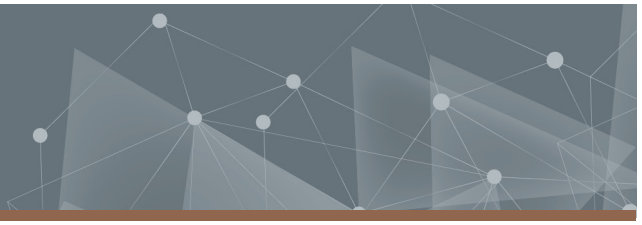
CHALMERS
UNIVERSITY OF TECHNOLOGY



Non-equilibrium stabilisation of proteins by chaperones

The Hsp70 molecular chaperone

Master's thesis in Complex Adaptive Systems

MARIA LI LÓPEZ BAUTISTA

DEPARTMENT OF LIFE SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

Master's thesis 2024

Non-equilibrium stabilisation of proteins by chaperones

The Hsp70 molecular chaperone

MARIA LI LÓPEZ BAUTISTA

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Life Sciences
Division of Chemical Biology
Chalmers University of Technology
Gothenburg, Sweden 2024

Non-equilibrium stabilisation of proteins by chaperones
The Hsp70 molecular chaperone
MARIA LI LÓPEZ BAUTISTA

Supervisor: Paolo De Los Rios, Laboratory of Statistical Biophysics, École Polytechnique Fédérale de Lausanne (EPFL)
Examiner: Pernilla Wittung-Stafshede, Department of Life Sciences, Chalmers University of Technology

Master's Thesis 2024
Department of Life Sciences
Division of Chemical Biology
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2024

Non-equilibrium stabilisation of proteins by chaperones
The Hsp70 molecular chaperone
MARIA LI LÓPEZ BAUTISTA
Department of Life Sciences
Chalmers University of Technology

Abstract

Within living organisms, proteins are essential components. They are responsible for carrying out almost every function in the cell. Proteins must fold into a specific three-dimensional shape to perform their diverse roles effectively. Indeed, improper folding is the root cause of many diseases.

Not surprisingly, there is a specific group of proteins whose function is to assist and safeguard the folding process of other proteins. These are known as molecular chaperones. Here, we focus on the 70 kiloDalton heat shock protein, Hsp70, a molecular chaperone that has been under the spotlight of scientists for decades due to its ubiquitous presence across all living systems and its assistance in a wide range of cellular processes. It is well accepted that the mechanism of action of Hsp70 chaperones consists of a biochemical energy-consuming cycle, which allows them to drive the system out-of-equilibrium and escape the inherent limitations of equilibrium thermodynamics to perform their functions efficiently.

Considering both the molecular details of chaperones and their client protein, along with a correct inclusion of the energy consumed in each step of the cycle and all relevant conformational transitions, we present a kinetic rate model for the description of the functional cycle of Hsp70 chaperones in protein folding that aims to elucidate the fundamental principles that govern their complex behaviour.

Keywords: Hsp70, protein folding, non-equilibrium systems, kinetic modelling.

Acknowledgements

I would like to express my most sincere gratitude to Paolo De Los Rios for welcoming me, a theoretical physicist at heart, and for introducing and guiding me through the complex world of biology. Likewise, I extend my appreciation to the rest of the group — Zoé, Shiling, Hugo, Mathieu, Satyam, Davide, and Alexandra. Without them, my time at Lausanne and this experience would not have been the same. Last but not least, I am grateful to Pernilla Wittung-Stafshede for willingly offering to go beyond the role of an examiner. On a personal note, I would also like to thank my family for being a constant source of motivation through their infinite support, patience, and encouragement throughout my academic journey.

Maria Li López Bautista, Gothenburg, June 2024

Contents

List of Abbreviations and Acronyms

Below is the list of abbreviations and acronyms that have been used throughout this thesis listed in alphabetical order:

ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
DNA	Deoxyribonucleic acid
FLuc	Firefly Luciferase
Hsp	Heat shock protein
KJE	DnaK/DnaJ/GrpE
JDP	J-domain protein
MD	Molecular Dynamics
NBD	Nucleotide Binding Domain
NEF	Nucleotide Exchange Factor
NESS	Non-equilibrium Steady State
RNA	Ribonucleic acid
PQS	Protein Quality Control
SBD	Substrate Binding Domain
sHsp	Small heat shock protein

List of Figures

List of Tables

1
Introduction

Usually, we think about proteins as something we eat to have a healthy and balanced diet or to help grow our muscles. Although it is true that they are important nutrients and the main components of muscles, they are much more than this. Indeed, proteins are complex biological molecules that are present in all living organisms. They are involved in a diverse range of cellular functions, from providing structural support to stimulating immune responses [1], which makes them essential to understand life and to be often considered as the workhorse molecules of the cell [2].

Such functional versatility arises from the fact that proteins come in a wide variety of shapes, ranging from elongated (fibrous) to ball-like (globular) structures [3], and their shape is closely related to their function. Usually, due to external influences, the same protein can be found in (often slightly) distinct conformations. While this attractive characteristic is sometimes crucial to perform its designated biological function, it can be dangerous in other cases. For instance, several diseases like diabetes type 2 or neurodegerative disorders are suspected to be caused by the incorrect folding (misfolding) of proteins followed by the formation of toxic aggregates [4]. More interestingly, proteins themselves contain the information to, at physiological conditions, fold into their three-dimensional functional conformation [5], namely the native state, which in turn must correspond to their free-energy minimum. Therefore, a protein should always be able to spontaneously fold correctly and be thermodynamically stable in that conformation. However, this is more of an ideal case since in the crowded environment of the cell the process of folding is prone to errors and proteins can populate non-native dysfunctional states corresponding to local minima [6]. In addition, stress conditions, such as an extreme temperature or exposure to chemical agents, can affect the stability of proteins causing them to lose their native conformation (unfold or misfold) and become non-functional. This is often referred to as protein denaturation [7]. So, in general, the native state is not guaranteed to always be the most stable state accessible to the protein.

In essence, while proteins must assume their native conformation and remain relatively stable to carry out their biological function, processes inherent to the physics of proteins can lead to unfolding or misfolding and further aggregation. These abnormalities in the folding of proteins can subsequently induce severe cell damage. Fortunately, cells are equipped with a protein quality control (PQC) system [8] to manage these off-pathway reactions and reduce the presence of harmful non-native proteins. The PQC system has two operational modes: degradation and repair [9]. On the one hand, irreversibly damaged proteins are disintegrated (degraded) by some

protein complexes. On the other hand, structurally altered proteins are natively refolded (repaired) with the help of a specific class of proteins known as molecular chaperones. In this project, we are interested in the latter.

Molecular chaperones were traditionally thought to simply function as a passive prevention mechanism against aggregation [10]. By binding to unstable non-native (unfolded or misfolded) proteins and 'holding' them, chaperones could avoid the formation of irreversible aggregates. Moreover, the reduction in the concentration of aggregation-prone proteins could induce a kinetic partitioning between the folding and aggregation pathways [11], decreasing the rate of aggregation since the process follows second-order kinetics[1], and rendering native folding the dominant process as it obeys first-order kinetics[2]. However, there is strong evidence suggesting that some chaperones are also able to actively rescue stable misfolded or aggregated proteins by consuming (hydrolysing[3]) ATP to conformationally remodel the client protein (substrate) they bind to [12, 13]. For example, among the various families of molecular chaperones known to date, the 70 kilo-Dalton heat shock proteins (Hsp70s) are widely recognised for their role in facilitating protein disaggregation and refolding [14, 15].

Hsp70s are particularly relevant due to their ubiquitous presence among living organisms [16] and their supervision of a diverse range of cellular processes [17, 18]. It is well accepted that their mechanism of action consists of a complex ATP-fuelled conformational cycle [19]. More precisely, the constant energy consumption by Hsp70s drives the system's dynamics away from equilibrium, leading to the establishment of a non-equilibrium steady state that is non-trivially governed by all the kinetic rates of the chemical reactions involved in the chaperone's cycle [20, 21]. Several experimental and theoretical studies suggest that this allows the enhancement of the native protein substrates stability even under denaturing conditions, in which the native state is thermodynamically less stable than the non-native species [22, 23, 24].

1.1 Objective

Based on the knowledge about the mechanism of action of Hsp70s in protein folding, the main purpose of this master's thesis is to create a kinetic model of the Hsp70 biochemical cycle that accurately captures the key characteristics observed in theoretical and experimental models to date. At the same time, we will include both the molecular details of Hsp70s and substrates along with the energy consumed at each step of the cycle, considering all relevant conformational transitions. To ensure thermodynamic consistency, the model should also hold under equilibrium conditions, even though most interesting phenomena occur under non-equilibrium conditions.

[1] A reaction that depends quadratically on the concentration of reactants.

[2] A reaction that depends linearly on the concentration of reactants.

[3] Chemical reaction where a phosphate bond on adenosine triphosphate (ATP) is broken by water, thereby releasing energy and resulting in a adenosine diphosphate (ADP) and an inorganic phosphate (P_i).

Research questions and limitations

Even though numerous experimental observations and groundbreaking discoveries have led to a general consensus on the action cycle of Hsp70s, some aspects still need a better overall understanding. In this project, with the help of physics principles, we hope to provide an answer to the key determining features of the Hsp70 chaperone cycle in protein folding.

Notwithstanding the above, it should be remembered that we use a coarse-grained computational model to describe a complex system, with the inherent benefits and limitations that it entails. Furthermore, while there are far more detailed models that explicitly account for all the chemical reactions involved in protein folding, see for example [25], we only consider the fundamental ones. This decision is to avoid an overly complicated model that is unsuitable for an intuitive interpretation since adding too many details from the start can hinder the essential underlying principles.

1.2 Thesis outline

First, we provide a biological insight into the problem of protein folding and the role of molecular chaperones, with a focus on Hsp70s. This is followed by the physical concepts necessary to describe a biological system using kinetic rate equations, including both in and out of equilibrium constraints derived from thermodynamic principles. Based on all this information, it is presented a simple model in which the fundamental reactions of the functional cycle of Hsp70s in protein folding are included. Subsequently, two models for protein folding are described. Finally, all the results obtained from the simulations of the models are presented in the end.

2

Biological background

In this chapter, we provide the information necessary for a better understanding of the problem of protein folding and the role of molecular chaperones, in specific Hsp70s.

2.1 Proteins

As mentioned before, proteins form the main building blocks of life as they undertake the majority of the work happening in the cell. Proteins are synthesised by a process called translation, in which the genetic information (originally encoded in the DNA[1]) transcribed into a messenger RNA[2] (mRNA) molecule and other cellular components intervene. As a result[3], proteins are composed of basic units, called amino acids, which are attached to each other in long chains. The particular order of this sequence of amino acids, combined with intramolecular forces, leads to the folding of the protein in a specific three-dimensional structure, which ultimately determines its function.

Since proteins are integral to numerous cellular processes, their proper folding is critical. Indeed, to achieve biological activity, proteins must adopt a specific three-dimensional conformation known as the native state. This precise folding is essential for the proteins to perform their designated roles within the cell effectively.

In vitro[4] experiments [26, 27] suggest that most proteins are able to reach their native structure spontaneously (without assistance of other molecules) in a physiological environment, following a process determined purely by their amino acid sequence. Therefore, the native state is the one that corresponds to the minimum of the Gibbs free-energy (Figure 2.1 i) and hence is the most thermodynamically stable configuration. Moreover, it is widely accepted that protein folding takes place over a funnel-like free-energy landscape, representing the steadily decreasing number of conformations available to the unstructured protein (high free-energy) as the folding process advances toward the native state [28].

[1] Deoxyribonucleic acid.

[2] Ribonucleic acid.

[3] We acknowledge that we do not give a very precise biologically description here, but we believe that it is sufficient within the scope of this project.

[4] Outside of the living organism, often in a test-tube.

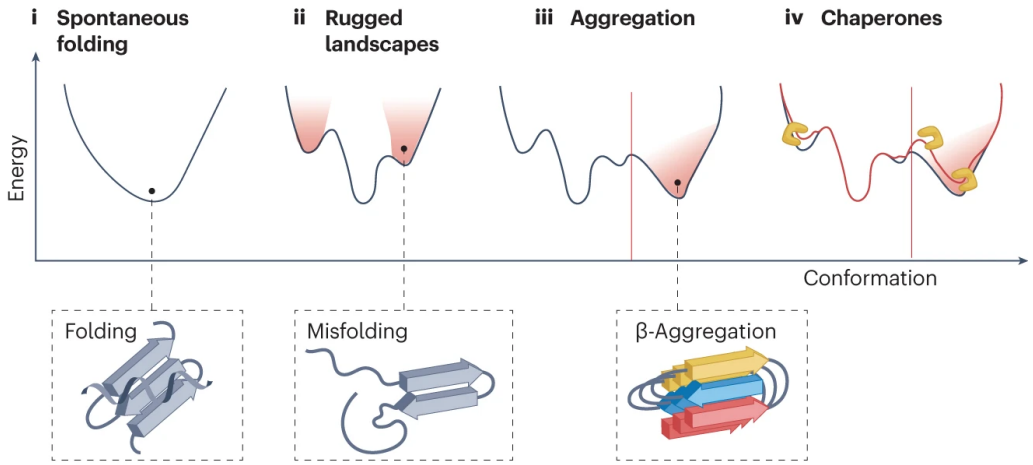


Figure 2.1: Free-energy landscape of protein folding, misfolding, and aggregation. i) In the ideal case, the folding process exhibits a smooth funnel energy landscape. ii) In reality, the landscape is often rugged and intermediate non-native (misfolded) states can be populated. Sometimes, this causes proteins to leave their original native folding trajectory, resulting in the iii) formation of aggregates. iv) Molecular chaperones guide proteins through the landscape by interacting and remodelling non-native, but not native, conformations, thus avoiding aggregation and promoting native folding. Figure adapted from [29].

Nevertheless, in vivo[5] and in vitro, it has also been observed that many proteins can acquire non-native conformations [6]. This indicates that the spontaneous folding process is prone to errors and local minima in the otherwise smooth free-energy landscape, where proteins are kinetically trapped in metastable misfolded states (Figure 2.1 ii). In such scenario, the accumulation of misfolded proteins can lead to the formation of inactive aggregates (Figure 2.1 iii), which can be thermodynamically more stable than the native state, as may occur at high protein concentrations. So, even though evolution has tried to select protein sequences that are stable in their native configuration at physiological conditions [30], their stability and foldability can be challenged by the appearance of some ruggedness in the free-energy landscape.

In fact, under certain environmental conditions, like heat shocks or chemical stress, the free-energy landscape of proteins can be perturbed in such a way that the native state is no longer the most stable configuration but is intrinsically less stable than the non-native conformations, therefore leading to protein misfolding and aggregation, which can eventually result in the formation of toxic aggregates that have deleterious effects on cellular functions and overall organismal health [31, 29]. Indeed, protein aggregates have been associated to several diseases, such as type 2 diabetes and neurodegenerative disorders like Parkinson's or Alzheimer's [32, 33].

Given the fatal consequences of protein misfolding and aggregation while being unavoidable processes for cells since they stem from the physics of proteins, organisms,

[5]Inside of the living organism.

in response, developed the cellular protein quality control (PQC) system [8] with the aim of minimising the presence of non-native proteins. One of the main components of this elaborate machinery is a specific group of proteins called molecular chaperones [34], whose main role is to avoid the aggregation of misfolded proteins and, in some cases, also to bring them back to the pathway of proper native folding [35]. Roughly speaking, molecular chaperones prevent or promote structural transitions with the aim of directing proteins toward its native conformation (Figure 2.1 iv). A more extensive description of these proteins is given in the following section.

2.2 Molecular chaperones

Recapitulating the description given in the previous section, molecular chaperones are a class of proteins that ensure the correct (native) folding of other proteins and thus their functionality, avoiding their misfolding and aggregation.

Molecular chaperones comprise a very diverse group of proteins. Several chaperone families have been identified, among the main conserved[6] ones are Hsp60 (GroEL), Hsp70 (DnaK), Hsp90 (HtpG), and Hsp100 (ClpB) [36] (bacterial homologues names in brackets). The acronym Hsp stands for heat shock protein and it is a reminder of their experimental correlation with thermal stresses, such conditions provided the first evidence of their existence and hinted their contribution to the cell, though they also participate in the response to other stresses and the regulatory (housekeeping) processes of cells under non-stressful conditions. The number indicates their approximate molecular mass in kiloDaltons (kDa). While having completely different structures and molecular mechanisms of action, they do share some commonalities: they bind to the client protein (usually called substrate) and hydrolyse ATP, classifying them as ATPases. Likewise, they use the energy produced in the ATP hydrolysis to perform mechanical work and remodel the conformation of the substrate. As a consequence, they are also known as ATP-driven or ATP-dependent molecular chaperones [37, 38].

Here, it is worth mentioning that in the cell the production of energy is continuous since the concentrations of ATP and ADP are kept fixed to maintain the cell alive [39]. This is particularly significant, since this constant supply of energy places the system out-of-equilibrium, driving it after a relaxation time to a non-equilibrium steady state (NESS) [40] in which the distribution of conformational states between the chaperones and their substrate is different from that in equilibrium [20, 21]. More details on this in the case of Hsp70s are given in the next section.

To complete the list of main conserved chaperone families, we must also mention the small (around 12-43 kDa) heat shock proteins (sHsps) [41]. These, at variance with the previous heat shock proteins, are not ATPases (i.e. they do not hydrolyse ATP) and their effect is thus constrained to the bounds imposed by equilibrium

[6] In an evolutionary perspective, namely, molecular chaperones have a high degree of sequence conservation across most domains of life.

thermodynamics. In addition, they are traditionally assumed to take the passive role of the holding mechanism in the sense that they bind to non-native proteins to avoid their aggregation until conditions are again favourable for native folding or an active chaperone takes over. However, some studies have provided a different insight, potentially picturing them as active protein holders [42, 43] due to their ability to adapt their concentration depending on the amount of non-native proteins present. This once again reiterates the complexity entailed to the functioning of chaperones, which will be also noticeable as we focus on the Hsp70 family of chaperones in the following section.

2.3 The Hsp70 chaperone family

Among the highly conserved families of molecular chaperones present in living systems, Hsp70s are very attractive to study because of their ubiquitous presence in eukaryotic and prokaryotic (mainly in bacterial) cells [17] and their assistance in a large amount of fundamental cellular processes beyond protein folding [18, 44]. Examples include, but are not limited to, protein transport [45], signalling [46], and protein degradation [47].

The participation of Hsp70s in such diverse cellular functions is particularly interesting as, at first, it seems to be in contradiction with the fact that Hsp70s are structurally and functionally conserved in evolution [48]. Indeed, across and within species Hsp70s show a high sequence identity. For instance, the eukaryotic Hsp70 shares around 50% amino acid identity with the prokaryotic Hsp70 protein DnaK. And, the human Hsp70 (sub)family has at least a group of eight members that are very similar, having between 52-99% amino acid identity [49]. So, in general, many Hsp70s have a common domain structure, which results in functional redundancy; They share a single biochemical mechanism of action: an ATP-driven cycle of client-binding and release. Intuitively, it makes sense that Hsp70s should not be specific as in principle any protein could need their assistance. Yet, the earlier mentioned functional diversity requires correct selection of substrates: Hsp70s should associate promiscuously with misfolded proteins while being selective with folded proteins. The apparent paradox is solved by noting that generally molecular chaperones do not work alone but with other chaperones and Hsp70s are no exception. Precisely, they are usually assisted by the so-called cochaperones[7], which can regulate Hsp70s interaction with substrates and thus grant them specificity [50]. Additionally, Hsp70s functionality can benefit from the cooperation with other chaperone systems. For example, they are known to collaborate with Hsp100s during protein disaggregation [51].

Hsp70s consist of two distinct domains, a nucleotide-binding domain (NBD) and a substrate-binding domain (SBD) attached by a linker (see Figure 2.2). These two domains are allosterically coupled such that the type of nucleotide bound to the

[7]Molecular chaperones are usually accompanied by other proteins whose main role is to provide support to the chaperone, and therefore are usually referred as cochaperons.

NBD influences the structure of the SBD, which in turn affects the affinity[8] for the substrate [19]. In particular, when the NBD is bound to ADP, the SBD is in a closed configuration. While when the NBD is bound to ATP, the SBD is in an open configuration, hence is more accessible to the substrate. As a consequence, the ADP-bound state of Hsp70 binds to the substrate more stably (with higher affinity) than the ATP-bound state. Furthermore, this structural difference also has an effect on the binding/unbinding kinetics. Indeed, when ATP is bound, substrate association to and dissociation from the chaperone occurs at high rates, which are several orders of magnitudes faster than when ADP is bound [52].

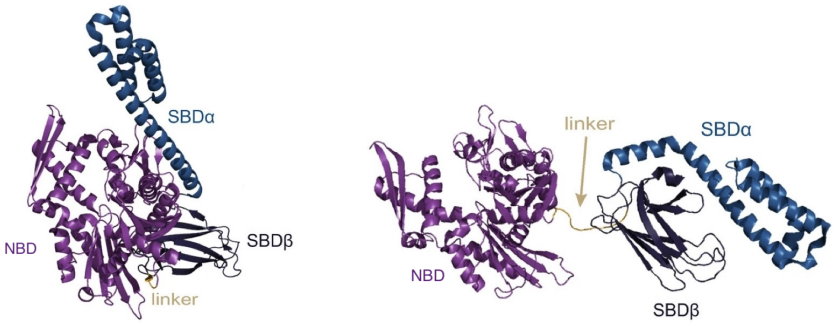


Figure 2.2: Hsp70s conformational states. Cartoon representation of the structure of Hsp70 in the ATP-bound state (left) and in the ADP-bound state (right). Figure adapted from [53].

