The physics of the protein folding problem

Amélie Coulon

Supervisor: Prof. Paolo de Los Rios, Head of the Laboratory of Statistical Biophysics

Laidlaw Research Internship 2024

Abstract

Proteins are essential to cellular functions, acting as the primary agents for various tasks such as transport, storage, communication, defense, and energy transformation. To perform these roles, proteins must fold into specific structures, known as their native states. Protein misfolding and aggregation are closely associated with neurodegenerative diseases like Alzheimer's and Parkinson's. This report investigates the current understanding of protein folding from a physics perspective, exploring protein structure, folding dynamics, and the role of molecular chaperones in assisting correct folding.

Introduction

Proteins are chains of amino acids which are truly essential to the functioning of the cell. They are the workers of the cell, and perform various functions such as the transport and storage from electrons to macromolecules, the transmission of information between cells and organs, the defense of the organism (as antibodies), the control of gene expression, the transformation of chemical energy to mechanical energy in muscles allowing us to make movements, and those are just a few examples [1]. They are present in every living being, driving nearly every essential process that sustains life. In essence, proteins don't just contribute to life's complexity—they are the very tools that make it possible.

In order to perform their functions, they need to fold in a very specific shape, called the native state of the protein. Abnormal protein folding can lead to aggregation, which is also a hallmark of neurodegenerative diseases such as Alzheimer's, Huntington's, Parkinson's and motor neuron disorder [2]. This makes the understanding of how proteins fold and how cells protect their proteins under different stresses a key question in biology, attracting great interest because of its complexity and importance for medicine. My goal during the internship was to explore our current understanding of this complex mechanism from a physics perspective and to identify the key challenges that remain.

1 Protein Structure and Folding Dynamics

General structure

Before considering the process of folding, we shall remind some basics about protein structure. As mentioned above, proteins are chains of amino acids, ranging from a few amino acids to several thousands. The twenty-one different amino acids can be found in Fig. 1.

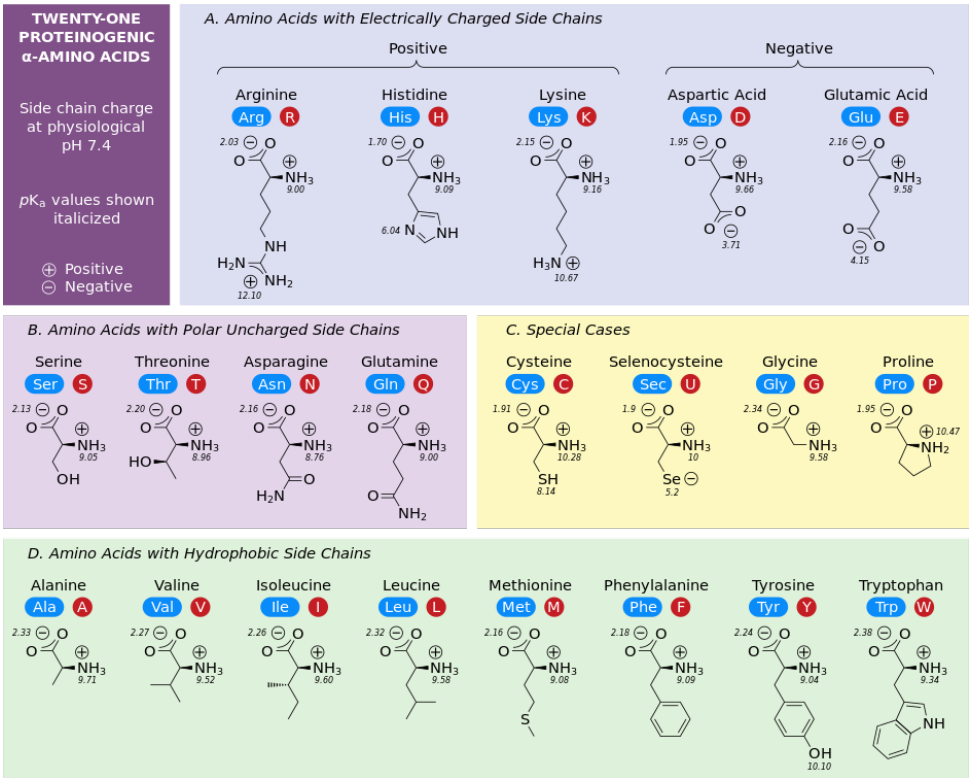

Figure 1: All different special chemical groups of amino acids

The process of creation of a protein can be decomposed into two different steps. The first step is the transcription, where the messenger RNA is created from the DNA with the help of the RNA polymerase enzyme. The second step is translation, where RNA is read by ribosome with the genetic code and form a protein. Proteins are linear heteropolymers, meaning they are an assembly of different monomers, the amino acids. Those units are all formed with the same basis, an α-carbon attached to three chemical groups. The assembly of those common units is called the backbone of the protein. The fourth group that can be attached to the α-carbon is different for each amino acid. Some can be hydrophobic chemical groups such as Proline or Alanine, some can be polar as Cysteine or hydrophilic such as Glycine. These properties influence how the protein folds in an aqueous environment like the cell's cytoplasm. We will explore further this part later.

