The Physics of Protein Folding

1 Abstract

The physics of protein folding has been poorly understood since its inception in the late 1960's. This paper provides a brief history on the physics of protein folding, a description of some understood and accepted mechanisms and models of protein folding such as hydrophobic collapse or the two state model, as well as providing an overview of the interdisciplinary field of misfolded protein or amyloid fibril research. In addition to the explanation of the above theories and research areas, an understanding of some of the more prevalent experimental techniques is also presented, including Fluorescence Resonance Energy Transfer and Circular Dichroism.

2 Introduction

Even though the physics of protein folding has been studied since 1968[1], there is still very little known about the actual procedure. C. Levinthal suggested that if proteins were always found to exist in a specific native or ground state which corresponded to its energy minimum, then "it would be a consequence of biological evolution"[1], not a question of randomly testing possible folding configurations. The suggestion of a pathway of events which led to the specific conformation of each protein was intriguing and has fueled theoretical and experimental research in the field for over 40 years.

The actual folding process, and the physics which governs the proper folding is very important to understand, as proteins aid in almost all biological processes, and only perform their function in their native/folded state. Misfolded proteins do not only cease to carry out their biological function, but also aggregate and form structures called *amyloid fibrils.* These structures are known to be associated with physical ailments such as Alzheimer's disease, Parkinson's disease, and Kreutzfeldt-Jakobs disease.[2,3]

A protein is simply a polymer composed of a specific sequence of 20 standard amino acids.[4] A protein does not necessarily contain all 20 standard amino acids, but each one has a unique order of amino acids which dictates its uniqueness and function. Each

amino acid has a carboxyl group (CO) and an amide group (NH), as well as a side chain which determines the amino acid.[4]

The protein structure is studied on many different levels in modern biochemistry and physics. There is the primary structure which constitutes the sequence of amino acids, secondary which describes the 3D folding domains for the protein, tertiary describing the spatial arrangement of the secondary structure and quaternary structure describing the spatial arrangement of subunits of chain sequences in the protein.[4]

Although the order structure is important for describing the overall arrangement of the protein in its native state, it is important to understand the main types of secondary structures which may form when a protein folds. The secondary structure follows mostly from hydrogen bonds between the carboxyl and amide groups contained at the ends of each amino acid. The three main secondary structures a protein may form are random coil, α-helix and β-sheet. The random coil is self explanatory, and is just when the protein is not structured and coils due to entropy like a regular polymer. The α-helix structure is formed when the carboxyl groups form hydrogen bonds with the amide groups 4 amino acids away on the protein chain. This allows the side chains to stick out the side of the helical structure, keeping them as separated as possible. There is also the β-sheet structure which occurs when the chain is straightened, and then falls back upon itself creating a lined structure, while the CO and NH groups stick out 90 degrees from this and form hydrogen bonds to connect the layers. The α-helix and β-sheet structures are given in schematic form in Figure 1.

2.1 The Levinthal Paradox

The original suggestion of a folding pathway which was asserted by C. Levinthal was originally suggested because proteins were found to refold from a disordered state in seconds. Therefore, for a protein to reach its' biologically active, or native state, whether that state exists at a total energy minima or in a metastable local energy minima, it most likely follows a pathway.[1] This is because if the protein were to sample random configurations in order to reach its energy minimum, even for a short protein chain of only ~100 amino acids, it would have to test $2^{100} \sim 10^{30}$ conformations. Therefore for only a picosecond test time of each conformation, it would take 10^{18} seconds, or over 30 billion years, which is on the order of the age of our universe, as opposed to the seconds

which are taken to fold proteins. Therefore, it can be assumed that due to the speed at which proteins fold, a pathway must exist in order for the protein to fold quickly without error.[1,5]

2.2 Kinetic and Thermodynamic Hypotheses

The kinetic and thermodynamic hypotheses were suggested in the original article by C. Levinthal, and postulate on the relative energy of the native state to other possible conformations for the protein.[1]

The thermodynamic hypothesis is that the native state exists as the state in normal physiological conditions which minimizes its Gibbs free energy globally. This means that the native state is a unique minimum determined by the protein's amino acid sequence as well as the environment it is in, regardless of the energy bestowed to the protein.[5] Anfinsen et al. provided support for this theory in early work using a mixture of mostly improperly folded protein due to inappropriate environmental conditions (being in 8M urea). The proteins in this environment were only operating at ~1% the activity of the native protein. However, when the protein was removed from the urea and placed in a proper solvent, the same group of proteins eventually formed a homogeneous solution of protein in the native state. This process was determined solely by the free energy minimized by the proteins folding into a more stable energy state, and this state is determined by the amino acid sequence only.[5]

The kinetic hypothesis is an opposite view, and predicts that the native state is not at a total energy minimum, but is actually at a local minimum which is determined by the folding pathway which achieves a low free energy the quickest.[1] The thermodynamic hypothesis has lots of support, as when it comes to computational work, assumptions are always made that the thermodynamic hypothesis is correct to simplify calculations, yielding results which are usually quite descriptive, and give very strong agreement with experimental findings.[6,7,8,9] However, even though sequence structure of proteins gives accurate results for finding the structure of proteins theoretically at room temperature, there seems to be validity in the kinetic hypothesis.[10] As will be discussed later, amyloid fibrils have been shown to be much more stable than the native state of the protein, and thus it is possible that the amyloid fibrils are actually in a more global minimum of free

energy, but require extra energy (higher temperatures) to form, supporting the kinetic hypothesis.[10]

3 Proposed Folding Mechanisms and Models

3.1 Hydrophobic Collapse

An original consequence of the thermodynamic model was that since the sequence of amino acids in the proteins determines the native state and secondary structure of the protein, the protein folds due to "local" interactions. This means that it forms the secondary structure first, and then arranges itself spatially into its tertiary structure, as shown in Figure 2A.[9] However, even though proteins are fairly complex molecules, they are still fundamentally polymers. Therefore, it is natural for them in regular physiological conditions to undergo hydrophobic collapse in which they form a globule as opposed to staying as a long stretched chain for an extensive period.[10,11,30] This was studied computationally by Sali et al., where they used a Monte Carlo algorithm in order to computationally study the process of folding for a short 27 bead self-avoiding chain.

They used a system where they knew that the native state for the chain would be at a global minimum, in conjunction with the thermodynamic hypothesis. In Figure 3 is an example of one folding sequence using this method during the study. You can see that initially, the particle undergoes a massive hydrophobic collapse which is on a much shorter timescale than the rest of the folding process. Then on a much larger timescale, the particle seems to search for the intermediary transition folding state which will lead it to its native state.[11] This result agrees with the idea of a specific folding pathway for the protein, as the rate limiting step for the folding process to the native state is the search for the intermediary state.