The conversion between these two states can occur through two different processes: ATP hydrolysis[9] and nucleotide exchange. However, the spontaneous transition from one state to the other is extremely slow since the Hsp70s' ATP hydrolysis basal rate is low and the nucleotides (ATP and ADP) typically bind stably to Hsp70s (see Table 4.1). This reiterates the necessity of cochaperones for the regulation of Hsp70s action. In particular, Hsp70s most relevant cochaperones are the J-domain proteins (JDPs) [55] and the nucleotide exchange factors (NEFs) [56]. JDPs target the substrate and in synergism with the substrate enhance the rate of ATP hydrolysis whilst NEFs promote the exchange of ADP to ATP after hydrolysis, thus facilitating the release of the substrate [57].

Here it should be remarked that while from the ATP hydrolysis (non-equilibrium process) there is energy consumption by Hsp70s, from the nucleotide exchange there is not since it is an equilibrium process driven by thermal fluctuations. This has crucial implications in the dynamics of the system, as the populations of ATP-bound and ADP-bound states are no longer thermodynamically fixed by their free-energy difference, but are kinetically influenced by the ATP hydrolysis and nucleotide exchange rates.

[8] In chemistry, affinity is the tendency of different chemical species to form compounds.
[9] The rate of ATP synthesis by Hsp70s is negligible [54].

2.3.1 Hsp70s exploitation of ATP hydrolysis

Hsp70s activity depends on the nucleotide-regulated alternation between these two conformational states. Substrate binding not only benefits from the low-affinity of the ATP-bound state and high-affinity of the ADP-bound state, but also, and more importantly, from the time-scale separation between the binding/unbinding kinetics in the two states. To be exact, the fast-association rate of the ATP-bound state in combination with the slow-dissociation rate of the ADP-bound state results in a non-equilibrium (energy-consuming) affinity, commonly termed as ultra-affinity [20], that can go beyond the possible affinities in either the ATP- or ADP-bound states of Hsp70s (see Figure 2.3). Therefore, escaping the bounds imposed by equilibrium thermodynamics [58].

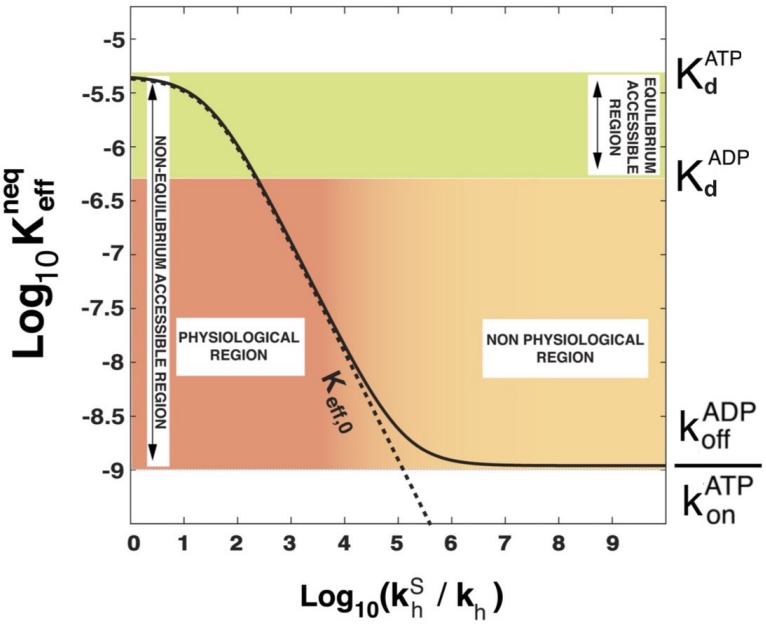


Figure 2.3: Non-equilibrium enhancement of Hsp70 affinity (ultra-affinity). Effective dissociation constant (product of the concentrations of free substrate and chaperone divided by the concentration of substrate-bound chaperone) of Hsp70 at non-equilibrium conditions as a function of the hydrolysis acceleration ratio $k_{\mathrm{h}}^{\mathrm{s}}/k_{\mathrm{h}}$. The dissociation constants (ratio of unbinding and binding rates) of the ATP- and ADP- bound states, respectively $K_{\mathrm{d}}^{\mathrm{ATP}}$ and $K_{\mathrm{d}}^{\mathrm{ADP}}$, are shown for reference. Likewise, the limit case where binding/unbinding to/from the ADP-bound state are negligible (black dashed line) and the lower bound (achieved when exclusively the substrate binds to the ATP-bound state and is only released from the ADP-bound state) of the dissociation constant are also included. Figure adapted from [20]. Refer to [20, 58] for further details.

Of note, on top of the difference in the (fast) ATP-bound and the (slow) ADP-bound timescales of substrate-exchange, from Figure 2.3 it can be observed that ATP hydrolysis must be stimulated several orders of magnitude to enter the ultra-

affinity regime. This reflects the influence of ATP hydrolysis on the chemical cycle of Hsp70s and the importance of energy consumption in their mechanism of action, as the ratio $k_{\mathrm{h}}^{\mathrm{s}}/k_{\mathrm{s}}$ is related to the total energy consumed [20]. Furthermore, in the expression of the effective non-equilibrium dissociation constant, refer to [20], it can be noted a dependence on the concentration of substrate, which results in the need of having the chaperone in excess over the substrate to reach ultra-affinity. The explanation of this condition is that in case of an excess of substrate there would be dominant substrate binding to the ADP-bound state, which would make impossible the utilisation of the time-scale separation in the binding/unbinding kinetics to achieve ultra-affinity.

In general, all these conditions are met in the cell by canonical[10] Hsp70s. Therefore, at physiological conditions, Hsp70s are able to exploit the energy provided by hydrolysis of ATP to overcome the inherent limitations of equilibrium thermodynamics.

Hsp70s conversion of chemical energy into mechanical work

As mentioned before, Hsp70s are ATPases (hydrolyse ATP) and use the energy produced to induce conformational changes in their substrates, which have a certain flexibility. In fact, the binding of Hsp70s to large proteins (polypeptides) provides their expansion [59], which depends on the presence of ATP [60]. Moreover, in cellular conditions, Hsp70s are efficient converters of ATP hydrolysis energy into mechanical work to unfold the substrate [61]. This relies on the inherent non-equilibrium dynamics associated to ATP hydrolysis in the cell, where the concentration of ATP is maintained in excess over the ADP one [62]. More precisely, Hsp70s generate a force due to excluded volume effects, namely of entropic origin (entropic pulling [63, 64, 65]), that requires the conversion of ATP chemical energy into ultra-affinity to guarantee the binding. This ultimately results in the unfolding of the attached substrate to Hsps70s (see Figure 2.4).

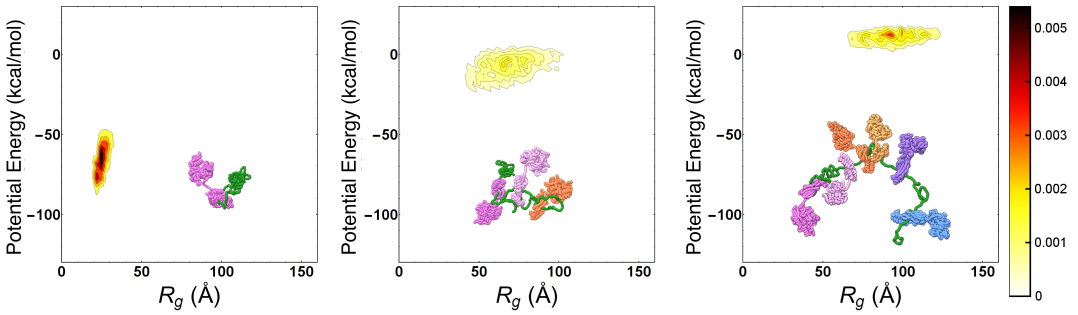


Figure 2.4: Expansion of the substrate (in green) by the binding of Hsp70s (in different colours). Probability density maps of the substrate potential energy and radius of gyration, R_g, for Hsp70-substrate complexes with one (left), three (center), and six (right) bound chaperones. Figure from [61].

[10]Structurally and functionally similar to the bacterial Hsp70, DnaK.

In particular, it can be observed that the progressive expansion of the chaperone-substrate complex and disruption of the attractive intramolecular interactions of the protein places the substrate at larger gyration radius[11] and higher potential energies. This increase in the free energy of the conformation penalises the binding of Hsp70s to the substrate, reflecting the growing difficulty of Hsp70s to bind to very spatially constrained (prone to steric clashes) regions.

It should be noted that while the binding of several Hsp70s leads to the stretching of the bound protein due to the steric effects between them, the binding of a single Hsp70 is also enough to induce a conformational change in the locked protein since there is also mutual exclusion of molecular volumes between the Hsp70 and the substrate.

2.3.2 Hsp70 chaperone in protein folding

Previously, we have mentioned that a protein must acquire its native conformation given that only in this state it is functional, i.e., it can perform its designated biological function. However, sometimes, in the cell the protein's native state is not the most stable state accessible to the protein. In fact, external stresses, such as an abrupt increment of temperature (heat shock) or addition of chemical agents like urea[12] (chemical stress) can interfere in the natural folding process, leading the protein to a dysfunctional non-native conformation. Moreover, such conditions can also be denaturing since they can be enough to destabilise the protein, causing it to lose its native structure and become non-functional. As a consequence, molecular chaperones are needed, either to facilitate the proper folding of the protein or to enhance (assist) its refolding (if removal of the stress is not enough for the protein recovery). In any case, the ultimate task presented to the molecular chaperone, namely Hsp70, is the same: ensure that the protein achieves and maintains its native conformation.

Within the role of Hsp70s in protein folding it is implicit that they should be able to act on a broad range of substrates. Hsp70s are attracted to a certain regions in the protein that consists of a sequence of 5-7 sticky water-repellent (hydrophobic) amino acids. Such segments are typically exposed on the surface when the protein is in a non-native conformation and buried inside the protein's core when correctly folded [66]. As a consequence, Hsp70s interact with unfolded, misfolded, or aggregated proteins but not with their folded (native) counterparts.

It is well accepted that the mechanism of action of Hsp70 follows a cycle of substrate (client protein) binding, remodelling (mechanically unfolding but not folding [59]) and release (see Figure 2.5). The precise steps of this cycle are the following: (1) JDP cochaperone targets the non-native protein and mediates the delivery of the substrate to the ATP-bound state of Hsp70, which is followed by (2) a rapid tran-

[11] In biology, the radius of gyration of a protein is considered to give a measure of the structure compactness and size of the protein. It is defined as the root mean square distance of all atoms from the centre of mass the protein.

[12] A colourless substance found in urine.

sition, as the hydrolysis rate is enhanced due to the presence of the substrate and the JDP, to the high affinity ADP-bound state of Hsp70. Hence, it needs the (3) NEF cochaperone to induce the dissociation of ADP and (4) binding of ATP in exchange, which leads to the (5) release of the unfolded substrate, thus allowing it to spontaneously refold or misfold, causing the substrate to undergo another cycle in the latter.

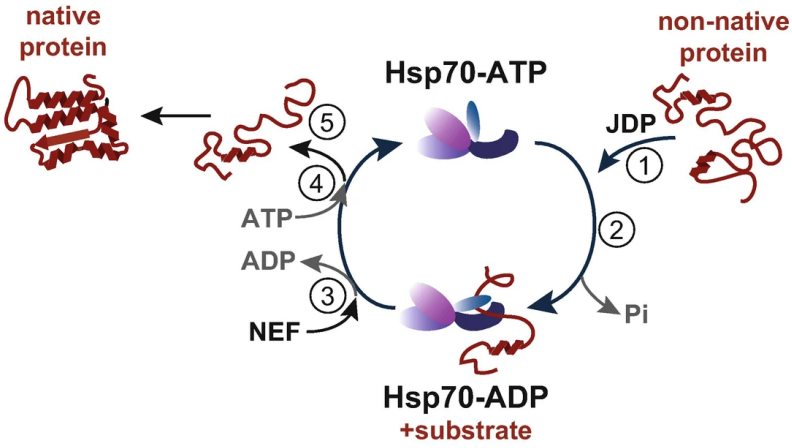


Figure 2.5: Functional cycle of the Hsp70 chaperone. Schematic representation of the conformational transitions in the canonical model of Hsp70 in protein folding. Figure adapted from [53]. Refer to the main text for further information about the steps along the cycle.

Here, it should be clarified that we do not specialise in a specific member of the Hsp70 chaperone family and rather assume the so-called canonical model for the functioning of Hsp70s [66]. This includes the regulating action of Hsp70s' cochaperones: JDPs delivering the protein and NEFs driving its release.

Regulation by JDPs

JDPs, sometimes referred to as the Hsp40 chaperone family, share a conserved stretch of amino acids, called the J-domain (named after the bacterial Hsp40, DnaJ), but not the substrate-binding domains [67]. The first domain is involved with the ATPase activity of Hsp70s, while the latter is essential in aiding the accomplishment of the mentioned Hsp70s' multifunctionality [66].

Until this point, it has been implicitly considered that during the functional cycle of Hsp70s JDPs interact both with substrates and Hsp70s. Indeed, the assumed role of JDPs cochaperones is to recognise Hsp70s' substrates and enable their interaction with the chaperones. Moreover, in experiments it has been observed that their presence affects the ATP hydrolysis rate, slightly enhancing it at cellular conditions [68] where the concentration of JDP is an order of magnitude lower than the one of Hsp70 [69].

However, and more interestingly, the joint presence of JDPs and the substrate (which

accelerates ATP hydrolysis about an order compared to the basal rate [70]) drastically stimulates ATP hydrolysis, with their resulting impact higher than the combination of their individual effect: the ATP hydrolysis rates is increased by several orders of magnitude [70].

Regulation by NEFs

NEFs, recognised to date, comprise four different types of protein families with no sequence similarity and structurally unrelated [66, 44]. As a consequence, it is known that they use different mechanisms in their task of inducing ADP release, which in turn enables ATP binding and substrate dissociation from Hsp70s. Furthermore, NEFs can also in part control the timing of nucleotide exchange.

Nevertheless, it has been observed that some Hsp70s may function independently of the presence of NEFs as they have high intrinsic ADP dissociation rates [71, 72].

3

Physical theory

Biological systems are often described using kinetic networks, in which the nodes represent the possible states of the system and the biochemical reactions of the system configure the connections (transitions) between them. Through this coarse-grained description, one can model the dynamical evolution of the system, that is, how the population of each state changes over time.

Within the framework of this project, it is assumed that the time evolution of the concentrations of the states is dictated by deterministic mass-action kinetics. Any possible fluctuations are neglected on the basis that the concentrations of molecular chaperones and substrates are high enough in standard cellular conditions. Otherwise, a more detailed stochastic description of the process should be considered [40].

Before moving into the mathematical details needed for the description of these systems, let us review the meaning of mass-action kinetics.

The law of mass action

The law of mass-action, introduced by Guldberg and Waage [73, 74, 75], states that the rate at which a chemical reaction occurs is proportional to the probability of collision of the reactants. This probability, assuming spatial homogeneity, is in turn proportional to the concentration of the reactants, each elevated to the power equal to its coefficient in the reaction.

To get a clearer idea of this definition, let us consider a reversible reaction[1] of the form,

$$a\,\mathrm{A} + b\,\mathrm{B} \underset{k_b}{\overset{k_f}{\rightleftharpoons}} c\,\mathrm{C} + d\,\mathrm{D}, \tag{3.1}$$

where k_f and k_b are, respectively, the rate constants (generally independent of space and time) of the forward and backward reactions.

The reaction rate (net flux) is defined as

$$r = r_f - r_b, \tag{3.2}$$

[1] A chemical reaction is reversible if it can proceed in a forward a backward direction, i.e., both the conversion of reactants to products and the conversion of products to reactants are possible. Conversely, a reaction is irreversible if only one of the direction is possible.

where are the forward and backward rates (fluxes) are given by

$$\begin{aligned} r_f &= k_f[\mathrm{A}]^a[\mathrm{B}]^b \\ r_b &= k_b[\mathrm{C}]^c[\mathrm{D}]^d. \end{aligned} \tag{3.3}$$

At this point, it is also beneficial to introduce the equilibrium constant, a quantitative measure that characterises the relation between the concentrations of reactants at equilibrium.

Equilibrium constant

The equilibrium constant of the reaction is defined as the ratio of reactant and product concentrations at equilibrium, that is, when the forward and backward rates are equal.

$$K_{\mathrm{eq}} \equiv \frac{k_f}{k_b} = \frac{[\mathrm{C}]^c_{\mathrm{eq}}[\mathrm{D}]^d_{\mathrm{eq}}}{[\mathrm{A}]^a_{\mathrm{eq}}[\mathrm{B}]^b_{\mathrm{eq}}}. \tag{3.4}$$

3.1 Rate equations

The kinetic scheme of a biological system under the law of mass action (Equations (3.2), (3.3)) leads to a system of coupled ordinary differential equations, commonly referred to as rate equations.

As a result, the time evolution of a biological system of N states with first-order kinetics is as follows

$$\frac{dc_i}{dt} = \sum_{j=1}^{N} k_{ji}c_j - \sum_{j=1}^{N} k_{ij}c_i, \tag{3.5}$$

where the solution, $c_i(t)$, is the concentration of state i ($i = 1, ..., N$) and k_{ij} is the transition rate from state i to state j, assuming always $j \neq i$. The first sum represents the total flux of molecules from other states toward state i, namely the gain term. While the second sum represents the total flux of molecules from state i toward other states, namely the loss term.

Let us remark some relevant properties of the system.

Conservation of mass

The law of conservation of mass in Equation (3.5) is implicitly implied as by construction

$$\sum_i \frac{dc_i}{dt} = 0, \tag{3.6}$$

and hence, $\sum_i c_i = \text{const.}$ at all times.

Steady-state condition

A well-known asymptotic behaviour of dynamical systems is to become stationary. In the case of system in Equation (3.5) it means that after a sufficiently long time the concentrations no longer depend on time, i.e.,

$$\frac{dc_i}{dt} = 0 \ \forall i. \tag{3.7}$$

This condition defines the steady-state distribution of the system.

3.2 Equilibrium and detailed balance

In physics, a system is considered to be in equilibrium when there are no net flows or changes going through it. Thus, the system maintains a fixed configuration of states unless externally perturbed.

For the system in Equation (3.5) to be in equilibrium, the fluxes between any connected states need to be individually balanced, i.e.,

$$k_{ij} c_i^{\mathrm{eq}} = k_{ji} c_j^{\mathrm{eq}} \ \forall i, j \tag{3.8}$$

where c_i^{eq} is the equilibrium concentration of state i. This condition is known as the principle of detailed balance. Here it should be noticed that when detailed balanced is respected, so is the steady-state condition (Equation (3.7)). Therefore, the system is in an equilibrium steady state.

And, for a any reaction from state i to j the equilibrium constant (Equation (3.4)) is given by

$$K_{ij} = \frac{k_{ij}}{k_{ji}} = \frac{c_j^{\mathrm{eq}}}{c_i^{\mathrm{eq}}}, \tag{3.9}$$

which draws a connection between the kinetic description of biological systems and thermodynamics. Further details are provided in the next section.

In the presence of a closed loop (cycle) in the chemical network, the detailed balance conditions for the states involved are satisfied only if

$$\frac{\prod_i k_{i,i+1}}{\prod_i k_{i+1,i}} = 1, \tag{3.10}$$

where the states in the cycle are ordered so that state i is connected with state $i+1$ by a reversible reaction with nonzero rates $k_{i,i+1}$ and $k_{i+1,i}$, with the exception of the last state that is connected to the first state. The derivation to go from Equation (3.8) to Equation (3.10) in the case of a system of three components is showed in Appendix A.1.

The meaning of Equation (3.10) is that in a cycle the product of the rates in one direction is equal to the product of rates in the opposite direction. This is the Wegscheider-Lewis cycle condition [76, 77].

3.2.1 Reaction kinetics and thermodynamics

From the principles of (reversible) thermodynamics, the Gibbs free energy[2] of a chemical system at constant temperature and pressure is

$$G = \sum_i n_i \mu_i, \tag{3.11}$$

where n_i is the population of state i and μ_i is the chemical potential of state i, with the latter given by

$$\mu_i = \mu_i^\circ + k_B T \ln \frac{c_i}{c^\circ}, \tag{3.12}$$

where μ_i° is the standard[3] chemical potential, $c^\circ (= 1\text{M})$ is the standard concentration, k_B is the Boltzmann's constant and T is the absolute temperature of the system.

Following from Equation (3.11), the free energy of an arbitrary state i can be defined as $G_i = n_i \mu_i$. Considering the formula of the chemical potential (Equation (3.12)), the Gibbs free energy difference between states i and j is the following

$$\Delta G_{ij} = G_j - G_i = \Delta G_{ij}^\circ + k_B T \ln \frac{c_j}{c_i}, \tag{3.13}$$

where it has been assumed that the difference in chemical potential is equal to the difference in free energy both at standard and non-standard conditions, i.e., $\Delta G_{ij}^\circ = \Delta \mu_{ij}^\circ$ and $\Delta G_{ij} = \Delta \mu_{ij}$, since it only implies a change in the units that the free energy is expressed.

