Non equilibrium thermodynamics of protein organization

THÈSE Nº 8507 (2018)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Alberto Stefano SASSI

EPFL

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

Ovviamente,
per Elisa.

Acknowledgements

I prefer to use the italian language for this part.

Abstract

Three quarters of the thesis will be devoted to the discussion of non equilibrium systems. We show how certain biological systems cannot be described by standard thermodynamics. The reason is that the energy consumption due to the hydrolysis of ATP imposes to the systems a directionality and therefore a chemical flux that is different from zero. More specifically, we will apply the formalism to molecular chaperones, that are proteins involved in a plethora of biological processes, with a particular in-depth analysis on the mechanisms of protein unfolding and refolding mediated by the two families of chaperones Hsp70 and GroEL. Moreover, with the same mathematical tools, we will provide a qualitative description of some phenomena regarding the Atp-binding cassette transporters, that are important membrane protein complexes used to translocate small molecules in and out of the cell.

In the remaining quarter we will analyze the deformation of a polymer chain when it is pulled by an external force. We will explain how it is possible to quantify the orientation along the force and the shrinking along the other directions. It will be also shown that the transverse section shrinks isotropically once the orientation along the force is complete. The main quantities that are used to identify the shape and the orientation of a polymer chain can all be written as a function of a universal quantity that is independent, in first approximation, by the number of monomers, by the rigidity, and even by the volume exclusion.

Key words: non equilibrium physics | molecular chaperones | ABC transporters | polymer theory

Sommario

Tre quarti della tesi saranno dedicati alla trattazione di sistemi di non equilibrio. Si mostra come alcuni sistemi biologici non possono essere descritti con gli strumenti della termodinamica standard. Il motivo è che il consumo di energia dovuto all'idrolisi di ATP impone al sistema una direzionalità e perciò un flusso chimico che è diverso da zero. Più precisamente, si applicherà il formalismo ai chaperone molecolari, che sono proteine impiegate in un gran numero di processi biologici, con particolare riguardo al meccanismo di dispiegamento e ripiegamento mediato dalle due famiglie di chaperone Hsp70 e GroEL. Inoltre, con gli stessi strumenti matematici, sarà fornita una descrizione qualitativa di alcuni fenomeni concernenti i trasportatori ABC, che sono importanti proteine di membrana usate per trasportare piccole molecole all'interno e all'esterno della cellula.

Nel rimanente quarto della tesi verrà analizzata la deformazione di una catena polimerica quando è sottoposta ad una forza di tensione esterna. Si spiegherà come è possibile quantificare l'orientazione lungo la forza e l'assottigliamento lungo le altre direzioni. Si mostrerà, inoltre, che la sezione trasversa viene ridotta in modo isotropico ona volta che l'orientazione verso la forza è completata. Le principali grandezze che sono usate per identificare la forma e l'orientazione di una catena polimerica possono essere tutte scritte come funzioni di una quantità universale che è indipendente, in prima approssimazione, dal numero di monomeri, dalla rigidità, e persino dal volume escluso.

Parole chiave: fisica di non equilibrio | chaperone molecolari | trasportatori ABC | teoria dei polimeri

Contents

Contents

List of Figures

List of Tables

1 Introduction

Besides its intrinsic scientific appeal and its indisputable importance in many biological processes, the topic of protein organization and dynamics has an ever growing interest on the theoretical physics perspective. There are several reasons for this fact, we will just mention the ones that directly motivated our work. First, like other macromolecules, proteins can be seen as *mesoscopic* systems, which means[1] that even though they are formed by a large number of components, they are subjected to non negligible fluctuations [2]. When they are unfolded, the can be modeled with the tools of polymer physics [3, 4] and several conformational properties can be studied, like the scaling laws that regulate the relation between the end-to-end vector and the number of monomers or the response to confinement and other external constraints [5]. On the other hand, when they are functional they often take part to processes in which the thermal fluctuations not only are non negligible, but they are the real driving force that makes the system work. The most evident example are the molecular motors, that are important biological systems distinguished by the peculiarity that they use chemical energy and thermal fluctuations to perform mechanical work [6, 7]. Moreover, proteins represent a case study for some groundbreaking accomplishments in modern statistical physics, like the fluctuation theorem [8] and the important equalities obtained by Jarzynski and Crooks [9, 10], concerning non equilibrium transformations between two states at the same temperature. Second, the path through which a protein assumes its native configuration is also interesting on the point of view of theoretical physics: in order to be functional, it has to reach a complex tertiary structure that in many cases is univocally determined by its amino acids sequence [11]. This leads to the development of statistical methods that are devised to infer the contacts between amino acids on the sole basis of the protein sequence, like co-evolutionary analysis [12, 13]. In addition, protein folding is an intriguing challenge that can be investigated by means of molecular dynamics simulations (MD)[14]. Third, and this is probably the most important reason for our field, proteins are involved in several cellular mechanisms in which there is a continuous energy consumption. Here, the word *continuous* is particularly significant: what we want to highlight is that the systems under our consideration do not

---

[1]We recognize that here we are not giving a very formal definition, but we hope it is sufficient to explain our point.

simply gain a certain amount of energy in order to overcome a kinetic barrier, or escape from a metastable state, so that they can reach more easily the equilibrium state. On the contrary, we are focusing on those systems that are found in states that would not be populated without the support of an energetic source. As a consequence, these systems cannot be described with the formalism of the statistical physics of equilibrium [15, 16, 17]. However, it would be a mistake to conclude that for this reason they are necessarily time dependent. Indeed, after a relaxation time they can be found in *non equilibrium steady states* (NESS), with the main relevant quantities that are constant in time but do not correspond to a minimum in the free energy[17, 18]. An example of systems of this kind is a metallic bar that is kept at different temperatures at its extremities [2]: after a certain interval, there is a net flux of heat along the bar that does not change in time and that would start decreasing as soon as the temperature gradient is removed. There are thus essentially two different kinds of non equilibrium states. One is a *transient state*, which may or may not converge to an equilibrium state after a given time, and the other is the NESS, which does not depend on time and requires the constant supply of external energy. The cell itself can be considered a non equilibrium system. The difference with respect to the metallic bar is that the thermodynamic force is not generated by a temperature gradient but rather by a chemical gradient. The main source of energy in the cell is provided by ATP [19]. The reaction of hydrolysis has a standard free energy $\Delta G^0 \simeq -32.6 \, kJ/mol$ [20]. However, in the cell there is a continuous production of ATP, so that in physiological conditions the ratio between the concentration of ATP and the one of ADP is kept at values of the order of 10 [21].

Among the different mechanisms taking place in the cell, we were particularly interested in the one of protein folding. It is mostly accepted in the biological community that many proteins are capable of spontaneously assuming the functional structure *in vitro* [11]. However, it is fair to argue that the experimental conditions that are normally used to perform *in vitro* experiments do not match the real cellular environment [22]. In physiological conditions there are indeed several factors, like heat or chemical stresses, that can represent an obstacle to the folding process. Therefore, protein folding is helped by molecular chaperones, which are ATP consuming systems that interact with the exposed hydrophobic residues of unfolded and misfolded substrates [1, 23]. We will discuss two of them: the 70-kilodalton heat shock protein (Hsp70) and the bacterial chaperonin GroEL. For Hsp70 we start with a more abstract approach, by showing the theoretical constraints that the rate constants follow when the chaperone is able to efficiently enhance the refolding process. We then move to a model that is closer to experiments, trying to reproduce the results obtained in the group of Ben Schuler [24]. They used the technique of Förster energy transfer spectroscopy (FRET) to show that denatured rhodanese was expanded by Hsp70 in the presence of ATP. With a combination of MD simulations and a rate model[2] we showed how it was possible to relate the substrate expansion to the energy consumption, as well as to obtain FRET efficiencies that are in agreement with the experiments.
In the case of GroEL the problem was tackled in a different way. Above all, we could work

---

[2]The MD simulations were devised and implemented by S. Assenza and A. Barducci.

closely with the group of Pierre Goloubinoff, from the University of Lausanne, so that we were able devise with them the experiments that are more relevant for our message. We used as substrate an enzyme that is functional at 25 °C but it misfolds and it is susceptible to aggregation at 37 °C. However, GroEL, together with its co-chaperone GroES and ATP, was able to recover the functional activity at 37 °C, suggesting that molecular chaperones can stabilize the native state even when it does not correspond to a minimum in free energy in equilibrium conditions.

Even if it is less advanced in the development, we also want to discuss our application of the methods to ABC transporters. We show analytically that the free energy available because of the supply of ATP in the cell constitutes a higher bond to the maximal efficiency of the transporters defined as the ratio between the concentration of molecules on the two sides of the cell membrane. We also reproduce an experiment with exporters in which a concentration gradient in the wrong direction was artificially imposed. In the absence of ATP, an uptake was detected. However, as soon as ATP was injected into the system, the efflux was recovered.

The last part will be only weakly related to the others, but still definitely inherent to the main message. We study the deformation of a polymer chain when it is stretched by an external force, that is a situation that is often encountered both in experiment [25] and in some biological phenomena, like protein translocation [26, 16]. With both analytical calculations and numerical simulations, we show that the transverse section of the ellipsoid which best approximates the polymer starts shrinking isotropically after the orientation. In addition, we show that the important quantities like the projection of the eigenvectors on the direction of the force and the asphericity can be written in terms of a universal parameter, corresponding to the force contribution to the free energy of the system.

The next chapter will be devoted to a brief summary of the main results in non equilibrium physics, with a particular emphasis on their application to biochemical cycles. In chapter 3 we will discuss the application of the formalism to molecular chaperones. More precisely, in the first part we will focus on Hsp70 and in the second part on GroEL. In chapter 4 we will explain how the energy is used by the Atp-binding cassette transporter (ABC) in order to mediate the uptake and the efflux of small molecules through cell membranes. In the last chapter we will describe the shape and the orientation of a polymer subjected to an external tension.

2 Basic principles on non equilibrium systems

In this chapter we want to summarize some of the most important results concerning non equilibrium systems. In the particular case of biochemical cycles, the source of non equilibrium is usually a *chemostat*, that is an external agent which stimulates the injection of a particular type of molecules into the system and the ejection of another type. In the biological processes that we will consider in this thesis, the chemostat is usually the regeneration system of ATP. As we will describe below, the presence of a chemostat implies that some of the rules that are usually valid for equilibrium systems, like the detailed balance, are broken. The main consequence is the emergence of fluxes of reactions, and chemical cycles that are followed in a preferential direction. In any case, the main characteristic of non equilibrium systems, which is also the reason of the obstacles for their systematic statistical description, is their non locality. While for two equilibrium states the difference in free energy provides, in principle, a sufficient indication of the way the two states will be populated after a given relaxation time, in non equilibrium the rate constants of the whole system are important to define the final distribution. We will give some more details on this aspect, then we will discuss the main properties of non equilibrium biochemical cycles, with a particular attention to their steady state solutions.

2.1 Non locality of non equilibrium systems

Before going into the formal description, we want to show an example in which the uniqueness of non equilibrium systems is evident.

Consider an enzyme which can be found in three different states A, B and C. When transforming from B to C and vice versa, the enzyme must always pass through state A (see figure 2.1, left).

We assume that, in the absence of ATP and ADP, B is more stable than C, but the transformation between C and A is much faster than the one between B and A. We also assume that in state B and C the enzyme can bind ATP. In addition, the hydrolysis always leads to a transformation from B, or C, to A (see figure 2.1, right). Both the binding with ATP and the hydrolysis have the same rates for B and for C. Therefore the only
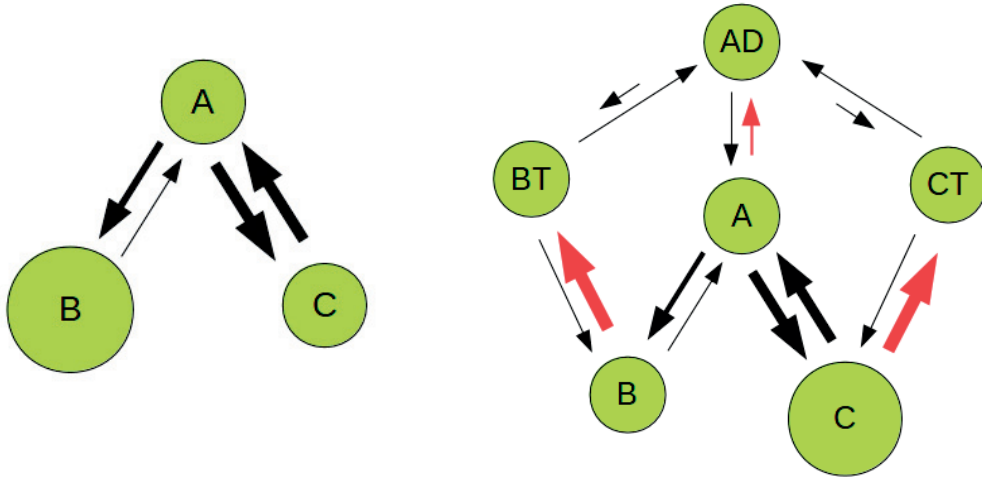
Figure 2.1 – (left) Scheme for the enzyme in equilibrium. The state B and C can both transform reversibly to state A. B is favored in equilibrium, even if C has faster rates of reaction. (right) Scheme for the enzyme out of equilibrium. As soon as the interaction with the nucleotide is introduced, state C becomes more populated than state B.

asymmetries are given by the fact that the rates between A and C are faster than the rates between A and B, even though the equilibrium constant, defined for a generic reaction from i to j as

$$K_{ij} = \frac{k_{ij}}{k_{ji}}, \tag{2.1}$$

where k_{ij} and k_{ji} are the rates from i to j and from j to i respectively, is lower for $A \leftrightarrow C$ than for $A \leftrightarrow B$. The system is kept out of equilibrium by a chemostat which continuously removes ADP and injects new ATP into the system. In figure 2.2 we plot the relative concentration of state C and state B as a function of the concentration of ATP, keeping fixed the concentration of ADP.

As we can see, C is significantly stabilized with respect to B for large values of $[ATP]$. This stabilization is only due to the fact that the transformations between A and C are faster than the ones between A and B, even though the reaction constants would foster a stabilization of the state B instead of C in equilibrium. Therefore, we can conclude that the NESS has the peculiarity of being *non local*, meaning that the kinetics of a component of the biochemical network can have a significant influence on the whole system and considering the values of the equilibrium constants is not sufficient anymore.

2.2   Stationary states in biochemical systems

We can now recall some basic principles on biochemical systems. When we consider a system with M components with probabilities p_i $(i = 1, ..., M)$, the probabilities are
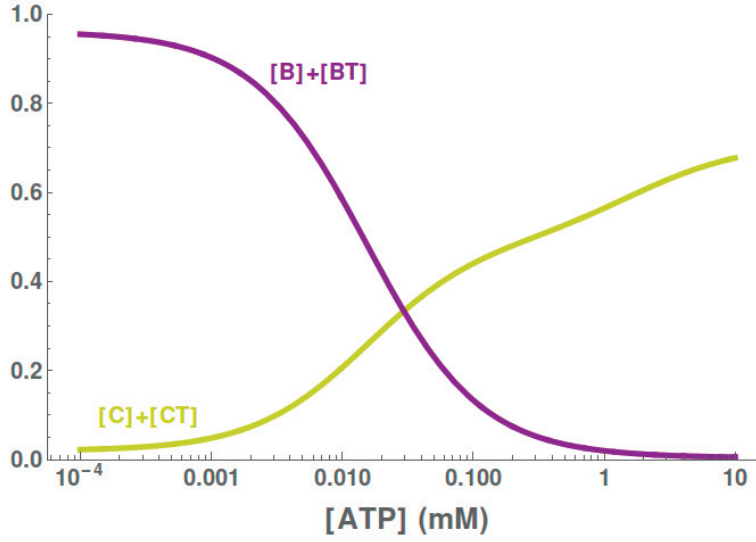
Figure 2.2 – Relative concentration of the states B and C of the enzyme as a function of $[ATP]$ when the concentration of ADP is kept fixed at $0.1\,mM$.

solutions of master equations of the form (see [17, 18])

$$\frac{dp_i}{dt} = \sum_j \left(p_j k_{ji} - p_i k_{ij} \right) , \tag{2.2}$$

where k_{ij} and k_{ji} follow the definition given above. After a certain relaxation time, a stationary state will eventually be reached, characterized by probabilities that are not dependent on time:

$$\frac{dp_i}{dt} = 0 \qquad \text{for all } i \tag{2.3}$$

If the stationary state is an equilibrium state, the detailed balance condition is respected:

$$p_j k_{ji} = p_i k_{ij} . \tag{2.4}$$

We can thus say that the probability of observing the transformation $i \rightarrow j$ is the same as the probability of observing the transformation $j \rightarrow i$ or, in other words, there are no *probability fluxes* for any couple (i, j), where a flux is defined as

$$J_{ij} = p_j k_{ji} - p_i k_{ij} . \tag{2.5}$$

Looking at 2.4, it is quite natural to define a *free energy change* ΔG_{ij} [27] as

$$\Delta G_{ij} = kT \log \left[\frac{p_j k_{ji}}{p_i k_{ij}} \right] , \tag{2.6}$$

so that the free energy change is zero at equilibrium. We will call *cycle* any set of N components that are connected by non zero reaction rates in such a way that they form

a single closed loop. We can reorder the components of the cycle so that component i is connected with component $i + 1$ with the only exception of component N, which is connected with component 1 by construction. The implication of 2.4 is that the product of the reaction rates following a cycle in one direction is equal to the one that is obtain in the opposite direction[1], that is

$$\gamma = \prod_i^N \frac{k_{i,i+1}}{k_{i+1,i}} = 1 \,, \tag{2.7}$$

where we have imposed $k_{N,N+1} = k_{N,1}$. The parameter γ is the *affinity coefficient* and it is related to the free energy of a cycle ξ by the relation

$$\Delta G_\xi \equiv \sum_{i \in \xi} \Delta G_{i,i+1} = kT \ln \gamma_\xi \,. \tag{2.8}$$

As we said, if the system is not in contact with a chemostat, the stationary state is an equilibrium state. Otherwise, after the relaxation time the system is in a non equilibrium stationary state (NESS) with a net flux

$$J_{ij} \equiv p_j k_{ji} - p_i k_{ij} \neq 0 \,. \tag{2.9}$$

We can define the entropy as [17]

$$S = -k \sum_i p_i \ln p_i \,. \tag{2.10}$$

2.3 Entropy flow and entropy production

Let us now assume that the rates do not depend on time and that the master equations are linear. Computing the derivative of this expression with respect to time we obtain

$$\frac{dS}{dt} \;=\; -k \sum_i \frac{dp_i}{dt} (\ln p_i + 1) \;=\; -k \sum_i \frac{dp_i}{dt} \ln p_i \,, \tag{2.11}$$

where we have used the normalization condition on the probabilities:

$$\sum_i p_i = 1 \,. \tag{2.12}$$

---

[1]We can convince ourselves by simply multiplying and dividing for the product of the probabilities of the cycle.

Using the Master equation for p_i and playing a bit with the sums, we have

$$
\begin{aligned}
\frac{dS}{dt} = & -k \sum_{i,j} (k_{ij}p_i - k_{ji}p_j) \ln p_i = \\
& -\frac{k}{2} \left(\sum_{i,j} (k_{ij}p_i - k_{ji}p_j) \ln p_i - \sum_{i,j} (k_{ij}p_i - k_{ji}p_j) \ln p_j \right) = \\
& -\frac{k}{2} \left(\sum_{i,j} (k_{ij}p_i - k_{ji}p_j) \ln \left(\frac{p_i}{p_j} \right) \right),
\end{aligned}
\tag{2.13}
$$

by multiplying and dividing inside the logarithm for $\left(\frac{k_{ij}}{k_{ji}} \right)$, we can separate the derivative of S with respect to time in two contributions:

$$
\frac{dS}{dt} = -\frac{k}{2} \sum_{i,j} (k_{ij}p_i - k_{ji}p_j) \ln \left(\frac{k_{ij}}{k_{ji}} \right) + \frac{k}{2} \sum_{i,j} (k_{ij}p_i - k_{ji}p_j) \ln \left(\frac{p_i k_{ij}}{p_j k_{ji}} \right) \equiv S^f + S^p .
\tag{2.14}
$$

The first term is called *entropy flow* and the second term is the *entropy production*. Note that by construction we always have

$$
S^p \geq 0 .
\tag{2.15}
$$

In the stationary state the two terms must be equal in absolute value and with opposite sign so that

$$
\frac{dS}{dt} = 0 .
\tag{2.16}
$$

In the particular case of equilibrium both terms are zero, since the detailed balance is satisfied for every branch of the reaction network. The entropy production can be used to quantify how far the system is from the equilibrium conditions. Using the definition in (2.9) we can write it as the sum of products between the flux and the difference in free energy [27] for every transformation:

$$
S^p = \frac{1}{2} \sum_{i,j} J_{ij} \Delta G_{ij} .
\tag{2.17}
$$

2.4 Stationary state solution

Let us now go back to the stationary state. There is very synthetic way to write the solution of the stationary state, in the case of linear systems [18]. We represent the system as a graph G whose components are the vertices. Two vertices are connected if between the corresponding components the rate is different from zero.

A *spanning tree* of the graph is a portion of the graph with the following properties:

- all the edges of the tree are edges of the graph, all the vertices of the graphs are also vertices of the tree;

- the tree is connected;

- the tree does not contain closed loops.

For example, from the graph in figure 2.3 (a) it is possible to construct the eight graphs shown in figure 2.3 (b).
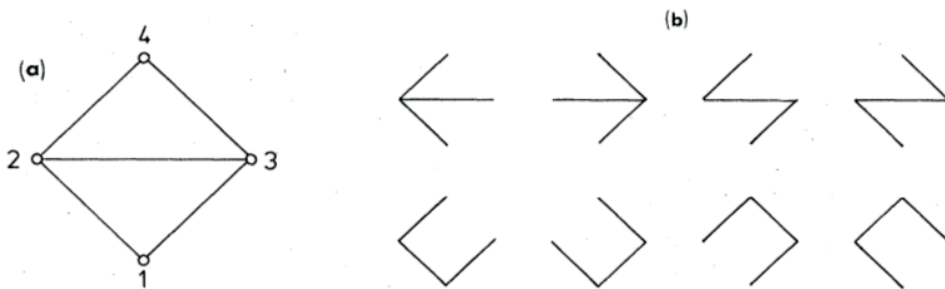
Figure 2.3 – The eight graphs in (b) are the spanning trees of the graph shown in (a). Figure taken from [18].

We will indicated each tree with $T^\mu(G)$. A tree can be i-oriented (univocally, thanks to the fact that the trees are not connected) by taking for all edges the particular direction that points toward i. Following the example above, the spanning trees oriented toward node 1 are shown in figure 2.4.
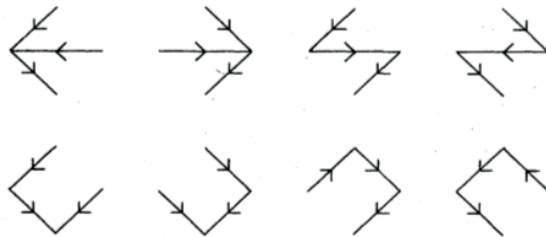
Figure 2.4 – Spanning trees oriented toward node 1 of the graph in figure 2.3 (a). Figure taken from [18].

The i-oriented tree is indicated as $T_i^\mu(G)$. Each edge of a given oriented tree has thus a precise orientation, therefore we can associate to it its corresponding rate. We now give to each oriented tree an algebraic value $A(T_i^\mu(G))$, given by the product of all the transition rates associated to the edges of the tree. For the solution of the system we can use the Kirchhoff's theorem. The probabilities are

$$p_i = \frac{S_i}{S} \ , \tag{2.18}$$

with

$$S_i = \sum_\mu A(T_i^\mu(G)) \tag{2.19}$$

and

$$S = \sum_i S_i \,. \tag{2.20}$$

2.5 The source of the non equilibrium

We have shown that the equilibrium regime is determined by a condition on the rate constants. Thermodynamics impose those constraints on whatever system can be considered isolated, and thus which does not exchange heat or particles with the environment. When this is not the case, it is usually possible to identify those rates that are influenced by external agents. An example that we will frequently use is the tuning of the effective exchange rates, in reactions that involve the hydrolysis of ATP. Systems like molecular chaperones have conformations that can vary according to the nucleotide that is bound. The protein, or the enzyme, can release a nucleotide and bind a new one. However, in most cases the concentration of ATP and ADP in physiological is large enough that the protein is almost never found in the absence of nucleotide (often referred to as the *apo-state*). Therefore, it is possible to define effective rate constants that take into account, at the same time, both the release and the rebinding. If the rates of binding and unbinding of ATP and ADP are k_{T+}, k_{T-}, k_{D+} and k_{D-} respectively, the effective rate constants of exchange are

$$k_{TD}^{ex} = k_{T-} \frac{k_{D+}}{k_{D+} + \frac{[ATP]}{[ADP]}k_{T+}} \,,$$

$$k_{DT}^{ex} = k_{D-} \frac{\frac{[ATP]}{[ADP]} k_{T+}}{k_{D+} + \frac{[ATP]}{[ADP]}k_{T+}} \,. \tag{2.21}$$

In this way the dependency on the actual energy source, that is the ratio between ATP and ADP, can be written explicitly. All the rate constants must be chosen in such a way that, when the ratio is equal to the ratio that we would have in the absence of a regenerating system for ATP, the constraint of formula 2.7 is satisfied.

2.6 Michaelis-Menten kinetics

We now consider a very simple and well-known example of enzyme kinetics [17]. Suppose there is a substrate S that can bind with an enzyme E and be converted into product P:

$$E + S \rightleftarrows ES \rightarrow E + P \,. \tag{2.22}$$

We assume that k_f is the binding rate, k_r is the unbinding rate and k_c is the rate for the reaction that converts the substrate into the product. In the stationary state, since the total concentration of enzyme $[E_0] = [E] + [ES]$ does not change in time, we must have

$$[ES] = [E][S]\frac{k_f}{k_r}\,. \tag{2.23}$$

Therefore we can write

$$[E] = \frac{[E_0]}{1 + \frac{k_f}{k_r}[S]}\,. \tag{2.24}$$

The rate of the production of P is given by

$$\frac{d[P]}{dt} = [ES]k_c\,. \tag{2.25}$$

Using equation 2.23 and 2.24 we obtain

$$\frac{d[P]}{dt} = [S]\frac{k_c[E_0]}{[S] + \frac{k_r}{k_f}} \equiv \frac{[S]V_{max}}{[S] + K_d}\,, \tag{2.26}$$

where we have defined the maximum reaction velocity V_{max} and the dissociation constant for the enzyme-substrate complex K_d. We can see that in the case the concentration of the substrate is much lower than the dissociation constant, there is a linear dependency of the rate of formation of the product from the concentration $[S]$. Formula 2.26 is called *Michaelis-Menten equation* [17].

3 Molecular Chaperones

Introduction

The vast majority of works involving the problem of protein folding usually give for granted the so called *Anfinsen's dogma* [11], that is a postulate according to which a protein sequence determines univocally the tertiary structure of the native state, a conformation that can be reached spontaneously, that is unique and kinetically accessible. However, the free energy landscape relative to the tertiary structure of a protein is rugged and proteins are prone to be kinetically trapped in non functional states, corresponding to local minima in the free energy. A cell is a very crowded environment and in physiological conditions it is often not possible for the protein to spontaneously cross the kinetic barrier in relevant time scales. Moreover, in the case of extraordinary conditions such as thermal or chemical stresses, it is not even assured that the functional, or native state is actually the state of the absolute minimum in the free energy. Proteins that are *misfolded*, namely that are not in their functional state, apart from the obvious issue of not being able to perform their biological task, could also form cytotoxic aggregates [28] responsible for many diseases, such as Alzheimer and Parkinson [29]. Thankfully, the folding is usually supervised by a particular class of specialized proteins called *chaperones*, that are defined as proteins that provoke a conformational change in a substrate without taking part in its final structure [1]. Among the variety of chaperone families, many of them are *heat-shock proteins* (Hsps) because for historical reasons they have been initially associated to the factor that strongly enhances their expression, which is the sudden increase in temperature (heat shock) responsible for the unfolding [30]. Even though heat shock proteins have been under study for more than three decades and there has been an overwhelming improvement in the understanding of the field, there are still many questions that at present remain unanswered. For example, it is unclear the precise mechanism that drives the substrate to the folded state: do chaperones actively catalyze protein unfolding or they are rather passive holding machines that are able to prevent aggregation during the folding process? Moreover, it is well known that ATP is consumed during the process, but it is a matter of debate how the energy of ATP hydrolysis is actually used. Is it possible for molecular chaperones to stabilize client proteins that would not be stable in equilibrium? In this work,

we will try to answer some of these questions with the help of non equilibrium statistical physics.