Structure types

Protein structure is classified into four scale-related categories, in ascending order (from local to global). The *primary structure* is the linear amino acid sequence of a protein's polypeptide chain (a subunit of a protein). They are linked by covalent peptide bonds, which are strong and maintain the integrity of the polypeptide chain.
Secondary structure refers to the local spatial conformation of a protein subunit. The two most common types of secondary structures are α-helices and β-sheets. In an α-helix, the chain twists into a right-handed coil. The carbonyl group of each ith residue forms a hydrogen bond with the amino group of the residue located four positions ahead $(i + 4)$ in the sequence. It has approximately 3.6 residues per turn. There exists also other helices that are less common, such as 3-helix, which as approximately 3 residues per turn, or π-helix, with 4.4 residues per turn [1]. A schematic of an α-helix can be found in Fig. 2.
On the other hand, β-sheets consist of strands lying side by side, with hydrogen bonds forming between the strands. They can be parallels or antiparallels, as depicted in Fig. 3.

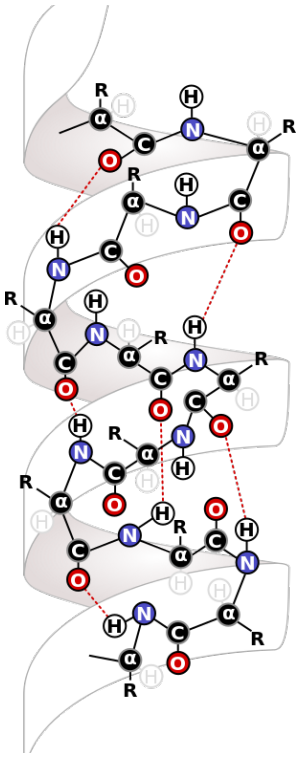

Figure 2: α-helix

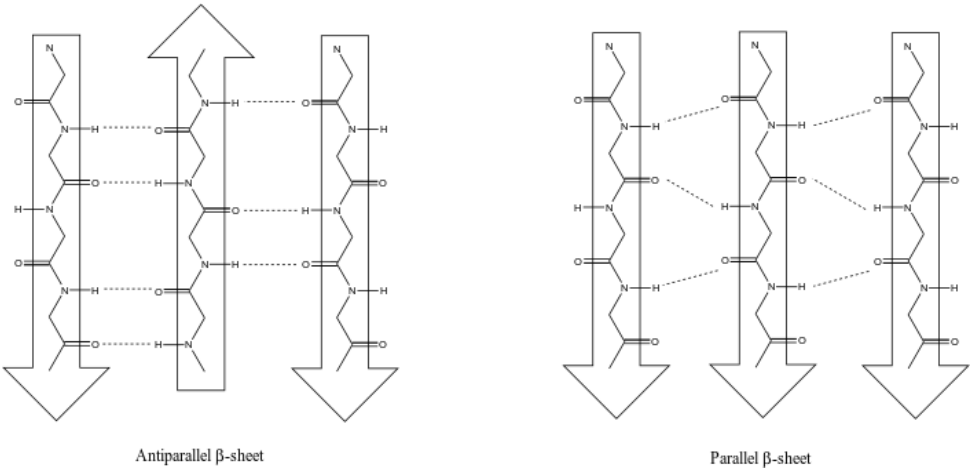

Figure 3: β-sheet

Tertiary structures describe the overall three-dimensional shape of a single subunit chain. This structure is stabilized by various interactions, such as hydrophobic interactions.

Quaternary structure arises when two or more subunits chains come together to form a functional protein complex. Each subunit may be identical or different, and their interactions are stabilized similarly to tertiary structures. Hemoglobin, the oxygen-transporting protein in blood, is a classic example of a protein with quaternary structure, consisting of four subunits. This protein is depicted in Fig. 4.

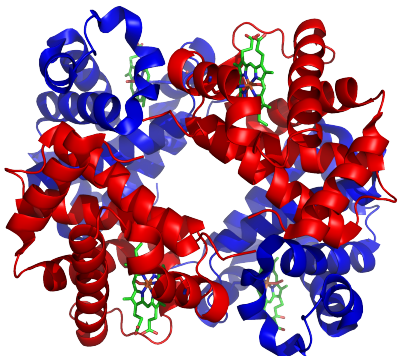

Figure 4: Hemoglobin's native state

A view of the protein folding problem

In 1957, Anfinsen and other scientists made an experiment to have a better idea of what makes a protein fold and whether it occurs randomly. Anfinsen denatured the ribonuclease enzyme using urea and mercaptoethanol, disrupting its structure, and making it unfold. Upon removal of these agents, the enzyme spontaneously refolded into its active form, demonstrating that the information necessary for folding is contained within the amino acid sequence itself. This pivotal experiment provided strong evidence for the principle that a protein's native structure is determined solely by its primary sequence.

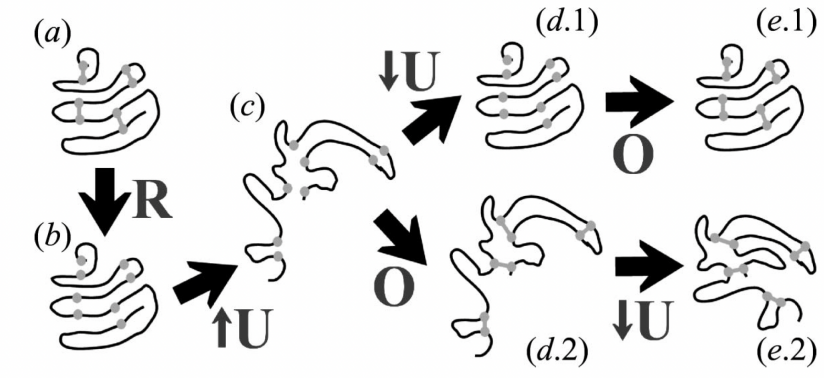

Figure 5: Schematic of Anfinsen's experiment [1]. Black arrows indicate key steps: (R) reducing agent breaks disulphide bonds, (O) oxidizing agent reforms them, (↑U) urea concentration is increased, and (↓U) decreased. The protein's backbone is shown as a black line, cysteines as grey circles, and disulphide bonds as connecting lines. The labels represent different states of the protein, from the active enzyme (a), unfolded or scrambled forms (b, c, d.1, d.2), to the refolded enzyme (e.1, e.2).