This free-energy difference can be understood as the (potential) energy needed to go from state i to state j. And hence, it gives a measure of the spontaneity of the chemical reaction between the two states. Indeed, the reaction is spontaneous if $\Delta G_{ij} < 0$ and there is energy release. On the contrary, the reaction is not energetically favourable if $\Delta G_{ij} > 0$, and needs external energy to happen.

From the perspective of thermodynamics, a system has reached its equilibrium distribution when there is no net macroscopic flows of matter or energy within the system and thus no work can be extracted. So, a system is in equilibrium if and only if $\Delta G_{ij} = 0$. Introducing this constraint in Equation (3.13) leads to

$$\Delta G_{ij}^\circ = -k_B T \ln \frac{k_{ij}}{k_{ji}} \tag{3.14}$$

where the definition of the equilibrium constant between two connected states i and j (Equation (3.9)) has been applied. It should be noted that this condition is just another way of expressing the detailed balance condition (Equation (3.8)).

[2] In physics, the change in Gibbs free energy is defined as the energy available to perform non-expansion (without volume changes) work in a closed system at constant temperature and pressure.

[3] At standard conditions. In chemistry, defined by a temperature $T = 298$ °K (25 °C), and a pressure $p = 1$ bar.

The free-energy difference of a cycle is then given by

$$\Delta G_{\text{cycle}} = \sum_i \Delta G_{i,i+1} = \sum_i \Delta G^{\circ}_{i,i+1} = -k_B T \ln \frac{\prod_i k_{i,i+1}}{\prod_i k_{i+1,i}}, \tag{3.15}$$

where the cycle follows the same definition given above.

Therefore, a chemical system that contains a cycle is in equilibrium if $\Delta G_{\text{cycle}} = 0$.

3.3 Non-equilibrium and NESS

It has been mentioned that the fulfilment of the detailed balance condition gives a condition to have an equilibrium stationary state. However, the opposite is not always true. The existence of a steady state does not guarantee that detailed balance holds and that the system is in equilibrium.

In fact, the system can reach a state where the concentrations are independent of time due to a compensation between the total gain and loss, but the net flux is not zero along all branches of the network. Thus, the detailed balance condition is not satisfied and the system is not in equilibrium. This is called as the non-equilibrium steady state (NESS).

Such state is maintained by a circular balance (cyclic flux) that usually requires a external driving force [78, 40]. As a consequence, the detailed balance condition in Equation (3.14) generalises to the so-called local detailed balance [79]

$$\frac{k_{ij}}{k_{ji}} = e^{-\beta(\Delta G^{\circ}_{ij} - F_{ij})}, \tag{3.16}$$

where $\beta = 1/k_B T$ and F_{ij} is the non-equilibrium driving force between the states i and j.

3.3.1 Source of non-equilibrium: ATP hydrolysis

In living systems, the primary source of energy comes from the hydrolysis of ATP. During ATP hydrolysis a phosphate bond on ATP is broken by water (H_2O), which results in a ADP and an inorganic phosphate (P_i). The chemical reaction is then

$$\text{ATP} \; + \; \text{H}_2\text{O} \underset{k_{\text{s}}}{\overset{k_{\text{h}}}{\rightleftharpoons}} \; \text{ADP} \; + \; \text{P}_i, \tag{3.17}$$

where k_{h} is the rate of hydrolysis. And, it has also been included the reversed process: ATP synthesis, which has a rate k_{s}.

The available energy from the reaction, using Equations (3.13) and (3.14), is given by

$$\Delta G_{\text{ATP}} = -\Delta G_{\text{ATP,ADP}} = k_B T \ln \left\{ \frac{[\text{ATP}]}{[\text{ADP}][\text{P}_i]} - \frac{[\text{ATP}]_{\text{eq}}}{[\text{ADP}]_{\text{eq}}[\text{P}_i]_{\text{eq}}} \right\}. \tag{3.18}$$

It should be noted that the concentration of H_2O does not appear in the expression for the free-energy difference since it is assumed that the solutions are dilute so that the concentration of H_2O does not change during the reaction.

In the previous chapter, it has been constantly mentioned that in the cell ATP hydrolysis is a non-equilibrium process given that the concentration of ATP is maintained in excess over the concentration of ADP. To clarify this idea, let us consider the case of a chemical system of two states.

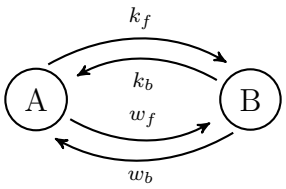


Figure 3.1: Network representation of a chemical reaction system of two states connected through two different pathways.

Applying the detailed balance condition, Equation (3.8), to the chemical newtork in Figure 3.1

$$\begin{aligned} \frac{k_f}{k_b} &= \frac{[\mathrm{A}]_{\mathrm{eq}}}{[\mathrm{B}]_{\mathrm{eq}}} \\ \frac{w_f}{w_b} &= \frac{[\mathrm{A}]_{\mathrm{eq}}}{[\mathrm{B}]_{\mathrm{eq}}} \end{aligned}, \tag{3.19}$$

and both ratios are equal to $e^{-\beta \Delta G^{\circ}_{\mathrm{AB}}}$. Moreover, it can be easily observed that the detailed balance condition for the cycle is also satisfied, i.e., $k_f w_b / w_f k_b = 1$.

If we include the hydrolysis of ATP in the top pathway, from the local detailed balance in Equation (3.16)

$$\frac{k_f}{k_b} = e^{-\beta \Delta G^{\circ}_{\mathrm{AB}}} e^{\beta \Delta G_{ATP}} \tag{3.20}$$

and then,

$$\frac{k_f w_b}{w_f k_b} = e^{\beta \Delta G_{ATP}}. \tag{3.21}$$

Therefore, unless $\Delta G_{\mathrm{ATP}} = 0$, the system is out-of-equilibrium. Indeed, this is the case in the cell since the concentrations of ATP and ADP are fixed (i.e., they do not reach their equilibrium values) to have a continuous production of energy.

In general, having $\Delta G_{\mathrm{ATP}} \neq 0$ is a necessary and sufficient condition to have a biological system displaced from its equilibrium distribution. This actually relies on the inherent presence of a cycle in the chemical network, as then different pathways connecting the same states lead to different equilibrium configurations and thus equilibrium can not be established since the detailed balance condition is broken.

4

Methods

In this chapter, we will provide a description of the kinetic models used for the characterisation of the role of the Hsp70 chaperone in protein folding. Here it should be reminded that we do not specialise in any specific member of the Hsp70 family. Instead, our models are based on the already known biochemical and biophysical information about canonical Hsp70s. Furthermore, to guarantee thermodynamic consistency we will consider all reactions reversible, even if the rate in one of the directions is much slower than in the other. Likewise, while most of the rates will be taken from the literature, some of them will be defined and/or constrained by relations imposed from the detailed balance condition (Equation (3.8) or Equation (3.14)) between pairs of connected states and in cycles (Equation (3.10)). These must hold in equilibrium conditions so that the kinetic model is consistent both at equilibrium and away from it.

4.1 Kinetic rate models

To begin with, let us recall the rate equations that describe a biochemical system of N states with first-order kinetics

$$\frac{dc_i}{dt} = \sum_j k_{ji} c_j - \sum_j k_{ij} c_i, \tag{4.1}$$

where the solution $c_i(t)$ is the concentration of state i $(i = 1, ..., N)$ and k_{ij} is the transition rate from state i to state j, assuming $j \neq i$.

This system of equations will be used for the description of the kinetic models of the Hsp70 chaperone system, as the reactions involved can be considered to follow first-order kinetics, and hence the transition rates k_{ij} are independent of time.

It is worth mentioning that by defining a vector $\mathbf{c} = (c_1, ..., c_N)^\top$ and a certain matrix $\mathbf{A}$ that only contains the transition rates and not the system concentrations since all the reactions follow first-order kinetics, the system in Equation (4.1) can be rewritten in a matrix formulation as

$$\frac{d\mathbf{c}}{dt} = \mathbf{A}\mathbf{c}, \tag{4.2}$$

whose general solution is

$$\mathbf{c}(t) = e^{\mathbf{A}t}\mathbf{c}(t = 0). \tag{4.3}$$

It should be noted that this solution is valid only under the constraint that the transitions rates are independent of time.

4.1.1 Hsp70s and ultra-affinity

In order to study the non-equilibrium effect of ATP-hydrolysis in the functional cycle of Hsp70s, we started by considering a simple model in which the substrate is in a misfolded state (M) that can bind to either a Hsp70 in the ATP-bound state (MT) or a Hsp70 in the ADP-bound state (MD), see Figure 4.1. These are coarse grained representations of all the possible structural conformations of these states at the molecular level.

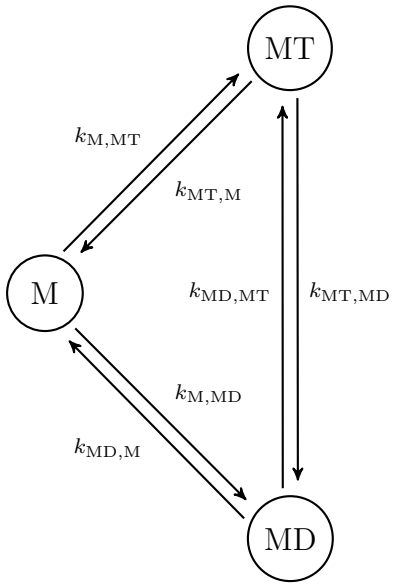


Figure 4.1: Kinetic model for ultra-affinity. Network representation of the biochemical system consisting of the misfolded substrate (M), the misfolded substrate bound to the ATP-bound state of Hsp70 (MT), and the misfolded substrate bound to the ADP-bound state of Hsp70 (MD). Refer to the main text for more details about the indicated transition rates.

The concentrations of this system are determined by the following rate equations

$$\begin{aligned}
\frac{d[\mathrm{M}]}{dt} &= -(k_{\mathrm{M,MT}} + k_{\mathrm{M,MD}})[\mathrm{M}] + k_{\mathrm{MT,M}}[\mathrm{MT}] + k_{\mathrm{MD,M}}[\mathrm{MD}] \\
\frac{d[\mathrm{MT}]}{dt} &= -(k_{\mathrm{MT,M}} + k_{\mathrm{MT,MD}})[\mathrm{MT}] + k_{\mathrm{M,MT}}[\mathrm{M}] + k_{\mathrm{MD,MT}}[\mathrm{MD}], \\
\frac{d[\mathrm{MD}]}{dt} &= -(k_{\mathrm{MD,M}} + k_{\mathrm{MD,MT}})[\mathrm{MD}] + k_{\mathrm{M,MD}}[\mathrm{M}] + k_{\mathrm{MT,MD}}[\mathrm{MT}]
\end{aligned} \tag{4.4}$$

where the transition rates (in s^{-1}) are defined as

$$\begin{aligned}
k_{\mathrm{M,MT}} &= k_{\mathrm{on}}^{\mathrm{ATP}}[\mathrm{Hsp70}\cdot\mathrm{ATP}]e^{-\beta\Delta G_{\mathrm{M,M\,70}}} & k_{\mathrm{MT,M}} &= k_{\mathrm{off}}^{\mathrm{ATP}} \\
k_{\mathrm{M,MD}} &= k_{\mathrm{on}}^{\mathrm{ADP}}[\mathrm{Hsp70}\cdot\mathrm{ADP}]e^{-\beta\Delta G_{\mathrm{M,M\,70}}} & k_{\mathrm{MD,M}} &= k_{\mathrm{off}}^{\mathrm{ADP}} \\
k_{\mathrm{MT,MD}} &= k_{\mathrm{ex,TD}}^{\mathrm{eff}} + k_{\mathrm{h}}^{\mathrm{s}} & k_{\mathrm{MD,MT}} &= k_{\mathrm{ex,DT}}^{\mathrm{eff}} + k_{\mathrm{s}}^{\mathrm{s}}
\end{aligned} \tag{4.5}$$

Here, it has been considered that the Hsp70 binding/unbinding rates to/from the substrate ($k_{\text{on}}^{\text{ATP}}, k_{\text{off}}^{\text{ATP}}, k_{\text{on}}^{\text{ADP}}, k_{\text{off}}^{\text{ADP}}$) depend on the type of nucleotide bound to the chaperone. Moreover, for the binding rates the conformational free-energy cost of Hsp70 binding ($\Delta G_{\text{M,M}\,70}$) as well as the concentration of Hsp70 in its two possible conformational states ($[\text{Hsp70}\cdot\text{ATP}], [\text{Hsp70}\cdot\text{ADP}]$) have been included. The total conversion rate between these two is given by the sum of the hydrolysis/synthesis rates ($k_{\text{h}}^{\text{s}}, k_{\text{s}}^{\text{s}}$) in the presence of the substrate, assuming JDP co-localisation, and the nucleotide exchange rates ($k_{\text{ex,TD}}^{\text{eff}}, k_{\text{ex,DT}}^{\text{eff}}$), which effectively describe the exchange process as the unbinding of a nucleotide and the binding of a different one are simultaneously taken into account.

Such system, actually, considers all relevant biochemical reactions of the Hsp70-substrate system. That is, chaperone binding/unbinding to/from the non-native misfolded protein (transitions between M and MT/MD), hydrolysis/synthesis and exchange (transitions between MT and MD) of the nucleotide bound to Hsp70. Indeed, these reactions form the fundamental basis of the functional cycle of the Hsp70 chaperone. So, let us provide more details about their rates.

Mathematical details

The core rates for Hp70s are taken from the literature in studies about the bacteria Escherichia coli (E. coli), see Table (4.1). In this organism, the Hp70-cochaperones system is formed by the DnaK/DnaJ/GrpE trio, usually referred to as the KJE system [23].

k_{+}^{ATP}	$1.3 \times 10^{5}\ \text{M}^{-1}\text{s}^{-1}$	k_{-}^{ATP}	$1.33 \times 10^{-4}\ \text{s}^{-1}$
k_{+}^{ADP}	$2.67 \times 10^{5}\ \text{M}^{-1}\text{s}^{-1}$	k_{-}^{ADP}	$0.022\ \text{s}^{-1}$
$k_{\text{on}}^{\text{ATP}}$	$1.28 \times 10^{6}\ \text{M}^{-1}\text{s}^{-1}$	$k_{\text{off}}^{\text{ATP}}$	$2.31\ \text{s}^{-1}$
$k_{\text{on}}^{\text{ADP}}$	$1000\ \text{M}^{-1}\text{s}^{-1}$	$k_{\text{off}}^{\text{ADP}}$	$2 \times 10^{-3}\ \text{s}^{-1}$
k_{h}	$6 \times 10^{-4}\ \text{s}^{-1}$	k_{h}^{s}	$1.8\ \text{s}^{-1}$

Table 4.1: Parameters of the kinetic rate models. Rates for nucleotide binding/unbinding to/from the chaperone: $k_{+}^{\text{ATP}}, k_{-}^{\text{ATP}}, k_{+}^{\text{ADP}}, k_{-}^{\text{ADP}}$, respectively. Rates for chaperone in the ATP and ADP states binding/unbinding to/from the substrate: $k_{\text{on}}^{\text{ATP}}, k_{\text{off}}^{\text{ATP}}, k_{\text{on}}^{\text{ADP}}, k_{\text{off}}^{\text{ADP}}$, respectively. And, rates for ATP hydrolysis without and with substrate: respectively k_{h} and k_{h}^{s}, with the latter considering also the co-localised presence of the JDP cochaperone. Taken from [52, 70, 60].

It can be observed that some of the rates mentioned in Equation (4.5) are not present in the table above. For instance, the rate for ATP synthesis. This, in general, can not be measured in experiments due to its very low value. Nevertheless, since we wanted to consider the reversibility of all reactions we derived an expression for it.

Further details are given below.

Let us start with the nucleotide exchange process as it is related to the rates of the other reactions.

- **Nucleotide exchange**

  During the exchange of the nucleotide bound to Hsp70, first the nucleotide bound is released and then the other nucleotide can bind. The reaction can be described by the following scheme

$$\mathrm{Hsp70\cdot ATP} + \mathrm{ADP} \underset{k_+^{\mathrm{ATP}}}{\overset{k_-^{\mathrm{ATP}}}{\rightleftharpoons}} \mathrm{Hsp70} + \mathrm{ATP} + \mathrm{ADP} \underset{k_-^{\mathrm{ADP}}}{\overset{k_+^{\mathrm{ADP}}}{\rightleftharpoons}} \mathrm{Hsp70\cdot ADP} + \mathrm{ATP}\,. \quad (4.6)$$

  It is shown, refer to [20], that solving the rate equations of this reactions under the assumption that the concentration of the nucloetide-free chaperone, [Hsp70], is the steady state one leads to an effective description of the process, which can be expressed by the following reaction scheme

$$\mathrm{Hsp70\cdot ATP} \underset{k_{\mathrm{ex,DT}}^{\mathrm{eff}}}{\overset{k_{\mathrm{ex,TD}}^{\mathrm{eff}}}{\rightleftharpoons}} \mathrm{Hsp70\cdot ADP}, \quad (4.7)$$

  where $k_{\mathrm{ex,TD}}^{\mathrm{eff}}$ and $k_{\mathrm{ex,DT}}^{\mathrm{eff}}$ take the following definitions

$$\begin{aligned} k_{\mathrm{ex,TD}}^{\mathrm{eff}} &= \frac{k_+^{\mathrm{ADP}} k_-^{\mathrm{ATP}}}{k_+^{\mathrm{ADP}} + k_+^{\mathrm{ATP}} \frac{[\mathrm{ATP}]}{[\mathrm{ADP}]}} \\ k_{\mathrm{ex,DT}}^{\mathrm{eff}} &= \frac{k_-^{\mathrm{ADP}} k_+^{\mathrm{ATP}} \frac{[\mathrm{ATP}]}{[\mathrm{ADP}]}}{k_+^{\mathrm{ADP}} + k_+^{\mathrm{ATP}} \frac{[\mathrm{ATP}]}{[\mathrm{ADP}]}} \end{aligned} \,. \quad (4.8)$$

  These effective constants correctly capture the nucleotide exchange process as the nucleotide-free chaperone state is generally non-populated in comparison with the ATP- and ADP-bound states given that the nucleotide binding rates are much faster than the nucleotide release rates, as exposed in Table 4.1, hence the transitions are rapid, and the concentrations of free nucleotides (ATP, ADP) are in high abundance at the cell, usually several orders higher than the chaperone concentration. Analogous expressions can be obtained in the presence of the substrate.

- **Hydrolysis/synthesis**

  In the previous chapter, it has been mentioned that presence of either the substrate or the JDP cochaperone can stimulate the rate for ATP hydrolysis, and that the presence of both results in a synergistic enhancement. In our models, we did not consider the explicit reactions behind these effects but implicitly included their contribution through the choice of the rate constants, see Table 4.1.

To obtain the ATP synthesis rates we considered a chemical network that contains the free nucleotides and the Hsp70 ATP- and ADP-bound states, see Figure 4.2.

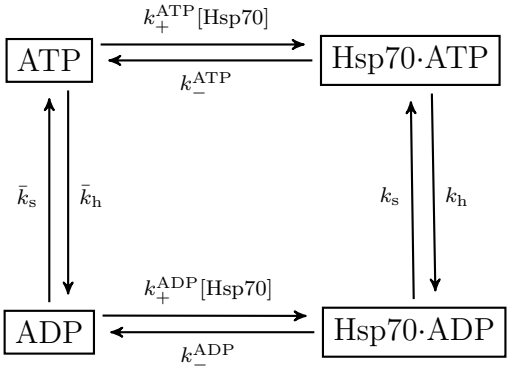


Figure 4.2: Network representation of a chemical system that contains the free nucleotides and the Hsp70 ATP- and ADP-bound states. The parameters $\bar{k}_\mathrm{h}$ and $\bar{k}_\mathrm{s}$, respectively, denote the hydrolysis and synthesis rates between the free nucleotides, whose values are not specified as these reactions are not explicitly considered in our models.