Dill et al. recognized patterns set forth by groups such as Sali et al, and proposed a model which contained the above described properties in opposition to previous models governed by local interactions, forming secondary structures first.[10] He called this the "collapse" model in opposition to Ptitsyn's "framework" model,[9] and postulated that folding relied on non-local instead of local interactions. A diagram of the two models may be seen in Figure 2. This collapse model made sense in conjunction with previous physics of polymers, as hydrophobic collapse in proper physiological media is consistent with other polymer conformation and theory. The model he set forth was a "simple

exact" model, meaning it approximated each amino acid as a lattice point, or bead, as in the previous study discussed by Sali et al.[11]

They discussed years of experimental evidence which led to the theory that collapse, or non-local interactions, was the dominant force in protein folding.[10] This includes evidence that β-sheet structures had very few local interactions involved in its structure, the free energy difference for helix formation is small, making their formation unstable at physiological temperatures,[12,13] and the tendency for formation of secondary structures is more dependant on the solvent than on the amino acid sequence.[14] The last result indicated that the local interactions which supposedly caused secondary structure formation initially is not very strong compared to the force imposed by the natural environment. This once again agrees with the well known drive for polymers to undergo hydrophobic collapse in proper solvents in order to achieve maximum entropy and minimize free energy.

3.2 Multiple Folding Pathways and Intermediary States

Studies have shown that proteins may fold along multiple folding pathways in order to change conformation from the random state to native state and vice versa.[15-18] Therefore instead of a specific pathway leading to the native state, there are many pathways which lead to the folding of the protein in a quick fashion. Originally proposed by Dill et al. was that the energy landscape for the folding of the protein was much like a funnel, hence the name "funnel landscapes".[18] This idea of funnel landscapes came from the thermodynamic hypothesis which stated the native state was at an energy minimum. Therefore for a protein to fold, it needed to follow a directional path across its energy landscape. Since it had been shown that regardless of the path, the protein in the proper conditions would eventually find its way from the unfolded state to the native state,[5] the energy landscape would take on a funnel shape, where the folding would be inevitably directional toward its energy minimum.

An example of a funnel landscape is given in Figure 4. You can see that the folding may follow one of many routes by this model, and some will be quicker than others, but eventually the protein will conform to its native state when under the proper environmental conditions.

Rhoades et al. studied single adenylate kinase protein molecules which were placed in conditions favouring spontaneous transitions between folded and unfolded states as found in previous ensemble measurements of adenylate kinase.[15] Single molecule studies are a fairly recent possibility in protein folding.[15,19] In this particular study, they used a technique called Fluorescence Resonance Energy Transfer (FRET) which features a donor dye attached to one end of the protein chain and an acceptor dye to the other. The protein is illuminated with laser light which excites the donor dye. When the donor dye is close enough to the acceptor dye, there is energy transfer without radiation to the acceptor dye, which is then excited and emits radiation of a separate wavelength. This is dependant on the acceptor dye having an excitation energy in the range of the emission energy for the donor dye.

By measuring the intensity of fluorescence emitted by the donor and acceptor dye, we define an efficiency of energy transfer as follows:

$$E_{ET} = \frac{I_a}{I_a + I_d}$$

where I_a is the acceptor emission intensity detected, and I_d is the donor emission intensity detected, we can determine how folded the molecule is, as the energy transfer from donor to acceptor is extremely sensitive to distance of separation:

$$E \sim \frac{R_0^6}{R_0^6 + r^6}$$

where R_0 is the Forster's radius, usually on the order of 1-5 *nm*, and r is the distance of separation for the donor and acceptor dyes. It is worth noting here that for high efficiency, the distance between dyes must be small, and thus the protein is in a folded form.

In order to measure only single molecules, Rhoades et al. studied the protein inside a unilamellar vesicle, which is essentially just a biological holding chamber. The vesicle was then tethered to a glass slide for observation so they could monitor the folding of the protein inside for extended time, as individually the single proteins would move too fast for observation. Before reporting their results, the group was sure to ensure that the protein did not interact with the vesicle wall in any way by measuring the rotation of the protein, and ensuring that it was not binding to vesicle surface.[15,22]

In Figure 5 we see two examples of experiments done with single protein molecules. In Figure 5A-B, we see that the protein goes from the folded state (high E_{ET} means dyes are close together) to some intermediate folded states, as opposed to completely unfolding. These states must be in local minima and fairly stable under the conditions of the experiment. However in Figure 5C-D, we see that the protein seems to be in the folded state, and then unfold quite quickly to a random state where E_{ET} is very low.

These two different protein folding kinetics for the same conditions show that there are many paths for the protein to follow when folding to the native state. In addition to these, there were also experiments where there was slow unfolding/folding kinetics found for individual molecules which may be seen in Figure 6.[15] Overall, they considered many causes for the heterogeneity of the folding pathways, and either disproved the possibilities or concluded that they were not large enough errors to cause such deviations from a single path, revealing that multiple pathways was the only option.

Recently, in another study, Peng et al. discovered through atomic force microscopy that the T4 lysozome unfolded with parallel unfolding pathways as well.[16] In order to make sure the AFM tip did not affect the native state of the T4 lysozome, they attached four other proteins on each side of the T4 lysozome. The AFM tip was attached to these outside protein molecules, and the molecule was then pulled away from the AFM tip, and the chain of T4 lysozome surrounded by the other proteins was unfolded by the pulling force. Examples of force-extension curves from the study are given in Figure 7.[16] You can see that in the curve, there are 8 spikes, which seem to be fairly evenly spaced. These spikes constitute the unfolding of the 8 protein units attached to the T4 lysozome.

Therefore, we see that the unfolding preceding these 8 unfolding events must be due to the unfolding of the T4 lysozome. Now we see that there is a two state unfolding in these diagrams, but in Figure 8, we can note that there is 3 state unfolding for the same molecule, giving more solid evidence that multiple unfolding pathways exist. The three-state unfolding also comes in various forms, indicating that not only is there different numbers of unfolding steps, but also many different combinations which seem equally desirable when under the same conditions. This sums up to the same length extension

before the first attached protein (GB1 units) unfolds, meaning the final result of unfolding is always the same despite the path taken.

Consequently, multiple pathways have a large amount of support from theoretical and experimental studies on protein folding, indicating that the original idea of a single folding pathway is incomplete.[1]

3.3 Two State Folding

Since protein folding is complex, and there is evidence of multiple folding pathways, it is very difficult to analyze and interpret protein folding data. Therefore in order to obtain quantitative understanding of protein folding, groups have studied the two state folding model in conjunction with short stranded proteins.[20-24]

The two state folding model relies on the fact that the protein may exist in only two states: folded (F) and unfolded (U). The model is set up much like a chemical reaction, with a before and after state, with folding and unfolding rates. Thus we can describe the kinetics of folding as well as the equilibrium transitions in a straightforward method.

Essentially, the model treats the two states as a reaction with an activation barrier. Therefore, the equilibrium condition can be easily seen to be from elementary chemistry:

$$\frac{[f]}{[u]} = e^{-(E_f - E_u)/kT} \qquad (1)$$

where E_f and E_u are the energy barriers for folding and unfolding respectively, k is Boltzmann's constant, and T is the temperature of the system.