3.1 Hsp70

Hsp70 is a very abundant chaperone family that is involved in protein folding together with other very important cellular functions, as protein translocation across cell membranes [16, 31] and protein degradation [32]. In the Hsp70 system the heat shock protein is usually helped by other proteins called *cochaperones*: a J domain protein (JDP), which carries the substrate to the complex and stimulates the hydrolysis [33, 34], and a nucleotide exchange factor [35] (NEF), responsible for the release of ADP after the hydrolysis and the subsequent binding of a new molecule of ATP. Despite the remarkable versatility of Hsp70, its sequence is pretty conserved, mostly because the specificity is usually guaranteed by the cochaperones [33, 35].

Hsp70 is formed by two domains [36]: a nucleotide binding domain (NBD, represented in blue in figure [37]), of approximately 43 kDa, which binds and releases ATP and ADP, and a substrate binding domain (SBD), of approximately 27 kDa, which interacts with the substrate protein.
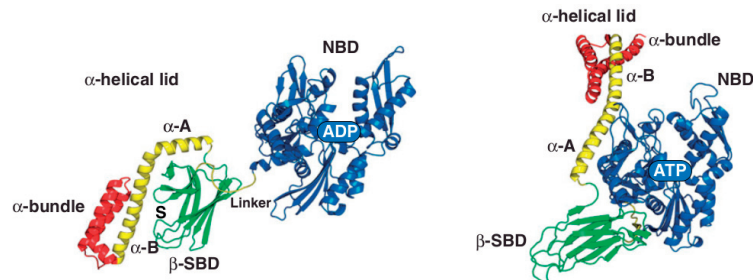
Figure 3.1 – The structure of Hsp70 when it is bound with ADP (left) and with ATP (right). The ADP state is characterized by a closed conformation of the SBD, implying lower binding and unbinding rates for the substrate. Figure adapted from [37].

The two domains are connected by a short linker of about 5 amino acids. The SBD is in turn composed by different domains (green, yellow and red in the figure) with a lid-like structure, which can assume either a closed or an open conformation, depending on whether the NBD is bound with ADP or ATP, respectively [38]. The rates of binding and unbinding of Hsp70 with the substrate are much larger when the lids are open [36]. The switch between open and closed conformation is regulated by the hydrolysis of ATP, which is catalysed by the substrate protein [21]. With this mechanism, the effective affinity of Hsp70 for the substrate is significantly increased compared to the binding without nucleotides, being the chaperone able to bind the substrate protein when the lids are open and then, after hydrolysis, to close the lids, clamping the substrate and lowering the rate of unbinding.

As we said, in equilibrium conditions ADP is much more stable than ATP. However, the

cell is a system that is strongly out of equilibrium, and in physiological conditions we can have a ratio $\alpha = \frac{[ATP]}{[ADP]} \in (1,10)$. Previously [21], it was shown that the switching between a configuration with large binding rates and one with small binding rates can lead to an *ultra-affinity*, that is an effective affinity between the chaperone and the substrate that is much larger than the affinity for the ATP or the ADP state of Hsp70 and this phenomenon is possible only when the system is out of equilibrium. This result would suggest that the binding is indeed the step in the reaction cycle that is most affected by the energy consumption and α is thus a parameter that quantifies how far the system is from equilibrium. This section will be divided in two. In the first part we will consider a simple model with a single binding site and derive some constraints on the rate constants necessary for the catalysis of the unfolding. Afterwards, we will work on a model with several binding sites, focusing on the particular case of rhodanese, motivated by some experimental data [24]. Before starting we want to give an intuitive explanation on the mechanism used by Hsp70 for the unfolding of the substrate.

3.1.1 Entropic pulling

Here we want to give a qualitative description of the mechanism that Hsp70 uses for the translocation and the unfolding of substrate proteins. We consider the translocation as an example because it allows a simpler representation, but a similar approach can be used for the unfolding[16, 15]. The idea is that thermal fluctuations induce the substrate to increase its freedom by reducing the effects of the excluded volume. Let us model the translocating substrate as a Gaussian chain and consider the case in which the chain is constrained to stay above a fixed rigid surface [1]. In the last chapter of the thesis the shape of a random chain will be discussed, here we just anticipate that the end-to-end vector, that is the distance between the first and the last monomer of the chain, is normally distributed when there is no interaction between the monomers. However, it is possible to show that in the presence of a constraint that prohibit the monomers to be in a certain region, the form of the distribution is different. In our example the monomers are not allowed to be in the region in which $z < 0$. In this case, the distribution of the end to end vector $\mathbf{R}_e$ becomes

$$P(\mathbf{R}_e) = \omega R_z \exp\left[-\frac{3}{2}\frac{R_e^2}{Nb^2}\right],\tag{3.1}$$

where N is the number of monomers in the chain that have already crossed the membrane, ω is a constant and R_z is the z-component of $\mathbf{R}_e$. If a chain with N monomers has A_0 possible conformations in free space, then in the presence of the membrane it will have a number A of possible conformations where

$$A = A_0 N \int_{-\infty}^{+\infty} dz \int_{-\infty}^{+\infty} dy \int_{0}^{+\infty} dz\, P(\mathbf{R}_e).\tag{3.2}$$

---

[1]This calculation can be found in [16]

Now we assume that at the end of the chain an Hsp70 is bound. Therefore, when we integrate over z we must consider the volume that is forbidden by the chaperone (see figure 3.2):

$$A_{70} = A_0 N \int_{-\infty}^{+\infty} dz \int_{-\infty}^{+\infty} dy \int_{R_{70}}^{+\infty} dz \, P(\mathbf{R}_e) = A \exp\left[-\frac{3R_{70}^2}{2Nb^2}\right] . \tag{3.3}$$

As a consequence, the free energy gained by the addition of one chaperone is

$$\Delta F = -k_b T \left(\ln\left[A_{70}\right] - \ln\left[A\right]\right) = \frac{3R_{70}^2}{2Nb^2} , \tag{3.4}$$

and we can see that the increment is inversely proportional to the number of monomers N. We can conclude that to minimize ΔF the chain is enhanced to increase the number of monomers that cross the membrane.
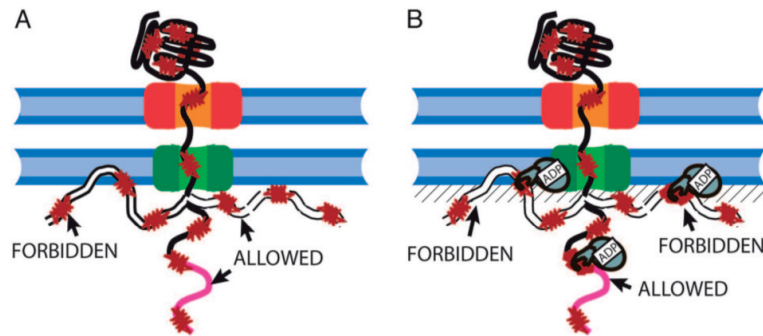
Figure 3.2 – Schematic representation of the excluded volume effect of Hsp70. The number of conformations that are forbidden is lower in the presence (B) with respect to the absence (A) of Hsp70. This picture is taken from [16].

This mechanism is called *entropic pulling*. Even though we presented it for the translocation, the same mechanism can be applied to the chaperone-mediated unfolding and disaggregation. The role of the surface is played in this case by the misfolded protein itself or by the aggregate, that can be modeled as rigid spheres. Hsp70 can specifically bind to exposed hydrophobic regions in the substrate sequence that are not hidden in the inner part of the substrate/aggregate. In this way, the excluded volume interaction between the chaperone and the substrate or the aggregate stimulates an increase in the number of monomers that form the portion of the chain that is exposed and thus that is not misfolded anymore.

3.1.2 Rate model for refolding

In this section we want to derive the constraints that the rate constants must respect so that the chaperones can properly do their work as unfoldase. In order to model a process of refolding catalysed by Hsp70, we had to make some assumptions. We assumed that each protein can be found in one of four conformational states, that are coarse grained ensembles of all the possible configurations (a similar approach has been used, for example,

in [39]): the protein can be native (N), partially folded (X), misfolded (M) or unfolded (U) (see figure 3.3).
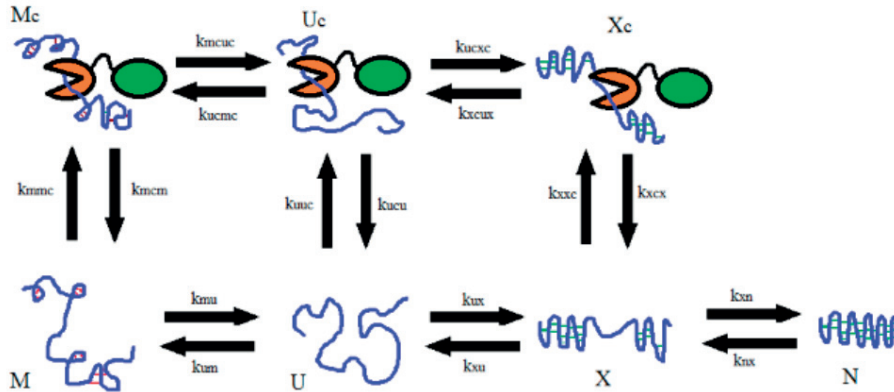
Figure 3.3 – Initial kinetic cycle, with four different conformational states for the protein. Under some assumptions it can be reduced to a simpler cycle, without the unfolded state (the second from the left).

The native state N is a state that is completely folded and we assume, for this reason, that almost all the hydrophobic residues are hidden in the inner regions and it is impossible for the chaperone to bind the protein [40]. The partially folded state is an intermediate step in the process of refolding. In proteins in state X a large number of amino acids are correctly arranged, but some of them are not and, as a consequence, there is still a certain number of exposed hydrophobic residues to which the chaperone can bind. The misfolded state is a non-functional state with non native hydrogen bonds and a certain number of exposed hydrophobic residues. It corresponds to a local minimum in the free energy and it is separated by a high energy barrier from the native state. The unfolded state has few contacts among the amino acids and the majority of the hydrophobic residues are exposed. We also assumed that the unfolded state is very unstable and thus the rates of misfolding and refolding from state U are large. We can show that, under these hypotheses, the unfolded state can be left out of the cycle and the state M can be connected directly with the state X, but with new effective rates that are a function of the rates from U to X and to M.

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

$$[\dot{M}] = [U]k_{um} + [M_c]k_{mcm} - [M](k_{mu} + k_{mmc})$$
$$[\dot{M_c}] = [M]k_{mmc} + [U_c]k_{ucmc} - [M_c](k_{mcm} + k_{mcuc})$$
$$[\dot{U}] = [M]k_{mu} + [U_c]k_{ucu} + [X]k_{xu} - [U](k_{um} + k_{uuc} + k_{ux})$$
$$[\dot{U_c}] = [U]k_{uuc} + [M_c]k_{mcuc} + [X_c]k_{xcuc} - [U_c](k_{ucu} + k_{ucmc} + k_{ucxc}) \tag{3.5}$$
$$[\dot{X}] = [U]k_{ux} + [X_c]k_{xcx} + [N]k_{nx} - [X](k_{xu} + k_{xxc} + k_{xn})$$
$$[\dot{X_c}] = [X]k_{xxc} + [U_c]k_{ucxc} - [X_c](k_{xcx} + k_{xcuc})$$
$$[\dot{N}] = [X]k_{xn} - [N]k_{nx} \; .$$

Note that we are assuming that the concentration of free chaperones $[c]$ is large enough that it can be considered constant. Therefore, the rates that we have written for the binding reactions must be interpreted as the product between the actual binding rates and the concentration of free chaperones, for example $k_{mmc} = [c]k^0_{mmc}$ where k^0_{mmc} is the binding rate and k_{mmc} has dimension s^{-1}. The total concentration of the substrate $[S_t]$ is fixed:

$$[M] + [M_c] + [U] + [U_c] + [X] + [X_c] + [N] = [S_t] \, . \tag{3.6}$$

The steady state is obtained by imposing on all the concentrations to be constant in time:

$$[\dot{M}] = [\dot{M_c}] = [\dot{U}] = [\dot{U_c}] = [\dot{X}] = [\dot{X_c}] = [\dot{N}] = 0 \; . \tag{3.7}$$

If we now assume that the rates k_{um}, k_{ux}, k_{ucmc} and k_{ucxc} are much larger than all the other rates, then we can pass from a cycle of the form of figure 3.3 to a cycle of the form of figure 3.4.
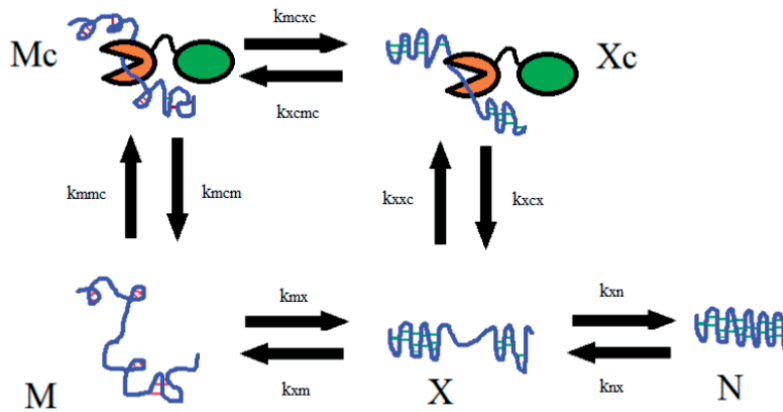
Figure 3.4 – Kinetic cycle used in this work, after the simplification.

Intuitively, a simplification of this kind is possible because the proteins spend few time in the unfolded state and once they unfold they rapidly refold or misfold again. The

concentrations of the new cycle are solutions of the following system:

$$
\begin{aligned}
0 &= [X]k_{xm} + [M_c]k_{mcm} - [M](k_{mx} + k_{mmc}) \\
0 &= [M]k_{mmc} + [X_c]k_{xcmc} - [M_c](k_{mcm} + k_{mcxc}) \\
0 &= [M]k_{mx} + [X_c]k_{xcx} + [N]k_{nx} - [X](k_{xm} + k_{xxc} + k_{xn}) \\
0 &= [X]k_{xxc} + [M_c]k_{mcxc} - [X_c](k_{xcx} + k_{xcmc}) \\
0 &= [X]k_{xn} - [N]k_{nx} \ ,
\end{aligned}
$$
(3.8)

where we have defined

$$
\begin{aligned}
k_{mx} &= \frac{k_{mu}k_{ux}}{k_{ux} + k_{um}} \\
k_{xm} &= \frac{k_{xu}k_{um}}{k_{ux} + k_{um}} \\
k_{mcxc} &= \frac{k_{mcuc}k_{ucxc}}{k_{ucxc} + k_{ucmc}} \\
k_{xcmc} &= \frac{k_{ucmc}k_{xcuc}}{k_{xcuc} + k_{mcuc}} \ .
\end{aligned}
$$
(3.9)

Once the simplification on state U has been done, the remaining states are connected in the kinetic cycle shown in figure 3.4.

When a misfolded protein refolds, it necessarily passes through the intermediate state X before reaching the native state N. Hsp70s are able to catalyse the reaction $M \leftrightarrow X$ via the path $M \leftrightarrow M_c \leftrightarrow X_c \leftrightarrow X$, but for a complete refolding they must leave the client protein and the reaction $X \leftrightarrow N$ is always spontaneous.

As we said, Hsp70 is always helped by the cochaperones. Here, the effects of the cochaperones are implicitly taken into account via a modulation of the rate constants. The hydrolysis of ATP is stimulated by the combined interaction of Hsp70 with JDP and the client protein: when they are both bound with Hsp70, the hydrolysis is much faster than in the presence of either the substrate or the J protein alone: while JDP and substrate can stimulate the hydrolysis of 2-5 fold, when they both interact with Hsp70 at the same time the hydrolysis can be increased of up to three orders of magnitude. [33, 41, 42, 43]. At a first sight, this is slightly counter-intuitive. As we will better explain later, our guess is that the chaperone in the ATP state can have two different conformations and only in one of them it stimulates the hydrolysis. The J-protein and the substrate are responsible for the stabilization of that conformation with respect to the other. In this way, it is possible to explain the particularly large enhancement of the ATP hydrolysis.

A consequence of this phenomenon is that the catalysis of the hydrolysis must depend on the conformation of the substrate. Indeed, if the interaction with JDP depends on the conformation of the client protein, as we can expect, then it is reasonable to assume that also the hydrolysis of ATP is stimulated at a different amount by proteins that are in state M or in state X. In the next section we will discuss how it is possible to write an effective

binding affinity that takes into account the hydrolysis of ATP and the interaction between chaperone, substrate and cochaperone.

The mechanism by which the switch between the different conformational states of Hsp70 influences its interaction with the client protein will be discussed later and here we collect in one state all the configurations of Hsp70. The rates of the hydrolysis of ATP are included in the effective rates of binding with Hsp70, k_{mmc} and k_{xxc}. Therefore, with M_c and X_c we indicate the misfolded and partially folded proteins when they are bound with Hsp70, independently on the conformation of Hsp70 and the nucleotide that is bound to it.

We consider a system in which the concentrations are low, so that aggregation can be neglected. Intuitively, if the rates toward X following a path that include the binding are faster, the concentration of native proteins relative to the concentration of misfolded proteins must be larger with Hsp70, compared to a system in the absence of chaperones. More formally, we define the affinity coefficient for this particular system:

$$\gamma = \frac{k_{xm}\ k_{mmc}\ k_{mcxc}\ k_{xcx}}{k_{mx}\ k_{xxc}\ k_{xcmc}\ k_{mcm}}\ . \tag{3.10}$$

The concentrations in the stationary state can be calculated by solving 3.8. The result can be used to obtain the ratio between the concentrations of the two states with respect to the same ratio in the case of spontaneous refolding (without Hsp70):

$$\frac{[N]_{ch}}{[M]_{ch}} = \frac{[N]_0}{[M]_0}(1 + C(\gamma - 1))\ , \tag{3.11}$$

where

$$C = \frac{k_{xxc}k_{xcmc}k_{mcm}}{k_{xxc}k_{xcmc}k_{mcm} + k_{xcmc}k_{mcm}k_{xm} + k_{xm}k_{xcx}(k_{mcm} + k_{mcxc})}\ , \tag{3.12}$$

is a positive constant, $[N]_0$ and $[M]_0$ are the concentrations of native protein in the absence of chaperones and $[N]_{ch}$ and $[M]_{ch}$ are the corresponding concentrations for a system with chaperones. We can see that, as soon as $\gamma > 1$, the native state is stabilized with respect to the misfolded state in the presence of Hsp70:

$$\frac{[N]_{ch}}{[M]_{ch}} > \frac{[N]_0}{[M]_0}\ . \tag{3.13}$$

However, the fact that the relative concentration is increased does not mean that the absolute concentration $[N]$ is also increased. For the general case, the necessary condition on the rate constants in order to have $[N]_{ch} > [N]_0$ can be written as

$$\gamma > 1 + \gamma \left[\frac{k_{mx}}{k_{xcx}} + \frac{k_{mx}}{k_{mcxc}} \right] + k_{mx} \left[\frac{1}{k_{xcmc}} + \frac{1}{k_{mcm}} \right]\ , \tag{3.14}$$

and we thus confirm the fact that it is necessary to have $\gamma > 1$. In [21] it has been shown that the role of the hydrolysis of ATP is to tune the effective affinity of Hsp70 for the

substrate. If we assume that the energy of the hydrolysis is used only to perform this task, then the condition $\gamma > 1$ implies that, for an improvement of the refolding in the steady state, it is necessary to increase the affinity for proteins in state M of a larger amount than for proteins in state X, because that would be the only way to increase γ, that depends on the ratio k_{mmc}/k_{xxc}. Note that $\gamma = 1$ is not necessarily a signal of equilibrium, but it rather implies that, if there is hydrolysis, the affinity is increased of the same amount for the reaction $M \leftrightarrow M_c$ and $X \leftrightarrow X_c$. Playing a bit with the algebra and using the definition of γ, we can rewrite the inequality as

$$\boxed{\gamma\left[\tau_{mx} - \tau_{mcx}\right] > \tau_{xcmc} + \tau_{mcm} + \tau_{mx}} \tag{3.15}$$

where

$$\tau_{mcx} = \tau_{mcxc}\left[1 + \frac{\tau_{xcx}}{\tau_{xcmc}}\right] + \tau_{xcx} \tag{3.16}$$

is the time needed for the reaction $M_c \to X_c \to X$ to occur and $\tau_{ij} = k_{ij}^{-1}$. Since γ is always positive, the condition can be satisfied only if $\tau_{mx} > \tau_{mcx}$, namely if the refolding in the presence of Hsp70 is faster than the spontaneous refolding, as we should expect. If this condition is not satisfied, the concentration $[N]_{ch}$ will not be larger than $[N]_0$ in any time interval, even during the transient state.

In figure 3.5 we show the concentration of native proteins $[N]$ normalized by the total concentration $[S_T]$ in three different conditions.
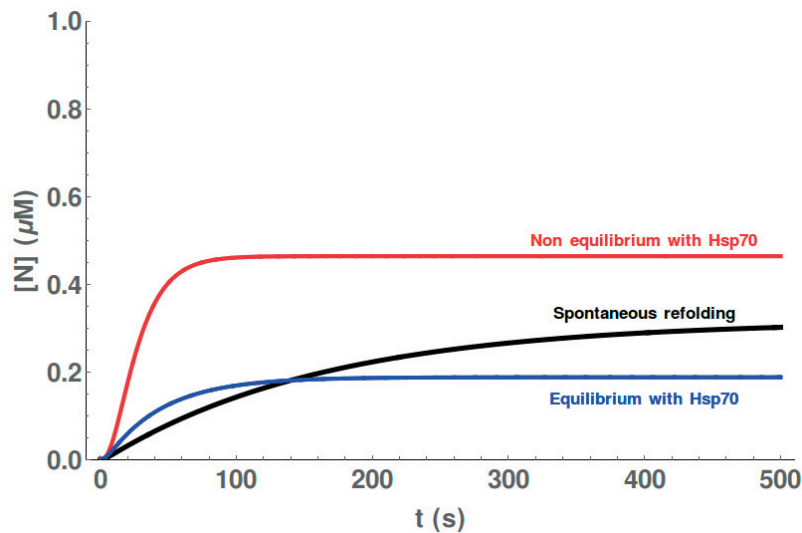
Figure 3.5 – Concentration of native proteins as a function of time in the absence of Hsp70 (black), in the presence of Hsp70 but in equilibrium conditions (blue) and in the presence of Hsp70 in non equilibrium conditions (red).

We can see that the presence of Hsp70 in non equilibrium conditions (red curve) not only stimulates a faster refolding, but also leads to a larger concentration of native proteins in the stationary state with respect to the case in which the chaperones are absent (blue

curve). On the contrary, if $\gamma = 1$ (black curve), even though during the transient state a larger concentration is possible, the yield of native protein in the stationary state is always larger without Hsp70.

Even though the hydrolysis of ATP is mainly used to increase the effective affinity for the substrate protein, it is worth to mention that the affinity must not be too large, because Hsp70 needs to release the substrate protein in order to allow it to spontaneously refold. In figure 3.6 we plot the relative concentration of native protein in the stationary state as a function of the concentration of Hsp70 in the system. As we can see, if the concentration of chaperones is too large, the yield of native proteins becomes lower than in the case of spontaneous refolding (orange line).
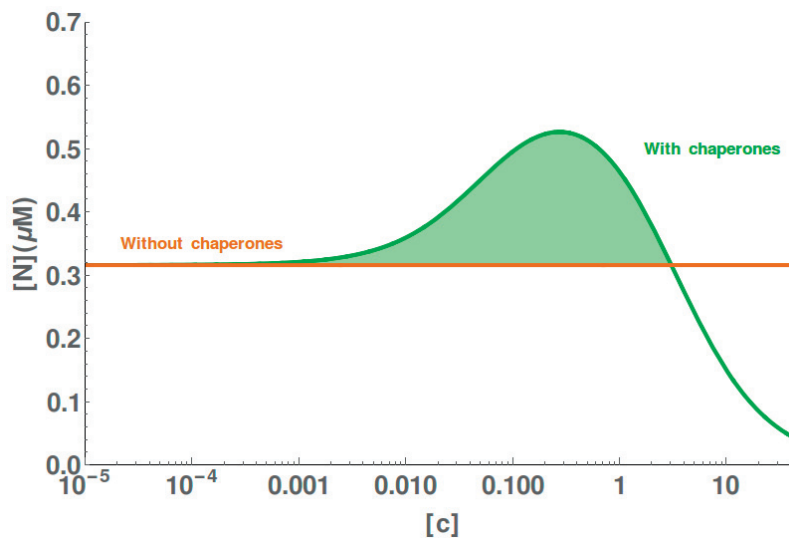
Figure 3.6 – Concentration of native proteins as a function of the concentration of Hsp70 (green). The light green region represents the improvement in the yield of native proteins due to the presence of Hsp70.

3.1.3 Role of the cochaperones in the binding affinity

Since in our model we have not taken into account explicitly the conformation of the chaperone, as well as the nucleotide that is bound to it, we want to give some details on the dependence of the binding affinity from the hydrolysis of ATP. In order to do so, there is the need of a brief discussion about J-domain proteins.

Hsp70 never works alone: even though sometimes it can function without the nucleotide exchange factor, up to now it has never been shown to work in the absence of a J-domain protein (JDP) as cochaperone [33]. Hsp70 is a very conserved protein, but at the same time it has to interact with a plethora of very different client proteins [33]. JDP is the protein that mediates this interaction with the substrate [34], having a much larger diversity in its protein binding domains. On the other hand, the J-domain, that is used to catalyze the

ATPase activity of Hsp70, is quite conserved among different JDPs. The mechanism by which the J-domain stimulates the hydrolysis of ATP is still a matter of debate.

We want to study in more detail how JDP influences the binding affinity between the chaperone and the substrate. JDP has two binding sites, located in the C-terminal domain of its two monomers. We assume that the client protein binds separately to each domain. Therefore, we can define two states, SJ_1 and SJ_2, the first one representing the configuration with only one monomer of JDP bound to the client protein, the second one the configuration with both monomers attached to it. The binding reaction is

$$S + J \rightleftarrows SJ_1 \rightleftarrows SJ_2 \, . \tag{3.17}$$

In the steady state, we have

$$[SJ_1] = n \frac{k_{on}}{k_{off}} [S][J] \, , \tag{3.18}$$

and

$$[SJ_2] = \frac{(n-1)\hat{k}_{on}}{k_{off}} [SJ_1] = \frac{n(n-1)\hat{k}_{on}k_{on}}{k_{off}^2} [S][J] \, . \tag{3.19}$$

where k_{off} is the rate of unbinding (the same for the two domains), n is the number of available binding sites, k_{on} is the binding rate at a precise site for the first domain and $\hat{k}_{on}$ is the binding rate for the second, that is larger because, after the first monomer binds, JDP is localized in the neighborhood of the client protein. If we define the total concentration of client proteins bound with JDP as

$$[SJ] = [SJ_1] + [SJ_2] \, , \tag{3.20}$$

we end up with the following relation:

$$[SJ] = n \frac{k_{on}}{k_{off}} \left(1 + \frac{(n-1)\hat{k}_{on}}{k_{off}} \right) [S][J] \, . \tag{3.21}$$

Hsp70 can interact with the substrate either when it is free or when it is bound with JDP, with approximately the same binding rate. However, the hydrolysis is stimulated at different amounts by SJ and S. We will call H^T and H^D the Hsp70 in the ATP and ADP state respectively. Let us assume that the Hsp70 can bind with SJ only when it is in the ATP state. Then we have

$$[SJH^T] = \frac{k_{jk}^0}{k_{kj}} [SJ][H^T] \equiv \frac{1}{K_d} [SJ][H^T] \, , \tag{3.22}$$

where k_{jk}^0 and k_{kj} are the rate constants for the reaction $SJ + H^T \rightleftarrows SJH^T$ and $K_d = \frac{k_{kj}}{k_{jk}^0}$ the dissociation constant of SJH^T. It is considered to be approximately equal to the dissociation constant of the bound state without JDP, indicated as SH^T. Since the

reaction of synthesis is negligible, the interchange between the ATP and the ADP states are determined by the rates of hydrolysis and of exchange:

$$[SJH^D] = \frac{k_h^j}{k_{ex}}[SJH^T] = \frac{k_h^j}{k_{ex}\,K_d}[SJ][H^T]\ , \tag{3.23}$$

where k_{ex} is the rate of exchange from ADP to ATP and k_h^j is the rate of hydrolysis in the presence of SJ. Moreover, we are assuming that the hydrolysis is the leading contribution in the reaction $H^T \to H^D$, being much faster than the exchange from ATP to ADP. The total concentration of the complex formed by the three proteins is given by

$$[SJK] = [SJH^T] + [SJH^D] = \frac{1}{K_d}\left(1 + \frac{k_h^j}{k_{ex}}\right)[SJ][H^T]\ . \tag{3.24}$$

If we now plug 3.21 into 3.24, we obtain

$$[SJK] = \frac{k_{on}}{k_{off}}\frac{n}{K_d}\left(1 + \frac{k_h^j}{k_{ex}}\right)\left(1 + \frac{(n-1)\hat{k}_{on}}{k_{off}}\right)[J][S][H^T] \equiv \frac{1}{K_d^e}[S][H^T]\ . \tag{3.25}$$

where we have defined the effective dissociation constant as

$$K_d^e = \left[\frac{k_{on}}{k_{off}}\frac{n}{K_d}\left(1 + \frac{k_h^j}{k_{ex}}\right)\left(1 + \frac{(n-1)\hat{k}_{on}}{k_{off}}\right)[J]\right]^{-1}\ . \tag{3.26}$$

We can see that the dissociation constant is decreased by an increase in either the concentration of JDP or in the rate of hydrolysis.