Combining the detailed balance conditions for the reaction between free nucleotides (Equation (3.8)) and for the cycle (Equation (3.10)) in Figure 4.2

$$\begin{aligned} &\bar{k}_\mathrm{h}[\mathrm{ATP}]_\mathrm{eq} = \bar{k}_\mathrm{s}[\mathrm{ADP}]_\mathrm{eq} \\ &\frac{\bar{k}_\mathrm{s} k_+^{\mathrm{ATP}}[\mathrm{Hsp70}] k_\mathrm{h} k_-^{\mathrm{ADP}}}{\bar{k}_\mathrm{h} k_+^{\mathrm{ADP}}[\mathrm{Hsp70}] k_\mathrm{s} k_-^{\mathrm{ATP}}} = 1 \end{aligned} \tag{4.9}$$

with the definitions of the effective nucleotide exchange rates (Equation (4.8)), results in

$$\frac{k_\mathrm{h}}{k_\mathrm{s}} = \frac{[\mathrm{ADP}]_\mathrm{eq}}{[\mathrm{ATP}]_\mathrm{eq}} \frac{[\mathrm{ATP}]}{[\mathrm{ADP}]} \frac{k_{\mathrm{ex,TD}}^{\mathrm{eff}}}{k_{\mathrm{ex,DT}}^{\mathrm{eff}}}. \tag{4.10}$$

And thus, the synthesis rate is given by

$$k_\mathrm{s} = k_\mathrm{h} \frac{k_-^{\mathrm{ADP}} k_+^{\mathrm{ATP}}}{k_+^{\mathrm{ADP}} k_-^{\mathrm{ATP}}} \frac{[\mathrm{ATP}]_\mathrm{eq}}{[\mathrm{ADP}]_\mathrm{eq}} \tag{4.11}$$

It can be showed that that the same relationship holds when the substrate is considered. Therefore, it can be assumed that the ratio between the rate of hydrolysis and synthesis is not affected by the presence of the substrate

$$\frac{k_\mathrm{h}}{k_\mathrm{s}} = \frac{k_\mathrm{h}^\mathrm{s}}{k_\mathrm{s}^\mathrm{s}}. \tag{4.12}$$

It is worth noting that when the ratio [ATP]/[ADP] is the equilibrium one ($[\mathrm{ATP}]_\mathrm{eq}/[\mathrm{ADP}]_\mathrm{eq} \sim 10^{-9}$ [80]), then Equation (4.10) is written as

$$\frac{k_\mathrm{h}}{k_\mathrm{s}} = \frac{k_{\mathrm{ex,TD}}^{\mathrm{eff}}}{k_{\mathrm{ex,DT}}^{\mathrm{eff}}}. \tag{4.13}$$

This, in fact, is the detailed balance condition in the cycle defined by the ATP- and ADP-bound states connected through two possible reactions: hydrolysis/synthesis and nucleotide exchange. So, when it holds, the two pathways lead to the same equilibrium distribution. Moreover, the ratio [ATP]/[ADP] is closely related to the energy released by ATP hydrolysis, see Equation (3.18). Therefore, the ratio [ATP]/[ADP] is a natural parameter to determine how far from equilibrium the system is.

Throughout the testing of our models, for the non-equilibrium scenario we considered a ratio of [ATP]/[ADP] = 10, which approximately reproduces the conditions at the cell [62].

- **Binding/unbinding**

The rates for binding/unbinding, hydrolysis/synthesis and nucleotide exchange are collectively constrained by thermodynamic relations since in equilibrium all cycles of the network must satisfy the detailed balance condition (Equation (3.10)). In fact, the ATP- and ADP-bound states of Hsp70 form a cycle with their respective conformations in the presence of the substrate via the hydrolysis of the bound nucleotides to the chaperone. Therefore, the following relation should hold

$$\frac{k_{\text{on}}^{\text{ATP}} k_{\text{h}}^{\text{s}} k_{\text{off}}^{\text{ADP}} k_{\text{s}}}{k_{\text{on}}^{\text{ADP}} k_{\text{s}}^{\text{s}} k_{\text{off}}^{\text{ATP}} k_{\text{h}}} = \frac{k_{\text{on}}^{\text{ATP}} k_{\text{off}}^{\text{ADP}}}{k_{\text{on}}^{\text{ADP}} k_{\text{off}}^{\text{ATP}}} = 1, \tag{4.14}$$

where Equation (4.12) has been introduced in the first equality. We can arrive at the same expression by considering only the nucleotide exchange process.

However, this conditions is not satisfied using the chaperone binding/unbinding rates in Table 4.1. So, we adopted the correction proposed in [61] and modified the binding and unbinding rates in the following way

$$r \equiv \frac{k_{\text{on}}^{\text{ATP}} k_{\text{off}}^{\text{ADP}}}{k_{\text{on}}^{\text{ADP}} k_{\text{off}}^{\text{ATP}}} \tag{4.15}$$

$$\begin{aligned} k_{\text{on}}^{\text{ATP}}, k_{\text{off}}^{\text{ADP}} &\to r^{1/4} k_{\text{on}}^{\text{ATP}}, r^{1/4} k_{\text{off}}^{\text{ADP}} \\ k_{\text{on}}^{\text{ADP}}, k_{\text{off}}^{\text{ATP}} &\to \frac{1}{r^{1/4}} k_{\text{on}}^{\text{ADP}}, \frac{1}{r^{1/4}} k_{\text{off}}^{\text{ATP}}. \end{aligned} \tag{4.16}$$

Furthermore, we have included two more modifications to the chaperone binding rates.

First, the concentrations of Hsp70 in the ATP- and ADP-bound states have been multiplied to the respective intrinsic binding rates. In particular, these concentrations are determined once and remain fixed during the temporal evolution of the system. This relies on the assumption that the concentration of chaperones is constant given that it is large enough with respect to the substrate concentration.

The concentrations in both Hsp70 conformational states can be obtained by solving the rate equations of the following three-state system

$$\mathrm{Hsp70} + \mathrm{ATP} \underset{k_-^{\mathrm{ATP}}}{\overset{k_+^{\mathrm{ATP}}}{\rightleftharpoons}} \mathrm{Hsp70 \cdot ATP} \underset{k_{\mathrm{s}}}{\overset{k_{\mathrm{h}}}{\rightleftharpoons}} \mathrm{Hsp70 \cdot ADP} \underset{k_+^{\mathrm{ADP}}}{\overset{k_-^{\mathrm{ADP}}}{\rightleftharpoons}} \mathrm{Hsp70} + \mathrm{ADP}\,. \quad (4.17)$$

Since we are under the assumption of excess of Hsp70 in the system, it can be considered that [Hsp70 · ATP] and [Hsp70 · ADP] are also constant in time. Imposing the steady state condition (Equation (3.7)), one can arrive to

$$\frac{[\mathrm{Hsp70 \cdot ATP}]}{[\mathrm{Hsp70 \cdot ADP}]} = \frac{k_{\mathrm{ex,DT}}^{\mathrm{eff}} + k_{\mathrm{s}}}{k_{\mathrm{ex,TD}}^{\mathrm{eff}} + k_{\mathrm{h}}} \quad (4.18)$$

Thus, considering that the total concentration of Hsp70 is given by

$$[\mathrm{Hsp70}] = [\mathrm{Hsp70 \cdot ATP}] + [\mathrm{Hsp70 \cdot ADP}], \quad (4.19)$$

the concentrations of Hsp70 in the ATP- and ADP-bound states can be determined.

Second, the effect of substrate expansion upon chaperone binding has been modulated in the following way

$$\hat{k}_{\mathrm{on},ij} = k_{\mathrm{on},ij} e^{-\beta \Delta G_{ij}}, \quad (4.20)$$

where ΔG_{ij} is the conformational free-energy difference[1] between states i and j. That is, it represents the energetic cost to go from the structure at state i to the one at state j.

This conformational free-energy cost of chaperone binding and the other relevant in our models, namely substrate unfolding with and without the presence of a chaperone, were determined using the results from Molecular Dynamics (MD) simulations that studied the action of DnaK on the substrate (rhodanese), which was modeled as a polypeptide with six binding sites that were identified through bioinformatic algorithms. In brief, from steering MD trajectories, the distribution for the work needed to go from the initial equilibrium conformation, $R_{g,\mathrm{eq}}$, to an elongated structure with a certain gyration radius, R_g, was obtained. Furthermore, the Jarzynski equality [81] was used to calculate the conformational free-energy difference between these two states, $\delta G(R_g) = G(R_g) - G(R_{g,\mathrm{eq}})$, for the free substrate and several chaperone-substrate complexes. For the models in this work, we took the values for both the free substrate and the substrate with one Hsp70 bound (see Figure 4.3), as we only considered the case of single chaperone binding.

[1] Reminding that, in physics, the Gibbs free-energy difference is the energy available to perform non-expansion work in a closed system at constant temperature and pressure. In this project, we treat the Gibbs and conformational free-energy differences interchangeably as the effects on the overall volume of the system due to the protein conformational changes are negligible.

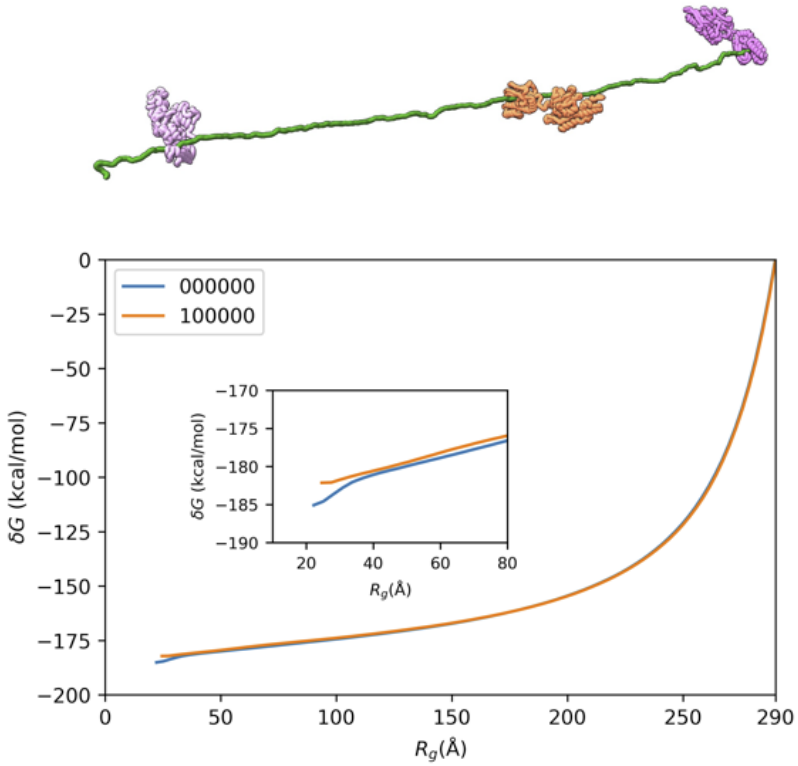


Figure 4.3: Free-energy difference δG between equilibrium conformations and the fully-stretched state (represented at the top for a given chaperone-substrate complex, figure adapted from [61]) for the free substrate (blue line) and the substrate with one Hsp70 bound (orange line) as a function of the corresponding radius of gyration R_g. An amplification of the low values of R_g is depicted in the inset. Data taken from [61] (under the courtesy of S. Assenza).

Since the initial equilibrium radius was defined as the one in which the conformation remains unperturbed, $\delta G(R_g^{\text{eq}}) = 0$, we assumed to it to be the radius of the misfolded protein, $R_{g,\text{m}} \equiv R_{g,\text{eq}} \sim 20$ Å.

In order to calculate the conformational free-energy cost of Hsp70 binding, we took as reference the fully-stretched conformation, $R_{g,\text{fs}} \sim 290$ Å, as the effect of chaperone binding on this conformation is negligible due to the large intermolecular distances. As a result,

$$\Delta G_{ij} = \delta G(R_{g,j}) - \delta G(R_{g,i}), \tag{4.21}$$

with $\delta G(R_g)$ redefined as $\delta G(R_g) = G(R_g) - G(R_{g,\text{fs}})$. And, then the was calculated as

$$\Delta G_{\text{M,M}\,70} = \delta G_{70}(R_{g,\text{m}}) - \delta G(R_{g,\text{m}}), \tag{4.22}$$

where the subscript in δG indicates the case of the substrate with one Hsp70 bound.

For the determination of the conformational free-energy costs of unfolding the misfolded protein with and without a chaperone bound, the following expressions were used

$$\Delta G_{\mathrm{M,U}} = \delta G(R_{g,\mathrm{u}}) - \delta G(R_{g,\mathrm{m}}), \tag{4.23}$$

$$\Delta G_{\mathrm{M}\,70,\mathrm{U}\,70} = \delta G_{70}(R_{g,\mathrm{u}}) - \delta G_{70}(R_{g,\mathrm{m}}). \tag{4.24}$$

In these, $R_{g,\mathrm{u}} \sim 40$ Å is the radius of the unfolded substrate. This value was chosen so that the values obtained were reasonable and the following constraint was satisfied

$$\Delta G_{\mathrm{M,U}} = \Delta G_{\mathrm{M,M}\,70} + \Delta G_{\mathrm{M}\,70,U70} + \Delta G_{\mathrm{U}\,70,\mathrm{U}}. \tag{4.25}$$

The values obtained are shown in Table 4.2. This condition comes from thermodynamics since the free-energy is a state function, namely it depends on the start and end states but not on the path taken.

Furthermore, it was assumed that from an energetic point of view there was no cost to bind a chaperone to the substrate once it was unfolded, namely $\Delta G_{\mathrm{U,U}\,70} = 0$. Therefore, $\Delta G_{\mathrm{U}\,70,\mathrm{U}} = 0$.

$\Delta G_{\mathrm{M,U}}$	$\Delta G_{\mathrm{M,M}\,70}$	$\Delta G_{\mathrm{M}\,70,\mathrm{U}\,70}$	$\Delta G_{\mathrm{U}\,70,\mathrm{U}}$
4 kcal/mol	2.5 kcal/mol	1.5 kcal/mol	0 kcal/mol

Table 4.2: Conformational free-energy cost of unfolding the substrate from a misfolded conformation ($\Delta G_{\mathrm{M,U}}$), binding a chaperone to the misfolded substrate ($\Delta G_{\mathrm{M,M}\,70}$), unfolding a misfolded substrate with a chaperone bound to it ($\Delta G_{\mathrm{M}\,70,\mathrm{U}\,70}$), and binding a chaperone to the unfolded substrate ($\Delta G_{\mathrm{U,U}\,70}$).

It should be mentioned that to guarantee the stability and precision of the values shown in Table 4.2 further means over different chaperone-substrate conformations, in which the chaperone is fixed at other binding sites, should be performed. However, within the scope of this project, the order of magnitude of these quantities were enough given the level of coarse graining in the models. Moreover, they were also partially determined to avoid falling into a hyperparameter space in the models.

4.1.2 Hsp70s in protein folding

Based on the canonical cycle of Hsp70s depicted in Figure 2.5, we constructed a first model that included the main conformational states present in the cycle. The action of JDPs was considered but not the one of the NEFs, as it is known that some Hsp70s may function without their assistance. Their intervention was included in the later and, for now, final model.

It should be noted that in both models we did not consider the process of aggregation on the basis that substrate concentration was small, generally orders of magnitude lower with respect to the concentration of Hsp70.

Initial model

For the initial description of the role of Hsp70s in protein folding, we considered a kinetic model in which the substrate protein could be in five states: mifolded (M), misfolded bound to the ATP-bound state of Hsp70 (MT), misfolded bound to the ADP-bound state of Hsp70 (MD), unfolded (U), and in its native conformation (N), see Figure 4.4. As before, these are coarse grained representations of all the possible structural conformations of these states at the molecular level.

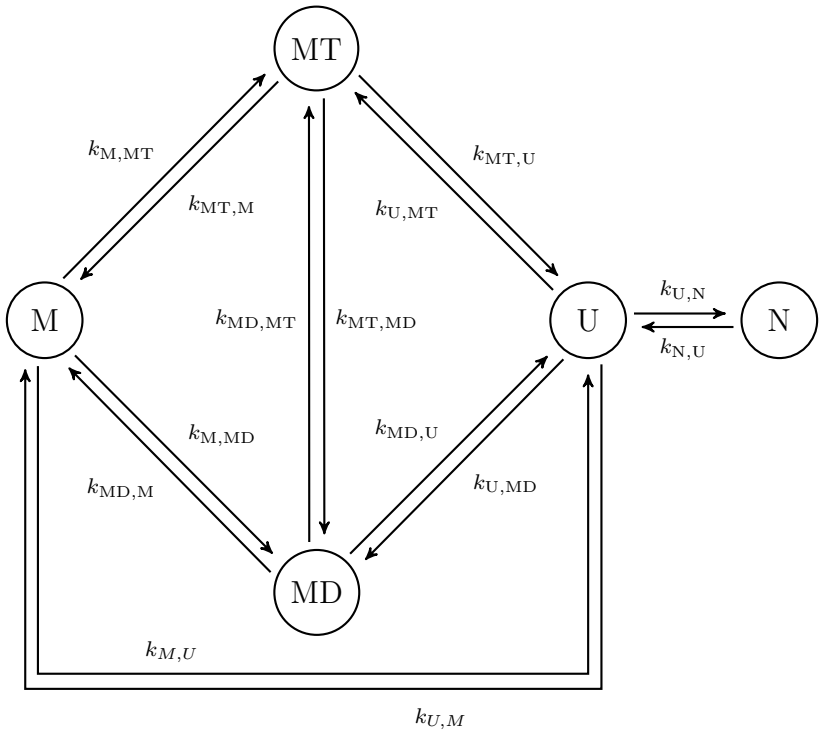


Figure 4.4: Initial kinetic model for Hsp70s in protein folding. Network representation of the biochemical system with five states for the client protein: misfolded (M), misfolded bound to the ATP-bound state of Hsp70 (MT), misfolded bound to the ADP-bound state of Hsp70 (MD), unfolded (U), and native conformation (N). Refer to the main text for more details about the indicated transition rates.

It can be noticed that this model is basically an amplification of the previous three-state system, which here is the left triangle formed by the nodes M,MT and MD in Figure 4.4. Therefore, the information provided about the transition rates between these states can be applied here.

The concentrations of this system are given by the following rate equations

$$\begin{aligned}
\frac{d[\mathrm{M}]}{dt} &= -\left(k_{\mathrm{M,MT}} + k_{\mathrm{M,MD}} + k_{\mathrm{M,U}}\right)[\mathrm{M}] \\
&\quad + k_{\mathrm{MT,M}}[\mathrm{MT}] + k_{\mathrm{MD,M}}[\mathrm{MD}] + k_{\mathrm{U,M}}[\mathrm{U}] \\
\frac{d[\mathrm{MT}]}{dt} &= -\left(k_{\mathrm{MT,M}} + k_{\mathrm{MT,MD}} + k_{\mathrm{MT,U}}\right)[\mathrm{MT}] \\
&\quad + k_{\mathrm{M,MT}}[\mathrm{M}] + k_{\mathrm{MD,MT}}[\mathrm{MD}] + k_{\mathrm{U,MT}}[\mathrm{U}] \\
\frac{d[\mathrm{MD}]}{dt} &= -\left(k_{\mathrm{MD,M}} + k_{\mathrm{MD,MT}} + k_{\mathrm{MD,U}}\right)[\mathrm{MD}] \\
&\quad + k_{\mathrm{M,MD}}[\mathrm{M}] + k_{\mathrm{MT,MD}}[\mathrm{MT}] + k_{\mathrm{U,MD}}[\mathrm{U}] \\
\frac{d[\mathrm{U}]}{dt} &= -\left(k_{\mathrm{U,MT}} + k_{\mathrm{U,MD}} + k_{\mathrm{U,M}} + k_{\mathrm{U,N}}\right)[\mathrm{U}] \\
&\quad + k_{\mathrm{MT,U}}[\mathrm{MT}] + k_{\mathrm{MD,U}}[\mathrm{MD}] + k_{\mathrm{M,U}}[\mathrm{M}] + k_{\mathrm{N,U}}[\mathrm{N}] \\
\frac{d[\mathrm{N}]}{dt} &= -\,k_{\mathrm{N,U}}[\mathrm{N}] + k_{\mathrm{U,N}}[\mathrm{U}]
\end{aligned} \quad , \tag{4.26}$$

where the reaction rates (in s^{-1}) are defined as

$$\begin{aligned}
k_{\mathrm{M,MT}} &= k_{\mathrm{on}}^{\mathrm{ATP}}[\mathrm{Hsp70}\cdot\mathrm{ATP}]e^{-\beta\Delta G_{\mathrm{M,M\,70}}} & k_{\mathrm{MT,M}} &= k_{\mathrm{off}}^{\mathrm{ATP}} \\
k_{\mathrm{M,MD}} &= k_{\mathrm{on}}^{\mathrm{ADP}}[\mathrm{Hsp70}\cdot\mathrm{ADP}]e^{-\beta\Delta G_{\mathrm{M,M\,70}}} & k_{\mathrm{MD,D}} &= k_{\mathrm{off}}^{\mathrm{ADP}} \\
k_{\mathrm{MT,MD}} &= k_{\mathrm{ex,TD}}^{\mathrm{eff}} + k_{\mathrm{h}}^{\mathrm{s}} & k_{\mathrm{MD,MT}} &= k_{\mathrm{ex,DT}}^{\mathrm{eff}} + k_{\mathrm{s}}^{\mathrm{s}} \\
k_{\mathrm{MT,U}} &= k_{\mathrm{off}}^{\mathrm{ATP}}e^{-\beta\Delta G_{\mathrm{M\,70,U}}} & k_{\mathrm{U,MT}} &= k_{\mathrm{on}}^{\mathrm{ATP}}[\mathrm{Hsp70}\cdot\mathrm{ATP}]\,. \\
k_{\mathrm{MD,U}} &= k_{\mathrm{off}}^{\mathrm{ADP}}e^{-\beta\Delta G_{\mathrm{M\,70,U}}} & k_{\mathrm{U,MD}} &= k_{\mathrm{on}}^{\mathrm{ADP}}[\mathrm{Hsp70}\cdot\mathrm{ADP}] \\
k_{\mathrm{U,M}} &= \alpha & k_{\mathrm{M,U}} &= k_{\mathrm{U,M}}e^{-\beta\Delta G_{\mathrm{M,U}}} \\
k_{\mathrm{U,N}} &= k_{\mathrm{U,M}} & k_{\mathrm{N,U}} &= \gamma\cdot k_{\mathrm{M,U}}
\end{aligned} \tag{4.27}$$

For the transition rates to the added states U (unfolded) and N (native), the following was considered.