The kinetics may also be studied, as the rates of folding and unfolding are:

$$\begin{aligned} k_f &= Ae^{-E_f/kT} \\ k_u &= Ae^{-E_u/kT} \end{aligned} \qquad (2)$$

where k_f and k_u are the folding and unfolding rates respectively, and A is just a constant. Using these kinetic rates, groups may obtain a quantitative understanding of the energy barrier for folding of the small proteins.

Yang et al. carried out a study which used an ensemble of a small protein, meaning a large amount of them were studied at the same time.[23] The experiment consisted of perturbing a large group of protein molecules in their native state with a sudden increase in temperature. This temperature increase forced the proteins to unfold,

and when the solution was placed back at physiological temperature, the relaxation rate of the proteins back into the folded state was measured. By taking the ratio of the unfolding rate to the folding rate, they were able to determine the activation barrier over which the folding occurs for the proteins. Figure 9 shows an example of the exponential relaxation of the particle from the unfolded to the folded state, and gives the average amount of time a particle spends unfolded when placed in proper folding conditions. This example gives a rate of ~0.05 μs^{-1} and indicates an activation barrier on the order of $2.1kT$.[23]

Ensemble measurements allowed groups to test the two state theory and to give quantitative description for the kinetics of folding. However, a real revolution for the two state model came when single molecule measurements were able to show directly that it was a valid theory.[22] In ensemble measurements, they could show that there were two states available, but not say with complete certainty that there were never any intermediates. These results were then compared readily to previous ensemble measurements of the same protein, and the results were comparable, therefore giving validity to all previous ensemble measurements.[24]

Rhoades et al. used the same technique described earlier for single molecule study inside vesicles with FRET by trapping Thermotoga Maritima (CspTm) protein inside a tethered vesicle.[22] Using FRET measurements, they were able to tell when the protein inside was in its folded or unfolded state as in the earlier experiment. An example of the efficiency measurement for a trial is given in Figure 10A,B, displaying single or multiple transitions. The binning time for the measurements was 20ms, but even when binned at 100 μs it was found that transition from one state to the next was immediate. There were no intermediates found, and thus the transition was necessarily on a timescale of less than 100-200 μs.[22]

Figure 11 gives the histogram of the different time intervals preceding transitions between states. Fit with an exponential, they found a rate constant of 0.62 ± 0.26 s^{-1} compared to the ensemble study result of 0.39 ± 0.02 s^{-1}.[24] Therefore the results are in agreement within error. The error for the single molecule measurement is understandable, as we see that the amount of counts for the histogram are low when looking for a statistical value.

The two state model is a valid method for evaluating quantitatively the folding of small, fast folding proteins. Although, in reality you are almost guaranteed to see intermediate folding steps if you decrease your detection time enough. The model does however give us a way to quantify the folding process, and ignore the complexity of the other models for folding. This simplicity is desirable in a field of research where complexity is far too apparent.

4 Amyloidoses – Misfolded Proteins Causing Disease

Understanding how proteins fold is an interesting and important topic, as it is an essential part for biological function and is poorly understood. However, it may be more important to understand exactly why some proteins do not fold.

Misfolded proteins(forming amyloid fibrils) are extremely dangerous to physical and mental health. Most notable are the neurodegenerative diseases which arise from misfolded proteins such as Alzheimer's disease (caused by amyloid β-protein), Parkinson's disease (α-Synuclein) and Creutzfeldt-Jakob disease (Prion protein).[2]

Amyloid fibrils are not individual misfolded proteins, but are aggregates of the same misfolded protein.[2] All amyloid fibrils, independent of the misfolded protein forming them, are mostly β-sheet structure.[2,3,25] Therefore, even for predominantly α-helical proteins, when forming amyloid fibrils they will go through a α-helix to β-sheet transition.[25] Booth et al. found this to be true for lysozyme amyloid fibril formation, and proposed a mechanism for the formation of the fibril which is displayed in Figure 12.[25] This proposed mechanism proposes that the aggregation of the misfolded proteins and formation of the fibril arises from proteins in their intermediate, molten globule state.[25] This means that they are not completely denatured, but are more free form with some secondary structure. When left in that form, the proteins will aggregate to form a large beta sheet structure to reduce free energy creating an undefined branch structure from the leftover parts of the chain not used in the sheet.[25]

Figure 13 gives a smattering of results which are combined in the study by Fandrich et al. to show the measured spacing of the side chain and the main chain for different amyloid fibril structures in relation to their volume.[26] It is shown that the side chain spacing stays essentially the same regardless of the Van der Waals volume of the molecule, which is just the volume of the side chains. However, the main chain spacing

depends linearly upon the Van der Waals volume. This data is more proof that the beta structure is almost identical in the formation of all amyloid fibrils. The beta sheet structure which is pictured in Figure 1 shows that the main chains line up and hydrogen bond together, and the side chains would just extend out from the main chains. This means that regardless of the size of the side chain, they will still be just as near to each other, however the main chains will have larger distances between them as some side chains will take up more space in between main chain strands, pushing the beta structure apart.[26]

In studies of amyloid fibrils, Fandrich et al. used results from an experiment by Santoro et al. which studied proteins before and after temperature increases at different concentrations of denaturant solution.[26,27] Denaturant is just a solvent which causes a material to alter from its native state, so protein in the presence of denaturant unfolds and tends to take on the stretched and free form shown in Figure 12. Denaturant studies have been done many times, however in this study they used knowledge given by Blondelle et al. that an increase in temperature in the presence of some proteins will cause them to form amyloid fibrils. This means that when the proteins were cooled back to room temperature after being held at an elevated heat, the amyloid fibril will form, and stay in that state.[28]

Performing this irreversible process in different concentrations of denaturant, they were able to study the form of the protein molecule in its natural and amyloid state to see how stable the form of the protein was. They used Circular Dichroism (CD) in order to determine the secondary structure of the protein and fibril at the different denaturant concentrations, and thus the stability.[26]

CD is a quick and cheap experimental method which can be used to determine the general secondary structure of protein. The procedure relies on the interaction of the protein structure with circularly polarized light. In order for circularly polarized light to interact with a material, it must be chiral in nature. Fortunately, amino acids are naturally chiral, and each of the three secondary structure forms: α-helix, β-sheet and random coil are also chiral. Figure 14 shows CD spectra for proteins made up primarily of each of the above structures. Notice that the α-helix has a largely negative value at 222*nm* for the CD

spectra, the Beta sheet is slightly less negative, and the random coil polymer has a value of $\sim 0 \ \frac{\deg \cdot cm^3}{dmol}$.

