} k_{tot}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} k_{f,H_T}^{(1)} \\
& \left. - k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_s^{S_1 \cdot H_D \rightarrow S_2 \cdot H_D} k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_{b,H_T}^{(2)} \right] \quad (\text{A.71})
\end{aligned}$$

$$\begin{aligned}
 R = & \frac{1}{\Sigma_{TOT}} \left[ k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_s^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_{f, H_D}^{(1)} \right] \\
 & \cdot \left( 1 - \frac{P_{S_1}(S_1 \cdot H_T) P_{S_2}(S_2 \cdot H_D)}{P_{S_1}(S_1 \cdot H_D) P_{S_2}(S_2 \cdot H_T)} \frac{k_{b, H_D}^{(2)} k_s^{S_1 \cdot H_T \rightarrow S_2 \cdot H_T}}{k_{f, H_D}^{(1)} k_s^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T}} \right) \\
 & + \frac{1}{\Sigma_{TOT}} \left[ k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_s^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D} k_{tot}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} k_{f, H_T}^{(1)} \right] \\
 & \cdot \left( 1 - \frac{P_{S_1}(S_1 \cdot H_D) P_{S_2}(S_2 \cdot H_T)}{P_{S_1}(S_1 \cdot H_T) P_{S_2}(S_2 \cdot H_D)} \frac{k_{b, H_T}^{(2)} k_s^{S_1 \cdot H_D \rightarrow S_2 \cdot H_D}}{k_{f, H_T}^{(1)} k_s^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D}} \right)
 \end{aligned} \tag{A.72}$$

$$\begin{aligned}
 R = & \frac{1}{\Sigma_{TOT}} \left[ k_{tot}^{S_1 \cdot H_T \rightarrow S_1 \cdot H_D} k_s^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T} k_{tot}^{S_2 \cdot H_D \rightarrow S_2 \cdot H_T} k_{f, H_D}^{(1)} \right] \\
 & \cdot \left( 1 - \frac{P_{S_1}(S_1 \cdot H_T) P_{S_2}(S_2 \cdot H_D)}{P_{S_1}(S_1 \cdot H_D) P_{S_2}(S_2 \cdot H_T)} \cdot \frac{k_{b, H_D}^{(2)} k_s^{S_1 \cdot H_T \rightarrow S_2 \cdot H_T}}{k_{f, H_D}^{(1)} k_s^{S_2 \cdot H_T \rightarrow S_1 \cdot H_T}} \right) \\
 & + \frac{1}{\Sigma_{TOT}} \left[ k_{tot}^{S_2 \cdot H_T \rightarrow S_2 \cdot H_D} k_s^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D} k_{tot}^{S_1 \cdot H_D \rightarrow S_1 \cdot H_T} k_{f, H_T}^{(1)} \right] \\
 & \cdot \left( 1 - \frac{P_{S_1}(S_1 \cdot H_D) P_{S_2}(S_2 \cdot H_T)}{P_{S_1}(S_1 \cdot H_T) P_{S_2}(S_2 \cdot H_D)} \cdot \frac{k_{b, H_T}^{(2)} k_s^{S_1 \cdot H_D \rightarrow S_2 \cdot H_D}}{k_{f, H_T}^{(1)} k_s^{S_2 \cdot H_D \rightarrow S_1 \cdot H_D}} \right)
 \end{aligned} \tag{A.73}$$

# Bibliography

- [1] E. Schrodinger. *What Is Life?: With Mind and Matter and Autobiographical Sketches*. Canto Classics. Cambridge: Cambridge University Press, 2012.
- [2] Encyclopædia Britannica. "Life". Ed. by B. Academic. URL: [academic.eb.com/levels/collegiate/article/life/106478](http://academic.eb.com/levels/collegiate/article/life/106478) (visited on 09/15/2022).
- [3] Science et Vie, ed. *Les origines de la vie*. Hors-série 245 (June 2018).
- [4] G.-P. Dotto. "To Be or Not to Be". In: *EMBO reports* 21.7 (July 3, 2020), e50861.
- [5] J. Monod. *Le Hasard et la Nécessité. Essai sur la philosophie naturelle de la biologie moderne*. Média Diffusion, Dec. 25, 2014. 155 pp.
- [6] J. Monod. *Chance and Necessity: An Essay on the National Philosophy of Modern Biology*. Collins, 1974. 187 pp.
- [7] A. Einstein, H. Born, and M. Born. *Briefwechsel 1916-1955*. Rowohlt, 1972. 260 pp.
- [8] S. I. Walker and P. C. W. Davies. "The Algorithmic Origins of Life". In: *Journal of The Royal Society Interface* 10.79 (Feb. 6, 2013), p. 20120869.
- [9] W. R. Loewenstein. *The Touchstone of Life: Molecular Information, Cell Communication, and the Foundations of Life*. Illustrated edition. New York: Oxford University Press, Dec. 14, 2000. 388 pp.
- [10] J. H. van Hateren. *What Does Maxwell's Demon Want from Life? When Information Becomes Functional and Physical*. Apr. 28, 2015. URL: <http://arxiv.org/abs/1407.8314>.
- [11] L. Cronin. *Im a chemist, and Im building a universal robot to make life and find aliens*. Ed. by Big Think. URL: <https://bigthink.com/the-well/life-chemistry-aliens/> (visited on 09/15/2022).
- [12] S. A. S. Eddington. *The Nature of the Physical World*. University Press, 1948.
- [13] J. C. Maxwell. *Theory of Heat*. Longmans, Green, and Company, 1872.
- [14] W. Ehrenberg. "Maxwell's Demon". In: *Scientific American* 217.5 (1967), pp. 103–111.