As we said, the binding rate of the chaperone with the free substrate is not very different from the one in the presence of JDP. However, in the absence of JDP the hydrolysis of ATP is much slower[33]. Therefore, the steady state is given by the following relation:

$$[SH^T] = \frac{1}{K_d}\left(1 + \frac{k_h^s}{k_{ex}}\right)[S][H^T] \simeq \frac{1}{K_d}[S][H^T]\ , \tag{3.27}$$

where k_h^s is the rate of hydrolysis in the absence of the J-domain. The last equality in 3.27 follows from the fact that $k_h^s \ll k_{ex}$. The total concentration of the bound state $S^{tot}K$ is the sum of the two contributions:

$$[S^{tot}K] = [SJK] + [SH^T] = \left(\frac{1}{K_d^e} + \frac{1}{K_d}\right)[S][H^T]\ . \tag{3.28}$$

3.1.4 Binding with JDP and hydrolysis

We have just seen how the hydrolysis can change the effective binding affinity between the chaperone and the substrate. The fact that JDP stimulates the hydrolysis in cooperation with the substrate, that is supported by experimental observation [34], up to now in this

work has just been given for granted. We now want to implement a model that can explain this behavior.

We assume that the chaperone can be found in one of two conformational states, A and B, with a free energy difference ΔG_{AB}. Let us also assume that the chaperone is able to stimulate the ATP hydrolysis only when it is in the state with larger free energy (say A) whereas it does not work as ATPase when it is in the other state (B). The state that stimulates the hydrolysis has a large affinity for the J-domain protein, whereas the other state barely interact with it. We will call $k_{h,max}$ the rate of hydrolysis stimulated by Hsp70. In the absence of J and S, the effective rate of hydrolysis is equal to $k_{h,max}$ multiplied by the probability that Hsp70 is in the higher energy state:

$$k_{h,eff} = \frac{\exp\left(-\beta \Delta G\right)}{1 + \exp\left(-\beta \Delta G\right)} k_{h,max} \; . \tag{3.29}$$

For the free Hsp70, the difference in free energy is about $7\,kT$ and thus

$$k_{h,eff} \simeq 10^{-3} \, k_{h,max} \, . \tag{3.30}$$

The binding of substrate or JDP leads to a state characterized by a lower difference in free energy with respect to state A. Suppose that the chaperone can bind with the cochaperone or with the substrate and when they are both bound they can also interact with one another. The expression of the effective rate must be modified:

$$k_{h,eff} = \frac{\exp\left(-\beta \Delta G\right)\left(1 + \frac{[J]}{K_A^{jc}} + \frac{[S]}{K_A^{sc}} + \frac{[J]}{K_A^{jc}} \frac{[S]}{K_A^{sc}} \frac{\rho_{loc}}{K^{js}}\right)}{\left(1 + \frac{[J]}{K_B^{jc}} + \frac{[S]}{K_B^{sc}} + \frac{[J]}{K_B^{jc}} \frac{[S]}{K_B^{sc}} \frac{\rho_{loc}}{K^{js}}\right) + \exp\left(-\beta \Delta G\right)\left(1 + \frac{[J]}{K_A^{jc}} + \frac{[S]}{K_A^{sc}} + \frac{[J]}{K_A^{jc}} \frac{[S]}{K_A^{sc}} \frac{\rho_{loc}}{K^{js}}\right)} k_{h,max} \; , \tag{3.31}$$

where K_I^{uv} is the dissociation constant of state UV (where U and V can be JDP (J), substrate (S) or Hsp70 (C)) when Hsp70 is in conformation I (either A or B). In principle, we should add the terms corresponding to the other possible three body interactions, that are when protein and JDP interact before they are both bound to Hsp70. However, they are not particularly relevant for the effect that we are going to show, so we will omit them. ρ_{loc} is the *local* concentration, the number of binding sites of S that are available to the JDP when they are both bound with Hsp70, over the volume that can be reached by JDP without unbinding from Hsp70 (same order of magnitude of its radius of gyration). If we write 3.31 in the following way:

$$k_{h,eff} = \exp\left(-\beta \Delta G\right) \frac{D}{F + \exp\left(-\beta \Delta G\right) D} k_{h,max} \; , \tag{3.32}$$

with

$$D = \left(1 + \frac{[J]}{K_A^{jc}}\right)\left(1 + \frac{[S]}{K_A^{sc}}\right) + \frac{[J]}{K_A^{jk}} \frac{[S]}{K_A^{sc}}\left(\frac{\rho_{loc}}{K^{js}} - 1\right) \; , \tag{3.33}$$

and

$$F = \left(1 + \frac{[J]}{K_B^{jc}}\right)\left(1 + \frac{[S]}{K_B^{sc}}\right) + \frac{[J]}{K_B^{jk}}\frac{[S]}{K_B^{sc}}\left(\frac{\rho_{loc}}{K^{js}} - 1\right) \; , \tag{3.34}$$

we can see that D and F are the sum of a contribution given by the product of the terms that we would have with either only J or only S and a contribution with the coupling between J and S. It is now time to make some estimates[2]. The local concentration ρ_{loc} can be around $1 - 10\,mM$. Substrate and JDP are usually in concentrations of the order of $1\,\mu M$. The dissociation constants for the binding between the substrate and the chaperone is also of the order of the micromolar, while for the interaction between S and JDP we can assume $K^{js} \sim 10^{-2}\mu M$. We can also assume $K_A^{jc} \sim 1\,\mu M$, whereas in first approximation there is no binding between JDP and Hsp70 in state B. Considering that we have

$$\exp\left[-\Delta G\right] \sim 10^{-3} \, , \tag{3.35}$$

we can see that, when there are either JDP alone or the substrate alone in the system, only F is relevant in the denominator of formula 3.32. Therefore in the absence of cooperation between JDP and S the rate of hydrolysis can be improved of a factor

$$z = 1 + \frac{[J]}{K_A^{jc}} \, , \tag{3.36}$$

which is the order of unity. According to the observation, the addition of JDP alone can lead to an increase in the rate of hydrolysis of up to 5-fold [36, 43]. However, in the presence of both the substrate and JDP the second term in the denominator of 3.32 is not negligible anymore. We have indeed

$$\exp\left(-\beta\Delta G\right)D \sim 10^2 - 10^3 \, . \tag{3.37}$$

This is in agreement with the general opinion that the rate of hydrolysis is increased by up to 10^3 times in the presence of both the substrate and the cochaperone.

3.1.5   The expansion of rhodanese

Until now we have considered a model with a single binding site. However, in many cases the substrate is bound at the same time by more than one chaperone and in general a substrate protein can be bound in more than one region along its amino acid sequence [24, 44]. The other problem of the previous model is that even though it gives a qualitative description of the conditions for Hsp70 to work in the most efficient way, it is not able to directly relate the effect of the chaperone to the substrate with the amount of energy that can be used. Indeed, this is only possible when the change in conformation of the chaperone itself is taken into account. We decided to focus on the substrate protein rhodanese, which

---

[2]The values that will be chosen are all in agreement with those reported in the article of Mayer, Hu and Tomita [40].

has been proven to expand under the influence of Hsp70 by Schuler and co-workers [24]. In 2014, they measured the expansion of denaturated rhodanese in the absence and in the presence of DnaK, that is a particular family of prokaryotic Hsp70, its cochaperone DnaJ, that is a J-domain protein, and ATP using a method called *Förster resonance energy transfer spectroscopy* (FRET). This technique is based on the fact that an excited chromophore, *the donor*, can transfer energy to another chromophore, *the acceptor*, via a nonradiative dipole-dipole coupling [45]. The efficiency of this transfer of energy goes with the inverse of the sixth power of the distance between the two chromophores [45, 46]. Therefore, with a measurement of the transfer efficiency it is possible to deduce with good accuracy the distance between the chromophores.

In the experiment, they attached two fluorescent dyes in two positions along the sequence of the substrate protein and measured the transfer efficiency. The idea is that when the substrate protein expands the distance between the dyes increases and thus there is a net drop in the transfer efficiency.

As we can see in figure 3.7, the presence of chaperones provokes a decrease in the efficiency that is indicative of an expansion of the substrate.
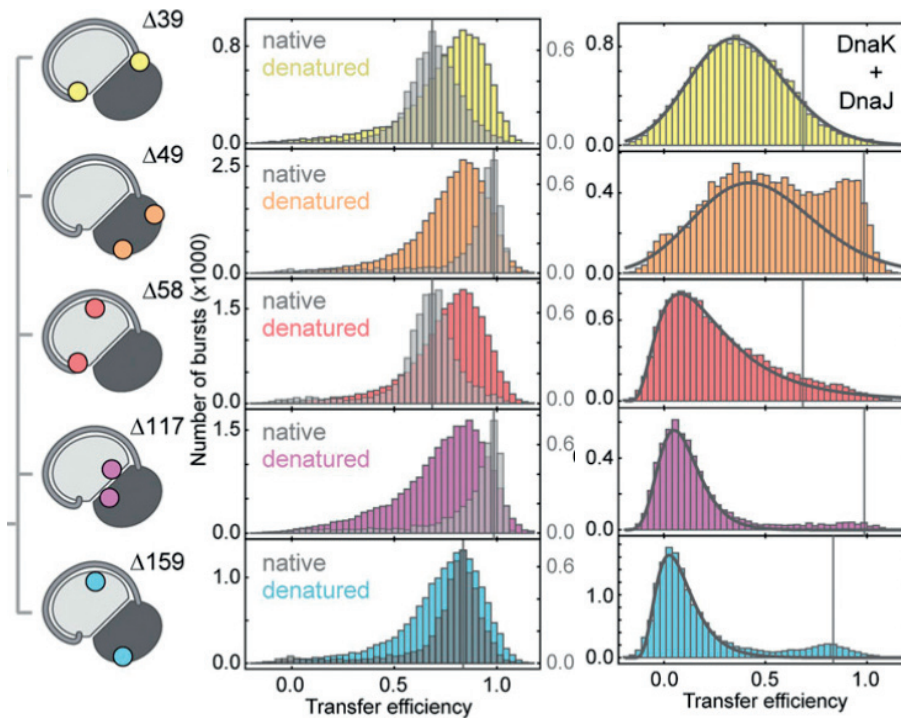
Figure 3.7 – Histogram for the FRET efficiencies for five different positions of the fluorescent dyes, in the absence (left) and in the presence (right) of DnaK and DnaJ. The histograms for the native state are shown in gray. Figure adapted from [24].

The mean values of the FRET efficiency for a system in the presence of chaperones is plotted in figure 3.8. We will soon discuss the validity of our model by a comparison of our prediction with these data.
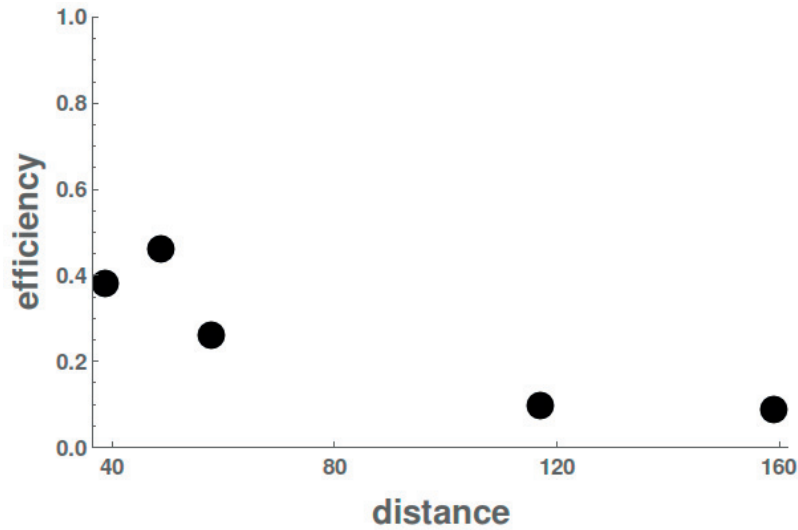
Figure 3.8 – Mean values of the FRET efficiencies in the presence of DnaK. Data taken from [24].

We wanted to model this system, and the first step to achieve this goal was to understand on a quantitative point of view the response of the substrate protein to the binding of a given number of chaperones.

This was done with molecular dynamic simulations (MD), with a coarse-grained model in which the residues interact via steric repulsion and hydrophobic interactions. In the model the substrate has 6 binding sites, located in the regions that were predicted at the same time with the algorithm of van Durme et al. [47] and with the one of Rüdiger et al. [44]. Each binding site could be occupied by a copy of a structure based model for Hsp70: each specific binding pattern was considered separately, so that the total number of possible configurations is 64. In this way, a density map was obtained from equilibrium simulations using the internal energy and the radius of gyration as parameters. The histograms for four different configurations[3] are reported in figure 3.9. When the substrate is not bound, the attractive interaction between the monomers of the chain wins over the entropic contribution to the free energy and the protein assumes a compact conformation, characterized by a small radius of gyration with a distribution that is strongly peaked on its average value. For a larger number of chaperones bound, there is an increase in both the radius of gyration and the internal energy, suggesting that the conformation of the substrate is more similar to the one of a random coil. Note that the distribution of the radius of gyration is quite broad even in the case in which all the binding sites are occupied, indicating that the entropic pulling with six chaperones is not yet able to force a completely elongated conformation.

With MD simulations the free energy increment ΔG due to the conformational rearrange-

---

[3]for example, with "100100" we indicate a substrate that has the first and the fourth binding sites occupied.
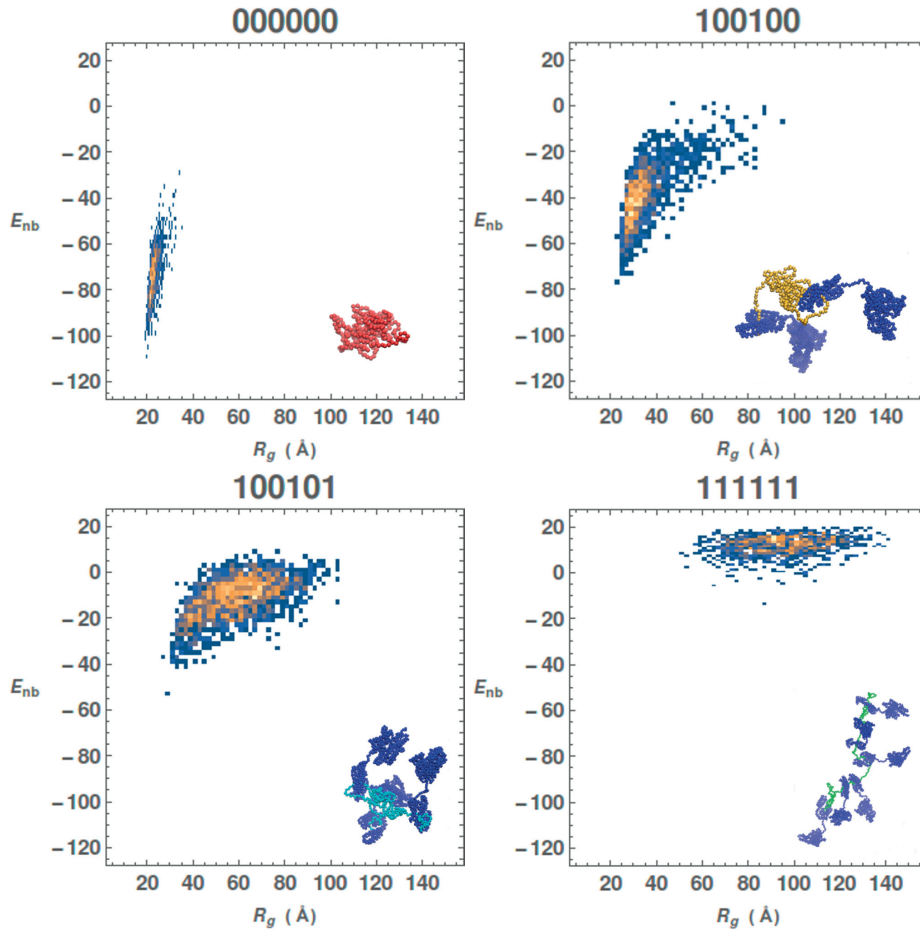
Figure 3.9 – 2D Histograms showing the radius of gyration and the internal energy for four different configurations. An increase of the number of bound chaperones leads to a larger internal energy and a larger radius of gyration.

ment of the substrate upon binding was also derived. The procedure that needs to be used for that derivation is to perform pulling simulations on the substrate and deduce the free energy from a measurement of the work W needed in each realization using the Jarzynski equality:

$$\exp\left[-\beta \Delta G\right] = \langle \exp\left[-\beta W\right]\rangle \ , \tag{3.38}$$

where the average must be intended over all realizations. The pulling was done with an harmonic potential acting on the radius of gyration:

$$U = \frac{1}{2}k_{el}\left[R_g - \lambda\right]^2 \ , \tag{3.39}$$

where k_{el} is a constant and λ is the control parameter. We can see the free energy for a given configuration in figure 3.10. In order to compare the free energies of substrates that are bound in different binding sites, the following assumption was made: when the rhodanese
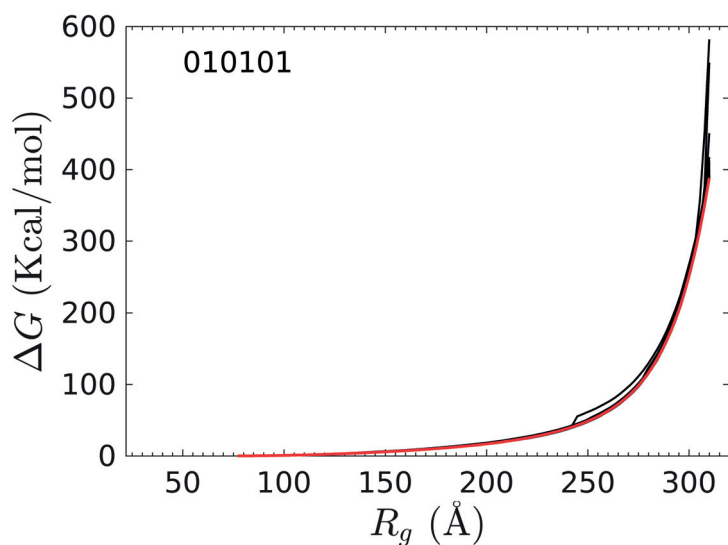
Figure 3.10 – Free energy from a pulling simulation for the configuration with the 2nd, 4th and 6th binding sites occupied.

is completely stretched, its free energy can be considered to be the same independently on the binding configuration. Therefore the curves relative to different configurations were superimposed so that the limiting values for large pulling forces were the same (an example is shown in figure 3.11).
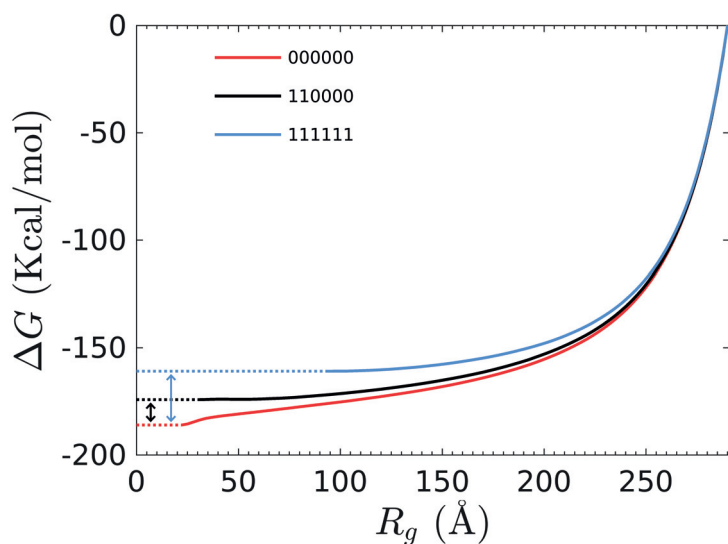
Figure 3.11 – Superposition of the averaged curves for three configurations.

Once we had the partial free energy relative to a precise value of the control parameter we evaluated the thermodynamic average to obtain the total free energy required for the rate model.

In figure 3.12 ΔG is plotted as a function of the number of bound chaperones n, for each configuration. The increment in free energy corresponding to the addition of one chaperone
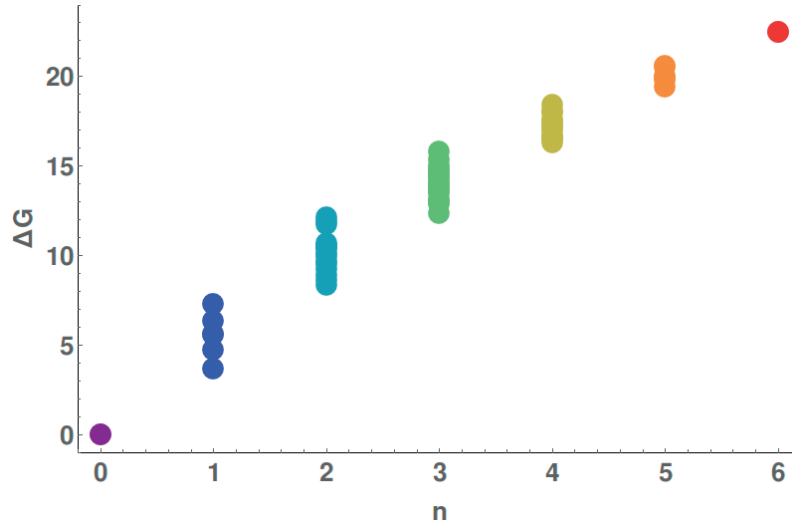
Figure 3.12 – Free energy of each configuration as a function of the stoichiometry.

is taken into account in the value of the dissociation constant:

$$K_d(i,j) \rightarrow K_d^0(i,j) \exp\left[-\Delta\Delta G_{i,j}\right] , \tag{3.40}$$

where $K_d(i,j)$ is the dissociation constant for the binding reaction between state i and state j, $K_d^0(i,j)$ is the dissociation constant for the isolated binding site (without considering the change in conformation of the whole substrate) and $\Delta\Delta G_{i,j}$ is the relative increment in free energy. Experimental evidence suggests that the unbinding rate of Hsp70 from large proteins is similar to the one from single peptides, whereas the binding rate can be up to two orders of magnitude smaller [24, 38, 48]. Therefore, the rate that must be different for the protein and for the single peptide is the binding rate.

We can now focus on the thermodynamics of the binding interaction between Hsp70 and the substrate protein, addressing in particular how the binding of the chaperones is influenced by the amount of energy that is available in the form of ATP. In an article of Barducci and De Los Rios [21] the dependence of the binding affinity on the non equilibrium regime induced by the hydrolysis of ATP was discussed in the case of single binding sites. In that work it was explained how chaperones can achieve an effective binding affinity for the client peptide that is larger than the affinity that they would have in the ADP or in the ATP state and this *ultra-affinity* is only possible if the system is out-of equilibrium. Here, we need to do a step further and consider as a substrate a protein with the six binding sites that have been mentioned above. We implemented a rate model in which each state corresponds to a single configuration, identified by the positions of all the bound chaperones in the ATP and in the ADP state. Therefore, since each of the six binding sites can be free, bound with a chaperone in the ATP or in the ADP state, we have in total

$3^6 = 729$ configurations. A schematic representation of two portions of the rate model is shown in figure 3.13 and 3.14.
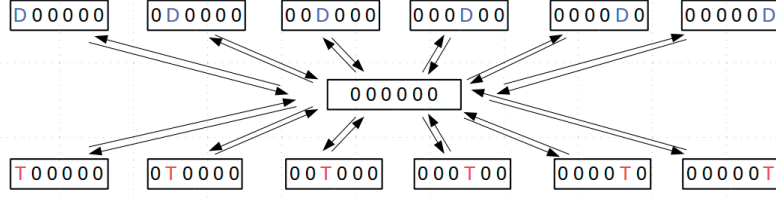
Figure 3.13 – Schematic representation of all the reactions related to the free state. We indicate a binding site that is free with 0, that is bound with a chaperone in th ADP state with D and in the ATP state with T.
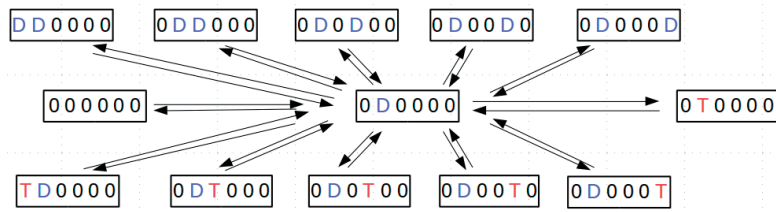
Figure 3.14 – Schematic representation of all the reactions related to the state with the second binding site occupied by a chaperone in the ADP state.

It is possible for Hsp70 to pass from the ATP to the ADP state via hydrolysis or via exchange. For the exchange we derived effective rates that depend on the ratio between the concentrations of ATP and ADP. We work under the assumption that the concentration of ATP and ADP is large enough that the concentration of the apo-state can be neglected. When this is the case, the effective rates of exchange are [21]

$$
k_{TD}^{ex} = k_{T-}\frac{k_{D+}}{k_{D+} + \alpha k_{T+}}\,,
$$
$$
k_{DT}^{ex} = k_{D-}\frac{\alpha\, k_{T+}}{k_{D+} + \alpha k_{T+}}\,,
\tag{3.41}
$$

where k_{D+} and k_{D-} are the rates of binding and unbinding of ADP respectively, k_{T+} and k_{T-} are the rates of binding and unbinding of ATP and we recall that $\alpha = \frac{[ATP]}{[ADP]}$.

In the Appendix are discussed the details of the model with all the rate constants, that have been taken from [24] and [38]. We calculated the probability of each configurations, then we summed the probabilities of configurations relative to the same number of chaperones. In figure 3.15 we plot the probability of each stoichiometric complex in the stationary state as a function of α. For low values of α, namely in conditions close to equilibrium, the vast majority of the substrate proteins are free. For values of α between 10^{-3} and 1, the complexes with less than 5 chaperones are populated. For a larger ratio, their probability decreases and the majority of the complexes contain more than 4 chaperones, with an

average number between 5 and 6. It is possible to define an effective dissociation constant
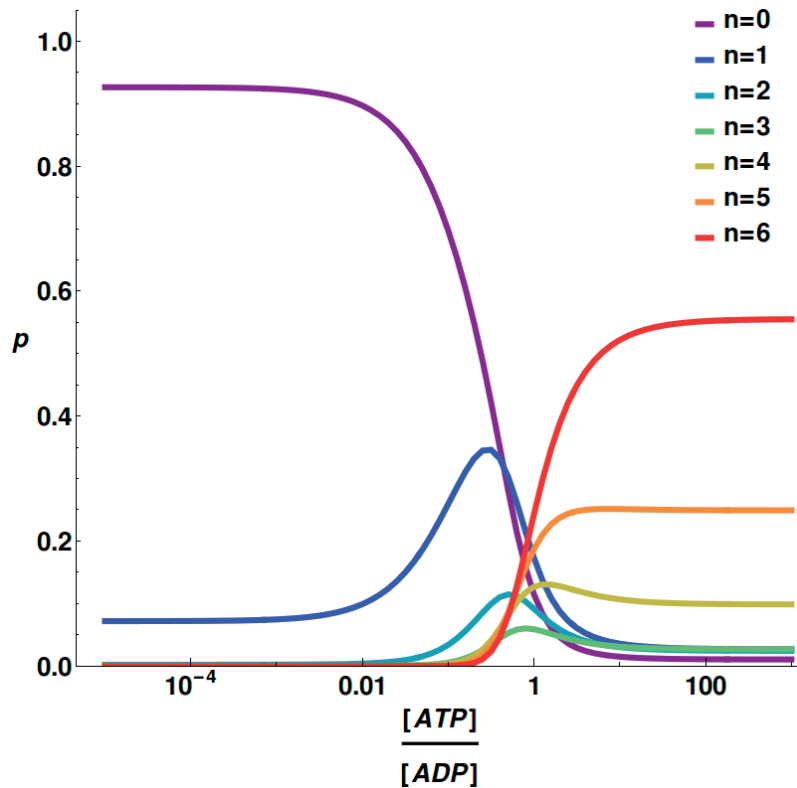
Figure 3.15 – Probability of each stoichiometry as a function of the ratio $\frac{[ATP]}{[ADP]}$. To an increase in $\frac{[ATP]}{[ADP]}$ corresponds a larger number of bound chaperones.