By studying the values of the CD spectra at $\lambda \sim 220\ nm$, they obtained the data in Figure 15. You can see that the amyloid fibril form of the protein starts off at a less negative value, as expected since it has β-sheet structure as opposed to the α-helix structure of the native protein.[26] More importantly though, the β-sheet amyloid fibril is much less dependant on the denaturant concentration than the native protein. The consequence of this is that the amyloid fibril structure must be in a much more stable state than the native protein. Figure 16 gives a schematic representation of how this might work in the idea of energy levels and activation barriers.[26]

These results seem to contrast the concept of the thermodynamic hypothesis. As seen in Figure 16, the amyloid structure at room temperature has a much lower energy than the native state, but requires much more energy to overcome the activation barrier. This gives support to the kinetic hypothesis, as the energy of the native state is not at a global minimum, but one which is quickest and easiest to access at a low energy.[1,5] Even though amyloid structure requires the aggregation of many proteins, not just the folding of a single protein molecule, it is still worth commenting that the thermodynamic hypothesis could be incorrect in certain scenarios. If we consider the possibility of proteins interacting, the amyloid fibril structure is a much lower but harder to access energy minimum.[26]

Recently, Jahn et al. studied the idea of folding vs. the formation of aggregate amyloid fibrils.[29] This review gave a similar view point to that of Fandrich et al.,[26] indicating that the amyloid fibril state is at an extremely deep energy minima as shown in their energy landscape of Figure 17. It can be seen that the surrounding intermediates for the amyloid fibril have higher activation barriers, and thus are more stable intermediates than those surrounding the native state. Therefore it makes sense under normal physiological conditions for unaltered proteins to fold along the native state path, as it is a funnel type path, forcing the protein to its folded state. Also, it requires much higher entropy to even access the region of the amyloid fibrils, and thus requires some sort of mutation of the protein sequence or different physiological conditions in order to form.[29]

Although a general understanding is obtained from previous results pertaining to the energy landscape associated with folding and misfolding of proteins, very little is understood to why the proteins aggregate and misfold.[29] Understanding the energy landscapes is an initial step toward understanding amyloid fibrils, however in physiological conditions the temperature is not increased spontaneously by 30 ^{o}C and the pH isn't shifted by +2. Therefore, the initiating mechanism for the aggregation and misfolding inside biological systems, as well as the mutations which cause the formation of amyloid fibrils are ill understood or not understood at all, and are of great interest.[29]

5 Conclusions

Overall, it is clear that the physics of protein folding is still quite poorly understood, despite all the current research being performed in the field. Even the most basic argument which established the field of protein folding (kinetic vs. thermodynamic hypothesis) is still unresolved over 40 years later.

Due to paper limitations, I was able to review only a few more established theories and mechanisms for protein folding. The two state model was originally used to study most protein folding, but upon greater resolution times of experiments, it was found that it only properly described extremely short protein chains for quick scale time resolution.[22] It is seen that many different mechanisms are possible for protein folding, depending on the structure, sequence and environment conditions. There is no current model which may completely describe the folding process for even one protein exactly, especially for longer and more complex proteins. This is currently an area of intense research, attempting to develop a model of the folding pathways for a specific protein.[16]

I was also able to give an overview of the misfolding processes involved with the fabrication of amyloid fibrils.[26-29] The physics which governs this process is extremely important to understand, as it has direct implications to human physical and mental health. Progress has been made in understanding the structure of the misfolded proteins, as well as a general understanding of the kinetics of misfolding. However some of the most important questions are still remaining to be answered, such as how is misfolding initiated in physiological environments where environment may not be altered?[29] Studies include changes in temperature, pH and level of denaturant in order to create amyloid fibrils, but none of these drastic changes are likely in a physiological system.[26-29]

In conclusion, protein folding is an essential function in biological systems, and the implications of the folding procedure are widespread. A complete understanding of protein folding is under development, and further research of the physics of protein folding is crucial for other scientific fields to advance their knowledge of the process, and eventually develop preventions for many amyloidoses.

6 Diagrams and Figures

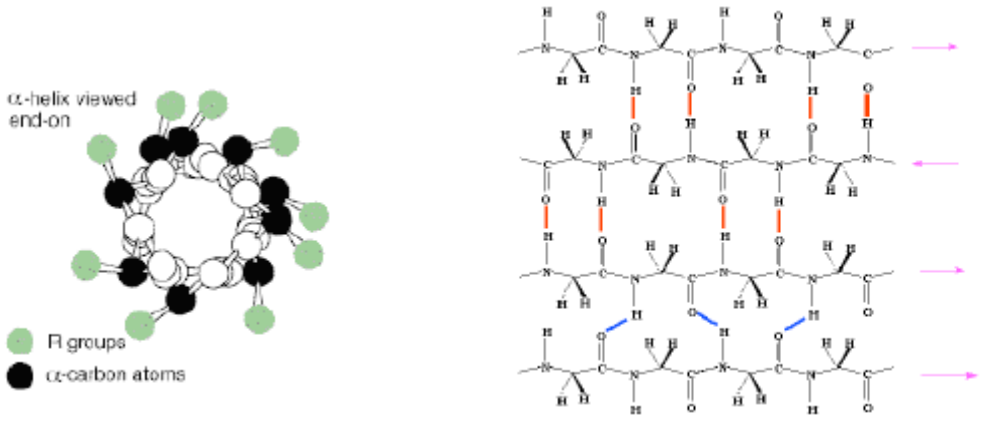

Figure 1: α-helix structure (left) shows the spiral of the central carbon atoms (black), and the side groups (green) extending outside from the ring. The helix extends into the page. The right gives the β-sheet structure created by the main carbon chain of amino acids extending parallel to each other and connecting by hydrogen bonds (blue and red lines). Adapted from [4].

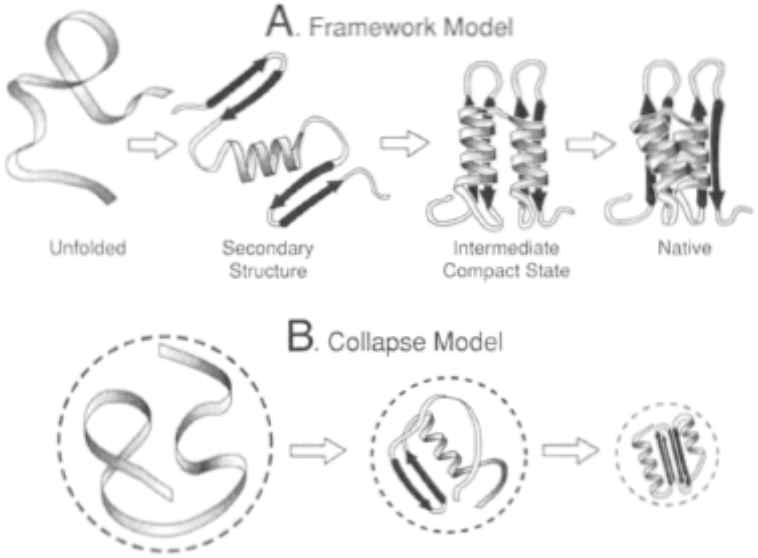

Figure 2: The framework model (A) describes the model outlined by Ptitsyn et al. [9] which dictates that secondary structure form first, and is followed by tertiary structure formation. The collapse model (B) outlines the initial collapse of the molecule into a smaller structure with tertiary spatial organization, and then initiates secondary structure formation. Adapted from [10]