First, it was assumed that the intrinsic rates for Hsp70 chaperone binding/unbinding to/from the substrate were the same, regardless of the latter being in an unfolded conformation. And, as explained before, the binding rates needed to be multiplied by the concentration of Hsp70 in the ATP- and ADD bound states. Moreover, at the beginning it was also supposed that there should not be any energy penalty in the transition from the misfolded conformations bound to a chaperone (MT,MD) to the unfolded state (U) since it is known that the binding of Hsp70s facilitates the unfolding of the substrate [61]. Otherwise said, it was considered that $\Delta G_{\mathrm{M\,70,U}}$ was zero and hence $\Delta G_{\mathrm{M,M\,70}} = \Delta G_{\mathrm{M,U}}$. The implications of this choice will be shown in the next chapter.

Second, the rates between the unfolded and misfolded states ($k_{\mathrm{U,M}}, k_{\mathrm{M,U}}$) and the unfolded and native states ($k_{\mathrm{U,N}}, k_{\mathrm{N,U}}$) were chosen so that the unfolded state was less stable than the misfolded and native states, i.e., $k_{UM}/k_{MU} > 1$ and $k_{UN}/k_{NU} > 1$

as in equilibrium the free-energy difference[2] of two states is $\Delta G^{\circ}_{ij} = k_B T \ln(k_{ij}/k_{ji})$ (detailed balance condition, Equation (3.14)). At the same time, it was also considered that the misfolded state needed to be more stable than the native state, namely $k_{\mathrm{U,M}}/k_{\mathrm{M,U}} > k_{\mathrm{U,N}}/k_{\mathrm{N,U}}$ since then

$$\Delta G^{\circ}_{\mathrm{M,N}} = G^{\circ}_{\mathrm{N}} - G^{\circ}_{\mathrm{M}} = -k_B T \left[\ln\left(\frac{k_{\mathrm{U,N}}}{k_{\mathrm{N,U}}}\right) - \ln\left(\frac{k_{\mathrm{U,M}}}{k_{\mathrm{M,U}}}\right) \right] > 0. \tag{4.28}$$

This condition represents the denaturation of the protein.

These constraints were combined with the detailed balance condition to have two free parameters (α and γ) instead of four.

Final model

In order to be able to add the effect of NEFs in the functional cycle of Hsp70s, we added to the initial model two more possible states for the substrate: unfolded bound to Hsp70 in the ATP-bound state (UT) and unfolded bound to Hsp70 in the ADP-bound state (UD), see Figure 4.5.

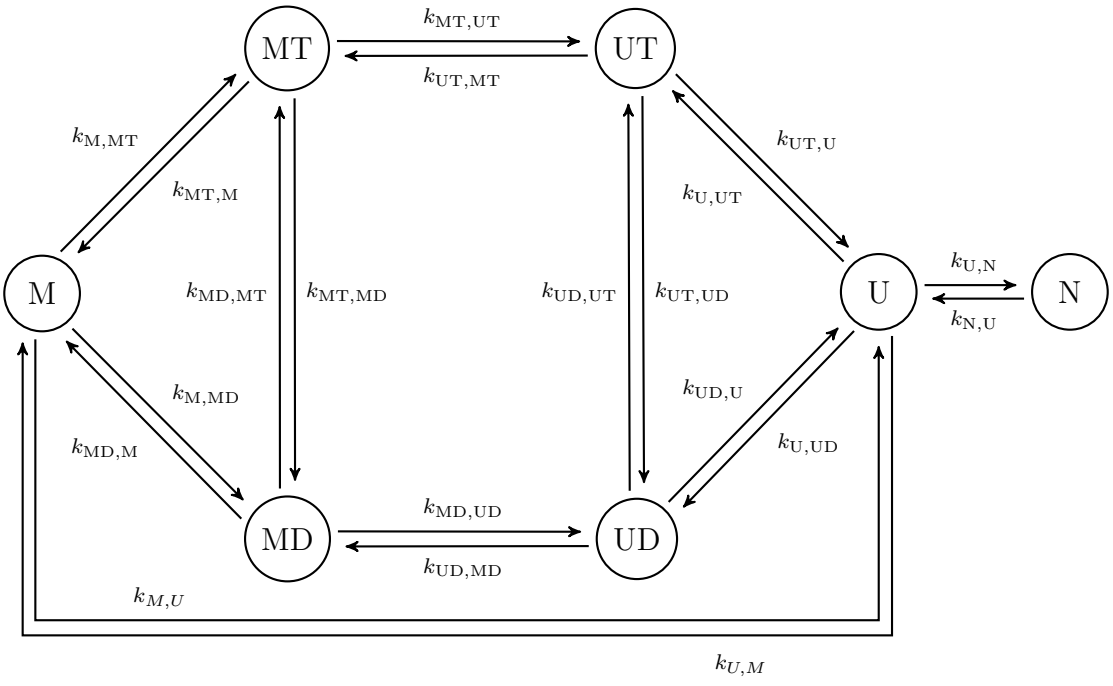


Figure 4.5: Final kinetic model for Hsp70s in protein folding. Network representation of the biochemical system with seven states for the client protein: misfolded (M), misfolded bound to the ATP-bound state of Hsp70 (MT), misfolded bound to the ADP-bound state of Hsp70 (MD), unfolded bound the ADP-bound state of Hsp70 (UT), unfolded the ADP-bound state of Hsp70 (UD), unfolded (U), and native conformation (N). Refer to the main text for more details about the indicated transition rates.

[2]Notice that here the Gibbs and conformational free-energies are interchangeably used.

The concentrations of the system are given by the following rate equations

$$\begin{aligned}
\frac{d[\mathrm{M}]}{dt} &= -(k_{\mathrm{M,MT}} + k_{\mathrm{M,MD}} + k_{\mathrm{M,U}})[\mathrm{M}] \\
&\quad + k_{\mathrm{MT,M}}[\mathrm{MT}] + k_{\mathrm{MD,M}}[\mathrm{MD}] + k_{\mathrm{U,M}}[\mathrm{U}] \\
\frac{d[\mathrm{MT}]}{dt} &= -(k_{\mathrm{MT,M}} + k_{\mathrm{MT,MD}} + k_{\mathrm{MT,UT}})[\mathrm{MT}] \\
&\quad + k_{\mathrm{M,MT}}[\mathrm{M}] + k_{\mathrm{MD,MT}}[\mathrm{MD}] + k_{\mathrm{UT,MT}}[\mathrm{UT}] \\
\frac{d[\mathrm{MD}]}{dt} &= -(k_{\mathrm{MD,M}} + k_{\mathrm{MD,MT}} + k_{\mathrm{MD,UD}})[\mathrm{MD}] \\
&\quad + k_{\mathrm{M,MD}}[\mathrm{M}] + k_{\mathrm{MT,MD}}[\mathrm{MT}] + k_{\mathrm{UD,MD}}[\mathrm{UD}] \\
\frac{d[\mathrm{UT}]}{dt} &= -(k_{\mathrm{UT,U}} + k_{\mathrm{UT,UD}} + k_{\mathrm{UT,MT}})[\mathrm{UT}] \\
&\quad + k_{\mathrm{U,UT}}[\mathrm{U}] + k_{\mathrm{UD,UT}}[\mathrm{UD}] + k_{\mathrm{MT,UT}}[\mathrm{MT}] \\
\frac{d[\mathrm{UD}]}{dt} &= -(k_{\mathrm{UD,U}} + k_{\mathrm{UD,UT}} + k_{\mathrm{UD,MD}})[\mathrm{UD}] \\
&\quad + k_{\mathrm{U,UD}}[\mathrm{U}] + k_{\mathrm{UT,UD}}[\mathrm{UT}] + k_{\mathrm{MD,UD}}[\mathrm{MD}] \\
\frac{d[\mathrm{U}]}{dt} &= -(k_{\mathrm{U,UT}} + k_{\mathrm{U,UD}} + k_{\mathrm{U,M}} + k_{\mathrm{U,N}})[\mathrm{U}] \\
&\quad + k_{\mathrm{UT,U}}[\mathrm{UT}] + k_{\mathrm{UD,U}}[\mathrm{UD}] + k_{\mathrm{M,U}}[\mathrm{M}] + k_{\mathrm{N,U}}[\mathrm{N}] \\
\frac{d[\mathrm{N}]}{dt} &= -k_{\mathrm{N,U}}[\mathrm{N}] + k_{\mathrm{U,N}}[\mathrm{U}]
\end{aligned} \quad , \tag{4.29}$$

where the reaction rates (in s^{-1}) are defined as

$$\begin{aligned}
k_{\mathrm{M,MT}} &= k_{\mathrm{on}}^{\mathrm{ATP}}[\mathrm{Hsp70}\cdot\mathrm{ATP}]e^{-\beta\Delta G_{\mathrm{M,M\,70}}} & k_{\mathrm{MT,M}} &= k_{\mathrm{off}}^{\mathrm{ATP}} \\
k_{\mathrm{M,MD}} &= k_{\mathrm{on}}^{\mathrm{ADP}}[\mathrm{Hsp70}\cdot\mathrm{ADP}]e^{-\beta\Delta G_{\mathrm{M,M\,70}}} & k_{\mathrm{MD,M}} &= k_{\mathrm{off}}^{\mathrm{ADP}} \\
k_{\mathrm{MT,MD}} &= k_{\mathrm{ex,TD}}^{\mathrm{eff}} + k_{\mathrm{h}}^{\mathrm{s}} & k_{\mathrm{MD,MT}} &= k_{\mathrm{ex,DT}}^{\mathrm{eff}} + k_{\mathrm{s}}^{\mathrm{s}} \\
k_{\mathrm{MT,UT}} &= k_{\mathrm{M,U}}^{\mathrm{ATP}} e^{-\beta\Delta G_{\mathrm{M\,70,U\,70}}} & k_{\mathrm{UT,MT}} &= k_{\mathrm{MT,UT}} e^{+\beta\Delta G_{\mathrm{M\,70,U70}}} \\
k_{\mathrm{MD,UD}} &= k_{\mathrm{M,U}}^{\mathrm{ADP}} e^{-\beta\Delta G_{\mathrm{M\,70,U\,70}}} & k_{\mathrm{UD,MD}} &= k_{\mathrm{MD,UD}} e^{+\beta\Delta G_{\mathrm{M\,70,U70}}} \\
k_{\mathrm{UT,U}} &= k_{\mathrm{off}}^{\mathrm{ATP}} e^{-\beta\Delta G_{\mathrm{U\,70,U}}} & k_{\mathrm{U,UT}} &= k_{\mathrm{on}}^{\mathrm{ATP}}[\mathrm{Hsp70}\cdot\mathrm{ATP}] \\
k_{\mathrm{UD,U}} &= k_{\mathrm{off}}^{\mathrm{ADP}} e^{-\beta\Delta G_{\mathrm{U\,70,U}}} & k_{\mathrm{U,UD}} &= k_{\mathrm{on}}^{\mathrm{ADP}}[\mathrm{Hsp70}\cdot\mathrm{ADP}] \\
k_{\mathrm{UT,UD}} &= k_{\mathrm{ex,TD}}^{\mathrm{eff}} + k_{\mathrm{h}}^{\mathrm{s}} & k_{\mathrm{UD,UT}} &= k_{\mathrm{ex,DT}}^{\mathrm{eff}} + k_{\mathrm{s}}^{\mathrm{s}} \\
k_{\mathrm{U,M}} &= \alpha & k_{\mathrm{M,U}} &= k_{\mathrm{U,M}} e^{-\beta\Delta G_{\mathrm{M,U}}} \\
k_{\mathrm{U,N}} &= k_{\mathrm{U,M}} & k_{\mathrm{N,U}} &= \gamma \cdot k_{\mathrm{M,U}}
\end{aligned} \quad . \tag{4.30}$$

It can be noticed that we conserved all the definitions from the previous model, and just shifted the rates defined for $k_{\mathrm{MT,U}}, k_{\mathrm{U,MT}}, k_{\mathrm{UT,U}}, k_{\mathrm{U,UT}}$ to the ones for $k_{\mathrm{MD,U}}, k_{\mathrm{U,MD}}, k_{\mathrm{UD,U}}, k_{\mathrm{U,UD}}$, respectively.

For the transition rates between the misfolded conformations bound to Hsp70 (MT, MD) and the added unfolded states bound to Hsp70 (UT, UD), we chose to not make any distinction between the two pathways. That is, the unfolding was considered to

be independent of having the chaperone in the ATP- or ADP-bound state. Furthermore, we defined $k^{\mathrm{ATP}}_{\mathrm{M,U}} = k^{\mathrm{ADP}}_{\mathrm{M,U}} = \frac{1}{\tau} e^{+\beta \Delta G_{\mathrm{M}\,70,U70}}$ so that $k_{\mathrm{MT,UT}} = k_{\mathrm{MD,UD}} = 1/\tau$, with τ being approximately the transition time from state the chaperone bound misfolded state (MT or MD) to the chaperone bound unfolded state (UT or UD). This was also a free parameter of the network. While for the rates in the opposite direction, $k_{\mathrm{UT,MT}}$ and $k_{\mathrm{UD,MD}}$, we imposed the detailed balance condition (Equation (3.14)).

5

Results

In this chapter, we present the results obtained for each model and include a discussion about them.

Since our aim is to understand the action of Hsp70s in guiding non-native proteins to their native conformation, in the simulations we always begin with all the substrates in their misfolded conformation. Hence, for the integration of the rate equations of the kinetic models, only the misfolded state (M) has an initial non-zero concentration. In general, unless otherwise specified, we have that $[\mathrm{M}]_{t=0} = 10$ nM as we are under the assumption of low concentrations of misfolded substrates to avoid the formation of aggregates. Furthermore, we consider that $[\mathrm{Hsp70}] = 10$ µM, which is close to the value at physiological conditions. And then, in our system, the chaperone is in excess over the substrate.

Here it should be mentioned that the rate equations are both integrated numerically to high accuracy using the function `odeint` from the SciPy library in Python and solved analytically through the Equation (4.3) to guarantee the reliability of the results. Moreover, the resulting concentrations are normalised by the total concentration of substrate and converted to percentages to facilitate the analysis.

5.1 Exploring ultra-affinity

It has been explained that Hsp70s exploit the energy from ATP hydrolysis to achieve a non-equilibrium affinity (ultra-affinity) for the substrate that goes beyond the bounds imposed by equilibrium thermodynamics.

A general measure of how well Hsp70s can bind to the substrates through their cycle is the effective dissociation constant $K_{\mathrm{d}}^{\mathrm{eff}}$, which is defined as

$$K_{\mathrm{d}}^{\mathrm{eff}} = \frac{[\mathrm{S}][\mathrm{Hsp70}]}{[\mathrm{Hsp70} \cdot \mathrm{S}]}, \tag{5.1}$$

where [S] is the concentration of free substrate, [Hsp70] is the total concentration of chaperone in the system ($[\mathrm{Hsp70}] = [\mathrm{Hsp70} \cdot \mathrm{ATP}] + [\mathrm{Hsp70} \cdot \mathrm{ADP}]$), and $[\mathrm{Hsp70} \cdot \mathrm{S}]$ is the concentration of chaperone-substrate complex ($[\mathrm{Hsp70} \cdot \mathrm{S}] = [\mathrm{Hsp70} \cdot \mathrm{ATP} \cdot \mathrm{S}] + [\mathrm{Hsp70} \cdot \mathrm{ADP} \cdot \mathrm{S}]$).

In order to observe the non-equilibrium effect on the chaperone's affinity for the substrate, we use the three-state model 4.4. We should recall that is formed by the misfolded substrate (M) that can bind to the Hsp70 ATP-bound state (MT) or to the Hsp70 ATP-bound state (MD). We start by comparing the time evolution of this system at equilibrium ($[ATP]_{eq}/[ADP]_{eq} = 10^{-9}$) and non-equilibrium conditions ($[ATP]/[ADP] = 10$), see Figure 5.1.

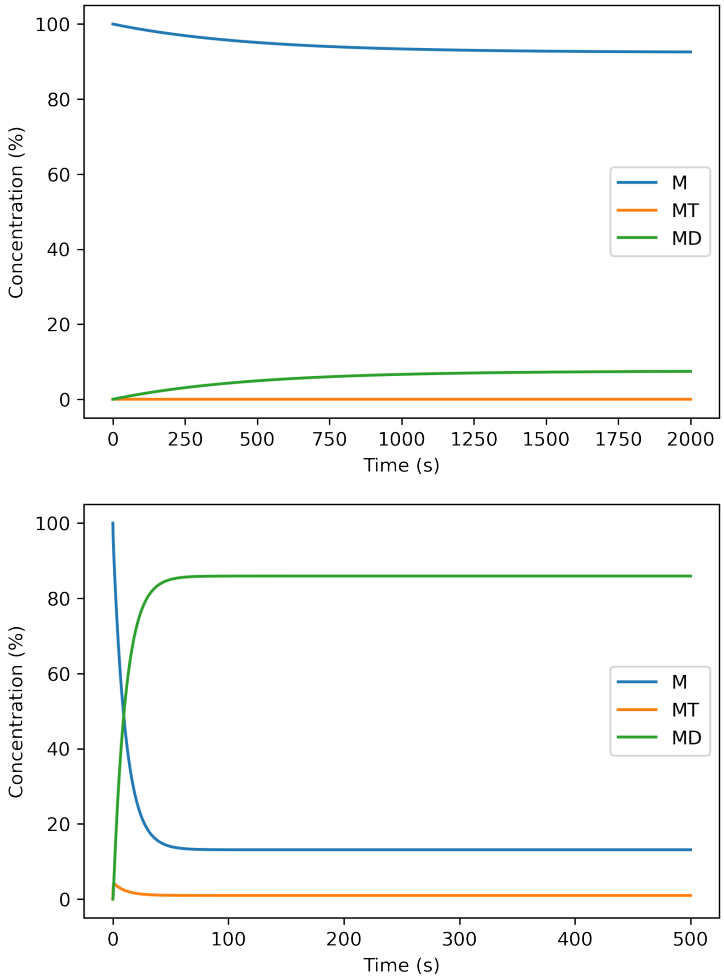


Figure 5.1: Ultra-affinity: equilibrium versus non-equilibrium conditions. Time evolution of the concentrations in the system 4.4 under equilibrium conditions (top panel) and non-equilibrium conditions (bottom panel).

From this simulations, we can observe that in equilibrium conditions chaperone binding to the substrate is very low; the concentration of mislfoded substrate bound to Hsp70 in the ADP state, [MD], is less than 20%. On the contrary, when the system is placed out of equilibrium chaperone binding is high; [MD] around 85 %. The substrate is locked in the ADP-bound state of Hsp70 since this state has a higher affinity than the ATP-bound state of Hsp70.

Considering the concentrations obtained at the steady state, we can calculate the

effective dissociation constant (Equation (5.1)) for this system as

$$K_{\mathrm{d}}^{\mathrm{eff}} = \frac{[\mathrm{M}][\mathrm{Hsp70}]}{[\mathrm{MT}] + [\mathrm{MD}]} \tag{5.2}$$

Thus, in equilibrium conditions the effective dissociation constant is $K_{\mathrm{d}}^{\mathrm{eff,eq}} = 124.48$ µM while in non-equilibrium conditions is $K_{\mathrm{d}}^{\mathrm{eff,neq}} = 1.51$ µM, which is much lower.

To directly relate this quantity with the dissociation constant of the ATP- and ADP-bound states: $K_{\mathrm{d}}^{\mathrm{ATP}} = k_{\mathrm{off}}^{\mathrm{ATP}}/k_{\mathrm{off}}^{\mathrm{ATP}} = 4.4$ µM, $K_{\mathrm{d}}^{\mathrm{ADP}} = k_{\mathrm{off}}^{\mathrm{ADP}}/k_{\mathrm{off}}^{\mathrm{ADP}} = 0.47$ µM (see Table 4.1), we can consider the scenario in which there is not a conformational free-energy in chaperone binding, $\Delta G_{\mathrm{M,M\,70}} = 0$. For instance, this would correspond to the case of a small substrate (peptide), which does not undergo through structural changes upon chaperone binding. From the simulations, the following values were obtained: $K_{\mathrm{d}}^{\mathrm{eff,eq}} = 1.90$ µM and $K_{\mathrm{d}}^{\mathrm{eff,neq}} = 0.023$ µM. Therefore, we can clearly see that $K_{\mathrm{d}}^{\mathrm{eff,neq}} < K_{\mathrm{d}}^{\mathrm{ATP}}$ and $K_{\mathrm{d}}^{\mathrm{eff,neq}} < K_{\mathrm{d}}^{\mathrm{ADP}}$, as expected.

It should also be noticed the difference in the timescales between the equilibrium and non-equilibrium scenarios. While the system is in equilibrium the dynamics are slow but once the system is driven away from it, the system is able to reach its steady-state configuration faster. This can be reasoned by considering that the concentrations of Hsp70 in the ATP and ADP states depend on the ratio [ATP]/[ADP] (Equation (4.18)), which in turn affect the chaperone binding rates as we multiply them to its intrinsic binding rates, see Equation (4.5).

Furthermore, in Figure 5.2, it is shown how the value of the ratio [ATP]/[ADP] affects the steady-state distribution of the system.