## Bibliography

---

- [15] M. Smoluchowski. “Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularphänomene”. In: *Pisma Mariana Smoluchowskiego* 2.1 (1927), pp. 226–251.
- [16] A. Rex. “Maxwell’s Demon A Historical Review”. In: *Entropy* 19.6 (May 2017), p. 240.
- [17] P. A. Skordos and W. H. Zurek. “Maxwell’s Demon, Rectifiers, and the Second Law: Computer Simulation of Smoluchowski’s Trapdoor”. In: *American Journal of Physics* 60 (Oct. 1992), pp. 876–882.
- [18] L. Szilard. “On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings”. In: *Behavioral Science* 9.4 (1964), pp. 301–310.
- [19] L. Brillouin. “Maxwell’s Demon Cannot Operate: Information and Entropy. I”. In: *Journal of Applied Physics* 22.3 (Mar. 1951), pp. 334–337.
- [20] L. Brillouin. “Physical Entropy and Information. II”. In: *Journal of Applied Physics* 22.3 (Mar. 1951), pp. 338–343.
- [21] R. Landauer. “Information Is Physical”. In: (1991), p. 8.
- [22] C. H. Bennett. “The Thermodynamics of Computation - a Review”. In: *International Journal of Theoretical Physics* 21.12 (Dec. 1982), pp. 905–940.
- [23] C. E. Shannon. “A Mathematical Theory of Communication”. In: (1948), p. 55.
- [24] S. Ciliberto. “Experiments in Stochastic Thermodynamics: Short History and Perspectives”. In: (2017), p. 26.
- [25] S. Toyabe et al. “Experimental Demonstration of Information-to-Energy Conversion and Validation of the Generalized Jarzynski Equality”. In: *Nature Physics* 6.12 (Dec. 2010), pp. 988–992.
- [26] G. Boël et al. “Omnipresent Maxwell’s Demons Orchestrate Information Management in Living Cells”. In: *Microbial Biotechnology* 12.2 (2019), pp. 210–242.
- [27] M. A. Hediger et al. “The ABCs of Membrane Transporters in Health and Disease (SLC Series): Introduction”. In: *Molecular Aspects of Medicine. The ABCs of Membrane Transporters in Health and Disease (SLC Series)* 34.2 (Apr. 2013), pp. 95–107.
- [28] J. A. MacKenzie and R. M. Payne. “Mitochondrial Protein Import and Human Health and Disease”. In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1772.5 (May 2007), pp. 509–523.
- [29] D. C. Gadsby, P. Vergani, and L. Csanády. “The ABC Protein Turned Chloride Channel Whose Failure Causes Cystic Fibrosis”. In: *Nature* 440.7083 (Mar. 2006), pp. 477–483.
- [30] S. C. Hyde et al. “Structural Model of ALP-binding Proteins Associated with Cystic Fibrosis, Multidrug Resistance and Bacterial Transport”. In: 346 (1990), p. 4.
- [31] B. Alberts et al. *Molecular Biology of the Cell*. Fourth. Garland Science, 2002.

- 
- [32] L. R. Forrest, R. Krämer, and C. Ziegler. “The Structural Basis of Secondary Active Transport Mechanisms”. In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1807.2 (Feb. 2011), pp. 167–188.
- [33] E. M. Wright. “The Intestinal Na<sup>+</sup>/Glucose Cotransporter”. In: *Annual Review of Physiology* 55 (1993), pp. 575–589.
- [34] L. Peliti. *Statistical Mechanics in a Nutshell*. Princeton University Press, Aug. 2011.
- [35] U. Seifert. “Stochastic Thermodynamics: An Introduction”. In: *Nonequilibrium Statistical Physics Today: Proceedings of the 11th Granada Seminar on Computational and Statistical Physics*. 2011, pp. 56–76.
- [36] R. Clausius. “Ueber Verschiedene Für Die Anwendung Bequeme Formen Der Hauptgleichungen Der Mechanischen Wärmetheorie”. In: *Annalen der Physik* 201.7 (1865), pp. 353–400.
- [37] I. Prigogine and P. V. Rysselberghe. “Introduction to Thermodynamics of Irreversible Processes”. In: *Journal of The Electrochemical Society* 110.4 (Apr. 1963), p. 97C.
- [38] I. Prigogine and G. Nicolis. “Self-Organisation in Nonequilibrium Systems: Towards A Dynamics of Complexity”. In: *Bifurcation Analysis: Principles, Applications and Synthesis*. Ed. by M. Hazewinkel, R. Jurkovich, and J. H. P. Paelinck. Dordrecht: Springer Netherlands, 1985, pp. 3–12.
- [39] X.-J. Zhang, H. Qian, and M. Qian. “Stochastic Theory of Nonequilibrium Steady States and Its Applications. Part I”. In: *Physics Reports. Stochastic Theory of Nonequilibrium Steady States and Its Applications: Part I* 510.1 (Jan. 2012), pp. 1–86.
- [40] C. V. den Broeck and M. Esposito. “Ensemble and Trajectory Thermodynamics: A Brief Introduction”. In: *Physica A: Statistical Mechanics and its Applications* 418 (Jan. 2015), pp. 6–16.
- [41] U. Seifert. “Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem”. In: *Physical Review Letters* 95.4 (July 2005), p. 040602.
- [42] G. M. Wang et al. “Experimental Demonstration of Violations of the Second Law of Thermodynamics for Small Systems and Short Time Scales”. In: *Physical Review Letters* 89.5 (July 2002), p. 050601.
- [43] D. M. Carberry et al. “Fluctuations and Irreversibility: An Experimental Demonstration of a Second-Law-Like Theorem Using a Colloidal Particle Held in an Optical Trap”. In: *Physical Review Letters* 92.14 (Apr. 2004), p. 140601.
- [44] O. Mazonka and C. Jarzynski. *Exactly Solvable Model Illustrating Far-from-Equilibrium Predictions*. Dec. 1999.
- [45] R. van Zon and E. G. D. Cohen. “Extended Heat-Fluctuation Theorems for a System with Deterministic and Stochastic Forces”. In: *Physical Review E* 69.5 (May 2004), p. 056121.