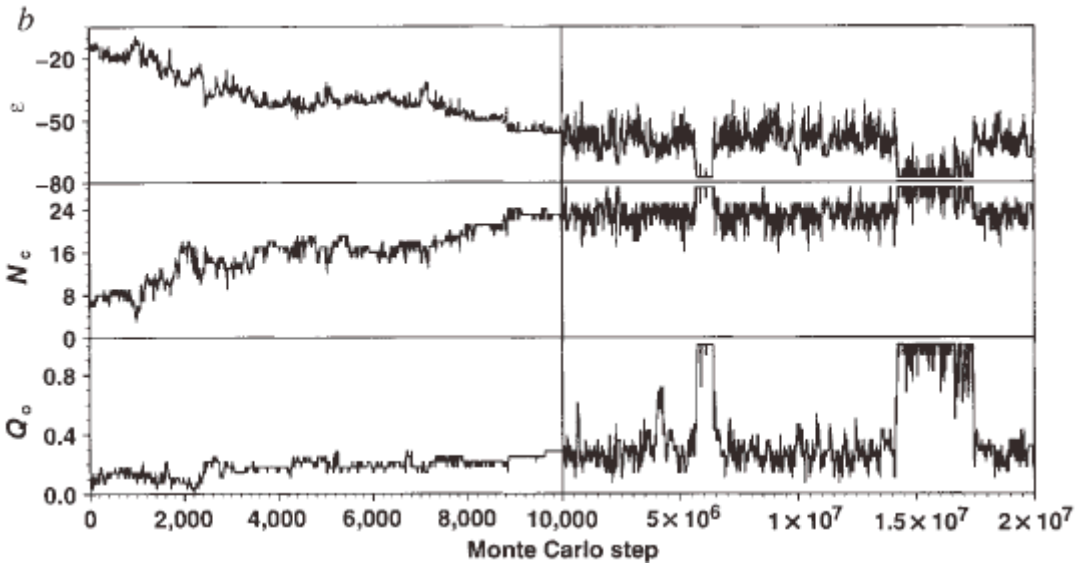

Figure 3: Example of one folding sequence found with Monte Carlo method. Note that initially there is a very fast hydrophobic collapse before taking an extended period of time to find a transition state to the native state (energy minimum assumed). The top is a graph of energy vs. Monte Carlo step, N_c gives the number of contacts, and Q_o gives the fraction of contacts which are the same as the known native state. Adapted from [11].

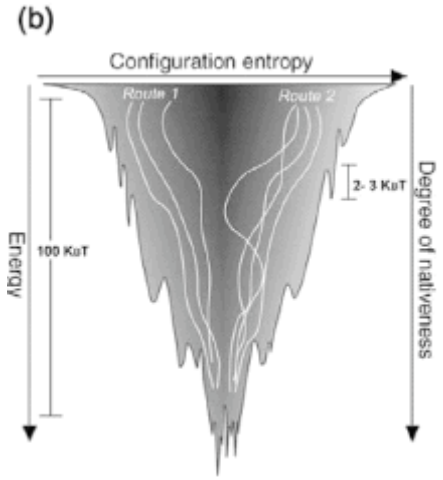

Figure 4: Example of funnel landscape adapted from [18]. You can see that many routes are possible toward native ground state, but some are quicker than others, and some involve possible intermediate states (local minima at edge of funnel.)

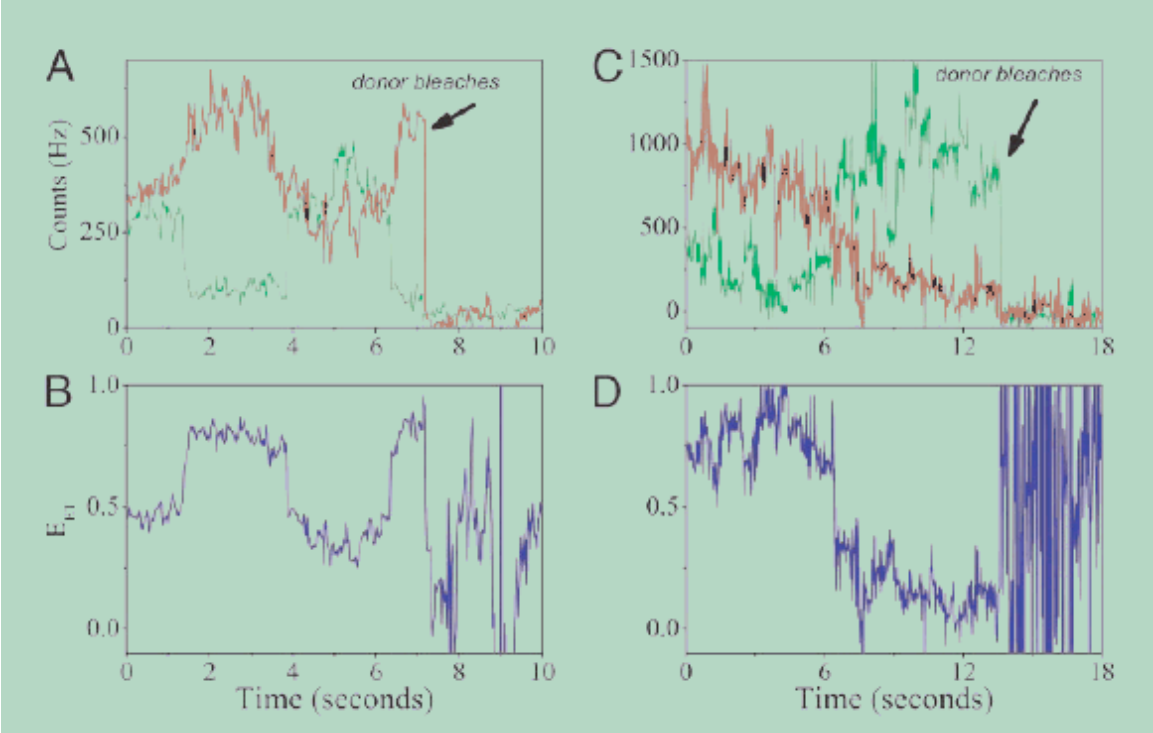

Figure 5: Two examples of FRET measurements recorded for single molecule experiments for AK protein folding. One of them shows an efficiency plot that indicates the protein bounces between intermediate and native state (A-B), where the other plots indicate the protein transitioning from the folded to the completely unfolded state (C-D). E_{ET} is the energy transfer efficiency, outlined in an equation above. Adapted from [15].