K_d as

$$K_d = \left[\frac{6}{\langle n\rangle} - 1\right][\text{DnaK}], \tag{3.42}$$

where $\langle n\rangle$ is the average number of occupied binding sites. In figure 3.16 we report the dissociation constant divided by its value in equilibrium as a function of α.
In [21] the rate of hydrolysis was used to quantify the degree of non equilibrium. Here, we decided to use the parameter α, since it is more directly related to the energy that the chemostat can inject into the system.
In equilibrium conditions K_d is equal to the dissociation constant of a chaperone in ADP state. When the concentration of ATP is larger, the dissociation constant is well below that value, as predicted in [21]. The red region in the figure is the ultra-affinity region, namely the gain in affinity caused by the non equilibrium.

A direct measure of the expansion of rhodanese is provided by the radius of gyration. The average value was measured in the equilibrium MD simulation for each configuration. We then calculated the average over all the configurations weighting each term with the concentrations obtained with the rate model. When calculating the main quantities of
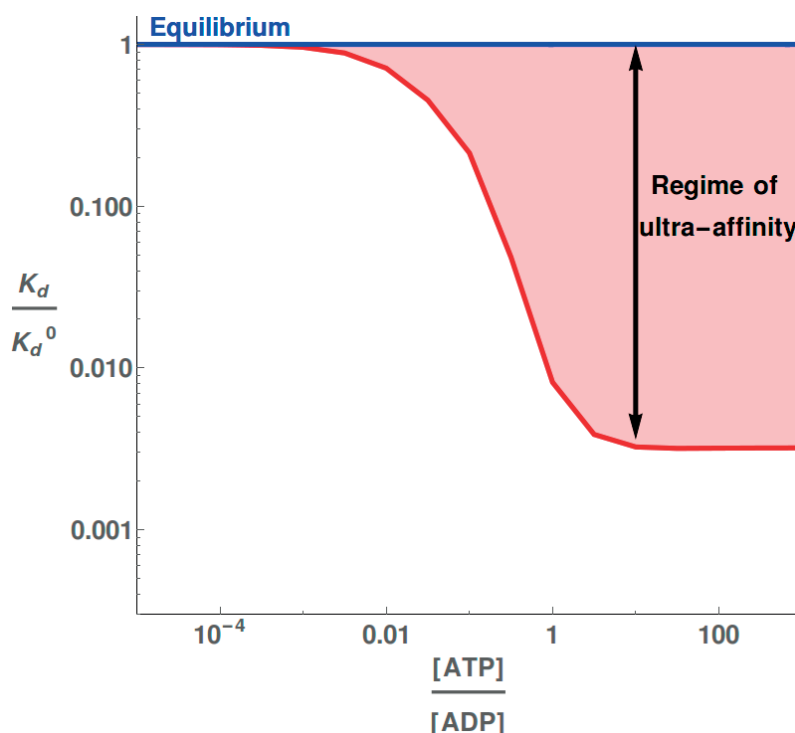
Figure 3.16 – Dissociation constant as a function of α.
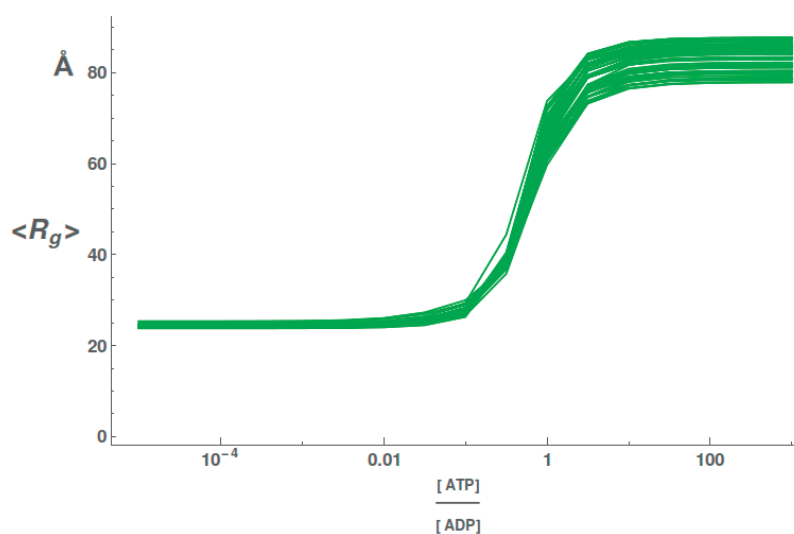
Figure 3.17 – Radius of gyration as a function of α. Each curve corresponds to a different set of free energy increases for all the configuration. The set was randomly generated from a Gaussian distribution centered in the values obtained with the MD simulations.

interest we had to face the problem estimating the error for each curve. Since the largest uncertainty was on the increment in the free energy, we randomly generated a value for the increment from a normal distribution with a standard deviation of 0.3 kcal/mol for each step in the reaction and obtained in this way 50 curves for the radius of gyration. The result is the plot in figure 3.17: for equilibrium values of α the rhodanese has a compact

denaturated conformation, with an average radius of gyration $R_g \simeq 25$Å, and then for physiological values of α it reaches an expanded conformation with $R_g \simeq 80$Å.

The solution of the chemical network has also been used to calculate the average efficiencies of virtual FRET dyes placed at the same positions as in the experiment [24]. The value of the FRET efficiency for each configuration was averaged using the results of the rate model. The efficiency was measured during the MD simulations. In equilibrium the FRET efficiency is around 0.8 for all values of the distance between the fluorescent dyes (figure 3.18 blue). In the non equilibrium stationary state, the expansion of the substrate leads to an efficiency that is significantly lower (figure 3.18 red) and close to the experimental values (figure 3.18 black).

At this point, it is natural to wonder what is the free energy gain of the whole system as
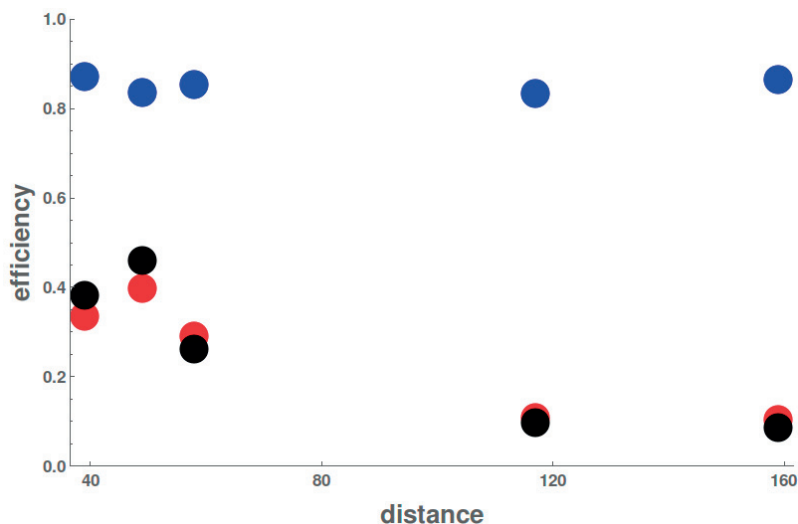
Figure 3.18 – Mean values of the FRET efficiencies in our model in the presence of DnaK, for equilibrium (blue) and non equilibrium (red) values of α, compared to the result of the experiments (black).

a consequence of the non equilibrium and whether the system is optimized in physiological conditions. The free energy of the system Ω is defined in the following way

$$\Omega = \sum_i p_i g_i + RT \sum_i p_i \ln p_i \tag{3.43}$$

where g_i is the difference in free energy between the ith bound state and the free state and p_i is the probability of such a state [4]. We define a parameter called *effectiveness*, θ, that is the ratio between the increase in the free energy with respect to equilibrium and the free

---

[4]Do not confuse g_i, that is the increment in free energy taking into account both the change in conformation of the protein and the intrinsic affinity between the chaperone and the binding site, and ΔG_i, that is instead the conformational free energy.

energy that is available in the form of ATP:

$$\theta = \frac{\Omega - \Omega_{eq}}{6 \ln \left[\frac{\alpha}{\alpha_{eq}} \right]} \, , \tag{3.44}$$

where α_{eq} is the ratio between the concentration of ATP and ADP in equilibrium conditions and the number 6 is due to the fact that there are in total six binding sites, therefore in principle the hydrolysis of six ATP molecules could take place at the same time. In figure
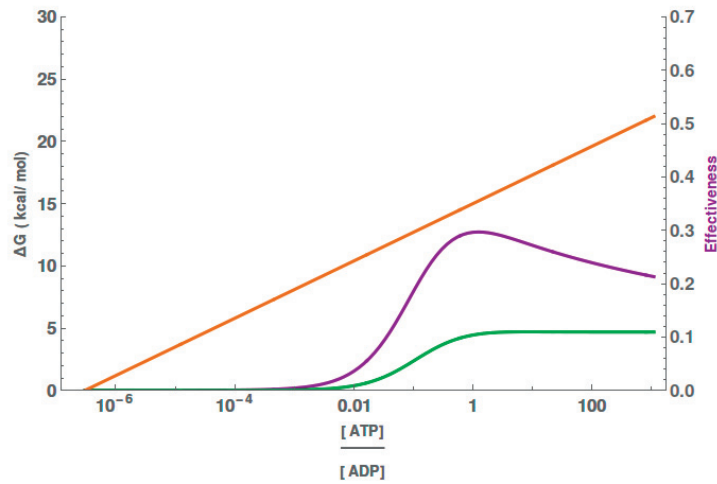
Figure 3.19 – Effectiveness (purple) as a function of α, together with its numerator (green) and denominator (orange).

3.19 we plot the effectiveness (purple), together with its numerator (green), that is the free energy difference with respect to equilibrium, and its denominator (orange), the free energy available in the form of ATP. The increase of the free energy of the system reaches a plateau when the ratio α is close to the one of physiological conditions ($\alpha \in (1, 10)$). Therefore, we can conclude that the physiological conditions are in an interval of parameters such that the improvement in the binding affinity is achieved at an optimal energetic cost.

3.1.6 Brief summary

We used the tools of non equilibrium statistical physics to describe some aspects the Hsp70 system. First of all, with a simplified model we derived the constraints that the rate constants must follow in order for Hsp70 to perform its role as unfoldase. We were able, in this way, to show that the non equilibrium is a necessary condition to improve the yield of native proteins in the stationary state. If the purpose of Hsp70 is the refolding of the misfolded polypeptide, there is the need to properly tune the rates of binding and unbinding in such a way that the protein stays in the bound state for a time interval that is long enough to modify its conformation but at the same time it is not too long, because the unbinding is necessary to give the protein the chance to spontaneously refold. Cells naturally perform this fine tuning with the use of co-chaperones. We briefly discussed the most important of them, the J-domain protein, focusing in particular in its role in the catalysis of the hydrolysis of ATP. It was then possible to obtain more quantitative results by considering a particular protein, rhodanese, as substrate for Hsp70, and confront them with the available experimental data and molecular dynamic simulations. For this latter problem, we used a model in which the substrate protein has 6 sites, and we discriminated between the different states of Hsp70. We showed that the substrate expands upon the interaction with more than one Hsp70 and we deduced values for the FRET efficiency of fluorescent dyes located in two regions of the substrate that were in perfect agreement with what was measured experimentally.

3.2 Chaperonin

The results presented in this section have been published in [49]. Many of the figures have been adapted from the article, in accordance with the Creative Commons Attribution License used by Nature Chemical Biology.

3.2.1 General description

Hsp70 is surely not the only biological system involved in protein folding in the cell. Another very important chaperone family is the one of chaperonins, that are large proteins present in almost the totality of bacteria and eukaryotes [1]. There are different types of chaperonin, here we will focus on the bacterial GroEL, that is the first discovered and by far the most studied [23]. Even if the final outcome is similar to the one of Hsp70, namely the refolding of misfolded proteins, the mechanism by which this happens is completely different. GroEL is formed by two domains [1, 23, 30], composed by seven subunits each, arranged in a cage-like structure (see figure 3.20). The subunits have a mass of about 60 kDa, so that the whole complex has a mass of more than 800 kDa. All the seven subunits have a binding site for ATP, and the binding reaction is cooperative, which means that as soon as one molecule binds, the binding affinity for all the others is strongly enhanced [1]. Each ring has two conformations, one with large and one with low affinity
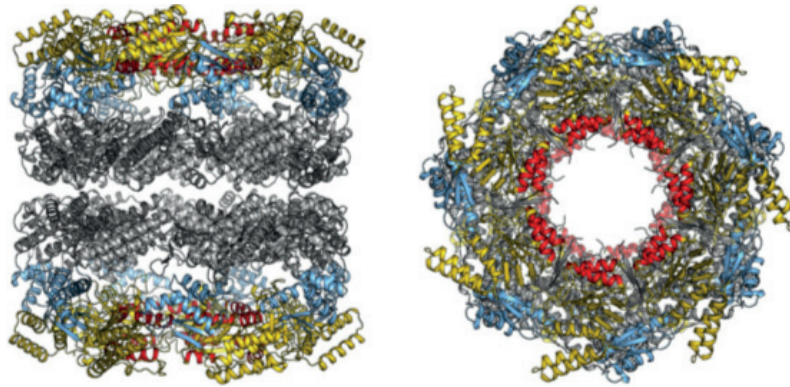
Figure 3.20 – Structure of GroEL. Figure taken from [1].

for ATP. The binding of a nucleotide stabilizes the former with respect to the latter. The way the nucleotides are used to tune the affinity for the client protein is the opposite of the one of Hsp70 [50, 51], because the ADP state has lower affinity than the ATP state [1, 23], therefore the hydrolysis stimulates the unbinding rather than the binding reaction. The cage also has a cover, a protein called GroES, formed by seven subunits of about 10 kDa. GroES is responsible to close the cage after the substrate protein has entered [1]. We will often indicate the whole complex GroEL/GroES by simply writing GroELS. The precise mechanism of refolding is still a matter of animate discussions: we can say that the models that by now have more popularity in the chaperones community are essentially three: Anfinsen cage (AC), active folding (AF) and iterative annealing (IA). According to the AC model, as the name suggests, GroEL does nothing but protecting the substrate protein from aggregation, leaving it in an environment that is suitable for folding, even though the protein refolds at the same rates that it would have in the case of a spontaneous refolding in conditions in which there is not the obstacle of aggregation. Therefore, the role of GroEL would be to deplete the misfolded state in order to reduce the mean speed of the aggregation process. However, several experiments have shown that it is possible for GroEL to accelerate the refolding, instead of just increasing the yield of the functional conformation once the stationary state has been reached [52, 53]. The reply of the supporters of the AC model is that the aggregation might be *a reversible process*: if this is the case, the apparent acceleration of the refolding can be explained as the result of the rescue of proteins from aggregates. In an article published in *Pnas* [54], Apetri and Horwich monitored the folding of Rubisco in different conditions. They did confirm that there was an acceleration of the refolding in the presence of GroEL. However, they also showed that Rubisco was able to spontaneously refold when the aggregation was obstructed by the presence of bovine serum albumin (BSA). Moreover, they showed that when they used the maltose binding protein (MBP) as a substrate the refolding had the same speed when it was spontaneous and when it was helped by the chaperonin, providing that they artificially created the conditions necessary to avoid aggregation. They thus concluded that the role of GroEL was to protect the substrate proteins from reversible aggregation and that the chaperone was not able to actively enhance their conformational remodeling.

The strongest argument against the AC model lies in the fact that GroEL has been shown to successfully stimulate protein refolding even at concentrations of the substrate that are so low that the aggregation can be neglected [52, 55]. This is the main reason why models postulating an active role of GroEL in the refolding process have been proposed [23, 1].

In the AF model the hydrophilic walls of the chaperonin participate to the folding of the substrate protein via steric repulsion [56, 1]. The reduced available volume inside the cage favors the formation of the bonds required for the tertiary structure of the native state. The role of confinement in protein folding has been demonstrated with MD simulations by Baumketner and co-workers [57]. The power of the AF model is that it is able to explain why the refolding is accelerated even for very small concentration of substrate proteins, which indicates a low probability of aggregation, and even when GroELS is in sub-stoichiometric amount. In 2014 Gupta and co-workers [52] demonstrated with fluorescence resonance energy transfer spectroscopy (FRET) that GroEL can accelerate the refolding of a mutant of the maltose binding protein at concentrations as low as 100 pM, arguing for this reason that the AF is likely the most promising model, since aggregation is not a relevant factor at those concentrations.

Another possible explanation, non necessarily in conflict with the previous one, is the iterative annealing (IA). According to this model, GroEL stimulates the unfolding of the misfolded proteins and gives them a chance to properly refold spontaneously, either inside or outside the cage, and this can be accomplished after several binding/release cycles. Lin and co-workers [53] used FRET spectroscopy to show how a misfolded substrate unfolds in the presence of GroEL and ATP. As a substrate they used Ribulose-1,5-bisphosphate carboxylase/oxygenase (known as RuBisCO), which is an abundant enzyme involved in the process of carbon fixation[58], that after denaturation is found in misfolded conformations that are susceptible to aggregation even at 25 °C. The unfolding time of misfolded RuBisCO is longer than the average time in which it is bound with GroEL [53, 59], therefore they concluded that the transformation that leads to a recovery of the functional state happens via several cycles of binding and release and it is thus unlikely that the active refolding could be the only process by which the functional state is recovered.

3.2.2 Out-of-equilibrium stabilization of the functional state

Rather than trying to validate or disprove any model, we wanted to deepen a concept that is maybe slightly undervalued by part of the chaperone community, that is the out-of-equilibrium intrinsic nature of the chaperone systems and GroELS in particular. We wanted to show that even in some stress conditions in which the functional conformation of the substrate does not correspond to a minimum in the free energy landscape, and thus it is not the one that is favored in equilibrium, it is possible for the GroELS system to stabilize the functional conformation in a non equilibrium steady state (NESS). We devised this work in collaboration with the group of Goloubinoff in the University of Lausanne, which also performed all the experiments, and with Alessandro Barducci in the University

of Montpellier.

The choice of the substrate went on malate dehydrogenase (MDH) in mitochondria. MDH is a widely used substrate for GroEL that is properly folded, forming a functional dimer, at 25 °C, but loses its structure at higher temperatures [55]. In the misfolded state, it is also prone to aggregation. If we try represent schematically the different possible state of the protein at 37 °C we obtain what is shown in figure 3.21.
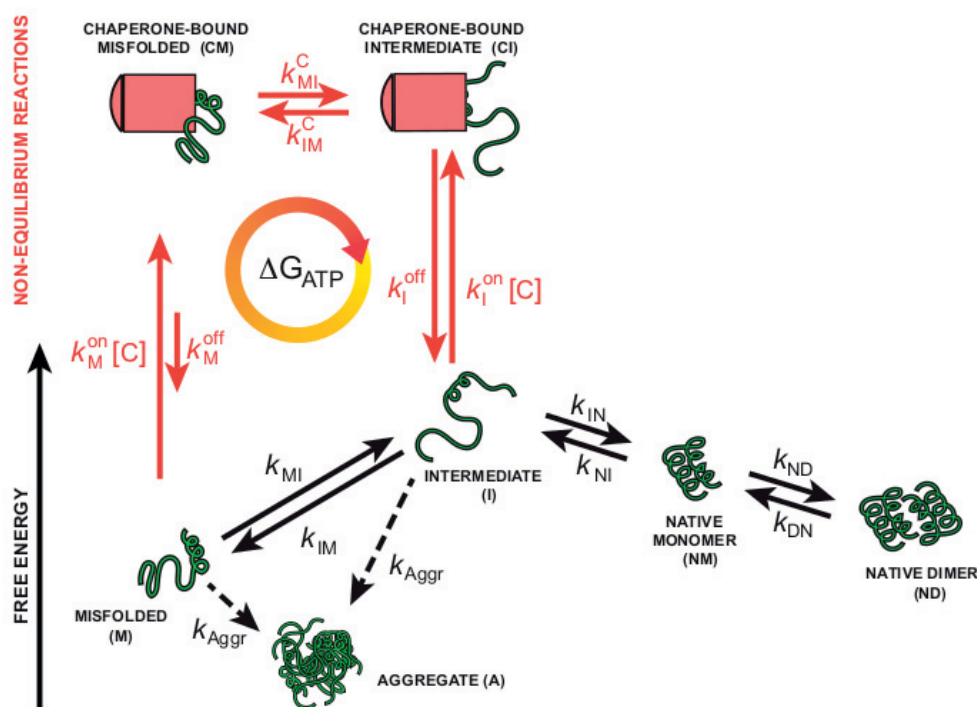
Figure 3.21 – Schematic representation of the different conformations of MDH, organized on the basis of their free energy at 37 °C. At this temperature, the native state is less stable than the misfolded state.

Obviously, apart from the native dimer, the states that are shown in the figure do not correspond to well defined tertiary structures but must rather be intended as coarse grained ensembles of several conformations sharing the same structural properties. As we can see in the scheme, at 37 °C the misfolded state (M) is at a lower free energy than the native state (N), as well as than the functional dimer (D). The aggregates are even more stable but the kinetics to reach them is rather slow at the concentrations that were used in the experiments ($< 1\,\mu M$). The arrows shown in red indicate the reactions in which the ATP is hydrolyzed, and thus the energy is consumed. In the experiment, the presence of the functional dimer was determined by monitoring the substrate's enzymatic activity. At 25 °C the substrate maintained its native structure and formed the dimer without being helped, as expected (figure 3.22, left). However, there was a significant decrease in activity

at 37 °C, as we can see in figure 3.22(right). If only GroELS was added, the denaturation still occurred at the same rate and the activity was not recovered (blue data). On the contrary, when ATP and a regeneration system were present[5], a clear reactivation was measured and MDH remained stable for more than 2 hours (red data).
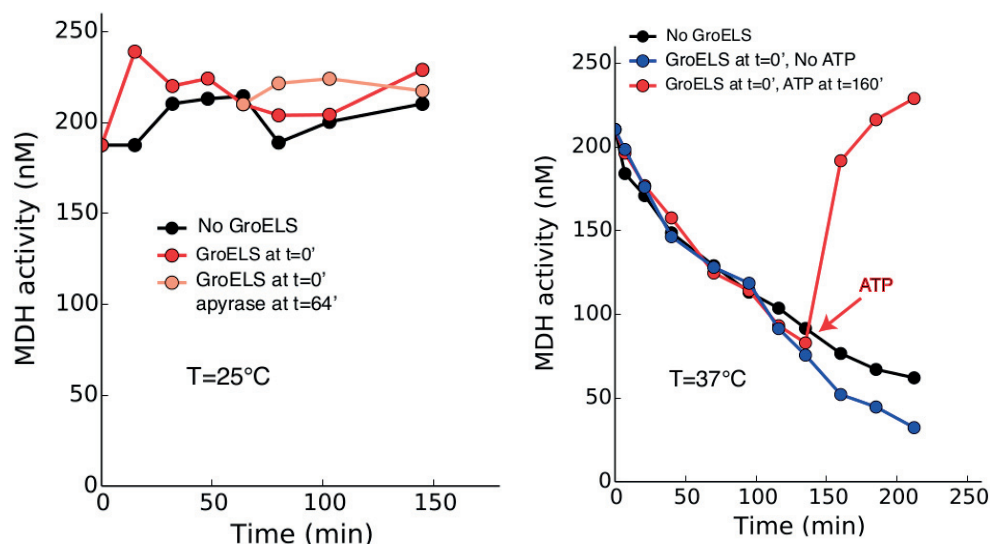
Figure 3.22 – (left) Enzymatic activity of MDH at 25 °C as a function of time, for different experimental set-ups. (right) Enzymatic activity of MDH at 37 °C as a function of time, for different experimental set-ups. For both the blue and the red curve GroELS was added after 160 min, but only for the red curve there was also an addition of ATP.

To determine which fraction of the substrate was not possible to recover because it was trapped in irreversible aggregations, an experiment was performed in which GroELS and ATP where added at different times. The aggregation started to be significant after a quite large time (see figure 3.23, left), suggesting that the *irreversible* aggregation was a slow process.

As a consequence, unless we assume that there is a reversible aggregation that is fast enough to compete with the refolding process and thus to slow the net refolding rate, we must conclude that the conformational rearrangement of the substrate protein is actively enhanced by GroELS. In figure 3.23(right) we show how the different conformations are distribute once the stationary state is reached, for different value of the time in which GroELS and ATP are added to the system. Note that the native monomer is included in the non-recoverable fraction, since native MDH monomers need to form dimers in order to be functional and thus to contribute to the enzymatic activity. In the presence of chaperones and ATP, the native state could be maintained for a long time (figure 3.24, left), or it could be recovered if at the initial time MDH was denaturated (figure 3.24, right). However, upon the addiction of apyrase, an enzyme known to massively catalyze

---

[5]Which means that the ATP concentration could be considered constant for the time scales under consideration.
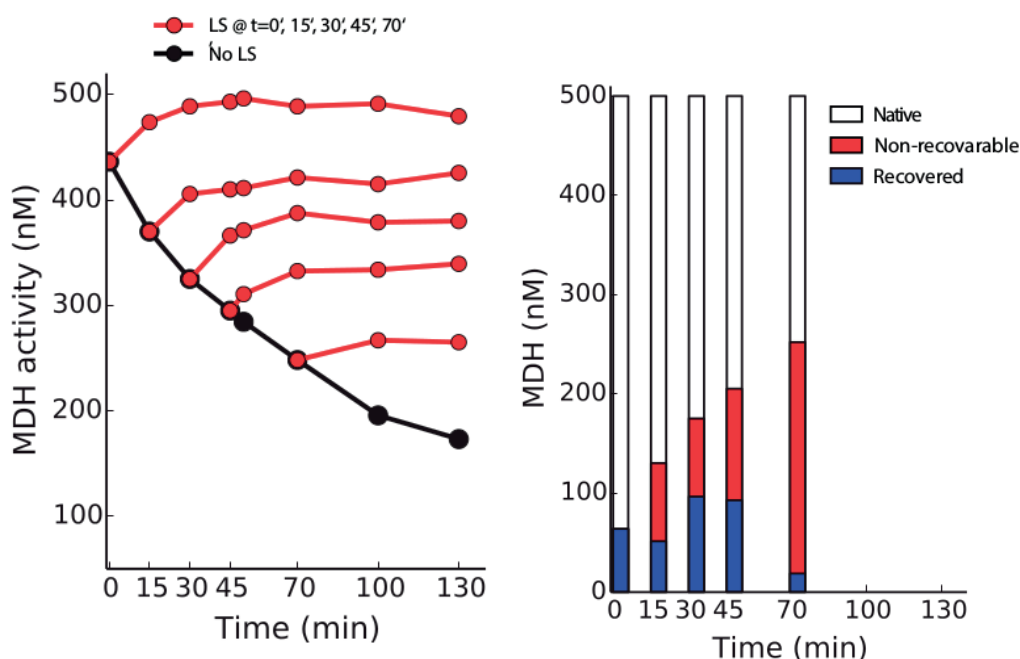
Figure 3.23 – (left) Recovery of MDH activity with the help of GroELS, added in the system at 5 different times. (right) Fraction of native, recovered and non-recoverable MDH in the stationary state for different values of the time at which GroELS was added.

the hydrolysis of ATP so that there was no ATP left in the system, the activity was quickly lost again.
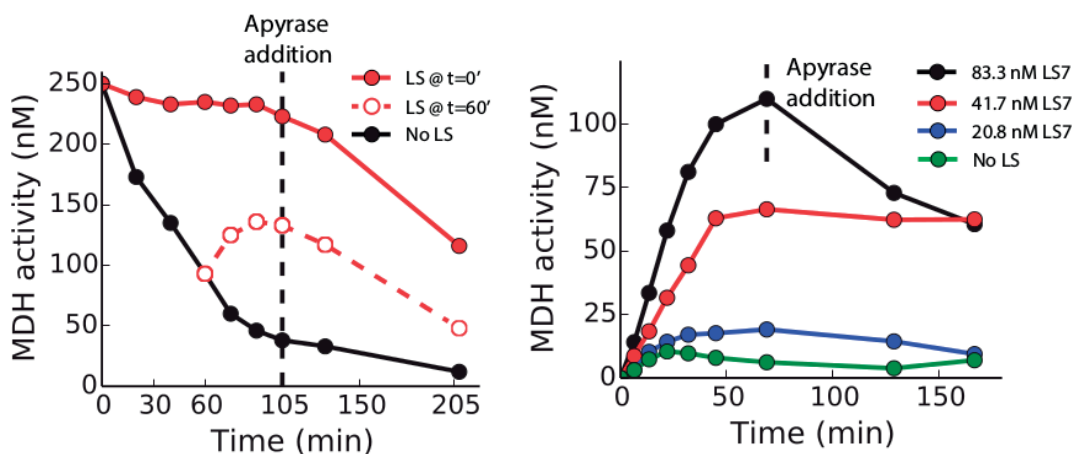
Figure 3.24 – (left) Enzymatic activity of $250\,nM$ of MDH as a function of time in the presence and in the absence of $3.5\,\mu M$ of GroELS. Apyrase was added after 105 min. (right) Enzymatic activity of $500\,nM$ of MDH pre-denaturated with urea, for different concentrations of GroELS. Apyrase was added after an hour.