Several questions arose from this experiment, the main one being *What could explain such a complex native structure from a disordered amino-acid chain?* A primary coarse grained model has been found to describe this sequence-structure relationship, the HP model (Hydrophobic-Polar model). In this model, proteins are represented as chains of H and P monomers, which interact based on their hydrophobic or polar nature. The model assumes that hydrophobic residues tend to cluster together to minimize exposure to water, driving the protein to fold into a compact structure. The HP model helps explain how a sequence of amino acids determines the protein's native fold by highlighting the role of hydrophobic interactions in stabilizing the structure.

It also appears that the time required for a protein to fold is in the order of microseconds or milliseconds. *What could explain that a protein fold so quickly?* If a protein would explore all the possible configuration ($\approx 10^{300}$), it would take a huge amount of time. This problem is known as Levinthal's paradox. There must be a path that a protein must follow to avoid exploring every possible configuration. In fact, when placed in folding conditions, the protein chain begins to randomly develop more and more hydrophobic-hydrophobic (HH) contacts (and forming α-helices and β-sheets locally), which reduces the number of possible conformations it can explore. Thus the folding process is not purely random; it is largely guided by the number of conformations remaining to be explored, which decreases rapidly as the protein becomes more compact.

Some quite general law we need to remind is that a biological system tries to minimize its energy and maximize its entropy. It is like when a ball is placed on the top of a hill, it will go down to the valley. But as we are at the nano-scale, we also need to take into account the thermal forces, which leads to the maximization of the entropy. At nano-scale, the mechanical forces are indeed compensated by the thermal forces. Thus due to the thermal shocks, the ball in the valley could go back up to the top of the hill. If we combine both phenomenon, we find that a biological system tries to minimize its free-energy, defined by its energy minus its entropy.

Figure 6: Schematic of a ball going down to a valley to minimize its energy

This analogy with the hill is useful to understand the model used to describe the path followed by a protein as it folds: the (free-)energy landscape. It is a conceptual model used to describe the range of possible conformations a protein can adopt during folding, represented as a surface where different conformations correspond to points with varying free-energy levels. In this landscape, the native state of the protein—a stable and functional conformation—lies in a low-energy region (a valley), while unfolded states are located in higher-energy regions (hills). Proteins traverse this energy landscape as they fold, moving from high-energy, disordered states to low-energy, ordered states. Importantly, the landscape can be rugged, with bumps and valleys that represent folding intermediates or misfolded states, creating kinetic barriers that slow the folding process. Examples of energy landscape can be found in Fig. 7. We will come back later on this subject.

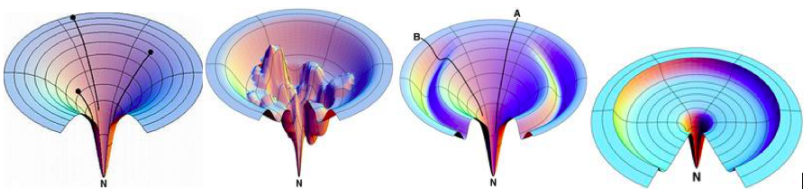

Figure 7: Examples of energy landscapes

2 Chaperones as Protein Folding Assistants

Protein folding is not always straightforward, especially in the crowded environment of the cell. Chaperones, a specialized group of proteins, play a vital role in assisting other proteins to fold correctly, preventing them from misfolding and aggregating, which can lead to severe diseases such as Alzheimer's, Parkinson's, and Huntington's [4].

The basics of chaperones

As we saw, Anfinsen's experiment showed that a protein is able to fold by itself. But even in a test tube, larger proteins take much more time to fold than shorter ones. This folding process becomes significantly more complicated within the cell because of its crammed environment. Inside the cell, proteins interact with countless other molecules, which increases risks of aggregation [5]. This is because hydrophobic molecules tend to stick together to reduce contact with water, which can lead some proteins to aggregate. As mentioned previously, the energy landscape is often rugged, which means that proteins can be stuck in kinetic barriers, where they are partially folded or misfolded. Partially folded states are particularly common, especially for proteins with more than 100 amino acids (90% of all proteins in a cell [4]), and are highly susceptible to collapse into amorphous aggregates. This is why many proteins need some help to reach their native state within the cell, which is provided by a specific type of proteins, namely molecular chaperones. Literature defines them as any protein that interacts with, stabilizes, or assists another protein in reaching its functional conformation, without becoming part of the final structure [6]. However, taking a step back, one may say that this is a broad definition, which highlights the fact that we have still many things to understand about them. They have several roles: to guide the folding process by helping proteins to pass trough kinetic barriers, to prevent aggregation by sticking to proteins, tackle abnormal protein aggregates and break them down into individual monomers, take care of protein degradation. Different types of chaperones will be presented below, and a schematic of their roles can be found in Fig.8.