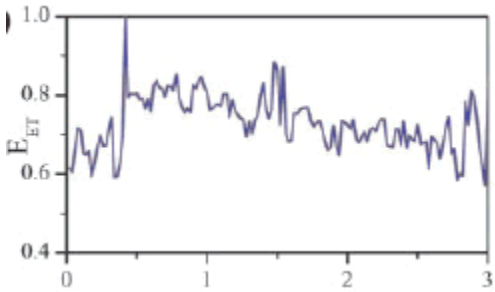

Figure 6: Slow folding kinetics. Note that the unfolding after the initial folding at ~0.5 seconds is slow and fairly constant. Since both data sets in Figure 5 and the data here are for the same protein, it is easy to see that there are many folding pathways, and some take much longer than others, sometimes requiring intermediate states between folding and unfolding. Adapted from [15]

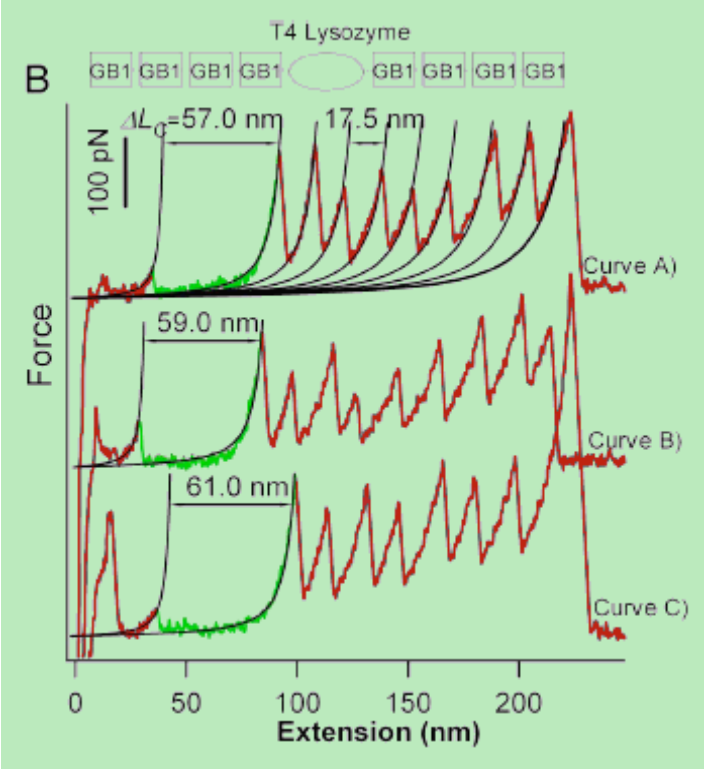

Figure 7: Force-extension curves for AFM measurements of the unfolding of the T4 lysozyme. The 8 evenly separated spikes in the data seen for all three sets are the unfolding of the GB1 proteins attached to the lysozyme, and therefore the unfolding features prior to this are those of the T4 Lysozyme. The unfolding seems to occur in a two state fashion for these curves, but they occur for different lengths, and thus the unfolding follows different paths for the two step unfolding. Adapted from [16].

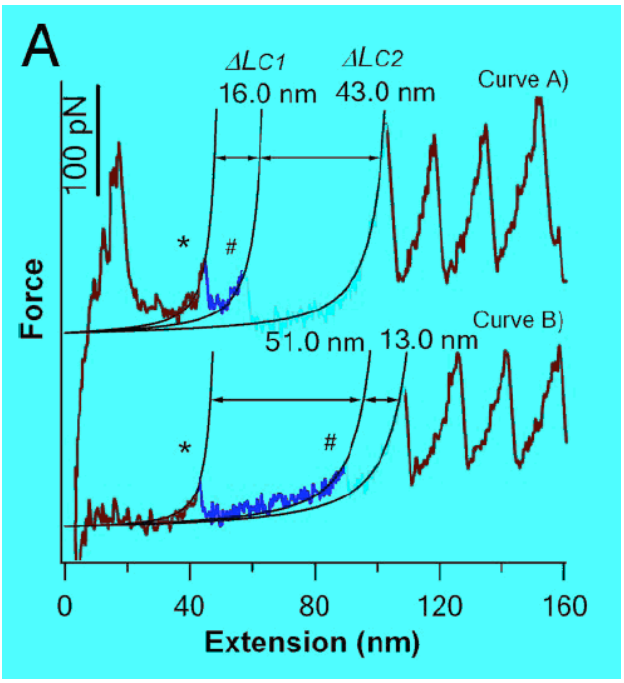

Figure 8: Three state process of unfolding for same T4 Lysozyme protein as Figure 7. This three state process is seen by the three spikes in the force extension curve, indicating three partial unfolding events. This indicates multiple pathways of unfolding, supporting the funnel landscape model as opposed to the more basic classical model of linear folding. Adapted from [16]

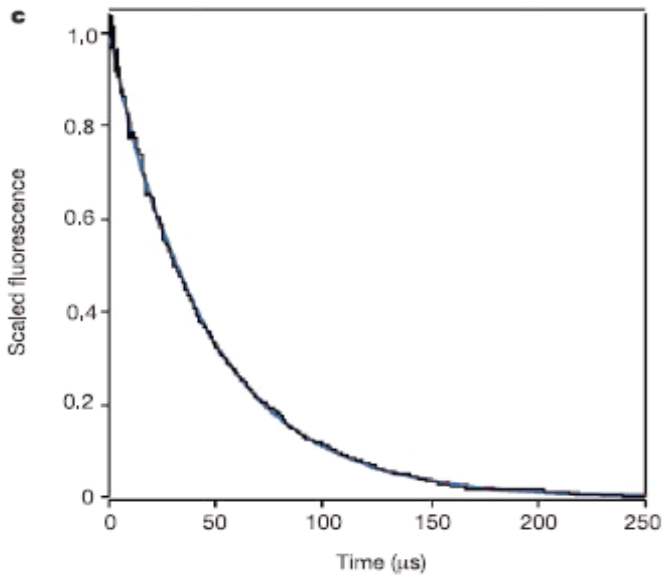

Figure 9: After excitation to the unfolded state by increased temperature, the fluorescence of the protein is monitored over time, where fluorescence occurs for the unfolded state, but is not possible for the native state. Therefore a decay rate was found to be ~0.05 μs^{-1}, indicating an activation barrier for native state folding for the protein of ~2.1kT using the two state folding model, applicable to small proteins. Adapted from [23]

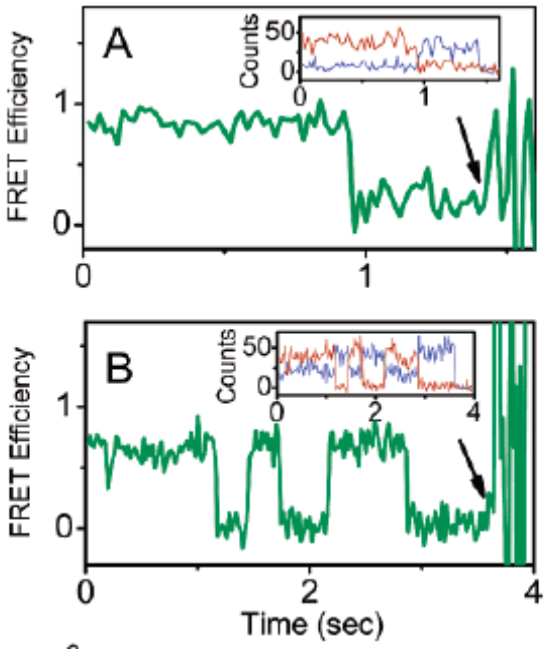

Figure10: FRET efficiency graphs indicating sudden jumps between two states, the unfolded (lower efficiency) and folded (higher efficiency) states for Thermotoga Maritima (CspTm). It is clear that the sudden jumps indicate a two state folding process. The insets are the donor and acceptor fluorescence intensity (blue and red respectively). Adapted from [22].