The implication is that molecular chaperones were continuously needed to keep the functional conformations stable, rather than behaving simply as kinetic means to reach them

from the misfolded state. The free energy ΔG_h that is available thanks to the hydrolysis of ATP is given by

$$\Delta G_h = k_b T \ln\left[\frac{[ATP]k_h}{[ADP][P_i]k_{sy}}\right] = k_b T \ln\left[\frac{[ATP]}{[ADP][P_i]}\right] - k_b T \ln\left[\frac{[ATP]_{eq}}{[ADP]_{eq}[P_i]_{eq}}\right] \, , \quad (3.45)$$

where with k_h and k_{sy} we indicate the rates of hydrolysis and synthesis and with the subscript *eq* we indicate concentrations in equilibrium. Note that in equilibrium conditions we have $\Delta G_h = 0$, as we should expect, since the detailed balance is satisfied:

$$[ATP]_{eq}k_h - [ADP]_{eq}[P_i]_{eq}k_{sy} = 0 \, . \quad (3.46)$$

We can see that the free energy is a function of the ratio between the concentration of ATP and ADP. As a consequence, a part from being lower for lower concentrations of ATP, it also decreases if the concentration of ATP does not change while we increase the concentration of ADP. Without the use of the free energy, a more intuitive explanation is that if there is overabundance of ADP with respect to ATP there will be too many chaperones in the ADP state and the kinetic cycle will not be effectively followed in the right direction. This phenomenon is shown in figure 3.25 (right), where the enzymatic activity of MDH at 37 °C is plotted as a function of time for different values of the concentration of ADP, while the ATP was kept at a concentration of $400 \, \mu M$. We can see that in a first stage GroELS is able to successfully refold the misfolded MDH, which recovers its activity. However, since this experiment was performed without a regeneration system of ATP, after a certain time the energy supplied starts decreasing until the system is found in equilibrium, where as we know the substrate is mostly misfolded and thus the enzymatic activity is very low.

Refolding processes of this kind can be rationalized with a rate model. In this case a non linear approach is needed, because both the dimerization and the aggregation are processes that depend on multiplications of two concentrations. We assumed that only proteins in the intermediate or misfolded state can bind to the chaperone. This assumption is substained by experimental observations showing that a substrate in the native state has a negligible affinity for GroEL [52]. We also assumed that proteins in misfolded or intermediate state can form aggregates irreversibly. Moreover, proteins that are already in aggregates can aggregate with other non native proteins. However, we did not discriminate between aggregates composed of different numbers of proteins. The chemical network is
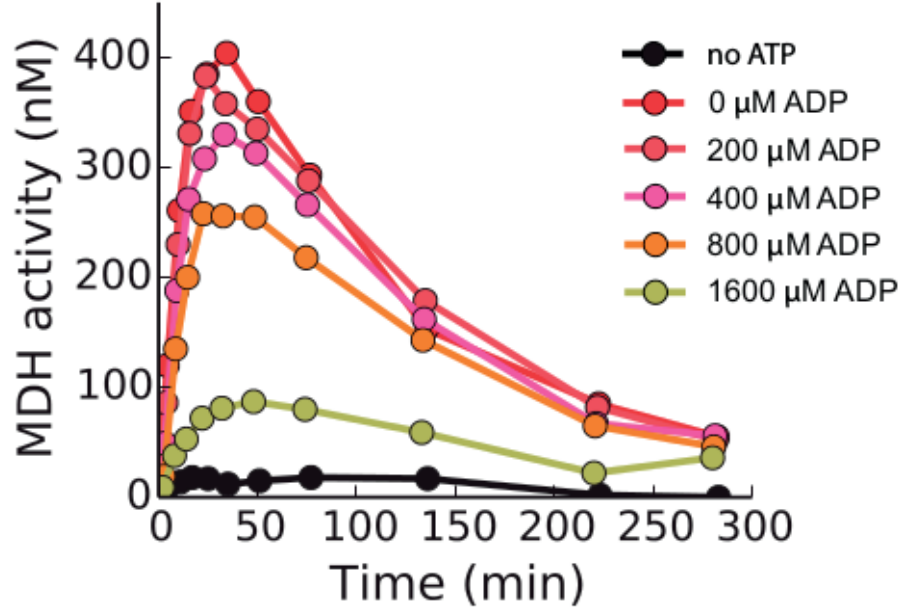
Figure 3.25 – MDH activity as a function of time, for different values of ADP. The concentration of ATP was kept at an initial value of $400\,\mu M$ (with the exclusion of the black curve), but the experiments were performed without a regeneration system of ATP.

regulated by the following system of equations:

$$\frac{d[D]}{dt} = -2k_{DN}[D] + 2k_{ND}[N]^2$$

$$\frac{d[N]}{dt} = -k_{NI}[N] - 2k_{ND}[N]^2 + 2k_{DN}[D] + k_{IN}[I]$$

$$\frac{d[I]}{dt} = -(k_{IN} + k_{IM})[I] + k_{NI}[N] + k_{MI}[M]+$$
$$- k_{agg}([M] + 2[I] + [A])[I] + k_I^{off}[CI] - k_I^{on}[C][I]$$

$$\frac{d[M]}{dt} = -k_{MI}[M] + k_{IM}[I] - k_{agg}(2[M] + [I] + [A])[M]+$$
$$+ k_M^{off}[CM] - k_M^{on}[C][M]$$

$$\frac{d[A]}{dt} = 2k_{agg}\left([M][I] + [M]^2 + [I]^2\right) + k_{agg}([M] + [I])[A] \qquad (3.47)$$

$$\frac{d[CI]}{dt} = -\left(k_I^{off} + k_{IM}^C\right)[CI] + k_I^{on}[C][I] + k_{MI}^C[CM]$$

$$\frac{d[CM]}{dt} = -\left(k_M^{off} + k_{MI}^C\right)[CM] + k_M^{on}[C][M] + k_{IM}^C[CI]$$

$$\frac{d[C]}{dt} = -k_I^{on}[C][I] - k_M^{on}[C][M] + k_I^{off}[CI] + k_M^{off}[CM].$$

The notation is the following:

- $[D]$ is the concentration of proteins forming native dimers;

- $[N]$ is the concentration of native proteins that do not form a dimer;

- $[I]$ is the concentration of free proteins in the intermediate state;

- $[M]$ is the concentration of free misfolded proteins;

- $[A]$ is the concentration of proteins that are in aggregates;

- $[CI]$ is the concentration of proteins in the intermediate state that are bound with a chaperone;

- $[CM]$ is the concentration of proteins in the misfolded state that are bound with a chaperone;

- $[C]$ is the concentration of free chaperones;

- The rate $k_{\alpha\beta}$ represents the reaction $\alpha \to \beta$, the apex *on* and *off* indicate binding and unbinding respectively and C the presence of the chaperone.

Note that we used a single state for GroEL, even though usually the different conformations of the chaperone are treated separately. However, since we are more interested in a discussion about the non equilibrium nature of the system, rather than in the details of the chaperone's conformations, this simplification has no direct consequences for the purpose of this work. The effects of the hydrolysis of ATP are implicitly taken into account in the values of the rate constants of binding and unbinding. We know that the affinity coefficient γ defined in formula 2.7 is related to the free energy injected into the system by the following expression:

$$\Delta G = \frac{1}{\beta} \ln \gamma \,, \tag{3.48}$$

where $\beta = \frac{1}{kT}$. For this particular system, we have

$$\gamma = \exp\left[\beta\Delta G\right] = \frac{k_M^{on} k_{MI}^C k_I^{off} k_{IM}}{k_M^{off} k_{IM}^C k_I^{on} k_{MI}} \,, \tag{3.49}$$

and in this way we can connect the concentration of ATP with the rate constants of the model.

The rates of aggregation, as well as the rates of the unfolding and misfolding reactions, were chosen so that at the same time they were comparable to the ones reported in the literature [40] and they could faithfully reproduce the curve of spontaneous reaction and the third curve in figure 3.23. The same rate constants have been used for all the other plots. All the rates used in the reactions are reported in a table in the Appendix. With these values the dimer is more stable than the native monomer and the non-native proteins, namely the combination of misfolded and unfolded intermediates, are more stable than the

proteins in the native states. The aggregation is irreversible but its rate is small compared to the others. However, its effect are evident in the time scales under consideration.

The rate of unfolding stimulated by GroEL has been chosen in such a way that the chaperone enhances a more unstructured conformation. Therefore, even though the misfolded conformation is still more stable, the free energy barrier is lower when GroEL is bound.

Another aspect that we want to stress is that the rates of binding have been chosen to be the same for the misfolded proteins and the intermediates: $k_M^{on} = k_I^{on}$. In fact, we do not have any experimental evidence that suggests that for GroEL this is not the case. In figure 3.26 (left) we can see that the model reproduces well the renaturation of MDH at 37 °C. Moreover, it was possible to discriminate between two different types of *non recoverable* proteins, that it was not possible to distinguish experimentally. Indeed, in addition to the fraction of MDH that is irreversibly aggregated, among the proteins which do not show enzymatic activity it is also necessary to consider the native monomers. We can see the
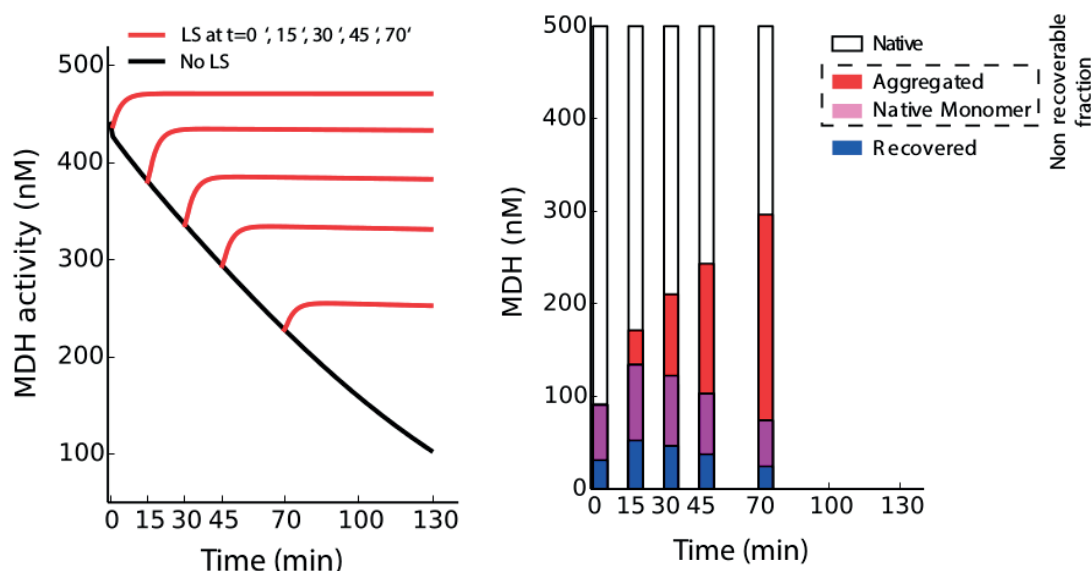
Figure 3.26 – Results of the rate model. (left) Reactivation of MDH when GroELS is added at different times. (right) Fraction of MDH in the different states in the stationary state for different times of the GroEL addition.

difference by comparing figure 3.23 (right) and 3.26 (right). Especially in the initial phase, a large fraction of the proteins that do not contribute to the enzymatic activity are native monomers that have not formed functional dimers. The apyrase addition corresponds, in the model, to setting the free energy of ATP to zero (figure 3.27). In the same spirit, it was possible to change the ratio between the concentration of ATP and ADP and reproduce the experiment performed with several different concentrations of ADP (confront figure 3.25 with figure 3.28). The absence of the regenerating system imposed the need to estimate the decrease in time of the concentration of ATP. It was assumed that the rate was proportional to the actual flux of dissociation between the chaperone and the protein in the intermediate

state:

$$\frac{d[ATP]}{dt} = -7\chi \left(k_I^{off}[CI] - k_I^{on}[C][I] \right), \tag{3.50}$$

where χ is a constant whose value is chosen in order to best reproduce the experimental results. We have $\chi = 12$, suggesting that for every productive cycle there are about 11 unproductive ones. From the previous results it is quite evident that the process of
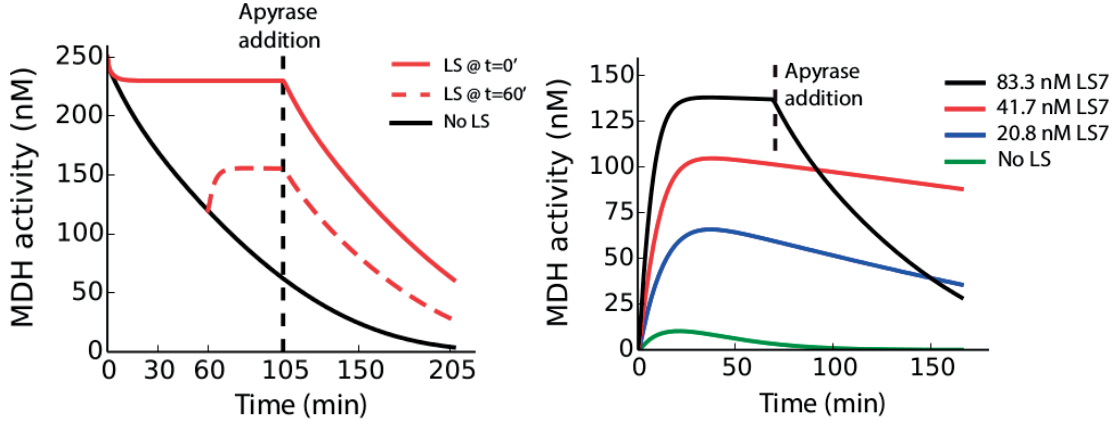
Figure 3.27 – Results of the rate model. (left) Enzymatic activity of $250\,nM$ of MDH as a function of time in the case MDH is in its functional state at the initial time. Apyrase was added at 105'. (right) MDH activity as a function of time for different concentration of GroELS, in the case MDH is non-native at the initial time. Apyrase was added after one hour.

aggregation is slow in the case of MDH. Therefore, we considered the non equilibrium stationary state of the system assuming that the aggregation can be neglected. We were particularly interested in a quantification of the stabilization of the native states with respect to the non-native ones. In a NESS we can define, with a slight abuse of terminology, a *nonequilibrium free energy* as

$$\Delta G_{nat} = -k_b T \ln \left[\frac{[P_{nat}]}{[P_{nonnat}]} \right], \tag{3.51}$$

where

$$[P_{nat}] = [N] + [D], \tag{3.52}$$

and

$$[P_{nonnat}] = [M] + [I] + [MC] + [IC]. \tag{3.53}$$

We wanted to see the dependence of this quantity from the free energy that is available because of the ATP hydrolysis, ΔG_{ATP}, for a system at 37 °C. In figure 3.29 we show that for low values of ΔG_{ATP} the non native state is the most stable one, but by increasing the free energy of ATP it is possible to significantly stabilize the native state.
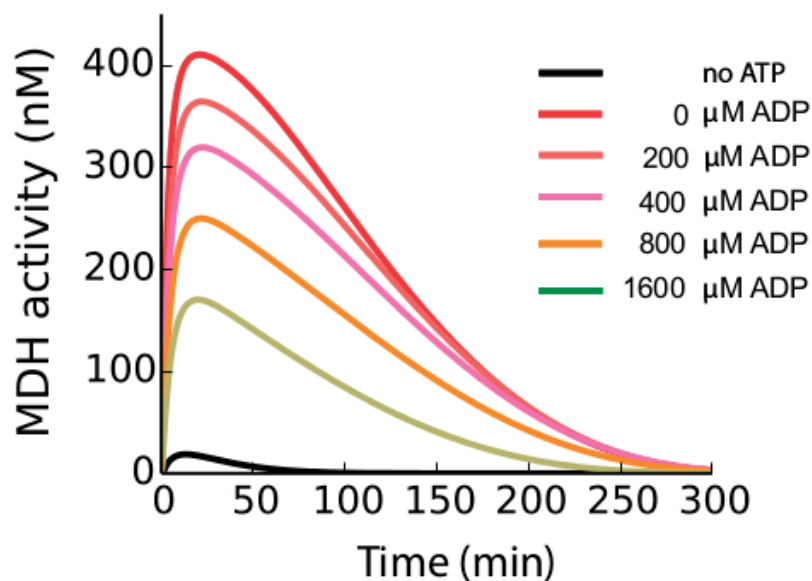
Figure 3.28 – Results of the rate model. MDH activity as a function of time for different concentrations of ADP. The system was considered to be without a regenerating system, so that the concentration of ATP decreases over time.
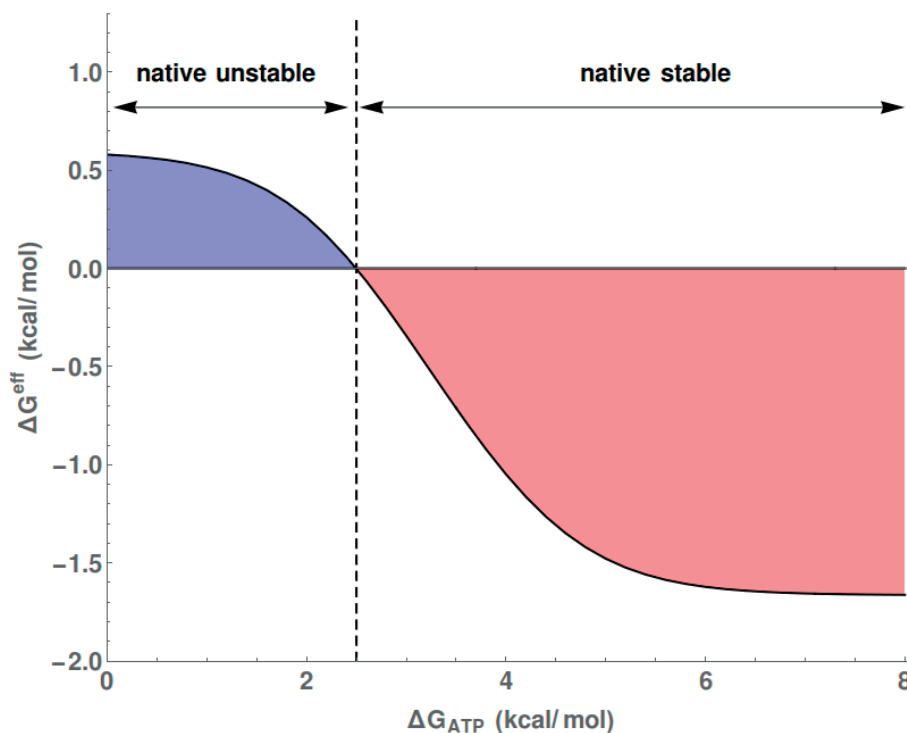
Figure 3.29 – Results of the rate model in the NESS. The nonequilibrium free energy of the native ensemble is plotted as a function of the energy that is available in the form of ATP.

The equilibrium stability of the native state can be tuned by changing, for example, the rate constant of misfolding. We define the free energy difference between the native and non native ensemble:

$$\Delta G_{nat}^{eq} = -k_b T \ln \left[\frac{[P_{nat}]_{eq}}{[P_{nonnat}]_{eq}} \right] , \tag{3.54}$$

which we can see as a quantitative estimation of the stress factors that inhibit a spontaneous refolding of the substrate. We were interested to see in which interval of ΔG_{nat}^{eq} there was a stronger non equilibrium stabilization of the functional state. In figure 3.30 we plot the fraction of native MDH as a function of ΔG_{nat}^{eq}. The gray region represents the gain in native structure due to the energy consumption of the ATP hydrolysis. For energies gaps larger than $5\,kcal/mol$ the substrate is so prone to the misfolded state that even in the presence of GroELS it is not possible to stabilize the native state. On the contrary, when $\Delta G_{nat}^{eq} < 2\,kcal/mol$ the functional state is stable also in equilibrium. For values of the free energy inside that interval GroELS is able to use the non equilibrium regime to significantly increase the yield of native proteins.
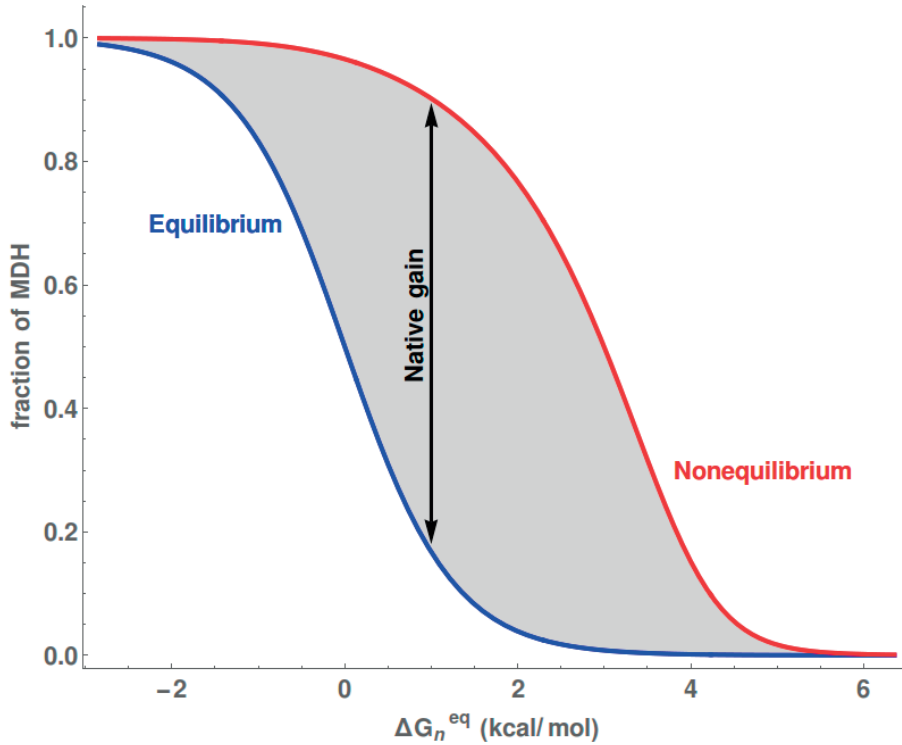
Figure 3.30 – Results of the rate model in the NESS. Fraction of native MDH as a function of the difference in free energy between native and misfolded ensembles.

3.2.3 Brief summary

We wanted to demonstrate that GroELS can stabilize the functional conformation of a given substrate protein even under stress conditions such that it would be normally found in a misfolded state. The substrate that was used, the enzyme MDH, was inactive at 37 °C. However, in the presence of GroELS and ATP the enzymatic activity was recovered, but only as long as ATP was available. The experimental results were reproduced by a rate model that includes aggregation, even though it turned out to be slow enough to be neglected in a first approximation. With an analysis of the non equilibrium stationary state we showed the dependence of the refolding activity of GroELS from the free energy of ATP and from the intrinsic stability of the substrate in equilibrium conditions.

3.3 Conclusions

Despite the great interest that chaperone systems have among biologists, there are still very few works in which thermodynamics and statistical physics are used to investigate their mechanism. With our work we tried to give our contribution to fill this gap. Thanks to the constant effort directed to the characterization of the structure and the dynamics of chaperone systems, lasted more than three decades, there are some general facts that are now universally accepted:

- There are particular conditions in which proteins are not capable to reach their functional structure by themselves.

- In these cases they can be helped, in some way, by molecular chaperones.

- In order to take part in the refolding process (regardless of what they actually do) chaperones must consume ATP.

Unfortunately, many questions concerning the facts just mentioned are still unanswered today. There is no general consensus on whether the aggregation is the only obstacle to proper folding in stress conditions or proteins can actually be more stable in a misfolded state independently of aggregation. It is still unclear how chaperones favor the refolding of their substrate: do they actively stimulates the refolding? Do they just break the non-native bonds leaving the protein the chance to refold spontaneously? Do they rather just prevent aggregation? Moreover, there is still the lack of an accepted model that is able to explain precisely how the hydrolysis of ATP is stimulated and in some cases there is not even an agreement on its consequence on the chaperone dynamics [1, 23].

Since our main goal was to determine what could be understood by simply considering basic thermodynamic principles, inevitably we had to abandon the level of molecular detail that is often used. However, we managed, in this way, to use an approach that could lead to more general results, that are true independently of the precise mechanism and sometimes even of the chaperone system under study. For example it was possible to deduce some

constraint on the rate constants, as well as to demonstrate that the hydrolysis of ATP is not just a contingent step in a complicated series of molecular transformation but rather a necessity that naturally follows from the rules of non equilibrium thermodynamics.

Despite that, we also obtained some results that correctly reproduce the experimental observations, both in the case the GroELS system and in the case of Hsp70.

4 Atp-binding cassette transporters

4.1 Introduction

In order to move through the cell membrane, small molecules usually need the help of specific membrane proteins called *transporters*. The Atp-binding cassette transporters (ABC) are one of the most abundant family of transporters [60, 61, 62]; they use the energy of ATP to drive the transport of several molecules, like lipids, sugars, sterols and vitamins. Moreover, they are of particular interest for biomedical applications, since it has been demonstrated [63, 64] that they can mediate the efflux of chemotherapeutic drugs from the cell, promoting in this way the multidrug resistance. ABC can be exporters or importers, but until now it has not been found any transporter that is able to perform both tasks [61]. The exporters are present in eukaryotes and prokaryotes, whereas the importers can only be found in prokaryotes [20, 65].

ABC are formed by two transmembrane domains (TMD, in yellow and cyan in figure 4.1), which are less conserved because they need to specifically bind to members of the huge variety of available substrate molecules, and two nucleotide binding domains (NBD, in purple and green in figure 4.1), more conserved, with the binding sites for ATP or ADP [66]. The crucial point of the mechanism of ABCs is the allosteric interaction between the NBD and the TMD, that is usually described with the so called *ATP-switch model* [61, 67]. In the model, the TMD can assume mainly two conformations: outward or inward facing. Experimental evidence suggests that for the majority of transporters when ATP is bound with the NBD the outward facing conformation is preferred, whereas when the ADP is bound the inward facing conformation is dominant [65, 67]. Therefore, the hydrolysis of ATP and the exchange are the reactions that enhance the conformational change in the ABC transporter.

4.1.1 Importers

In importers, the substrate is usually carried to the complex by a substrate binding protein (SBP) [68, 69]. For example, in the maltose uptake performed by the transporter MalFGK
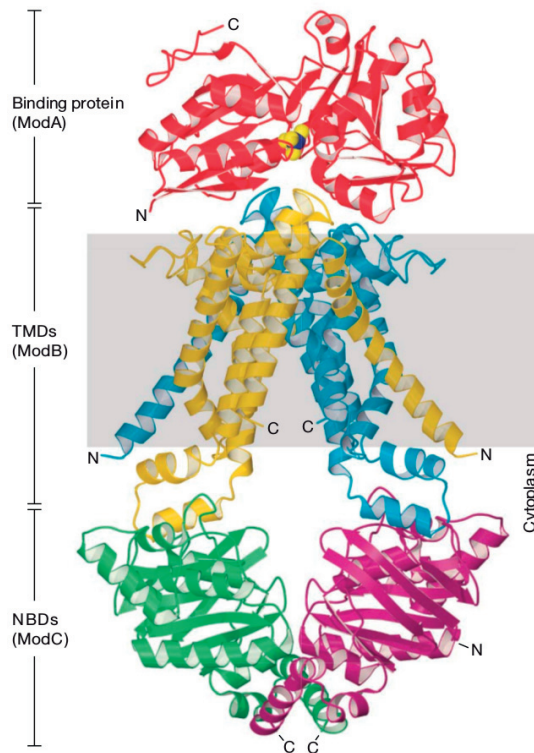
Figure 4.1 – The structure of an ABC importer, together with a binding protein (red). Figure taken from [66].

an important role is played by the maltose-binding protein expressed from the gene malE [70], which in addition to bringing the substrate to the transporter also stimulates the hydrolysis of ATP. Another very well known system in which the binding protein is extremely important is the BtuCD in *E. coli*, that assists the import of vitamin B_{12} with the help of the binding protein BtuF [71]. Usually, the loaded SBD interacts with the transporter when the latter is in a conformation that would not be optimal for the binding of the substrate alone, and then it stimulates the conformational change required to increase the binding affinity [65]. The PBD, the ATP and the substrate are also responsible for the closing of the NBD that stimulates the hydrolysis of ATP [70, 72]. After the hydrolysis of ATP, the ABC returns in the inward facing conformation [73], where it remains until the substrate is released.

4.1.2 Exporters

The exporters behave in a different way: the allosteric interaction between the TMD and the NBD is similar to the one of the importers, but in this case the substrate binds when the ABC is inward facing and unbinds when it is outward facing, therefore the reaction cycle is followed in the opposite direction with respect to the importers. The exporters work without the need of SBP. The binding of the substrate molecule to an inward facing

ABC increases the affinity between the NBD and ATP. The binding of ATP stimulates, together with the substrate, the closing of the NBD[61]. The direct consequence is an opening of the TMD toward the exterior of the cell. Then the substrate unbinds and only after that step the hydrolysis of ATP takes place [61]. However, it has been demonstrated [20, 63] that when a gradient in concentration directed toward the outer side of the cell is imposed, ATP-depleted exporters can mediate the uptake of molecules instead of the efflux.