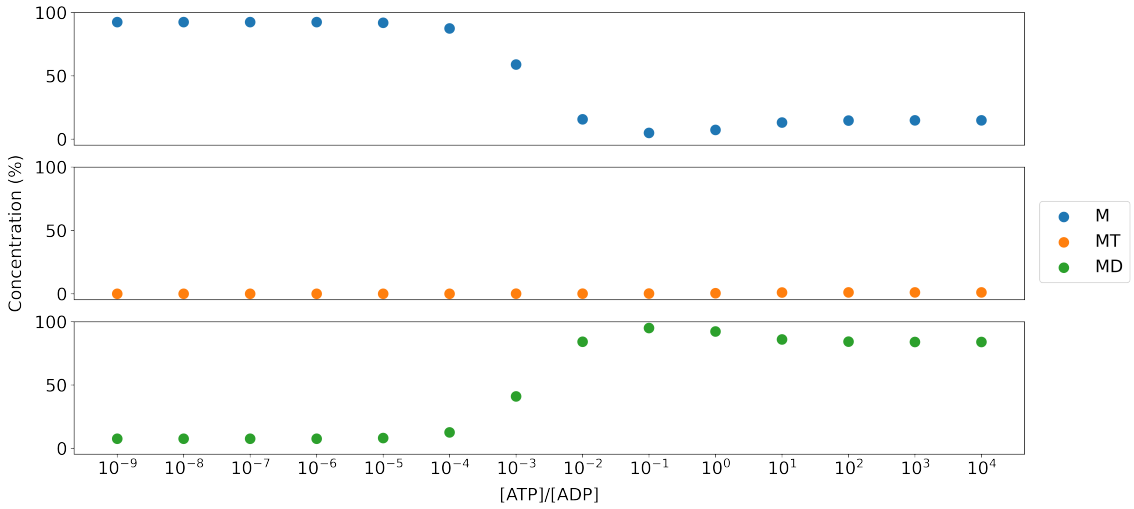


Figure 5.2: Effect of the ratio [ATP]/[ADP] on the dynamics. Steady-state concentrations of the system 4.4 as a function of the ratio [ATP]/[ADP].

It can be observed that, as expected, as the ratio [ATP]/[ADP] goes from low to high values, so it does the concentration of substrate bound to Hsp70 in the ADP-bound state. However, we can see that there is a small decrease at high values of

ATP. This is assumed to be explained by the low affinity of the Hsp70 ATP-bound state for the substrate, which slightly disfavours the binding. However, a complete explanation for this phenomena at the molecular level it is not yet determined.

However, it should be remarked that this (ultra) enhancement of the chaperone affinity for the substrate is not only a result of the excess of ATP over ADP or the difference between the affinities of the ATP and ADP states of Hsp70, but is a non-equilibrium effect driven by the energy produced from hydrolysing ATP. Indeed, in Figure 5.3, we can observe that if we neglect the chaperone ATPase activity ($k_{\mathrm{h}}^{\mathrm{s}} = k_{\mathrm{s}}^{\mathrm{s}} = 0$) while keeping the other parameters unchanged, then ultra-affinity can not be achieved.

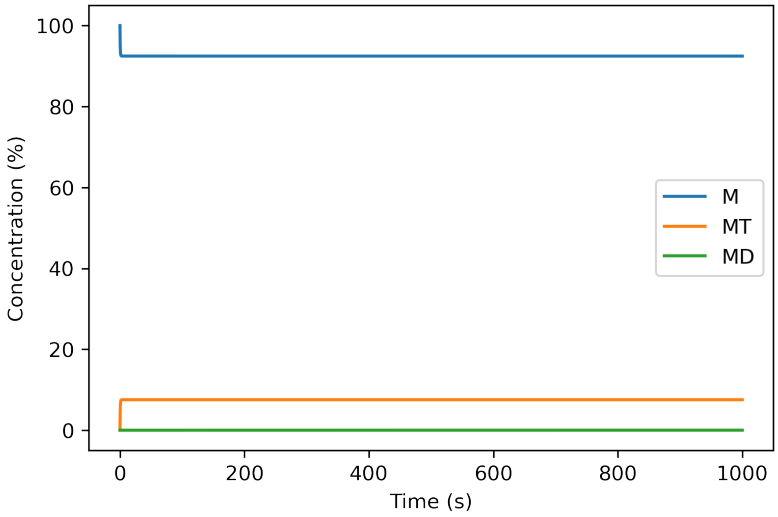


Figure 5.3: Influence of the ATPase activity in ultra-affinity. Time evolution of the concentrations in the system 4.4 when the Hsp70 ATPase activity is neglected, namely $k_{\mathrm{h}}^{\mathrm{s}} = k_{\mathrm{s}}^{\mathrm{s}} = 0$.

Hence, the results from the simulations are in agreement with the fact that it is the energy of ATP hydrolysis what permits the population of certain states that otherwise would not be accessible.

5.2 Protein folding

In this section, we only show the results of the models for the non-equilibrium case as it is the relevant one to study the action of Hsp70s in protein folding. However, in the background, we still consider the equilibrium case to check that the detailed balance conditions are satisfied.

For the mentioned free parameters in the rates of the models, at first, we consider $\alpha = 1$ and $\gamma = 5$ in Equations (4.27) and (4.30). This results in $\Delta G_{\mathrm{M,N}} \simeq 1$ kcal/mol. It should be mentioned that this choice is arbitrary as it is not based in any experimental observation.

5.2.1 Initial model

From the simulations of the initial model 4.26, we obtain that when there is no free-energy cost in the unfolding of the misfolded substrate bound to a chaperone ($\Delta G_{\mathrm{M}\,70,\mathrm{U}} = 0$ in Equation (4.27)), the binding is penalised compared to the case in which we distribute the conformational free-energy cost of unfolding the misfolded substrate between the steps of chaperone binding and the unfolding (see Figure 5.4). However, in the second scenario, we were unable to promote the misfolded substrate bound to Hsp70 in the ADP state toward its native state.

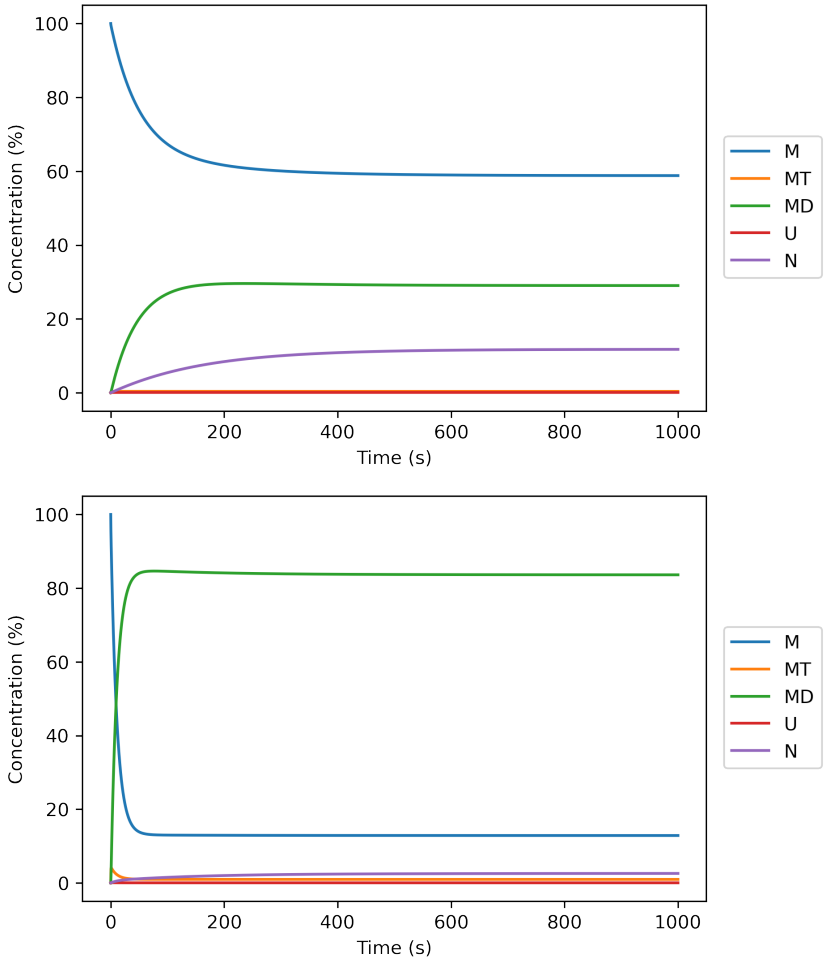


Figure 5.4: Effect of the distribution of the conformational free-energies. Time evolution of the concentrations of the system 4.26 when there is no free-energy cost to unfold the misfolded substrate bound to the chaperone (top panel) and when there is (bottom panel).

It should also be mentioned that while the concentration of substrate in its native state appears to be higher in the first scenario, this is mainly because of the spontaneous (without the intervention of the chaperone) unfolding of the protein, followed by the transition to the native state. Indeed, if we increase the free-energy difference between the unfolded and native states, then the concentration of native states decreases while the concentration of bounded substrates remains almost unchanged.

5.2.2 Final model

For the final model 4.29, we consider $\tau = 60$ s. That is, the time to transition from the misfolded bound to a chaperone state to the unfolded bound to a chaperone state is around one minute in our model. Here, as before, this choice does not rely on any observation of this process.

From the simulations of this system, we observe that if we keep the parameters from the initial model in the scenario with an energetic penalty in the transition from the misfolded to the unfolded bound to Hsp70 states, the results obtained are similar. The only change is an almost imperceptible concentration of unfolded substrate bound to Hsp70 in the ADP state, see Figure 5.5.

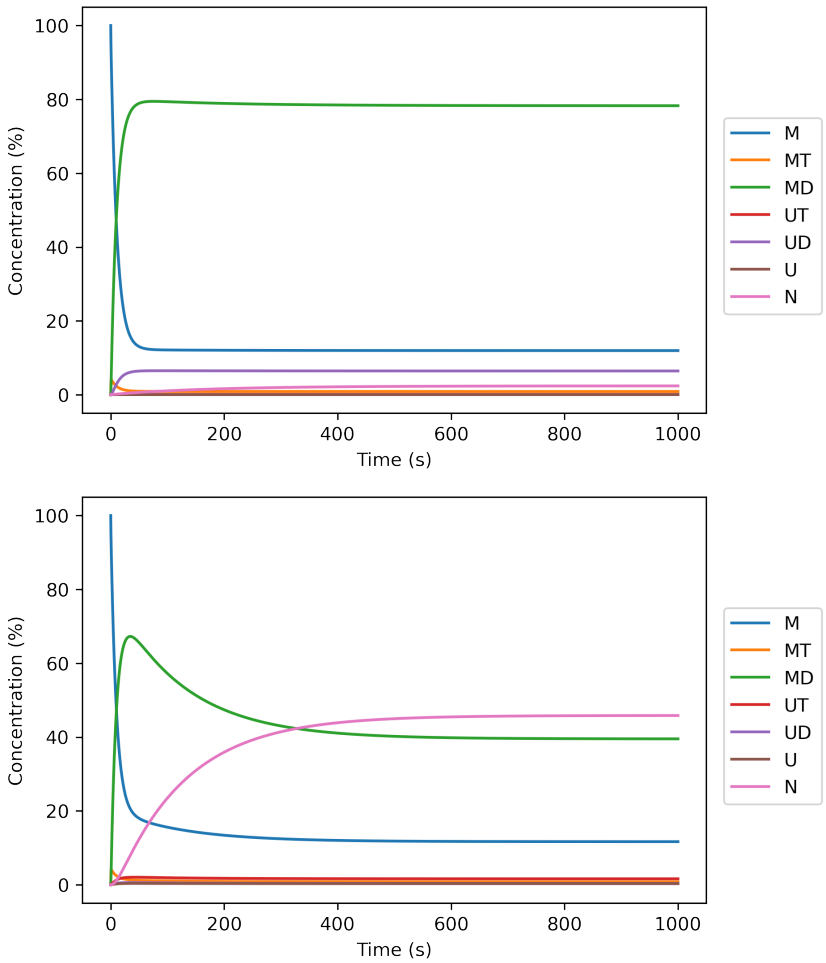


Figure 5.5: Effect of the asymmetry in the hydrolysis and nucleotide exchange rates. Time evolution of the concentrations of the system 4.29 considering the parameters of the initial model (top panel) and adding the effect of cochaperones on the hydrolysis/synthesis and nucleotide exchange rate (bottom panel).

However, in this case, it is also shown that we are able to promote the misfolded substrate toward its native configuration by decreasing the hydrolysis/synthesis rates

and increasing the rates for nucleotide exchange when the substrate is unfolded. In fact, this corresponds to taking into account that JDPs might leave after the hydrolysis of ATP and the intervention of NEFs, which ultimately facilitate the release of the substrate. Particularly, the recovery of misfolded substrates is higher when these rates are modified so that the transition rates between the states MT and MD are asymmetric with respect to the ones between the states UT and UD.

Exploitation of the asymmetry

To investigate the main determinants of this asymmetry, first, we define basal and maximal rates for the hydrolysis/synthesis and nucleotide exchange reactions (see Table 5.1). These are derived by increasing or decreasing the original rates of the reactions depending on the presence of a cochaperone. Second, we define modified rates for these reactions in the following way

$$\begin{aligned}\hat{k}_{ij}(\mathrm{y}) &= k_{ij}^{\mathrm{basal}}\left[1 - P(\mathrm{x\,S})\right] + k_{ij}^{\mathrm{max}} P(\mathrm{x\,S}) \\ &= k_{ij}^{\mathrm{basal}} \frac{K_{\mathrm{d}}^{\mathrm{x}}(\mathrm{y})}{K_{\mathrm{d}}^{\mathrm{x}}(\mathrm{y}) + [\mathrm{x}]} + k_{ij}^{\mathrm{max}} \frac{[\mathrm{x}]}{K_{\mathrm{d}}^{\mathrm{x}}(\mathrm{y}) + [\mathrm{x}]}\end{aligned} \tag{5.3}$$

where x denotes the cochaperone (JDP or NEF) bound to the substrate (S), y is the conformational state of the substrate, misfolded (M) or unfolded (U), and $P(\mathrm{x\,S})$ is the probability[1] of having the cochaperone bound to the substrate. Third, for some fixed concentrations of the cochaperones, we exploit the asymmetry between the MT-MD and UT-UD pathways through varying the dissociation constants of the cochaperones from the substrate, $K_{\mathrm{d}}^{\mathrm{x}}(\mathrm{y})$. Finally, we measure the asymmetry from hydrolysis and nucleotide exchange of ADP through the following ratios

$$\frac{\hat{k}_{\mathrm{h}}(\mathrm{M})}{\hat{k}_{\mathrm{h}}(\mathrm{U})}\ ,\ \frac{\hat{k}_{\mathrm{ex,DT}}(\mathrm{U})}{\hat{k}_{\mathrm{ex,DT}}(\mathrm{M})}. \tag{5.4}$$

Here it has been considered that the influence of synthesis and nucleotide exchange of ATP are negligible as their rates are low. Furthermore, for convenience, we take the logarithm of these ratios as then when there is no asymmetry (the rates are equal for both pathways) this quantity is zero. The results are shown in Figure 5.6.

$k_{\mathrm{h}}^{\mathrm{basal}}$	$0.0018\ \mathrm{s}^{-1}$	$k_{\mathrm{s}}^{\mathrm{basal}}$	$1.45 \times 10^{-10}\ \mathrm{s}^{-1}$
$k_{\mathrm{h}}^{\mathrm{max}}$	$1.8\ \mathrm{s}^{-1}$	$k_{\mathrm{s}}^{\mathrm{max}}$	$1.45 \times 10^{-7}\ \mathrm{s}^{-1}$
$k_{\mathrm{ex,TD}}^{\mathrm{basal}}$	$2.27 \times 10^{-5}\ \mathrm{s}^{-1}$	$k_{\mathrm{ex,DT}}^{\mathrm{basal}}$	$0.0183\ \mathrm{s}^{-1}$
$k_{\mathrm{ex,TD}}^{\mathrm{max}}$	$2.27 \times 10^{-3}\ \mathrm{s}^{-1}$	$k_{\mathrm{ex,DT}}^{\mathrm{max}}$	$1.83\ \mathrm{s}^{-1}$

Table 5.1: Basal and maximal rates for hydrolysis ($k_{\mathrm{h}}^{\mathrm{basal}}, k_{\mathrm{h}}^{\mathrm{max}}$), synthesis ($k_{\mathrm{s}}^{\mathrm{basal}}, k_{\mathrm{s}}^{\mathrm{max}}$), and nucleotide exchange ($k_{\mathrm{ex,TD}}^{\mathrm{basal}}, k_{\mathrm{ex,TD}}^{\mathrm{max}}, k_{\mathrm{ex,DT}}^{\mathrm{basal}}, k_{\mathrm{ex,DT}}^{\mathrm{max}}$).

[1]The derivation of this expression can be found in Appendix A.2

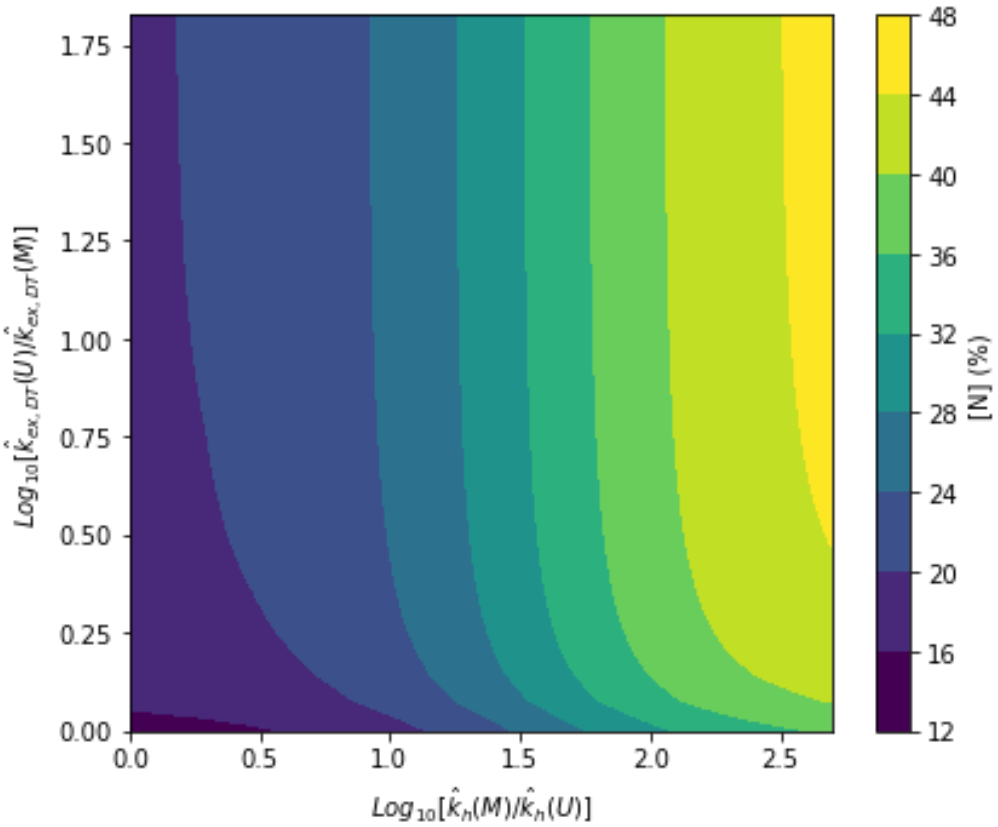


Figure 5.6: Comparison between the asymmetry from hydrolysis and from nucleotide exchange. Contour map of the concentration of native substrate at the steady state as a function of the asymmetry from hydrolysis ($\hat{k}_{\mathrm{h}}(\mathrm{M})/\hat{k}_{\mathrm{h}}(\mathrm{U})$) and from nucleotide exchange ($\hat{k}_{\mathrm{ex,DT}}(\mathrm{U})/\hat{k}_{\mathrm{ex,DT}}(\mathrm{M})$). In the simulations, the dissociation constants of the cochaperones were ranged from 0.01 µM to 1000 µM while considering [JDP] = 1 µM and [NEF] = 5 µM, which are values close to physiological conditions.

It can be observed that if there is asymmetry from the hydrolysis but not from the nucleotide exchange, the yield of native substrates is still high. On the contrary, if there is no asymmetry from the hydrolysis but only from the nucleotide exchange, then the yield of native proteins is severely reduced. So, the asymmetry from ATP hydrolysis has a larger effect on the final concentration of native than the asymmetry from nucleotide exchange. This can be reasoned by considering that the functional cycle of Hsp70 has a certain direction and that ATP hydrolysis comes first.

Furthermore, these results also indicate that it is essential that the enhancement of ATP hydrolysis by JDPs only happens when the substrate is misfolded. This observation again connects with the idea that JPDs unbinding from the substrate likely occurs after hydrolysis. Nevertheless, this does not mean that the action of NEFs is irrelevant. In fact, the concentration of native proteins is maximal when there is asymmetry from both hydrolysis and nucleotide exchange.

Connection with experiments

First, we consider the experiments from Imamoglu et al. [23]. In these, they showed that the bacterial Hsp70 system, DnaK/DnaJ/GrpE (KJE), is able to resolve non-native firefly luciferase (FLuc) in natural conditions and stabilise the protein in its native state under otherwise denaturing conditions, caused by high temperatures.