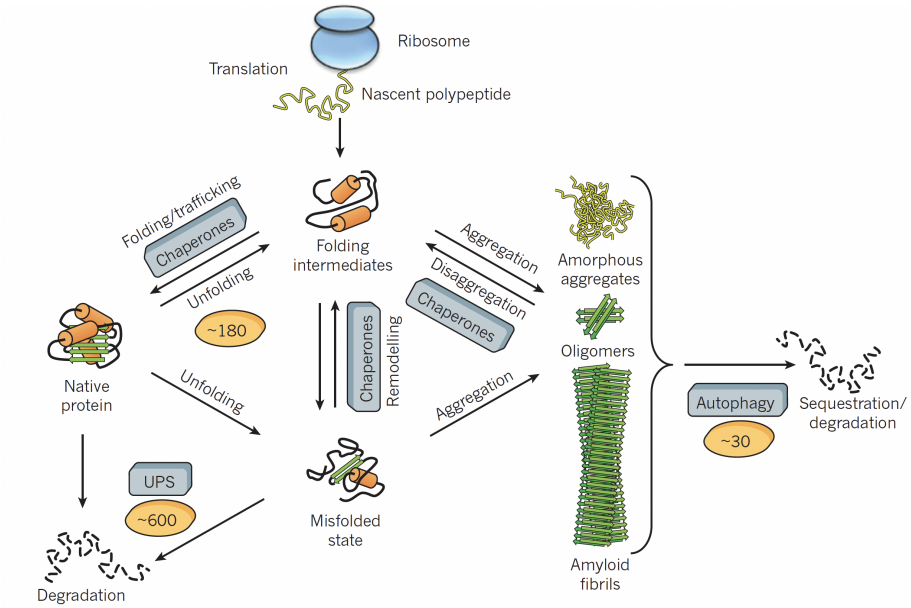

Figure 8: Summary of different chaperone roles within the cell [4]

Their role in proteome maintenance

Most chaperone proteins are stress-protein or heat-shock proteins (HSP) as they act on a stress-condition of the cell, in which the concentrations of aggregation-prone folding intermediates increase. They are named after their molecular mass (HSP60, HSP70, HSP90...). Chaperones assist in proteome maintenance through several mechanisms presented above. Their cycle often uses ATP to function, but *how* they use it is more or less understood depending on the protein considered. They also can collaborate with smaller chaperones (that don't need ATP) to function [7]. They can for example recognize hydrophobic amino-acid, which are exposed when protein is misfolded, and stick to this part of the protein to prevent other proteins to bind to it, and thus avoid aggregation.

HSP70

Let us look at the HSP70, an important chaperone that help to prevent aggregation. This protein plays a key role in protein folding and maintaining proteostasis—we will come back to this concept. It can appear in several forms: some are always present (HSC70 or HSPA8) and some appear during stress. HSP70 binds to and releases proteins by changing shape, with the help of HSP40. When ATP is attached, HSP70 is open and can quickly grab onto proteins. When ATP is broken down to ADP (releasing energy in the process, ADP being the product of the reaction), HSP40 assists in closing HSP70 around a protein, ensuring it remains securely bound. Later, a mechanism replaces ADP with ATP, causing HSP70 to reopen and release the protein, allowing it to fold properly. If more time is needed for folding, the protein can rebind to HSP70, preventing aggregation. This cycle is illustrated in Fig.9, and it is to be noted that it has recently been reviewed, and the order of protein states has been revised from 'unfolded, partially folded, native' to 'partially folded, unfolded, native.' This better reflects the actual sequence of events, while maintaining the same HSP70 cycle.

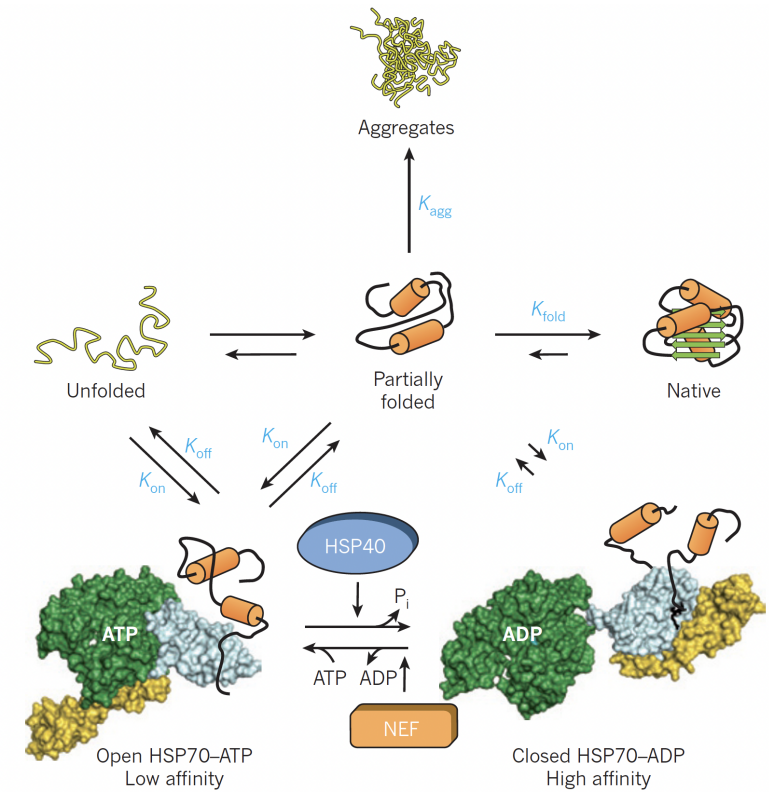

Figure 9: HSP70 chaperone cycle [4]

Chaperonins

If HSP70 is insufficient for proper protein folding, chaperonins can provide additional help. These large complexes help proteins fold by enclosing them in a protective cage. There are two main types of chaperonins, with the GroEL-GroES system—depicted in Fig. 10—being the most notable. In this system, GroEL forms a cage, and GroES acts as a lid, creating an environment where proteins can fold undisturbed by cellular interactions. Initially, the substrate protein binds to GroEL, and ATP triggers a conformational change that allows GroES to encapsulate the substrate inside the GroEL chamber. The protein then has around 10 seconds to fold as ATP is hydrolyzed. Recent research has shown that two proteins can be encapsulated simultaneously. A schematic of this cycle is shown in Fig. 11.