## Bibliography

---

- [46] K. Sekimoto. “Langevin Equation and Thermodynamics”. In: *Progress of Theoretical Physics Supplement* 130 (Jan. 1998), pp. 17–27.
- [47] G. E. Crooks. “Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences”. In: *Physical Review E* 60.3 (Sept. 1999), pp. 2721–2726.
- [48] K. Sekimoto. *Stochastic Energetics*. Vol. 799. Lecture Notes in Physics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.
- [49] L. Peliti and S. Pigolotti. *Stochastic Thermodynamics: An Introduction*. Princeton University Press, July 2021.
- [50] D. M. Busiello. “Entropy Production in Non-Equilibrium Systems”. PhD Thesis. Università degli Studi di Padova, Jan. 19, 2018.
- [51] J. Kurchan. “Fluctuation Theorem for Stochastic Dynamics”. In: *Journal of Physics A: Mathematical and General* 31.16 (Apr. 1998), pp. 3719–3729.
- [52] C. Maes. “On the Origin and the Use of Fluctuation Relations for the Entropy”. In: *Poincaré Seminar 2003*. Ed. by J. Dalibard, B. Duplantier, and V. Rivasseau. Basel: Birkhäuser Basel, 2004, pp. 145–191.
- [53] U. Seifert. “Stochastic Thermodynamics, Fluctuation Theorems and Molecular Machines”. In: *Reports on Progress in Physics* 75.12 (Nov. 2012), p. 126001.
- [54] T. L. Hill. “Theoretical Formalism for the Sliding Filament Model of Contraction of Striated Muscle Part I”. In: *Progress in Biophysics and Molecular Biology* 28 (Jan. 1974), pp. 267–340.
- [55] C. Vidal, A. Pacault, and H. Haken, eds. *Non-Equilibrium Dynamics in Chemical Systems: Proceedings of the International Symposium, Bordeaux, France, September 37, 1984*. Vol. 27. Springer Series in Synergetics. Berlin, Heidelberg: Springer Berlin Heidelberg, 1984.
- [56] A. F. Huxley. “Muscle structure and theories of contraction”. In: *Progress in Biophysics and Biophysical Chemistry* 7 (1957), pp. 255–318.
- [57] C. Bustamante, J. Liphardt, and F. Ritort. “The Nonequilibrium Thermodynamics of Small Systems”. In: *Physics Today* 58.7 (July 2005), pp. 43–48.
- [58] H. Ge, M. Qian, and H. Qian. “Stochastic Theory of Nonequilibrium Steady States. Part II: Applications in Chemical Biophysics”. In: *Physics Reports. Stochastic Theory of Nonequilibrium Steady States (Part II): Applications in Chemical Biophysics* 510.3 (Jan. 2012), pp. 87–118.
- [59] C. Gardiner. *Stochastic Methods, A Handbook for the Natural and Social Sciences*. Ed. by Springer. Fourth edition. 2009.



- [60] T. L. Hill. *State Probabilities and Fluxes in Terms of the Rate Constants of the Diagram*. Ed. by T. L. Hill. New York, NY: Springer, 1989, pp. 39–88.
- [61] S. Chaiken and D. J. Kleitman. “Matrix Tree Theorems”. In: *Journal of Combinatorial Theory, Series A* 24.3 (May 1978), pp. 377–381.
- [62] M. Chakraborty et al. “Algorithms for Generating All Possible Spanning Trees of a Simple Undirected Connected Graph: An Extensive Review”. In: *Complex & Intelligent Systems* 5.3 (Oct. 2019), pp. 265–281.
- [63] *International Union of Pure and Applied Chemists (IUPAC), Compendium of Chemical Terminology, "Principle of microscopic reversibility"*.
- [64] K. J. Laidler. “A Glossary of Terms Used in Chemical Kinetics, Including Reaction Dynamics (IUPAC Recommendations 1996)”. In: *Pure and Applied Chemistry* 68.1 (Jan. 1996), pp. 149–192.
- [65] R. D. Astumian. “Microscopic Reversibility as the Organizing Principle of Molecular Machines”. In: *Nature Nanotechnology* 7.11 (Nov. 2012), pp. 684–688.
- [66] R. D. Astumian. “How Molecular Motors Work Insights from the Molecular Machinist’s Toolbox: The Nobel Prize in Chemistry 2016”. In: *Chemical Science* 8.2 (Jan. 2017), pp. 840–845.
- [67] L. Boltzmann. *Lectures on Gas Theory*. Ed. by S. G. Brush. July 2022.
- [68] *International Union of Pure and Applied Chemists (IUPAC), Compendium of Chemical Terminology, "Detailed balancing (principle of)"*.
- [69] D. M. Busiello, D. Gupta, and A. Maritan. “Entropy Production in Systems with Unidirectional Transitions”. In: *Physical Review Research* 2.2 (Apr. 2020), p. 023011.
- [70] A. Pal, S. Reuveni, and S. Rahav. “Thermodynamic Uncertainty Relation for Systems with Unidirectional Transitions”. In: *Physical Review Research* 3.1 (Mar. 2021), p. 013273.
- [71] D. Gupta and D. M. Busiello. “Tighter Thermodynamic Bound on Speed Limit in Systems with Unidirectional Transitions”. In: *Physical Review E* 102.6 (Dec. 2020), p. 062121.
- [72] M. Le Bellac, F. Mortessagne, and G. G. Batrouni. *Equilibrium and Non-Equilibrium Statistical Thermodynamics*. Cambridge: Cambridge University Press, 2004.
- [73] T. L. Hill. “Survey of the Elements of Free Energy Transduction”. In: *Free Energy Transduction and Biochemical Cycle Kinetics*. Ed. by T. L. Hill. New York, NY: Springer, 1989, pp. 1–37.
- [74] R. D. Astumian. “Kinetic Asymmetry Allows Macromolecular Catalysts to Drive an Information Ratchet”. In: *Nature Communications* 10.1 (Aug. 23, 2019), p. 3837.
- [75] J. Schnakenberg. “Network Theory of Microscopic and Macroscopic Behavior of Master Equation Systems”. In: *Reviews of Modern Physics* 48.4 (Oct. 1976), pp. 571–585.