In 2004, Balakrishnan and co-workers made interesting observations on a particular kind of exporter called LmrA [20]. They used ethidium as a substrate, which has the property of emitting fluorescent light when it binds to the DNA inside the cells. They imposed a concentration gradient in order to force a flux of ethidium in cells and proteoliposomes containing LmrA. For low concentrations of ATP, an increase in fluorescence was measured, indicating that the exporters mediate the uptake of ethidium, performing in this way the opposite task of what they normally do. However, as soon as glucose was added, so that the system could be fueled by the metabolic energy, there was a clear decrease in fluorescence: the usual activity of exporters was restored and the ABC returned to accelerate the efflux of the substrate (see figure 4.2). It is unlikely that the efflux was simply do to the binding of the nucleotides, because it was previously shown that the effect was not present when a non-hydrolyzable analogue of ATP, that is AMP-PNP, was used [63].
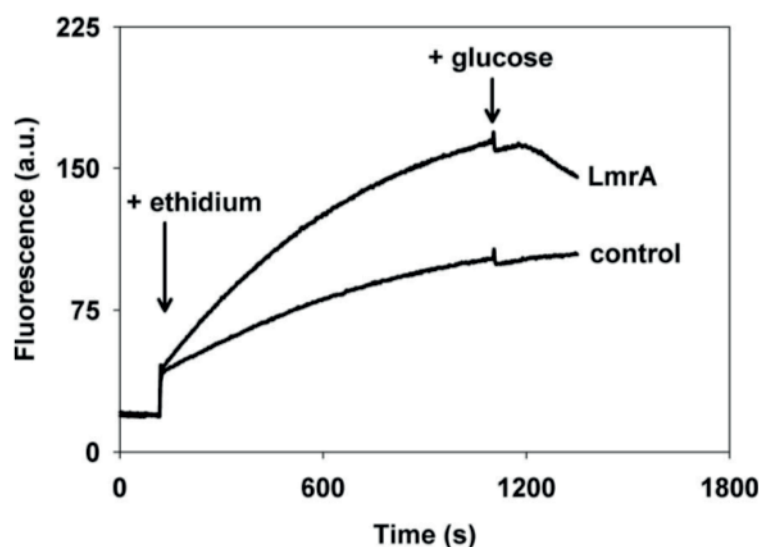
Figure 4.2 – When ethidium is added to a system containing ATP-depleted cells, there is an increase in fluorescence, indicating uptake of ethidium. After the addition of glucose the fluorescence starts decreasing. In the control cells there are no particular changes upon the addition of glucose. Figure taken from [20].

4.2   The model

The system is studied using a rate model with four states: the transporter can be in an ATP or an ADP state and in both cases it can be either bound with the substrate or free, as we schematically represent in figure 4.3. A simplification of this kind has already been used previously [67]. In the figure, we use a red color for the two reactions that contain the hydrolysis of ATP, because those are the reaction in which the source of energy is actually used. We can see that the apo state of the transporter has not been taken into account. Indeed, because of the large concentration of nucleotides, the vast majority of the transporters are bound and we can use effective exchange rates [21]. Those are the rates that are directly influenced by the energy consumption of the hydrolysis of ATP.
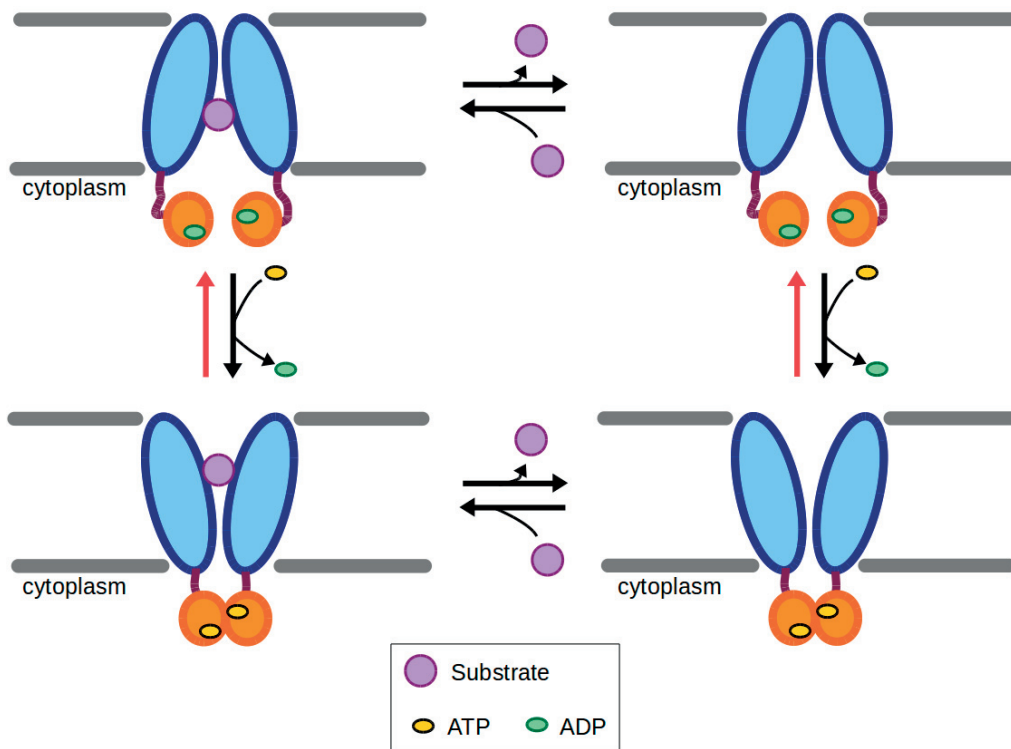
Figure 4.3 – Simplified version of the kinetic cycle of an ABC transporter in the model presented. We assume that the transporter has always an outer facing conformation when it is bound with ATP and an inner facing conformation when it is bound with ADP. The transmembrane domain is represented in blue and the nucleotide binding domain in orange.

With the binding reactions, the hydrolysis and the exchange, the model has in total 16 rate constants:

- k_{on}^{Tc} and k_{off}^{Tc} are the binding and unbinding rates to an ABC in the **ATP** state for a molecule coming from **inside** the cell;

- k_{on}^{Dc} and k_{off}^{Dc} are the binding and unbinding rates to an ABC in the **ADP** state for a molecule coming from **inside** the cell;

- k_{on}^{Tt} and k_{off}^{Tt} are the binding and unbinding rates to an ABC in the **ATP** state for a molecule coming from **outside** the cell;

- k_{on}^{Dt} and k_{off}^{Dt} are the binding and unbinding rates to an ABC in the **ADP** state for a molecule coming from **outside** the cell;

- k_h and k_s are the rates of **hydrolysis** and **synthesis** in the **absence** of the substrate;

- k_h^s and k_s^s are the rates of **hydrolysis** and **synthesis** in the **presence** of the substrate;

- k_{TD}^e and k_{DT}^e are the **exchange** rates from ATP to ADP and from ATP to ADP respectively, when the substrate is **not bound**;

- $k_{TD}^{e,s}$ and $k_{DT}^{e,s}$ are the exchange rates when the substrate is **bound**.

They are not independent, because they must respect the constraints imposed by the thermodynamics of equilibrium. In equilibrium the concentration of the molecules inside (indicated as c) is equal to the concentration of the molecules outside the cell (indicated as c). Therefore, the equilibrium conditions are

$$\gamma_{eq} = \frac{[c]_{eq}\, k_{on}^{Tc}\, k_h^s\, k_{off}^{Dt}\, k_s}{[t]_{eq}\, k_{on}^{Dt}\, k_s^s\, k_{off}^{Tc}\, k_h} = \frac{k_{on}^{Tc}\, k_h^s\, k_{off}^{Dt}\, k_s}{k_{on}^{Dt}\, k_s^s\, k_{off}^{Tc}\, k_h} = 1 \ , \tag{4.1}$$

$$\gamma_{eq} = \frac{[t]_{eq}\, k_{on}^{Tt}\, k_h^s\, k_{off}^{Dt}\, k_s}{[t]_{eq}\, k_{on}^{Dt}\, k_s^s\, k_{off}^{Tt}\, k_h} = \frac{k_{on}^{Tt}\, k_h^s\, k_{off}^{Dt}\, k_s}{k_{on}^{Dt}\, k_s^s\, k_{off}^{Tt}\, k_h} = 1 \ , \tag{4.2}$$

$$\gamma_{eq} = \frac{[c]_{eq}\, k_{on}^{Tc}\, k_h^s\, k_{off}^{Dc}\, k_s}{[c]_{eq}\, k_{on}^{Dc}\, k_s^s\, k_{off}^{Tc}\, k_h} = \frac{k_{on}^{Tc}\, k_h^s\, k_{off}^{Dc}\, k_s}{k_{on}^{Dc}\, k_s^s\, k_{off}^{Tc}\, k_h} = 1 \ , \tag{4.3}$$

where with $[c]_{eq}$ and $[t]_{eq}$ we indicate concentrations in equilibrium.

In our simplified model, the only difference between importers and exporters is the choice of the parameters that is influenced by the presence of the substrate molecule: in the case of importers the parameter is the hydrolysis rate, whereas in the case of exporters it is the binding of ATP [65, 61]. To the rate model we associate a system of master equations (see Appendix). The numerical solution with the parameters valid for the exporters was used to plot the concentration $[c]$ of molecules inside the cell as a function of time. Imitating the steps followed in the experiment with ethidium that we previously described, we started with a concentration of ATP that was similar to the one of thermodynamic equilibrium. The concentration of molecules inside the cell increased as expected. We then changed the value of the ratio $\alpha = \frac{[ATP]}{[ADP]}$ at a time $t = 1200s$, in order to reproduce the effect of the addition of ATP. As we can see in figure 4.4, for $t > 1200$ there is a decrease in the concentration of substrate inside the cell, indicating that the flux of molecules is directed outward, as it should, since we were using the parameters of exporters.

We then considered the non equilibrium stationary state. In our system the NESS is
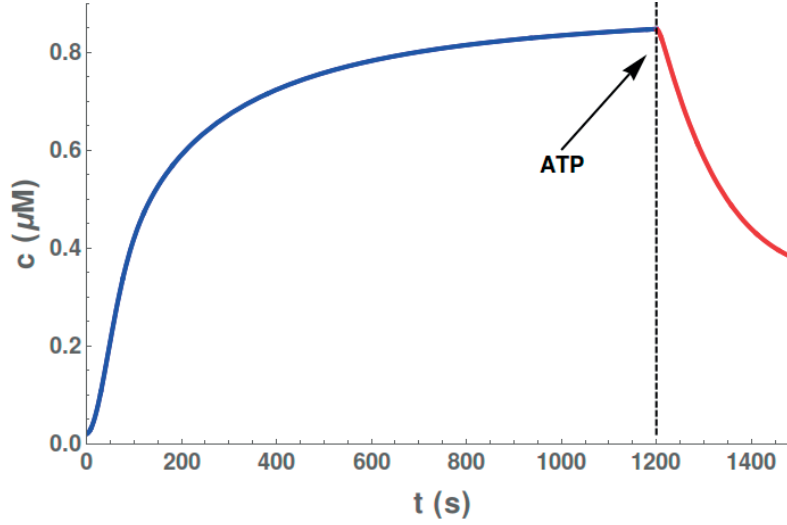
Figure 4.4 – The concentration of molecules inside the cell is shown as a function of time. For $t < 1200s$ equilibrium value $[ATP]/[ADP] \simeq 10$ is used. For $t > 1200s$ we used $[ATP]/[ADP] = 10$.

only due to a net flux of hydrolysis of ATP, whereas the total concentration of substrate molecules is fixed once for all and we assume that there are no exchanges of substrate molecules with the surrounding. Therefore, for the stationary state we imposed

$$J_c = 0, \tag{4.4}$$

where the flux J_c is defined as

$$J_c = [c] \left([T]k_{on}^{Tc} + [D]k_{on}^{Dc} \right) - \left([TS]k_{off}^{Tc} + [DS]k_{off}^{Dc} \right). \tag{4.5}$$

In figure 4.5 we show two phase diagrams in which the ratio $\omega = \frac{[c]}{[t]}$ is shown as a function of $\alpha = \frac{[ATP]}{[ADP]}$ and $[s_{tot}]$ in the case of exporters (left) and importers (right). We can see that an increase in α leads to an increase in the chemical gradient between the two sides of the membrane. This effect is larger for smaller values of the total concentration of the substrate molecules and if they are in large excess the transporters can not work properly, leaving the ratio ω close to 1. It is interesting to see some asymptotic behaviors: first of all, we can determine a limiting case for optimal hydrolysis rate and optimal binding rate from the inner side of the cell. Indeed, in these conditions we have:

$$\omega \to \frac{A\frac{[ADP]_{eq}}{[ATP]_{eq}} + B}{A\frac{[ADP]}{[ATP]} + B}, \tag{4.6}$$

where A and B are two positive constants (see Appendix). Here, with $[ADP]_{eq}$ and $[ATP]_{eq}$ we indicate the concentrations in equilibrium conditions, that are valid for an isolated system after the relaxation time. The implication of 4.6 is that the following inequality
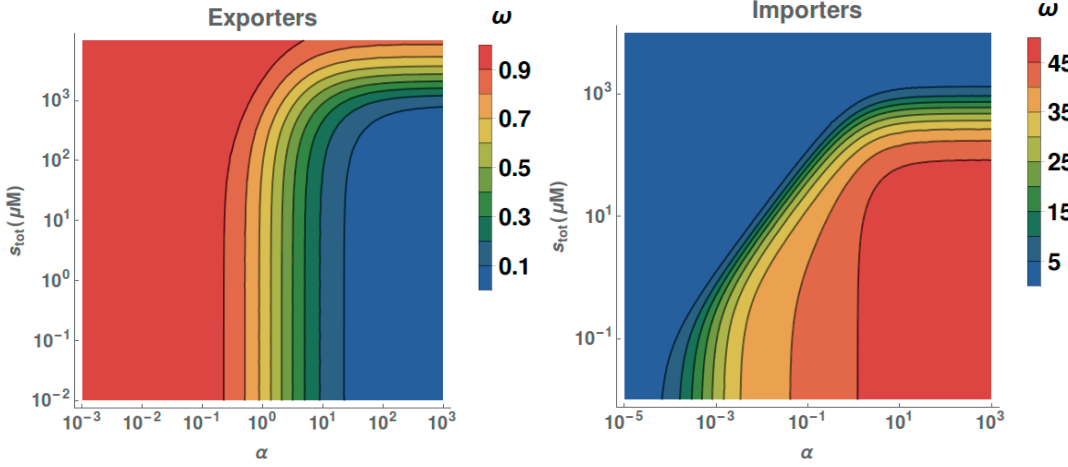
Figure 4.5 – Contour plot showing the ratio $\omega = \frac{[c]}{[t]}$ corresponding to values of $\alpha = \frac{[ATP]}{[ADP]}$ and $[s_{tot}] = [c] + [t]$, in the case of exporters (left) and importers (right).

holds:

$$\omega \leq \frac{[ADP]_{eq}[ATP]}{[ATP]_{eq}[ADP]}. \tag{4.7}$$

The inequality 4.7 gives a boundary to the possible concentration gradient that can be achieved with a given amount of available free energy. Note that it is a thermodynamic limit that is valid regardless of the intrinsic properties of the ABC system. If instead we assume that the available energy is infinite, that is

$$\frac{[ATP]}{[ADP]} \to \infty, \tag{4.8}$$

we have the following asymptotic behavior

$$\omega \to \frac{k_{on}^{Tt} k_{off}^{Dc}}{k_{off}^{Tt} k_{on}^{Dc}}. \tag{4.9}$$

Therefore, in that case the efficiency of the transporter is uniquely determined by the binding rates, that are the time limiting steps in the reaction.

4.3 Conclusions

The non equilibrium thermodynamic formalism is particularly fit for the description of the mechanisms of the ABC transporters, in which the prominent role of the ATP consumption has been highlighted even during experiments [20]. We have shown that the requirement for the energy consumption is not a simple mechanistic tool, but rather a constraint that naturally comes from basic principles of thermodynamics. With our coarse-grained model it has been possible to reproduce qualitatively an experimental result. It has been done

starting from a simple assumption that was already made in other works [67], namely that in the case of exporters the substrate is able to stimulate the exchange by interacting with the NBD [61], whereas in the case of importers it is rather the hydrolysis of ATP to be accelerated. In addition, we obtained the phase diagrams for importers and exporters. In both cases, the transporters work in a more efficient way for large concentrations of ATP and for low concentrations of substrate molecules.

5 Shape of a stretched polymer

The main results presented in this chapter have been published in [74]. Many of the figures have been adapted from the article, in accordance with the Creative Commons Attribution License used by Physical Review Letters.

Introduction

In the previous chapters we have discussed the mechanisms and the interactions that lead polymer chains to reach their functional state, characterized by a very complex and unique tertiary structure. Now we would like to study their conformation when they are instead in an unfolded state.

The shape of a polymer chain is strongly dependent both on the intrinsic characteristics of the chain, like the rigidity and all the interactions among its monomers, and on the properties of the solvent. Roughly speaking, we can say that the main classification of different types of polymers is determined on the basis of how much the monomers are attracted to the molecules of the solvent with respect to their interaction with one another. The parameter that is commonly used for this classification is the second virial coefficient, defined as follows [5]:

$$v = \int d^3r \left(1 - \exp\left[-\beta U(r)\right]\right) , \tag{5.1}$$

where $U(r)$ is the energy necessary to bring two monomers to a distance r from very far away and $\beta = (kT)^{-1}$ is the thermal energy. When $v < 0$ the attractive interaction among the monomers is stronger and therefore the polymer tends to assume a more compact shape. When $v > 0$ the monomers tend to stay far from one another and the solvent is usually called *good*. If the monomers repel each other as if they were hard spheres the virial coefficient is precisely the excluded volume. When instead $v = 0$, the molecules of the solvent perfectly compensate the interaction among the monomers and the polymer behaves as an ideal chain. We are in the regime of θ-*solvent*. This condition is of particular interest because the polymer chain can be treated with the same formalism that is used

for the random walk. When this is possible, the chain is usually called *freely-jointed chain* (FJC).

In this chapter we will discuss some results regarding the conformational shape of a polymer chain when it is pulled by an external force, a condition that is often encountered both *in vivo*, for example in protein translocation across the cell membrane [26, 16], and *in vitro*, e.g. in pulling experiments performed with optical or magnetic tweezers [75]. We will show that the transverse section of the ellipsoid that approximates the polymer chain shrinks isotropically after the orientation. We will then show that there is a new model independent behavior of some main quantities when they are written in terms of the contribution of the force to the free energy.

Next section will be devoted to a brief summary of the main results in polymer physics. In the following section we will derive the end to end vector and the radius of gyration for a FJC under tension. In section three we will study the polymer shape in more details, with the inertia tensor and its eigenvalues. We will also show the exact value of the asphericity as a function of the force. In section four we will extend the discussion to models that are different from the FJC, discussing also some universal behavior that is valid for models with rigidities and excluded volume.

5.1   State of the art

The simplest way to model a polymer chain is to assume that each monomer behaves independently of all the others and has a fixed length. In order to study the shape of the chain in three dimensions the quantity that comes more natural to define is the end-to-end vector

$$\mathbf{R}_e \equiv \sum_{i=1}^{N} \mathbf{r}_i \,, \tag{5.2}$$

where $\mathbf{r}_i$ is a tangent vector corresponding to the i-th monomer and N is the number of monomers. A well known result of polymer physics is that the probability distribution of $\mathbf{R}_e$ for a polymer of N monomers of length b is a normal distribution:

$$P\left(\mathbf{R}_e\right) = \left(\frac{3}{2\pi N b^2}\right)^{3/2} \exp\left[-\frac{3}{2}\frac{R_e^2}{N b^2}\right] \,. \tag{5.3}$$

As a consequence, using spherical coordinates it is easy to show that the distribution of the distance R_e is

$$P\left(R_e\right) = 4\pi R_e^2 \left(\frac{3}{2\pi N b^2}\right)^{3/2} \exp\left[-\frac{3}{2}\frac{R_e^2}{N b^2}\right] \,. \tag{5.4}$$

Formula 5.4 implies that the mean square of the end to end vector is proportional to the number of monomers:

$$\left\langle R_e^2 \right\rangle = b^2 N \,. \tag{5.5}$$

Another very important parameter is the square of the radius of gyration [5], that is the average distance of the monomers from the position of the center of mass of the chain:

$$R_g^2 \equiv \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{R}_i - \mathbf{R}_{cm} \right)^2 \,, \tag{5.6}$$

where

$$\mathbf{R}_i = \sum_{j=1}^{i} \mathbf{r}_j \qquad \text{for } i \neq 0 \,, \tag{5.7}$$

and

$$\mathbf{R}_{cm} = \frac{\sum_{i=1}^{N} \mathbf{R}_i}{N} \,. \tag{5.8}$$

Roughly speaking, the radius of gyration is the radius of the sphere that best approximates the polymer chain. In a freely-jointed chain the mean square of the radius of gyration is[1]

$$\left\langle R_g^2 \right\rangle = \frac{Nb^2}{6} \,, \tag{5.9}$$

smaller than $\langle R_e^2 \rangle$. As this result suggests, there must be a significant asymmetry in the shape of an ideal chain, since its ends are often outside the sphere enclosing the majority of the monomers. Obviously, if we overlap a large number of random chains we obtain a symmetric object, as we can see in figure 5.1 (left), because no direction is favored with respect to the others. The asymmetries are capture by the inertia tensor:

$$\mathcal{T}_{\alpha\beta} = \frac{1}{N} \sum_{i=1}^{N} \left(\alpha_i - \alpha_{cm} \right) \left(\beta_i - \beta_{cm} \right) \,, \tag{5.10}$$

where α and β indicate space coordinates. The trace of the inertia tensor is the square of the radius of gyration. The eigenvalues of the inertia tensor are proportional to the average value of the square of the axes of the ellipsoid that best approximate the shape of the polymer chain, and the eigenvectors determine the orientation of the ellipsoid. In figure 5.1 (right) we show the same chains as before, with the only difference that they have been rotated toward their principal axes before the overlap. The eigenvalues of the inertia tensor in a FJC are in ratios 11.8 : 2.7 : 1 [76].

---

[1]To simplify the notation, in many cases we will omit the brackets for the average values.
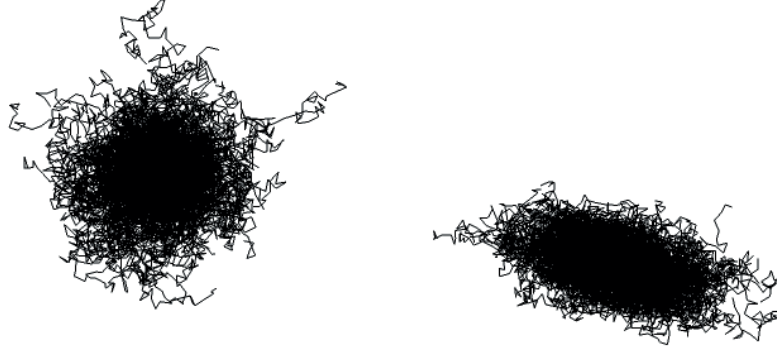
Figure 5.1 – (left) Superposition of random chains. (right) Superposition of the same chains after a proper orientation.

The parameter that is commonly used to quantify the asymmetry is the *asphericity*:

$$\mathcal{A} \equiv \frac{\sum_s \left\langle \left(\lambda_s - \bar{\lambda} \right)^2 \right\rangle}{6 \left\langle \bar{\lambda}^2 \right\rangle} , \tag{5.11}$$

where λ_s are the eigenvalues and $\bar{\lambda} = \frac{\sum_s \lambda_s}{3}$. Note that $\mathcal{A} = 0$ in the case of perfect spherical symmetry, where all the eigenvalues are equal, and $\mathcal{A} = 1$ in the case of a rigid rod, with $\lambda_2 = \lambda_3 = 0$. In a freely jointed chain $\mathcal{A} = \frac{10}{19} \simeq 0.526$ [77].

There are two main features that are usually added to a FJC in order to reproduce more faithfully the behavior of real chains: monomers interactions and bending rigidity. For the former, the potential that is normally used is the Lennard Jones potential, which has the form:

$$V_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{x} \right)^{12} - \left(\frac{\sigma}{x} \right)^6 \right] , \tag{5.12}$$

where x is the distance between two monomers and σ and ϵ are two constant parameters. The first term in the square brackets is due to the very short range repulsion due to the Pauli exclusion principle and the second term to the attractive interaction at long ranges (for example the van der Waals force). In our case we used a truncated Lennard Jones potential and for all practical purposes the monomers behaved as if they were identical hard spheres. We will refer to polymers of this kind, namely in which a couple of monomers cannot be at a distance lower than $2b$ but there are no long range interactions, with the term *self avoiding walk* (SAW). For the description of SAWs a mean field model was developed by Flory in the 50s [78], consisting in writing a free energy that is the sum of

the entropic term that we would have even in the Gaussian regime and a term that takes into account the hard sphere repulsion in the same way it would treat it for a gas confined in a volume $w \sim R^3$, where R is the size of the chain:

$$F = a\frac{R^2}{N} + b\frac{N^2}{R^3} \,, \tag{5.13}$$

where a and b are two constant parameters. Minimizing the free energy with respect to R we obtain the scaling law

$$R \sim N^{3/5} \,. \tag{5.14}$$

Given the simplicity of the model, the critical exponent $\nu = \frac{3}{5} = 0.6$ has a remarkable accuracy, since the correct exponent is $\nu \simeq 0.588$ [5]. Self avoiding polymers are on average more elongated than FJCs: the ratios between the eigenvalues are [79] 14.05 : 2.96 : 1 and the asphericity is $\mathcal{A} \simeq 0.55$.

For the bending rigidity a correlation between neighboring tangent vectors is introduced, usually with a potential of the form

$$V = -\frac{B}{b} \sum_{i=1}^{N-1} \mathbf{r}_i \cdot \mathbf{r}_{i+1} = -\frac{B}{b} \sum_{i=1}^{N-1} \cos\theta_i \,, \tag{5.15}$$

where B is a constant parameter and θ_i is the angle between $\mathbf{r}_i$ and $\mathbf{r}_{i+1}$. This is known as the *Kratky-Porod model*. In most cases the cosine is expanded and an harmonic potential is used:

$$V = k \sum_{i=1}^{N-1} \theta_i^2 \,, \tag{5.16}$$

where k is a new parameter that from now on will be called *bending constant*. The correlation between tangent vectors can be quantified using the persistence length [5] :

$$l_p = -\frac{1}{\ln\left[\langle\cos\theta_i\rangle\right]} \,. \tag{5.17}$$

Note that in our case, since the monomers are considered to be identical, l_p is just a number that does not depend on the position along the chain. It can be shown that the persistence length is related to the bending constant via the relation

$$l_p \simeq \frac{2bk}{k_bT} \,. \tag{5.18}$$

Concerning polymer stretching, there are some known result we need to recall. We will show in the next section that when a FJC is subjected to an external force $\mathbf{f} = f\hat{\mathbf{e}}_z$ its

mean extension is

$$\langle \mathbf{R}_z \rangle = Nb\,\mathcal{L}(\gamma)\,\mathbf{e_z}\,, \tag{5.19}$$

where

$$\mathcal{L}(\gamma) \equiv \coth(\gamma) - 1/\gamma \tag{5.20}$$

is the Langevin function and

$$\gamma \equiv \beta f b\,. \tag{5.21}$$

However, if the bending rigidity is taken into account, the force vs extension relation becomes [80]

$$\frac{fl_p}{k_{\mathrm{b}}T} = \frac{R_z}{L} + \frac{1}{4\left(1 - \frac{R_z}{L}\right)^2} - \frac{1}{4}\,, \tag{5.22}$$

where L is the contour length. In figure 5.2 we compare the curve relative to the formula we just wrote with the Langevin function.
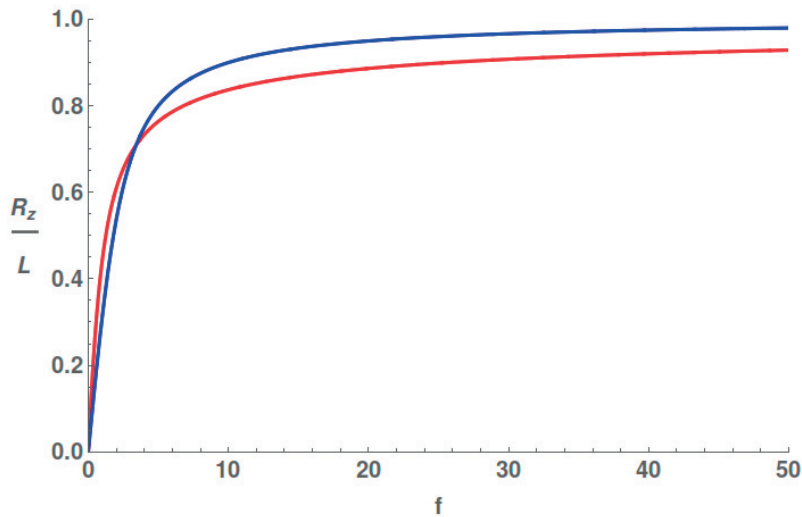

Figure 5.2 – Comparison of the force vs extension curve in the case with (red) and without rigidity (blue).