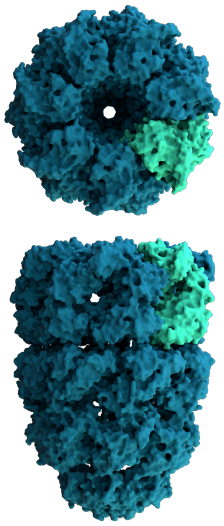

Figure 10: GroES-GroEL system structure

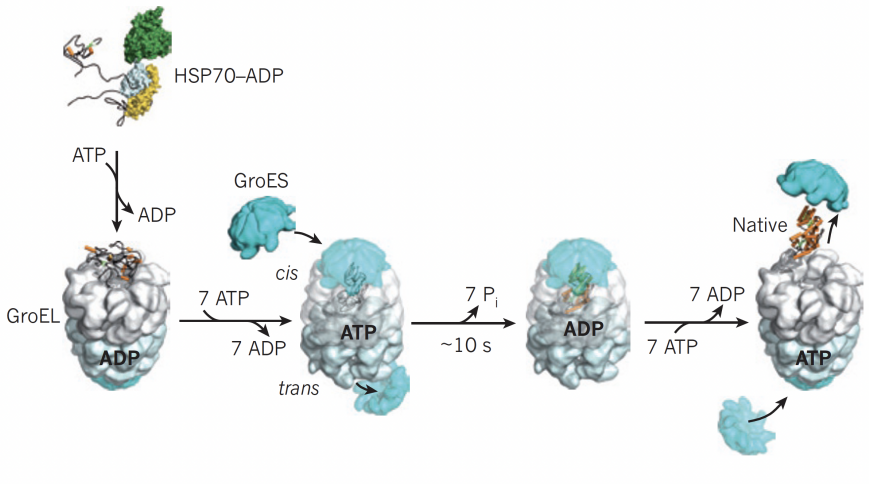

Figure 11: GroEL-GroES cycle [4]

The second type of chaperonins, such as the TRiC complex, have a similar function but don't require a separate lid like GroES. Instead, they have built-in "iris-like" structures that open and close to enclose proteins. It is slower than GroEL-GroES system.

HSP90

HSP90 is a highly conserved molecular chaperone, meaning that its structure and function have remained very similar across different species throughout evolution, highlighting the fact that it is important for survival. It has three main domains: the N-terminal domain, which binds and hydrolyzes ATP; the middle domain, responsible for client protein interactions; and the C-terminal domain, which is involved in dimerization (process by which two identical or similar molecules, called monomers, bind together to form a larger, more complex structure known as a dimer). The structural flexibility of HSP90 is crucial for its function, as it undergoes conformational changes during its ATP-driven cycle, switching between open and closed states to assist in protein folding. When ATP binds to HSP90, the N-terminal domains of two HSP90 molecules come together to form a dimer, creating

a closed shape. This closed form allows HSP90 to effectively assist in folding other proteins. When ATP is broken down, HSP90 releases the client proteins it was helping and gets ready to start the process again. HSP90 does not work in isolation but is part of a broader chaperone system, collaborating closely with other chaperones such as HSP70 and HSP40 [8]. The cycle of HSP90 is shown in Fig. 12.

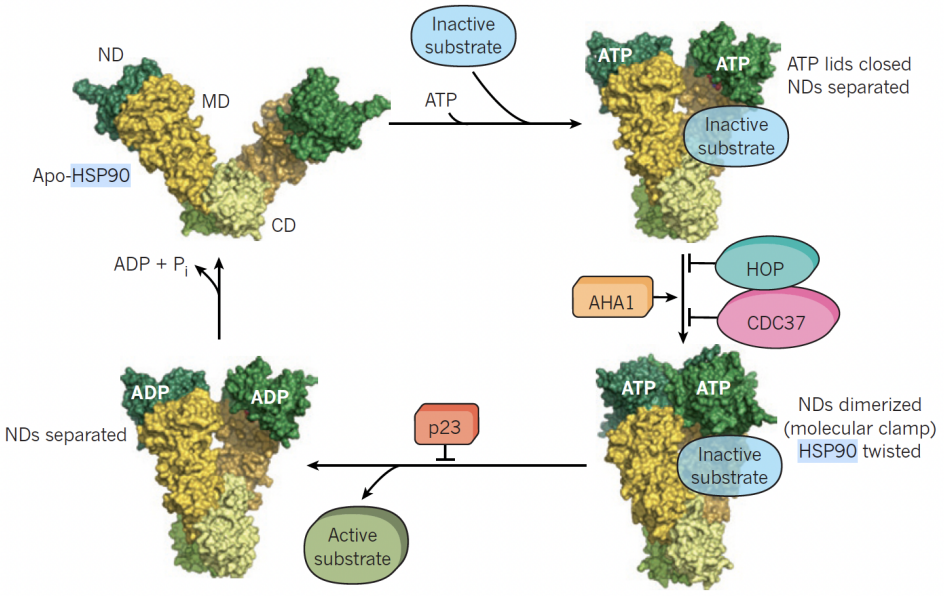

Figure 12: HSP90 cycle [4]