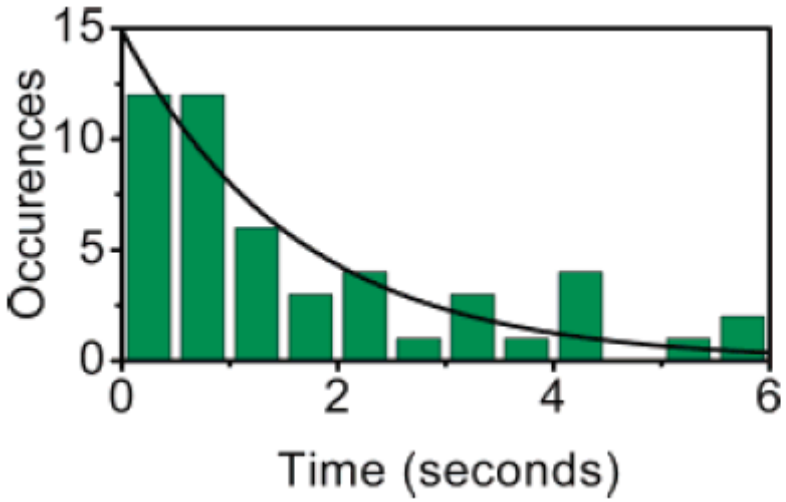

Figure 11: Histogram of times preceding folding process. Exponential fit gives close fit to rates found for same protein in previous bulk protein experiments, with many proteins monitored at once. Gives folding rate of $0.62 \pm 0.26\ s^{-1}$ compared to ensemble rate of $0.39 \pm 0.02\ s^{-1}$. Adapted from [22]

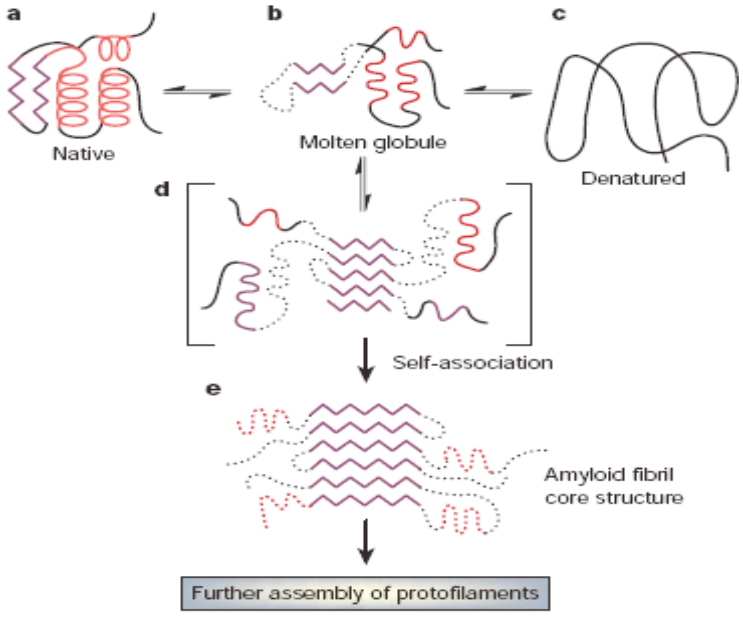

Figure 12: β-sheet amyloid structure forms from the molten globule state, where secondary structures are partially formed (β-sheet is blue, α-helix is red), but some randomness of protein structure is still evident (broken lines). The fully denatured state is not useful for the misfolding, as you need beta structure, and thus molten globule may find that aggregation and formation of amyloid fibril minimizes energy. Adapted from [25]

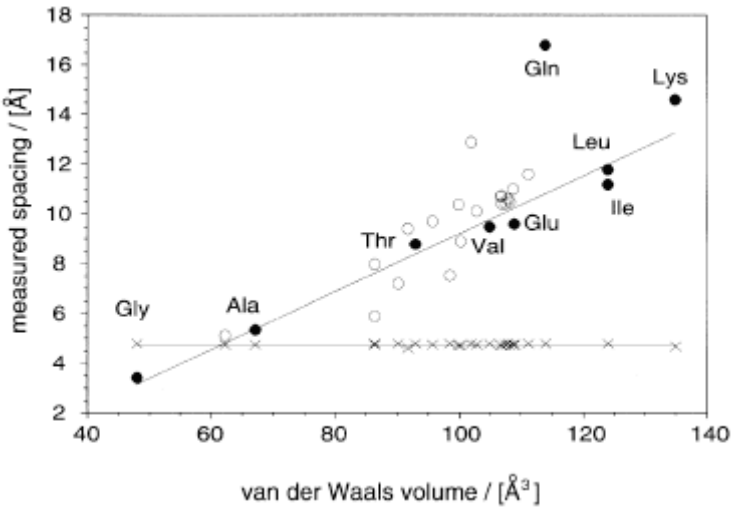

Figure 13: Measured spacing of main chain elements which are consistent for all protein structures (circles) and side chain elements varying dependant on protein sequences (crosses). Plotted as function of van der Waals volume, or size of side chains. Adapted from [26]

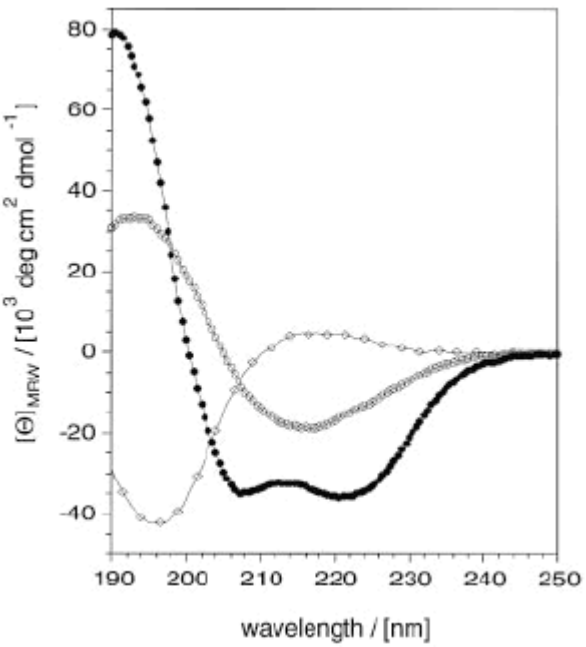

Figure 14: CD spectra for majority α-helix (open circles), β-sheet (black circles) and random coil (diamond) structure. Not the large differences of values at ~220 *nm*. This is used in later study to distinguish between the structures of the protein under denaturant conditions. Adapted from [26].