5.2 Exact Computation of $\langle R_e^2 \rangle$ and $\langle R_g^2 \rangle$

The force applied to the polymer chain corresponds to an internal energy $E = -\mathbf{f} \cdot \mathbf{R}_e$ and thus to a Boltzmann factor

$$\exp\left[-\beta E\right] = \exp\left[\beta \mathbf{f} \cdot \mathbf{R}_e\right] = \exp\left[\sum_{i=1}^{N} \beta \mathbf{f} \cdot \mathbf{r}_i\right] = \prod_{i=1}^{N} \exp\left[\beta \mathbf{f} \cdot \mathbf{r}_i\right]. \tag{5.23}$$

Therefore, the force is coupled separately with each tangent vector and we can calculate the partition function for a single tangent vector, that is [81]

$$q(\beta f b) = \int d\boldsymbol{r} \exp(\beta \boldsymbol{f} \cdot \boldsymbol{r_i}) \delta\left(|\boldsymbol{r_i}| - b\right) =$$
$$= 4\pi b^2 \sinh(\beta f b)/\beta f b, \tag{5.24}$$

where we used a Dirac delta to impose a constraint on the length of the tangent vector. The partition function of the whole system is just q^N. We now assume that a force is oriented along the z axis. Using the partition function, we can derive the known result of the mean extension

$$\langle \mathbf{R}_e \rangle = N b \, \mathcal{L}(\gamma) \, \mathbf{e_z}, \tag{5.25}$$

In the same way [2], it is possible to calculate the variances of the three components of $\mathbf{R}_e$, that are

$$\sigma_x^2 = \sigma_y^2 = N b^2 g, \tag{5.26}$$
$$\sigma_z^2 = N b^2 a, \tag{5.27}$$

where $g \equiv \mathcal{L}/\gamma$ and $a \equiv 1 - 2g - \mathcal{L}^2$. The end to end vector is a combination of variables that are independent and identically distributed, therefore it follows the central limit theorem and its mean square is given by

$$\left\langle \mathbf{R}_e^2 \right\rangle = \langle \mathbf{R}_e \rangle^2 + \sigma^2 = N b^2 (1 - \mathcal{L}^2) + N^2 b^2 \mathcal{L}^2. \tag{5.28}$$

As we can see, the end to end vector is the sum of two contributions, modulated by the term $\mathcal{L}^2$. For low forces, the unstretched FJC dominates, whereas as the force increases the end to end vector becomes similar to the one of a rigid rod, that is $R_e \simeq N b$. The force at which the second component becomes larger than the first one is

$$f = \frac{kT}{b} \frac{3}{\sqrt{N+1}}. \tag{5.29}$$

The expression of the end to end vector can be used for the derivation of the radius of gyration. First of all, we note that it can be written in terms of the distance between

---

[2]This derivation, like many others in this chapter, has been published in [74].

couples of monomer along the chain:

$$
\left\langle R_g^2 \right\rangle = \left\langle \frac{1}{N} \sum_{i=1}^{N} (\mathbf{R}_i - \mathbf{R}_{cm})^2 \right\rangle =
$$
$$
= \left\langle \frac{1}{2N^2} \sum_{i,j} (\mathbf{R}_i - \mathbf{R}_j)^2 \right\rangle = \tag{5.30}
$$
$$
= \frac{1}{N^2} \sum_{j>i} \left\langle (\mathbf{R}_i - \mathbf{R}_j)^2 \right\rangle .
$$

The average distance between two position vectors is equal to the end to end vector of the subchain between them. Therefore, we can calculate it using 5.28. With this method we obtain

$$
\left\langle R_g^2 \right\rangle = \frac{1}{6} N b^2 + \frac{1}{12} N^2 b^2 \mathcal{L}^2 . \tag{5.31}
$$

Note that even in the case of the radius of gyration the two regimes are weighted by $\mathcal{L}^2$.

The increase in the radius of gyration for larger forces does not give any information about the directionality of the elongation. Obviously, we expect that when the forces are large enough the elongation takes place only in one preferred direction and beyond a certain threshold there should be a shrinking in the section orthogonal to the force. In order to give a more insightful description we need to use the inertia tensor.

5.3 Inertia tensor and asphericity

We implemented a FJC for different values of N and different forces using Monte Carlo simulations (MC). For each chain we calculated the inertia tensor and its eigenvalues and eigenvectors. There are essentially two scales of the force that must be considered. When the force is small, the polymer is slightly deformed even though it is not oriented in the same direction as the force. As soon as the force becomes large enough to orient the polymer in its direction there is an elongation in that direction and a shrinking in the other two. The orientation can be represented with the projection of the force on the eigenvectors. In figure 5.3 we can see that the eigenvector $\mathbf{v}_1$ relative to the largest eigenvalue starts at random orientation, with projection $\langle \cos \psi_1 \rangle = 0.5$, where ψ_1 is the angle between $\mathbf{v}_1$ and $\mathbf{f}$, and then it rotates until it is parallel to the force.
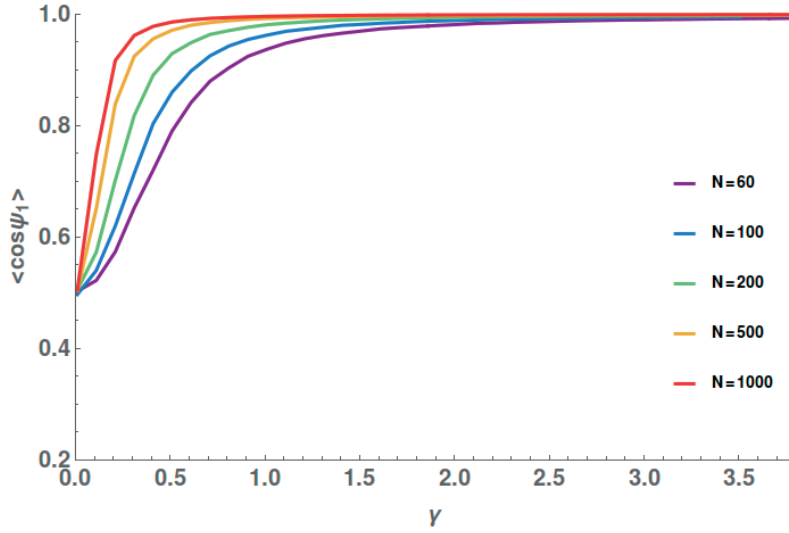
Figure 5.3 – Projection of the first eigenvector to the force, for five different values of N.

As we should expect, longer polymer are easier to rotate. This is not surprising: the expression

$$\langle \mathbf{R}_e \rangle = Nb\,\mathcal{L}\,\mathbf{e_z} \tag{5.32}$$

for small forces become

$$\langle \mathbf{R}_e \rangle \simeq \frac{Nb\gamma}{3}\,, \tag{5.33}$$

meaning that the polymer behaves as a spring with force

$$\mathbf{f}_{polymer} = -\frac{3kT\,\langle \mathbf{R}_e \rangle}{Nb}\,, \tag{5.34}$$

and thus with an elastic constant that decreases as N increases:

$$K = \frac{3kT}{Nb}\,. \tag{5.35}$$

As we said, the trace of the inertia tensor, and thus the sum of its eigenvalues, must be equal to the radius of gyration. This property allowed us to make some predictions about the behavior of the eigenvalues, because we do have the analytic expression of R_g. When the force is large enough, the radius of gyration should be close to the largest eigenvalue, since the polymer is already oriented in the direction of the external force. As a consequence, we expect $\lambda_1 \simeq \frac{N^2 \mathcal{L}^2}{12}$, namely the largest eigenvalue is almost equal to the square of the radius of gyration. This is indeed the case, as we can see in figure 5.4, where the renormalized eigenvalue $\frac{\lambda_1}{N^2 b^2}$ is plotted as a function of γ for different values of N. The black dashed line represents the curve $h(\gamma) = \frac{\mathcal{L}^2}{12}$. Note that the curves for different values of N do not perfectly collapse, but they get closer to $h(\gamma)$ as N is increased. Indeed, for larger N the

critical force at which the rod-like behavior becomes dominant is smaller.
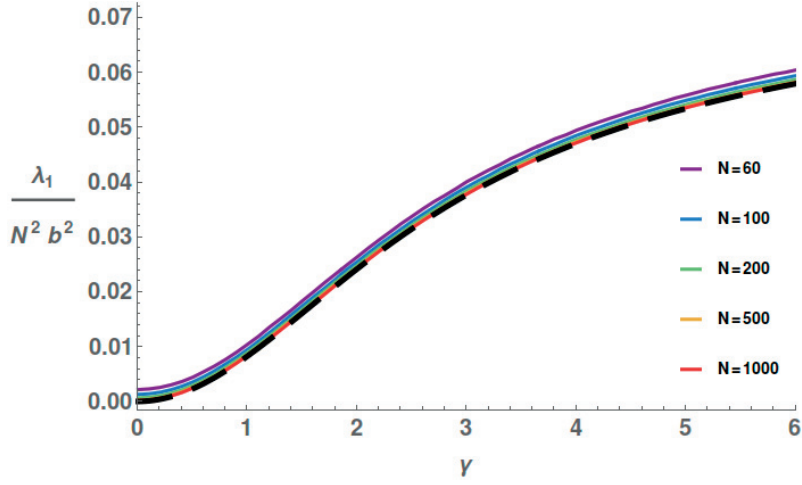
Figure 5.4 – Renormalized eigenvalue λ_1 as a function of the force, for five different values of N. The black dashed line is the curve $h(\gamma) = \frac{\mathcal{L}^2}{12}$.

In the same regime, the other two eigenvalues should instead depend on the force in the same way as the variances of the end to end vector in the directions orthogonal to the force, which means $\lambda_{2,3} \sim N\mathcal{L}/\gamma$. If we consider figure 5.5 and 5.6, we can see that there is an initial increase in the value of the mean eigenvalues for small forces.
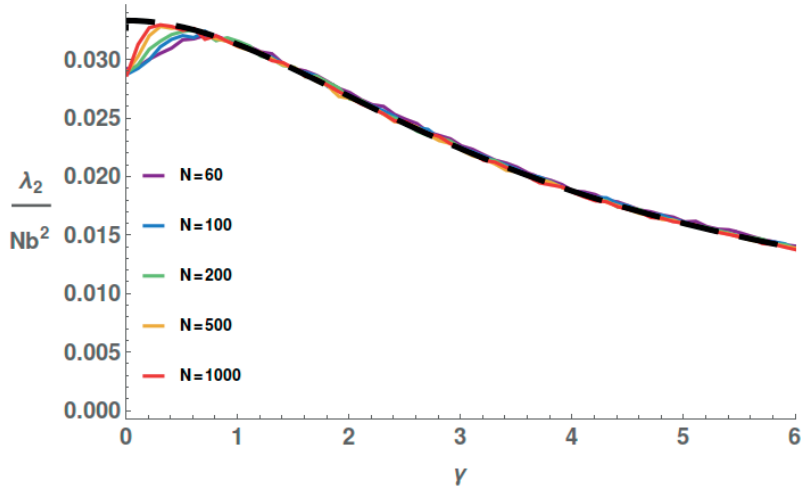
Figure 5.5 – Renormalized eigenvalue λ_2 as a function of the force, for five different values of N. The black dashed line is the curve $h(\gamma) = 0.1\frac{\mathcal{L}}{\gamma}$.

This regime indicates forces that are low enough that the polymer is not well oriented, therefore the elongation takes place in a random direction, increasing, on a statistical point of view, all the eigenvalues. After the orientation, the stretching is directed along the major axis of the ellipsoid, and the smaller eigenvalues decrease as predicted. Fitting with curves of the form $\lambda_i = c_i N\mathcal{L}/\gamma$, we obtained the parameters $c_2 = 0.1$ and $c_3 = 0.034$ for

the second and the third eigenvalue respectively. As the results suggest, for large forces
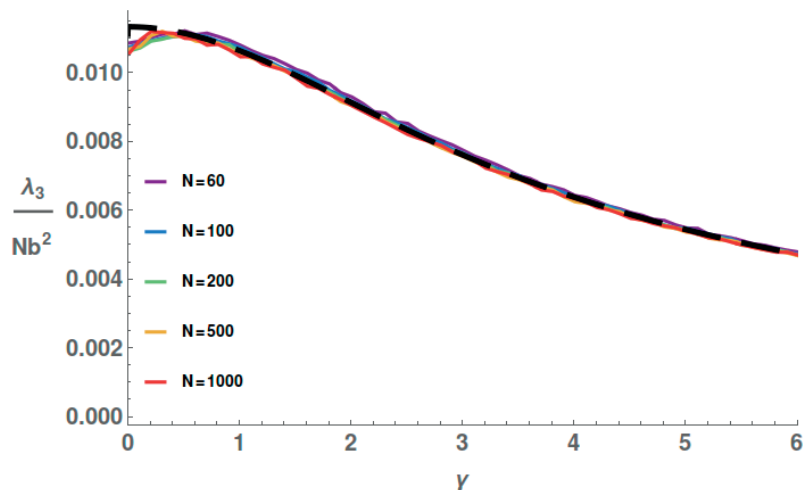
Figure 5.6 – Renormalized eigenvalue λ_3 as a function of the force, for five different values of N. The black dashed line is the curve $h(\gamma) = 0.034\frac{\mathcal{L}}{\gamma}$.

the ratio between the two smaller eigenvalues reaches a plateau that is independent on the force but that is different from the case at zero force (figure 5.7).
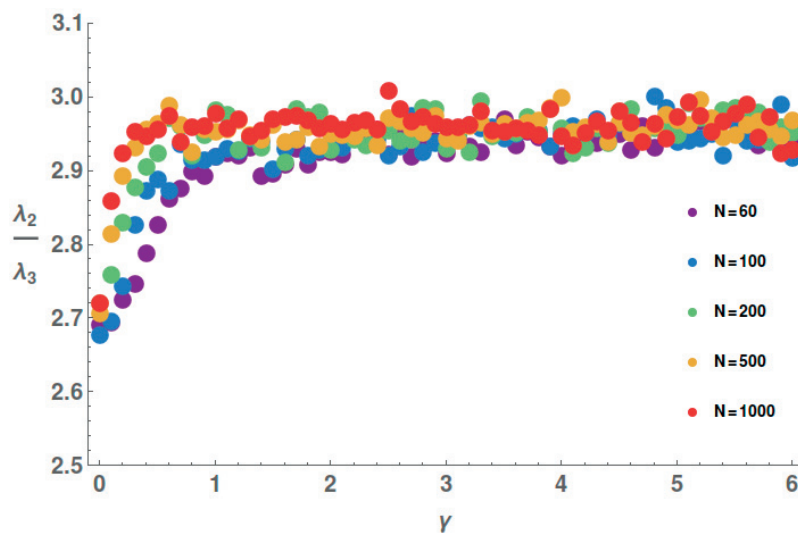
Figure 5.7 – Ratio between the two smaller eigenvalues as a function of the force, for five different values of N.

With the MC is also possible to calculate the asphericity as a function of γ and N. However, we calculated its expression analytically using the same method that was used for the

radius of gyration [3]. The exact formula is the following:

$$\mathcal{A} = 1 - \frac{72ag + 36g^2 + 24Ng\mathcal{L}^2}{36a^2 + 80ag + 112g^2 + 4\mathcal{L}^2\left(11a + 10g\right)N + 5N^2\mathcal{L}^4} \, . \tag{5.36}$$

We show the curve in figure 5.8. The derivative of the asphericity with respect to γ is a
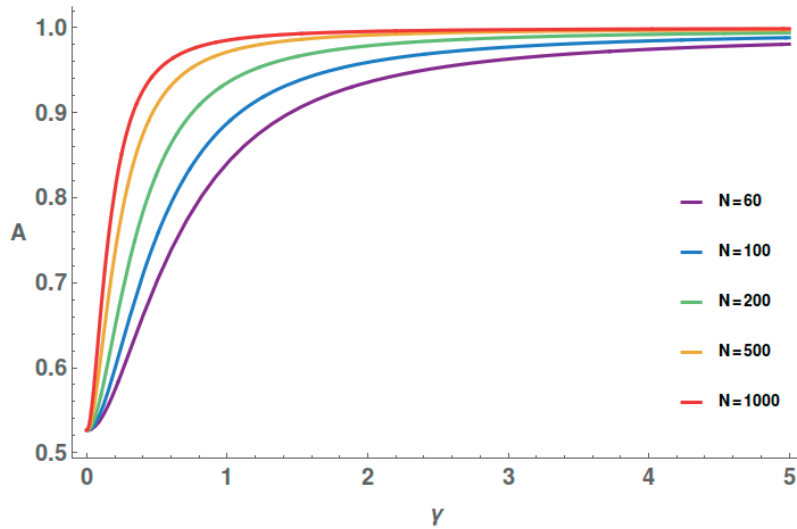
Figure 5.8 – Asphericity as a function of γ for different values of N.

measure on the sensibility of the polymer chain to a change in the external force. As we can see in figure 5.9, the maximum of this value is at lower forces for larger N.
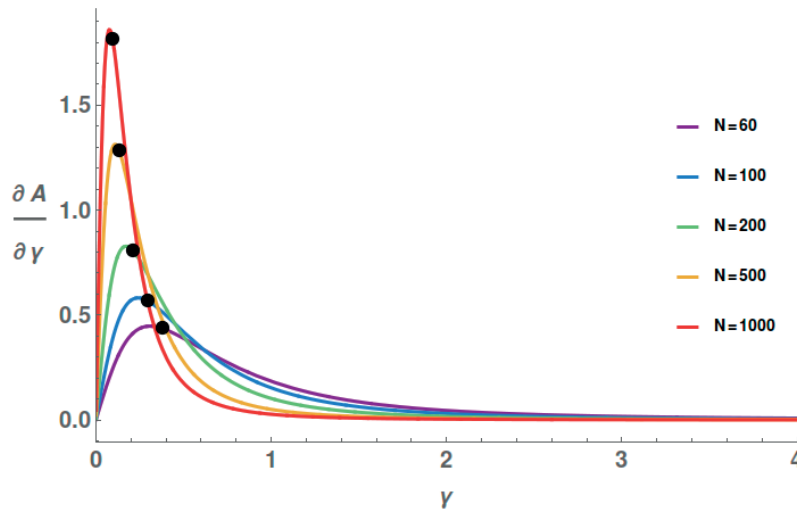
Figure 5.9 – Derivative of the asphericity as a function of γ for different values of N. Black dots correspond to $f = \frac{3}{N+1}$.

We indicated with black dots the values corresponding to the critical force at which the

---

[3]the whole calculation can be found in the supplementary information of [74].

rod-like behavior starts being dominant with respect to the one of unbiased Gaussian chain: $f = \frac{3}{\sqrt{N+1}}$. As we should expect, it is very close to the maximum.

What we found interesting of formula (5.36) were the two asymptotic behaviors, that are

$$\mathcal{A} \simeq 1 - \frac{24}{5N\gamma\mathcal{L}}\,, \tag{5.37}$$

for large forces and

$$\mathcal{A} \simeq \frac{10}{19} + \frac{25}{1083}N\gamma^2\,, \tag{5.38}$$

for small forces. The reason why we found the result interesting is that for low forces we have $N\gamma^2 \sim N\gamma\mathcal{L}$. Therefore, both the asymptotic behaviors depend on the combination $\eta \equiv N\gamma\mathcal{L}$, which is nothing but the contribution of the force to the free energy, since $N\gamma\mathcal{L} = \mathbf{f} \cdot \langle \mathbf{R_e} \rangle / k_b T$. This suggested that the whole asphericity could be written, in first approximation, as a function of the sole parameter η. Indeed, after a bit of algebra and keeping only the leading terms, we have

$$\mathcal{A} \simeq \frac{120 + 60\eta + 5\eta^2}{228 + 84\eta + 5\eta^2}\,. \tag{5.39}$$

5.4 Universal behaviors

When we obtained formula (5.39) we found it natural to investigate whether the result was somehow more general and applicable to other models or it was instead limited to the very specific case of FJC. Models with excluded volume or rigidity were too complicated to be treated analytically. We implemented molecular dynamics (MD) simulations for models in which the rigidity of the polymer and the excluded volume are taken into account. In figure 5.10 we plot the asphericity as a function of η for different parameters of the rigidity, with and without volume exclusion and with $N = 10^3$. We can see that in first approximation there is a collapse onto the same curve for models in which the excluded volume is taken into account (E) and in which it is not (P), and for every value of the rigidity. To be precise, for very small values of the force there is a small difference between the self-avoiding walk (SAW) and the other models since the latter is slightly more elongated. Increasing the force, the effect of the excluded volume becomes less important and the SAW is even closer to the other models. Therefore, we can conclude that in a θ-solvent it is possible to infer the asphericity of a polymer from its end to end vector, that is the accessible quantity in pulling experiments.

A similar result can be found for the projection of the eigenvectors on the direction of the force. As we have previously pointed out, for large forces the mean largest eigenvalue goes
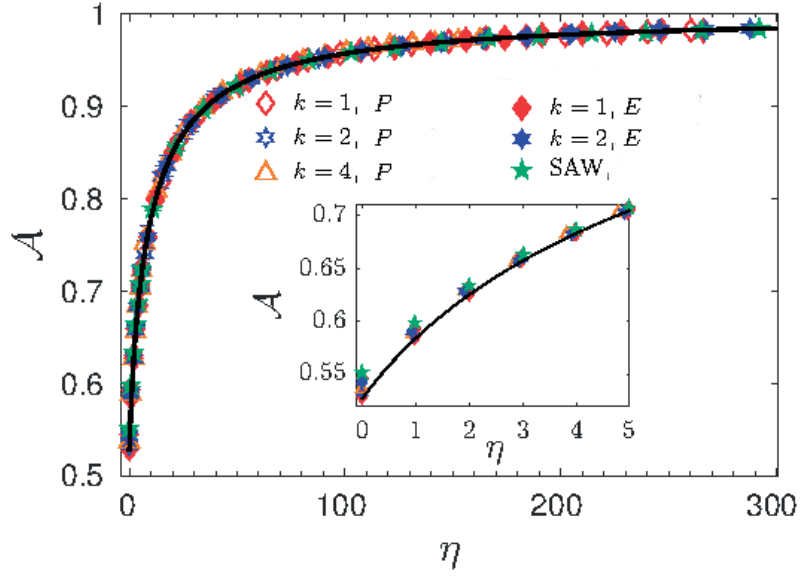
Figure 5.10 – (main) Asphericity as a function of η for different models. (inset) Zoom for small values of the force. The black curve corresponds to equation 5.39. Image taken from [74].

as

$$\lambda_1 \simeq \frac{(N\mathcal{L}b)^2}{12}\,. \tag{5.40}$$

We must now find the projection of the vector $\lambda_1 \mathbf{v}_1$ on the xy plane. We can do this by taking the contribution to the radius of gyration that are orthogonal to the force, that are

$$\frac{1}{(N+1)^2} \sum_{j>i} \left\langle (X_j - X_i)^2 \right\rangle + \left\langle (Y_j - Y_i)^2 \right\rangle \simeq \frac{1}{3} N b^2 g\,, \tag{5.41}$$

and subtract the part of the two smaller eigenvalues, that has been discussed above. In this way we obtain

$$(\lambda_1 \mathbf{v}_1 \cdot \hat{\mathbf{x}})\,\hat{\mathbf{x}} + (\lambda_1 \mathbf{v}_1 \cdot \hat{\mathbf{y}})\,\hat{\mathbf{y}} \simeq N b^2 g \left(\frac{1}{3} - c_2 - c_3 \right) \simeq \frac{N b^2 g}{5}\,. \tag{5.42}$$

Therefore the mean value of $\sin \psi_1$ is

$$\langle \sin \psi_1 \rangle \simeq \sqrt{\frac{N b^2 g}{5 \lambda_1}} \simeq \sqrt{\frac{12}{5 N \gamma \mathcal{L}}} = \sqrt{\frac{12}{5 \eta}}\,. \tag{5.43}$$

The behavior of the eigenvalues suggested an analogy with the electric dipole, when it is immersed in a constant electric field. Here, we are slightly giving up the mathematical rigor, since the analogy makes sense only if the dipole is already partially oriented in the correct direction, otherwise its orientation would be completely flipped, differently from the case of the polymer. However, a part from this detail, we think that it is useful to

better understand the physical meaning of some quantities. Imagine an electric dipole with dipole moment:

$$\mathbf{d} = \frac{\alpha \mathbf{R}_e}{kT} \tag{5.44}$$

subjected to a force $\mathbf{f}$ due to the electric field. The projection of the dipole in the direction of the force will be

$$\langle \cos \psi_1 \rangle \simeq \frac{1}{1 - e^{-\alpha\eta}} - \frac{1}{\alpha\eta}, \tag{5.45}$$

and thus in the orthogonal direction

$$\langle \sin \psi_1 \rangle \simeq \sqrt{\frac{2}{\alpha\eta}}. \tag{5.46}$$

With a comparison between the two expressions for $\langle \sin \psi_1 \rangle$, we obtain

$$\alpha \simeq 0.83. \tag{5.47}$$

With a more involving calculation, it is also possible to show that

$$\langle \cos \psi_2 \rangle = \langle \cos \psi_3 \rangle = \frac{I_1(\alpha\eta)}{\sinh(\alpha\eta)}, \tag{5.48}$$

where I_1 is the modified Bessel function of the first kind. We plotted formula 5.45 and 5.48 together with the results of the simulations, in figure 5.11.
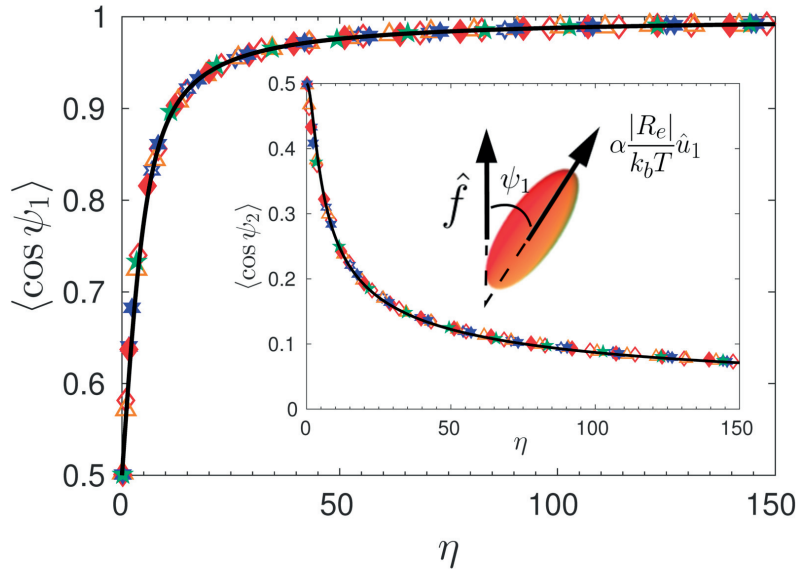

Figure 5.11 – Projection of the first (main) and the second (inset) eigenvalues. The legend of the points is the same as in the figure for the asphericity. Plot taken from [74].

First of all, we have the confirmation that the collapse that was shown for the asphericity

is also valid for the projections of the eigenvector. Moreover, the curves that were derived with the dipole analogy fit in an excellent way the simulations, even for rather small values of η.

5.5 Conclusions

We studied the response of a polymer chain under the influence of an external pulling force. Starting from the distribution of the end to end vector, it has been possible to derive analytically both the expression of the radius of gyration and the one of the asphericity in the case of a FJC. Using numerical simulations performed with MD and MC methods, we also obtained the average eigenvalues and eigenvectors of the inertia tensor, that give us some more detailed information about the orientation and the shrinking of the polymer chain. The first result that we found particularly intriguing is the fact that the lower eigenvalues do not have a monotonic dependence from the force, but they rather increase for small forces, they reach a maximum and then after the polymer is significantly oriented they start decreasing as intuitively one should expected. This behavior is less pronounced for longer polymers, since they are easier to orient along the direction of the force. More importantly, once the polymer is oriented the transverse section shrinks isotropically with the lower eigenvalues that reach a new constant value that is different from the one at zero force. The turning point of our work was when we realized that the analytical formula for the asphericity could be written, keeping only the leading terms, as a function of the sole parameter $\eta \equiv N\gamma\mathcal{L}$. After this discovery, it immediately became natural to wonder whether other quantities could also be written in terms of the same parameter. We thus showed that the average projections of the eigenvectors on the direction of the force corresponding to different numbers of monomers collapsed onto the same curve when plotted as a function of η. Moreover, we realized that it was more universal than what we expected, since it the result could be extended to polymer models that are more complicated than the FJC.

6 Concluding remarks

In this work we used methods of statistical physics to model biological systems.

We wanted to illustrate how the physics of non equilibrium can be a useful tool for the description of those biological processes that intrinsically contain a directionality. We did it by choosing the emblematic example of the molecular chaperones, with particular attention to their role in protein folding. The concept that the path through which the proteins are able to reach their functional structure is accessible in equilibrium is quite deeply rooted in the scientific community [11]. For this reason, the requirement of molecular chaperones in stress conditions has often been considered as a temporary necessity to avoid aggregation or in some cases to simply accelerate the refolding process [1, 54]. Our goal was to challenge this position, promoting the idea that there are situations in which the native conformation of the proteins is not the most thermodynamically stable, but that even in those cases it is possible for them to recover their functional structure, at the condition that a certain amount of energy is continuously supplied to the system, keeping it out of equilibrium.