Importance of chaperones in translation process
During the process of translation, proteins are synthesized on ribosomes, but this step is relatively slow, exposing the nascent polypeptide chains to a high risk of misfolding. Because the ribosomal exit channel is narrow and prevents the formation of large structural elements, newly formed proteins can only start folding after they emerge from the ribosome. However, the incomplete proteins that exit the ribosome are highly susceptible to forming incorrect structures or aggregating due to their exposed hydrophobic regions. Chaperones play a crucial role here by binding to these nascent chains, preventing premature folding and aggregation. This allows the proteins to remain in a folding-competent state until they are fully synthesized and able to fold into their native structures.

3 Mathematical and physical framework for protein folding

Now that we have a good overview of the problem, we can dive into the framework used to properly model it. What is used to describe the problem is statistical mechanics. It is a branch of physics that connects the microscopic properties of particles (atoms, molecules) to the macroscopic properties we observe (such as temperature, pressure, and energy). It uses probabilistic methods to describe the behavior of systems with a large number of particles, where a purely mechanical approach would be too complex.

The system we consider is a protein surrounded by N_w water molecules, and we suppose that this number do not change, as well as the temperature T. Each microscopic state of our system is completely described by the coordinates and momenta (or velocities) of the atoms of the protein (denoted by x_μ and p_μ respectively, for the μth atom

of the protein, and with $\mu = 1, ..., N$) and those of water molecules (X_m and P_m, with $m = 1, ..., N_w$). This last assumption can be made if we suppose that the system obeys classical mechanics.

The space we use to work with these quantities is called a phase space. It is a way to describe the complete state of a physical system. For each particle in the system, it includes all possible positions and momenta the particle can have. In simple terms, it's like a map that shows every possible way a system can be arranged, both in terms of where things are and how they are moving. For a system with n particles, phase space is a $6n$-dimensional space that combines all these positions and momenta into one unified description.

Here, the phase space designates the space of all microscopic states, and is defined by $\Gamma := \Gamma_p \times \Gamma_w$, with Γ_p the phase space of the protein and Γ_w the one of the water molecules. In physics, we often tackle problems by considering the energy of the underlying systems. This makes the Hamiltonian, that plays the role of an energy function, a fundamental quantity. It is defined by the positions and momenta of the system and has a kinetic term (with momenta) and a potential term (with positions):

$$H(x_\mu, X_m, p_\mu, P_m) = \sum_\mu \frac{p_\mu^2}{2M_\mu} + \sum_m \frac{P_m^2}{2M_m} + V(x_\mu, X_m), \tag{1}$$

where we used M_m and M_μ as the atomic masses. Written like this, the Hamiltonian makes the calculations very complex and time-consuming. As we are interested specifically in describing the behavior of the protein, we can get rid of the water-related terms by averaging them out. This leads to an effective Hamiltonian that depends only on the protein's atoms and the temperature of the system, rather than on the exact positions and momenta of the water molecules. This effective Hamiltonian is easier to work with and helps reduce the computational effort required to study the protein. The result is that we capture the main effects of water on the protein without needing to track every detail of the water molecules. The potential energy in this effective Hamiltonian therefore includes the average influence of water on the protein. This approach works because, in many cases, the water molecules respond much more quickly to changes than the protein does, thus the protein's movement dominates the system. The new effective Hamiltonian is

$$H_{\text{eff}}(x_\mu, p_\mu; T) = \sum_\mu \frac{p_\mu^2}{2M_\mu} + W(x_\mu; T), \tag{2}$$

with $W(x_\mu; T)$ the new potential energy.

We discussed previously the concept of energy landscapes, which are used to describe the range of possible conformations a protein can adopt during folding. We also saw that in general, a biological system tries to minimize its free-energy. For reasons that go beyond the scope of this report, the effective potential energy $W(x_\mu; T)$ exactly corresponds to the free-energy of the system, and writes as

$$W(x_\mu; T) = -k_B T \ln \left(\int \exp[\ \frac{1}{k_B T} V(x_\mu, X_m)\ ] \mathrm{d}X_m \right), \tag{3}$$

with $V(x_\mu, X_m)$ the potential of the complete hamiltonian in (1), k_B is the Boltzmann constant, an ubiquitous constant in statistical mechanics. This expression of the free-energy can therefore be interpreted as an explicit definition of the energy landscapes.

Even though considering particles individually is not tractable, as systems are always moving and changing, it is still interesting to consider the probabilities of each possible state of a particle, in order to assess its average behavior. Instead of exact characteristics about each particle, we calculate average quantities based on all the possible states the system can be in. Statistical mechanics gives us tools, like probability distributions, to calculate these averages. This approach is essential for understanding complex systems like proteins, where we care more about the overall shape or energy of the system rather than the exact position of every atom at every moment. For example here we can consider the probability distribution of the position of the μth atom of the protein:

$$p(x_\mu) = \frac{1}{Z} \exp\left(-\frac{W(x_\mu; T)}{k_B T}\right), \tag{4}$$

where Z is called the partition function ensuring normalization i.e. $\int p(x_\mu) \mathrm{d}x_\mu = 1$. This probability density function can also be derived from the free-energy, as it is defined as the energy minus the entropy, and the latter is directly related to probabilities. Conformations with lower $W(x_\mu; T)$ are exponentially more likely. Hence, the energy landscape dictates the protein's structural preferences, meaning that the protein will most likely adopt conformations that minimize $W(x_\mu; T)$, leading to thermodynamic stability. When misfolding, proteins get stuck in a local minimum, and often can't escape by themselves and need chaperones.