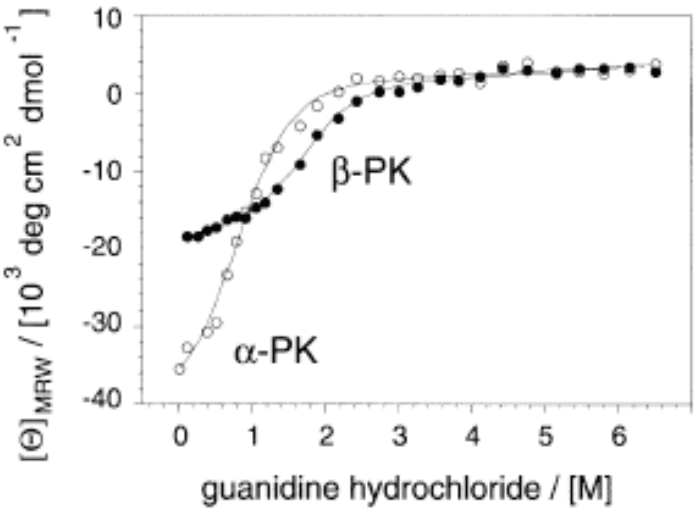

Figure 15: Value of CD spectra at ~220 *nm* for PK protein in amyloid and native form. At different denaturant concentrations, the protein was measured in native state, and then heated up to form amyloid fibrils, and cooled to see what structure it had. Since at 220 *nm*, CD spectra for random coil is ~0, the quicker the form of the protein approaches zero with increasing denaturant, the less stable it is. Can see that amyloid form (β-PK) is more stable under denaturant, giving energy landscape theory in Figure 16. Adapted from [27]

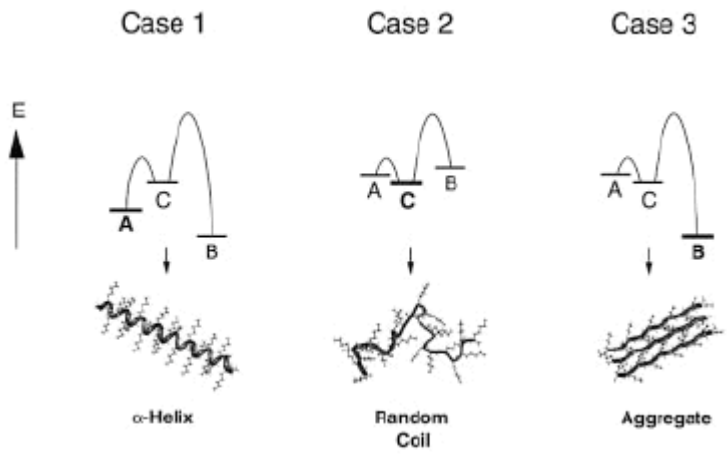

Figure 16: Case 1 shows that the protein takes on alpha helix form (A) at regular temperatures, as even though amyloid form (B) is at lower energy, energy barrier is too high. In Case 2, when denaturant is increased enough, energy of random coil state (C) becomes favorable over both the amyloid/aggregate state and α-helix(A). At high temperatures, amyloid fibril may be formed as increased energy allows for overcoming the energy barrier. Upon lowering temperature, we see that due to the deep well, it makes sense the protein stays in amyloid fibril form.

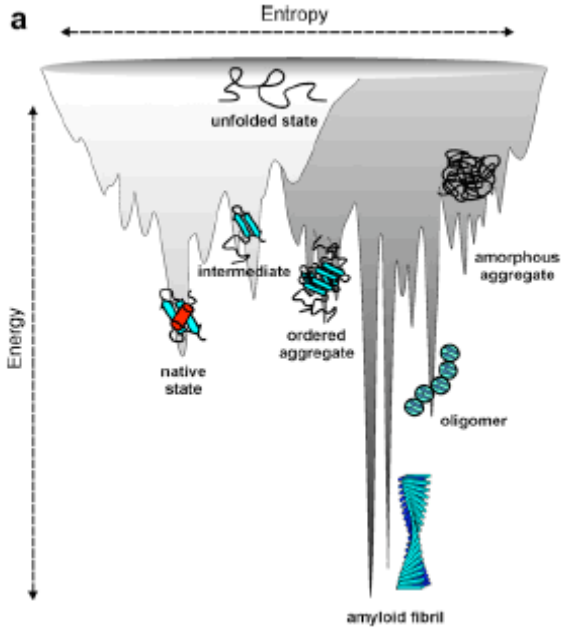

Figure 17: Energy landscape for amyloid fibrils and native state, showing that at high entropy situations, if you can fold from there, amyloid fibrils may be formed in deep energy wells. However, more likely, in natural lower entropy situations, the native state has a funnel landscape with some possible intermediate states. Adapted from [29].

7 References

1) Levinthal, C. *Extrait du Journal de Chimie Physique*. **65,** 44-45 (1968)

2) Selkoe, D.J. *Nature.* **426**, 900-904. (2003)

3) Fandrich, M. et al. *Nature*. **410,** 165 (2001)

4) Voet, D. et al. *"Fundamentals of Biochemistry"*. John Wiley & Sons, Inc. (1999)

5) Anfinsen, C. *Science*. **181**, 223-230 (1973)

6) Onuchik, J.N. et al. *Current Opinion in Structural Biology.* **14,** 70-75. (2004)

7) Park, S. et al. *Biophysical Journal*. **93**, 4108-4115. (2007)

8) Bai, Y. *Biochemical and Biophysical Research Communications.* **305**, 785-788. (2003)

9) Ptitsyn, O. *J. Protein Chem.* **6**, 273-293 (1987)

10) Dill, K.A. et al. *Protein Science*. **4**, 561-602. (1995)

11) Sali, A. et al. *Nature*. **369**, 248-2521 (1994)

12) Scholtz, J.M. et al. *Annu. Rev. Biophys. Biomol. Struct.* **21,** 95-118 (1992)

13) Zhang, X.J. et al. *Biochemistry*. **30,** 2012-2017 (1991)

14) Reed, J. et al. *Natl Acad Sci USA.* **90,** 6761-6765. (1993)

15) Rhoades, E. et al. *PNAS.* **100**, 3197-3202. (2003)

16) Peng, Q. et al. *PNAS.* **105,** 1885-1990 (2008)

17) Matysiak, S. et al. *Archives of Biochemistry and Biophysics*. **469,** 29-33 (2008)

18) Dill, K.A. et al. *Nature Struc. Biol.* **4**, 10-19 (1997)

19) Lipmann, E.A. et al. *Science*. **301**, 1233-1235. (2003)

20) Jackson, S.G. *Folding and Design*. **3**, R82-R91. (1998)

21) Jackson, S.G. et al. *Biochemistry.* **30,** 10428-10435. (1991)

22) Rhoades, E. et al. *Journal of American Chemical Society.* **126**, 14686-14687 (2004)

23) Yang, W. et al. *Nature.* **423,** 194-197 (2003)

24) Schuler, B. et al. *Nature*. **419**, 743-747. (2002)

25) Booth, D.R. et al. *Nature* **385,** 787-793 (1997)

26) Fandrich, M. et al. *EMBO*. **21**, 5682-5690. (2002)

27) Santoro, M.M. et al. *Biochemistry*. **27,** 8063-8068 (1988)

28) Blondelle, S.E. *Biochemistry*. **36,** 8393-8400 (1997)

29) Jahn, T.R. et al. *Archives of Biochemistry and Biophysics*. **469,** 100-117 (2008)

30) Nath, U. et al. *Biochemistry*. **36**, 8602-8610. (1997)