## Bibliography

---

- [76] E. Penocchio, F. Avanzini, and M. Esposito. “Information Thermodynamics for Deterministic Chemical Reaction Networks”. In: *The Journal of Chemical Physics* 157.3 (July 21, 2022), p. 034110.
- [77] M. Esposito and C. V. den Broeck. “Second Law and Landauer Principle Far from Equilibrium”. In: *EPL (Europhysics Letters)* 95.4 (Aug. 2011), p. 40004.
- [78] S. Deffner and E. Lutz. *Information Free Energy for Nonequilibrium States*. Jan. 2012.
- [79] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Nov. 2012.
- [80] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa. “Thermodynamics of Information”. In: *Nature Physics* (Feb. 2015), pp. 131–139.
- [81] T. Sagawa and M. Ueda. “Generalized Jarzynski Equality under Nonequilibrium Feedback Control”. In: *Physical Review Letters* 104.9 (Mar. 2010), p. 090602.
- [82] J. M. Horowitz and S. Vaikuntanathan. “Nonequilibrium Detailed Fluctuation Theorem for Repeated Discrete Feedback”. In: *Physical Review E* 82.6 (Dec. 2010), p. 061120.
- [83] J. M. Horowitz and J. M. R. Parrondo. “Designing Optimal Discrete-Feedback Thermodynamic Engines”. In: *New Journal of Physics* 13.12 (Dec. 2011), p. 123019.
- [84] K. Mallick and B. Duplantier. “Thermodynamics and Information Theory”. In: *Information Theory: Poincaré Seminar 2018*. Ed. by B. Duplantier and V. Rivasseau. Progress in Mathematical Physics. Cham: Springer International Publishing, 2021, pp. 1–48.
- [85] C. Jarzynski. “Nonequilibrium Equality for Free Energy Differences”. In: *Physical Review Letters* 78.14 (Apr. 1997), pp. 2690–2693.
- [86] M. Ponmurugan. “Generalized Detailed Fluctuation Theorem under Nonequilibrium Feedback Control”. In: *Physical Review E* 82.3 (Sept. 2010), p. 031129.
- [87] J. M. Horowitz and J. M. R. Parrondo. “Thermodynamic Reversibility in Feedback Processes”. In: *EPL (Europhysics Letters)* 95.1 (June 2011), p. 10005.
- [88] U. Seifert. “Stochastic Thermodynamics: Principles and Perspectives”. In: *The European Physical Journal B* 64.3 (Aug. 1, 2008), pp. 423–431.
- [89] R. Rao and M. Esposito. “Detailed Fluctuation Theorems: A Unifying Perspective”. In: *Entropy* 20.9 (Sept. 2018), p. 635.
- [90] R. Phillips et al. *Physical Biology of the Cell*. Ed. by G. Science. Second Edition. 2013.
- [91] S. Flatt et al. “ABC Transporters Are Billion-Year-Old Maxwell Demons”. In: (Dec. 3, 2021), p. 2021.12.03.471046.
- [92] C. F. Higgins. “ABC Transporters: From Microorganisms to Man”. In: *Annual Review of Cell Biology* 8 (1992), pp. 67–113.

- [93] D. C. Rees, E. Johnson, and O. Lewinson. “ABC Transporters: The Power to Change”. In: *Nature Reviews Molecular Cell Biology* (Mar. 2009), pp. 218–227.
- [94] A. L. Davidson et al. “Structure, Function, and Evolution of Bacterial ATP-binding Cassette Systems”. In: *Microbiology and molecular biology reviews: MMBR* 72.2 (June 2008), 317–364, table of contents.
- [95] Y. Li and W. A. Prinz. “ATP-binding Cassette (ABC) Transporters Mediate Nonvesicular, Raft-modulated Sterol Movement from the Plasma Membrane to the Endoplasmic Reticulum \*”. In: *Journal of Biological Chemistry* 279.43 (Oct. 2004), pp. 45226–45234.
- [96] M. M. Gottesman and S. V. Ambudkar. “Overview: ABC Transporters and Human Disease”. In: *Journal of Bioenergetics and Biomembranes* 33.6 (Dec. 2001), pp. 453–458.
- [97] J. I. Fletcher et al. “ABC Transporters in Cancer: More than Just Drug Efflux Pumps”. In: *Nature Reviews Cancer* 10.2 (Feb. 2010), pp. 147–156.
- [98] M. Benadiba and Y. Maor. “Importance of ABC Transporters in Drug Development”. In: *Current Pharmaceutical Design* 22.38 (2016), pp. 5817–5829.
- [99] P. M. Jones and A. M. George. “The ABC Transporter Structure and Mechanism: Perspectives on Recent Research”. In: *Cellular and Molecular Life Sciences CMLS* 61.6 (Mar. 2004), pp. 682–699.
- [100] A. L. Davidson and J. Chen. “ATP-binding Cassette Transporters in Bacteria”. In: *Annual Review of Biochemistry* 73 (2004), pp. 241–268.
- [101] J. ter Beek, A. Guskov, and D. J. Slotboom. “Structural Diversity of ABC Transporters”. In: *Journal of General Physiology* 143.4 (Mar. 2014), pp. 419–435.
- [102] C. Orelle et al. “A Multidrug ABC Transporter with a Taste for GTP”. In: *Scientific Reports* 8.1 (Feb. 2018), p. 2309.
- [103] P. M. Jones, M. L. O’Mara, and A. M. George. “ABC Transporters: A Riddle Wrapped in a Mystery inside an Enigma”. In: *Trends in Biochemical Sciences* 34.10 (Oct. 2009), pp. 520–531.
- [104] C. Thomas et al. “Structural and Functional Diversity Calls for a New Classification of ABC Transporters”. In: *FEBS Letters* 594.23 (2020), pp. 3767–3775.
- [105] K. P. Locher, A. T. Lee, and D. C. Rees. “The E. Coli BtuCD Structure: A Framework for ABC Transporter Architecture and Mechanism”. In: *Science (New York, N.Y.)* 296.5570 (May 2002), pp. 1091–1098.
- [106] M. Hohl et al. “Crystal Structure of a Heterodimeric ABC Transporter in Its Inward-Facing Conformation”. In: *Nature Structural & Molecular Biology* 19.4 (Mar. 2012), pp. 395–402.
- [107] C. A. J. Hutter et al. “The Extracellular Gate Shapes the Energy Profile of an ABC Exporter”. In: *Nature Communications* 10.1 (May 2019), p. 2260.

## Bibliography

---

- [108] K. J. Linton and C. F. Higgins. “Structure and Function of ABC Transporters: The ATP Switch Provides Flexible Control”. In: *Pflügers Archiv - European Journal of Physiology* 453.5 (Feb. 2007), pp. 555–567.
- [109] K. P. Locher. “Mechanistic Diversity in ATP-binding Cassette (ABC) Transporters”. In: *Nature Structural & Molecular Biology* (June 2016), pp. 487–493.
- [110] P. C. Smith et al. “ATP Binding to the Motor Domain from an ABC Transporter Drives Formation of a Nucleotide Sandwich Dimer”. In: *Molecular Cell* 10.1 (July 2002), pp. 139–149.
- [111] E. O. Oloo and D. P. Tieleman. “Conformational Transitions Induced by the Binding of MgATP to the Vitamin B12 ATP-binding Cassette (ABC) Transporter BtuCD”. In: *The Journal of Biological Chemistry* 279.43 (Oct. 2004), pp. 45013–45019.
- [112] A. M. George and P. M. Jones. “Perspectives on the Structure-Function of ABC Transporters: The Switch and Constant Contact Models”. In: *Progress in Biophysics and Molecular Biology* 109.3 (Aug. 2012), pp. 95–107.
- [113] C. F. Higgins and K. J. Linton. “The ATP Switch Model for ABC Transporters”. In: *Nature Structural & Molecular Biology* (Oct. 2004), pp. 918–926.
- [114] K. Nikaido and G. E.-L. Ames. “One Intact ATP-binding Subunit Is Sufficient to Support ATP Hydrolysis and Translocation in an ABC Transporter, the Histidine Permease\*<sup>†</sup>”. In: *Journal of Biological Chemistry* 274.38 (Sept. 1999), pp. 26727–26735.
- [115] G. Kuhnke et al. “Stimulation of the ATPase Activity of the Yeast Mitochondrial ABC Transporter Atm1p by Thiol Compounds”. In: *Molecular Membrane Biology* 23.2 (Jan. 2006), pp. 173–184.
- [116] J.-P. Changeux and A. Christopoulos. “Allosteric Modulation as a Unifying Mechanism for Receptor Function and Regulation”. In: *Cell* 166.5 (Aug. 2016), pp. 1084–1102.
- [117] H. N. Motlagh et al. “The Ensemble Nature of Allostery”. In: *Nature* 508.7496 (Apr. 2014), pp. 331–339.
- [118] C. Thomas and R. Tampé. “Structural and Mechanistic Principles of ABC Transporters”. In: *Annual Review of Biochemistry* 89 (June 2020), pp. 605–636.
- [119] K. Hollenstein, R. J. Dawson, and K. P. Locher. “Structure and Mechanism of ABC Transporter Proteins”. In: *Current Opinion in Structural Biology. Membranes / Engineering and Design* 17.4 (Aug. 2007), pp. 412–418.
- [120] O. Jardetzky. “Simple Allosteric Model for Membrane Pumps”. In: *Nature* 211.5052 (Aug. 1966), pp. 969–970.
- [121] S. Wilkens. “Structure and Mechanism of ABC Transporters”. In: *F1000 Prime Reports* (Feb. 2015).

- [122] A. E. Senior and D. C. Gadsby. “ATP Hydrolysis Cycles and Mechanism in P-glycoprotein and CFTR”. In: *Seminars in Cancer Biology* 8.3 (June 1997), pp. 143–150.
- [123] P. M. Jones and A. M. George. “A Reciprocating Twin-Channel Model for ABC Transporters”. In: *Quarterly Reviews of Biophysics* 47.3 (Aug. 2014), pp. 189–220.
- [124] R. J. P. Dawson, K. Hollenstein, and K. P. Locher. “Uptake or Extrusion: Crystal Structures of Full ABC Transporters Suggest a Common Mechanism”. In: *Molecular Microbiology* 65.2 (2007), pp. 250–257.
- [125] A. S. Sassi. “Non Equilibrium Thermodynamics of Protein Organization”. PhD Thesis. École Polytechnique fédérale de Lausanne, Feb. 2018.
- [126] R. M. Bibini. *Cell Biology by the Numbers*. URL: <http://book.bionumbers.org>.
- [127] R. D. Astumian and M. Bier. “Mechanochemical Coupling of the Motion of Molecular Motors to ATP Hydrolysis.” In: *Biophysical Journal* 70.2 (Feb. 1996), pp. 637–653.
- [128] S. Gerber et al. “Structural Basis of Trans-Inhibition in a Molybdate/Tungstate ABC Transporter”. In: *Science (New York, N.Y.)* 321.5886 (July 2008), pp. 246–250.
- [129] N. Grossmann et al. “Mechanistic Determinants of the Directionality and Energetics of Active Export by a Heterodimeric ABC Transporter”. In: *Nature Communications* 5.1 (Nov. 2014), p. 5419.
- [130] M. de Boer, T. Cordes, and B. Poolman. “Kinetic Modelling of Transport Inhibition by Substrates in ABC Importers”. In: *Journal of Molecular Biology* 432.20 (Sept. 2020), pp. 5565–5576.
- [131] T. A. Rapoport, L. Li, and E. Park. “Structural and Mechanistic Insights into Protein Translocation”. In: *Annual Review of Cell and Developmental Biology* 33 (Oct. 2017), pp. 369–390.
- [132] F. A. Agarraberes and J. F. Dice. “A Molecular Chaperone Complex at the Lysosomal Membrane Is Required for Protein Translocation”. In: *Journal of Cell Science* 114.Pt 13 (July 2001), pp. 2491–2499.
- [133] L. Sagan. “On the Origin of Mitosing Cells”. In: *Journal of Theoretical Biology* 14.3 (Mar. 1, 1967), 225–IN6.
- [134] R. Zimmermann et al. “Protein Translocation across the ER Membrane”. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes*. Including the Special Section: Protein Translocation across or Insertion into Membranes 1808.3 (Mar. 2011), pp. 912–924.
- [135] D. Mokranjac. “How to Get to the Other Side of the Mitochondrial Inner Membrane the Protein Import Motor”. In: *Biological Chemistry* 401.6-7 (May 2020), pp. 723–736.
- [136] E. A. Craig. “Hsp70 at the Membrane: Driving Protein Translocation”. In: *BMC Biology* 16.1 (Jan. 2018), p. 11.

## Bibliography

---

- [137] P.-J. Kang et al. "Requirement for Hsp70 in the Mitochondrial Matrix for Translocation and Folding of Precursor Proteins". In: *Nature* 348.6297 (Nov. 1990), pp. 137–143.
- [138] R. Kityk et al. "Pathways of Allosteric Regulation in Hsp70 Chaperones". In: *Nature Communications* 6.1 (Sept. 2015), p. 8308.
- [139] V. Kohler and C. Andréasson. "Hsp70-Mediated Quality Control: Should I Stay or Should I Go?" In: *Biological Chemistry* 401.11 (Oct. 2020), pp. 1233–1248.
- [140] D. Mokranjac et al. "Tim14, a Novel Key Component of the Import Motor of the TIM23 Protein Translocase of Mitochondria". In: *The EMBO Journal* 22.19 (Oct. 2003), pp. 4945–4956.
- [141] M. Linxweiler, B. Schick, and R. Zimmermann. "Let's Talk about Secs: Sec61, Sec62 and Sec63 in Signal Transduction, Oncology and Personalized Medicine". In: *Signal Transduction and Targeted Therapy* 2.1 (Apr. 2017), pp. 1–10.
- [142] D. Mokranjac et al. "Structure and Function of Tim14 and Tim16, the J and J-like Components of the Mitochondrial Protein Import Motor". In: *The EMBO Journal* 25.19 (Oct. 2006), pp. 4675–4685.
- [143] H. C. Schneider et al. "The Nucleotide Exchange Factor MGE Exerts a Key Function in the ATP-dependent Cycle of Mt-Hsp70-Tim44 Interaction Driving Mitochondrial Protein Import." In: *The EMBO Journal* 15.21 (Nov. 1996), pp. 5796–5803.
- [144] W. Neupert and M. Brunner. "The Protein Import Motor of Mitochondria". In: *Nature Reviews Molecular Cell Biology* 3.8 (Aug. 2002), pp. 555–565.
- [145] P. D. L. Rios et al. "Hsp70 Chaperones Accelerate Protein Translocation and the Unfolding of Stable Protein Aggregates by Entropic Pulling". In: *Proceedings of the National Academy of Sciences* 103.16 (Apr. 2006), pp. 6166–6171.
- [146] M. Horst et al. "The Mitochondrial Protein Import Motor: Dissociation of Mitochondrial Hsp70 from Its Membrane Anchor Requires ATP Binding Rather than ATP Hydrolysis: ATP Requirement of the Mitochondrial Import Motor". In: *Protein Science* 5.4 (Apr. 1996), pp. 759–767.
- [147] P. Goloubinoff and P. D. L. Rios. "The Mechanism of Hsp70 Chaperones: (Entropic) Pulling the Models Together". In: *Trends in Biochemical Sciences* 32.8 (Aug. 2007), pp. 372–380.
- [148] P. De Los Rios and A. Barducci. "Hsp70 Chaperones Are Non-Equilibrium Machines That Achieve Ultra-Affinity by Energy Consumption". In: *eLife* 3 (May 2014). Ed. by A. K. Chakraborty, e02218.
- [149] B. Nguyen et al. "Thermodynamic Bounds on the Ultra- and Infra-affinity of Hsp70 for Its Substrates". In: *Biophysical Journal* 113.2 (July 2017), pp. 362–370.

- [150] S. Rüdiger, A. Buchberger, and B. Bukau. “Interaction of Hsp70 Chaperones with Substrates”. In: *Nature Structural Biology* 4.5 (May 1997), pp. 342–349.
- [151] S. Rüdiger et al. “Substrate Specificity of the DnaK Chaperone Determined by Screening Cellulose-Bound Peptide Libraries”. In: *The EMBO journal* 16.7 (Apr. 1997), pp. 1501–1507.
- [152] R. Sousa and E. M. Lafer. “The Physics of Entropic Pulling: A Novel Model for the Hsp70 Motor Mechanism”. In: *International Journal of Molecular Sciences* 20.9 (May 11, 2019), E2334.
- [153] P. D. D’Silva et al. “J Protein Cochaperone of the Mitochondrial Inner Membrane Required for Protein Import into the Mitochondrial Matrix”. In: *Proceedings of the National Academy of Sciences* 100.24 (Nov. 2003), pp. 13839–13844.
- [154] A. K. Corsi and R. Schekman. “The Luminal Domain of Sec63p Stimulates the ATPase Activity of BiP and Mediates BiP Recruitment to the Translocon in *Saccharomyces Cerevisiae*”. In: *The Journal of Cell Biology* 137.7 (June 1997), pp. 1483–1493.
- [155] W. Schwenke et al. “Mitochondrial and Cytosolic ATP/ADP Ratios in Rat Liver in Vivo”. In: *The Biochemical journal* 200 (Dec. 1981), pp. 405–8.
- [156] H. W. Heldt, M. Klingenberg, and M. Milovancev. “Differences between the ATP/ADP Ratios in the Mitochondrial Matrix and in the Extramitochondrial Space”. In: *European Journal of Biochemistry* 30.3 (1972), pp. 434–440.
- [157] F. A. Aprile et al. “Hsp70 Oligomerization Is Mediated by an Interaction between the Interdomain Linker and the Substrate-Binding Domain”. In: *PLOS ONE* 8.6 (June 2013), e67961.
- [158] J. E. Takakuwa et al. “Oligomerization of Hsp70: Current Perspectives on Regulation and Function”. In: *Frontiers in Molecular Biosciences* 6 (2019).
- [159] R. Rosenzweig et al. “The Hsp70 Chaperone Network”. In: *Nature Reviews Molecular Cell Biology* 20.11 (Nov. 2019), pp. 665–680.
- [160] D. Kim, Y. J. Lee, and P. M. Corry. “Constitutive HSP70: Oligomerization and Its Dependence on ATP Binding”. In: *Journal of Cellular Physiology* 153.2 (1992), pp. 353–361.
- [161] J. I.-J. Leu et al. “Structural Basis for the Inhibition of HSP70 and DnaK Chaperones by Small-Molecule Targeting of a C-Terminal Allosteric Pocket”. In: *ACS Chemical Biology* 9.11 (Nov. 2014), pp. 2508–2516.
- [162] N. Wadhwa and H. C. Berg. “Bacterial Motility: Machinery and Mechanisms”. In: *Nature Reviews Microbiology* (Sept. 2021).
- [163] N. Wadhwa et al. “A Multi-State Dynamic Process Confers Mechano-Adaptation to a Biological Nanomachine”. In: *Nature Communications* 13.1 (Sept. 2022), p. 5327.

## Bibliography

---

- [164] W. Liebermeister, T. A. Rapoport, and R. Heinrich. “Ratcheting in Post-Translational Protein Translocation: A Mathematical Model”. In: *Journal of Molecular Biology* 305.3 (Jan. 19, 2001), pp. 643–656.
- [165] K. E. S. Matlack et al. “BiP Acts as a Molecular Ratchet during Posttranslational Transport of Prepro- $\alpha$  Factor across the ER Membrane”. In: *Cell* 97.5 (May 28, 1999), pp. 553–564.
- [166] S. Assenza, P. De Los Rios, and A. Barducci. “Quantifying the Role of Chaperones in Protein Translocation by Computational Modeling”. In: *Frontiers in Molecular Biosciences* 2 (2015), p. 8.
- [167] S. Assenza et al. “Efficient Conversion of Chemical Energy into Mechanical Work by Hsp70 Chaperones”. In: *eLife* 8 (Dec. 17, 2019). Ed. by A. K. Chakraborty, D. Weigel, and E. Papaleo, e48491.
- [168] B. Hu, M. P. Mayer, and M. Tomita. “Modeling Hsp70-Mediated Protein Folding”. In: *Biophysical Journal* 91.2 (July 2006), pp. 496–507.



# Solange Marie FLATT

DATE OF BIRTH: 15th July 1995  
ADDRESS: Rue du Bugnon 22, 1005 Lausanne, Switzerland  
PHONE: +41 78 606 57 86  
EMAIL: [solange\\_flatt@hotmail.com](mailto:solange_flatt@hotmail.com)  
[solange.flatt@epfl.ch](mailto:solange.flatt@epfl.ch)



## EDUCATION

---

2015 - 2017 | Master in PHYSICS - Minor in COMPUTATIONAL SCIENCE AND ENGINEERING (CSE),  
Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

2012 - 2015 | Bachelor in PHYSICS,  
Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

## LANGUAGES

---

FRENCH: Native proficiency  
ENGLISH: Fluent  
High School level (B2 Attestation in 2012)  
June 2018 : *Academic Writing for Doctoral Students*  
PhD Intensive Course (B2-C1 Students), 2ECTS Credits, Grade 6 out of 6.  
GERMAN: High School level (B2 Attestation in 2012)

## ACADEMIC EXPERIENCE

---

JUN 2018 - NOV 2022 | PhD Candidate in the **Laboratory of Statistical Biophysics (LBS)** - EPFL  
with Prof. P. De Los Rios

JAN 2018-MAY 2018 | Scientific collaborator in the **Physics of Complex Systems Laboratory (PCSL)** - EPFL with Prof. M. Wyart

SEP 2017-DEC 2017 | Valorisation of Master Project as trainee **Physics of Complex Systems Laboratory (PCSL)** - EPFL with Prof. M. Wyart

FEB-JUN 2017 | Master Project in the **Physics of Complex Systems Laboratory (PCSL)** - EPFL - under the supervision of R. Ravasio & under the direction of Prof. M. Wyart  
Rigidity properties and mechanics of allosteric proteins

SEP-DEC 2016 | Minor Project (CSE) in the **Laboratory for Reactor Physics and Systems Behaviour (LRS)** - EPFL - under the supervision of Dr. M. Hursin  
Application of Lattice Boltzmann Method to neutron transport

SEP 2015-JUN 2016 | Semester Projects in the **Laboratory for Quantum Magnetism (LQM)** - EPFL - under the supervision of Prof. H.M. Rønnow  
Spin wave dispersion relation in  $\text{Sr}_2\text{Cu}_3\text{O}_4\text{Cl}_2$  lattice structure (Jan-June 2016)  
 $J_1 - J_2 - J_\square$  model on a square lattice (Sept-Dec 2015)

SEP-DEC 2015 | Project in the **Laboratory for Reactor Physics and Systems Behaviour (LRS)** - EPFL - under the supervision of Dr. M. Hursin  
Comparison of methods to solve the 1D Boltzmann Equation for neutron transport

## CONFERENCES

---

- |               |   |
|---------------|---|
| 5-8 SEP 2022  | Workshop in Bressanone: <i>Physics of Biomolecules: Structure, Dynamics and Function</i> . Oral presentation<br>Title: ABC Transporters are billion-year old Maxwell Demons                 |
| 3-7 JUL 2022  | Conference in Barcelona: <i>Biological evolution and nonequilibrium physics: close encounters</i> . Presentation of a poster<br>Title: ABC Transporters are billion-year old Maxwell Demons |
| 8-26 JUL 2019 | Boulder School for Condensed Matter and Materials Physics (USA): <i>Theoretical Biophysics</i>  |
| 7-11 OCT 2018 | Summer school in Weizmann Institute (Israel): Introduction to biological physics for students of science and engineering  |

## PUBLICATIONS

---

- R. Ravasio, S. M. Flatt, L. Yan, S. Zamuner, C. Brito & M. Wyart (2019).  
Mechanics of Allostery: Contrasting the Induced Fit and Population Shift Scenarios.  
Biophysical journal, 117(10), 1954–1962. <https://doi.org/10.1016/j.bpj.2019.10.002>
- S. M. Flatt, D. M. Busiello, S. Zamuner & P. De Los Rios (2022, Under review).  
ABC Transporters are billion-year-old Maxwell Demons.  
<https://doi.org/10.1101/2021.12.03.471046> (BioRxiv)

## TEACHING EXPERIENCE

---

- |                              |  |
|------------------------------|--|
| SEP-DEC 2019                 | Teaching assistant: "Quantum Physics I"  |
| FEB-JUN 2019<br>FEB-JUN 2021 | Teaching assistant: "Analytical Mechanics"   |
| SEP-DEC 2018                 | Teaching assistant: "Statistical Physics of Biomacromolecules"                                       |
| FEB-JUN 2017                 | Teaching assistant: "Physique générale IV" (sections of Electrical engineering and Microengineering) |
| SEP-DEC 2016                 | Teaching assistant: "Physique générale I" (section of Mechanical engineering)                        |

## COMPUTER LANGUAGES AND SOFTWARE

---

- Fluent use of **Matlab**, **Mathematica**, **C++** and **Python**
- Basic knowledge in **Fortran**, **Wordpress**, **HTML** and **CSS**

## INTERESTS AND ACTIVITIES

---

Handball	Player (2010-2022) Coaching with certification J+S (2017-2022), "Jugend+Sport" - Licence E Volunteering : Webmaster and Community Manager in the Handball club.
First Aid	First Aid volunteer at EPFL with IAS1 and IAS2 certification (2017 - Present) Certified head of event first aid post (coordinating the team of first aiders) (2021-Present)

Creative hobbies (Drawing, Photography), Cooking, Sports and Traveling

## REFERENCES

---

P. De Los Rios, Associate Professor in Laboratory of Statistical Biophysics (LBS), EPFL

[paolo.delosrios@epfl.ch](mailto:paolo.delosrios@epfl.ch)

M. Hursin, Senior Scientist in Laboratory of Reactor Physics and Systems Behaviour (LRS), EPFL

[mathieu.hursin@epfl.ch](mailto:mathieu.hursin@epfl.ch)

M. Wyart, Associate Professor in Physics of Complex System Laboratory (PCSL), EPFL

[matthieu.wyart@epfl.ch](mailto:matthieu.wyart@epfl.ch)