To address Levinthal's paradox, we can consider models that are related to the kinetics of the folding process. One such model, introduced by Thirumalai [9], predicts the folding rate

$$k_f \sim \mathrm{e}^{-\sqrt{N}} \tag{5}$$

where N is the number of amino acids. This equation reflects the fact that larger proteins take longer to fold, due to the increased number of possible intermediate states that must be navigated. The kinetics of folding are also influenced by the height of the energy barriers between intermediates.

4 Challenges in protein misfolding and aggregation

Protein misfolding and aggregation are linked to numerous neurodegenerative diseases, such as Alzheimer's, Parkinson's, and Huntington's, making this a critical area of study.

Mechanisms of Misfolding and Aggregation
In the folding process, as it explores the energy landscape, a protein might get stuck in local-minima, some of which may correspond to states exposing hydrophobic parts, resulting in a misfolded conformation favoring aggregation with other hydrophobic molecules. Indeed, as we already mentioned previously, hydrophobic particles tend to stick together in aqueous environment, and this may lead to aggregates of multiple misfolded proteins clumped together. A notable example of this phenomenon is the formation of amyloid fibrils—highly structured, insoluble aggregates characterized by β-sheets aligned perpendicular to the fibril axis. An overview of all this is visible in Fig13.

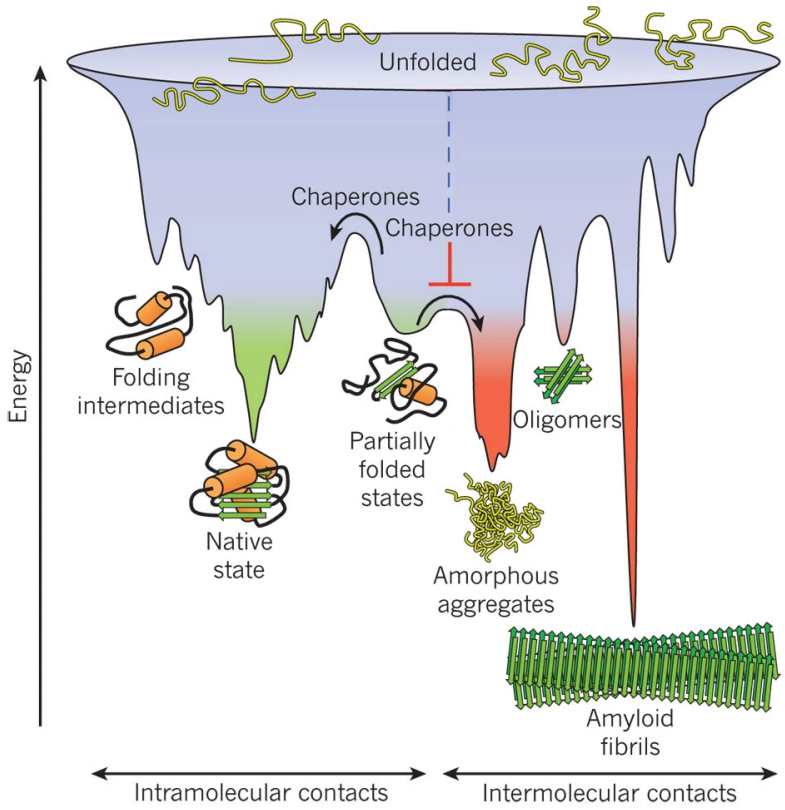

Figure 13: Illustration of an energy landscapes and different examples of conformations [4]

Many neurodegenerative diseases are caused by the accumulation of misfolded proteins. For instance, in Alzheimer's disease, amyloid-β peptides misfold and aggregate, forming plaques in the brain. Similarly, in Parkinson's disease, α-synuclein aggregates into fibrils that interfere with normal cellular processes. These aggregates disrupt cellular homeostasis, leading to cell death [4]. By facilitating correct folding or targeting misfolded proteins for degradation, chaperones help maintain proteostasis. Also, HSP90 plays a critical role in cancer by stabilizing and assisting the proper folding of numerous proteins involved in cell signaling pathways, including those that regulate cancer cell growth and survival. Its ability to interact with a wide range of client proteins, such as kinases and transcription factors, makes it an essential chaperone for the maintenance of oncogenic proteins, thereby promoting tumor progression. Understanding the balance between folding, misfolding, and aggregation is key to developing future therapeutic interventions.

Why is Proteostasis Crucial for Cell Health?
Proteostasis, or protein homeostasis, refers to the network of cellular processes that maintain the balance between protein synthesis, folding, refolding, and degradation. This network ensures that proteins are correctly folded and that damaged or misfolded proteins are efficiently degraded to prevent the accumulation of toxic aggregates [4].

The proteostasis network includes molecular chaperones, the ubiquitin-proteasome system (UPS), and autophagy. The UPS tags misfolded proteins for degradation through a process called ubiquitination, where proteins are marked with ubiquitin molecules and directed to the proteasome for destruction. Autophagy, on the other hand, removes larger aggregates or protein complexes by engulfing them in vesicles and degrading them within lysosomes [4].