We started with a more abstract approach, with a calculation that allowed to derive the constraints that the parameters follow in an Hsp70 system that stimulates the unfolding of misfolded proteins. From these constraints the importance of the non equilibrium was implied by the requirement for the affinity coefficient to be larger than one. Still with Hsp70, we then proceeded with a model that is closer to experiments. Motivated by the FRET experiments performed on denaturated rhodanese in the group of Ben Schuler [24], we showed the dependence on the energy due to the hydrolysis of ATP of the expansion of rhodanese under the influence of Hsp70. In addition, our model was able to faithfully reproduce the average values of FRET efficiency.

In a similar fashion, we showed that GroEL could rescue misfolded proteins even at a temperature at which they could not refold spontaneously. More importantly, the conservation of the native conformation was possible only in the constant presence of ATP. Indeed, as soon as adenosine diphosphatase was added, the enzymatic activity of the substrate proteins dropped drastically. These results could be nicely explained with the use of very simple non equilibrium models.

The third system that we found intriguing for its relation to non equilibrium physics is the one of the ABC transporters. Our research in that field is more recent than in the others discussed in the thesis, and so we could say that it is still somehow *work in progress*. However we could demonstrate that their necessity of ATP follows naturally from basic thermodynamic considerations. Moreover, we reproduced qualitatively the experiment in which a concentration gradient is imposed to the exporters, in the absence of ATP. The initial uptake of the substrate molecule is substitute by the efflux as soon as ATP is added in the system.

Finally, we described the deformation of a polymer chain when it is under the effect of an external pulling force. The polymer is expanded in random directions for low forces, as the increase in the lower eigenvalues suggests. Afterwards, when the polymer is oriented, the transverse section start shrinking isotropically, with the ratio between the two lower eigenvalues that reaches a new constant value that is different from the one at zero force. We think that our analysis can provide not only some speculations of purely theoretical interest (which, for the sake of the argument, would be fine as well), but also predictions that can be tested in experiments. For example, we demonstrated that a parameter like the asphericity can be inferred from the knowledge of the end to end vector, that is a quantity that can be directly measured or even tuned in single molecule experiments, performed with techniques like optical or magnetic tweezers, or atomic force microscopy.

A Appendix

A.1 Hsp70

A.1.1 Rate constants in the model with rhodanese

We assumed that the concentration of chaperones was $[Hsp70] = 10\,\mu M$ and the concentration of nucleotides $[ATP] + [ADP] = 1\,mM$. The rate constants were similar to the ones of the [40] and [38]. We report them in the table,

$k_h = 6 \cdot 10^{-4}\,s^{-1}$	$k_h^s = 1.8\,s^{-1}$
$k_{on}^T = 1.2\,M^{-1}\,s^{-1}$	$k_{on}^D = 11.7 \cdot 10^{-3}\,M^{-1}\,s^{-1}$
$k_{off}^T = 2.31\,s^{-1}$	$k_{off}^D = 2.5 \cdot 10^{-3}\,M^{-1}\,s^{-1}$
$k_-^D = 1.4 \cdot 10^{-3}\,s^{-1}$	$k_-^T = 1.33 \cdot 10^{-4}\,s^{-1}$
$k_+^D = 0.27\,s^{-1}$	$k_+^T = 0.13\,s^{-1}$
$k_-^{s,D} = (1/\delta)k_-^D$	$k_-^{s,T} = \delta\,k_-^T$
$k_+^{s,D} = \delta\,k_+^D$	$k_+^{s,T} = (1/\delta)k_+^T$

The parameter $\delta \simeq 1.7$ was a correction necessary in order to have $\gamma = 1$ in equilibrium.

A.2 Chaperonin

A.2.1 Rate constants in the model

In the model the rate constants in the following table were used.

$k_{ND} = 1.5 \cdot 10^6 \, M^{-1} s^{-1}$	$k_{DN} = 5 \cdot 10^{-4} \, s^{-1}$
$k_{NI} = 8 \cdot 10^{-3} \, s^{-1}$	$k_{IN} = 13 \, s^{-1}$
$k_{MI} = 2 \cdot 10^{-3} \, s^{-1}$	$k_{IM} = 80 \, s^{-1}$
$k_{agg} = 4 \cdot 10^3 \, M^{-1} s^{-1}$	$k_{MI}^{C} = 8 \cdot 10^{-2} \, s^{-1}$
$k_{IM}^{C} = 0.8 \, s^{-1}$	$k_{I}^{on} = 10^6 \, M^{-1} s^{-1}$
$k_{M}^{on} = 10^6 \, M^{-1} s^{-1}$	$k_{I}^{off} = 1 \, s^{-1}$

A.3   ABC transporters

A.3.1   Explicit expression for the exchange rates

Since we are assuming that the concentration of ATP and ADP are large enough to neglect the concentration of the apo-state, it is possible to reduce the system by properly coarse-graining the exchange between the ATP and ADP state using effective rate constants [21]. These effective rate constants can be written in terms of the binding constants and the concentrations of ATP and ADP:

$$k_{DT}^e = k_-^D \frac{k_+^T [ATP]}{k_+^T [ATP] + k_+^D [ADP]} \tag{A.1}$$

and

$$k_{TD}^e = k_-^T \frac{k_+^D [ADP]}{k_+^T [ATP] + k_+^D [ADP]} \ , \tag{A.2}$$

and we have similar equations in the case of bound substrate. Therefore, we have

$$\frac{k_{TD}^e}{k_{DT}^e} = \frac{[ADP] \, K_d^T}{[ATP] \, K_d^D} ,$$
$$\frac{k_{TD}^{e,s}}{k_{DT}^{e,s}} = \frac{[ADP] \, K_d^{T,s}}{[ATP] \, K_d^{D,s}} \tag{A.3}$$

where K_d^T and K_d^D are the dissociation constants of the ATP and the ADP respectively.

A.3.2   Master equations for the rate model

We report here the Master equations used to determine the concentrations in the stationary state. We use the letter "B" to indicate the transporter and "s" the substrate protein. As

usual, "T" and "D" indicate ATP and ADP respectively.

$$\frac{d[BT]}{dt} = -[BT]\left[[c]k_{on}^{Tc} + [t]k_{on}^{Tt} + k_h + k_{TD}^e\right] + [BTs]\left[k_{off}^{Tc} + k_{off}^{Tt}\right] +$$
$$+ [BD]\left[k_x + k_{DT}^e\right]$$

$$\frac{d[BTs]}{dt} = -[BTs]\left[k_{off}^{Tc} + k_{off}^{Tt} + k_h^s + k_{TD}^{(s,e)}\right] + [BT]\left[[c]k_{on}^{Tc} + [t]k_{on}^{Tt}\right] +$$
$$+ [BDs]\left[k_x^s + k_{DT}^{(s,e)}\right]$$

$$\frac{d[BD]}{dt} = -[BD]\left[[c]k_{on}^{Dc} + [t]k_{on}^{Dt} + k_x + k_{DT}^e\right] + [BDs]\left[k_{off}^{Dc} + k_{off}^{Dt}\right] +$$
$$+ [BD]\left[k_x + k_{DT}^e\right] \tag{A.4}$$

$$\frac{d[BDs]}{dt} = -[BDs]\left[k_{off}^{Dc} + k_{off}^{Dt} + k_x^s + k_{DT}^{(s,e)}\right] + [BD]\left[[c]k_{on}^{Dc} + [t]k_{on}^{Dt}\right] +$$
$$+ [BDs]\left[k_s^s + k_{TD}^{(s,e)}\right].$$

When we want to consider the NESS, we impose

$$\frac{d[BT]}{dt} = \frac{d[BTs]}{dt} = \frac{d[BD]}{dt} = \frac{d[BDs]}{dt} = 0\,, \tag{A.5}$$

as usual.

A.3.3 Asymptotic results

The constants that appear in 4.6 are

$$A = k_{off}^{Tt}k_-^{s,D}k_+^{s,T}$$
$$B = k_{off}^{Dt}k_-^{s,T}k_+^{s,D}\,. \tag{A.6}$$

A.3.4 Rates in the model

For the rate model we used parameters that were in agreement with the prediction of Senior and Gadsby [82]. We report them in the following table.

Appendix A. Appendix

$k_{on}^{Tc} = 10^{-4}\,s^{-1}\,\mu M^{-1}$	$k_{on}^{Dt} = 10^{-4}\,s^{-1}\,M^{-1}$
$k_{on}^{Tt} = 0.5\,s^{-1}\,\mu M^{-1}$	$k_{on}^{Dc} = 0.5\,s^{-1}\,\mu M^{-1}$
$k_{off}^{Dc} = 0.01\,s^{-1}$	$k_{off}^{Tt} = 0.01\,s^{-1}$
$k_{off}^{Tc} = 2 \cdot 10^{-6}\,s^{-1}$	$k_{off}^{Dt} = 2 \cdot 10^{-6}\,s^{-1}$
$k_h = 10^{-3}\,s^{-1}$	$k_{sy} = 1.47 \cdot 10^{-9}\,s^{-1}$
$k_h^s = 2 \cdot k_h$	$k_{sy}^s = 2.94 \cdot 10^{-9}\,s^{-1}$
$k_-^D = 10^{-3}\,s^{-1}$	$k_-^T = 10^{-3}\,s^{-1}$
$k_-^{s,D} = 10\,k_-^D$	$k_-^{s,T} = k_-^T$
$k_+^D = 0.5\,s^{-1}$	$k_+^T = 0.5\,s^{-1}$
$k_+^{s,D} = 10\,k_+^D$	$k_+^{s,T} = k_+^T$

They were chosen so that $\gamma = 1$ in equilibrium, for every closed cycle in the biochemical network, as we explain in the main text.

Bibliography

[1] M. Hayer-Hartl, A. Bracher, and F. Hartl, "The groel–groes chaperonin machine: A nano-cage for protein folding," *Trends in Biochemical Sciences*, vol. 41, pp. 62–76, 2015.

[2] X. Zhang, H. Qian, and M. Qian, "Stochastic theory of nonequilibrium steady states and its applications," *Physics Reports*, vol. 510, p. 1–86, 2012.

[3] H. Hofmann, A. Soranno, K. G. A. Borgia, D. Nettels, and B. Schuler, "Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single molecule spectroscopy," *Proc. Nat. Acad. Sci*, vol. 109, pp. 16155–16160, 2012.

[4] P. Gennes, *Scaling Concepts in Polymer Physics.* Cornell University Press, 1979.

[5] M. Rubinstein and R. Colby, *Polymer Physics.* Oxford University Press, 2003.

[6] C.Bustamante, Y. Chemla, N. Forde, and D. Izhaky, "Mechanical processes in biochemistry," *Annu. Rev. Biochem.*, vol. 73, pp. 705–748, 2004.

[7] R. Astumian and M. Bier, "Fluctuation driven ratchets: Molecular motors," *Phys. Rev. Lett.*, vol. 72, pp. 1766–1769, 1994.

[8] D. Evans, E. Cohen, and G. Morriss, "Probability of second law violations in shearing steady states," *Phys. Rev. Lett.*, vol. 71, p. 2401, 1993.

[9] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. T. Jr, and C. Bustamante, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, vol. 78, pp. 2690–2690, 1997.

[10] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. T. Jr, and C. Bustamante, "Verification of the crooks fluctuation theorem and recovery of rna folding free energies," *Nat Med.*, vol. 437, p. 231–234, 2005.

[11] C. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, pp. 223–230, 1973.

[12] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. Marks, C. Sander, R. Zecchina, J. Onuchic, T. Hwa, and M. Weigt, "Direct-coupling analysis of residue coevolution captures native contacts across many protein families," *eLife*, vol. 6, p. e23471, 2017.

Bibliography

[13] D. Malinverni, A. Lopez, P. D. L. Rios, G. Hummer, and A. Barducci, "Modeling hsp70-hsp40 interaction by multi-scale molecular simulations and coevolutionary sequence analysis," *eLife*, vol. 6, p. e23471, 2017.

[14] M. Karplus and J. McCammon, "Molecular dynamics simulations of biomolecules," *Nat. Struct. Mol. Biol.*, vol. 9, pp. 646–652, 2002.

[15] P. Goloubinoff and P. D. L. Rios, "The mechanism of hsp70 chaperones: (entropic) pulling the models together," *Trends. Biochem. Sci.*, vol. 32, p. 372–380, 2007.

[16] P. D. L. Rios, A. Ben-Zvi, O. Slutsky, A. Azem, and P. Goloubinoff, "Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling," *Proc. Natl. Acad. Sci.*, vol. 103, p. 6166–6171, 2005.

[17] H.Ge, M.Qian, and H.Qian, "Stochastic theory of nonequilibrium steady states. part ii: Applications in chemical biophysics," *Physics Report*, vol. 510, pp. 87–118, 2012.

[18] J. Schnakenberg, "Network theory of microscopic and macroscopic behavior of master equation systems," *Rev. Mod. Phys.*, vol. 48, pp. 571–585, 1976.

[19] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*. Garland Science, 2007.

[20] L. Balakrishnan, H. Venter, R. Shilling, and H. van Veen, "Reversible transport by the atp-binding cassette multidrug export pump lmra," *J. Biol. Chem.*, vol. 279, pp. 11273–11280, 2004.

[21] P. D. L. Rios and A. Barducci, "Hsp70 chaperones are non-equilibrium machines that achieve ultra-affinity by energy consumption," *eLife*, vol. 3, p. e02218, 2014.

[22] K. Hingorani and L. Gierasch, "Comparing protein folding in vitro and in vivo: Foldability meets the fitness challenge," *Curr Opin Struct Biol.*, vol. 0, pp. 81–90, 2014.

[23] D. Thirumalai and G. Lorimer, "Chaperonin-mediated protein folding," *Annu. Rev. Biophys.*, vol. 30, p. 245–69, 2001.

[24] R. Kellner, H. Hofmann, A. Barducci, B. Wunderlich, D. Nettels, and B. Schuler, "Single-molecule spectroscopy reveals chaperone-mediated expansion of substrate protein," *Proc. Natl. Acad. Sci.*, vol. 111, pp. 13355–13360, 2014.

[25] K. Neuman and A. Nagy, "Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy," *Nature Methods*, vol. 5, pp. 491–505, 2008.

[26] W. Chirico, M. Waters, and G. Blobel, "70k heat shock related proteins stimulate protein translocation into microsomes," *Nature*, vol. 332, pp. 805–810, 1988.

[27] G. Lebon, D. Jou, and J. Casas-Vázquez, *Understanding Non-equilibrium Thermody-namics.* Springer, 2008.

[28] C. Dobson, "Protein folding and misfolding," *Nature*, vol. 426, pp. 884–890, 2003.

[29] C. Ross and M. Poirier, "Protein aggregation and neurodegenerative disease," *Nat Med.*, vol. 10, pp. S10–S17, 2004.

[30] B. Bukau and H. Horwich, "The hsp70 and hsp60 chaperone machines," *Cell*, vol. 92, p. 351–366, 1998.

[31] M. Pilon and R. Schekman, "Protein translocation: How hsp70 pulls it off," *Cell*, vol. 97, p. 679–682, 1999.

[32] M.R..Fernández, Gragera, L. Ibarrola, Q. Gallardo, and M.Valpuesta, "Hsp70 - a master regulator in protein degradation," *FEBS Lett.*, vol. 591, pp. 2648–2660, 1999.

[33] H. Kampinga and E.A.Craig, "The hsp70 chaperone machinery: J proteins as drivers of functional specificity," *Nat. Rev.*, vol. 11, pp. 579–592, 2010.

[34] T. Laufen, M. Mayer, C. Beisel, D. Klostermeier, A. Mogk, J. Reinstein, and B. Bukau, "Mechanism of regulation of hsp70 chaperones by dnaj cochaperones," *Proc. Natl. Acad. Sci.*, vol. 96, p. 5452–5457, 1999.

[35] J. Rauch and J. Gestwicki, "Binding of human nucleotide exchange factors to heat shock protein 70 (hsp70) generates functionally distinct complexes in vitro," *J. Biol. Chem.*, vol. 289, p. 1402–1414, 2014.

[36] M. Mayer, "Hsp70 chaperone dynamics and molecular mechanism," *Trends in Bio-chemical Sciences*, vol. 38, pp. 223–230, 2013.

[37] A. Zhuravleva, E. Clerico, and L. Gierasch, "An interdomain energetic tug-of-war creates the allosterically active state in hsp70 molecular chaperones," *J. Biol. Chem.*, vol. 151, pp. 1296–1307, 2012.

[38] M. Mayer and B. Bukau, "Hsp70 chaperones: Cellular functions and molecular mechanism," *Cell Mol Life Sci.*, vol. 62, pp. 670–684, 2005.

[39] R. Tehver and D. Thirumalai, "Kinetic model for the coupling between allosteric transitions in groel and substrate protein folding and aggregation," *J. Mol. Biol.*, vol. 377, pp. 1279–1295, 2008.

[40] B. Hu, M. Mayer, and M. Tomita, "Modeling hsp70-mediated protein folding," *Biophys. J.*, vol. 91, pp. 496–507, 2006.

[41] A. Karzai and R. McMacken, "A bipartite signaling mechanism involved in dna-j medi-ated activation of the e. coli dna-k protein," *J. Biol. Chem.*, vol. 271, p. 11236–11246, 1995.

[42] F. Rodriguez, F. Arsène-Ploetze, W. Rist, S. Ruediger, J. Schneider-Mergener, M. Mayer, and B. Bukau, "Molecular basis for regulation of the heat shock transcription factor σ^{32} by the dnak and dnaj chaperones," *Molecular Cell*, vol. 32, pp. 347–358, 2008.

[43] R. Kityk, J. Kopp, and M. Mayer, "Molecular mechanism of j-domain-triggered atp hydrolysis by hsp70 chaperones," *Molecular Cell*, vol. 69, pp. 227–237, 2018.

[44] S. Rüdiger, L. Germerot, J. Schneider-Mergener, and B. Bukau, "Substrate specificity of the dnak chaperone determined by screening cellulose-bound peptide libraries," *EMBO J.*, vol. 16, pp. 1501–1507, 1997.

[45] C. dos Remedios and P. Moens, "Fluorescence resonance energy transfer spectroscopy is a reliable ruler for measuring structural changes in proteins: dispelling the problem of the unknown orientation factor," *J. Struct. Biol.*, vol. 115, pp. 175–185, 1995.

[46] P. D. L. Rios, A. Ben-Zvi, O. Slutsky, A. Azem, and P. Goloubinoff, "Single-molecule fret spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins," *Annu. Rev. Biophys.*, vol. 45, p. 207–231, 2016.

[47] J. V. Durme, S. Maurer-Stroh, R. Gallardo, H. Wilkinson, F. Rousseau, and J. Schymkowitz, "Accurate prediction of dnak-peptide binding via homology modelling and experimental data," *Plos*, vol. 5, pp. e1000475–e1000475, 2009.

[48] D. Schmid, A. Baici, H. Gehring, and P. Christen, "Kinetics of molecular chaperone action," *Science*, vol. 263, pp. 971–973, 1994.

[49] P. Goloubinoff, A. Sassi, B. Fauvet, A. Barducci, and P. D. L. Rios, "Chaperones convert the energy from atp into the non equilibrium stabilisation of native proteins," *Nat. Chem. Biol.*, vol. 14, pp. 388–395, 2018.

[50] N. Ranson, D. Clare, G. Farr, D. Houldershaw, A. Horwich, and H. R. Saibil, "Allosteric signalling of atp hydrolysis in groel-groes complexes," *Nat Struct Mol Biol.*, vol. 13, p. 147–152, 2006.

[51] M. Preuss and A. Miller, "The affinity of the groel/groes complex for peptides under conditions of protein folding," *FEBS*, vol. 466, pp. 75–79, 2000.

[52] A. Gupta, S. Haldar, G. Mili, F. Hartl, and M. Hayer-Hartl, "Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level," *J. Mol. Biol.*, vol. 426, pp. 2739–2754, 2014.

[53] Z. Lin, J. Puchalla, D. Shoup, and H. Rye, "Repetitive protein unfolding by the trans ring of the groel-groes chaperonin complex stimulates folding," *J. Biol. Chem.*, vol. 288, p. 30944 –30955, 2013.

[54] A. Apetri and A. Horwich, "Chaperonin chamber accelerates protein folding through passive action of preventing aggregation," *Proc. Natl. Acad. Sci.*, vol. 105, p. 17351–17355, 2008.

[55] D. Hartman, B. Surin, N. Dixon, N. Hoogenraad, and P. Hoj, "Substoichiometric amounts of the molecular chaperones groel and groes prevent thermal denaturation and aggregation of mammalian mitochondrial malate dehydrogenase in vitro," *Proc. Natl. Acad. Sci.*, vol. 90, p. 2276–2280, 1993.

[56] A.Brinker, G. Michael, J.K.Dean, J.Naylor, F.U.Hartl, and M.Hayer-Hartl, "Modeling hsp70-mediated protein folding," *Biophys. J.*, vol. 107, p. 223–233, 2001.

[57] A. Baumketner, A. Jewett, and J. Shea, "Effects of confinement in chaperonin assisted protein folding: Rate enhancement by decreasing the roughness of the folding energy landscape," *J. Mol. Biol.*, vol. 332, p. 701–713, 2003.

[58] U. Feller, I. Anders, and T. Mae, "Rubiscolytics: fate of rubisco after its enzymatic function in a cell is terminated," *J. Exper. Bot.*, vol. 59, p. 1615–1624, 2008.

[59] U. Feller, I. Anders, and T. Mae, "Groel-groes cycling: Atp and nonnative polypeptide direct alternation of folding-active rings," *Cell*, vol. 97, pp. 325–338, 1999.

[60] B. Holland, S. Cole, K. Kuchler, and C. Higgins, *ABC Proteins: From Bacteria to Man.* Academic Press, 2015.

[61] C. Higgins and K. Linton, "The atp switch model for abc transporters," *Nat. Struct. Mol. Biol.*, vol. 10, pp. 918–926, 2004.

[62] C. Higgins, "Abc transporters: physiology, structure and mechanism – an overview," *Res. Microbiol.*, vol. 152, p. 205–210, 2001.

[63] S. Velamakanni, Y. Yao, D. Gutmann, and H. van Veen, "Multidrug transport by the abc transporter sav1866 from staphylococcus aureus," *Biochemistry*, vol. 47, p. 9300–9308, 2008.

[64] R. Shilling, H. Venter, S. Velamakanni, A. Bapna, B. Woebking, S. Shahi, and H. van Veen, "New light on multidrug binding by an atp-binding-cassette transporter," *TRENDS in Pharmacological Sciences*, vol. 27, pp. 195–204, 2006.

[65] R. Dawson and K. P. Locher, "Structure of a bacterial multidrug abc transporter," *Nature*, vol. 443, pp. 180–185, 2006.

[66] K. Hollenstein, D. Frei, and K. Locher, "Structure of an abc transporter in complex with its binding protein," *Nature*, vol. 446, pp. 213–216, 2007.

[67] D. Rees, E. Johnson, and O. Lewinson, "Abc transporters: the power to change," *Nature*, vol. 10, pp. 218–227, 2009.

[68] O. Lewinson, A. Lee, K. Locher, and D. Rees, "A distinct mechanism for the abc transporter btucd-btuf revealed by the dynamics of complex formation," *Nat. Struct. Mol. Biol.*, vol. 17, p. 332–338, 2010.

Bibliography

[69] L. Bishop, R. Agbayani, S. Ambudkar, P. Maloney, and G. Ames, "Reconstitution of a bacterial periplasmic permease in proteoliposomes and demonstration of atp hydrolysis concomitant with transport," *Proc. Natl. Acad. Sci.*, vol. 18, pp. 6953–6957, 1989.

[70] A. Davidson, H. Shuman, and H. Nikaido, "Mechanism of maltose transport in escherichia coli: transmembrane signaling by periplasmic binding proteins," *Pnas*, vol. 89, pp. 2360–2364, 1992.

[71] E. Borths, B. Poolman, R. Hvorup, K. Locher, and D. Rees, "In vitro functional characterization of btucd-f, the escherichia coli abc transporter for vitamin b12 uptake," *Biochemistry*, vol. 44, pp. 16301–16309, 2005.

[72] M. Oldham and J. Chen, "Crystal structure of the maltose transporter in a pretranslocation intermediate state," *Science*, vol. 332, pp. 1202–1205, 2011.

[73] S.Wilkens, "Structure and mechanism of abc transporters," *F1000Prime Reports*, vol. 7, p. 14, 2015.

[74] A. Sassi, S. Assenza, and P. D. L. Rios, "Shape of a stretched chain," *Phys. Rev. Lett.*, vol. 119, p. 037801, 2017.

[75] J. Moffitt, Y. Chemla, S. Smith, and C. Bustamante, "Recent advances in optical tweezers," *Annu. Rev. Biochem.*, vol. 77, pp. 205–228, 2008.

[76] J. Rudnick and G. Gaspari, "Elements of the random walk," *Library of Congress*, 2004.

[77] J. Rudnick and G. Gaspari, "The aspherity of random walks," *J. Phys. A: Math. Gen.*, vol. 19, pp. L191–L193, 1986.

[78] P. J. Flory, "Principles of polymer chemistry," *Cornell University Press*, 1953.

[79] S. Sciutto, "The shape of self-avoiding walks," *J. Phys. A: Math. Gen.*, vol. 29, p. 5455–5473, 1996.

[80] J. Marko and E. Siggia, "Stretching dna," *Macromolecules*, vol. 28, pp. 8759–8770, 1995.

[81] R. Neumann, "Nonequivalence of the stress and strain ensembles in describing polymer-chain elasticity," *Phys. Rev. A*, vol. 31, pp. 3516–3517, 1985.

[82] A. Senior and D. Gadsby, "Atp hydrolysis cycles and mechanism in p-glycoprotein and cftr," *Seminars in Cancer Biology*, vol. 8, pp. 143–150, 1997.

Alberto Sassi

MSc in Physics

BSP 519 Rte de la Sorge
CH-1015 Lausanne
Switzerland
☎ *+41 21 69 38347*
✉ *alberto.sassi@epfl.ch*
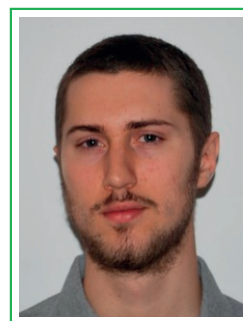
Education

2014–2018 **PhD in Theoretical Biophysics**, *École Polytechnique Fédérale de Lausanne*, Lausanne.
Thesis title: *Non equilibrium thermodynamics of protein organization*

2012–2014 **MSc degree in Physics**, *École Polytechnique Fédérale de Lausanne*, Lausanne.
Thesis title: *Asymmetries in 3-dimensional unbranched polymer chains*

2009–2012 **BSc degree in Physics**, *University of Turin*, Turin.

Professional Experiences

Laboratory of Statistical Biophysics, École Polytéchnique Fédérale de Lausanne

- Implementation of Monte Carlo simulations of polymer chains in a good solvent;
- Theoretical models of non equilibrium thermodynamics applied to chaperone systems;
- Theoretical models of non equilibrium thermodynamics applied to ABC transporters;

Languages

Italian **Native**
English **Very good**
French **Good**

Skills

- Programming in C++ and Mathematica;
- Numerical methods for the solution of non-linear equations;
- Basic knowledge in cell biology;
- Basic knowledge in protein folding and chaperone systems;

Publications

- A.S. Sassi, S. Assenza, and P. De Los Rios, *Shape of a Stretched Polymer*, Phys. Rev. Lett. 119, 037801 (2017);
- P. Goloubinoff, A.S. Sassi, B. Fauvet, A. Barducci and P. De Los Rios, *Chaperones convert the energy from ATP into the non-equilibrium stabilisation of native proteins*, Nat. Chem. Biol.,

publishing (2018);

Oral Presentations

- *Active unfolding of RNA by chaperones*, Swiss Soft Day 2015, Fribourg;
- *DEAD-box proteins actively unfold RNA*, FisMat 2015, Palermo;
- *Molecular chaperones as cellular non-equilibrium machines*, Physics of Living Systems Seminar, 2017, Lausanne;

Teaching Experiences

- 2014–2015, Teaching assistant in the course *Mathematical methods for physicists*, BSc course for physicists;
- 2015-2016 and 2016-2017, Teaching assistant in the course *Classical electrodynamics*, BSc course for physicists;
- 2014-2015, Teaching assistant in the course *General physics I*, BSc course for chemists;
- 2015-2016 and 2016-2017, Teaching assistant in the course *Analytical Mechanics*, BSc course for physicists.

References

- Prof Paolo De Los Rios, École Polytéchnique Fédérale de Lausanne, Institute of Physics, +41 21 69 30510, paolo.delosrios@epfl.ch
- Dr. Alessandro Barducci, Centre de Biochimie Structurale de Montpellier, CNRS, +33 467 4179 11, barducci@cbs.cnrs.fr