In the simulations, we modify the free parameters α and γ (i.e., we alter the stability of the misfolded and native state) to obtain the same behaviour in the case of spontaneous refolding, namely without the intervention of chaperones. Furthermore, to reach maximal yield of native proteins, the case of maximal asymmetry from both hydrolysis and exchange of ADP is considered. Hence, we assume the following dissociation constants of the JDP and NEF cochaperones from the substrate

$$\begin{aligned} K_{\mathrm{d}}^{\mathrm{JDP}}(\mathrm{M}) &= 0.01\ \mu\mathrm{M} \quad K_{\mathrm{d}}^{\mathrm{JDP}}(\mathrm{U}) = 1000\ \mu\mathrm{M} \\ K_{\mathrm{d}}^{\mathrm{NEF}}(\mathrm{M}) &= 1000\ \mu\mathrm{M} \quad K_{\mathrm{d}}^{\mathrm{NEF}}(\mathrm{U}) = 0.01\ \mu\mathrm{M} \end{aligned}, \tag{5.5}$$

which reflect the very low or very high affinity for the substrate depending on its conformation: misfolded (M) or unfolded (U).

- **Natural conditions (T = 25 °C).** Spontaneous refolding and assistance by adding KJE-ATP: [DnaK] = 0.3 µM, [DnaJ] = 0.1 µM, [GrpE] = 0.5 µM, and [ATP] = 5 mM.

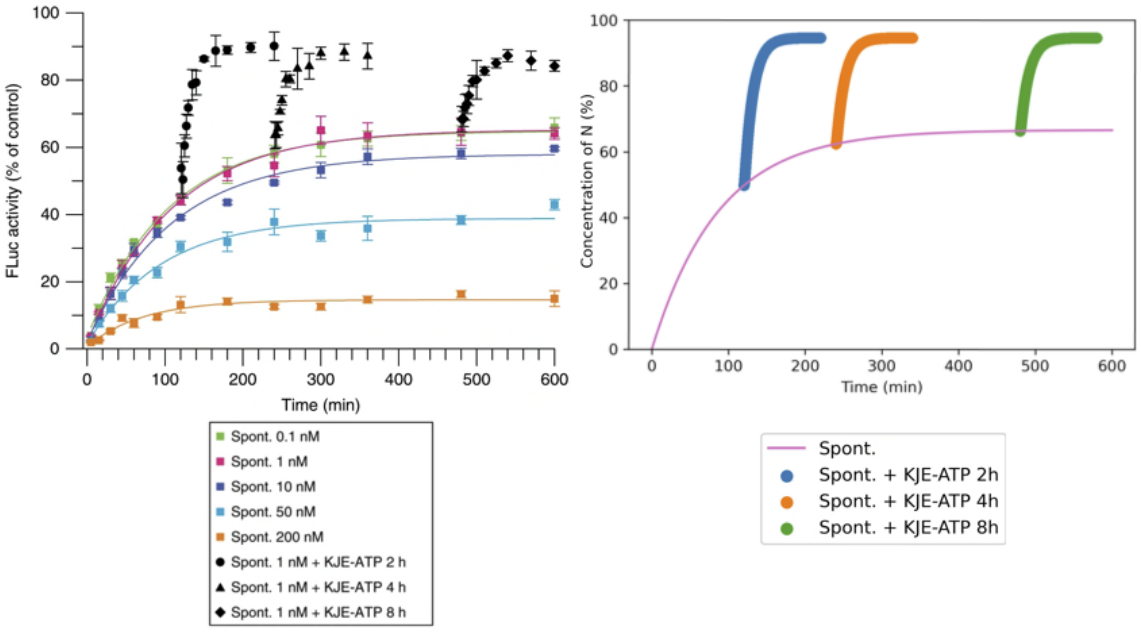


Figure 5.7: Natural conditions. Rescue of spontaneous folding of misfolded substrate (FLuc) by adding KJE-ATP. FLuc activity as a function of time (left panel, figure adapted from [23]). Results from the simulations (right panel).

From the experimental observations at natural conditions, we focus on the case of having 1 nM of misfolded (aggregation is avoided due to the low concentrations)

substrate. It can be observed that the spontaneous refolding has a yield of native proteins of around 80% at t = 600 min. To simulate this scenario, we consider $\alpha = 0.2$ and $\gamma = 0.5$. Furthermore, we also add the assistance of KJE-ATP at the same time than in the experiments. It should be noted that to ease the visualisation we use different colours to represent the assisted refolding, as we do not have discrete measures but a continuous integration over time.

- **Denaturing conditions (T = 37 °C)**. Refolding of 10 pM of misfolded substrate in the presence or absence of KJE-ATP. Assistance by adding: [DnaK] = 1 µM, [DnaJ] = 0.33 µM, [GrpE] = 1.5 µM, and [ATP] = 5 mM.

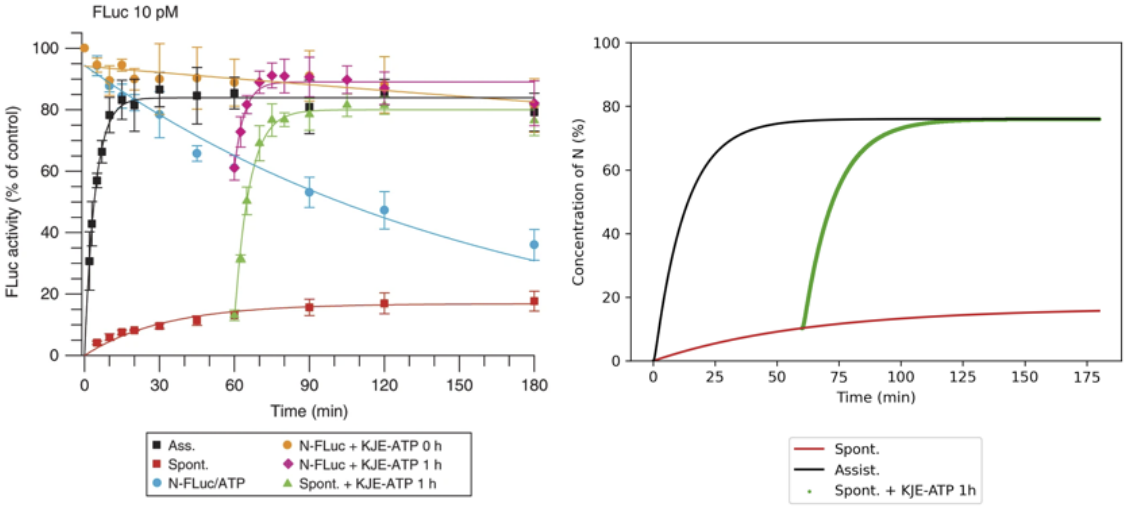


Figure 5.8: Denaturing conditions. Refolding of 10 pM of misfolded substrate (FLuc) in the presence (black line), absence (red line) or assistance (green line) of KJE-ATP. FLuc activity as a function of time (left panel, figure adapted from [23]). Results from the simulations (right panel).

From the experimental observations at denaturing conditions, it can be observed that the spontaneous refolding of 10 pM of misfolded substrate has a yield of native proteins of around 15% at t = 180 min. To simulate this scenario, we consider $\alpha = 0.07$ and $\gamma = 5$. Furthermore, we also add the assistance of KJE-ATP at the same time than in the experiments.

From Figures 5.7 and 5.8, it can be observed that there is a good agreement between the experimental observations and the results of the simulations. It should be remarked apart from the free parameters α and γ that determine the stability of the misfolded (M) and native (N) state, there is no fine-tuning or modification in the rest of the system's rates.

Second, motivated by unpublished observations from Pierre Goloubinoff, we investigate the effect of a large addition of DnaJ in the recovery of 200 nM of urea-preunfolded MLucV (a fluorescent multi-domain luciferase). In the experiments, it is noticed that an overexpression of DnaJ with respect to the concentration of

DnaK (4000 nM) is detrimental (see Figure 5.9). And, while it is well-known that GrpE works in combination with DnaJ in the folding of non-native proteins, in these experiments it is also observed that the addition of GrpE can help to reduce the inhibitory effects of excessive amounts of DnaJ.

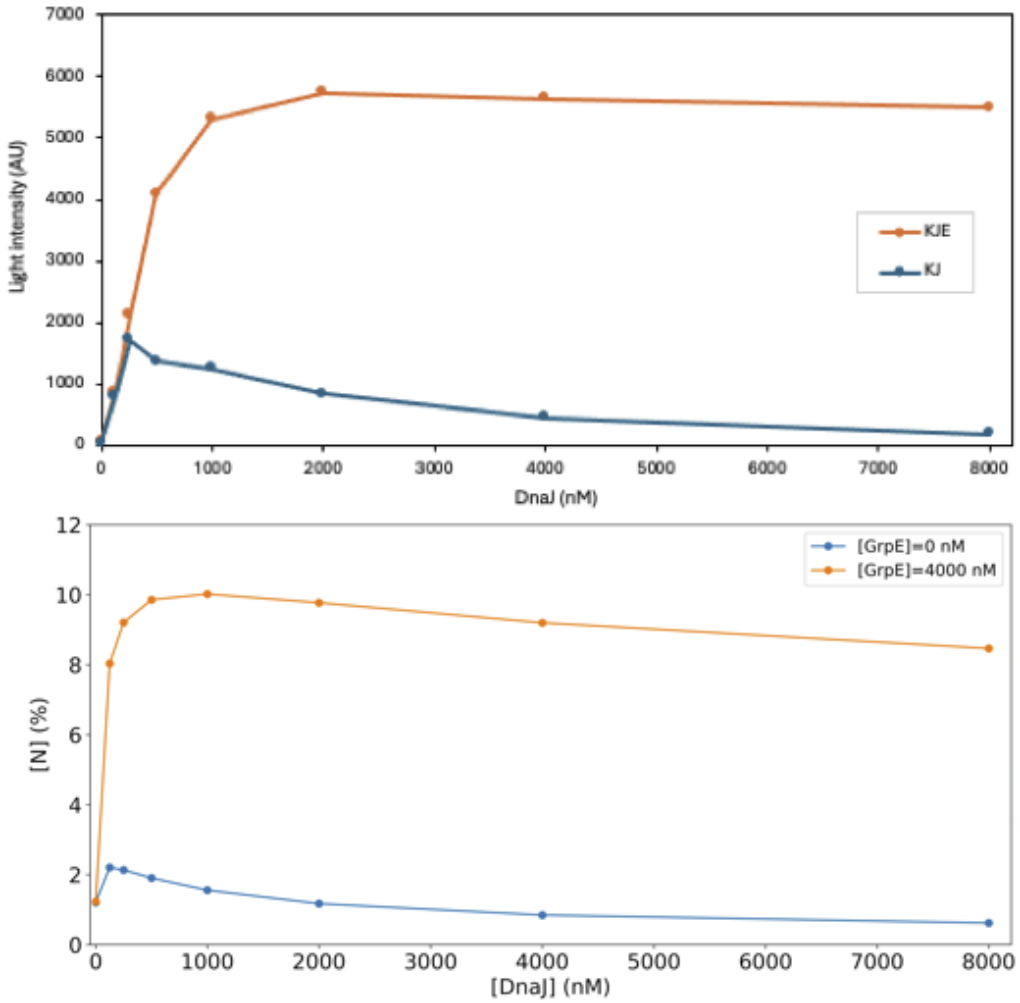


Figure 5.9: Detrimental effect due to the excess of DnaJ. Light intensity (photon emission produced from the luciferase reaction) of the substrate (MLucV) as a function of the concentration of DnaJ (from 0 to 8000 nM) in the absence (blue line) or presence (orange line) of GrpE (top panel, from unpublished observations of P. Goloubinoff). Concentration of native proteins at the steady state as a function of the concentration of DnaJ from the simulations. Both the cases of absence or presence of GrpE are shown (bottom panel).

For the reproduction of the presented behaviours in the simulations, the dissociation constants of the cochaperones are adapted in the following way:

$$\begin{aligned} K_{\mathrm{d}}^{\mathrm{JDP}}(\mathrm{M}) &= 0.1\ \mu\mathrm{M} \quad K_{\mathrm{d}}^{\mathrm{JDP}}(\mathrm{U}) = 10\ \mu\mathrm{M} \\ K_{\mathrm{d}}^{\mathrm{NEF}}(\mathrm{M}) &= 100\ \mu\mathrm{M} \quad K_{\mathrm{d}}^{\mathrm{NEF}}(\mathrm{U}) = 0.01\ \mu\mathrm{M}. \end{aligned} \tag{5.6}$$

It should be noticed that presence of asymmetry from hydrolysis and ADP exchange is still needed. Furthermore, since in this case we did not have the data for spon-

taneous refolding, we independently define $k_{U,M} = 0.0007\ s^{-1}$ and $k_{U,N} = 0.4\ s^{-1}$, with $k_{N,U} = 0.04\ s^{-1}$ so the misfolded state is still more stable than the native configuration.

In this case, we were unable to have the same units on the y-axis. Therefore, the comparison between the experimental data and the simulation results is not based on the final yield of native proteins but rather on the relative behaviour between the cases with and without GrpE, which is to a great extent accurately reproduced.

Given these observations, we decided to simulate what happens in the case of an excess of GrpE over the concentration of DnaK (see Figure 5.10).

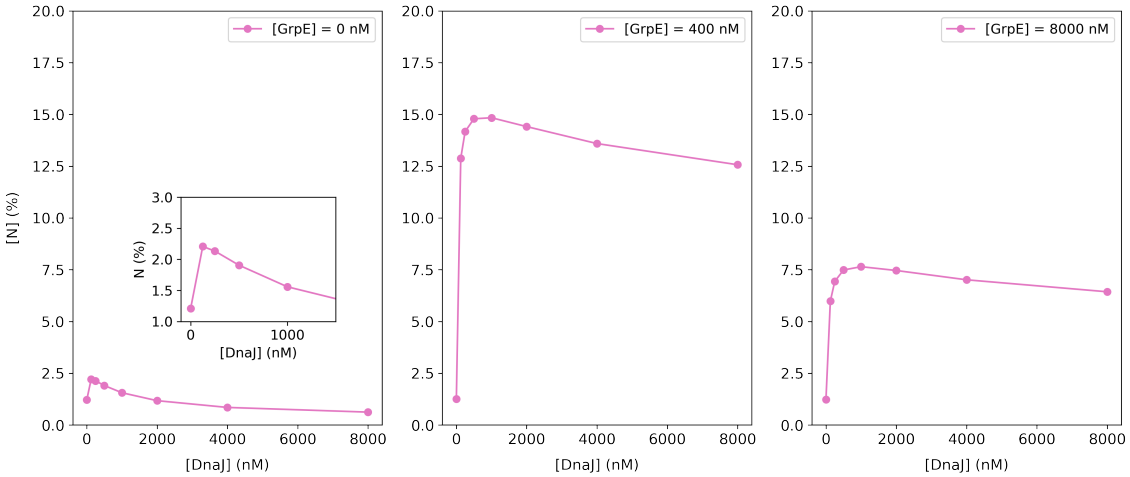


Figure 5.10: Effects of an excess of cochaperones in the system. Concentration of native substrate at the steady state as a function of the concentration of DnaJ, without GrpE (left panel; an inset for low values of DnaJ is also included), with GrpE lower than DnaK (center panel), and with an overexpression of GrpE (right panel).

It can be observed that a certain amount of JDPs and NEFs is needed to reach a maximal yield of native proteins. However, when they are in excessive amounts the final concentration of native substrates is negatively affected. These results are in agreement with some other experimental findings. Particularly, it has been observed that an increase of the concentration of JDP while maintaining fixed the concentration of Hsp70 leads to a decrease in the yield of refolded substrates [68]. Analogously, it has been also noted that folding efficiency declines when there is an overexpression of NEFs in the system [82]. Furthermore, in other studies it has also been remarked the importance of the fine-tuning of the cochaperones, as an overstimulation of ATP hydrolysis by JDPs may prevent the capture of substrates and an excess of NEFs may cause the premature release of captured clients [66].

In essence, while it has been observed that JDPs and NEFs are necessary factors for the functional cycle of Hsp70s in protein folding, as almost everything in life, an excessive amount of them with respect to Hsp70s might be counterproductive.

6
Conclusion

From the simple model with the substrate protein in three conformational states, imposing thermodynamic consistency, we have been able to derive the transitions rates for the core biochemical reactions of the cycle of Hsp70s in protein folding. That is, chaperone binding/unbinding, ATP hydrolysis/synthesis and nucleotide exchange. Furthermore, from the simulations of this model we have observed that ATP hydrolysis is strictly necessary for Hsp70s to achieve an enhanced non-equilibrium affinity (ultra-affinity) for the substrate, which facilitates chaperone binding.

Based on this information and the widely accepted canonical functional cycle for Hsp70s in protein folding, we proposed an initial model consisting of five possible conformational states for the substrate. Even though this model was unable to reproduce the expected recovery of misfolded proteins, it has indicated the need of considering in more detail the unfolding action of Hsp70s.

Finally, the last model proposed for protein folding suggests that an asymmetry in the reaction network is needed to promote the transition toward the native state. In particular, it seems necessary for the cochaperones assisting Hsp70s (JDPs and NEFs) to have different affinities for the substrate depending on its conformation (misfolded or unfolded). Likewise, we have observed that the asymmetry from the hydrolysis of ATP has a higher effect on the yielding of native proteins than the asymmetry from the nucleotide exchange of ADP. While these ideas seem reasonable from the perspective of the system's network and how the cycle is directed through it, the underpinning reason at the molecular level has not yet been determined. Furthermore, with this model we have been able to reproduce experimental observations. Indeed, it has been shown that it can model the Hsp70 chaperones rescue of misfolded proteins in natural conditions and their stabilisation under otherwise denaturing conditions, and the counterproductive effects of an excess of JDPs and NEFs in the yield of native proteins.

Even though our model, at the coarse-grained level used, seems to capture the intricacies of the mechanism of Hps70s in protein folding, it still needs improvement and could benefit from experimental observations. For instance, by measuring the time it takes for a substrate protein to unfold once it is bound to a chaperone or by performing a study on the dissociation constants of the cochaperones depending on the conformational state of the substrate. Indeed, adding these details would eliminate the majority of the freedom in the rate constants and thus be determinant in assessing the reliability of the dynamics observed in the model up to this point.

Moreover, since the proposed model is intended to explain the generic behaviour and activity of canonical Hsp70 systems, it would also be interesting to see if it can reproduce observations from experiments using Hsp70 chaperones other than the bacterial one, such as those present in yeast or the human body.

Bibliography

[1] Wiesława Widłak. Protein structure and function. pages 15–29, 2013.

[2] Giulio Chiesa, Szilvia Kiriakov, and Ahmad S. Khalil. Protein assembly systems in natural and synthetic biology. *BMC Biology 2020 18:1*, 18:1–18, 3 2020.

[3] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. The shape and structure of proteins. 2002.

[4] Asra Nasir Khan and Rizwan Hasan Khan. Protein misfolding and related human diseases: A comprehensive review of toxicity, proteins involved, and current therapeutic strategies. *International Journal of Biological Macromolecules*, 223:143–160, 12 2022.

[5] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 7 1973.

[6] Margarita Brilkova, Martina Nigri, Harshitha Santhosh Kumar, James Moore, Matilde Mantovani, Claudia Keller, Amandine Grimm, Anne Eckert, Dimitri Shcherbakov, Rashid Akbergenov, Petra Seebeck, Stefanie D. Krämer, David P. Wolfer, Thomas C. Gent, and Erik C. Böttger. Error-prone protein synthesis recapitulates early symptoms of alzheimer disease in aging mice. *Cell Reports*, 40:111433, 9 2022.

[7] Lisa J. Lapidus. Protein unfolding mechanisms and their effects on folding experiments. *F1000Research*, 6:1723, 2017.

[8] David Balchin, Manajit Hayer-Hartl, and F. Ulrich Hartl. In vivo aspects of protein folding and quality control. *Science*, 353, 7 2016.

[9] Bruno Fauvet, Mathieu E. Rebeaud, Satyam Tiwari, Paolo De Los Rios, and Pierre Goloubinoff. Repair or degrade: the thermodynamic dilemma of cellular protein quality-control. *Frontiers in Molecular Biosciences*, 8:768888, 10 2021.

[10] F. Ulrich Hartl. Molecular chaperones in cellular protein folding. *Nature 1996 381:6583*, 381:571–580, 1996.

[11] Fabrizio Chiti, Niccolò Taddei, Fabiana Baroni, Cristina Capanni, Massimo Stefani, Giampietro Ramponi, and Christopher M. Dobson. Kinetic partitioning of protein folding and aggregation. *Nature Structural Biology 2002 9:2*, 9:137–143, 1 2002.

[12] Andrija Finka, Rayees U.H. Mattoo, and Pierre Goloubinoff. Experimental

milestones in the discovery of molecular chaperones as polypeptide unfolding enzymes. *Annual Review of Biochemistry*, 85:715–742, 6 2016.

[13] Shruti Sharma, Kausik Chakraborty, Barbara K. Müller, Nagore Astola, Yun Chi Tang, Don C. Lamb, Manajit Hayer-Hartl, and F. Ulrich Hartl. Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell*, 133:142–153, 4 2008.

[14] Anne S. Wentink, Nadinath B. Nillegoda, Jennifer Feufel, Gabrielė Ubartaitė, Carolyn P. Schneider, Paolo De Los Rios, Janosch Hennig, Alessandro Barducci, and Bernd Bukau. Molecular dissection of amyloid disaggregation by human hsp70. *Nature 2020 587:7834*, 587:483–488, 11 2020.

[15] Anat Ben-Zvi, Paolo De Los Rios, Giovanni Dietler, and Pierre Goloubinoff. Active solubilization and refolding of stable protein aggregates by cooperative unfolding action of individual hsp70 chaperones. *Journal of Biological Chemistry*, 279:37298–37303, 9 2004.