A decline in proteostasis is associated with ageing and various neurodegenerative diseases. As organisms age, the efficiency of chaperone-mediated folding and the degrada-

tion pathways decrease, leading to the accumulation of misfolded proteins and aggregates. Thus, chaperones are crucial buffers that allow cells to manage stress, maintain protein balance, and minimize the toxic effects of misfolded proteins.

5 Advances in protein folding simulations and prediction

Recent advances in computational models have made significant strides in addressing the protein folding problem, driven by the application of modern artificial intelligence approaches.

From Molecular Dynamics to Coarse-Grained Simulations
Historically, molecular dynamics simulations were the primary computational tool used to model protein folding. These simulations track the movement of individual atoms over time by solving Newton's equations of motion. While these simulations provide atomic-level accuracy, they are computationally expensive, making it difficult to simulate folding processes. [3].

To overcome these limitations, coarse-grained models were developed. In these models, atoms are grouped into larger units, allowing for longer simulations with reduced computational costs. Coarse-grained models provide a more tractable approach to studying large proteins and complex folding processes by simplifying the representation of the protein without losing essential physical properties. For example, the Go model treats proteins as beads-on-a-string, where interactions are designed to favor the native state configuration [10]. These simplifications reduce the complexity of the energy landscape, helping researchers understand general folding pathways without the need for atomic-level detail.

AlphaFold and Deep Learning Approaches
The introduction of AI models, particularly AlphaFold, has revolutionized protein structure prediction. AlphaFold, developed by DeepMind, uses a deep learning approach to predict protein structures with remarkable accuracy [11].

A key feature of AlphaFold's architecture is the attention mechanism [12], which allows the model to capture long-range interactions between amino acids that are distant in sequence but close in 3D space. This is crucial for predicting the complex tertiary structure of proteins. Unlike traditional molecular dynamics simulations, AlphaFold bypasses the need for physical simulations by directly predicting the most probable structure based on statistical patterns learned from data.

Despite the significant advances, there are still limitations in current prediction methods. AlphaFold, for instance, excels at predicting the static structure of globular proteins, but it struggles with intrinsically disordered proteins and multi-domain proteins that do not have a single stable structure. Furthermore, dynamic aspects of proteins, such as folding pathways and intermediate states, are yet to be fully captured by AI models.

Conclusion

In conclusion, while substantial progress has been made in understanding protein folding and the roles of molecular chaperones, many challenges remain. Although we have identified general principles governing how proteins fold and how chaperones assist in

this process, the detailed mechanisms by which proteins avoid misfolding, and the precise molecular interactions that guide them through their energy landscapes, are still not fully understood. For instance, we know that chaperones help prevent aggregation and promote correct folding in crowded cellular environments, but the intricacies of how they discern between correctly folded and misfolded intermediates, and how they intervene in highly complex, multi-domain proteins, remain elusive. Additionally, while the energy landscapes of protein folding are conceptualized as funnel-like, with chaperones helping proteins overcome kinetic barriers, the specific pathways and molecular checkpoints that dictate folding fidelity, particularly in large, multi-domain proteins, require further exploration. Intrinsically disordered proteins, which defy traditional folding models, add further complexity. Another significant gap lies in the understanding of how certain proteins, especially those involved in neurodegenerative diseases like Alzheimer's and Parkinson's, become misfolded and aggregated, even in the presence of chaperones. Further research is needed to unravel how mutations, cellular stress, and aging influence these processes, and how chaperone networks might be modulated to prevent or reverse disease-related protein misfolding.

References

[1] Echenique, P. Introduction to protein folding for physicists. Contemporary Physics 48:2, 81–108 (2007). https://doi.org/10.1080/00107510701520843

[2] Koo, E. H., Lansbury, P. T., Jr., & Kelly, J. W. The structural basis of neurodegenerative diseases. Proceedings of the National Academy of Sciences 96(18), 9989–9990 (1999). https://doi.org/10.1073/pnas.96.18.9989

[3] Nassar, R., Dignon, G. L., Razban, R. M., & Dill, K. A. The protein folding problem: The role of theory. Journal of Molecular Biology 433, 167126 (2021). https://doi.org/10.1016/j.jmb.2021.167126

[4] Hartl, F., Bracher, A., & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. Nature 475, 324–332 (2011). https://doi.org/10.1038/nature10317

[5] Ellis, R. J. & Minton, A. P. Protein aggregation in crowded environments. Biol. Chem. 387, 485-497 (2006) https://doi.org/10.1515/BC.2006.064

[6] Hartl, F. U. Molecular chaperones in cellular protein folding. Nature 381, 571-580 (1996) https://doi.org/10.1038/381571a0

[7] Kampinga, H. H. & Craig, E. A. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nature Rev. Mol. Cell Biol. 11, 579-592 (2010). https://doi.org/10.1038/nrm2941

[8] Minari, K.; Balasco Serrão, V.H.; Borges, J.C. New Insights into HSP90 Structural Plasticity Revealed by cryoEM. BioChem 2024, 4, 62-89. https://doi.org/10.3390/biochem4020004

[9] Thirumalai, D., From minimal models to real proteins: time scales for protein folding kinetics. J. Phys. I, 5 (11), 1457–1467 (1995) https://doi.org/10.1051/jp1:1995209

[10] Taketomi H, Ueda Y, Gō N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. Int J Pept Protein Res. 7(6):445-59. PMID: 1201909 (1975)

[11] Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

[12] Vaswani, A. "Attention is all you need." Advances in Neural Information Processing Systems (2017)