[16] Mathieu E. Rebeaud, Saurav Mallik, Pierre Goloubinoff, and Dan S. Tawfik. On the evolution of chaperones and cochaperones and the expansion of proteomes across the tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, 118:e2020885118, 5 2021.

[17] Jürgen Radons. The human hsp70 family of chaperones: where do we stand? *Cell Stress and Chaperones*, 21:379–404, 5 2016.

[18] M. P. Mayer and B. Bukau. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences 2005 62:6*, 62:670–684, 3 2005.

[19] Matthias P. Mayer. Hsp70 chaperone dynamics and molecular mechanism. *Trends in Biochemical Sciences*, 38:507–514, 10 2013.

[20] Paolo De Los Rios and Alessandro Barducci. Hsp70 chaperones are non-equilibrium machines that achieve ultra-affinity by energy consumption. *eLife*, 2014, 5 2014.

[21] Alessandro Barducci and Paolo De Los Rios. Non-equilibrium conformational dynamics in the function of molecular chaperones. *Current Opinion in Structural Biology*, 30:161–169, 2 2015.

[22] Pierre Goloubinoff, Alberto S. Sassi, Bruno Fauvet, Alessandro Barducci, and Paolo De Los Rios. Chaperones convert the energy from atp into the nonequilibrium stabilization of native proteins. *Nature Chemical Biology 2018 14:4*, 14:388–395, 3 2018.

[23] Rahmi Imamoglu, David Balchin, Manajit Hayer-Hartl, and F. Ulrich Hartl. Bacterial hsp70 resolves misfolded states and accelerates productive folding of a multi-domain protein. *Nature Communications 2020 11:1*, 11:1–13, 1 2020.

[24] Satyam Tiwari, Bruno Fauvet, Salvatore Assenza, Paolo De Los Rios, and Pierre Goloubinoff. A fluorescent multi-domain protein reveals the unfolding mechanism of hsp70. *Nature Chemical Biology 2022 19:2*, 19:198–205, 10 2022.

[25] Evan T. Powers and Lila M. Gierasch. The proteome folding problem and cellular proteostasis. *Journal of Molecular Biology*, 433:167197, 10 2021.

[26] Christian B. Anfinsen, Edgar Haber, Michael Sela, and Frederick H. White. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, 47:1309–1314, 9 1961.

[27] Christian B. Anfinsen and Edgar Haber. Studies on the reduction and reformation of protein disulfide bonds. *Journal of Biological Chemistry*, 236:1361–1363, 5 1961.

[28] José Nelson Onuchic and Peter G. Wolynes. Theory of protein folding. *Current Opinion in Structural Biology*, 14:70–75, 2 2004.

[29] Nikolaos Louros, Joost Schymkowitz, and Frederic Rousseau. Mechanisms and pathology of protein misfolding and aggregation. *Nature Reviews Molecular Cell Biology*, 24:912–933, 12 2023.

[30] Cyrus Levinthal. Are there pathways for protein folding? *Journal de Chimie Physique*, 65:44–45, 1968.

[31] Christopher M. Dobson. Protein folding and misfolding. *Nature 2003 426:6968*, 426:884–890, 12 2003.

[32] Abhisek Mukherjee, Diego Morales-Scheihing, Peter C. Butler, and Claudio Soto. Type 2 diabetes as a protein misfolding disease. *Trends in Molecular Medicine*, 21:439–449, 7 2015.

[33] Christopher A. Ross and Michelle A. Poirier. Protein aggregation and neurodegenerative disease. *Nature Medicine 2004 10:7*, 10:S10–S17, 7 2004.

[34] Bernd Bukau, Jonathan Weissman, and Arthur Horwich. Molecular chaperones and protein quality control. *Cell*, 125:443–451, 5 2006.

[35] F. Ulrich Hartl, Andreas Bracher, and Manajit Hayer-Hartl. Molecular chaperones in protein folding and proteostasis. *Nature 2011 475:7356*, 475:324–332, 7 2011.

[36] Helen Saibil. Chaperone machines for protein folding, unfolding and disaggregation. *Nature Reviews Molecular Cell Biology 2013 14:10*, 14:630–642, 9 2013.

[37] Daniel K. Clare and Helen R. Saibil. Atp-driven molecular chaperone machines. *Biopolymers*, 99:846–859, 11 2013.

[38] Raphael Trösch, Timo Mühlhaus, Michael Schroda, and Felix Willmund. Atp-dependent molecular chaperones in plastids — more complex than expected. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1847:872–888, 9 2015.

[39] Yulia Kushnareva and Donald D. Newmeyer. Bioenergetics and cell death. *Annals of the New York Academy of Sciences*, 1201:50–57, 7 2010.

[40] Hao Ge, Min Qian, and Hong Qian. Stochastic theory of nonequilibrium steady states. part ii: Applications in chemical biophysics. *Physics Reports*, 510:87–118, 1 2012.

[41] Serena Carra, Simon Alberti, Patrick A. Arrigo, Justin L. Benesch, Ivor J. Benjamin, Wilbert Boelens, Britta Bartelt-Kirbach, Bianca J.J.M. Brundel, Johannes Buchner, Bernd Bukau, John A. Carver, Heath Ecroyd, Cecilia Emanuelsson, Stephanie Finet, Nikola Golenhofen, Pierre Goloubinoff, Nikolai Gusev, Martin Haslbeck, Lawrence E. Hightower, Harm H. Kampinga, Rachel E. Klevit, Krzysztof Liberek, Hassane S. Mchaourab, Kathryn A. McMenimen, Angelo Poletti, Roy Quinlan, Sergei V. Strelkov, Melinda E. Toth, Elizabeth Vierling, and Robert M. Tanguay. The growing world of small heat shock proteins: from structure to functions. *Cell Stress and Chaperones*, 22:601–611, 7 2017.

[42] Thorsten Rogalla, Monika Ehrnsperger, Xavier Preville, Alexey Kotlyarov, Gudrun Lutsch, Cécile Ducasse, Catherine Paul, Martin Wieske, André Patrick Arrigo, Johannes Buchner, and Matthias Gaestel. Regulation of hsp27 oligomerization, chaperone function, and protective activity against oxidative stress/-tumor necrosis factor α by phosphorylation. *Journal of Biological Chemistry*, 274:18947–18956, 7 1999.

[43] Jirka Peschek, Nathalie Braun, Julia Rohrberg, Katrin Christiane Back, Thomas Kriehuber, Andreas Kastenmüller, Seèil Weinkauf, and Johannes Buchner. Regulated structural transitions unleash the chaperone activity of αb-crystallin. *Proceedings of the National Academy of Sciences of the United States of America*, 110:E3780–E3789, 10 2013.

[44] Rina Rosenzweig, Nadinath B. Nillegoda, Matthias P. Mayer, and Bernd Bukau. The hsp70 chaperone network. *Nature Reviews Molecular Cell Biology 2019 20:11*, 20:665–680, 6 2019.

[45] Katharina Ernst, Johannes Schmid, Matthias Beck, Marlen Hägele, Meike Hohwieler, Patricia Hauff, Anna Katharina Ückert, Anna Anastasia, Michael Fauler, Thomas Jank, Klaus Aktories, Michel R. Popoff, Cordelia Schiene-Fischer, Alexander Kleger, Martin Müller, Manfred Frick, and Holger Barth. Hsp70 facilitates trans-membrane transport of bacterial adp-ribosylating toxins into the cytosol of mammalian cells. *Scientific Reports 2017 7:1*, 7:1–16, 6 2017.

[46] William B. Pratt and David O. Toft. Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Experimental biology and medicine (Maywood, N.J.)*, 228:111–133, 2003.

[47] Jörg Höhfeld, Douglas M. Cyr, and Cam Patterson. From the cradle to the grave: molecular chaperones that may choose between folding and degradation. *EMBO reports*, 2:885–890, 2001.

[48] Mads Daugaard, Mikkel Rohde, and Marja Jäättelä. The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. *FEBS Letters*, 581:3702–3710, 7 2007.

[49] Maureen E. Murphy. The hsp70 family and cancer. *Carcinogenesis*, 34:1181–1188, 6 2013.

[50] Erik R.P. Zuiderweg, Lawrence E. Hightower, and Jason E. Gestwicki. The remarkable multivalency of the hsp70 chaperones. *Cell Stress and Chaperones*, 22:173–189, 3 2017.

[51] Axel Mogk, Eva Kummer, and Bernd Bukau. Cooperation of hsp70 and hsp100 chaperone machines in protein disaggregation. *Frontiers in Molecular Biosciences*, 2:146947, 5 2015.

[52] Matthias P. Mayer, Hartwig Schröder, Stefan Rüdiger, Klaus Paal, Thomas Laufen, and Bernd Bukau. Multistep mechanism of substrate binding determines chaperone activity of hsp70. *Nature Structural Biology 2000 7:7*, 7:586–593, 7 2000.

[53] Ofrah Faust and Rina Rosenzweig. Structural and biochemical properties of hsp40/hsp70 chaperone system. *Advances in Experimental Medicine and Biology*, 1243:3–20, 2020.

[54] Rick Russell, Robert Jordan, and Roger McMacken. Kinetic characterization of the atpase cycle of the dnak molecular chaperone. *Biochemistry*, 37:596–607, 1 1998.

[55] Bartlomiej Tomiczek, Wojciech Delewski, Lukasz Nierzwicki, Milena Stolarska, Igor Grochowina, Brenda Schilke, Rafal Dutkiewicz, Marta A. Uzarska, Szymon J. Ciesielski, Jacek Czub, Elizabeth A. Craig, and Jaroslaw Marszalek. Two-step mechanism of j-domain action in driving hsp70 function. *PLoS computational biology*, 16, 6 2020.

[56] Andreas Bracher and Jacob Verghese. The nucleotide exchange factors of hsp70 molecular chaperones. *Frontiers in Molecular Biosciences*, 2:10, 4 2015.

[57] Robert T. Youker and Jeffrey L. Brodsky. Regulation of hsp70 function: Hsp40 co-chaperones and nucleotide exchange factors. *Cell Stress Proteins*, pages 209–227, 2007.

[58] Basile Nguyen, David Hartich, Udo Seifert, and Paolo De Los Rios. Thermodynamic bounds on the ultra- and infra-affinity of hsp70 for its substrates. *Biophysical Journal*, 113:362–370, 7 2017.

[59] Sandeep K. Sharma, Paolo De Los Rios, Philipp Christen, Ariel Lustig, and Pierre Goloubinoff. The kinetic parameters and energy cost of the hsp70 chaperone as a polypeptide unfoldase. *Nature Chemical Biology 2010 6:12*, 6:914–920, 10 2010.

[60] Ruth Kellner, Hagen Hofmann, Alessandro Barducci, Bengt Wunderlich, Daniel Nettels, and Benjamin Schuler. Single-molecule spectroscopy reveals chaperone-mediated expansion of substrate protein. *Proceedings of the National Academy of Sciences of the United States of America*, 111:13355–13360, 9 2014.

[61] Salvatore Assenza, Alberto S. Sassi, Ruth Kellner, Benjamin Schuler, Paolo

De Los Rios, and Alessandro Barducci. Efficient conversion of chemical energy into mechanical work by hsp70 chaperones. *eLife*, 8, 12 2019.

[62] Ron Milo and Rob Phillips. Cell biology by the numbers. *Cell Biology by the Numbers*, 12 2015.

[63] Paolo De Los Rios, Anat Ben-Zvi, Olga Slutsky, Abdussalam Azem, and Pierre Goloubinoff. Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. *Proceedings of the National Academy of Sciences of the United States of America*, 103:6166–6171, 4 2006.

[64] Pierre Goloubinoff and Paolo De Los Rios. The mechanism of hsp70 chaperones: (entropic) pulling the models together. *Trends in Biochemical Sciences*, 32:372–380, 8 2007.

[65] Paolo De Los Rios and Pierre Goloubinoff. Hsp70 chaperones use atp to remodel native protein oligomers and stable aggregates by entropic pulling. *Nature Structural & Molecular Biology 2016 23:9*, 23:766–769, 9 2016.

[66] Harm H. Kampinga and Elizabeth A. Craig. The hsp70 chaperone machinery: J proteins as drivers of functional specificity. *Nature Reviews Molecular Cell Biology 2010 11:8*, 11:579–592, 8 2010.

[67] Stefan Rüdiger, Jens Schneider-Mergener, and Bernd Bukau. Its substrate specificity characterizes the dnaj co-chaperone as a scanning factor for the dnak chaperone. *The EMBO Journal*, 20:1042–1050, 3 2001.

[68] Thomas Laufen, Matthias P. Mayer, Christian Beisel, Dagmar Klostermeier, Axel Mogk, Jochen Reinstein, and Bernd Bukau. Mechanism of regulation of hsp70 chaperones by dnaj cochaperones. *Proceedings of the National Academy of Sciences of the United States of America*, 96:5452–5457, 5 1999.

[69] Toshifumi Tomoyasu, Teru Ogura, Takashi Tatsuta, and Bernd Bukau. Levels of dnak and dnaj provide tight control of heat shock gene expression and protein repair in escherichia coli. *Molecular microbiology*, 30:567–581, 1998.

[70] Bin Hu, Matthias P. Mayer, and Masaru Tomita. Modeling hsp70-mediated protein folding. *Biophysical Journal*, 91:496–507, 7 2006.

[71] Dirk Brehmer, Stefan Rüdiger, Claudia S. Gässler, Dagmar Klostermeier, Lars Packschies, Jochen Reinstein, Matthias P. Mayer, and Bernd Bukau. Tuning of chaperone activity of hsp70 proteins by modulation of nucleotide exchange. *Nature Structural Biology 2001 8:5*, 8:427–432, 2001.

[72] Larry E. Vickery and Jill R. Cupp-Vickery. Molecular chaperones hsca/ssq1 and hscb/jac1 and their roles in iron-sulfur protein maturation. *Critical Reviews in Biochemistry and Molecular Biology*, 42:95–111, 3 2007.

[73] Cato M. Guldberg and Peter Waage. *Studier i affiniteten*, page 35. 1864.

[74] Cato M. Guldberg and Peter Waage. *Études sur les affinités chimiques*. Christiana, 1867.

[75] Cato M. Guldberg and Peter Waage. Ueber die chemische affinität. *Journal fur Praktische Chemie*, 19, 1879.

[76] Rud Wegscheider. Über simultane gleichgewichte und die beziehungen zwischen thermodynamik und reactionskinetik homogener systeme. *Monatshefte für Chemie*, 32:849–906, 8 1911.

[77] Lewis N. Gilbert. A new principle of equilibrium. *Proceedings of the National Academy of Sciences*, 11:179–183, 3 1925.

[78] Martin J. Klein. Principle of detailed balance. *Physical Review*, 97:1446, 3 1955.

[79] Christian Maes. Local detailed balance. *SciPost Phys. Lect. Notes*, 32, 2021.

[80] Robert A. Alberty. Thermodynamics of biochemical reactions. *Thermodynamics of Biochemical Reactions*, 2 2003.

[81] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78:2690, 4 1997.

[82] Ellen A. A. Nollen, Jeanette F. Brunsting, Jaewhan Song, Harm H. Kampinga, and Richard I. Morimoto. Bag1 functions in vivo as a negative regulator of hsp70 chaperone activity. *Molecular and Cellular Biology*, 20:1083, 2 2000.

A

Mathematical derivations

A.1 Detailed balance condition in a cycle of three states

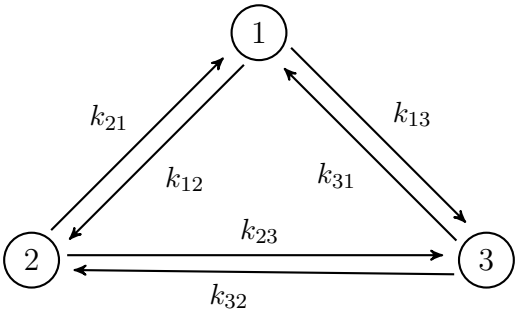


Figure A.1: Network representation of an arbitrary three-states system.

Applying the detailed balance condition (Equation (3.8)) in the kinetic scheme displayed in Figure A.1:

$$\begin{aligned} k_{21}[2]_{\text{eq}} &= k_{12}[1]_{\text{eq}} \\ k_{13}[1]_{\text{eq}} &= k_{31}[3]_{\text{eq}} \\ k_{23}[2]_{\text{eq}} &= k_{32}[3]_{\text{eq}}, \end{aligned} \tag{A.1}$$

where $[i]_{\text{eq}}$ denotes the concentration of state i at equilibrium.

Assuming that all concentrations and constant rates are non-zero. From the first equality, one obtains

$$\frac{[1]_{\text{eq}}}{[2]_{\text{eq}}} = \frac{k_{21}}{k_{12}}. \tag{A.2}$$

While from dividing the second and third equations

$$\frac{k_{13}[1]_{\text{eq}}}{k_{23}[2]_{\text{eq}}} = \frac{k_{31}}{k_{32}}. \tag{A.3}$$

Finally, substituting equation (A.2) in (A.3) leads to

$$k_{13}k_{32}k_{21} = k_{12}k_{23}k_{31}, \tag{A.4}$$

which corresponds to the detailed balance condition in a closed cycle (Equation (3.10).

A.2 Probability of the molecule complex AB

Considering the chemical components A,B that can form the complex AB through a binding process. The reversible chemical reaction that describes the process is

$$\mathrm{A} + \mathrm{B} \underset{k_{\mathrm{off}}^{\mathrm{A}}}{\overset{k_{\mathrm{on}}^{\mathrm{A}}}{\rightleftharpoons}} \mathrm{AB}, \tag{A.5}$$

where $k_{\mathrm{on}}, k_{\mathrm{off}}$ are the binding and unbinding rates.

From classical probability, the probability to have the complex AB is

$$P(\mathrm{AB}) = \frac{[\mathrm{AB}]}{[\mathrm{A}]_{\mathrm{t}}}, \tag{A.6}$$

where [AB] is concentration of the complex AB and $[\mathrm{A}]_{\mathrm{t}}$ the total concentration of A in the system.

The time evolution of the concentration of the complex AB (Equation (4.1)) is

$$\frac{d[\mathrm{AB}]}{dt} = k_{\mathrm{on}}[\mathrm{A}][\mathrm{B}] - k_{\mathrm{off}}[\mathrm{AB}], \tag{A.7}$$

where [A],[B] are the concentrations of free A and B.

Assuming that the concentration of AB is the steady state one, leads to

$$k_{\mathrm{on}}[\mathrm{A}][\mathrm{B}] - k_{\mathrm{off}}[\mathrm{AB}] = 0, \tag{A.8}$$

where the steady state condition (Equation (3.7) has been applied to Equation (A.7).

While, from the mass conservation law, the total concentrations of A and B in the system ($[\mathrm{A}]_{\mathrm{t}}, [\mathrm{B}]_{\mathrm{t}}$) are given by

$$\begin{aligned} [\mathrm{A}]_{\mathrm{t}} &= [\mathrm{A}] + [\mathrm{AB}] \\ [\mathrm{B}]_{\mathrm{t}} &= [\mathrm{B}] + [\mathrm{AB}] \end{aligned}. \tag{A.9}$$

So, for instance, considering that the total concentration of B is in excess over the total concentration of A

$$[\mathrm{B}]_{\mathrm{t}} \gg [\mathrm{A}]_{\mathrm{t}}$$

and that, in turn

$$[\mathrm{A}]_{\mathrm{t}} \geq [\mathrm{AB}],$$

we have that $[\mathrm{B}]_{\mathrm{t}} \gg [\mathrm{AB}]$. Therefore, we can assume $[\mathrm{B}]_{\mathrm{t}} \approx [\mathrm{B}]$.

Introducing it into Equation (A.8) leads to

$$k_{\mathrm{on}}\Big([\mathrm{A}]_{\mathrm{t}} - [\mathrm{AB}]\Big)[\mathrm{B}]_{\mathrm{t}} - k_{\mathrm{off}}[\mathrm{AB}] = 0, \tag{A.10}$$

and, solving for [AB] results in

$$[\mathrm{AB}] = \frac{k_{\mathrm{on}}[\mathrm{A}]_{\mathrm{t}}[\mathrm{B}]_{\mathrm{t}}}{k_{\mathrm{off}} + k_{\mathrm{on}}[\mathrm{B}]_{\mathrm{t}}} = \frac{[\mathrm{A}]_{\mathrm{t}}[\mathrm{B}]_{\mathrm{t}}}{K_{\mathrm{d}} + [\mathrm{B}]_{\mathrm{t}}}, \tag{A.11}$$

where $K_{\mathrm{d}} = k_{\mathrm{off}}/k_{\mathrm{on}}$ is the dissociation constant of the system.

Thus, introducing it into Equation (A.6), the probability of having the complex AB is

$$P(\mathrm{AB}) = \frac{[\mathrm{AB}]}{[\mathrm{A}]_{\mathrm{t}}} = \frac{[\mathrm{B}]_{\mathrm{t}}}{K_{\mathrm{d}} + [\mathrm{B}]_{\mathrm{t}}}. \tag{A.12}$$

Furthermore, the probability of not having the complex AB is given by its complementary probability

$$1 - P(\mathrm{AB}) = 1 - \frac{[\mathrm{B}]_{\mathrm{t}}}{K_{\mathrm{d}} + [\mathrm{B}]_{\mathrm{t}}} = \frac{K_{\mathrm{d}}}{K_{\mathrm{d}} + [\mathrm{B}]_{\mathrm{t}}}. \tag{A.13}$$

DEPARTMENT OF LIFE SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY