Energetics of linear polymerization: a study of amyloid fibril elongation

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Nicola Vettore

from Vicenza, Italy

Düsseldorf, November 2019

From the Institut für Physikalische Biologie
of Heinrich Heine University Düsseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Supervisor: Prof. Alexander Kai Büll
Co-supervisor: Prof. Gunnar Schröder

Date of the oral examination: December 16th 2019

I declare under oath that I have produced my thesis independently and without any undue assistance by third parties under consideration of the "Principles for the Safeguarding of Good Scientific Practice at Heinrich Heine University Düsseldorf"

Date	Signature

Acknowledgements

I would like to thank my PhD supervisor, Prof. Alexander Kai Büll, for his valuable guidance. Not only did he help me develop a methodical and critical thinking, but his enthusiasm and passion for science motivated me every day to face the challenges that research holds.

This project would not have been the same if not for my friend and colleague Emil Dandanell Agerschou: his dedication and work ethic inspired me to improve every day; his patience and unreserved support made it possible.

My thanks are due to Lena Nora Mangels and Alberto Coden, whose role in my work extends beyond their scientific contribution.

I would like to thank all the people that are and were part of the AG Büll, in particular Alessia Peduzzo, Rebecca Sternke-Hoffmann and Marie Schützmann.

I am grateful for the thorough and insightful comments of Dr. Lothar Gremer and the support of Prof. Gerhard Steger. Many thanks as well to Christine Röder, Prof. Gunnar Schröder and Prof. Wolfgang Hoyer for the collaboration, which was as fruitful as it was stimulating.

These years would have been less bright if it wasn't for the wonderful atmosphere of the Institut für Physikalische Biologie. My colleagues were always available for a conversation, and, when needed, a beer. I hope my future work environments will hold colleagues and chocolate cakes as precious as the ones of the IPB.

Finally, I thank my parents, Carlo e Bertilla, and my brother, Riccardo, for their love and support. Grazie per aver creduto in me.

SUMMARY

Supramolecular linear polymers are a class of non-covalent molecular assemblies of relevance in biology and chemistry. Proteins such as actin, tubulin, and flagellin assemble into filaments that are part of this class of polymers. In the cell they have different roles, from the spatial organization to motility. Due to the importance of these functions, thermodynamic characterization of these systems was carried out already in the 1950s. Over the last decades, a different class of protein polymers, amyloid fibrils, have garnered increasing interest due to their relevance in different pathologies, such as Alzheimer's and Parkison's disease. Amyloid fibrils are protein filaments characterised by a typical cross-β structure. The kinetics of amyloid formation are objects of numerous studies, but the thermodynamic behaviour is still insufficiently researched. This thesis builds on the current knowledge to lay the foundation for a detailed description of amyloid fibril elongation along its reaction coordinate.

In order to be able to describe a reaction it is necessary to evaluate the stability of the species involved. In the amyloid growth reaction this means characterizing the stability of the soluble and amyloid state. In Chapter 2, I compare two models to describe the stability of the amyloid state: the isodesmic and the cooperative one. The isodesmic model, the only one currently in use, proved to be insufficient in describing the complexity of amyloid polymerization. For this purpose, I introduced the cooperative model, through which I was able to characterize the stability of the amyloid state and to describe the salt dependency of the amyloid stability.

The different components that contribute to the stability of a species, namely entropy and enthalpy, can be studied by measuring how the stability changes with temperature. In Chapter 3, I analyse the thermal stability of two fibrillar systems, glucagon and PI3K-SH3, as a function of temperature. Two novel models that exploit kinetic out-of-equilibrium behaviour are applied to describe the enthalpic contribution to the stability of the two systems.

Alongside the characterization of the amyloid stability, in Chapter 4 a structure of PI3K-SH3 amyloid fibril solved by cryo-EM is presented. The architecture of amyloid fibrils has started to be resolved at atomic level just recently. This contribution to the field allows the rationalization of numerous previous findings related to the mechanism of PI3K-SH3 fibrillation.

Ultimately, in Chapter 5, I combine the novel knowledge from the previous chapters to provide the first description of the transition state of the PI3K-SH3 amyloid elongation reaction. By using five mutants to probe different parts of the sequence, the role of the mutated residues in the structural ensemble of the transition state can be inferred. This is possible through the use of the newly-established thermodynamic measurement and the already-established kinetic measurement techniques, which together allowed me to adapt the Φ-value analysis of protein folding to the amyloid elongation reaction. The possibility to investigate the transition state of the amyloid growth reaction paves the way for more detailed characterization of the amyloid fibrillation mechanism.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

Supramolecular assemblies are ordered complexes of molecules held together by non-covalent interactions [1]. Their formation is dictated by reversible and highly directional interactions among the monomers involved. This term is used to describe a wide class of molecules that can form spheres, sheets or rods. These assemblies are object of study in physics, chemistry and biology and their unique properties make them optimal building blocks for advanced functional materials [2, 3].

The self-assembly of linear supramolecular polymers is found in both synthetic and biological systems. Cellular motors, cyto-skeleton and semiconducting polymers are the result of non-covalent unidirectional polymerization of molecules such as proteins or small aromatic compounds [4–7]. Moreover, this class of polymers is related with diseases, as sickle cell and amyloidoses [8, 9].

Protein filaments were among the first classes of supramolecular linear polymer to be discovered [10, 11]. As the field of cell biology rapidly evolved from the 1950s, protein assemblies were found to have many different biological roles, like the formation of hairs, muscular contraction, cellular motility and spatial organisation [4, 10, 12–15]. It was clear that these assemblies needed a complex regulatory machinery to exert their function in their physiological environment, but at the same time some intrinsic property of the polypeptides involved had to be present, since not all proteins are able to assemble. Actin was the first filamentous protein assembly to be fully characterized, thanks to the ease of extraction and to the variety of roles in different biological contexts. Oosawa and collaborators applied and adapted the polymerization framework established by Flory to actin polymerization, leading to the first description of the thermodynamic and kinetic aspects of a polymerizing protein system [12, 16, 17]. Later two other main cellular components, tubulin and flagellin, were identified as exerting their functions as supramolecular polymers and were therefore characterised in this framework [4, 13, 14]. All functional filaments have well established roles in the composition of organelles, structures and tissues, and therefore their formation and dissolution are strictly regulated and coordinated at the cellular and organism level [18, 19]. The proteins involved in these filaments are able to interact with several other biological components, that, usually through phosphorylation and dephosphorylation of the polymerizing monomers, promote and inhibit the polymerization [20, 21]. Mis-regulations of these polymeric systems are usually hallmarks of cancer, in which the cells are often characterized by changes in plasticity and mobility [22, 23].

More recently, a new class of aromatic compounds raised interest for its ability to form supramolecular one-dimensional polymers [1]. These compounds, that for the sake of clarity I will call polymers of small aromatic molecules, show mechanical properties similar to covalent polymers and are capable to form semiconductors, all maintaining the versatility of a non-covalent system [2, 7]. These polymers' reversibility is due to the intermolecular interactions formed by hydrogen bonds and non-covalent stacking of aromatic rings. The stability of the molecules, the solubility in a wide range of buffers and the versatility of synthesis made these compounds ideal to study and characterize in depth different mechanisms of polymerization [24–26].

Finally, in the landscape of supramolecular assemblies, amyloid fibrils are nowadays one of the most researched, but still not fully understood systems [9]. Amyloid fibrils are filamentous assemblies of proteins with a link to the onset of pathologies such as Alzheimer's and Parkinson's disease. Furthermore, in the last decade, small peptides and hormones were found to be stored in the form of amyloid assemblies, while bacteria exploit amyloid fibrils as a class of functional filaments [9, 27, 28]. Amyloid fibrils are formed by proteins that adopt a new conformation as they assemble on the filaments. This conformational change is called misfolding, in opposition to the functional folding reaction, and the final conformation is known by the name of cross-β structure [29]. The misfolding step is what differentiates this class of protein supramolecular filaments from actin, tubulin and flagellin, that instead fold in their native structure and later participate in the formation of filaments [30]. Moreover, non-amyloid protein filaments have a well-established role in cells and have been selected through evolution to polymerize in a regulated way. In contrast, disease-related amyloid polymers form as a consequence of unregulated events.

Deeper understanding of the polymerization mechanism of amyloid fibrils is fundamental to tackle the disease-related issues [29]. In this task, the use of models and methodologies coming from the studies of functional proteins and small molecules assemblies can help to fill the gap. An overview of the current knowledge on amyloid polymerization will follow in the next sections, together with a comparative analysis of the different aspects that are shared with the other supramolecular linear assemblies.

1.1. Amyloid elongation

Amyloid fibrils are protein based polymers formed through a nucleation step, in which several monomers interact to initiate the "minimal fibril". The "minimal fibril" is defined as the smallest linear molecular assembly that can grow through monomer addition with the same rate of any bigger linear assembly. Growth of the fibril occurs by monomer addition in a reaction called elongation [29]. In the elongation reaction two molecular species are involved: a fibril and a monomer. The fibril is a linear assembly of n monomers (F_n) that by the end of the elongation reaction will incorporate a new monomer and reach the size n+1 (F_{n+1}). The monomer (M) starts from a more or less structured state in solution and misfolds, getting incorporated at the end of the fibril. The reaction is bi-molecular and its rate depends linearly on the species concentrations as follows [31]:

$$k_+[F][M] \tag{1.1}$$

where k_+ is the rate constant of elongation. This formulation simplifies the reaction, as it does not depict the variety of species in the population of fibrils. It is often used to study the reaction steady-state, in which the population of fibrils remains constant, while the concentration of

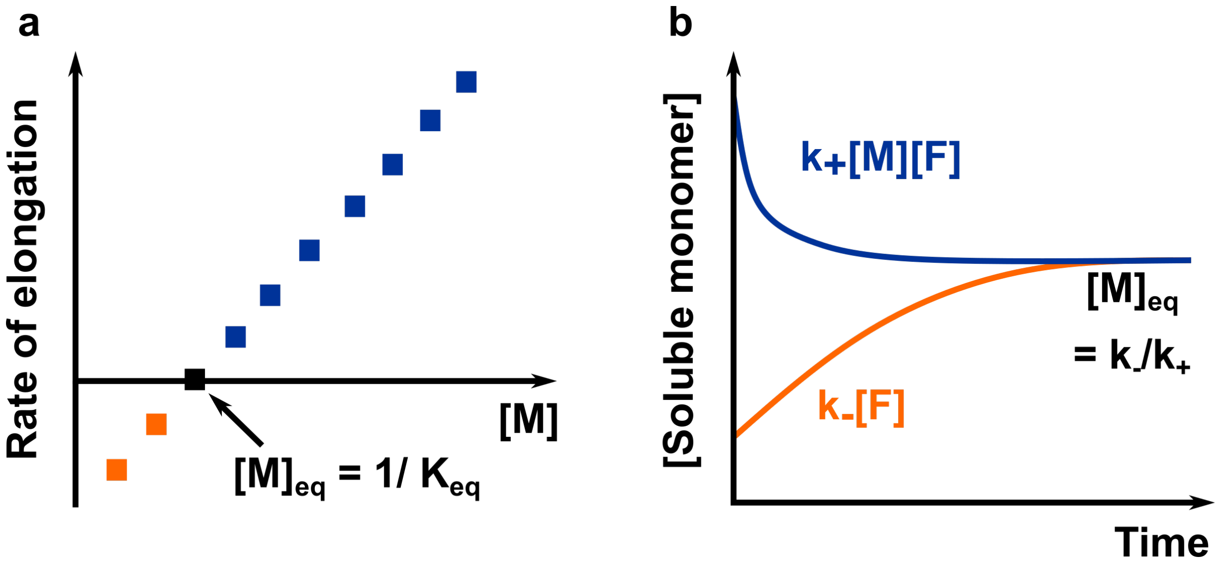

Figure 1.1: Correlation between critical concentration $[M]_{eq}$ and rate of elongation. **a**) The rate of elongation depends linearly on the concentration of monomer in solution. In blue the positive rates correspond to events of elongation. In orange, the negative rates show events of depolymerization. **b**) Change of monomer concentration ($[M]$) over time. An amyloid fibrillar system at equilibrium is characterized by a critical concentration of soluble monomer ($[M]_{eq}$). If the actual concentration of monomer is higher ($[M] > [M]_{eq}$), then the fibrils will elongate until the equilibrium is reached (blue line). If the actual concentration of monomer is lower ($[M] < [M]_{eq}$), then the fibrils will dissociate (orange line)

monomers is depleted over time. In the case of secondary phenomena such as fragmentation or secondary nucleation, this relation does not suffice. For the scope of this work these two secondary processes do not constitute a relevant topic, so they will not be further elaborated on.

The elongation reaction is reversible and the opposite reaction is called depolymerization or dissociation. It consists in the loss of a monomer from the fibril end. The dissociation reaction is dependent only on the concentration of the fibrils, and it is expressed as follows:

$$k_-[F] \tag{1.2}$$

where k_- is the rate constant of depolymerization. Also in this case the linear dependency of the depolymerization on the concentration of fibrils is explicit . In both the polymerization and depolymerization reaction, every fibril can potentially react at both extremities. However, single particle experiments showed a polarized growth with preference of an extremity of one order of magnitude over the other [32, 33].

This leads to the formulation of the effective rate of the reaction, which is defined by the change of the monomer concentration over time:

$$\frac{d[M]}{dt} = k_-[F] - k_+[F][M] \tag{1.3}$$

When the same amount of elongation and depolymerization events occur, the amount of monomer in solution does not change any longer and the equilibrium of the reaction is reached. This leads to a constant pool of soluble monomers being in equilibrium with the monomers composing the fibrils, that is formalized in the following equation:

$$k_+[F][M] = k_-[F] \tag{1.4}$$

$$K_{eq} = \frac{k_+}{k_-} = \frac{1}{[M]_{eq}} \tag{1.5}$$

where K_{eq} is the equilibrium constant that is inversely related to the concentration of monomers outside of the fibrils. The concentration of monomers at the equilibrium is defined as critical concentration ($[M]_{eq}$) and represent the concentration of monomers below which no fibrils are present (Fig. 1.1a). The formulation of the equilibrium in the terms above presented does not allow us to define how the system behaves below the critical concentration. The concentration of fibrils, $[F]$, is not described in terms of what it actually is, an ensemble of monomers, and therefore it cannot be entirely depleted. In the next sections the full expression of the species in solution is described, but for the description of the steady-state of the reaction it is currently not needed.

The steady-state is defined as the state in which the reaction proceeds with the same rate and $d[M]/dt$ is constant. For the elongation reaction this happens when $[M] >> [M]_{eq}$ and consists in the linear depletion of soluble monomers over time. The addition of monomers on the fibril ends continues with the same pace until $[M]$ approaches $[M]_{eq}$. In the opposite case, when $[M] << [M]_{eq}$ the reverse steady-state can be observed, in which the the amount of monomer linearly increases as the fibrils depolymerize (Fig. 1.1b).

The rate of addition of new monomers on the end of the fibrils in the steady-state depends on several factors. One factor is intrinsic to the reaction itself: the residency time [34]. The residency time is defined as the time in which the fibril end is occupied while the monomer rearranges on its end. The residency time defines the saturation behaviour of the reaction as a function of concentration. The longer the fibrils are occupied with a monomer rearranging on their ends, the lower will be the monomer concentration at which no fibrils are idle and available for a new reaction. A correlation was found with the length of the sequence, as well as it can be influenced by external factors such as denaturants and salts [35, 36].

Other external factors influencing the steady-state are salts, the degree of non-amyloid structure still present in the monomer and the temperature. Amyloid fibrils are homopolymers and salt screening of the repulsion between the monomeric units results in a faster rate [37]. The conformational rearrangement of the monomer can be modulated by the addition of denaturant [35, 38–40]. Finally the temperature influences the energy available to cross the activation barrier of the reaction [38, 41–51]. All these external agents can act in both directions, by speeding up or slowing down the reaction. The salts can salt out the monomers at high concentrations, causing a decrease of the rate. Denaturants can destabilize the fibrils if used in very high concentrations, possibly ending up to promote the depolymerization reaction. Similarly the temperature can act by speeding up the elongation reaction, but at high temperatures the depolymerization reaction may be favoured, resulting in an overall reduced elongation rate.

The influence of these factors on the steady-state of the elongation reaction is correlated with their effects on the energy level of the transition state ($\Delta G^{\ddagger}$). The rate constants of both elongation and depolymerization is defined by the height of the energetic barrier represented by $\Delta G^{\ddagger}$ in the following way:

$$k = \Gamma \exp - \frac{\Delta G^{\ddagger}}{RT} \tag{1.6}$$

where Γ is the kinetic pre-factor and R and T are respectively the gas constant and the temperature. The energy $\Delta G^{\ddagger}$ is a defined by two major contributions, enthalpy, $\Delta H^{\ddagger}$ and entropy $\Delta S^{\ddagger}$, accordingly to the relation:

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger} \tag{1.7}$$

This barrier is composed by two different conceptual steps, the conformational shift of the monomer and the formation of all the required contacts with the fibril ends, in order to become itself the new fibril end. The two steps together can be defined by only one barrier that defines a concerted and continuous shift from the soluble conformation to the cross-β one [34]. A detailed analysis of the influence of the temperature on $\Delta G^{\ddagger}$ shows that the transition state is mainly entropically driven, while its formation is influenced by electrostatic repulsion, as salts speed up the elongation rate [37, 49].

Further knowledge of the transition state of the elongation reaction has been reached only via simulations, while no further work has been conducted to elucidate it by an experimental point of view [52–54]. An existing framework to study the transition state of protein conformational changes is the Φ-value analysis [55]. Through this methodology it has been possible to study the folding pathways of many different proteins and to shed light on the complexity of the folding reaction. The experimental methodology consists in characterizing several mutants of a specific protein in their kinetic and thermodynamic aspects. The effect of the mutation is indicative of the role of the mutated residue in the transition state [56]. The amyloid elongation reaction can be considered conceptually as a folding reaction, with the peculiarity of being bi-molecular [30]. Methodologies to study the reaction kinetics are well assessed, but the field is still lacking a solid experimental framework to measure the stability of the amyloid fold. This work aims at filling this gap, building on the current knowledge of thermodynamic measurements of linear supramolecular assemblies.

1.2. Stability of a supramolecular one-dimensional assembly

Probing the equilibrium of a reaction is fundamental to understand a chemical system. The distribution of the molecules at the end of a reaction defines their relative stability and allows us to describe how the system behaves when out of the equilibrium. The relative stability of the species is usually modelled by the Gibbs free energy, ΔG^0, which is linked to the equilibrium constant K_{eq} by the following relation:

$$\Delta G^0 = -RT \ln K_{eq} \tag{1.8}$$

where R and T are the gas constant and the temperature. Knowing the equilibrium constant is therefore necessary to describe the stability. This is straightforward for reactions like the folding of a polypeptide chain in a two state system. As every molecule can be either in one or the other of two states, folded (N) or unfolded (U), the stability of the folded state is given by the ratio of the two populations at the equilibrium:

$$K_{eq} = \frac{[N]}{[U]} \tag{1.9}$$

Intuitively, the folded state is much more stable, as many more protein molecules are folded at the equilibrium under normal conditions. At the same time, being a mono-molecular reaction, the equilibrium is dependent on the isolated conformational change of every polypeptide chain. The equilibrium is therefore not influenced by the total protein concentration, if not by indirect effects like crowding at high concentrations [57].

Many of the polymeric filamentous systems described in the previous section are composed of molecules which are either part of the filament or in its monomeric soluble form [1, 17, 30, 58]. To describe the system stability all the species in solution have to be considered. As the filaments are quasi one dimensional, the soluble monomer $[A]$ is able to interact with other monomers, dimers $[A]_2$, trimers $[A]_3$, tetramers $[A]_4$, up to n-mers $[A]_i$. The simplest model to describe all these interactions assumes that all are regulated by the same equilibrium constant K_e:

$$\begin{aligned} K_e[A]^2 &= [A]_2 \\ K_e[A][A]_2 &= [A]_3 \\ K_e[A][A]_3 &= [A]_4 \\ &\dots \\ K_e[A][A]_i &= [A]_{i+1} \end{aligned} \tag{1.10}$$

As the equilibrium constant is connected to the stability of the system as in Eq. 1.8, every interaction is defined by the same energetics (Fig. 1.2a). For this reason the model is called isodesmic. The isodesmic model was first introduced by Oosawa and collaborators to describe the polymerization of protein filaments [17]. The solution of the equation system above presented allows us to describe the equilibrium constant as a function of the total concentration of protein $[A]_{tot}$:

$$K_e = \frac{1}{[A]} - \sqrt{\frac{1}{[A][A]_{tot}}} \tag{1.11}$$

If $[A]_{tot} >> [A]$, the solution of the equation converges to $K_e = 1/[A]$, which corresponds to the approximation of the system previously shown in Eq. 1.5. For both the approximation and the isodesmic model, the knowledge of the soluble monomer concentration at the equilibrium is fundamental to determine the stability of the polymeric system.

Through the isodesmic model, protein filaments like the one formed by tropomyosin can be fully described [59]. However, most of the other supramolecular linear assemblies composed by proteins show a different behaviour. Actin assemblies were studied both in the kinetic behaviour and at the equilibrium, leading to the observation that a nucleation step was necessary to be included in the description of the polymerization mechanism [12, 60]. The concept of a nucleus as the starting assembly for the formation of a filament was introduced first in the helical model

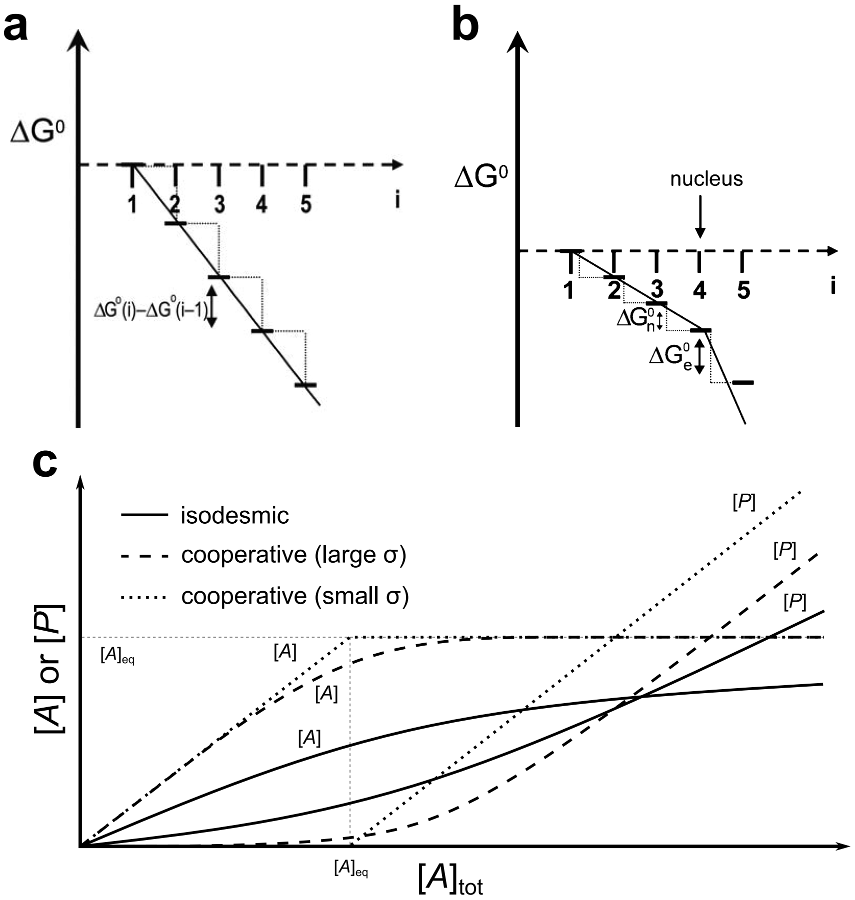

Figure 1.2: Comparison of isodesmic and cooperative model to describe supramolecular polymer stability. **a**) Energy diagram representing the equivalence of the interaction energy between a monomer and all the other polymeric species in an isodesmic system. **b**) Energy diagram representing the difference of interaction energy between a monomer and a polymeric specie below and above the nucleus size (4) in a cooperative system. **c**) Concentration of monomeric ($[A]$) and polymeric ($[P] = \sum_{j=2}^{i} j[A]_j$) species as a function of the total concentraton in an isodesmic and in two cooperative systems with different $\sigma = K_n/K_e$. Panels a and b adapted from [1].

[61]. The formation of a nucleus is considered unfavourable compared to the elongation of an existing filament, and this is reflected by the fact that elongated structures compose the biggest part of filaments solution. The helical model formalizes this by introducing a nucleation equilibrium constant, which is related to the equilibrium constant of elongation by the term $\sigma = \gamma(K_n/K_e)^2$. σ is much smaller than unity and allows us to describe the following dependence of the equilibrium once again on the soluble monomer concentration:

$$[A]_{tot} = [A] + \frac{\sigma[A]}{1-K_e[A]}^{2} \tag{1.12}$$

Also the helical model converges to $[A] = K_e^{-1}$ for total concentrations much bigger than the critical concentration. These works set the ground for the study of the energetics of interaction of supramolecular one-directional assemblies, leading to the characterization of more protein systems and of the dependence of them on the cellular energetics [20, 62, 63].

Polymers of small aromatic molecules follow similar patterns of polymerization, but more models were developed and tested to depict the wide variety of behaviours observed in this class of molecules [1, 64, 65]. The synthetic nature of these compounds makes them easier to modify compared to the bigger, more flexible and delicate proteins. This aspect allowed studying the aromatic polymers stability through the use of temperature changes and different solvent. Moreover, they exhibit tunable polymerization behaviour upon small modifications of the monomeric structure or upon the use of different solvents [66, 67].

A wider interpretation of the helical model has been found to be useful to describe the polymerization behaviour of different aromatic polymers. In the helical model of Oosawa the nucleus size is fixed and two conformation of the polymer are possible and in equilibrium. Goldstein and Stryer introduced the so called cooperative model, where the polymer can only adopt one conformation, but the presence of a nucleus of variable size n can be stated as follows (Fig. 1.2b):

$$\begin{aligned} K_n[A]^2 &= [A]_2 \\ K_n[A][A]_2 &= [A]_3 \\ K_n[A][A]_3 &= [A]_4 \\ &\dots \\ K_n[A][A]_{n-1} &= [A]_n \\ K_e[A][A]_n &= [A]_{n+1} \\ &\dots \\ K_e[A][A]_i &= [A]_{i+1} \end{aligned} \tag{1.13}$$

where K_n is the equilibrium constant that regulates the interactions among monomer and species smaller than a nucleus. K_n is related to the elongation rate constant by the term $\sigma = K_n/K_e$. Once more, the equilibrium constant converges to the inverse of the free monomer concentration for high total concentrations.

The difference between isodesmic and cooperative model is evident by looking at how they behave around the critical concentration (Fig. 1.2c). The steepness of the curve representing the population of filaments is much higher for the cooperative model, where due to the unfavourable interactions for the formation of nuclei almost no polymers are present below the critical concentration. Above this concentration, the monomeric population will not rise any longer, as the interactions between a monomer and a polymer of size above the nucleus is favoured. An isodesmic system instead will start to gradually form polymers below the critical concentration and the soluble monomer will slowly approach the critical concentration at total concentration much higher than the critical concentration itself.

Regardless of the system used, the determination of the critical concentration is the key to establish the stability of the supramolecular assembly. This concentration for stable polymeric systems is difficult to probe, as it tends to be below the detection levels of usual instrumentation. Besides, obtaining a dataset able to show the differences depicted in Fig. 1.2c is difficult, as a technique with a wide range of sensitivity is required.

Tuning the stability of these system proved to be best way to measure it reliably and to model the polymerization mechanism. The chemical stability of the aromatic compounds allows measuring them in conditions not available for proteins. Using high temperatures does not damage the molecular structure nor cause disordered aggregation. At the same time, the temperature destabilizes the assemblies, leading to solutions composed by monomers only. By cooling down the system at equilibrium, it is possible to distinguish between cooperative and isodesmic polymerization, all while maintaining the same total concentration of molecules [24]. Destabilizing the system is also possible by using different solvents that favour either the monomeric or the polymeric form [25, 67].

So far, the critical concentration of amyloid systems has been the object of a limited number of studies, mainly focused towards the protein Aβ [68–71]. Due to the high stability of amyloid fibrils, the critical concentration is difficult to access. The ease of degradation of the polypeptide chain renders the prolonged use of high temperatures problematic, so the limited amount of information obtained were modelled through indirect steady-state kinetic measurements based on the Eq. 1.5 (and how it is sketched in Fig. 1.1a). Similarly, the slow nucleation kinetics, which are also the biggest difference between amyloid fibrils and the other two systems here described, do not allow us to exploit the full polymerization kinetics like with actin and the other functional systems (although one tentative has been reported [72]). Amyloid stability has been modelled so far just in the framework of the isodesmic model, using chemical denaturation to access the critical concentration [73–75]. This approach, similar to the one proposed for aromatic compounds by Korevaar *et al.*, is inspired from the protein folding field. The use of a chemical denaturant shifts the reaction equilibrium, allowing experimental access to the species otherwise not measurable [76]. This can be either the unfolded monomer, for protein folding reactions, or the soluble monomer concentration, for supramolecular assemblies.

Chemical denaturation has been established since more than 50 years as a methodology to measure protein folding stability. Its introduction in the field of amyloid fibrils is more recent and adjustments still have to be made. The work shown in this thesis improves on the current framework and sets the necessary controls to be made in order to extrapolate meaningful stability parameters. In the next section a historical excursion of the technique is presented together with the current applications in the field of supramolecular assemblies.

1.3. Chemical denaturation as a tool to measure conformational stability

Denaturation upon specific changes of solution conditions was already found to be a generic property of proteins in the 1930s, when it was described as a process that results in the loss of various specific properties, like the ability to form crystals, immunological recognition, spectroscopical properties and function [79]. Through the use of chemicals or temperature it was possible to change, in some cases irreversibly, the so called "configuration" of the polypeptide. Knowledge about the molecular meaning of the "configuration" was still lacking, and it took more than 30 years to describe the loss of "configuration" as the unfolding of the native network of interaction of a protein in a two-state folding reaction [80]. Accessing the unfolded state through denaturation allowed measuring the unfolded protein population and, like in the case of the critical concentration in a polymeric systems, this means that a definition of stability became possible. Aiming in this direction, further experiments linked the denaturating action of additives like urea or guanidine chloride with the effective stability of the native state, introducing several models to extrapolate the stability parameters in the absence of denaturant [81, 82]. One of the first observations that was possible through the use of chemical denaturant was that the folded,

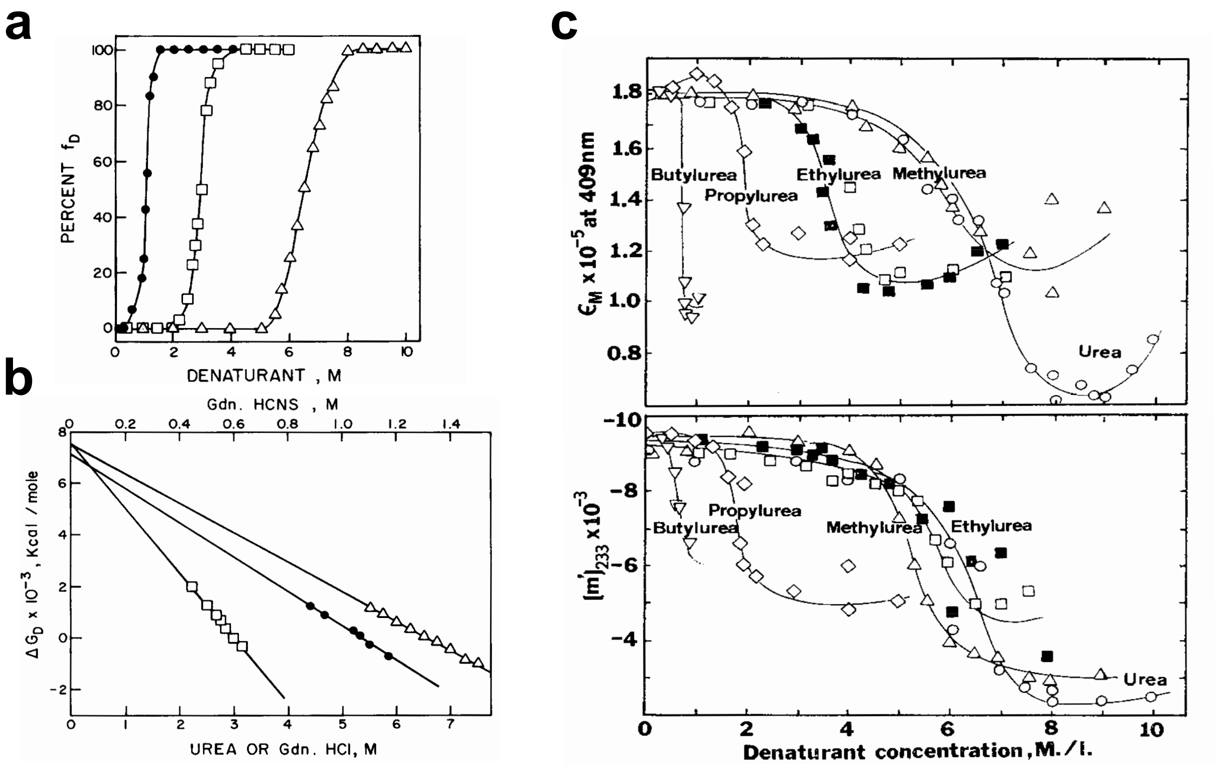

Figure 1.3: Denaturation curves of RNAse A (a,b) and sperm whale myoglobin (c). Different denaturants are able to denature the proteins with different efficacy, from the weakest urea to the strongest guanidine tiocyanate. **a**) Denaturation curves of RNAse A in presence of urea (empty triangles), guanidine chloride (empty squares) and guanidine tiocyanate (filled circles). **b**) ΔG_D for the transition phase of the denaturation curve are shown together with the linear extrapolation of the stability value in absence of denaturant. **c**) Denaturation of sperm whale myoglobin by various urea derivatives. The upper figure represent the change of molar extinction coefficient at 409 nm, while the lower one shows the mean residue rotation at 233 nm. Panels a and b adapted from [77] and c from [78].

functional state of a protein is only slightly more stable than the unfolded one, shedding light on the balance between enthalpic and entropic factors as driving forces towards the folded state [83].

The main model used to measure the stability of the folded system is based on the linear relation between stability and denaturant concentration. The basis for a linear dependency were established empirically through a series of comparisons between different proteins and different denaturants [77, 84]. The linear relation between the stability $\Delta G_{[D]}$ and the denaturant concentration $[D]$ is identified in the so called m-value:

$$\Delta G_{[D]} = \Delta G^0 + m[D] \tag{1.14}$$

where ΔG^0 the stability extrapolated in absence of denaturant. The introduction of the linear relation between denaturant and stability of the folded state allows avoiding to explicitly declare the mechanism behind denaturant induced unfolding. Further models rely on a binding constant between the protein and the denaturant, but still today evidence is missing in this direction [76].

Direct measurement of the unfolded state is possible when its population is of the same order of magnitude as that of the native state. To achieve this, different proteins require different amounts of denaturant, hence more stable proteins need more denaturant. At the same time different denaturants have different denaturing powers, leading to a measurable population of unfolded proteins at lower or higher concentrations, as it can be seen in Fig. 1.3 [77, 78]. The less denaturant is needed to unfold a protein, the higher the accuracy of the stability measure, as the extrapolation back to the absence of denaturant becomes less error-prone. Hence, stronger denaturants are preferred in protein folding, with guanidine chloride leading the charts as it is the best compromise between strength of denaturation and experimental usability [77, 85].

Chemical denaturants are also defined as chaotropic agents, molecules that alter the ordered pattern of the water structure in solution [86]. Guanidine salts are as well suggested in the Hofmeister series as the strongest solubilizing agent. This property made guanidine salts good candidates to study the dissolution of prion amyloid fibrils, a specific kind of highly stable amyloid polymers which causes different degenerative pathologies, such as the Creutzfeld-Jakob disease and BSE [87, 88]. Denaturation became an useful tool to distinguish different strains of this class of amyloid fibrils: the qualitative analysis of the denaturation midpoint of these fibrils allows us to differentiate the biological properties of the strain [89, 90].

The use of denaturants in the amyloid field was introduced in the previous section. The limited amount of literature in the topic used the isodesmic model to characterize the stability of the fibrils, even though it is known that it does not depict the complexity of amyloid polymerization [73–75]. It must be noted that the use of denaturants to describe the system stability can work only if the system behaves as a two state system at every concentration of denaturant, hence it must be able to acquire both states. This requirement is met by the vast majority of amyloid fibrillar systems, as they involve either intrinsically disordered proteins (IDP) or denatured states of structured proteins [29]. On the contrary, functional protein filaments like actin, tubulin and flagellin form as the result of the assembly of the native structures of monomers. The denaturation of these filamentous systems will not only result in the shift of the equilibrium between the monomeric and polymeric states, but will also coincide with the unfolding of the monomers, creating a three state system [14, 91].

While some amyloid systems involve polypeptide chains that are not structured and can form fibrils under physiological conditions, like α-synuclein, Aβ and IAPP, others need to be destabilized and acquire at least a partially unfolded conformation, like lysozyme, β2-microglobulin, β-lactoglobulin, PI3K-SH3 and glucagon [29]. The fibrillation of this second class of protein is usually induced by the use of denaturants, drastic pH shifts or temperature change [39, 92–95]. Denaturants play a a destabilizing action by denaturing the native structure of the protein, favouring kinetically the formation of amyloid fibrils. Guanidine salts exert a further stabilizing action by screening the charges of the different monomers in the filament [96]. A more detailed overview of the topic is later introduced in the first part of this work.

In order to characterize the stability of amyloid systems it is very important to ensure that no third state such as e.g. oligomers with a structure different from fibrils, is formed with the use of denaturant, i.e. to probe the equilibrium of the polymerization reaction only. In this work the stability of two amyloid systems is extensively characterized, PI3K-SH3 and glucagon, both forming amyloid fibrils in low pH. In the next sections the two proteins are described in term of the state-of-the-art knowledge of both their monomeric and polymeric states, with emphasis on the structure characterizing the soluble state.

1.3.1 PI3K-SH3

Phosphoinositide 3-kinases are a class of proteins involved in cellular signalling, with a central role in the transduction and amplification of external stimuli related to cell growth [97]. This central role makes this class of kinases a key component of the study of cancer propagation, to the point that elevated levels of its own signalling cascade are considered a hallmark of cancer. The protein itself is composed of different domains which mediate the interactions with other cellular components. The different domains can be split among two different macro-regions: p85-p55-p50, the regulatory unit, and p110, the catalytic unit [98]. The regulatory unit p85, which comes in the isoforms α and β, contains the adaptor domain SH3, which as many other homologue domains mediates the protein-protein interactions of various nature through the recognition of the P-X-X-P motif [99]. PI3K-SH3, a 84 residues long protein, is the major subject of this thesis and apart for its biological relevance and structure [100–102], it has been also used as a model protein to study protein folding [103, 104].

The finding of amyloid fibrillar aggregates formed by this protein without connection with amyloid-related diseases paved the way for the hypothesis of amyloid as the most stable conformational state of a protein [30, 105]. A detailed overview of the role of PI3K-SH3 in the establishment of some basic concept regarding the amyloid fibril formation are outlined in the second part of this thesis. Here I will discuss instead the unstructured monomer conformation in the acidic pH that has been used throughout the whole work.

The fibrils do not form under neutral pH, but are the result of incubation in $pH < 3$ [39]. The different behaviour between physiological and acidic pH resulted in the characterization of the monomeric conformation, which may have hinted towards the determinants of amyloidogenicity. Surprisingly, the monomeric pH 2 state turned out to present all the signatures of the denatured state, with the presence of possible hydrophobic clusters highlighted by ANS fluorescence [105]. The native fold is indeed lost upon decrease of pH, with a sharp loss of the conformation below the isoelectric point (4.7) which leads to an unstructured but yet compact state [39]. The compactness of the state, which seemed to increase upon acidification below pH 2.5, was not related to the acidification itself, but to the screening of the high monomeric charge (+12) derived from increased salt concentrations. No link between residual structure and compactness was determined, as the kinetics of pepsin digestion suggested equal accessibility to proteolysis of most extended states [106]. Moreover, very similar solution and fibrillation behaviour was recorded for a circularized version of PI3K-SH3, suggesting the lack of a specific structural preference in acidic environment [107]. Once the absence of a stable structure was established, more in depth studies were directed toward the characterization of a possible ensemble of rapidly inter-converting conformations [108, 109]. Long range weak interactions are still found in the acid conformation, but no remaining native structure can be observed [108]. At the same time, experiments involving point mutations on the sequence of PI3K-SH3 demonstrated the absence of specific interactions, hinting instead towards the presence of hydrophobic clusters that justify the binding and consequent fluorescence of the ANS dye [109].

1.3.2 Glucagon

Glucagon is a 29 residue peptide secreted by the α pancreatic cells in response to low glucose level in the blood [110]. The storage of the peptide seems to happen in multimeric form, possibly as a trimeric assembly resembling insulin [111] or even in fibrillar aggregate with cross-β conformation [27]. In the trimeric assembly, glucagon seems to adopt a predominant α-helix conformation, but when monomeric and soluble in solution it is largely unstructured with a propensity to form α-helix only at low temperatures or in the presence of apolar solvents and lipids [112, 113]. Being a peptide of great pharmaceutical interest, glucagon's biophysical properties are well characterized [95]. The isoelectric point is close to the physiological pH making it not very soluble unless in acidic or alkaline solutions [114]. At the same time, long storage in acidic or alkaline conditions lead to fibrillation of the peptide in β-sheet-rich aggregates [112].

The formation of amyloid structures by glucagon has been studied extensively, motivated in past by the high polymorphism of the fibrils [95]. Even though the protein is composed of only 29 residues, more than five different amyloid strains were characterized by spectroscopy [45, 115] and the atomic structure was solved for an additional one [116]. All the different strains exhibited different spectroscopical properties, highlighting a range of distinct molecular packings that translate into different fibrillar structure. The wide structural polymorphism leads to different stabilities, but at the same time some common properties are shared by all the strains. There is a clear exothermic signature of the stability as all fibrils depolymerize upon heating [45]. Similarly, the thermal stability of the strains seems to be increased in presence of salts that, by screening the charges of the monomers, decreases the inter-layer repulsion [117].

Glucagon fibrillation was also extensively studied by exploiting the intrinsic fluorescence of the single tryptophane residue present in its sequence [45]. Regardless of the fibril strain, the tryptophane residue is always buried inside the fibrillar core, causing a blue shift of the emission peak. The magnitude of the blue shift depends on the localization of the tryptophane, which together with circular dichroism spectroscopy allows us to distinguish the various fibrillar strains.

1.4. Aim of this work

The mechanisms of amyloid formation and growth have been widely studied in the last decades, increasing the knowledge via always more detailed kinetic studies of the reactions involved. Still, the field is lacking solid modelling of its thermodynamic aspects. Moreover, the misfolding step, in which the monomer shifts from its soluble conformation to the fibrillar one, has still to be understood at a single amino-acid residue level of detail [31]. To fill this gap it is necessary to obtain information about the transition state of the reaction. In this work I set up the framework to study the transition state by applying the Φ-value analysis to the elongation reaction. The thesis is divided in four chapters: the first three describe how to fulfil the fundamental requirements needed for the fourth chapter, the first Φ-value analysis of the PI3K-SH3 amyloid fibril elongation, performed on a preliminary set of mutants.

Establishing the Φ-value of the different residues requires to know the energetic levels of the states involved in the reaction, namely the soluble state, the transition state and the amyloid state. The first two part of this thesis describe the methodologies developed to measure the stability, both through chemical denaturation and temperature changes. In Chapter 2, a new step in the description of the thermodynamic stability of the amyloid fold is made. The cooperative model is for the first time introduced into the amyloid field and used to explain dataset not compatible with the isodesmic model. In Chapter 3 two frameworks are presented and tested to extract the stability over a wide range of temperatures, using both kinetic and out-of-equilibrium models.

To define the role of a residue in the transition state, it is important to know what is the role in the structured state. In this context, knowing the structure of the amyloid state is necessary. The structural characterization of PI3K-SH3 amyloid fibrils by cryogenic electron microscopy is presented in Chapter 4. The structure is supported by kinetic data highlighting the contribution of different residues in forming the amyloid network of interactions.

Finally, Chapter 5 exploits the knowledge obtained in the first two to describe the role of five residues in the transition state of the PI3K-SH3 amyloid elongation reaction. Five mutants involving residues in different portions of the sequence were studied in both their kinetic and thermodynamic behaviour to extract the corresponding Φ-values. This work creates the basis for a rigorous analysis of the transition state and lays down the basis for further analysis on the same and other amyloid systems.

2

THERMODYNAMICS OF AMYLOID FIBRIL FORMATION FROM CHEMICAL DEPOLYMERIZATION

Article information

Title: Thermodynamics of amyloid fibril formation from chemical depolymerization

Authors: Nicola Vettore[1] and Alexander K. Buell[1,2]

1) Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225, Düsseldorf, Germany

2) Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, 2800 Kgs, Lyngby, Denmark

Journal: Physical Chemistry Chemical Physics

DOI: 10.1039/C9CP04524D

Abstract

Amyloid fibrils are homo-molecular protein polymers that play an important role in disease and biological function. While much is known about their kinetics and mechanisms of formation, the origin and magnitude of their thermodynamic stability has received significantly less attention. This is despite the fact that the thermodynamic stability of amyloid fibrils is an important determinant of their lifetimes and processing *in vivo*. Here we use depolymerization by chemical denaturants of amyloid fibrils of two different proteins (PI3K-SH3 and glucagon) at different concentrations and show that the previously applied linear polymerization model is an oversimplification that does not capture the concentration dependence of chemical depolymerization of amyloid fibrils. We show that cooperative polymerization, which is compatible with the picture of amyloid formation as a nucleated polymerization process, is able to quantitatively describe the thermodynamic data. We use this combined experimental and conceptual framework in order to probe the ionic strength dependence of amyloid fibril stability. In combination with previously published data on the ionic strength dependence of amyloid fibril growth kinetics, our results provide strong evidence for the product-like nature of the transition state of amyloid fibril growth.

2.1. Introduction

Filamentous protein structures are ubiquitous in biology. They can fulfill functional roles, as in the case of the cytoskeletal proteins actin [118] and tubulin [119], or be associated with diseases, as in the case of sickle hemoglobin polymers [120] or amyloid fibrils [30]. For the cytoskeletal filaments, both the mechanical and thermodynamic stability, as well as the molecular origin of their reversibility have been the subject of extensive studies in the last decades [17]. Indeed, actin filaments were the first protein polymers the thermodynamic stability of which has been characterized in detail in seminal work by Oosawa [61]. Aktin filaments can be described as helical polymers, whereby each monomeric building block interacts not only with its nearest neighbors in the polymer, but also with building blocks further away. Actin monomers undergo relatively minor structural changes upon polymerization [121], and the same is also true for tubulin and sickle hemoglobin [122]. Amyloid fibrils, on the other hand, are protein polymers in which the protein building blocks usually adopt a very different structure inside the fibril compared to the isolated protein molecule [123]. Amyloid fibrils are mostly known as being the hallmark of a wide range of diseases, but mounting evidence demonstrates that amyloid fibrils can also play functional roles in biology [124]. Many different proteins entirely unrelated in structure (IDPs or folded proteins) and function (peptide hormones [27], lipid-binding proteins [125], milk proteins [126]) have been found to form amyloid fibrils, either *in vivo* or in the test tube. While much insight has been generated in the last two decades on the kinetics and mechanisms of amyloid fibril formation [127–130], much less is known about the thermodynamic stability of these structures [131], and to what extent common driving forces govern the amyloid formation of different proteins. It is very important to be able to quantify and rationalize the thermodynamic stability of amyloid fibrils because the thermodynamic stability is likely to be a decisive factor in determining whether amyloid fibrils of a certain protein can be cleared in a biological context. Such clearance could happen either through spontaneous binding of molecules, such as antibodies, to soluble or aggregated protein [132] or else through active, energy consuming processes, such as the action of chaperone complexes on amyloid fibrils [133].

Experimental data of amyloid fibril stability [73, 74] has so far been analysed in the framework of the linear polymerization model [61]. The simplest, so-called isodesmic form of this model contains only a single equilibrium constant, that for the addition of monomer to all possible species, including to another monomer. Under this assumption, the stability of the fibrils can be directly determined from a measurement of the free monomer concentration at equilibrium, for sufficiently high total concentrations. In many practical cases, the free monomer concentration is very low (nM range [134]) and difficult to determine accurately. In order to overcome this practical problem, amyloid fibrils can be destabilized by chaotropes, such as GdnSCN or GdnHCl [73, 74]. By assuming a linear relationship between the concentration of denaturant and the free energy difference ΔG between the monomeric and polymeric state, equivalent to the assumption in protein unfolding experiments [135], sigmoidal fibril depolymerization data can be fitted and the value of ΔG extrapolated to the absence of denaturant [73, 74]. While this method of determination of amyloid fibril stability is commonly used, its validity has not yet been thoroughly tested. Here we use intrinsic protein fluorescence to monitor the fraction of fibrillar vs. soluble protein. We investigate and rationalise the influence of the type of chemical denaturant, the ionic strength and the protein concentration on the depolymerization curves and show that a cooperative model provides a better description of the thermodynamics of amyloid fibril formation than the simple isodesmic polymerization model. The cooperative model, adapted from the field of supramolecular chemistry [24, 65], describes the thermodynamics of polymerization by two different equilibrium constants: a nucleation constant and an elongation constant. We find that experiments in which the peptide concentration is varied allow a better discrimination between the isodesmic and cooperative models than the standard experiments in which only the denaturant concentration is used.

We apply our insight and methodology to the analysis of the ionic strength dependence of amyloid fibril stability and compare it with the ionic strength dependence of the fibril growth kinetics [37]. Our combined analysis of this data provides strong evidence for the fact that the transition state of the fibril growth reaction is highly product-like with respect to the distance of the newly adding peptide to the fibril end.

2.2. Results

In order to establish a reliable methodology for the determination of the thermodynamic stability of amyloid fibrils, we chose two polypeptide systems that have been shown to form amyloid fibrils under acidic conditions, the peptide hormone glucagon [95, 136] and the protein PI3K-SH3 [39, 105]. Both of these polypeptides contain the amino acid tryptophan and therefore the use of intrinsic protein fluorescence can be explored for the distinction between monomeric and aggregated protein [45] (Suppl. Fig. 2.8). In several previous studies of amyloid fibril stability based on chemical depolymerization, the equilibrated samples were subjected to centrifugation and the protein concentration in the supernatant was determined [74, 75]. Experimental data confirming the equivalence between the data from concentration measurements and the intrinsic fluorescence data can be found in Supplementary Figures 2.12 and 2.13.

For both PI3K-SH3 and glucagon, homogeneous fibril preparations without appreciable amounts of non-fibrillar material (as judged by AFM imaging, see Supplementary Figure 2.7) can be produced by adjusting the solution conditions appropriately [45, 137]. In our experiments of fibril depolymerization we started from fully equilibrated fibrillar samples that we prepared by seeding, i.e. the addition of preformed fibrils to monomeric samples. We then homogenized the samples and shortened the average fibril length by subjecting the samples to ultrasonication (Supplementary Figure 2.7). Details about the sample preparation can be found in the methods section.

2.2.1 Choice of denaturant

Having chosen the polypeptide systems to investigate, we next proceeded to the choice of the chemical denaturant to be used to destabilise the fibrils. Previous reports mostly used strong denaturants, such as GdnHCl or GdnSCN [73–75] rather than milder ones, such as urea. It is well established that the ionic denaturants are more powerful than urea in the unfolding of proteins [138]. However, we found that in our case the fibrils could be dissociated by urea, and even that in the case of PI3K-SH3 amyloid fibrils, urea was a more powerful denaturant than GdnHCl (see Supplementary Figure 2.10). This is in contrast to the unfolding of monomeric PI3K-SH3 that was unfolded at lower concentrations of GdnHCl compared to urea, at neutral pH (see Supplementary Figure 2.10). Based on these initial results, we decided to chose urea as the denaturant of choice for our amyloid fibril depolymerization experiments, as this choice of a neutral denaturant allows us to explore the role of electrostatic interactions in amyloid fibril stability in more detail. The use of GdnHCl would lead to an almost complete screening of the electrostatic interactions at the high denaturant concentrations used in these experiments. It has to be noted that prolonged incubation of proteins in high urea concentratiosn can lead to carbamylation reactions [139]. The acidic pH conditions employed in the present study strongly disfavour this reaction. We showed by mass spectrometric analysis (Supplementary Figure 2.15) that even after 9 days incubation in 6 M urea, no sign of carbamylation was observed for PI3K-SH3, which required the longest equilibration times.

2.2.2 Analysis of chemical depolymerization with an isodesmic model

Equilibrium denaturation curves of folded proteins are often analysed with a two state model, whereby it is assumed that the denaturant linearly shifts (with proportionality constant m) the free energy difference between the folded and the unfolded polypeptide [56]. Due to the large number of aggregate species of different sizes, a two state model is not appropriate in the case of protein polymerization. It has first been proposed by Goto and coworkers [73] to apply the isodesmic form of the linear polymerization model [61] to fit equilibrium depolymerization curves of amyloid fibrils, and this method of analysis has been exclusively used to-date. The mathematical formulation of this model can be found in Supplementary Section 2. Individual sigmoidal depolymerization curves can be fitted and free energy differences between the fibrillar and the soluble states, as well as m-values, can be determined. The absolute concentration of peptide is an important parameter in the model that determines the shape of the depolymerization curve (see Supplementary Section 2). If depolymerization curves are acquired at different peptide concentrations, the extracted free energies and m-values are expected to be identical, within error. We have fitted data from experiments at three (glucagon) and two (PI3K-SH3) different peptide concentrations. It is convenient to display such data in a normalised way, i.e. as a fraction of depolymerised protein, rather than as an absolute concentration (Figure 2.1). In Supplementary Figure 2.14, we show the same data without normalisation. We find that while all the individual curves can be well-fitted (Figure 2.1, left panel), in the case of glucagon the fits at the three different concentrations yield significantly different values for the free energy difference ΔG_0 (-36.7 vs. -38.8 vs. -42.2 kJ/mol for the lowest, intermediate and highest concentrations, respectively). In the case of PI3K-SH3, the difference is similar (-64.7 vs. -71.1 kJ/mol), despite the fact that here the difference in concentrations is less significant. The reason for the difference between the individual fits is that both ΔG_0 and m are free parameters and can both vary between fits to the data sets at different protein concentrations. If, however, the data at the different concentrations are fitted globally (Figure 2.1, right panel), we find that in the case of PI3K-SH3, a satisfactory fit result is achieved and the value for the free energy of fibril stability is intermediate between the two values of the individual fits (-66.5 kJ/mol). On the other hand, a global fit to the data for glucagon is not as good and yields a value for the free energy comparable to that from a local fit of the lowest concentration (-36.7 kJ/mol). Tables 1 and 2 in the Supplementary Materials show all the free energy and m-values, when neither, either of these or both parameters are globally fitted. Therefore, already a global fit to denaturation curves at only two different protein concentrations reveals potential inadequacies of the isodesmic model for the description of chemical depolymerization experiments of amyloid fibrils.

2.2.3 Extension of the isodesmic model

The isodesmic version of the linear polymerization model assumes that all equilibria in a solution of polymers have the same equilibrium constant. However, this assumption is clearly not in agreement with the known mechanistic features of amyloid fibril formation. Amyloid fibrils form through a nucleated polymerization process [127, 129], whereby the formation of the initial oligomeric nucleus is energetically less favourable than the addition of a monomer to

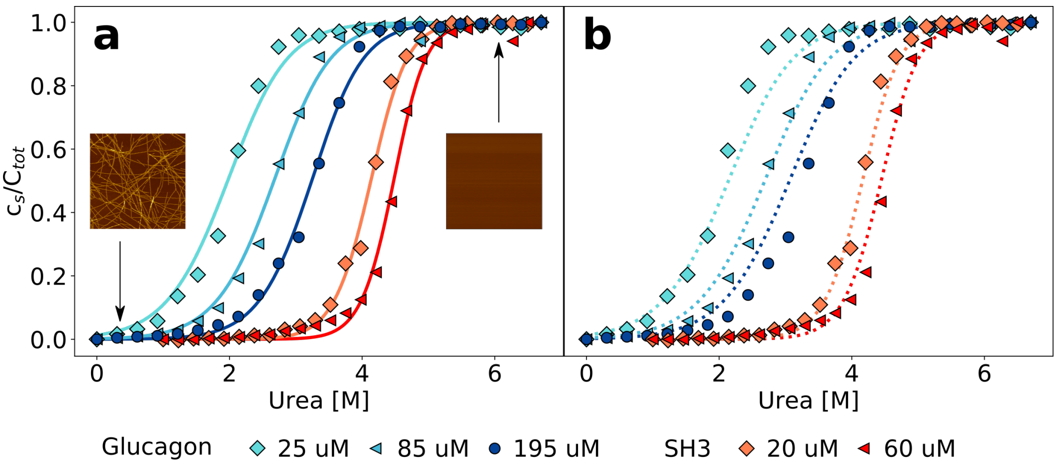

Figure 2.1: Equilibrium depolymerization profiles of glucagon and PI3K-SH3 fibrils at two different peptide concentrations. **a**) The data for each peptide concentration have been fitted individually (continuous lines). The insets are AFM images showing the sample at low denaturant and at high denaturant concentrations. The image size is 5x5 µm. **b**) Global fits to both concentrations simultaneously for each peptide (dotted lines).

a fully grown fibril [140]. We hypothesised that this simplification could be at the origin of the inability of the isodesmic model to quantitatively describe the concentration dependence of chemical depolymerization curves of amyloid fibrils (Figure 2.1). We therefore increased the complexity of the model by allowing a different equilibrium constant for monomer association to any aggregate below a certain aggregation number n, i.e. a distinct equilibrium constant of nucleation. The nucleation process is thus defined as a series of less favourable interactions between a monomer and any species up to an n-mer, where n defines the size of the nucleus. By allowing for two distinct equilibrium constants with a ratio $\sigma = \frac{k_n}{k_e}$, it is possible to account for the fact that the formation of a pre-fibrillar structure is thermodynamically less favourable than its growth.

This cooperative model, which is slightly more general than the helical polymerization model of Oosawa [17] (see Supplementary Section 2 for details on possible extensions of the isodesmic polymerisation model) has been successfully used to describe the aggregation process of supramolecular non-covalent polymers [65]. It has for example allowed to explain the differential effects of a gradual change in solvent conditions on the stability of several supramolecular polymer systems [25]. The building blocks of supramolecular polymers are usually simpler molecules than polypeptides, with much fewer degrees of freedom and less potential for polymorphism, allowing the systems in some cases to be characterised very accurately and in great detail [141]. Our concentration-dependent measurements of amyloid fibril depolymerization allow us for the first time to test the applicability of this type of model also in the case of amyloid fibrils. Figure 2.2 shows a comparison of fits of the same data sets as in Figure 2.1 to both the isodesmic and cooperative models. Compared to the isodesmic model, the cooperative model has two additional free parameters, σ and n. We performed fits for different fixed values of n, ranging from 1 to 50. The resulting fit parameters are recorded in tables 3 (PI3K-SH3) and 4 (glucagon) in the Supplementary Materials. σ and n are not independent of each other, the higher the fixed value of n, the closer σ is to unity. We also compared the model, whereby all species other than the monomer have the same spectroscopic signature (oligomer same as aggregate, 'osaa'), to the model whereby all species up to size n have the same spectroscopic signature as the monomer (oligomer same as monomer, 'osam'). Given the intrinsically disordered nature of

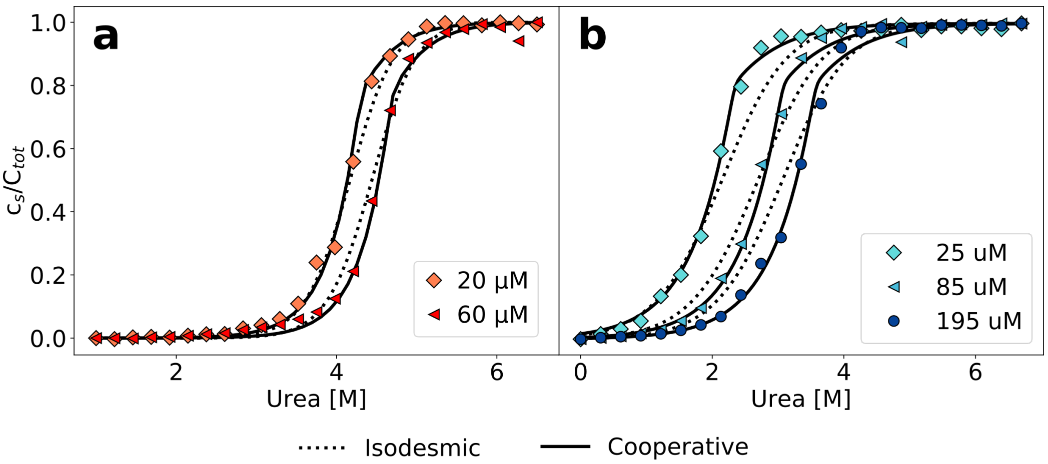

Figure 2.2: Cooperative polymerisation describes the concentration dependence of chemical depolymerization profiles of PI3K-SH3 (**a**) and glucagon (**b**) amyloid fibrils. The black lines show the best global fits of the two different linear polymerization models, corresponding to n = 4 in the case of the cooperative model for both proteins (see Supplementary Materials for a more detailed discussion of the effect of changes in n).

both PI3K-SH3 [39] and glucagon [142] under these solution conditions, we think it more likely that the oligomers will display a signature of intrinsic fluorescence closer to that of the fibrillar aggregate than to that of the monomer. Furthermore, the osaa model yields more consistent values for both ΔG_0 and σ . The best fits are achieved for small to intermediate values of σ (2-5), whereby the difference between the fit to the isodesmic model and the best fit to a cooperative model is more significant in the case of glucagon compared to PI3K-SH3 (Figure 2.2). The physical significance of n and σ are discussed in more detail in the Supplementary Materials.

2.2.4 Exploring the concentration dimension in chemical depolymerization

The finding that an increase in the concentration of urea allows to gradually depolymerize amyloid fibrils reflects that the denaturant shifts the equilibrium in a concentration dependent manner towards the soluble state. Therefore, while the critical concentration in the absence of denaturants can be very low and difficult to measure, at higher denaturant concentrations, it will eventually approach the total concentration of the sample. In order to explore this behaviour, we investigated the dependence of both glucagon and PI3K-SH3 amyloid fibril dissociation on the protein concentration at fixed concentrations of denaturant. A suitable denaturant concentration for each protein was chosen based on the data in Figures 2.1 and 2.2; we chose 3 M urea for glucagon and 4 M urea for PI3K-SH3. In Figure 2.3, we plot the concentration of soluble protein as a function of the total protein concentration. We performed these measurements by using the more conventional method of sample centrifugation, followed by measurement of the supernatant concentration. In Supplementary Figure 2.13, we show that for both proteins, the results from fluorescence and absorbance measurements are very similar, in particular at higher concentrations. We fit these data to both the isodesmic and the cooperative model and find that for both proteins, the cooperative model provides a significantly better fit than the isodesmic model. This is in particular also true for PI3K-SH3, for which both models gave very similar fits when two denaturant-dependent depolymerization curves were globally fitted at two different monomer concentrations (Figure 2.2, left panel).

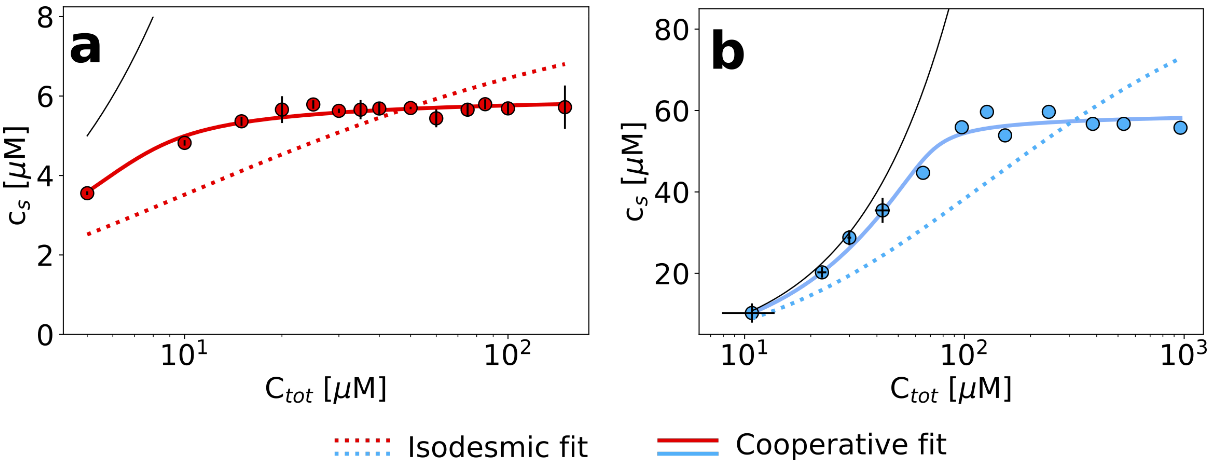

Figure 2.3: The dependence of the concentration of soluble protein on the total protein concentration at constant denaturant concentration. **a**) PI3K-SH3 in 4 M urea. **b**) glucagon in 3 M urea. The data was obtained from direct concentration measurements of the supernatant after centrifugation. The solid black line indicates the total protein concentration in both cases.

Therefore, the exploration of the protein concentration in addition to the denaturant concentration in amyloid fibril depolymerization experiments represents a powerful combination, allowing a more rigorous test and comparison of different models compared to an exploration of the denaturant concentration dimension alone.

2.2.5 Influence of ionic strength on amyloid fibril stability

Having established an experimental and conceptual framework in which to analyze the thermodynamic stability of amyloid fibrils quantitatively, we then proceeded to apply this methodology in order to probe the dependence of amyloid fibril stability on the ionic strength of the solution, by adding different concentrations of NaCl. We found that an increase in the concentration of NaCl stabilizes the amyloid fibrils of both PI3K-SH3 and glucagon (Figure 2.4 a and b), manifest through a shift of the depolymerization midpoint towards higher denaturant concentrations. This is expected, given the net positive charge of both proteins at the acidic pH of the study. The net charge carried by each monomer disfavors the polymerization process, and screening of the charges decreases this unfavorable contribution. We fitted the depolymerization curves by fixing the parameters of σ and n to the values determined in our previous best fits and plotted the logarithms of the equilibrium constants for elongation obtained from those fits against the square root of the total ionic strength of the solution (Figure 2.4 c and d, see Supplementary Section 3 for the mathematical analysis). We found that the stability of PI3K-SH3 amyloid fibrils depends more strongly on the solution ionic strength than the stability of glucagon fibrils, as evaluated from linear fits to these plots. This difference is consistent with the different (formal) net charge of PI3K-SH3 (+12) and glucagon (+5) at pH 2.

We have previously performed a detailed study of the dependence of the amyloid fibril elongation kinetics of PI3K-SH3 on the concentration of NaCl at a low background ionic strength of 10 mM [37] and we are now able to compare the ionic strength dependencies of both kinetics and thermodynamics of PI3K-SH3 amyloid fibril growth. The use of high urea concentrations in our chemical depolymerization experiments, however, requires a higher buffer concentration than that employed in the kinetic experiments and therefore the ionic strength ranges explored in the two cases overlap only slightly. However, by comparing the slope from the three highest ionic

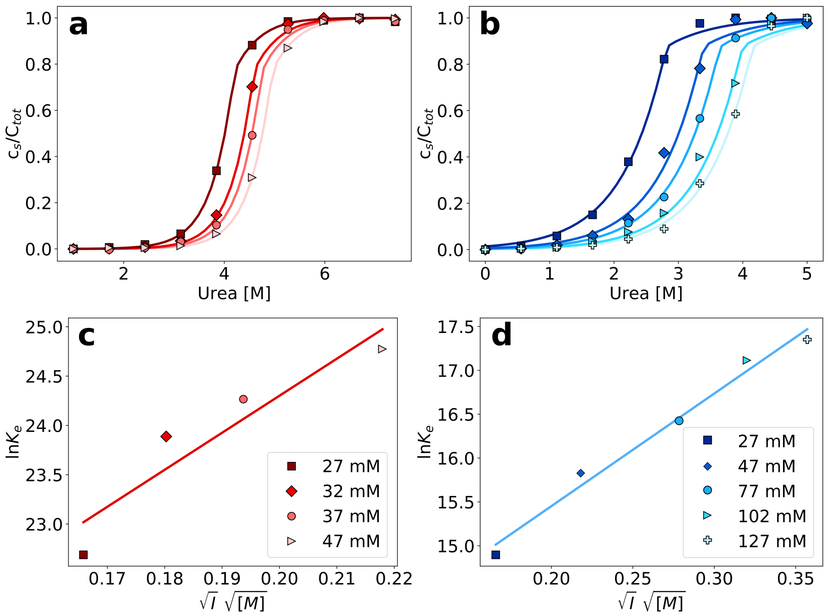

Figure 2.4: Dependence of amyloid stability on ionic strength. **a**) depolymerization profiles of PI3K-SH3 amyloid fibrils at different ionic strengths, adjusted by the addition of different NaCl concentrations. **b**) Denaturation profiles of glucagon amyloid fibrils at different ionic strengths. **c**) Plot of the logarithm of the equilibrium constants obtained from the fits in the top left panel, as a function of the square root of the ionic strengths for PI3K-SH3 (see Supplementary Section 3 for details). Parameters of the linear fit: slope = 37.5, y-intercept = 16.8. **d**) Plot of the logarithm of the equilibrium constants obtained from the fits in the top-right panel, as a function of the square root of the ionic strengths for glucagon. Parameters of the linear fit: slope = 12.84, y-intercept = 12.88.

strength values explored in the kinetic experiments and the three lowest ionic strength values explored in the thermodynamic experiments, it is possible to semi-quantitatively compare the ionic strength dependence of kinetics and thermodynamics (Figure 2.5 a). We find that the fitted slopes are very similar for the kinetic (43.5) and thermodynamic (57.2) plots.

In Figure 2.5 b), we illustrate with a schematic energy landscape of fibril elongation how screening of the electrostatic charges by salt ions in solution can lower the free energy barrier [37], as well as lower the energy of the fibril and hence stabilize it (see Supplementary Section 3 for details).

2.3. Discussion

2.3.1 Intrinsic fluorescence as a probe of amyloid fibril depolymerization

The determination of amyloid fibril stability from chemical depolymerization experiments requires a measurement of the relative populations of fibrillar and monomeric protein. This can be achieved through physical separation of aggregated from non-aggregated protein by centrifugation of the sample and measurement of the protein concentration in the supernatant [74, 75]. However, spectroscopic approaches, based on either circular dichroism [45, 73] or intrinsic protein fluorescence [45] have also been proposed, that do not require a physical separation of aggregates from monomeric protein. This feature is useful, as the separation of aggregated from soluble protein can depend on the size distribution of the aggregated species in conjunction with the applied centrifugal force. Fibrillar and monomeric protein can in most cases be expected to

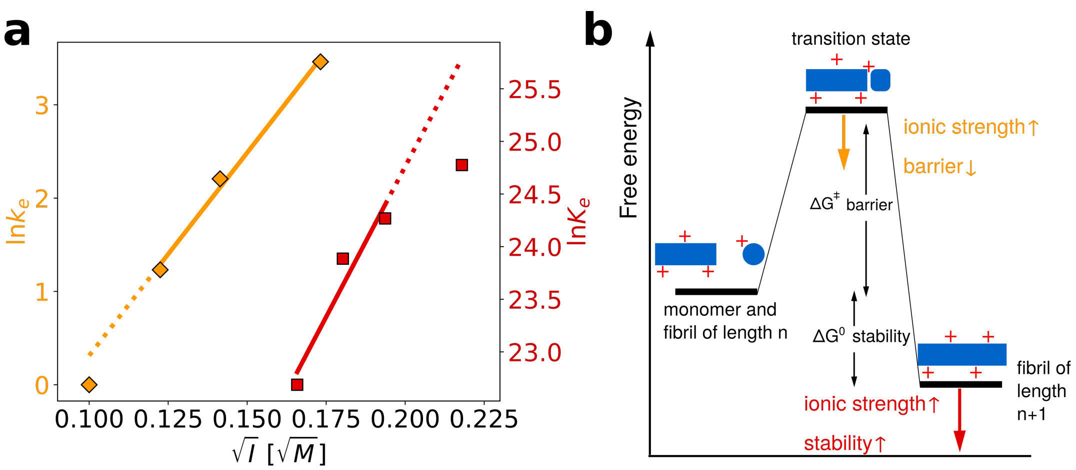

Figure 2.5: Comparison of the influence of salt on the kinetics and thermodynamics of the elongation of PI3K-SH3 amyloid fibrils. **a**) In orange, the data from [37], concerning the dependence of the elongation kinetics of PI3K-SH3 on the concentration of NaCl, while in red the thermodynamic data from this study. The linear fits have the following slopes: kinetics = 43.5; thermodynamics = 57.2 **b**) Schematic energy landscape illustrating the stabilizing effect that an increase in the ionic strength of the solution has on both the transition state and the final state of the elongation reaction.

have a different spectroscopic signature, due to differences in secondary structure and chemical environment of the aromatic amino acid residues. Therefore the change in relative populations can be monitored by analyzing the ratio of the fluorescence emission at two different wavelengths. We use this method here, based on Trp fluorescence, and show that it is equivalent to the direct concentration determination in the supernatant after centrifugation.

2.3.2 Nature of the denaturant

The equilibrium between the soluble and fibrillar conformation can be altered by the addition of chemical denaturant, similar to the denaturation of folded proteins. Interestingly, we find here that urea can be a more powerful denaturant for amyloid fibrils than GdnHCl, which is in contrast to the general finding that GdnHCl is a more powerful denaturant for protein unfolding than urea [138]. This inversion in the denaturant strength between monomeric proteins and amyloid fibrils can be explained through the differential importance of electrostatic effects. For the homo-molecular polymerization process of amyloid fibril formation, electrostatic interactions are generally unfavorable [143], and the addition of a denaturant salt, such as GdnHCl, will stabilise amyloid fibrils electrostatically at the same time as destabilizing them through competition for the hydrogen bonding network [144]. At comparably low denaturant concentrations, the stabilizing effect can be stronger than the destabilizing one, leading to a net stabilization. This stabilizing effect is virtually absent for non-ionic denaturants, such as urea, therefore explaining the depolymerization of some amyloid fibrils at lower urea than GdnHCl concentrations. It can be expected that this effect is particularly pronounced under solution conditions where the proteins carry a high net charge, such as the low pH values used here.

2.3.3 Appropriate model for the analysis of chemical depolymerization of amyloid fibrils

The theoretical framework for the description of the thermodynamics of linear polymers has been developed by Oosawa, initially for actin polymerization. There it was found that the system was best described by so-called helical polymers, whereby each monomer interacts with several other monomers in the fibril, not only with the two next neighbors, as expected for a perfectly one-dimensional polymer. This structural feature leads to the fact that the initial nucleus is difficult to form, as a monomer will not be able to form the same number of favorable interactions with the nucleus compared to when it adds onto a complete fibril. Amyloid fibril thermodynamics has so far exclusively been analyzed with the simplest form of the linear polymerization model [73–75], whereby a single equilibrium constant for monomer addition is postulated. As we show in this work, individual chemical depolymerization curves can be fitted very well with this simple model, but as soon as the total concentration of the protein is varied, it becomes obvious in some cases that this model cannot accurately describe the equilibrium behavior of amyloid fibrils. Inspired by the field of supramolecular polymerization, we find that a model that allows for a less favorable equilibrium constant of monomer attachment to species below a threshold size n provides a consistent theoretical description that allows to quantitatively account for the data. Interestingly, we find that the best fits are achieved for small values of n (approximately between 2 and 5) and for values of σ that are one to two orders of magnitude smaller than unity. We would like to stress here that this model is likely to still represent an oversimplification, given the real complexity of amyloid fibril formation, whereby in most cases different steps of monomer addition and structural rearrangement [145–147] lead to the formation of a minimal fibril. Nevertheless, it is probably the simplest extension to the basic linear polymerization model and the fact that our data can be well-fitted suggests that at least for the thermodynamic behaviour of amyloid fibrils, a description in terms of an effective equilibrium constant for the addition of monomers to small, pre-fibrillar species captures the essence of the process. It is interesting to note that the numerical values of the free energy of monomer addition and σ together suggest that also the addition of monomers to the smaller structures is favorable, albeit less so than the addition of monomer to a fully formed fibril. These results can be compared to the kinetics of amyloid fibril nucleation, which is generally found to be very slow and therefore the formation of the nucleus is viewed as being highly unfavorable. Indeed, the finding that in many amyloid systems, fibrils are found that consist of thousands of individual monomers suggests that the rate of nucleation of a fibril is at least 3-4 orders of magnitude slower than the rate of its growth. These observations can be reconciled with our present analysis by considering that for the kinetic behavior, the height of the free energy barriers, as well as the reaction order of nucleation and growth processes, have to be taken into account. A detailed analysis of these parameters in the case of the amyloid β peptide has recently revealed that the free energy barrier for primary nucleation is indeed several times higher than that for fibril elongation [51]. In our experiments, however, we probe the thermodynamic behavior of fibril formation. While the energy barrier for the formation of a dimer, trimer or tetramer might be much higher than for monomer addition to a fibril [51], our results suggest that once such a small oligomer is formed, it can be similarly stable, per monomer, as a fully grown fibril.

It is also interesting to note that the increasing number of available high resolution structures of amyloid fibrils [137, 148, 149] paints the consistent picture of a minimal fibril unit, consisting of between 2 and 8 monomers, which displays the full range of interactions of a fully grown fibril. It is plausible that structures smaller than this minimal unit are thermodynamically somewhat less stable than bigger structures, in contrast to classical nucleation theory, where structures smaller than the critical nucleus are considered unstable with respect to individual monomers.

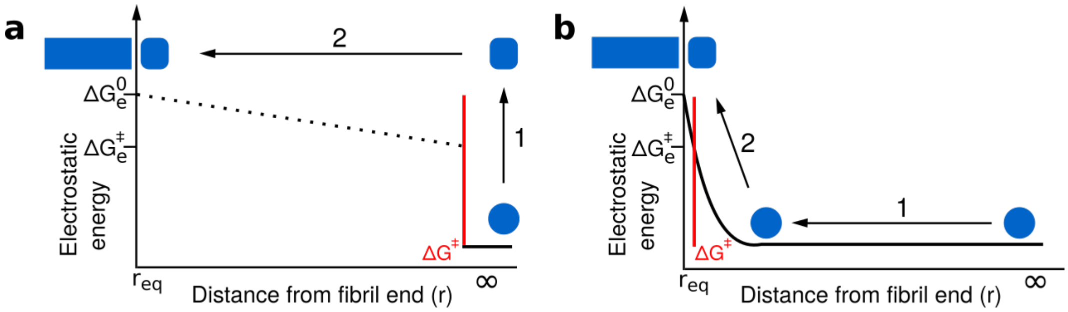

Figure 2.6: Comparison of two distinct scenarios of the transition state of amyloid fibril growth. The free energy barriers that define the kinetics of fibril growth are marked in red in both cases; however, we plot here only the electrostatic component of the free energy as a function of the center of mass separation between the approaching monomer and the fibril end. **a**) The transition state of fibril elongation corresponds to an isolated misfolding event of the free monomer (1), which is electrostatically unfavorable followed by a diffusive search of the fibril end (2). The electrostatic signature of the latter cannot be elucidated from kinetic experiments, because only the highest free energy barrier is probed [34, 150]. The final state (monomer incorporated) is electrostatically less favorable than the initial state and the transition state, but has the lowest free energy. **b**) The transition state of fibril elongation corresponds to a situation after a diffusional search (1), where the monomer is in close contact with the fibril end. The misfolding reaction is aided by the presence of the fibril end (2, templating effect). The transition state and the final state are similarly electrostatically unfavorable, because these two states have a very similar center-of-mass distance that dominates the electrostatic interactions. Therefore, we argue that our data on the ionic strength dependence of both kinetics and thermodynamics of fibril elongation is best compatible with a late, product-like transition state (scenario b).

2.3.4 Electrostatic effects to probe the nature of the transition state for amyloid fibril elongation

Our experimental and conceptual framework allows us to probe the effect of changes in solution ionic strength on the thermodynamic stability of amyloid fibrils. Here we study two proteins that form fibrils at acidic pH conditions, where the individual protein molecules carry a positive net charge that opposes homomolecular polymerization [37, 143, 151, 152]. The electrostatic repulsion of the individual monomers within the fibril can be expected to be screened by salt ions in solution, similar to what has been reported for the formation of surfactant micelles [153] or virus capsids [154]. The use of urea as a denaturant allowed us to probe this effect and we represent the resulting data by plotting the logarithm of the equilibrium constant against the square root of the total ionic strength (Figure 2.4 a). The slope of such a plot depends on the effective charge of the molecular interaction in question (see Supplementary Section 3 for a detailed discussion of the underlying model). In agreement with the higher formal net charge of PI3K-SH3 compared to glucagon at acidic pH, we find that the stability of PI3K-SH3 amyloid fibrils is more strongly affected by solution ionic strength than that of glucagon fibrils. It is insightful to compare the dependence of both PI3K-SH3 amyloid formation thermodynamics and kinetics [37] on NaCl concentration. For technical reasons (see above), the range of ionic strengths exploited in both studies overlaps only slightly, but if we extrapolate the respective slopes we are nevertheless able to compare kinetic and thermodynamic ionic strength dependencies. We find that the slopes of the kinetic and thermodynamic ionic strength dependency plots are very similar (Figure 2.5 a), with the thermodynamic slope being slightly higher. This result suggests that the monomer adopting the transition state for fibril elongation experiences very similar, albeit slightly less, electrostatic repulsion compared to the monomer fully incorporated into the fibril, thereby implying a high degree of similarity between the transition state and the final state of the fibril elongation reaction. Under conditions of acidic pH, where each monomer carries only positive charges and therefore a substantial net charge, electrostatic repulsion is probably mainly defined by the principal reaction coordinate [34], the center of mass distance between the monomer and the fibril end. Therefore, our results suggest that the transition state of PI3K-SH3 amyloid fibril elongation corresponds to a monomer in very close proximity of the fibril end, only marginally removed from its final position when incorporated into the fibril. This conclusion is in close agreement with the finding that the hydrophobic effect plays a major role in stabilizing the transition state of fibril elongation, as revealed through a strongly favorable entropy of activation [49, 51]. The picture that therefore emerges from this analysis is that the rate-limiting step of amyloid fibril elongation consists of a structural rearrangement of the monomer while being in close contact with the fibril end. Therefore, the defining energy barrier for fibril elongation is of an inter-molecular, rather than intra-molecular nature (see Figure 2.6 for a comparison of these two distinct scenarios). This result is an important contribution to the ongoing discussion about the intra- vs. intermolecular nature of the protein misfolding events that lead to amyloid fibril formation. It has been proposed for several amyloid systems, such as poly-glutamine [155] and tau [156] that the crucial and rate-limiting event along the pathway of amyloid fibril formation is a purely intra-molecular misfolding event to form an 'aggregation competent state' that can add onto a fibril without any significant barrier crossing. Our results for PI3K-SH3, on the other hand, are more compatible with the picture whereby the misfolding transition is a highly cooperative event between monomer and fibril end.

2.4. Conclusions

In summary, we have been able to show for the first time that the thermodynamics of amyloid fibril formation is best described by cooperative polymerization rather than by simple linear polymerization. This result is in excellent agreement with mechanistic insight into amyloid fibril formation, as well as with the emerging high resolution structural information on amyloid fibrils. Furthermore, we have been able to accurately probe the role of electrostatic effects in amyloid fibril stability. In combination with previously available data on the influence of solution ionic strength on the kinetics of amyloid fibril growth, we are able to reveal the product-like nature of the transition state-ensemble for amyloid fibril elongation by PI3K-SH3 monomers.

2.5. Materials and methods

2.5.1 Proteins

The human glucagon employed in the study was a kind gift from Novo Nordisk. The bovine PI3K-SH3 domain was purified according to the protocol in [39]. The constructs contains a 6xHis-tag linked to the protein by a thrombin cleavage site. The sequence of the WT protein after cleavage is the following, with the dipeptide Gly-Ser remaining as overhang from the cleavage:

GS MSAEGYQYRA LYDYKKEREE DIDLHLGDIL TVNKGSLVAL GFSDGQEAKP EEIG-WLNGYN ETTGERGDFP GTYVEYIGRK KISP

The protein was expressed in a BL21 *E. coli* strain with TB medium for auto induction containing 0.012% Glucose and 0.048% Lactose. The cells were grown for over 24 h and then harvested by centrifugation. After resuspension in sodium phosphate buffer (50 mM sodium phosphate pH

8, 5 mM Imidazole and 100 mM NaCl), the cells were disrupted by sonication, in the presence of protease inhibitors (cOmplete Mini EDTA-free, Roche) and DNAse (Sigma-Aldrich). The lysate was centrifuged, and the supernatant loaded on a Ni-NTA Superflow Cartridge (Qiagen, Venlo, Netherlands) equilibrated in 50 mM sodium phosphate pH 8, 5 mM Imidazole and 100 mM NaCl. The protein was eluted with a linear gradient from 5 to 300 mM imidazole in 50 mM sodium phosphate pH 8, 100 mM NaCl in 25 ml elution volume. Fractions containing the protein were collected and cleaved overnight at 7degC with 1 unit of thrombin (from bovine plasma, Sigma-Aldrich Saint Louis, Missouri, USA) per 1 mg of protein. The cleaved solution was then concentrated and loaded on a SEC HiLoad 26/60 Superdex 75 column (GE Healthcare, Chicago, Illinois, USA) equilibrated with 5 mM ammonium acetate pH 7. Fractions containing the PI3K-SH3 domain were collected and lyophilised for further use.

2.5.2 Fibril preparation

Glucagon fibrils were formed from protein solutions prepared after resuspension of the lyophilized peptide in 10 mM glycine hydrochloride pH 2 at 1 mM final concentration. This solution was incubated under shaking at 37°C for 1-2 hours. The fibrils obtained through this procedure were used as seeds for further solutions of monomeric protein, which were prepared at a concentration range between 1 and 2 mM and seeded at ca. 10% (monomer equivalents). The solution was left overnight at room temperature without shaking or stirring.

PI3K-SH3 fibrils were formed from protein solutions prepared from the lyophilized protein (produced as described above), resuspended in 10 mM glycine hydrochloride pH 2 at 200μM final concentration. This solution was incubated under shaking at 42°C overnight. The fibrils obtained through this procedure were then used as seeds for further solutions of monomeric protein, which were prepared at a concentration range between 200 and 300μM and then seeded at ca. 10% (equivalent monomer mass). The solution was left overnight at room temperature without shaking or stirring.

Such high concentrations of fibril stock solutions are needed, as the fibrils are strongly diluted upon addition of the denaturant, which is necessary to achieve denaturant concentrations high enough for depolymerization. Before preparing the samples, the fibril preparations were sonicated with a VialTweeter-sonotrode (Hielscher, Teltow, Germany). Glucagon fibril solutions were sonicated in a volume of at least 700μl, twice for 3 seconds, 100 % amplitude, with a pause of ca. 30 s. PI3K-SH3 fibrils were sonicated for 10 s at 100% amplitude in a volume of at least 700μl. The fibril preparations were imaged by AFM, both before and after the sonication protocol, to evaluate the effect of the sonication on the lengths of the fibrils.

2.5.3 Atomic force microscopy imaging

Atomic force microscopy (AFM) was performed on dried fibril samples. The fibril samples were diluted to ca. 3μM in the same buffer, then 20μl were placed on freshly cleaved mica and incubated at room temperature for 10 min. After 10 min the mica was washed with milliQ water. Imaging was performed with a Bruker Multimode 8 (Billerica, Massachusetts, USA) using OMCL-AC160TS cantilevers (Shinjuku, Tokyo, Japan).

2.5.4 Depolymerization experiments

The fibril samples were mixed with different volumes of urea or GdnHCl stock solutions and buffer. In order to maintain a solution pH of 2 constant throughout the whole denaturation series, an 8 M urea stock solution was prepared by dissolving 12 g of urea (SigmaAldrich) in 16 ml of concentrated buffer, to yield a final concentration of 10 mM glycine hydrochloride and 16 mM HCl. The HCl is necessary, as the urea has a weak buffering capacity. In order to keep the ionic strength constant in all samples, NaCl was added to the 10 mM glycine HCl buffer to a final concentration of 16 mM. In order to ensure that the samples have reached equilibrium when they are analyzed, we assessed the time of equilibration (see Supplementary Figure 2.11) and based on these assessments left the samples to equilibrate for one week in the case of glucagon fibrils and two weeks in the case of PI3K-SH3 fibrils.

2.5.5 Fluorescence measurements

Fluorescence spectra were recorded on a Tecan M1000pro instrument using a Greiner UV-transparent 96 well plates. The wells were filled with 140μl of solution and the temperature was maintained constant at 27°C. The fluorescence spectra were measured by top reading, exciting at 280 nm for glucagon samples and 290 nm for PI3K-SH3 samples (5 nm bandwidth), while the emission was recorded between 300 nm and 420 nm (5 nm bandwidth). A blank spectrum (buffer) resulting from the average of 10 different spectra is subtracted before analysis. In order to determine the relative populations of soluble vs. fibrillar protein, we computed the ratios of the fluorescence intensities at 340 and 320/310 nm (glucagon/PI3K-SH3) for each spectrum. This choice of wavelengths was based on the difference between the fibrillar and monomeric spectra in each case (Supplementary Figure 2.8).

2.5.6 Absorbance measurements

In order to measure the degree of aggregation of the fibrillar samples, they were centrifuged at 16.100 g for 1 hour. The supernatant was removed and its absorbance was measured between 220 and 350 nm with a Nanodrop 2000 (ThermoFisher) spectrophotometer. The extinction coefficients at 280 nm are 8480 M^{-1} for glucagon and 15930 M^{-1} for PI3K-SH3.

2.6. Acknowledgements

AKB and NV thank the Deutsche Forschungsgemeinschaft (DFG) for financial support. AKB thanks the Novo Nordisk Foundation for support through a Novo Nordisk Foundation professorship. The authors thank Sabine Metzger for help with the mass spectrometry, Florian Platten for helpful discussions and Novo Nordisk for a kind gift of glucagon.

2.7. Supplementary information

Supplementary results

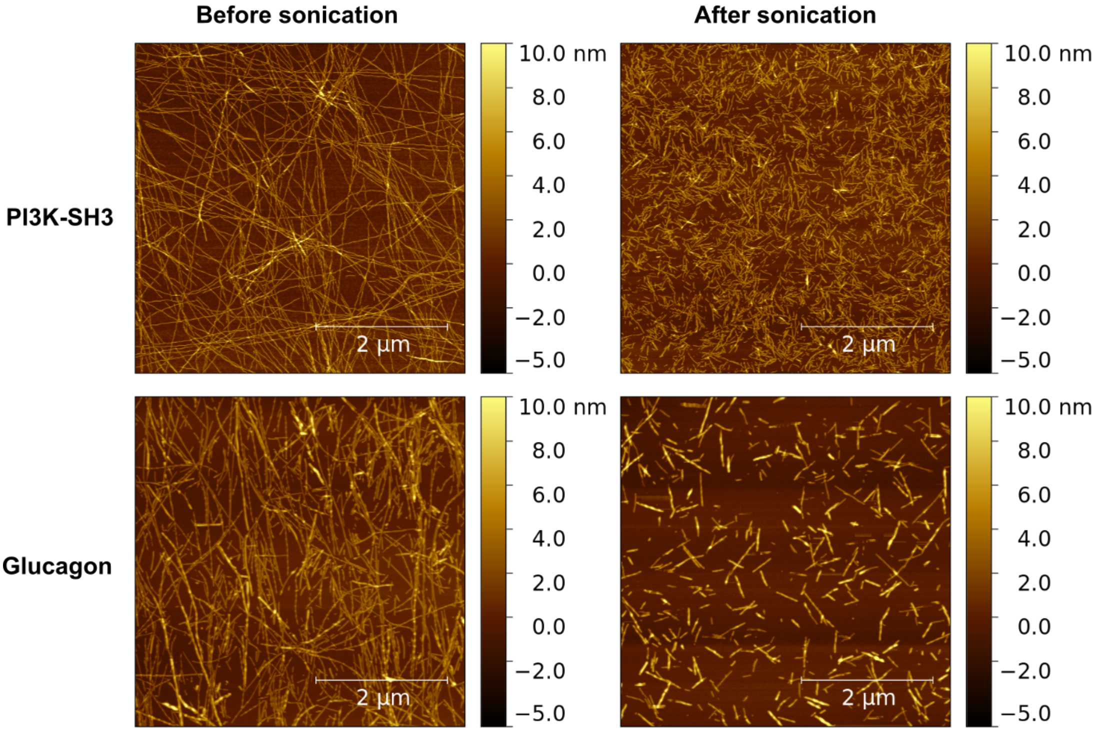

Figure 2.7: AFM images of PI3K-SH3 and glucagon fibrils, before and after sonication. See experimental methods section in main manuscript for details on the sonication and AFM imaging protocols.

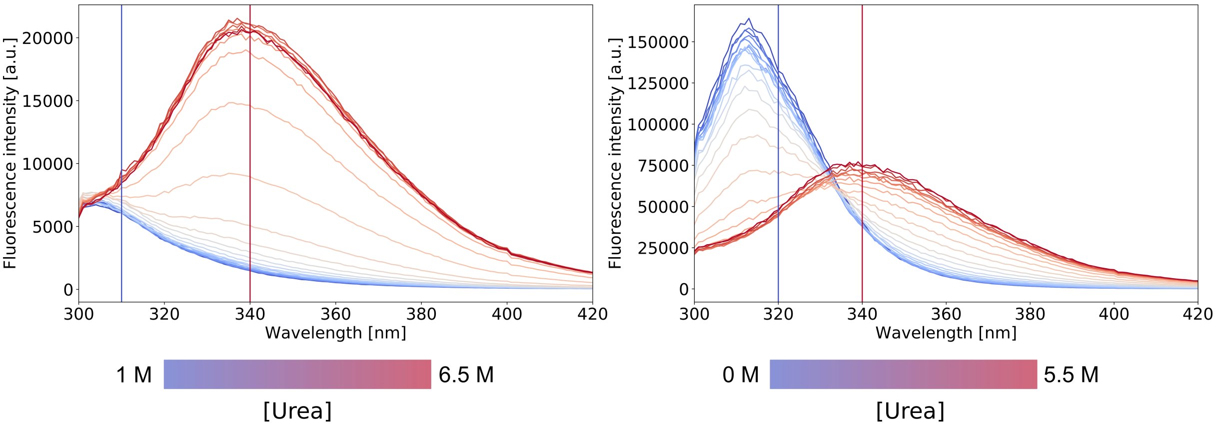

Figure 2.8: Intrinsic fluorescence spectra of PI3K-SH3 (left) and glucagon (right) at various denaturant concentrations. The fluorescence intensity ratios at the indicated wavelengths were analysed.

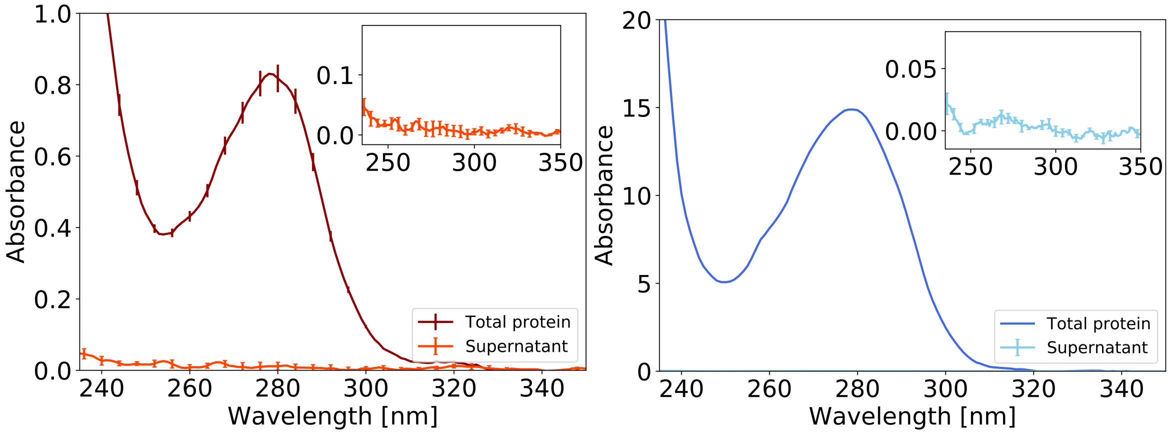

Figure 2.9: Quantification of soluble protein in supernatant before and after fibrillation, using absorbance spectroscopy. Left: PI3K-SH3; Right: glucagon. The errorbars represent the standard deviation on a sample size of n = 3. Details on the protocol of centrifugation can be found in the experimental methods section in the main manuscript. It can be seen from these results that the fibril formation raction is near-complete and that the protein concentration in the supernatant is too low to be reliably measurable, illustrating the need to destabilise the fibrils with chemical denaturants.

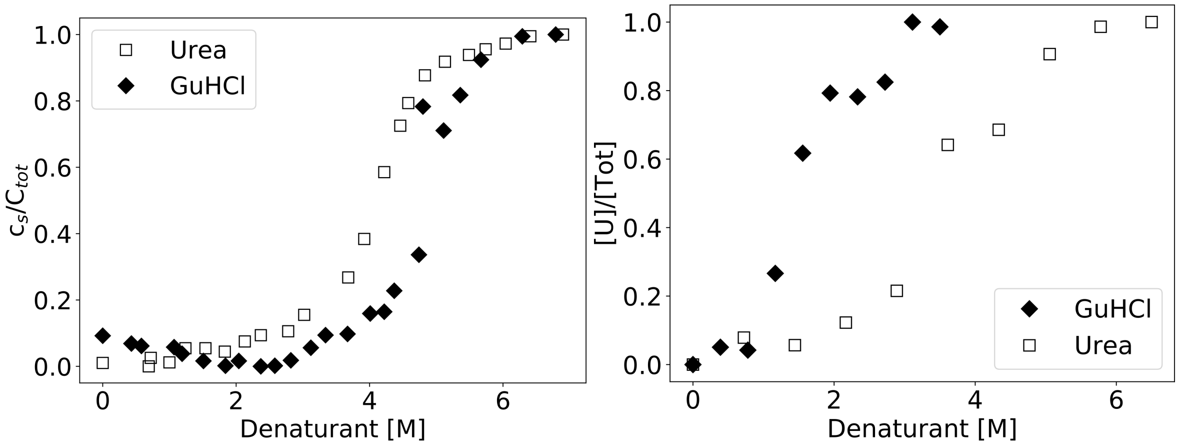

Figure 2.10: Comparison of urea and GdnHCl depolymerization of SH3 fibrils at a total protein concentration of 50µM. Left panel: amyloid fibrils; Right panel: monomers.

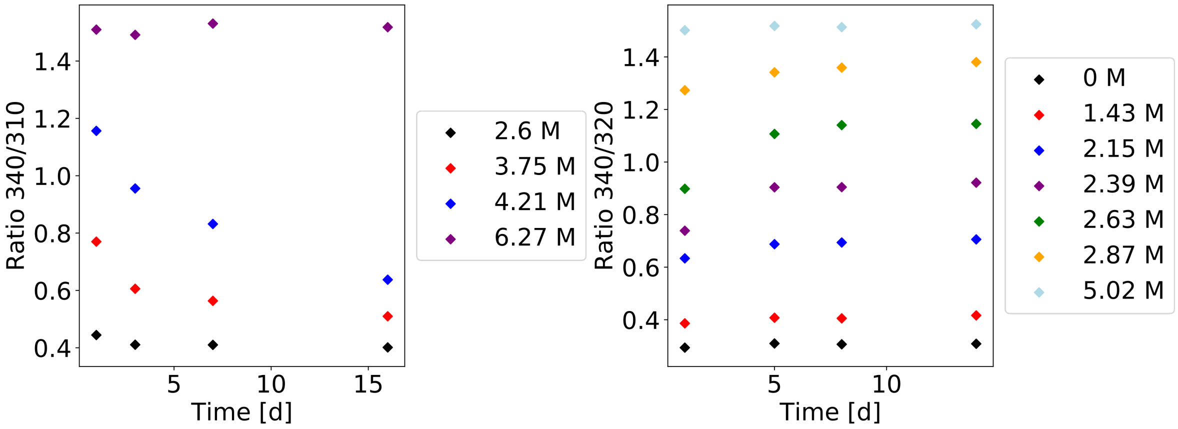

Figure 2.11: Time course of the fluorescence signal (expressed as intensity ratios at two wavelengths) of fibril samples with different urea molarities in order to monitor the time required for equilibration. Left: PI3K-SH3; Right: glucagon.

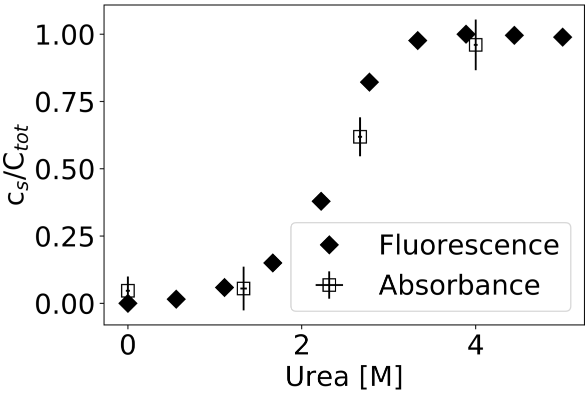

Figure 2.12: Direct comparison between absorbance (after centrifugation) and intrinsic fluorescence (intensity ratio 340/320 nm) data of urea depolymerization of glucagon amyloid fibrils. The errorbars represent the standard deviation on a sample size of n = 3.

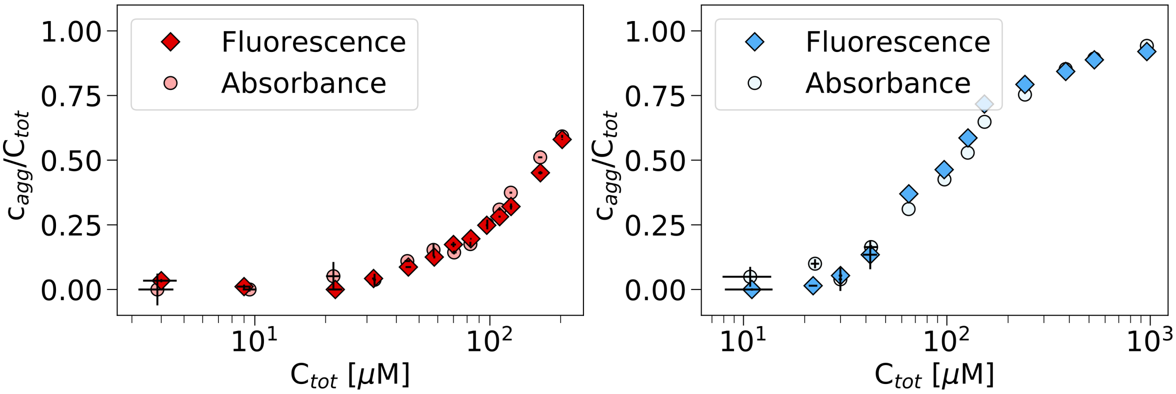

Figure 2.13: The degree of aggregation as a function of the total protein concentration for a given denaturant concentration. Data obtained from intrinsic fluorescence and direct quantification of soluble protein from supernatant absorbance are compared. In the fluorescence measuremnts the intensity ratios at 340/310 nm (PI3K-SH3) and 340/320 nm (glucagon) are monitored. Left panel: PI3K-SH3 amyloid fibrils equilibrated in 4 M Urea. Right panel: glucagon amyloid fibrils equilibrated in 3 M Urea.

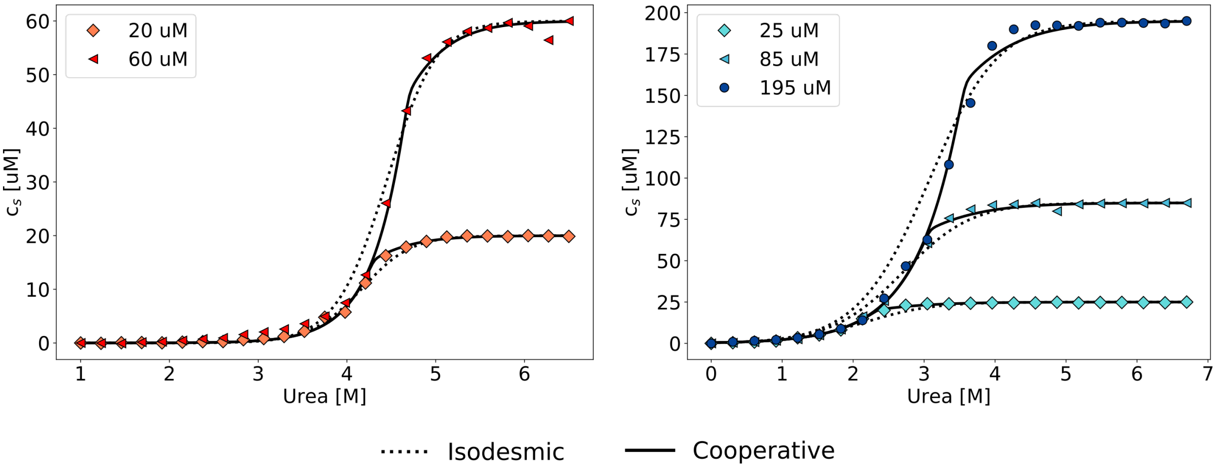

Figure 2.14: This is the same data set of chemical depolymerisation of PI3K-SH3 (left) and glucagon (right) as shown in Figures 1 and 2 of the main manuscript, but without normalising by the total protein concentration.

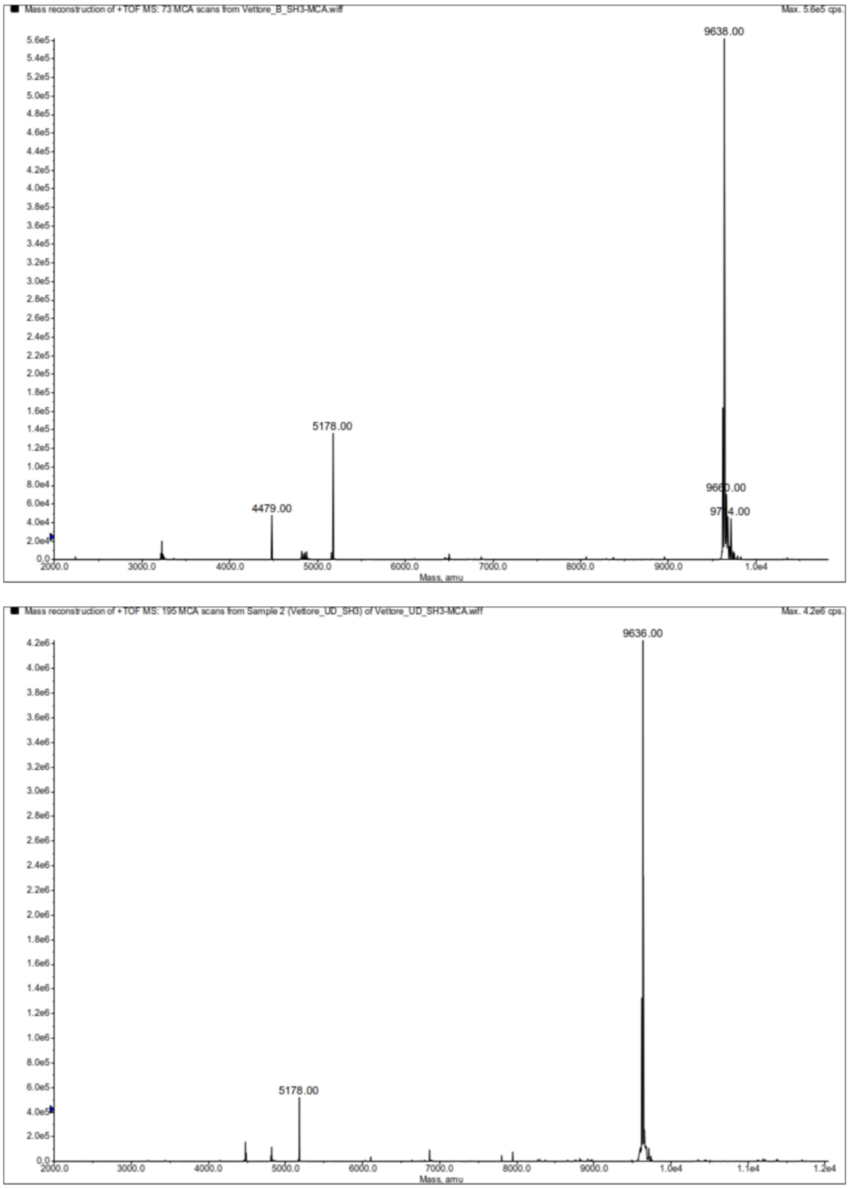

Figure 2.15: Mass spectrea of PI3K-SH3 to probe carbamydation. PI3K-SH3 monomers were incubated for 9 days at room temperature in both buffer, 10 mM Glycine HCl pH 2 (top), and 6M Urea in buffer (bottom), at a final concentration of 100 μM. Afterwards, the sample containing urea was extensively dialysed against buffer to remove the urea. Both samples were then analysed by ESI-MS to check for possible carbamylation of the basic moieties of the protein. A small amount of fragmented protein (at position 39) is present in both samples and likely stems from the production.

Models for data analysis

Within the theoretical framework of the linear polymerization model we can describe the thermodynamics of formation of a linear polymer, such as an amyloid fibril, as a series of equilibria:

$$\begin{aligned}
\mathrm{M} + \mathrm{M} &\xrightleftharpoons[\mathrm{k_{\text{-}2}}]{\mathrm{k_2}} \mathrm{M_2} \\
\mathrm{M_2} + \mathrm{M} &\xrightleftharpoons[\mathrm{k_{\text{-}3}}]{\mathrm{k_3}} \mathrm{M_3} \\
\mathrm{M_3} + \mathrm{M} &\xrightleftharpoons[\mathrm{k_{\text{-}4}}]{\mathrm{k_4}} \mathrm{M_4} \\
&\cdot \\
&\cdot \\
\mathrm{M}_{n-1} + \mathrm{M} &\xrightleftharpoons[\mathrm{k_{\text{-}n}}]{\mathrm{k_n}} \mathrm{M}_n \\
&\cdot \\
&\cdot \\
\mathrm{M}_{i-1} + \mathrm{M} &\xrightleftharpoons[\mathrm{k_{\text{-}i}}]{\mathrm{k_i}} \mathrm{M}_i
\end{aligned} \tag{2.1}$$

Each of these reactions is defined by an equilibrium constant which is defined as:

$$\mathrm{K}_i = \mathrm{k}_i/\mathrm{k}_{-i} = \frac{[\mathrm{M}_i]}{[\mathrm{M}_{i-1}][\mathrm{M}]} \tag{2.2}$$

Every $\mathrm{K_i}$ is then related to the free energy change, $\Delta \mathrm{G}_0$, by the relation:

$$\Delta \mathrm{G}^0 = -\mathrm{RT} \ln \mathrm{K_i} \tag{2.3}$$

The description of any real supra-molecular polymer system can then be based on this series of reactions by assuming certain relations between the different equilibrium constants, as will be shown in the following.

Isodesmic model

In the simplest form of the linear polymerization model, the so-called isodesmic model, it is assumed that the equilibrium constants $\mathrm{K_i}$ are equal for all the interactions. This means that the dimerization reaction will be associated with the same change in free energy as the interaction between a monomer and a polymer of dimensions *i*. This simplicity allows to obtain an analytical solution that relates the monomer concentration at equilibrium to the total concentration of in the following way:

$$\mathrm{K}_e[\mathrm{M}] = 1 + \frac{1}{2\,\mathrm{K}_e[\mathrm{M}]_{\mathrm{tot}}} - \sqrt{\frac{1}{\mathrm{K_e}[\mathrm{M}]_{\mathrm{tot}}} + \frac{1}{4(\mathrm{K_e}[\mathrm{M}]_{\mathrm{tot}})^2}} \tag{2.4}$$

where $\mathrm{K_e}$ is the equilibrium constant that defines all the interactions between monomers and the other species in solutions, and $[\mathrm{M}]_{\mathrm{tot}}$ is the total concentration of the sample. This equation therefore describes the evolution of the concentration of monomeric protein, [M], as a function of total concentration. It is clear from this expression that the monomer concentration in an isodesmic system only asymtotically approaches a constant value, the so-called critical concentration.

In fibril depolymerization experiments where the concentration of denaturant, [D], is varied at constant $[\mathrm{M}]_{\mathrm{tot}}$, $\mathrm{K_e}([\mathrm{D}])$, is the independent variable. The two following relations can therefore be defined, in direct analogy to the linear free energy relationships in protein folding [56]:

$$\mathrm{K}_e([\mathrm{D}]) = \exp - \frac{\Delta\mathrm{G}([\mathrm{D}])}{\mathrm{RT}} \tag{2.5}$$

$$\Delta\mathrm{G}([\mathrm{D}]) = \Delta\mathrm{G_0}^{\mathrm{Buff}} + \mathrm{m}[\mathrm{D}] \tag{2.6}$$

where $\Delta\mathrm{G_0}^{\mathrm{Buff}}$ is the free energy in the absence of denaturant. By introducing these terms in the equation 2.4, we can describe the depolymersiation curve as a function of the denaturant concentration as follows:

$$\mathrm{f}_s = \frac{[\mathrm{M}]}{[\mathrm{M}]_{\mathrm{tot}}} = \frac{2\,[\mathrm{M}]_{\mathrm{tot}} \exp - \frac{\Delta G + m[D]}{RT}{}_1{}^{+} - \sqrt{4[\mathrm{M}]_{\mathrm{tot}} \exp - \frac{\Delta\mathrm{G}+\mathrm{m}[\mathrm{D}]}{\mathrm{RT}} + 1}}{2\,[\mathrm{M}]_{\mathrm{tot}}{}^2 (\exp - \frac{\Delta G + m[D]}{RT})^2} \tag{2.7}$$

The dependence of the depolymerization profile on the total concentration allows to test the applicability of this theoretical framework to the system under study by globally fitting data at different protein concentrations, $[\mathrm{M}]_{\mathrm{tot}}$. In addition to testing the model, this procedure that we have employed in the present study, allows also to extract more robust thermodynamic parameters. Tables 2.1 and 2.2 below summarise the results of the possible combinations of glocal and local fits of the data shown in Figure 1 and 2 of the main manuscript.

Experimental data of amyloid fibril depolymerization by chemical denaturants has so far only been analyzed in the framework of this simplest model [73–75]. In the present study, however, we find that the isodesmic model is not able to quantitatively describe the ensemble of depolymerization curves at different protein concentrations. Furthermore, from a fundamental point of view, the isodesmic model is not a good description of a nucleated polymerization. Several simple extensions of the isodesmic model have therefore been introduced, all based on the idea that the equilibrium constant is size-dependent.

		Fit m [KJ mol^{-2}] as:				
		Local		Global		
Fit ΔG [KJ mol^{-1}] as:	Local	-64.7	9.5	-70.0	10.1	20 µM
		-71.1	10.8	-67.7		60 µM
	Global	-66.8	10.0	-66.5	9.9	20 µM
			9.9			60 µM

Table 2.1: Parameters obtained from fits of the PI3K-SH3 amyloid fibril depolymerisation by using the isodesmic model. The table summarises the results from all combinations of global and local fits of the parameters ΔG_0 and m.

		Fit m [KJ mol^{-2}] as:				
		Local		Global		
Fit ΔG [KJ mol^{-1}] as:	Local	-36.7	6.0	-35.7	5.7	25 µM
		-38.8	6.3	-37.2		85 µM
		-42.2	6.9	-38.2		195 µM
	Global	-36.9	6.1	-36.7	5.5	25 µM
			5.7			85 µM
			6.3			195 µM

Table 2.2: Parameters obtained from fits of the glucagon amyloid fibril depolymerisation by using the isodesmic model. The table summarises the results from all combinations of global and local fits of the parameters ΔG_0 and m.

Helical and cooperative polymerization models

In their studies of actin, Oosawa and co-workers noticed that the polymerization behaviour of this protein was not compatible with the isodesmic model and as a consequence introduced the helical polymerization model [17, 61]. In this model, a transition between the linear form of a trimer and a helical form is energetically unfavourable, due to elastic energy of deformation, but the addition of a monomer to the helical form is energetically more favorable than addition to the linear form, due to the larger number of inter-molecular contacts. The relevant equilibrium concentrations can be described as follows:

$$
\begin{aligned}
&[\mathrm{M}_{3h}] = \gamma[\mathrm{M}_3] = \gamma\mathrm{K}^{-1}(\mathrm{K}[\mathrm{M}])^3 \\
&[\mathrm{M}_{4h}] = \mathrm{K}_h\,[\mathrm{M}_{3h}]\,[\mathrm{M}] = \gamma\mathrm{K}^{-1}\mathrm{K}^3\,\mathrm{K}_h\,[\mathrm{M}]^4 \\
&\quad \dots. \\
&[\mathrm{M}_{\mathrm{ih}}] = \mathrm{K}_h\,[\mathrm{M}_{(i-1)h}]\,[\mathrm{M}] = \gamma(\mathrm{K}/\mathrm{K}_h)^2\,\mathrm{K}_h{}^{-1}\,(\mathrm{K}_h\,[\mathrm{M}])^i = \sigma\mathrm{K}_h{}^{-1}\,(\mathrm{K}_h\,[\mathrm{M}])^i
\end{aligned}
\tag{2.8}
$$

where $\sigma = \gamma(K/K_h)^2$, with K the equilibrium constant of addition to a linear polymer, K_h the equilibrium constant for addition to a helical polymer and $\gamma < 1$ the unfavorable Boltzmann factor for mechanical deformation of the linear into a helical polymer. Usually, $\sigma \ll 1$ and therefore the system displays nucleation behaviour. It can be shown [17] that the total protein concentration in such a helically polymerizing system can be written as

$$[M]_{tot} = \frac{[M]}{(1-K[M])^2} + \frac{\sigma[M]}{(1-K_h[M])^2} - \sigma[M] - 2\sigma K_h[M] \tag{2.9}$$

which, for $\sigma \ll 1$ and $K < K_h$ can be simplified to

$$[M]_{tot} = [M] + \frac{\sigma[M]}{(1-K_h[M])^2} \tag{2.10}$$

which predicts a well-defined critical concentration $[M_c] = K_h^{-1}$. The helical polymerization model in this form is elegant and well adapted to account for the specific structural features of actin filaments. It is, however, not the simplest extension of the isodesmic model. We chose to analyze our data on the chemical depolymerization of amyloid fibrils with an even simpler, and somewhat more general model, the so-called cooperative linear polymerization model [24, 65]. This model introduces the assumption of a minimal unit of the polymer, called nucleus, the energetics of formation of which differs from that of elongation according to the following relation:

$$\mathrm{K}_n = \sigma \mathrm{K}_e \tag{2.11}$$

Note that the definition of σ is different than in the helical model. In this model, all reactions between a monomer and species up to n-mers (where n is the nucleus size) are described by the equilibrium constant $\mathrm{K_n}$, while the interactions between monomers and species with a higher number of monomers than n follow the equilibrium dictated by the constant $\mathrm{K_e}$. A detailed description of this model can be found in the review from Zhao and Moore [65]. Here we limit ourselves to a discussion of its application to analyze data of amyloid fibril thermodynamics.

Introducing the dimensionless concentrations $\mathrm{x_{tot}=K_e[M]_{tot}}$ and $\mathrm{x=K_e[M]}$, the relation between the two is described by the following equation formulated by Korevaar et al.[25]:

$$\mathrm{x_{tot}} = \sigma^{-1}\left(\frac{(\sigma \mathrm{x})^{n+1}(\mathrm{n}\sigma\mathrm{x}-n-1)}{(\sigma\mathrm{x}-1)^2} + \frac{\sigma\mathrm{x}}{(\sigma\mathrm{x}-1)^2}\right) - \sigma^{n-1}\left(\frac{\mathrm{x}^{n+1}(\mathrm{nx}-n-1)}{(\mathrm{x}-1)^2}\right) \tag{2.12}$$

This equation is solved numerically through the use of the *least_square* module of the *scipy.optimize* python library. In order to fit experimental data from a concentration series (see e.g. Figure 3 of the main manuscript), the solution of this equation for multiple $\mathrm{x_{tot}}$ is fitted using $\mathrm{[M]_{tot}}$ as the independent variable and $\mathrm{K_e}$ as an open parameter (using the relation $\mathrm{x_{tot}=K_e[M]_{tot}}$).

The fitting procedure for a depolymerization experiment, whereby the denaturant concentration is varied requires instead a different approach. The independent variable is [D], while $\mathrm{[M]_{tot}}$ is kept constant and $\mathrm{K_e}$ is described in terms of $\mathrm{K}_e(\mathrm{[D]}) = \exp(\Delta \mathrm{G_0}^{\mathrm{Buff}} + \mathrm{m[D]})/(\mathrm{RT})$ with $\Delta \mathrm{G_0}^{\mathrm{Buff}}$, m and σ as fitting parameters, whereas n is kept constant for each given run.

Tables 2.3 and 2.4 below show the results of a systematic variation of the parameter n. We have performed fits for two types of models, which we call osaa and osam. The former (osaa, oligomers same as aggregates) corresponds to the scenario whereby the oligomers up to size n have the same fluorescence properties as the larger aggregates, whereas the latter (osam, oligomers same as monomers) corresponds to the scenario whereby all species up to size n have the same fluorescence properties as the monomer. It can be seen that the cooperative model, for all choices of n, provides a much better description of the concentration dependent depolymerization experiments than the isodesmic model. The fit parameters and fit quality are relatively insensitive to the exact choice of n, however, for both proteins the best results are achieved for small values of n of approximately 5. Therefore, we have used n = 5 for the analysis of the influence of changes in salt concentration on the stability of the aggregates, see below. Regarding the choice of model (osam vs. osaa), we conclude that the osaa is not only structurally more plausible (due to the disordered nature of the monomers of both PI3K-SH3 [39] and glucagon [142] under these conditions, it can be expected that the monomer will have a distinct spectroscopic signature from any oligomeric species, which are likely to be more compact, with solvent-protected Trp), but also yields lower values for σ. Smaller σ values correspond to a more distinct thermodynamic stability of small (<n) vs. large (>n) species, which we think is more plausible than σ value closer to unity.

We would like to stress here that n should not be regarded as a critical nucleus size. The values for ΔG_0 and σ we obtain from our fits suggest that also species below the size n are thermodynamically favourable (negative free energy of monomer attachment), albeit less so compared to monomer attachment to a species larger than n. In the framework of classical nucleation theory, the nucleus is considered a thermodynamically unstable species, i.e. the free energy of addition of a monomer to a species smaller than n is positive [140]. Our results suggest that the formation of small oligomeric species from amyloid proteins is thermodynamically favourable. However, it is still possible to explain the slow nucleation of amyloid fibrils, namely through the assumption of high free energy barriers for nucleation, as has been observed for example for the amyloid β-peptide [51]. When starting an aggregation experiment with freshly prepared monomeric protein, the establishment of the equilibrium distribution of small oligomers depends on the energy barriers rather than the thermodynamics of oligomer formation.

n	signal	ΔG_0 [KJ/mol]	m [KJ/mol^2]	σ	R^2
isodesmic		-66.9	10	NaN	99.6
1	osaa	-67.9	10	NaN	99.6
	osam	-67.1	9.3	NaN	99.7
2	osaa	-59.2	7.6	0.110	99.8
	osam	-60.3	7.8	0.229	99.8
3	osaa	-58.5	7.3	0.139	99.8
	osam	-62.0	7.8	0.506	99.8
4	osaa	-57.7	7.1	0.132	99.8
	osam	-63.3	7.8	0.640	99.8
5	osaa	-57.1	6.9	0.121	99.8
	osam	-64.3	7.9	0.715	99.8
8	osaa	-57.0	6.8	0.113	99.8
	osam	-66.3	7.9	0.821	99.7
10	osaa	-56.9	6.8	0.113	99.8
	osam	-67.3	7.9	0.860	99.7
20	osaa	-63.9	8.6	0.285	99.6
	osam	-70.5	7.8	0.930	99.7
40	osaa	-63.9	8.6	0.285	99.6
	osam	-73.7	7.8	0.962	99.7

Table 2.3: Fitting parameters ΔG_0, m and σ from a systematic variation of n, in the framework of the two models osaa and osam (see text for details) for PI3K-SH3.

Electrostatic contribution to amyloid fibril stability

The basic assumption of the approach we present here is that the dependence of the rates and equilibria of molecular processes in aqueous solution on the salt concentration (or, more precisely, on the ionic strength) of the solution informs about the importance of electrostatic interactions in these processes and states. Conceptually, this approach is rooted in the work of Debye and Hückel, who presented a theory of electrostatic interactions in dilute salt solutions [157]. Here we follow and extend a treatment that we have presented previously in the context of the ionic strength dependence of amyloid fibril elongation rates [37]. The rate of amyloid fibril growth, R, can be written as R = k_e[M][P], where k_e is the rate constant of elongation, [M] the monomer concentration and [P] the concentration of fibril ends. We have discussed previously [34, 37] that k_e can be written as $k_e = \Gamma e^{-\frac{\Delta G^\ddagger}{RT}}$, whereby Γ is a diffusive pre-factor and $\Delta G^\ddagger$ is the free energy barrier for fibril elongation. An important point to mention here is while the process of incorporation of a monomer into a fibril end can involve the sequential crossing of multiple barriers, it is the highest barrier only that determines the kinetics [34, 150]. This free energy barrier can be split into a non-electrostatic part ($\Delta G^\ddagger_{ne}$, breakage of H-bonds and other non-covalent interactions, hydrophobic effect, etc. [49]) and an electrostatic part ($\Delta G^\ddagger_e$). We postulate that the electrostatic contribution to $\Delta G^\ddagger$ can be written as a screened Coulomb potential [37, 154]:

$$\Delta G^\ddagger = \Delta G^\ddagger_{ne} + \Delta G^\ddagger_e = \Delta G^\ddagger_{ne} + \frac{N_A z_M z_F e^2}{4\pi\varepsilon_0\varepsilon_r r^\ddagger} e^{-\kappa r^\ddagger} \tag{2.13}$$

		ΔG_0 [KJ/mol]	m [KJ/mol^2]	σ	R^2
n	signal				
isodesmic		-36.7	5.5	NaN	98.6
1	osaa	-37.4	5.7	NaN	98.6
	osam	-37.4	5.3	NaN	99.0
2	osaa	-36.0	4.2	0.030	99.8
	osam	-36.2	4.2	0.041	99.8
3	osaa	-36.2	4.1	0.084	99.8
	osam	-37.0	4.2	0.212	99.8
4	osaa	-36.2	4.1	0.097	99.8
	osam	-37.9	4.2	0.350	99.8
5	osaa	-36.3	4.1	0.102	99.7
	osam	-38.7	4.2	0.44	99.8
8	osaa	-36.5	4.2	0.115	99.7
	osam	-40.4	4.2	0.620	99.8
10	osaa	-36.5	4.2	0.115	99.7
	osam	-41.3	4.2	0.682	99.8
20	osaa	-36.5	4.2	0.115	99.7
	osam	-44.4	4.2	0.833	99.8
40	osaa	-36.5	4.2	0.115	99.7
	osam	-47.6	4.2	0.910	99.8

Table 2.4: Fitting parameters ΔG_0, m and σ from a systematic variation of n, in the framework of the two models osaa and osam (see text for details) for glucagon.

where free energies are given in J/mol. Here, z_M and z_F denote the effective charges of the monomer and the fibril end, respectively, that lead to the electrostatic repulsion in the transition state. $\kappa = \sqrt{\frac{2000e^2 I N_A}{\varepsilon_0 \varepsilon_r k_B T}}$ is the inverse Debye length and $r^\ddagger$ corresponds to the average centre of mass distance between the monomer and the fibril end in the transition state ensemble [34, 37, 49]. For $\kappa r^\ddagger < 1$, we can linearize this expression, by using $e^{-x} \approx 1 - x$ for small x:

$$\Delta G^\ddagger = \Delta G^\ddagger_{ne} + \frac{N_A z_M z_F e^2}{4\pi\varepsilon_0\varepsilon_r r^\ddagger} - \frac{N_A z_M z_F e^2 \kappa}{4\pi\varepsilon_0\varepsilon_r} = \tag{2.14}$$

In this expression, only κ depends explicitly on the ionic strength; therefore, if we derive the logarithm of the fibril elongation rate with respect to the square root of the ionic strength, we obtain:

$$\frac{d\log R}{d\sqrt{I}} = \frac{N_A z_M z_F e^2}{4\pi\varepsilon_0\varepsilon_r RT}\frac{d\kappa}{d\sqrt{I}} = \frac{N_A z_M z_F e^2}{4\pi\varepsilon_0\varepsilon_r RT}\sqrt{\frac{2000e^2 N_A}{\varepsilon_0\varepsilon_r k_B T}} = A z_M z_F \tag{2.15}$$

We see, therefore, that the orange slope of a plot of the logarithm of the fibril elongation rate against the square root of the ionic strength is proportional to the product of the effective charges on the monomer and fibril end that are relevant for the interaction in the transition state of fibril elongation [37].

We next focus on the equilibrium constant for fibril elongation, which is related to the free energy

Ionic strength [$\sqrt{M}$]	ΔG_0 [KJ/mol] m glob. fit	ΔG_0 [KJ/mol] m fixed
0.17	-35.2	-36.8
0.22	-37.3	-39.1
0.28	-38.2	-40.6
0.32	-39.5	-42.3
0.36	-40.1	-42.9

Table 2.5: Glucagon amyloid fibril stabilities as a function of salt concentration. In the first column, sigma is kept constant at 0.064, n is kept constant at 5 and m is globally fitted among the whole dataset. In the second column, m is also fixed as well at 4400 KJ/mol, the value determined previously from the fit of the data set in the absence of salt.

change associated with the addition of one mole of monomer to fibril ends according to equation (3), $\Delta G^0 = -RT\ln K_e$. Again, we can divide ΔG^0 into a non-electrostatic contribution and an electrostatic contribution:

$$\Delta G^0 = \Delta G^0_{ne} + \Delta G^0_e = \Delta G^0_{ne} + \frac{N_A z'_M z'_F e^2}{4\pi\varepsilon_0\varepsilon_r r^{eq}} e^{-\kappa r^{eq}} \tag{2.16}$$

This expression is analogous to equation 2.13, except for the substitution of the ‡ ('double dagger') superscript (notation for transition state quantities) by the superscript 0. Here, r^{eq} denotes the center of mass distance of the last incorporated monomer from the previous fibril end, i.e. the center of mass of the second to last incorporated monomer. The effective charges have been primed in order to express the idea that the effective charges acting in the transition state ensemble and in the final state can be different. A rearrangement of this expression analogous to the one outlined for the kinetic expression above leads to:

$$\frac{d\log K_e}{d\sqrt{I}} = \frac{N_A z'_M z'_F e^2}{4\pi\varepsilon_0\varepsilon_r RT}\frac{d\kappa}{d\sqrt{I}} = \frac{N_A z'_M z'_F e^2}{4\pi\varepsilon_0\varepsilon_r RT}\sqrt{\frac{2000e^2 N_A}{\varepsilon_0\varepsilon_r k_B T}} = A z'_M z'_F \tag{2.17}$$

Our experimental data for PI3K-SH3 and glucagon (Figure 4 in the main manuscript, individual values of ΔG_0 can be found in tables 2.5 and 2.6 below) show that the slope is larger for PI3K-SH3, suggesting a higher effective charge that destabilizes the fibrils. This is consistent with the higher formal charge of PI3K-SH3 at this acidic pH, based on the amino acid sequence. In our analysis, we explored two different ways to fit the ionic strength dependence data, by keeping m fixed to the values determined previously and by fitting m globally to the ionic strength-dependent data set (see two columns in tables 2.5 and 2.6). We found that the general conclusions drawn here are independent of the exact manner in which the denaturation curves are fitted, and we show the results of the method with fixed m-value in the plots in the main manuscript.

We next proceed to relate the ionic strength dependencies of both kinetics and thermodynamics of fibril elongation. Our analysis above shows that the slopes for the kinetic and thermodynamic ionic strength dependencies are both independent of the characteristic distances, $r^{\ddagger}$ and r^{eq}, and

Ionic strength [$\sqrt{M}$]	ΔG$_0$ [KJ/mol] m glob. fit	ΔG$_0$ [KJ/mol] m fixed
0.17	-52.0	-56.0
0.18	-54.2	-59.0
0.19	-55.1	-59.9
0.22	-56.3	-61.2

Table 2.6: PI3K-SH3 amyloid fibril stabilities as a function of salt concentration. In the first column, sigma is kept constant at 0.121, n is kept constant at 5 and m is globally fitted among the whole dataset. In the second column, m is also fixed as well at 6900 KJ/mol, the value determined previously from the fit of the data set in the absence of salt.

have the same numerical pre-factor, $A = \frac{N_A e^2}{4\pi\varepsilon_0\varepsilon_r RT}\sqrt{\frac{2000e^2 N_A}{\varepsilon_0\varepsilon_r k_B T}}$. Therefore, in order to eliminate the numerical constant A, we can express the ratio of the slopes as:

$$\frac{d\log R}{d\sqrt{I}} \Bigg/ \frac{d\log K_e}{d\sqrt{I}} = \frac{z_M z_F}{z'_M z'_F} \tag{2.18}$$

We next perform a close analysis of the behavior of PI3K-SH3 for which both the dependence of the kinetics [37] and thermodynamics (this work) on solution ionic strength has been determined. We find that the slopes and therefore products of the effective charges are extremely similar in both the transition state and the final state (Figure 5 a of the main manuscript). This finding provides strong evidence for the fact that the transition state ensemble of fibril elongation, at least in this protein system, is highly product-like, regarding the center of mass distance component of the multi-dimensional reaction coordinate [34]. It is likely that under these conditions of high net charge, the electrostatic interactions are dominated by the center of mass distance between the newly incorporating monomer and the monomer that corresponds to the fibril end, rather than by the internal degrees of freedom. Therefore, no information on the similarity of the internal structures between transition state ensemble and final state can be obtained from our analysis.

3

THERMAL CHARACTERIZATION OF THE AMYLOID FIBRIL ELONGATION REACTION

3.1. Introduction

In order to determine the driving forces of a reaction it is fundamental to study its behaviour in the temperature dimension. By observing the reaction upon changes in temperature, it is possible to extract information about the enthalpic and entropic contribution to both the kinetics and stability of the system. The pioneering studies of actin and flagellin assemblies in the temperature dimension led to the observation that the energetic reaction signatures in the experimentally measured temperature ranges are not monotonic, highlighting the complexity of the system [14, 158]. Similarly, in the study of the supramolecular polymers composed by aromatic molecules, temperature is a key parameter in distinguish the assembly behaviour. Originally, the study of the polymerization model was done by measuring the polymer and monomer concentrations as a function of the total concentration. This meant monitoring the system over a wide range of concentrations, which could not be done with a single methodology due to limited detection ranges. Instead of working in the concentration dimension, the temperature dimension allows us to obtain information both on the model and on the enthalpic component of the assembly interactions. Cooling down the monomeric solution proved to be an easier and clearer way to tell whether the system behaves cooperatively or isodesmically, as the signature of a single concentration can be monitored throughout the whole temperature range with the same methodology [24, 66, 67].

Nowadays amyloid fibrils are one of the most researched protein assemblies, and many studies have addressed the temperature dependence of their formation with a focus on the thermal signature of the elongation reaction barrier [38, 41–51]. In all the monitored temperature ranges, the temperature dependence of the elongation reaction appeared to be monotonic, with some possible deviations at either high or low temperatures. Furthermore the stability of the fibrillar assemblies was probed as a function of the temperature, showing strong deviations from a monotonic behaviour for α-Synuclein, Aβ and β2 microglobulin [159].

The use of temperature can be further exploited to model both the thermodynamic and kinetic behaviour of the reaction once the equilibrium conditions are defined. Here I expand on the current knowledge of the temperature dependence of the amyloid elongation reaction, aiming to extract both kinetic and thermodynamic information through a differential temperature dependence of the two opposing reactions. Differently from previous experimental approaches, major

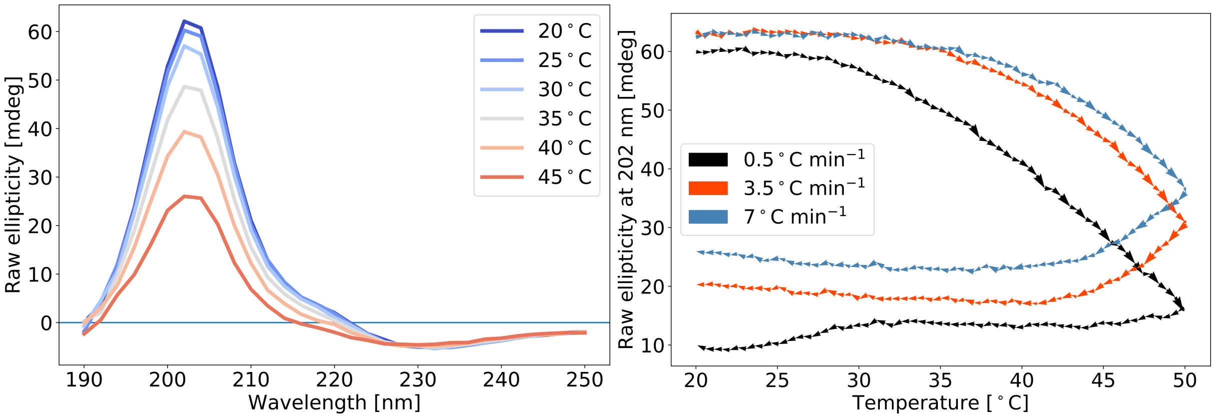

Figure 3.1: Temperature dependence of glucagon fibril monitored by CD spectroscopy. The sonicated fibrillar sample (25 µM) was measured every 1°C, with a time of equilibration of 15 minutes at each temperature step.

emphasis was placed on the use of the steady state of elongation and towards the dependence of the depolymerization on the temperature ramping speed. In order to achieve this, the temperature dependence of amyloid systems composed by glucagon and PI3K-SH3 amyloid fibrils was established.

Focusing on the time necessary for the system to equilibrate, I measured the intrinsic fluorescence of glucagon in various concentrations of GdnCl while ramping the temperature up and down at different rates. The stability in the range of temperatures between 20 and 60°C was extrapolated to the absence of denaturant, enabled by the long time of equilibration that allows us to explore both kinetic and thermodynamic parameters in one single experiment while maintaining the fibrils out of equilibrium.

The steady state of the fibril elongation was instead used as a means to extract thermodynamic informations by quartz crystal micro-balance (QCM) measurements. The rate of elongation of PI3K-SH3 fibrils was studied between 20 and 55°C. The observation of a biphasic temperature dependence allows us to investigate the differential role of the two opposite reactions and to obtain thermodynamic parameters for the whole temperature range.

These new methodologies open up the possibility to investigate the thermodynamic of the elongation reaction by exploiting temperature and kinetics, the most well established aspects of amyloid growth. The models proved to be robust and, even if more controls are required, they could be applied to different amyloid systems to probe their stability.

3.2. Results

3.2.1 Qualitative evaluation of the temperature dependence of fibril stability

To determine the effect of temperature on the elongation reaction of glucagon amyloid fibrils, the thermal denaturation of glucagon fibrils was first monitored acquiring circular dichroism (CD) spectra of sonicated fibril samples. To allow the system enough time to equilibrate, the temperature was ramped by 1°C every 15 minutes (Fig. 3.1, left panel). The spectrum at 20°C shows the characteristic β-turn curve, with a peak at 205 nm and a shoulder around 220 nm, probably due to the presence of a tryptophane in the sequence. The signal intensity of the

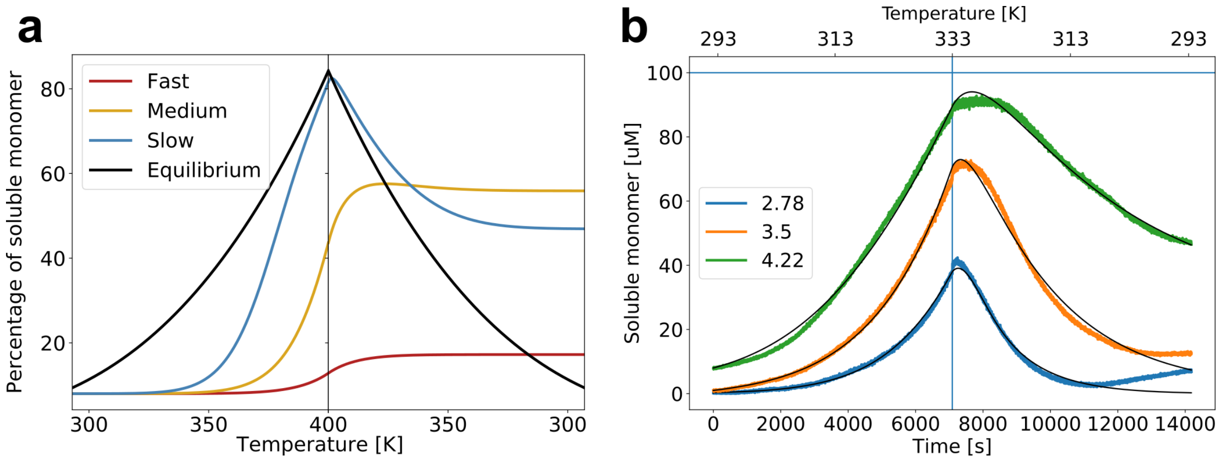

Figure 3.2: Non-equilibrium behaviour of a polymer temperature ramp. **a**) Simulation of the behaviour of a supramolecular polymeric system in a temperature ramping experiment. In a scenario where the polymer depolymerize as the temperature increases, the faster the ramping occurs, the less monomer is able to depolymerize in time. **b**) Temperature ramping of glucagon amyloid fibrils in different concentrations of denaturant at a speed of 0.1°C min^{-1}. The black lines show the best fit to the data. The legend shows the samples GdnCl concentrations. The blue vertical line highlights the time when the highest temperature was reached (90°C), while the horizontal blue line shows the total monomer concentration of the samples.

peak at 202 nm decreases as the temperature increases, highlighting the loss of the amyloid conformation.

The high stability of the amyloid fold is also due to the low rate of depolymerization, therefore I explored the time of equilibration as a variable to test the thermal denaturation profile of the glucagon fibrils. Different samples of fibrils with equimolar concentrations were ramped at different rate and sequentially cooled down with the same rate (Fig. 3.1, right panel). Upon heating to 50°C and cooling back to 20°C with different heating speed (from 0.5 to 7 °C min^{-1}), the degree of signal intensity that is lost correlates inversely with the rate of heating/cooling. As the heating/cooling rate increases, less time is given to the fibrils to equilibrate at higher temperatures. As the increased temperatures induces depolymerization of the fibrils, less time at higher temperatures results in a lower amount of soluble monomer present in solution.

By cooling back the solution the ellipticity continues to decrease, meaning that the fibrils continue to depolymerize to reach the equilibrium. The signal stabilizes itself when, during the cooling phase, the temperature corresponds to the one at which the actual soluble monomer concentration is the equilibrium one. The phenomenon is much more prominent as the heating/cooling rate increases, highlighting the fact that the system requires more time to reach the equilibrium at every temperature step then is available at high ramping rates.

3.2.2 Quantitative characterization of glucagon thermodynamics in out-of-equilibrium regime

As the systems "lags behind" while ramping the temperature, it is possible to study it outside of equilibrium. In order to do so, I analysed the behaviour of the fibrils in the presence of denaturant, in order to destabilize them and favour the depolymerization reaction. Glucagon fibril samples were diluted in different GdnCl and probed through intrinsic fluorescence between 20 and 60°C using the slowest heating/cooling speed possible, 0.1°C min^{-1}. The samples were

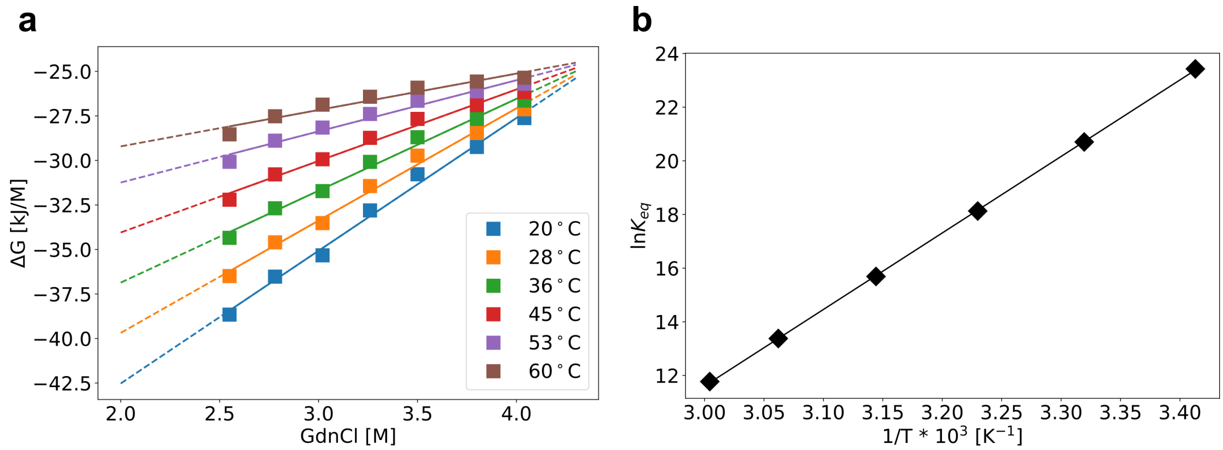

Figure 3.3: Temperature dependence of glucagon fibrils stability. **a**) The values of ΔG obtained from the fitting of the data presented in Fig. 3.2b are extrapolated to the absence of denaturant for different temperatures. **b**) Van't Hoff plot with linear fit to measure the enthalpy contribution to the amyloid state stability (ΔH = -237.1 kJ mol^{-1}; ΔS = -0.6 kJ mol^{-1} K^{-1}).

measured both while ramping the temperature up and down, so to probe the apparent critical concentration of monomer while maintaining the system out of equilibrium. If during the heating phase the increase in temperature happens in time intervals exceeding the equilibration time, the system will have enough time to reach the equilibrium. If the time intervals of heating are inferior to the equilibration time, the system will lag behind. The same reasoning applies to the cooling phase of the system.

In order to consider the reaction from a quantitative point of view, the ratio of fluorescence between 350 and 330 nm was translated to the amount of soluble monomer as described in the Method section. All the fibrillar samples reach their maximum of soluble monomer in solution during the cooling phase, hinting toward a non-equilibrium behaviour (Fig. 3.2b). As the temperature increases, the fibrils cannot depolymerize fast enough and reach the peak of soluble monomer during the cooling phase. As the equilibrium concentration of free monomer still has to be reached, the fibrils keep on depolymerizing as long as the soluble monomer concentration is lower than the one required by the actual temperature.

The fibrils' behaviour can be described as the result of the balance of the elongation and depolymerization reactions (black lines in Fig. 3.2b). The fit of the two rate constants as a function of their enthalpic and entropic contributions allows us to describe the behaviour of every sample in different concentrations of denaturant. It is then possible to extrapolate the stability of the fibrillar system in absence of denaturant by a linear *a posteriori* fitting (Fig. 3.3a). The stability, that can be approximated by the ratio of the elongation rate constant over the depolymerization rate constant, can be reliably measured for the denaturant concentrations which show a significant change of soluble monomer concentration in the temperature range here studied.

The temperature dependence of the fibril stability in the absence of denaturant can be visualized in a Van't Hoff plot (Fig. 3.3b). The stability dependence on the inverse of the temperature confirms the exothermic signature of the elongation reaction. The stability decreases as temperature increases, showing once again that the amount of soluble monomer increases with the temperature. The resulting enthalpy of the reaction, ΔH, is -237 kJ mol^{-1}, highlighting the drastic dependence

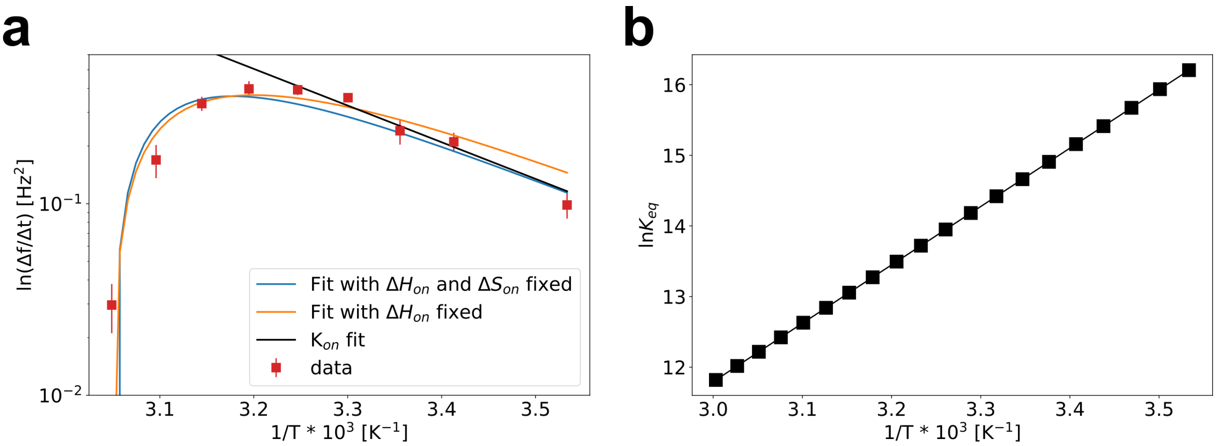

Figure 3.4: Temperature dependence of PI3K-SH3 elongation reaction. **a**) Thermal dependence of elongation rates of PI3K-SH3. The error bars represent the standard deviation on a triplicate measurement. **b**) Van't Hoff plot of the stability of PI3K-SH3 elongation reaction ((ΔH = -68.7 kJ mol^{-1}; ΔS = -0.1 kJ mol^{-1} K^{-1})). The values are obtained from the computation of the stability parameter of a finite number of temperatures obtained from the fitting described with the blue line in the panel a .

of the stability on the temperature in this range and the enthalpic drive of the amyloid elongation reaction.

3.2.3 Temperature dependence of PI3K-SH3 elongation kinetics

The rate of elongations of PI3K-SH3 fibrils was measured at different temperatures through QCM-D measurements. The measured elongation rates are modelled as the result of the competition between the forward and the backward reaction: the elongation itself and the depolymerization. Both rate constants are described as a function of the temperature, as shown in the Methods section.

First the dependence of the elongation on temperature was probed at a single concentration, namely 5 µM, in a range of temperatures between 10 and 55°C (Fig. 3.4a). The reaction rate exponentially increases with temperature until it reaches a maximum at ca. 40°C. The rate than rapidly decreases, with the fibrils almost not polymerizing any longer at 55°C.

The result of the net rate modelled as competition between the elongation and depolymerization reaction is shown in Figure 3.4a. The two opposite reactions are defined by different temperature dependencies, to depict the drastic increase of the depolymerization reaction with temperature compared to the elongation. The rate of elongation ($k_{+}[M]$) is obtained through fitting the four data points at the lowest temperatures (from 10 to 30°C), where the contribution of the depolymerization reaction is negligible (black line in Fig. 3.4). This first estimate of the elongation rate constant is then used to fit the rest of the dataset including the depolymerization reaction. The fitting is done both by keeping the enthalpy parameter ($\Delta H^{\ddagger}_{+}$) fixed alone or together with the entropy parameter ($\Delta S^{\ddagger}_{+}$).

The difference of temperature dependence between the two opposite reaction is determined by the enthalpic contribution to the barrier. The elongation $\Delta H^{\ddagger}_{+}$ is less than half of the depolimerization $\Delta H^{\ddagger}_{-}$ for both fitting approaches described before: the enthalpic barrier of elongation is 36.7 kJ mol^{-1}, while the enthalpic contribution of the backward reaction is 100.5 kJ mol^{-1} (in the case of both the $\Delta H^{\ddagger}_{+}$ and $\Delta S^{\ddagger}_{+}$ being fixed parameters) or 78.5 kJ mol^{-1} (if $\Delta S^{\ddagger}_{+}$ is an open parameter).

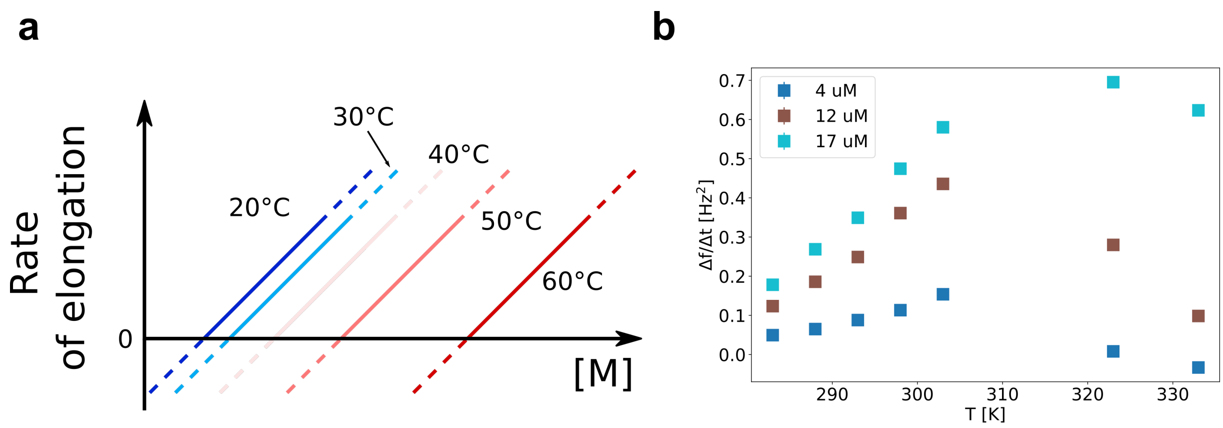

Figure 3.5: Concentration dependence of the elongation reaction in the temperature dimension. **a**) Schematic representation of **b**) Three different concentration of monomer are used to elongate the PI3K-SH3 fibrils attached on the sensor. The rates of elongation are normalized and compared by using a reference injection on all the sensors at the temperature of 20°C.

This finding is coherent with the fact that the contribution of the depolymerization reaction becomes significant only above 35°C. The resulting effect on the stability, which is proportional to the natural logarithm of the ratio of the constants, is an exothermic signature for the reaction for temperatures above 10°C (Fig. 3.4b). The enthalpy difference between the the soluble and the amyloid state can be extracted and it results to be ca. -68 kJ mol^{-1}.

To further demonstrate that the rate of elongation corresponds to the balance between the elongation and depolymerization rate, the temperature dependence of the reaction was measured using three different concentrations of monomer. The steady state of elongation was measured for temperatures ranging between 10 and 60°C with concentrations of 4, 12 and 17 µM (Fig. 3.5). While the elongation reaction is dependent on both the concentrations of fibril ends and the soluble monomer ($k_+ [P] [M]$), the depolymerization reaction has only a dependence on the fibril end concentration ($k_- [P]$). For both elongation and depolymerization, the fibril ends are the same, hence the apparent rate of elongation is defined by how many events of elongation happen compared to the events of depolymerization. As the temperature increases, the depolymerization process becomes intrinsically dominant, but the extent to which it influences the apparent rate is defined by the concentration of soluble monomer, which drives the events of elongation. The data reproduce this behaviour: the highest concentration (17 µM) is still elongating at high rates at 60°C, while the lowest concentration (4 µM) depolymerizes, having a net rate of "elongation" lower than zero.

3.3. Discussion

3.3.1 Characterization of the temperature-dependence of the glucagon amyloid fibrils elongation reaction

The response of glucagon fibrils to changes in temperature has been extensively studied by Otzen and collaborators [45, 95]. A wide range of responses to temperature changes was found for fibrillar samples prepared in different ways. This has been attributed to the high polymorphism of the system, which seems to be a prominent characteristic of glucagon fibrils. In the work here presented, the focus has been on the response of the fibrils on changes in temperature as a function of time, using fibrils named in the literature as "Type B" [95]. These fibrils show a β-turn CD spectrum, which with its peak around 202 nm closely resembles our samples (Fig. 3.1). These fibrils, formed in acidic environment with low salt and at room temperature, were shown to be mildly resistant to heat, which is consistent with the findings here presented.

The dependence of the depolymerization of these fibrils on the heating/cooling speed opened a new dimension to study amyloid fibrils. The study of aromatic polymers in supramolecular chemistry is widely based on this methodology, relying generally on the temperature dependence of the equilibrium to describe the polymerization mechanisms [24, 66, 160]. Here the temperature ramping experiment with glucagon fibrils showed how the elongation reaction can be studied by maintaining the amyloid system out of equilibrium. By ramping up the temperature, we favour the depolymerization of the system, due to the exothermic signature of the elongation reaction [45]. While the temperature rises, the system acquires more energy and slowly approaches the equilibrium state. In this experiment, the temperature scan is stopped once the temperature reaches 60°C, after which the cooling down phase begins. This causes the system composition to "catch up" with the temperature and eventually, at a given temperature, equilibrium will be reached. The further cooling slows down the reaction, causing the system to lag once again behind due to the decrease of the reaction kinetics.

As the temperature decreases, the fibrillar system reaches equilibrium at a given time point t, as the equilibrium soluble mass at that temperature is present in solution. The soluble mass is insufficient for the time point before, t-dt, and too abundant for the time point afterwards, t+dt. In Figure 3.2a the time point at which the expected soluble mass correspond to the measured soluble mass, the peak in concentration of measured soluble mass is observed.

Every curve, corresponding to a single concentration of denaturant, can be described by a set of two different rate constants, elongation (k_+) and depolymerization (k_-), both dependent on the temperature. From the ratio of the elongation rate constant over the depolymerization rate constant, an approximation of the equilibrium constant can be obtained, $K_e = k_+/k_-$. This is valid if there is no significant loss of fibril ends during the process, which is the reason why only the samples with at least 10% of fibrillar mass remaining at the maximal depolymerization are considered for the following analysis. In Figure 3.3a, the relation between $\Delta G = -RT \ln K_e$ and the denaturant concentration is shown at different temperatures. A linear fit describes the trend of the relation and extrapolates the fibrillar stability to the absence of denaturant. This kind of analysis does not imply *a priori* the linearity of the relation $\Delta G = f([D])$, otherwise implied in the non-linear fitting of the data in the equilibrium denaturation framework. If the result obtained with this method at 30°C is compared to the result obtained in the previous chapter through equilibrium denaturation, a difference can be easily spotted. The stability of glucagon fibrils obtained through chemical depolymerization obtained with the cooperative model is around 36-38 kJ mol^{-1}, lower than the value obtained with the non-equilibrium methodology, -50 kJ mol^{-1}. As in the non-equilibrium methodology the denaturant used is GdnCl, a better comparison should be drawn with the equilibrium data obtained at high salt concentration in the presence of urea. The stability of the fibrils in high ionic strength buffer (120 mM NaCl) increases up to -43 kJ mol^{-1}, approaching the value here obtained. As the Debye plot (Fig. 2.4) does not show any strong bending at higher salt concentration, a higher value for the stability in presence of complete screening could be expected, a situation which resembles more closely the high concentration of GdnCl used here for screening. Moreover, the modelling here employed approximates the elongation behaviour as the ratio of rates, that refers only to the growth of full length fibrils, while the data obtained through equilibrium depolymerization at a fixed temperature models the entire system as defined by two different equilibrium constants for nucleation and growth.

Previous experiments showed a non-Arrhenius behaviour for the fibril formation reaction [45], in an effort to characterize the entire fibrillation process, nucleation included. The elongation reaction alone was analysed by Buell *et al.*, who found a weak endothermic signature for the elongation reaction barrier of activation for temperatures inferior to 30°C. Both publications agree in describing a maximum rate for the amyloid growth in the range between 20 and 30°C. In this work, I probed the effect of temperature on the elongation reaction to measure the stability of the amyloid fold. In the temperature range probed by this experiment (20 to 60°C), the reaction signature is strongly exothermic (Fig. 3.3b). This is in contrast to previously published data that measured a enthalpy value one order of magnitude lower and a variation of the same with temperature [161]. In the work presented in this thesis a probable overstatement of the stability in absence of denaturant at lower temperature can be the cause of the magnitude of the enthalpic signature of the reaction. On the other hand, the model employed in the paper from Jeppesen *et al.* assumes that the enthalpic signature measured by calormietry arises from a complete conversion of the injected monomer to the amyloid fold, which may not be the case at higher temperatures. Still the difference is evident and it cannot be addressed with the available data.

The thermal stability of the fibrillar system is instead in agreement with the same data presented in the literature, since the fibrillar preparation here used resembles closely the Type B fibrils, which show a melting temperature of 55° in absence of salt. The presence of high quantities of salt (GdnCl) could further stabilize the fibrils, leading to increased melting temperature. Regarding the difference with the endothermic signature of the Arrhenius plot of the kinetic reaction, the increase of the depolymerization rate with the temperature could create an overall exothermic signature for the stability, while the elongation rate still increases with the temperature.

For a reliable use of the model here described, some additional factors have to be addressed. While primary nucleation events are improbable, due to high energetic barrier for their formation and to the constant presence of seeds in solutions, secondary nucleation has to be ruled out. Similarly, the effect of the denaturant could influence the secondary nucleation events during the cooling down phase, an accurate description of the latter being fundamental for the meaningful fitting of the dataset. In the end, the presence of a constant amount and type of fibrils, modelled as $[P]$, should be confirmed by a more in-depth study of the fibril length homogeneity.

A final comparison should be drawn with protein folding. Protein folding, being an intra-molecular reaction, does not rely on pre-existent native proteins to fold (in an *in vitro* environment). The intra-molecular nature of the reaction allows every monomer to exchange conformation between the unfolded and the folded state at every temperature, without the need of interactions with other molecules. Amyloid fibrils are polymer and are composed by many monomers involved in inter-molecular interactions. The different monomers that compose the fibril can change conformation on the basis of their position in the polymer. Depolymerization, even if favoured, occurs only at the fibril ends and therefore the equilibrium itself does not only depend on the molecular contacts, but also on the absolute concentration of the protein.

3.3.2 Temperature dependence of PI3K-SH3 elongation kinetics

The effect of temperature on the elongation reaction of amyloid fibrils has been widely documented [33, 38, 41–44, 46–51, 130]. Increased temperature seems to favour the elongation reaction for many different amyloid systems, with a monotonic exponential increase of the elongation rate up to temperatures in the range of 40-50°C (up to 70°C for Aβ fibrils in low pH [44]). The results shown here are in agreement with previous findings related to PI3K-SH3 [49], but extend the analysis further to include also higher temperatures. The information contained in

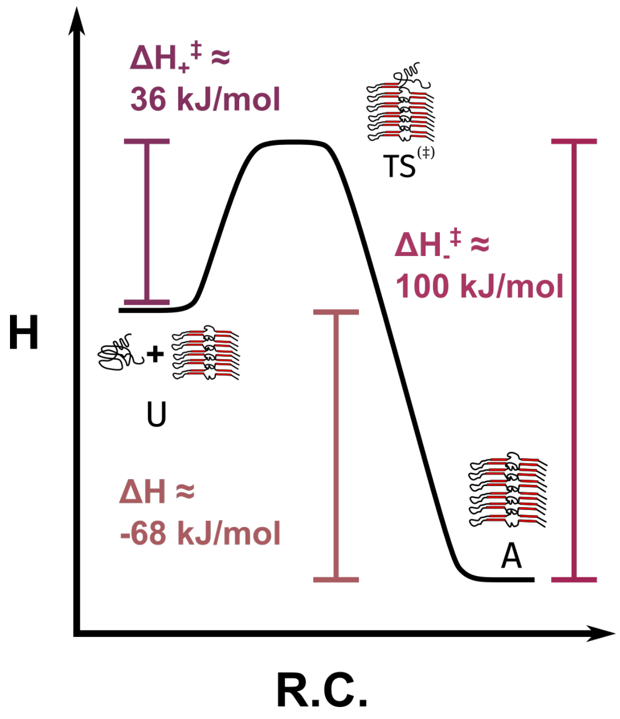

Figure 3.6: Enthalpy landscape of the PI3K-SH3 elongation reaction. The enthalpy barrier of elongation ($\Delta H_{+}^{\ddagger}$) and depolymerization ($\Delta H_{+}^{\ddagger}$) agree with the enthalpy difference between the unfolded and amyloid state (ΔH).

this higher temperature regime is fundamental to describe the complete thermal dependence of the reaction, both in its forward and backward components.

The $\Delta H_{+}^{\ddagger}$ parameter, obtained through the analysis of rates in the first four temperatures probed in this experiment, 36.7 kJ mol^{-1}, agrees with the value proposed by Buell *et al*, 42.1±8.5 kJ mol^{-1}. The further apparent deviation from a standard non-Arrhenius behaviour is here analysed in the framework of two opposite Arrhenius-like reactions. The different dependence of the two reactions, represented by the $\Delta H_{+}^{\ddagger}$ and $\Delta H_{-}^{\ddagger}$ parameter, leads to a maximum rate of elongation around 35-40°C. At lower temperature, the high stability of the amyloid fold is due to the negligible depolymerization reaction. The van't Hoff plot describes the stabiliy over the whole temperature range observed and allows us to describe the enthalpic differences between the soluble, amyloid and transition state as depicted in Fig. 3.6.

The high stability of the amyloid state allows us to probe the elongation reaction in the presence of very low soluble monomer concentrations, since the critical concentration of monomer is very low. The critical concentration of monomer is given by the ratio of the two rate constants, which is also the inverse of the equilibrium constant. The relation between the natural logarithm of the equilibrium constant and the inverse of the temperature is described in the van't Hoff plot in Fig. 3.4b. The similar behaviour of the two rate constants results in an exothermic signature for the whole temperature range. As in the case of glucagon, the bi-phasic kinetic behaviour is translated into a monotonic decrease of the stability over temperature. Moreover, the strong negative enthalpy contribution underlines the enthalpic drive towards the acquisition of the amyloid fold.

The critical concentration extracted from this fit for the maximum temperature, 60°, corresponds to ca. 7μM. This result is in excellent agreement with experiments performed with multiple monomer concentrations. By testing multiple concentrations it is possible to rule out other reactions, such as a structural rearrangement of fibrils or an effective lack of Arrhenius-like behaviour for the elongation reaction. In the low temperature regime, the linearity is maintained for all the concentrations, mirroring the behaviour that is expected from the relation $k_{+}\,[M]\,[P]$. At the two highest temperatures the behaviour changes radically, in agreement with the high $\Delta H^{\ddagger}_{-}$ of the depolymerization rate constant. As expected from the previous analysis, at 60°C the lowest concentration of protein does not elongate the fibrils, as it is inferior to the critical concentration. The medium and high concentration instead still elongate, even though without the same proportionality to concentration as at lower temperatures. The shift of the critical concentration is represented in the Fig. 3.5b, where the critical concentration is represented by the concentration of protein which crosses the $y = 0$ axis.

In the case of PI3K-SH3, as well as glucaon, there is a discrepancy between the stability parameters obtained though this methodology and the stability parameters obtained through chemical depolymerization. The stability of PI3K-SH3 probed at 30°C through chemical depolymerization in presence of 16 mM NaCl is ca. 56 kJ mol^{-1}, while the stability measured through this kinetic approach is ca. 37 kJ mol^{-1}, in absence of NaCl. As the charge of PI3K-SH3 in pH 2 is +12, the dependence of the stability on charge screening is high. By extrapolating the stability measured through chemical depolymerization in absence of NaCl, the stability goes down to ca. 47 kJ mol^{-1}. The difference between the two values is of ca. 25%, which is reasonable given the fact that two very different techniques are employed. The chemical denaturation approach aims at the description of the ensemble of species in solution, while the kinetic approach here presented focuses on the balance between the elongation and depolymerization reaction. Both approaches exploit two different dimensions to extrapolate the fibrillar stability: the denaturant concentration, for the chemical depolymerization methodology, and the temperature dimension, for the kinetic approach. Furthermore, other sources of error for this methodology can contribute to the observed difference. The normalization of the rates on a single temperature relies heavily on the accuracy of the measurements at that temperature. Moreover, the density of the measurements can be increased at higher tempeatures, leading to a higher resolution where the competition between the two opposite reactions becomes more prominent. In the end, an error of ca. 20% on the concentration measure of the monomer solution can be expected at this low concentrations. This translates in ca. 5% of error on the stability measure.

3.4. Conclusions

Through the study of the dynamic behaviour of the two fibrillar systems, glucagon and PI3K-SH3, I was able to describe the dependence of the kinetics on variations in the temperature dimension and obtain the thermodynamic parameters of the elongation reaction. The amount of information that can be drawn from this kind of analysis exploits both the thermal stability and the seeding potential of amyloid fibrils.

The non-equilibrium methodology employed with glucagon spans both the temperature and the denaturant dimension. Describing the full temperature dependence of the amyloid state is possible by exploiting the slow kinetics of the system and then extrapolating the results back to the absence of denaturant. This approach is particularly interesting when it comes to highlight the difference in behaviour between the classical folding and the amyloid elongation reaction.

The temperature dependence of the steady-state of elongation is a powerful technique to describe the behaviour of the elongation reaction over different concentrations, while characterizing the full thermal behaviour of both elongation and depolymerization rate constants. Applying it to the PI3K-SH3 amyloid elongation reaction led to both the extraction of the thermal dependence of the kinetic constants and to the description of the concentration criticality over a wide range of temperatures.

To summarize, measuring the kinetics of the elongation reactions over a wide range of temperatures provides information to describe the stability of the fibrillar system, while avoiding the use of denaturant. While further controls on the effective measurement of the isolated elongation reaction are required to apply this methodology more rigorously and to other fibrillar system, the possibility to dissect the different rate constants and to measure the thermodynamic parameters make this methodology a promising tool to characterize amyloid stability.

3.5. Methods

3.5.1 Fibril preparation

Glucagon and PI3K-SH3 fibrils were prepared accordingly to the protocol described in Section 2.5.2.

3.5.2 CD spectroscopy

The glucagon fibril spectra and kinetics were measured on an AVIV spectrophotomer (Aviv Biomedical, Lakewood, NJ, USA) using as parameters a bandwidth of 1 nm, averaging time of 1 second and three repetitions for each measurement. The measurements were all performed using a total protein concentration of 25 µM in 10 mM GlyHCl pH 2.5, with a 1 mm cuvette. The temperature ramping speed depends on the experiment and is specified in the result section.

3.5.3 Fluorescence measurement

The samples measured by fluorescence spectroscopy were composed of glucagon fibrils with a total protein concentration of 100 µM in 10 mM GlyHCl pH 2.5 buffer and different concentrations of GdnCl ranging from 1.4 M to 5.4 M. The fluorescence measurements were carried out with a Prometheus NT.Plex instrument (NanoTemper Technologies, München, Germany). The different samples were loaded on NT.Plex High Sensitivity capillaries (NanoTemper Technologies, München, Germany) and sealed with wax. The samples fluorescence was measured by excitation at 280 nm and recording at 330 and 350 nm. The temperature range was between 20°C and 60°C using a heating/cooling speed of 0.1°C min^{-1}.

3.5.4 Non-equilibrium temperature ramp kinetics

To estimate the fraction of soluble monomer, the ratio of the fluorescence intensity of 350/330 nm was converted to a fraction of soluble monomer by normalizing the signal between 0 and 1, using the highest and lowest molarity of GdnCl as references. To analyse the data, the fraction of soluble monomer was converted to concentration of soluble monomer and then fitted with the model described below.

As the system is not able to reach the equilibrium in the time of the experiment, an out-of-equilibrium framework was used to describe and fit the data. At every time point, the change of soluble mass, $d[M]/dt$, is described as:

$$\frac{d[M]}{dt} = [P]\,k_-(t) - [P]\,k_+(t)\,[M]$$

where $k_{-}(t)$ and $k_{+}(t)$ are described as a function of temperature according to the following relation:

$$k = \Gamma \exp - \frac{\Delta H^{\ddagger} - T(t)\Delta S^{\ddagger}}{RT(t)}$$

where $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are, respectively, the enthalpy and entropy of activation, R the gas constant and $T(t)$ the temperature as a function of time. The time dependence of the temperature is desribed as it follows:

$$T(t) = \begin{cases} T_0 + wt, & \text{while ramping up} \\ T_0 - wt, & \text{while ramping down} \end{cases} \tag{3.1}$$

where w is the ramping speed (0.00167°C s^{-1}), T_0 the starting temperature and t the time.

The relation between the time and the temperature is the key to define the non-equilibrium behaviour of the system, that acts according to $T(t)$ while it still has to reach the equilibrium of the temperature $T(t - dt)$. Both elongation and depolymerization events happen accordingly to the rate constants of $T(t)$, which define the stability as follows.

The stability of the fibrillar system is by the equilibrium definition, for which $d[M]/dt = 0$, stated as follows:

$$0 = [P]k_{-} - [P]k_{+}[M]$$
$$K_e = \frac{k_{-}}{k_{+}} = \frac{1}{[M]}$$
$$\Delta G_0 = -RT \ln K_e$$

Every concentration of denaturant has a stability parameter that can be measured at every temperature according to the determined values of k_{+} and k_{-} previously defined. The stability in absence of denaturant is then obtained for every temperature by linear extrapolation of the denaturant concentration range between 2.5 and 4 M GdnCl.

3.5.5 Temperature dependence of elongation kinetics

The dependence of the elongation reaction on temperature was probed by QCM. Different sensors were incubated with 5 µM PI3K-SH3 solution at different temperatures ranging from 10 to 70°C. The number of different injections that can be made on a single sensor is limited due to a maximal fibril length that can be sensed at constant sensitivity. For this reason, the different rates were compared to an injection made at the same temperature on all the sensors. The rates were then normalized with respect to this injection and the standard deviation was measured. Every temperature was measured at least in triplicate and its standard deviation calculated from that. The experiment was performed by Alberto Coden [162].

To analyze the data, the steady state of elongation was described as the combination of the elongation and depolymerization reaction:

$$r_{app} = k_{+}[M][P] - k_{-}[P] \tag{3.2}$$
$$r_{app} = k_{+}^{app}[M] - k_{-}^{app} \tag{3.3}$$

where r_{app} is the apparent elongation rate, k_{+} the rate constant of elongation, k_{-} the rate constant of depolymerization, $[M]$ the concentration of the injected monomer and $[P]$ the concentration of fibril ends. As $[P]$ is the same for both terms, it has been included into the apparent rate constants k_{+}^{app} and k_{-}^{app}. The temperature dependence of both k_{+}^{app} and k_{-}^{app} is described as it follows:

$$k^{app} = [P]\,\Gamma \exp\left(-\frac{\Delta H^{\ddagger} - T\Delta S^{\ddagger}}{RT}\right) \tag{3.4}$$
$$k^{app} = \Gamma^{app} \exp\left(\frac{-\Delta H^{\ddagger}}{RT}\right) \exp\left(\frac{T\Delta S^{\ddagger}}{RT}\right) \tag{3.5}$$

where Γ is the kinetic prefactor, $\Delta H^{\ddagger}$ the enthalpy of the reaction, $\Delta S^{\ddagger}$ the entropy of the reaction, R the gas constant and T the temperature. The term $\Gamma^{app} = \Gamma\,[P]$ includes the uncertainty on both the prefactor and the fibril end concentration. For simplicity, since the focus is on the determination of the enthalpic contribution of the reaction, the term Γ^{app} is considered to be identical in both the forward and backward reaction and its temperature dependence can be neglected.

The fitting was performed using $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ as free parameters. The rate constant k_{+}^{app} was fitted only in the low temperature regime, where the increase of the logarithm of the rate is linear with respect to the inverse of the temperature and where the contribution from k_{-}^{app} is negligible. Then the fitting on the whole dataset has been performed using the equation 3.3, in one case leaving the ΔS_{+} as an open parameter, while in the other case fixing it to the value found in the linear fit.

4

ATOMIC STRUCTURE OF PI3-KINASE SH3 AMYLOID FIBRILS BY CRYO-ELECTRON MICROSCOPY

Article information

Title: Atomic Structure of PI3-Kinase SH3 Amyloid Fibrils by Cryo-Electron Microscopy

Authors: Christine Röder[1,2], Nicola Vettore[2], Lena N. Mangels[1,2], Lothar Gremer[1,2], Raimond B. G. Ravelli[3], Dieter Willbold[1,2], Wolfgang Hoyer[1,2], Alexander K. Buell[2,4] and Gunnar F. Schröder[1,2,5]

1) Institute of Complex Systems, Structural Biochemistry (ICS-6) and JuStruct, Jülich Center for Structural Biology, Forschungszentrum Jülich, 52425, Jülich, Germany

2) Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225, Düsseldorf, Germany

3) The Maastricht Multimodal Molecular Imaging Institute, Maastricht University, Universiteitssingel 50, 6229 ER, Maastricht, The Netherlands

4) Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, 2800 Kgs, Lyngby, Denmark

5) Physics Department, Heinrich-Heine-Universität Düsseldorf, 40225, Düsseldorf, Germany

Journal: Nature Communications

DOI: 10.1038/s41467-019-11320-8

Abstract

High resolution structural information on amyloid fibrils is crucial for the understanding of their formation mechanisms and for the rational design of amyloid inhibitors in the context of protein misfolding diseases. The Src-homology 3 domain of phosphatidyl-inositol-3-kinase (PI3K-SH3) is a model amyloid system that plays a pivotal role in our basic understanding of protein misfolding and aggregation. Here, we present the atomic model of the PI3K-SH3 amyloid fibril with a resolution determined to 3.4 Å by cryo-electron microscopy (cryo-EM).

The fibril is composed of two intertwined protofilaments that create an interface spanning 13 residues from each monomer. The model comprises residues 1–77 out of 86 amino acids in total, with the missing residues located in the highly flexible C-terminus. The fibril structure allows us to rationalise the effects of chemically conservative point mutations as well as of the previously reported sequence perturbations on PI3K-SH3 fibril formation and growth.

4.1. Introduction

Alzheimer's and Parkinson's Disease as well as spongiform encephalopathies are prominent examples of protein misfolding diseases [30]. These disorders are characterised by the presence of amyloid fibrils [163]. Amyloid fibrils are straight and unbranched thread-like homopolymeric protein assemblies, which are stabilised by backbone hydrogen bonding between individual peptide molecules. These interactions lead to a highly ordered, repetitive cross-β architecture, in which the β-strands run perpendicularly to the fibril axis.

It has been shown that in the case of neurodegenerative protein misfolding diseases, the final amyloid fibrils are often not the most cytotoxic species, but that small, oligomeric precursors are more hydrophobic and more mobile and hence more prone to deleterious interactions with cellular components [164]. However, recent progress in the mechanistic understanding of amyloid fibril formation shows that the mature fibrils can be the main source of toxic oligomers, due to their role as catalytic sites in secondary nucleation processes [165]. Furthermore, in the case of systemic amyloidosis diseases, where amyloid fibrils form in organs other than the brain, the amyloid fibrils themselves are the deleterious species, as their presence in large quantities can disrupt organ functions [166].

Until recently, structural information on amyloid fibrils could only be obtained from relatively low-resolution methods, such as X-ray fibre diffraction [123], limited proteolysis [167] and H/D exchange [168]. High-resolution structural information on amyloid fibrils has only become available in recent years through progress in solid state NMR spectroscopy (ssNMR) [169–171] and cryo-electron microscopy (cryo-EM). In particular cryo-EM enables atomic resolution structures of amyloid fibrils to be determined, and this has indeed been achieved in a few cases so far [29, 148, 149, 172–180]. Such detailed structural information is crucial for the understanding of amyloid formation mechanisms, as well as for the rational design of inhibitors of the individual mechanistic steps, such as fibril nucleation and growth [181].

Here we present the high-resolution cryo-EM structure of amyloid fibrils of the Src-homology 3 domain of phosphatidyl-inositol-3-kinase (PI3K-SH3). SH3 domains are kinase subdomains of usually less than 100 amino acids length and have been found to be part of more than 350 proteins, ranging from kinases and GTPases to adaptor and structural proteins, within various organisms [182]. SH3 domains are known to play a significant role in several signalling pathways where they mediate protein-protein interactions by recognising PxxP sequence motifs [102, 183]. The structure of natively folded PI3K-SH3, a domain consisting of 86 amino acids from bovine PI3K, has been well-characterised by X-ray crystallography and NMR spectroscopy [100, 183].

Used initially as a model system for protein folding studies [103], PI3K-SH3 was among the first proteins discovered to form amyloid fibrils in the test tube, while not being associated to any known human disease [105]. Fibril formation was observed at acidic pH, where in contrast to the native fold at neutral pH [100, 102], monomeric PI3K-SH3 lacks a well-defined secondary structure [39, 105, 106]. Since this discovery, PI3K-SH3 has played a pivotal role in advancing our fundamental understanding of the relationships between protein folding, misfolding and aggregation. Indeed, the hypothesis of the amyloid fibril as the most generic ‘fold’ [30] that a polypeptide can adopt was significantly shaped by the finding that PI3K-SH3 forms amyloid fibrils. Many pioneering studies on the basic biochemical, structural and mechanistic features of amyloid fibrils have been performed with PI3K-SH3. Early cryo-EM measurements highlighted the need for conformational rearrangement of the sequence within the fibril [184]. The important role of the destabilisation of native secondary structure elements and the need for non-native contacts and extensive structural rearrangements during the formation of fibrillar aggregates was also observed for a related SH3 domain [185]. Despite not being related to any human disease, PI3K-SH3 aggregates were shown to be cytotoxic, suggesting sequence-independent toxic properties of amyloid fibrils and their precursors [186]. PI3K-SH3 also provided insight into the kinetics of molecular recycling of the monomeric building blocks of the fibril [187], as well as into the dynamics of the formation of oligomeric precursors of amyloid fibrils [188].

The structure of PI3K-SH3 fibrils we present here is in agreement with previous ssNMR data [189], but we find that its interface is orthogonal to that suggested previously based on a low resolution reconstruction [184]. Indeed, the inter-filament interface in PI3K-SH3 fibrils is large compared to that of other amyloid fibrils determined to-date, and is formed from residues distant in primary sequence. With the atomic model we can rationalise the effect of newly designed, as well as previously reported sequence variants of PI3K-SH3 on the kinetics of fibril growth. Our study therefore not only adds important insight into the structural variety of amyloid fibrils, but also demonstrates how such structures can be used to rationalise the dynamics of protein assembly processes.

4.2. Results

4.2.1 Structure Determination by Cryo-EM

Fibril formation by the full-length PI3K-SH3 domain under acidic solution conditions [105] led to long, straight fibrils of which the main population could be structurally determined by cryo-EM (Fig. 4.1). High overall homogeneity of the preparation has been shown by atomic force microscopy (AFM) and negative staining EM imaging (Fig. 4.2, Supplementary Figs. 4.7 and 4.8). Nevertheless, two different morphologies could be distinguished in both AFM and EM images (Fig. 4.2, Supplementary Fig. 4.7). The predominant, thick morphology has a diameter of 7–8 nm while the other, thinner morphology exhibits about half the diameter of the thick fibril (Supplementary Fig. 4.8).

For cryo-EM imaging, samples of PI3K-SH3 amyloid fibrils were flash-frozen on Quantifoil cryo-EM grids and imaged with a Tecnai Arctica microscope (200 kV) equipped with a Falcon 3 direct electron detector (Supplementary Fig. 4.9). Image processing and helical reconstruction were performed with Relion 2 [190–192]. A three-dimensional density map for the thick PI3K-SH3 fibril could be reconstructed to an overall resolution of 3.4 Å. The clear density of the fibril allowed us to build an atomic model for residues 1–77 out of a total of 86 amino acids (Fig. 4.1). The missing nine residues are located at the C-terminus, which shows blurred density likely due to substantial flexibility. Previously reported low-resolution cryo-EM [184] data are in good agreement with our structure (Supplementary Fig. 4.10).

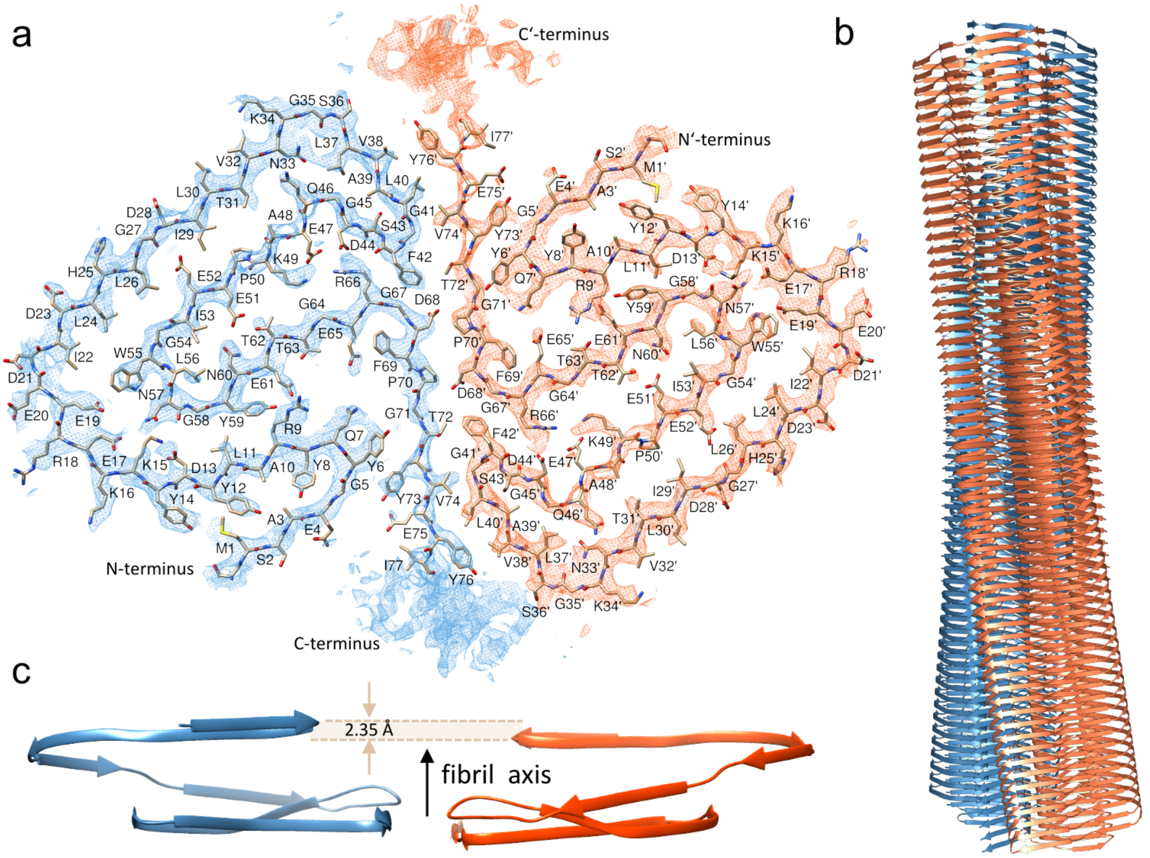

Figure 4.1: Double Filament SH3 fibril structure. **a** Cross-section of the double filament (DF) PI3K-SH3 fibril (two monomers). The different protofilaments are colored blue and orange. Density maps are shown at a contour level of 1.4 σ . The density map was sufficiently clear to model residues 1–77. **b** Side view of filaments twisting around each other displaying a total of 125 layers. **c** Relative arrangement of two adjacent PI3K-SH3 monomers within the fibril, showing a 2.35 Å shift between the protofilaments.

4.2.2 Architecture of the PI3K-SH3 Amyloid Fibril

The thick PI3K-SH3 fibril is a left-handed helical structure consisting of two intertwined protofilaments, and is thus called DF (double filament) fibril. The handedness of the density reconstruction was determined by comparison with AFM images (see Methods and Supplementary Fig. 4.11). From an analysis of the fibril height profiles in AFM images, we determined the helical pitch to be 170 ± 10 nm in reasonable agreement with a pitch of 140 nm obtained from the cryo-EM structure.

Protofilament subunits (PI3K-SH3 monomers) are stacked in a parallel, in-register cross-β structure. The spacing between the layers of the cross-β structure is around 4.7 Å and well visible in the density (Supplementary Fig. 4.12). The subunits in the two opposing protofilaments are not on the same z-position along the fibril axis but are arranged in a staggered fashion (Fig. 4.1C). The helical symmetry is therefore described by a twist of 179.4 ° and rise of 2.35 Å, which corresponds to an approximate 2_1-screw symmetry. The same staggering arrangement has been observed previously for other fibril structures, such as amyloid-β (1-42) [148] and paired-helical filaments of tau [149].

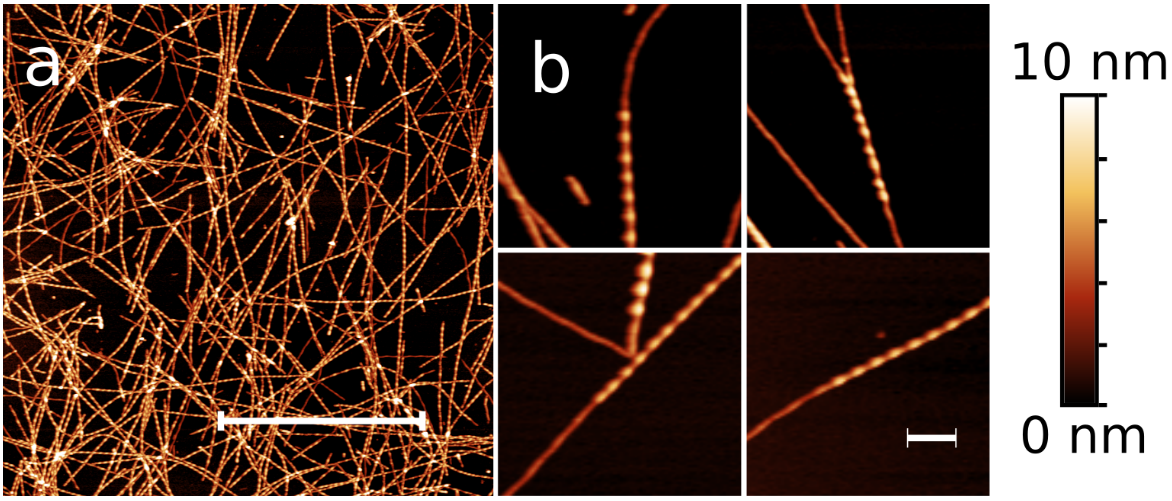

Figure 4.2: AFM images of PI3K-SH3 amyloid fibrils. **a** AFM image of PI3K-SH3 fibrils (scale bar, 3 μ m). **b** Four different close-up views of fibrils showing both the thick, double filament (DF) fibril and thin, single filament (SF) fibril morphology (scale bar, 300 nm).

Each protofilament subunit, or PI3K-SH3 monomer, consists of seven parallel in-register β-strands (β1, aa1-5; β2, aa7-19; β3, aa22-26; β4, aa28-34; β5, aa46-56; β6, aa59-68; β7, aa72-77) that are interrupted by either sharp kinks, glycine or proline residues – or a combination of those (Figs. 4.1A, 4.3, Supplementary Fig. 4.12). In particular glycine residues at kink positions have also been observed in other amyloid structures [175]. The total of seven kinks and turns (Fig. 4.3A) results in an amyloid key topology [178], which includes a structural motif similar to the bent β-arch described by Li *et al.* [175] for the α -synuclein fibril.

By encompassing 13 residues of each monomer, 26 in total (Supplementary Fig. 4.13), the inter-filament interface of the PI3K-SH3 DF fibril is very large. The protofilaments mainly interact through two identical symmetry-related hydrophobic patches (Fig. 4.4) at the periphery of their interface, between the bottom part of the bent β-arch motif and the respective C-terminal part of the opposing protofilament. The stability is provided by the hydrophobic clusters of Val^{38}, Leu^{40}, Gly^{41}, Phe^{42}, Val$^{74'}$, Tyr$^{67'}$ and Val$^{38'}$, Leu$^{40'}$, Gly$^{41'}$, Phe$^{42'}$, Val^{74}, Tyr^{67} (Figs. 4.1A, 4.4).

The amyloid-characteristic cross-β motif composed of parallel in-register β-sheets connects the different DF fibril layers and therefore contributes the largest share of intermolecular contacts (Fig. 4.5, Supplementary Fig. 4.12). This cross-β stacking is complemented by multiple inter- and intramolecular contacts including side chain interactions in homosteric and heterosteric zippers [194]. A further noteworthy feature of the structure is the fact that the PI3K-SH3 subunits are not planar but extent along the fibril axis (Fig. 4.5). The subunits within a protofilament therefore interact not only with the layer directly above (*i*+2) and below (*i*-2), but also with layers up to (*i*+6) and (*i*-6); the subunits are interlocked.

A single PI3K-SH3 monomer in the DF fibril exhibits an amyloid key topology (Fig. 4.1A), which is stabilised mainly by hydrophobic patches stretching between the strands β3–β5 (Ile^{22}(*i*), Trp^{55} (*i*-4), Leu^{24}(*i*), Leu^{26}(*i*)) (Fig. 4.4), and hydrogen bonds or salt bridges between strands β5 and β6 (Asp^{44}(*i*), Arg^{66}(*i*-2), Glu^{47}(*i*)) and β2- β6 (Arg^{9}(*i*), Glu^{61}(*i*-6)) (Figs. 4.1A, 4.3A). In the turn between strands β5 and β6, Asp^{13}(*i*+4) might bind to Lys^{15}(*i*+4) while Glu^{17}(*i*+4) might interact with Asn^{57}(*i*) (Figs. 4.1A, 4.3A). The bent β-arch motif between strands β4 and β5 is potentially strengthened by a contact between Asn^{33}(*i*) and Gln^{46}(*i*) that would tie the

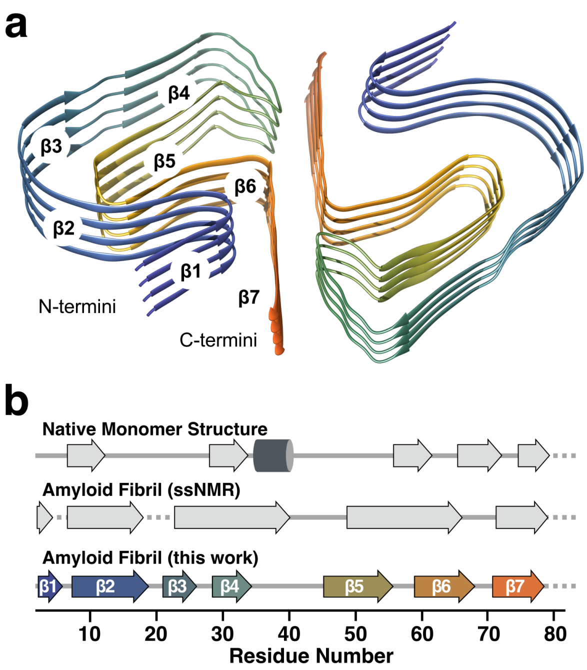

Figure 4.3: Secondary structure comparison of the DF PI3K-SH3 fibrilcryo-EM structure data obtained previously by solid-state NMR (fibrils) or liquid-state NMR (monomeric, native state). **a** Tilted cross-section of four DF SH3 fibril layers. Secondary structure is predominantly formed by seven cross-β sheets. **b** Secondary structure comparison to the native solution structure (PDB: 1PKS) [100] and ssNMR results (fibrils) modified according to Bayro *et al.* [189]. Flexible regions are shown as dashed lines, β-sheets as arrows, and the helix as a cylinder.

motif together (Figs. 4.1A, 4.3A). Further possible electrostatic interactions can be observed between strands $\beta 2$ and $\beta 6$ where the amino-group of Arg^{9}(i+4) might exhibit a salt bridge to Glu^{61}(i-2) (Figs. 4.1A, 4.3A). In addition, aromatic side chains are located in close proximity to glycine residues Tyr^{6}(i)-Gly^{71}(i+6), Phe^{42}(i)-Gly^{67}(i+2), Tyr^{73}(i)-Gly^{5}(i-6), indicating a potential involvement of glycine-aromatic $C\alpha$ -H$\cdots$ π -interactions [195, 196]. Further aliphatic-aliphatic and aromatic-aliphatic interactions comprise Ile^{29}(i)-Pro^{50}(i-2, i-4)-Ala^{48}(i-2, i-4), Met^{1}(i)-Tyr^{12}(i), and Leu^{11}(i)-Tyr^{59}(i-4). We could also observe possible interactions in-between monomer layers, so-called hydrogen bond ladders, e.g. with Gln^{7}, Gln^{46}, Asn^{57} and Asn^{60}. Additionally, these ladders could potentially be formed as well by side chains pointing outside of the fibril such as Glu^{4}, Glu^{20}, Asp^{21}, Asp^{23} and Asp^{28}. Remarkably, nearly all side chains in the outer $C\alpha$ -chain that are oriented towards the solvent are non-hydrophobic (Fig. 4.1A) (Ser^{0}-Glu^{4}, Tyr^{14}-Ser^{36}). Hence, the polar outside of the fibril shields the hydrophobic interface (Fig. 4.4). Only the hydrophobic residues Leu^{30} and Val^{32}, which are located on the fibril surface next to the bent β-arch motif, disrupt this pattern (Figs. 4.1A, 4.4).

Further hints towards the role of electrostatic interactions in maintaining the structure, comes from a pH-shift experiment. By changing the pH from 2 to 7.4, we observed that fibrils depolymerise almost completely after 1 hour (Supplementary Fig. 4.14), in agreement with the highly dynamic nature of PI3K-SH3 fibrils reported previously [197]. We propose three clusters that may be highly influenced by this pH shift: a) the space between strands $\beta 5$ and $\beta 6$ (Asp^{44}(i), Arg^{66}(i-2), Glu^{47}()); b) the interactions between Asp^{13}(i) and Lys^{15}(i), and between Glu^{17}(i+4) and Asn^{57}(i); c) the solvent-exposed and protonatable patch involving Asp^{21}(i), Asp^{23}(i) and His^{25}(i) (also highlighted as having a fundamental role in amyloid formation [198]).

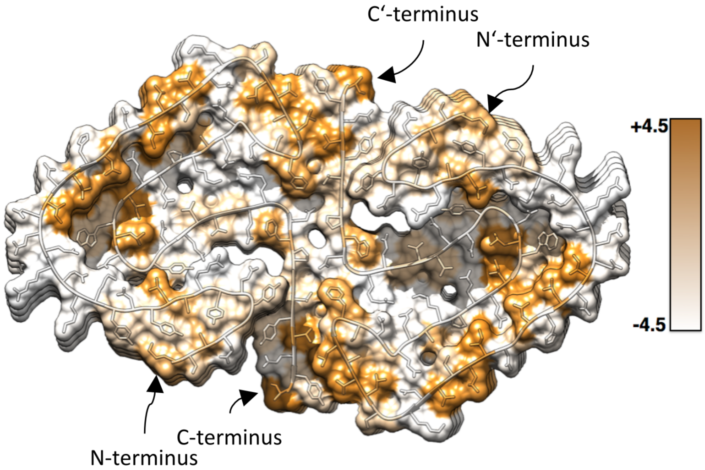

Figure 4.4: Hydrophobicity of the fibril cross-section. Hydrophobicity levels of the DF SH3 fibril cross-section are coloured according to Kyte-Doolittle [193]. Hydrophobic residues are mainly packed within the fibril core, while hydrophilic residues point towards the solvent. Hydrophobic patches in both monomers are clearly visible and are spanning several β-sheets.

The secondary structure of PI3K-SH3 fibrils and monomers has been analysed by comparing available solution NMR (monomeric native fold, PDB: 1PKS) [100] and ssNMR data (amyloid fold) [189] with our model (Fig. 4.3B). The only feature that is shared by all models is the flexible C-terminus starting around residue 80. Analysis of protein contact maps via the Contact Map WebViewer [199] of the DF fibril compared to the monomeric native structure [100] showed no consistent residue contacts in both structures, illustrating the substantially different conformation that the monomer unit has to adopt in order to incorporate into the fibril. The secondary structure of the native and amyloid fold differs substantially apart from a β-sheet between residues 70 to 80. While the monomeric native structure is characterised by one helix between residues 34 and 39, β-sheets and multiple flexible loop-regions, the DF fibril consists of seven β-sheets, exclusively, that are almost uninterrupted (Fig. 4.3B). The longest break in the β-sheet pattern of the fibril is the bent β-arch motif leading to a rigid loop between Gly^{35} and Gly^{45} (Figs. 4.1A, 4.3A). Our findings are consistent with former results by Bayro *et al.* [189] who proposed a PI3K-SH3 amyloid model based on solid-state NMR data. Both structures show β-sheets as the only secondary structure motif with most of the β -sheet regions corresponding (Fig. 4.3B). The main differences between the ssNMR and cryo-EM structures are found in the region between residues 25 and 60. Here, ssNMR data suggest two wide-spanning β-sheets while according to cryo-EM data this region consists of not two but four β-sheets, that are disrupted by a glycine residue, Gly^{27} (β3– β4) (Supplementary Fig. 4.12) and a sharp kink, Gly^{54}-Tyr^{59} (β5– β6) (Fig. 4.3A).

4.2.3 Impact of Mutations on the SH3 Aggregation

We probed the sensitivity of the fibril growth kinetics towards chemically conservative single point mutations by substituting isoleucine residues for alanine across the protein sequence. We expressed and purified the sequence variants and experimentally quantified the rates at which five different variants (I22A, I29A, I53A, I77A, I82A) elongated wildtype (WT) fibrils by quartz crystal microbalance (QCM) measurements (Supplementary Fig. 4.15) [38, 200]. This technology is ideally suited for such cross-seeding experiments. The fibril growth rates of WT

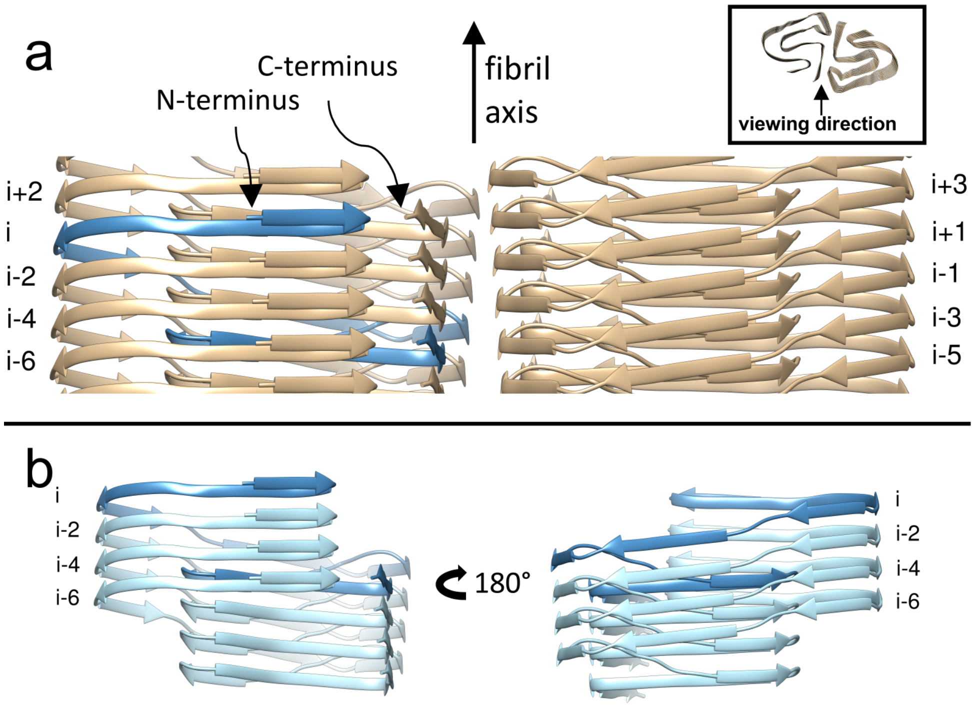

Figure 4.5: Side view of the secondary structure of the atomic model. **a** Single subunit *i* highlighted in blue with adjacent subunits in beige, described by even numbers, while the subunits of the opposite protofilament are described by odd numbers. **b** View of the minimal fibril unit (*i*) to (*i*-6) in one filament. The minimal fibril unit is displayed from two views by a turn of 180° highlighting the course of one monomer spanning several fibril layers.

and variant proteins can be directly compared, given that the same, constant ensemble of fibrils is monitored. An additional big advantage is that the use of WT seeds as templates ensures that the sequence variants adopt the same fibril structure as the WT, and therefore the change in fibril growth kinetics with respect to the wild type can be interpreted in terms of the perturbation induced by the sequence modification.

We measured the rates of WT fibril elongation by the different variants and expressed the rates relative to that of the elongation by WT monomer (Fig. 4.6, Supplementary Fig. 4.15). We found that the relative elongation rates differ by more than two orders of magnitude, with the mutations in the first third of the sequence, as well as around the middle of the sequence, displaying slow elongation rates (I22A = 0.02 ± 10.9% , I29A = 0.02 ± 7.5% , I53A = 0.003 ± 72.7%), whereas the mutations close to or within the disordered C-terminus display approximately the same rates as the wild type (I77A = 1.13 ± 9.3% , I82A = 1.59 ± 10.3%). We also took AFM images of the different single point mutants at a concentration of 100 μ M that were all seeded with WT-derived fibrils and incubated for 2 days at room temperature under quiescent conditions (Supplementary Fig. 4.16). From these images, it can be seen that all fibrils at the end of the experiment have a very similar morphology and length, except for those in the sample with I53A, where shorter fibrils are observed. The fibril growth of all the mutants seems to have come to completion within this time scale leading to very similar fibril lengths. Only in the case of I53A, the fibril growth rate is so slow that the available monomer was only partly incorporated into the seed fibrils during the course of the experiment. However, the very high mass sensitivity of the QCM permits to resolve even the growth rate of this slow growing mutant, and hence we base

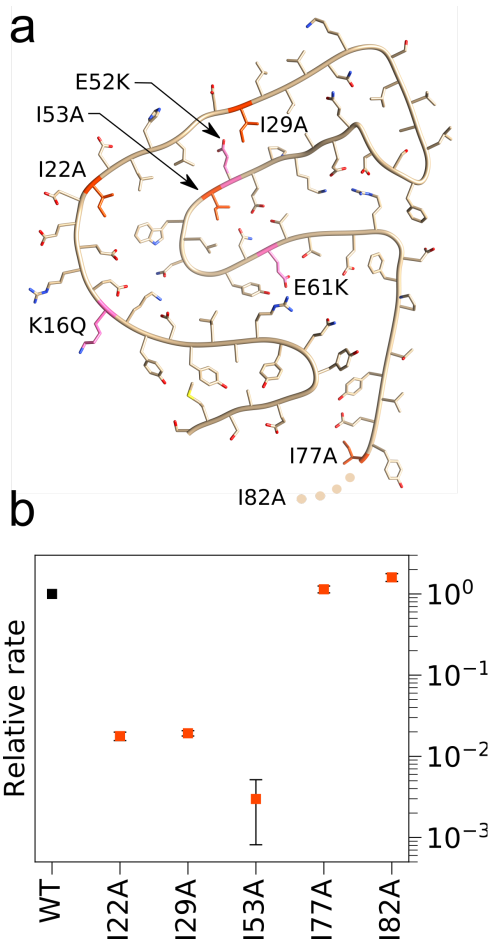

Figure 4.6: Point mutations and their effect on fibril elongation rates. **a** Top view of a PI3K-SH3 monomer derived from DF fibrils. Highlighted residues: Isoleucine residues mutated to alanine (I22A, I29A, I53A, I77A, I82A), from this work (red) and charged residues mutated by Buell *et al.* [200](K16Q, E52K, E61K, pink). Mutation I82A is missing in the model due to the flexible C-terminus. **b** Elongation rate of Ile-to-Ala mutants. The elongation rates are normalised to a WT rate of one.

our analysis of the relative growth rates exclusively on the QCM experiments (Supplementary Fig. 4.15).

4.3. Discussion

Concerning fibril architecture and polymorphism, we observed that in addition to the thick 7–8 nm DF SH3 fibril consisting of two filaments, also thin fibrils of about half the diameter are present in both AFM (Fig. 4.2, Supplementary Fig. 4.8) and negative stain EM images (Supplementary Fig. 4.7). We therefore hypothesise that the thin fibrils correspond to single filament fibrils (SF fibrils) of the same type that make up the DF fibrils. Our hypothesis is further supported by AFM images that suggest that long SF fibrils can contain stretches that appear to be identical to DF fibrils (Supplementary Fig. 4.8). It therefore seems likely that fragments of an additional protofilament can attach to or grow on a given SF fibril, to form DF fibrils. The incomplete cooperativity between the elongation of the individual filaments in a DF fibril suggests relatively weak interactions between the filaments. This, in turn, implies that one filament could well be stable without contact to another filament. We conclude that, at least in the case of PI3K-SH3, inter-filament interface contacts are not necessary for fibril formation.

A similar observation has been made for β2-microglobulin fibrils [201], which can comprise a single protofilament as well as two or more filaments resulting in at least six different polymorphs without major differences in the filament structure. Accordingly, a mixture of all polymorphs yielded only a single set of NMR resonances [201]. The stability of individual PI3K-SH3 filaments is particularly noteworthy, inasmuch as the inter-filament interface in the DF fibrils is comparatively large (Supplementary Fig. 4.13). A large inter-filament interface does therefore not indicate stable inter-filament interactions. We note here also that the identical conformation of every subunit along the fibril axis is in line with fibril growth by PI3K-SH3 monomer addition [34, 202].

A single subunit in the PI3K-SH3 DF fibril is not planar but winds itself along the protofilament axis (Fig. 4.5), which leads to interlocking within a protofilament. In contrast, in a planar subunit, the inter-subunit interactions within one protofilament would consist exclusively of the cross-β pattern, while all transverse interactions would be intramolecular. Here, however, one subunit within one protofilament is not only in contact with its direct neighbours (above and below) but with four other monomers in total, i.e. in addition to the longitudinal hydrogen bonds in the cross-β structure, there are other transverse interactions between the subunits. This staggered arrangement leads to an interlocking mechanism connecting several monomers within a protofilament, which very likely further stabilises the structure (Fig. 4.5). Interestingly, this interlocking mechanism is commonly seen in other amyloid fibril structures determined by cryo-EM [148, 175, 176, 178, 179]. It should be noted that NMR data can only distinguish between *intra*- and *inter*molecular contacts but cannot directly reveal a potential staggering of subunits along the fibril axis.

Given that the interlocking is observed in other fibril structures as well, it might in general contribute to the formation of stable fibrils by optimisation of the side chain packing [170]. In addition, the staggered architecture might in part be responsible for the templating effect during fibril elongation, as it establishes a rugged binding interface that may guide the incoming monomer into the fibril conformation, engaging it in more intermolecular contacts than a flat interface could.

As a measure for the interlocking of subunits we have previously defined the concept of a minimal fibril unit [148], which is the smallest fibril structure fragment in which the capping subunits at both ends would have established the same full contact interface with other constituting monomers as the capping subunits of an extended fibril. Since we hypothesise that one protofilament of the DF fibril could exist on its own and is identical to the SF fibril, here we describe the minimal fibril unit also for a single protofilament. For the PI3K-SH3 fibril, the minimal fibril unit has a size of four subunits when considering an individual protofilament, and a size of eight subunits in the case of the DF fibril (Fig. 4.5B).

In order to rationalise kinetic data of fibril formation and growth we substituted five different isoleucine residues with alanines, probing how these chemically conservative single point mutations at different positions affect the addition of new monomers to the fibrillar structure formed by the wild type sequence. Three mutations (I22A, I29A, I53A) showed a strong decrease of the elongation rate of two to three orders of magnitude compared to the wild type sequence (Fig. 4.6). In these positions, the side chains of the three isoleucines are all pointing towards the fibrillar core (Fig. 4.6). While the chemical nature of the amino acid substitution we chose is conservative (aliphatic to aliphatic), the bulkiness of the side chain decreases. The ability of these variants to elongate the WT structure, albeit significantly slower than the WT sequence itself, suggests that the formation of a cavity due to the reduced bulkiness is energetically tolerated. However, the free energy difference between the monomeric state and the transition state (structural ensemble) of the elongation reaction appears to be increased. Such an increase in energy difference could come either from a stabilisation of the monomeric state, or from a destabilisation of the transition state. The former seems less likely, due to the mostly disordered nature of PI3K-SH3 at pH 2, while the latter possibility could be caused by the reduction in hydrophobic contacts between the monomer and the fibril end. This conclusion, which assumes some degree of contact between the monomer and the fibril end in the transition state for fibril growth, is in excellent agreement with previous results that underline the importance of the sequence hydrophobicity for the magnitude of the elongation free energy barrier[buellDetailedAnalysisEnergy2012]. The remaining two mutations (I77A, I82A) (Fig. 4.6) are instead located close to or within the flexible C-terminus and show indeed a much weaker or no effect on the elongation rate. In both cases a chemical modification to alanine does not perturb any interactions crucial for the energetics of the transition state.

The availability of the high-resolution structure also allows us to rationalise the influence of previously reported single point mutations of PI3K-SH3 on the kinetics of fibril elongation. In a previous study [200], the effect of changes in charge at three different positions (Fig. 4.6; K16Q, E52K, E61K; residues depicted in pink) led to very different effects on the elongation rate. While these mutations, similar to the ones we have designed and studied in the present work, could lead to different fibrillar structures if induced to form fibrils *de novo*, the use of WT seed fibrils in both studies allows us to discuss here the effect of these mutations in the light of the present fibril structure. This is because the well-known templating effect in amyloid fibril growth imprints the structure of the fibril template onto the monomeric protein that adds onto the fibril end. The mutation E52K probably leads to the creation of a positive charge inside the hydrophobic fibril core, a highly unstable arrangement. In the WT fibril, Glu^{61} can form a salt bridge with Arg^{9}, which leads to an additional driving force for the deprotonation of Glu^{61}. If Glu^{61} is substituted by a lysine (E61K), two positive charges come into close proximity, again leading to a highly unstable situation. On the other hand, the mutation K16Q does not lead to any major change in the kinetics of fibril elongation, which is most likely due to the side chain pointing towards the outside of the fibril [37].

Our structural model can also help to understand the effects of further sets of previously investigated sequence changes [198]. With the aim of understanding which part of the sequence plays a major role in the amyloidogenicity of the protein, different portions of the sequence were replaced or mutated. In the vast majority of cases, the amyloidogenic behaviour was completely abolished, which we can now explain with our structural model: the introduction of bulky or charged residues facing the inner core of the structure destabilises the present amyloid fold, as evidenced by the mutants E17R/D23R and Q7E/R9K/E17R/D23R [198]. The only mutation that does not show a significant decrease in amyloidogenicity does not involve changing the charge of a buried residue: the mutant referred to as PI3-QMR (E17Q/D23M/H25R) modifies a single charge of outward pointing residue 25 (assuming Glu^{17} and Asp^{23} to be protonated at pH 2 [198]).

It should also be noted that the circularisation of the PI3K-SH3 sequence through the use of disulphide bridges causes a decrease of the elongation rate but does not prevent the circularised mutant to acquire the amyloid conformation [107]. The close proximity of the N- and C-termini (Fig. 4.1A), in combination with the flexibility of the C-terminal nine residues, likely allows the cyclised (disulfide bridge between Cys^{3} and Cys^{82}) sequence to form an amyloid structure very similar, if not identical to the one presented here.

A noteworthy example of substantial sequence modification is the grafting of the N-Src loop of SPC-SH3 onto PI3K-SH3, while simultaneously removing the stretch from residue 31 to 53 from PI3K-SH3 [203]. This operation did not remove the amyloidogenic properties from the modified PI3K-SH3 domain. The removed sequence stretch is part of the bent β-arch motif in the amyloid conformation. The removal of this prominent motif could have been expected to lead to a more significant impairment of amyloid fibril formation. However, the altered sequence is likely to be able to respond to this strong perturbation by forming an alternative structure.

In summary, we have determined the structure of an PI3K-SH3 amyloid fibril. The PI3K-SH3 fibril has been extensively studied in the past and the effect of many mutations on the kinetics of amyloid formation has been described. The atomic structure of the PI3K-SH3 fibril enables us now to rationalise the effect of these mutations, which is the basis for understanding the sequence-dependence of amyloidogenicity and ultimately the determinants of amyloid formation in general.

4.4. Methods

4.4.1 Protein production

WT and mutants of the bovine PI3K-SH3 domain were purified according to the protocol of Zurdo *et al.* [39]. All constructs contain a 6xHis-tag linked to the protein by a thrombin cleavage site. The sequence of the WT protein after cleavage is the following, with the peptide Gly-Ser remaining as overhang from the cleavage:

GS MSAEGYQYRA LYDYKKEREE DIDLHLGDIL TVNKGSLVAL GFSDGQEAKP EEIG-WLNGYN ETTGERGDFP GTYVEYIGRK KISP

The protein was expressed in a BL21 *E. coli* strain with TB medium for auto-induction containing 0.012 % Glucose and 0.048 % Lactose. The cells were grown for over 24 h and then harvested by centrifugation. After resuspension in sodium phosphate buffer (50 mM sodium phosphate pH 8, 5 mM Imidazole and 100 mM NaCl), the cells were disrupted by sonication, in presence of protease inhibitors and DNAse. The lysate was centrifuged, and the supernatant loaded on a Ni-NTA Superflow Cartridge (Qiagen, Venlo, Netherlands) equilibrated in 50 mM sodium phosphate pH 8, 5 mM Imidazole and 100 mM NaCl. The protein was eluted with a linear gradient from 5 to 300 mM imidazole in 50 mM sodium phosphate pH 8, 100 mM NaCl in 25 ml elution volume. Fractions containing the protein were collected and cleaved overnight at 7° C with 1 unit of thrombin (from bovine plasma, Sigma-Aldrich Saint Louis, Missouri, USA) per 1 mg of protein. The cleaved solution was then concentrated and loaded on a SEC HiLoad 26/60 Superdex 75 column (GE Healthcare, Chicago, Illinois, USA) equilibrated with 5 mM ammonium acetate pH 7. Fractions containing the PI3K-SH3 domain were collected and lyophilised for further use.

4.4.2 Fibril formation

The lyophilised protein was resuspended in 10 mM glycine-hydrochloride pH 2.5 buffer at a final concentration of ca. 200 μM. The solution was shaken in an Eppendorf tube at 1400 rpm at 42°C for 24 h to form seeds. These seeds were then sonicated in an Eppendorf tube in a volume of ca. 500 μl for 15 s (1 s “on” , 2 s “off” , 10 % amplitude) with a Bandelin Sonoplus using a M72 probe. To prepare the twisted fibrils, a new solution with ca. 100 μM monomer in 10 mM glycine-hydrochloride pH 2.5 was then mixed with 5 μM of equivalent seeds mass and incubated without stirring at 50 °C overnight.

4.4.3 AFM imaging

The fibril samples were diluted in 10 mM glycine-hydrochloride, pH 2.5 to a concentration of 5 μM and 10 μl were pipetted on a mica substrate. After 10 min of incubation, the mica was washed extensively with milliQ water and dried under a nitrogen gas flush. The pictures were taken in tapping mode on a Bruker Multimode 8 (Billerica, Massachusetts, USA) using OMCL-AC160TS cantilevers (Shinjuku, Tokyo, Japan).

4.4.4 Fibril elongation measurements with QCM

The elongation rate of PI3K-SH3 fibrils was measured through immobilisation of fibrils on a QCM sensor and subsequent incubation with monomer solution [34]. To immobilise the fibrils on the sensor, chemical modification is necessary. To achieve that, the fibrils were mixed at a final concentration of 50 μM in buffer (10 mM glycine-hydrochloride, pH 2) with EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride) (1M) and cystamine hydrochloride (0.5 mgml). After pelleting and washing the chemically modified fibrils, they were sonicated in an Eppendorf tube in a volume of ca. 500 μl with a Bandelin Sonoplus using a MS72 probe (10% amplitude, 15 s, 1 s “on” , 2 s “off”). The gold sensors (Biolin Scientific, Gothenburg, Sweden) were then incubated with the above-mentioned solution overnight in a 100% humidity environment. The measurements were performed with a QSense Pro (Biolin Scientific, Gothenburg, Sweden) by measuring the elongation rate as change in resonant frequency over time. With the temperature set at 25°C, the monomer solutions were injected for 30 s at a flow rate of 100 μl per second and the measurement lasted until a stable slope was reached. To obtain the relative rates, the protein solutions were injected in different sensor chambers after a WT injection, the latter being used as normalization reference. Two different triplicate measurements were performed for I53A. Two different duplicate measurements were performed for all the other mutants. The rate was measured as slope of the 3rd overtone and averaged among the multiple injections. The data are presented as average values with error bars indicating the standard deviation.

4.4.5 Fibril dissociation at pH 7.4

Fibril dissociation at pH 7.4 was probed by measuring ThT and intrinsic fluorescence change over time in two series of triplicates. The ThT measurements were performed by mixing 40 μl of 100 μM PI3K-SH3 fibrils and 50 μM ThT in 10 mM glycine-hydrochloride pH 2 with 60 μl of 100 mM sodium phosphate pH 7.4. The mixing was carried out using the injection system of a CLARIOstar plate reader (BMG LABTECH, Ortenberg, Germany) and measuring immediately afterwards by exciting at 440 nm and recording the signal intensity at 480 nm. The intrinsic tryptophan fluorescence measurements were carried out by mixing the same two solutions (without ThT) by pipetting, followed by the measurement of fluorescence spectra every 15 seconds by exciting at 290 nm and recording between 300 and 380 nm in 2 nm intervals.

The analysis of soluble peptide by concentration determination at the end of the dissolution experiment was performed after one night of equilibration after the mixing of the two solutions mentioned above (without ThT). The samples were spun down for 30 min at 16,100 g. The supernatant was then measured and the protein concentration was determined by measuring the absorbance at 280 nm together with the extinction coefficient of PI3K-SH3 of $\varepsilon_{280} = 15930$ $M^{-1}cm^{-1}$ using a V650 UV-Vis spectrophotometer (Jasco, Easton, MD, USA).

4.4.6 Negative stain and cryo-EM image acquisition

Negatively stained fibrils were prepared on 400 mesh carbon-coated copper grids (S160-4, Plano GmbH, Germany), stained with 1% uranyl acetate, and imaged using a Libra120 electron microscope (Zeiss) operated at 120 kV. Cryo-preparation was performed on glow-discharged holey carbon films (Quantifoil R 1.2/1.3, 300 mesh). The sample containing 50 μM PI3K-SH3 was 4x/10x/20x diluted with 10 mM glycine-hydrochloride (pH 2) to a final concentration of 12.5, 5 or 2.5 μM monomer equivalent. A total sample volume of 2.5 μl was applied onto the carbon grid and blotted for 3.5 s before being cryo-plunged using a Vitrobot (FEI). With 110,000-fold nominal magnification 622 micrographs have been recorded on a Tecnai Arctica electron microscope operating at 200 kV with a field emission gun using a Falcon III (FEI) direct electron detector in electron counting mode directed by EPU data collection software. Each micrograph was composed of 60 fractions. Each fraction contained 42 frames, i.e. in total 2520 frames were recorded per micrograph. The samples were exposed for 65 s to an integrated flux of 0.4 $e^{-}/Å^{2}/s$. Applied underfocus values ranged between 1.5 and 2.25 μm. The pixel size was 0.935 Å, as calibrated using gold diffraction rings within the powerspectra of a cross grating grid (EMS, Hatfield).

4.4.7 Cryo-EM Image Processing and Helical Reconstruction

MotionCor2 [204] was used for movie correction. Fitting CTF parameters for all 622 micrographs was performed using CTFFIND4 [205]. Further image processing and 3D reconstructions were done with Relion2 [190, 191]. Selection of 256 micrographs was done with CCTFFIND by estimating the maximum resolution at which Thon rings could be detected to be better than 5 Å. From these micrographs, 4540 fibrils were manually picked. From these fibrils, 103,733 segments were extracted using an overlap of 90 % between neighbouring segments. The size of the segment images is 220 pixels. For data set characterisation we performed 2D classification (Supplementary Fig. 4.17). As an initial model for the refinement we used a noise-filled cylinder.

After several rounds of 3D refinements with helical symmetry search, we found a problem with the tilt priors: the tilt angle distribution became bimodal with maxima at 85° and 95°. However, we would expect the tilt angles to show a unimodal distribution around 90°. The $\sim$ 4.7 Å cross-β pattern is a strong signal and substantially affects the alignment. If the helical rise parameter is slightly smaller than the correct value, the cross-β pattern can still be aligned by changing the tilt angles to higher or lower values (which accordingly reduces the spacing of the cross-β pattern). To overcome this problem, we fixed the tilt prior to 90° by usage of the Relion option *–helical_keep_tilt_prior_fixed*, and then first optimized the helical parameters. In subsequent refinements, the helical symmetry parameters were fixed and the tilt angles (together with the other angles) were optimized.

Since the automated 3D refinement in Relion did not yield high-resolution reconstructions, we performed gold-standard refinements by splitting the data into an even and an odd set by selecting entire fibrils (not just segments, as they are overlapping). The FSC curve (Supplementary Fig. 4.18) was computed between the two half-maps and yields a resolution (with the 0.143 criterion) of 3.4 Å.

The handedness of the fibril structure was determined by comparing the reconstructed density with AFM images (Supplementary Fig. 4.11). For this comparison, the 3D density map of the fibril was converted to a height profile using Chimera [206] as follows: set surface color by height and set the color scale to gray. Then set camera projection mode to orthographic and save the image. The alignment of the AFM image with the calculated height profiles yield a cross-correlation coefficient of 0.943 for the left-handed and 0.914 for the right-handed helix.

4.4.8 Model Building and Refinement

A single chain atomic model of PI3K-SH3 was built with Coot [207, 208]. Subsequently, seven copies of a single chain were placed into the EM density map. At residues Gln^{46} and Glu^{47}, between strands $\beta 5$ and $\beta 6$, the density map is slightly ambiguous and could possibly be in agreement with an alternative interpretation for the trace of the Cα -chain (Supplementary Fig. 4.19).

The final model containing seven helical symmetry related chains was used for further real space refinement in PHENIX [209]. Refinement was carried out using a resolution cut-off of 3.4 Å and NCS restraints between all seven subunits. At later stages of the refinement, hydrogen-bond restraints were defined for the cross-β sheets and Ramachandran restraints were used. The model-map FSC curve as obtained from *phenix.real_space_refine* is shown in Supplementary Fig. 4.18 (dashed line). The final statistics on the details of the refinement are shown in Supplementary Table 4.1. Molecular graphics and analyses were performed with Chimera [206].

Data availability

The structure of the PI3K-SH3 fibril has been deposited in the Protein Data Bank under accession code 6R4R [http://dx.doi.org/10.2210/pdb6R4R/pdb]. The 3.4 Å cryo-EM density map has been deposited in the Electron Microscopy Data Bank under accession code EMD-4727. The source data underlying Fig. 4.6b and Supplementary Fig. 4.14 are provided as a Source Data file. Other data are available from the corresponding authors upon reasonable request.

4.5. Acknowledgments

We thank P.J. Peters and C. López-Iglesias for advice and helpful discussions, H. Duimel for help with sample preparation, and the M4I Division of Nanoscopy of Maastricht University for microscope access and support. We thank Philipp Neudecker for helpful comments on the manuscript. The authors gratefully acknowledge the computing time granted by the Jülich Aachen Research Alliance High-Peformance Computing (JARA-HPC) Vergabegremium and VSR commission on the supercomputer JURECA at Forschungszentrum Jülich. Computational support and infrastructure was provided by the Center for Information and Media Technology (ZIM) at the University of Düsseldorf (Germany). A.K.B. and N.V. thank the Deutsche Forschungsgemeinschaft (DFG) for funding. D.W. was supported by grants from the Portfolio Technology and Medicine, the Portfolio Drug Design, and the Helmholtz-Validierungsfonds of the Impuls und Vernetzungs-Fonds der Helmholtzgemeinschaft. This study was funded in part by the DFG SFB 974 and SFB 1208 (to D.W.). Support from a European Research Council (ERC) Consolidator Grant (grant agreement no. 726368) to W.H. is acknowledged. A.K.B. thanks the Novo Nordisk Foundation for support through a Novo Nordisk Foundation Professorship.

Author contributions

A.K.B. and G.F.S. conceived the study. N.V. and L.N.M. performed the biochemical experiments. A.K.B., N.V., L.N.M. analysed the biochemical and kinetics data. R.G.B.R. performed cryo-EM experiments and initial data analysis. C.R. and G.F.S. performed image processing, reconstruction and model building. C.R., A.K.B., N.V., G.F.S., W.H., and L.G. wrote the manuscript. ll authors discussed results and commented on the manuscript.

4.6. Supplementary information

Data Collection	
Microscope	Tecnai Arctica
Camera	Falcon 3
Acceleration voltage (kV)	200
Nominal Magnification	110,000
Defocus range (µm)	1.5 to 2.25
Dose rate (e-/Å2/s)	0.4
Number of movie frames	2520
Exposure time (s)	65
Total electron dose (e-/Å2)	26.2
Pixel size (Å)	0.935
Reconstruction	
Box size (pixel)	200
Interbox distance (Å)	18.8
Number of extracted segments	103,733
Number of segments after 3D classification	27,681
Resolution based on the 0.143 FSC criterion (Å)	3.4
Map sharpening B-Factor (Å2)	150
Helical rise (Å)	2.3548
Helical twist (°)	179.436
Symmetry	C1 (pseudo-21 helical)
Model Composition	
Non-hydrogen atoms	4494
Number of chains	7
Model Refinement	
Resolution (Å)	3.4
Map CC (around atoms)	0.6457
RMSD bonds (Å)	0.008
RMSD angles (°)	1.426
All-atom clash score	6.73
Ramachandran outliers/favored (%)	0/88 %
Rotamer outliers	0/1.59 %
C-beta deviations	0/0
EMRinger score	4.98
Molprobility score	1.97

Table 4.1: Statistics of cryo-EM data collection, reconstruction and model building

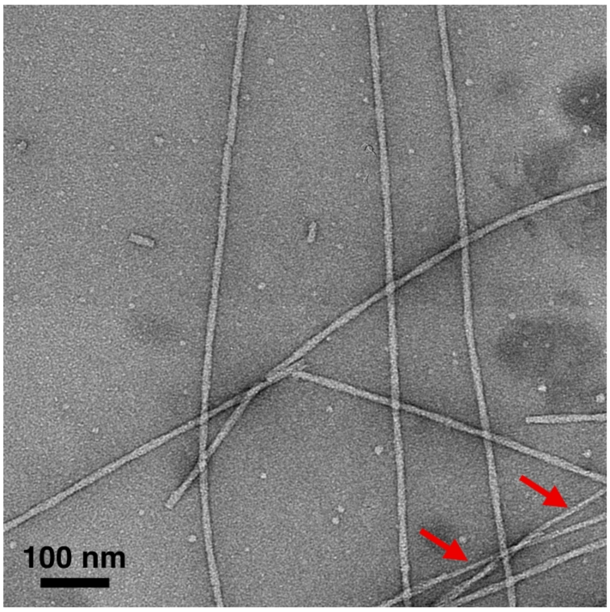

Figure 4.7: Exemplary negative stain micrograph showing the predominant thick double filament (DF) PI3K-SH3 fibrils and less dominant thin single filament (SF) fibrils (red arrows).

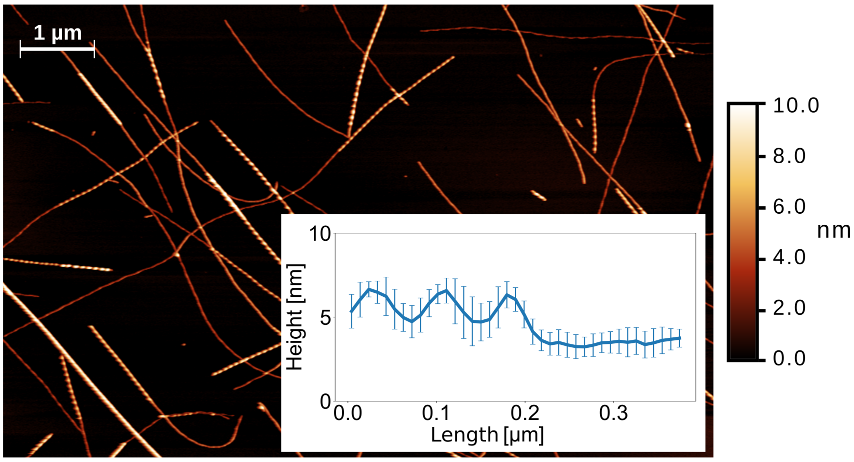

Figure 4.8: Averaged height pattern of different fibril profiles along their length. The different profiles were aligned on the peak position, the shortest profile was chosen as a minimum length, and their height averaged (error bars correspond to the standard deviation). The first three peaks are indicative of the double filament (DF) portion of the fibrils, while the flat part is indicative of the single filament (SF) portion.

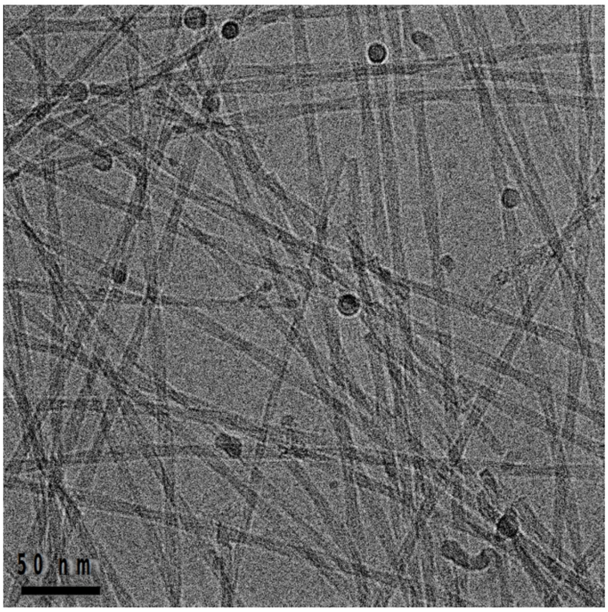

Figure 4.9: Representative cryo-EM micrograph of DF PI3K-SH3 fibrils. The fibril twist of the thick DF fibril is already well visible in this low-contrast image.

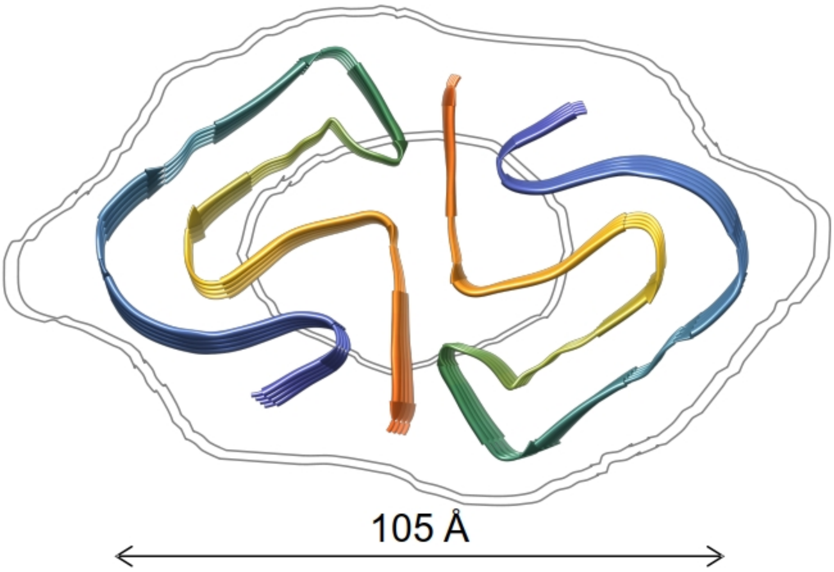

Figure 4.10: Overlay of the presented DF PI3K-SH3 fibril model with the low-resolution cryo-EM density (contour graphically extracted) from Jimenez et al. (1999). The model and the density are in good agreement, which suggests that both preparations likely yield the same structure.

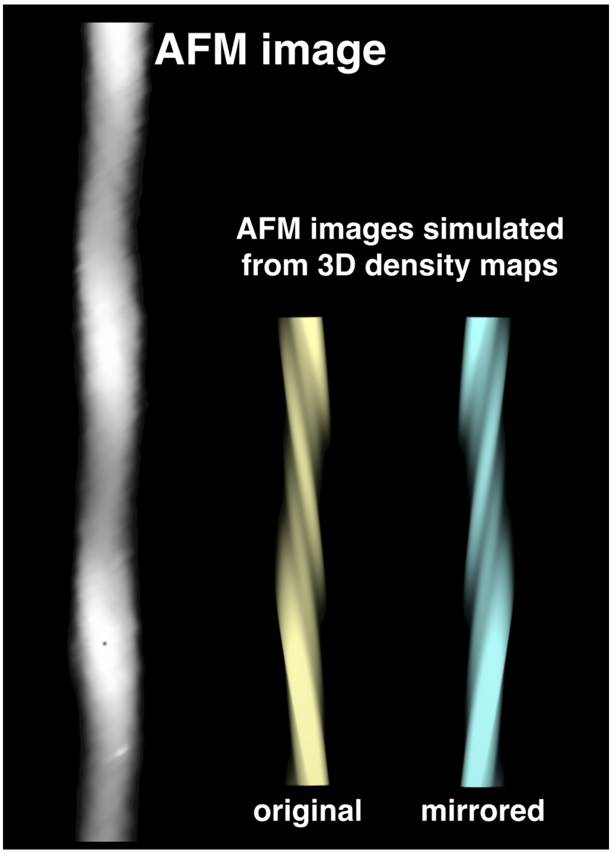

Figure 4.11: Comparison of an AFM image of a DF PI3K-SH3 fibril with height profiles computed from the 3D EM reconstructions. The original left-handed reconstruction (yellow) yields a higher cross-correlation coefficient (0.943) with the AFM images (grey) than the mirrored reconstruction (0.914, cyan), indicating that the DF PI3K-SH3 fibril is left-handed.

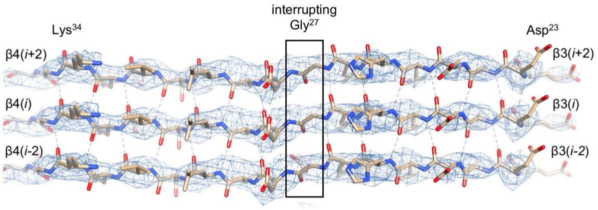

Figure 4.12: Side view of DF PI3K-SH3 amyloid fibril showing residues 22–35. Displayed are three layers of DF PI3K-SH3 encompassing parts of in-register sheets $\beta 3$ and $\beta 4$ with hydrogen bonds highlighted as dashed lines. The cross-β pattern between sheets $\beta 3$ and $\beta 4$ is only interrupted by residue Gly^{27}.

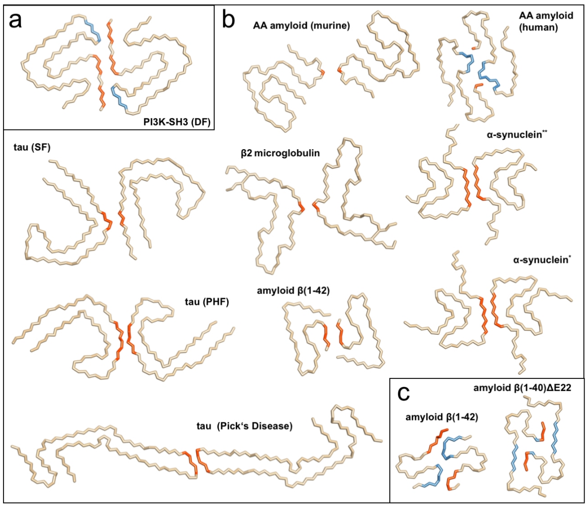

Figure 4.13: Interface gallery. a The protofilament interface of DF PI3K-SH3 fibrils is composed of two sequence regions (orange and blue) and is large compared to those of amyloid fibrils previously determined by cryo-EM (b) (Aβ(1–42), PDB: 5OQ56; β2 microglobulin, PDB: 6GK37; AA amyloidosis (human), PDB: 6MST8; AA amyloidosis (murine), PDB: 6DSO8; α-synuclein*, PDB: 6A6B9; α-synuclein**, PDB: 6H6B10; tau (PHF)11, PDB: 5O3L; tau (SF), PDB: 5O3T11; tau Pick's Disease, PDB: 6GX512) 6,7,10–13 . c Protofilament interfaces of different Aβ variants determined by solid-state NMR (Aβ(1-40)ΔE22, PDB: 2MVX14; Aβ(1-42), PDB: 5KK315) exhibit similar sizes and complexities as observed for DF PI3K-SH3 fibrils. Interfaces are shown as Cα-chain. Interface contacts have been defined as Cα-contacts with a cut-off of <10 Å. Beige, no interface contacts; orange, interface contact formed by first involved β-sheet; blue, interface contact formed by second β-sheet.

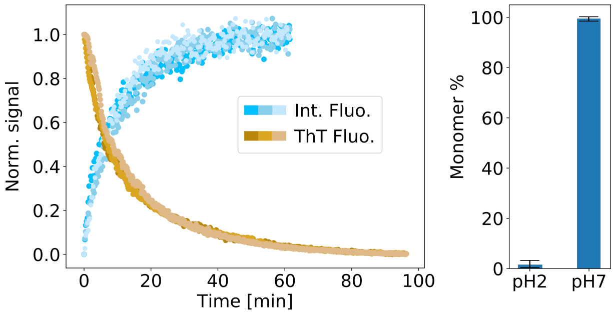

Figure 4.14: PI3K-SH3 amyloid fibril dissociation at pH 7. Left panel: ThT (brown data points) and intrinsic tryptophan fluorescence intensity ratio 340 nm/310 nm (blue data points) traces of fibril dissociation after dilution into pH 7.4 buffer (measurements are done in triplicate). Right panel: Concentration of protein in the supernatant as percentage of the total protein concentration after incubation for 24 h and centrifugation for 30 min at 16,100 g. Error bars represent the standard deviation on a triplicate measurement. Source data are provided as a Source Data file.

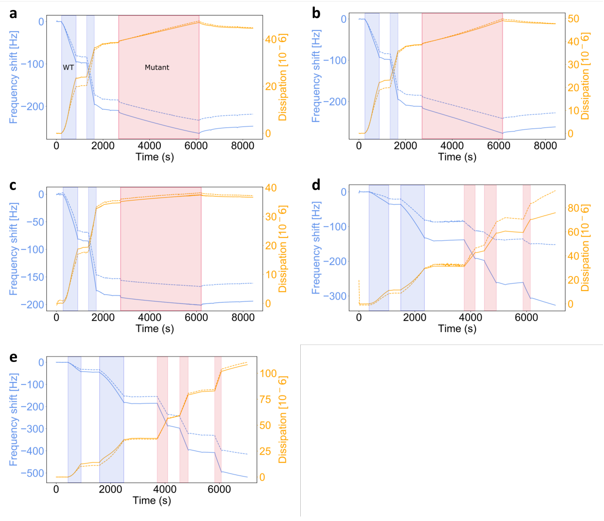

Figure 4.15: Example of QCM-D traces. Injections of monomeric wild-type (WT) PI3K-SH3 are highlighted with blue bands, and injections of monomeric Ile-to-Ala mutants of PI3K-SH3 are highlighted with red bands (**a** I22A; **b** I29A; **c** I53A; **d** I77A; **e** I82A). The white regions correspond to contact of the QCM sensor surface with buffer. Shown are both the changes in resonant frequency (blue), as well as dissipation (orange). In both cases, the signal for the two overtones N = 3 (dashed line) and N = 5 (solid line) are shown.

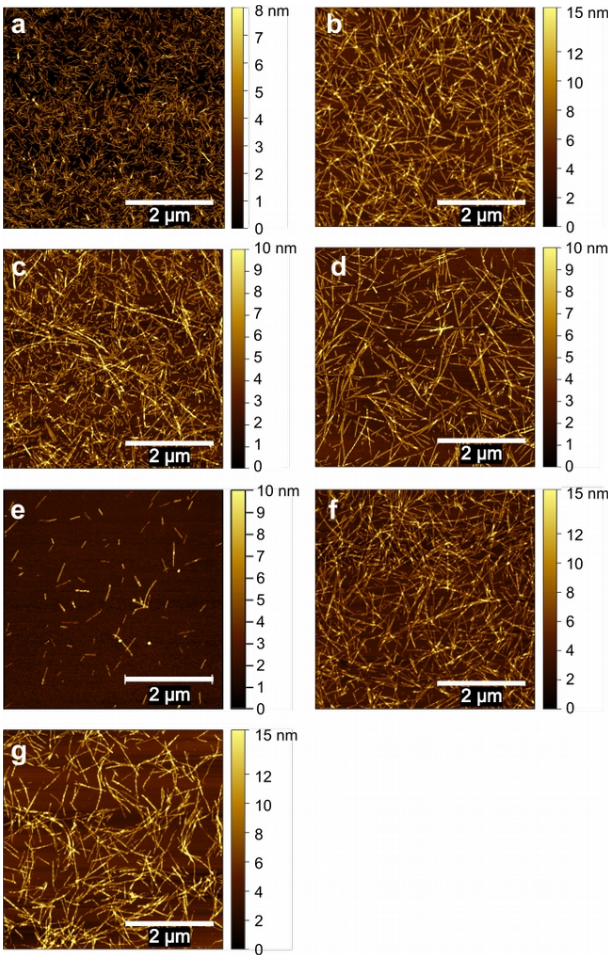

Figure 4.16: AFM images of **a** WT seeds before elongation experiments. **b** WT seeds incubated with WT monomer. **c-g** WT seeds incubated with monomer of I22A (**c**), I29A (**d**), I53A (**e**), I77A (**f**), I82A (**g**).

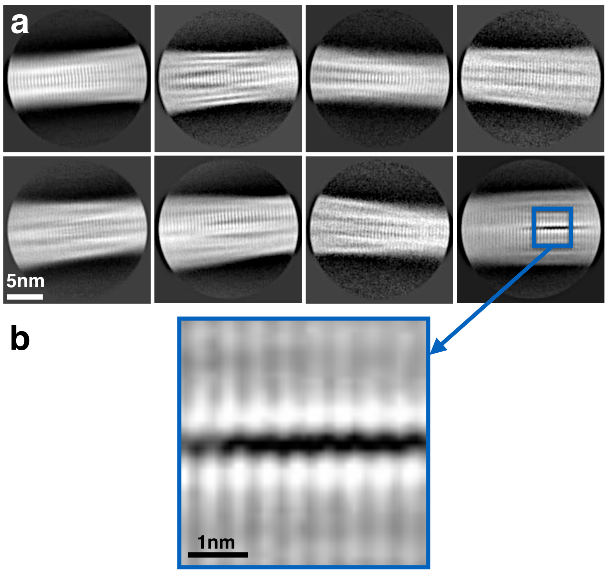

Figure 4.17: Exemplary 2D classes of DF PI3K-SH3 fibrils. **a** Overview of eight 2D classes comprising 146,215 particles. **b** Magnification of one of the classes displaying the characteristic β-sheet stacking.

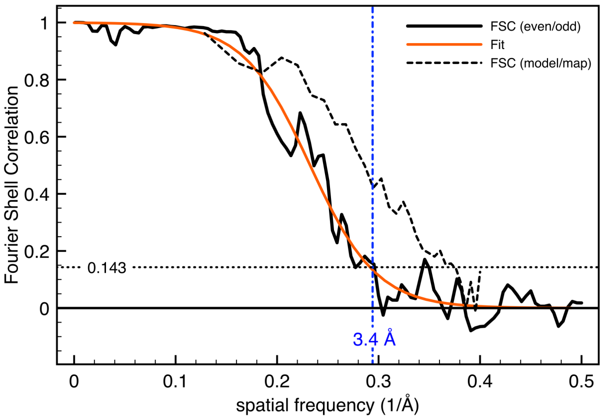

Figure 4.18: FSC Analysis. FSC curves from the even/odd test (solid black) from the gold-standard refinement yields a resolution of 3.4 Å (using the 0.143 criterion). The even/odd FSC curve is fitted (orange) with the model function 1/(1+e(x-A)/B)) (with A=0.2325 and B=0.03286) to obtain a more robust resolution estimate. The FSC curve comparing the density map computed from the atomic model with the full density reconstruction yields a cross-resolution of 3.4 Å (dashed black).

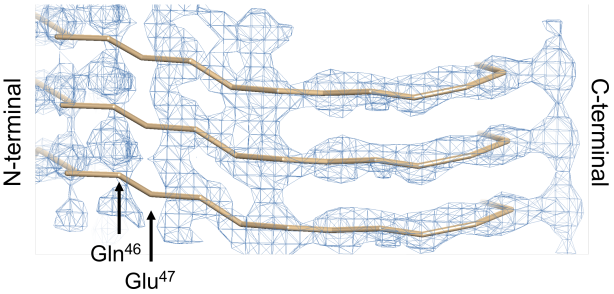

Figure 4.19: Details of map density between residues 45 and 55. The density around residues Gln^{46} and Glu^{47} does not allow for unambiguous chain tracing, which is mostly due to the weak density of the Glu^{47}. Weak densities of negatively charged residues are typically observed in cryo-EM density maps. The unclear density might also indicate the presence of a less populated second conformer. Detailed magnification of the density map shows the critical area. The Cα-chain trace is shown (brown).

5

Φ-VALUE ANALYSIS OF PI3K-SH3 AMYLOID FIBRIL ELONGATION

5.1. Introduction

Until the 1980s, one of the open fundamental problems in the mechanism of protein folding was the determination of the folding pathway [210]. The relation between sequence and structure had to be understood, but it was acknowledged that the connection between the two was to be found in the conformational space explored by the protein during the folding reaction.

In 1989, Sir Alan Fersht and collaborators proposed a new approach to study the pathway of protein folding by mapping the transition state of the enzyme Barnase [211]. The strategy aims at measuring the equilibrium stability and kinetic rates of folding of different mutants of the protein. If the mutants show differences compared to the WT, these will allow describing to which extent the mutated residues contribute to stabilize the transition state structural ensemble.

The methodology has been successfully employed to describe the transition state of many more proteins over the following years and it became known as Φ-value analysis, in analogy to the Bronsted β-value analysis [212]. Similarly to the latter, the Φ-value analysis relies on the conventional definition of the transition state, namely the maximum free energy point along the reaction coordinates. The transition state of protein folding can be described as the ensemble of states in the saddle point of the energy surface [56]. To reach this point on the folding reaction coordinates, the unfolded protein has to form and break several weak bonds. This property of the reaction allows studying it through structure-reactivity relationships, using a perturbation approach to observe the changes of the reaction upon modifications of the structure. The Φ-value analysis exploits this property, as it consists in the experimental analysis of the effect of mutations of individual residues. In this way the transition state (TS) can be investigated by obtaining atomic resolution details [56]. The resolution of this approach allowed the study of folding pathways of many proteins, filling the gap in the mechanistic understanding of such a complex reaction.

Similarly to protein folding in the second half of the 20th century, in the last three decades amyloid fibrils and the misfolding reaction received increasing interest. A new link was found between the amyloid conformation and its involvement with several neurodegenerative and systematic pathologies, while the cross-β conformation has proven to be common to many different polypeptide systems [9, 30]. Even though the formation of an aggregation-driven alternative folding was already proposed as a mechanism in 1935 by Astbury and Dickinson, today the reaction mechanism of misfolding is still largely unknown [213]. The huge variety of different pathways that lead to fibrillar aggregation is often at the base of the many different models proposed [214]. Among all, the one reaction with the deepest understanding seems to be so far the elongation reaction. The elongation reaction involves a fibril end and a soluble monomer. Throughout the reaction the monomer goes from soluble to fibrillar state, changing its conformation and reaching a more favourable energetic state [30]. Both the change in the energetic state and in conformation resemble closely the classical folding process, with the difference of being a bimolecular process.

While literature regarding kinetic aspects is wide and detailed for many different amyloid systems, the thermodynamic aspects are less characterized. Even less studied is the transition state, of which its mechanistic aspects so far have been investigated at atomic level just by simulations [52–54], relying on existing structures of the amyloid fibrils, even though lacking experimental guides. To better tackle the challenge of diseases correlated to amyloid structures, a detailed mechanistic understanding of the reaction is necessary [214].

In this work, PI3K-SH3 amyloid fibril elongation kinetics and thermodynamic are characterized at an unprecedented level of detail. By using a limited set of mutants, I present a strategy to study the influence of the different residues in the elongation process. The experimental part involves both kinetic measurements to study the kinetic barrier of elongation and equilibrium measurements aimed at characterizing the difference in stability of the two conformations. The effects of single point mutations on the two energetic states allow us to understand the role of the different mutated residues in forming the minimum core of interactions that guide the monomer to acquire the amyloid fold. By demonstrating the ability of this method to individuate some key residues of the elongation reaction, the hope is to set the ground to deeper mechanistic studies in the amyloid field, directed both toward functional and pathological amyloids.

5.2. Theoretical framework

The definition of the Φ-value relies on both kinetic and thermodynamic measurements [212]. Kinetic measurements are necessary to probe the energetic barrier that the unfolded protein has to overcome to reach the folded state. Thermodynamic measurements are instead aimed at describing the final distribution of the unfolded and folded protein population.

The energetic barrier, $\Delta G_{\ddagger-U}$, is defined as the difference in energy between the transition state, the highest energy state, and the unfolded state (Fig. 5.1). This energy difference is defined as:

$$\Delta G_{\ddagger-U} \propto -RT \ln k_f$$

where k_f is the rate constant of folding, R the gas constant and T the temperature.

The difference in stability between the native and the unfolded state, ΔG_{N-U}, is described by the energy difference between the two states. This is defined as follows:

$$\Delta G_{N-U} = -RT \ln K_{eq}$$

$$K_{eq} = \frac{[N]}{[U]}$$

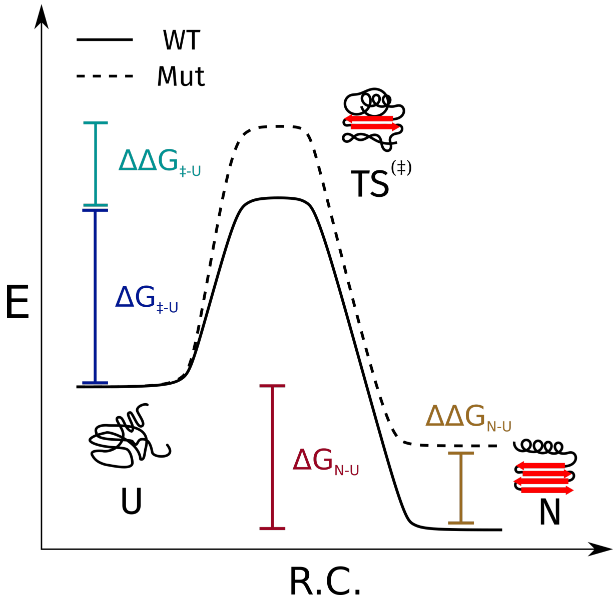

Figure 5.1: Schematic representation of the folding reaction. The reaction barrier ($\Delta G_{\ddagger-U}$) and the difference in stability (ΔG_{N-U}) of the two states are highlighted in the picture, together with the relative differences ($\Delta\Delta G_{\ddagger-U}$ and $\Delta\Delta G_{N-U}$).

where K_{eq} is the equilibrium constant, which is equal to the ratio of the two populations of folded and unfolded proteins ($[N]$ is the concentration of proteins in the native conformation, $[U]$ is the population of proteins in the unfolded conformation).

In both cases, these two energy values are defined by the difference of bonds between the two states. This allows us to define both energies as the sum of the contributions of each single residue i:

$$\Delta G = \sum_{i=1}^{n} \Delta G_i$$

As the Φ-value analysis is a perturbative approach, it is now necessary to define how the perturbation will influence both the transition state, TS, and the native state, N (Fig. 5.1). For what concerns this work, the nature of the perturbation is that of a non-disruptive mutation. Therefore, given a mutation in the residue i, we can define the new values $\Delta G'_{\ddagger-U}$ and $\Delta G'_{N-U}$. As the non-disruptive mutation consists of the removal of atoms from the original residue, the number of contacts that can be formed by the new residue will usually be inferior, as will be its contribution to the different energetic states assuming no structural rearrangement. This means that $\Delta G'_i < \Delta G_i$ for both states. The extent of bond formation in both states for the mutated residue i defines the two values $\Delta G'_{\ddagger-U}$ and $\Delta G'_{N-U}$.

The original residue i has a role in the native structure, contributing with its contacts to the energetic stability of the state itself. Also, the transition state in its classical interpretation is defined by a core of interactions that compose the minimum nucleus necessary to proceed with the formation of the native state. This core of interactions is part of the native state as well, and it is defined by the ensemble of residues that interact among them. Depending on the role of residue i in the transition state, the native contacts may be already formed or not. In the former case, the contribution of the residue to the transition state will be the same as its contribution to the native state. In the latter case, no native contacts are formed, and therefore its energetic contribution to the transition state will be zero. The comparisons between the energetic barrier and the stability of the mutant and the WT allow us to define this contribution. To do so, the following relations have to be formulated:

$$\Delta\Delta G_{\ddagger-U} = \Delta G_{\ddagger-U} - \Delta G'_{\ddagger-U}$$
$$\Delta\Delta G_{N-U} = \Delta G_{N-U} - \Delta G'_{N-U}$$

Three situation can be now delineated:

- When the residue *i* has formed all native contacts already in the transitions state, both the states will be destabilized equally by the mutation. This leads to $\Delta\Delta G_{\ddagger-U} \approx \Delta\Delta G_{N-U}$;
- If the residue *i* is not involved in the formation of the transition state, the energetic contribution will be $\Delta\Delta G_{\ddagger-U} \approx 0$;
- A partial participation of the residue *i* in the transition state leads to a partial destabilization of the same, but not in the same extent of the destabilization on the native state. This can be translated in the relation $\Delta\Delta G_{\ddagger-U} < \Delta\Delta G_{N-U}$.

Having stated all these cases, it is now possible to define the Φ-value as the ratio between the two energetic differences:

$$\Phi = \frac{\Delta\Delta G_{\ddagger-U}}{\Delta\Delta G_{N-U}}$$

The three cases enunciated above can now be depicted in terms of Φ-values:

$\Phi \approx 1$ The two differences of energy are approximately the same, and the residue *i* has already formed in the transition state all the contacts that persist in the native state;

$\Phi \approx 0$ The lack of a destabilizing effect in the transition state means that the residue *i* does not have a role in the formation of the structured regions of the transition state ensemble itself;

$0 < \Phi < 1$ The smaller destabilization of the barrier compared to the destabilization of the native state could mean that the residue *i* forms in the transition state a fraction of the contacts formed by it in the native state.

This model can be now compared and applied to the so-called "misfolding" reaction (Fig. 5.2). Similarly to the folding reaction, in the misfolding process a monomer, usually unstructured, assumes a structured conformation on the end of the fibril. As in the folding reaction, the monomer free in solution adopts a lower energy conformation upon structuring itself. However, unlike the folding reaction, the misfolding reaction is bi-molecular, as it involves both the monomer and the fibril end. As the fibril is already highly structured, it is assumed that it does not undergo a rearrangement. This reaction is indeed usually called elongation, as the fibril increases its length by one unit.

The barriers can be therefore described as follows:

$\Delta G_{\ddagger-U}$ This term correspond to the difference in energy between the transition state and the starting unfolded monomeric structure;

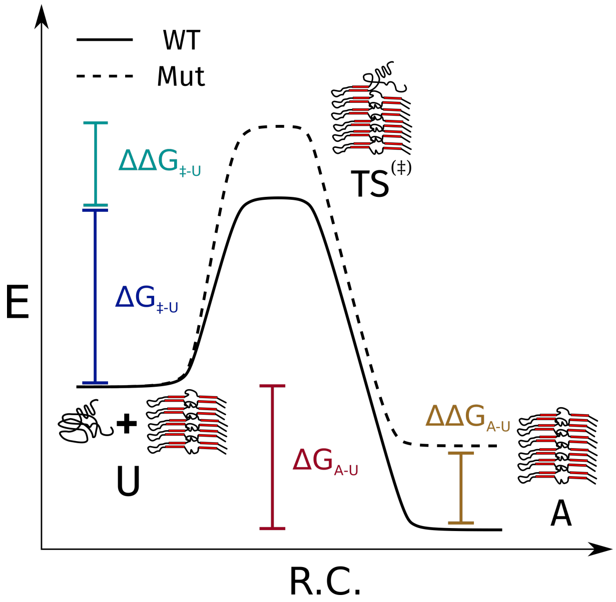

Figure 5.2: Schematic representation of the folding reaction. The reaction barrier ($\Delta G_{\ddagger-U}$) and the difference in stability (ΔG_{A-U}) of the two states are highlighted in the picture, together with the relative differences ($\Delta\Delta G_{\ddagger-U}$ and $\Delta\Delta G_{A-U}$).

ΔG_{A-U} This term corresponds to the folding ΔG_{N-U}, where the amyloid fibrillar state (A) substitutes the native state.

The differences between the mutants and the WT become $\Delta\Delta G_{\ddagger-U}$ for the elongation barrier and $\Delta\Delta G_{A-U}$ for the stability of the two states. The Φ-value is then defined as:

$$\Phi = \frac{\Delta\Delta G_{\ddagger-U}}{\Delta\Delta G_{A-U}}$$

The overall interpretation of the data is then made like in the analysis of the folding reaction, exploiting the Φ-value to define the role of the mutated residue in the transition state of the elongation reaction.

5.3. Results

5.3.1 Mutant fibrils preparation

Five different mutants of PI3K-SH3 were designed to probe the degree of structure of different part of the sequence in the transition state structural ensemble. The choice fell on the five isoleucine residues. These amino-acids are located in position 22,29,53,77 and 82 of the PI3K-SH3 WT sequence. Each mutant has one of the five isoleucine residues mutated to alanine, in order to measure how the different positions are perturbed by the same non-disruptive mutation.

WT PI3K-SH3 and the five different mutants were first tested for their ability to elongate WT seeds in a test tube. WT PI3K-SH3 seeds were used at a final concentration of 5% monomer equivalents in the presence of 200 μM protein of the different mutants. The solutions were incubated for 3 days and checked for effective fibril formation by AFM (Fig. 5.5). All the different mutants seem to elongate the WT fibrils, creating long fibrils with length superior to 5 μM in all cases except I82A. Similarly, the fibrillar morphology is maintained for all the mutants,

Table 5.1: Relative rate of elongation of PI3K-SH3 mutants. The rate is measured for both the 5 μM and 20 μM concentrations of monomer. The errors here reported correspond to a triplicate measurement of the rate on the third resonant frequency.

	Rel. rate 5μM	Rel. rate 20μM
I22A	0.018 ± 0.002	0.103 ± 0.017
I29A	0.019 ± 0.001	0.181 ± 0.003
I53A	0.003 ± 0.002	0.007 ± 0.003
I77A	1.142 ± 0.111	1.106 ± 0.082
I82A	1.601 ± 0.176	1.460 ± 0.104

with the exception of I82A that forms clusters composed by fibrils with slightly differnt height profiles. If the mutant would only elongate on the WT seeds, no change of structure should be observed. The different properties of the I82A fibrils may hint towards secondary pathways of fibril formation.

After sonication, all the fibrils reached a similar length (ca. 140 nm) showing that the sonication protocol can be employed without the risk of damaging the fibrillar preparation.

5.3.2 Elongation kinetics

The elongation rates of the mutants were probed by quartz crystal micro-balance, a surface-based technique [215]. Seeds of PI3K-SH3 WT fibrils were immobilized on all the sensors and solutions of the different mutants were flushed over different sensors to measure the elongation rate. Since the immobilization protocol does not guarantee that the same amount of seeds is immobilized on every sensor, every sensor was first elongated with PI3K-SH3 WT so to have a normalization standard for the rate of elongation of the mutant, allowing comparability between sensors.

The rate of elongation of the different mutants on the WT seeds represents a measure the height of the reaction barrier. A faster elongation rate, which is measured as higher amount of mass attaching on the sensor per unit of time, means that more monomeric protein attaches to the fibrils fixed on the sensor, acquiring the amyloid fold. The rates of elongation were measured for two different concentrations of monomeric protein, 5 and 20 μM (Fig. 5.4a). PI3K-SH3 WT and the mutants elongate at different rates, but two classes can be identified for both the concentrations (Tab. 5.1). I77A and I82A monomers elongate the WT fibrils at rates similar to the WT monomer, with a slightly increased rate of the two mutants within the same order of magnitude. I22A, I29A and I53A monomers instead elongate the WT fibrils at a much lower rate. The rates of I22A and I29A are one to two orders of magnitude slower and I53A is two to three orders of magnitude slower than the WT. The lower concentration, 5 μM, corresponds to the slower kinetics, due to the probable proximity of the monomer concentration to the critical one, explaining the deviation from the normal linear concentration dependence of elongation.

From these rates, it is possible to measure the difference of the barrier heights as the ratio of rates without the need to explicitly calculate the height of the barriers (further explanations in the Methods section). As slow rates are the result of an increased barrier of elongation, the difference between the two barriers, $\Delta\Delta G_{\ddagger-U}$, will be related to the magnitude of the decrease of the rate itself. The same can be stated for faster mutants, which have a lower barrier of elongation.

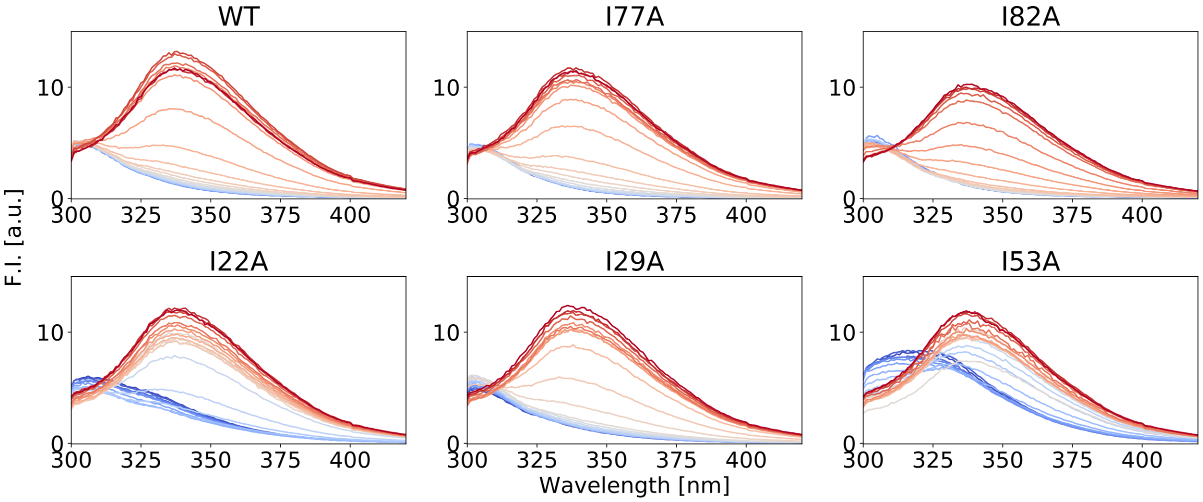

Figure 5.3: Fluorescence spectra of the denaturation series of PI3K-SH3 WT and mutants. The spectra are recorded with a bandwidth of 5 nm after excitation at 290 nm with a bandwidth of 5 nm. The gradient of color goes from blue for the lowest concentration of urea, 0M, to red for the highest, 6M.

The values of $\Delta\Delta G_{\ddagger-U}$ are reported in the Tab. 5.2. The energetic differences are calculated for both the rates coming from incubation with 5 and 20 µM. $\Delta\Delta G_{\ddagger-U}$ ranges between values around 0.3 kJ mol^{-1} for I77A, which is the mutant elongating with the rate closest to the WT one, to I53A, -12/-14 kJ mol^{-1}. For both the concentrations, the three slower mutants show an elongation barrier that differs more than 4 kJ mol^{-1} from the WT one, while I77A and I82A elongation barriers differ less than 1 kJ mol^{-1}.

5.3.3 Equilibrium denaturation

The stability of the amyloid state is measured through equilibrium denaturation experiments. In order to measure Φ-values in a meaningful manner, the PI3K-SH3 WT fibril structure has to be maintained. In order to do so, the different mutants were incubated at a total protein concentration of 200 µM with 5% monomer equivalents of sonicated WT seeds. The so-formed fibrils were then sonicated and diluted to 40 µM in different denaturant concentrations. After an equilibration time of 2 weeks, the fraction of soluble monomer was measured as ratio of fluorescence between 340 and 310 nm (Fig. 5.3).

The stability of the amyloid state is extrapolated to the absence of denaturant and, unlike the elongation barrier, it can be defined in its absolute value. Depending on the stability of the amyloid fold, which is defined by the concentration of soluble monomer, different amounts of denaturant are required to shift the reaction equilibrium so to be able to measure the concentration of the soluble monomer itself. This results in a denaturation profile of which the transition phase depends on the stability. More stable amyloid systems will present a transition phase shifted to the right. Likewise, a less stable amyloid state will result in a denaturation profile shifted to the left.

PI3K-SH3 WT and its mutants show different denaturation profiles, once again classifiable into two different categories (Fig. 5.4b). The three slowest mutants, I22A, I29A and I53A, show also a destabilization of the amyloid fold, as all the three denaturation profiles are significantly shifted to the left. I77A, which was slightly faster to elongate than the WT, here shows a slight destabilization of the amyloid fold. I82A instead once again shows a slightly more stable fold than WT. The absolute values range from -43.8 kJ mol^{-1} in the case of I53A, to -67.1 kJ mol^{-1} for I82A, and are all described in Tab. 5.2.

The difference of stability between the amyloid fold of PI3K-SH3 WT and the mutants is easily calculated as the absolute values are available (Tab. 5.2). $\Delta\Delta G_{A-U}$ is for all the mutants above 1 kJ mol^{-1}, ranging from -2.2 kJ mol^{-1} for I77A to -19.6 kJ mol^{-1} for I53A. Once again, I77A shows the smallest difference compared to the WT, while I53A shows the biggest destabilization. A positive $\Delta\Delta G_{A-U}$ is observed for I82A, fibrils of which are indeed stabilized compared to the WT by 3.8 kJ mol^{-1}.

Even though all mutants were analysed within the same framework, some notable differences can be observed in the fluorescence spectra of the amyloid state (Fig. 5.3). While the spectra of the amyloid fold (blue lines) seems to be conserved among WT, I29A, I77A and I82A, both I22A and I53A show a shoulder at ca. 340 nm, in proximity to the soluble monomer fluorescence peak. This difference could be due to several reasons. The critical concentration of the elongation reaction could be high enough to be measurable even in the absence of denaturant, meaning that the shoulder reflects the soluble monomer population. This option, in combination with the plateau at low denaturant concentration, would be incompatible with the hypothesis of linearity of the stability with the denaturant concentration. To check whether this was the case, a sample of PI3K-SH3 fibrils elongated with I53A monomers was centrifuged by ultra-centrifugation and the supernatant absorbance was measured, confirming that the soluble monomer concentration is lower than 3 µM. Similarly, secondary nucleation events could occur, but no significant surface adsorption was recorded by QCM measurements. As Ile53 is situated close to the only tryptophane residue, it could induce a slight rearrangement that could cause a different fluorescent signature for the buried fluorophore (Fig. 5.4c).

5.3.4 Φ-value calculation

By comparing the effect of the different mutants on the energetic barrier of elongation and on the stability of the amyloid fold, information on the role of the mutated residues on the transition state can be obtained. The Φ-values were calculated for both the concentrations used for the kinetic experiments and are presented in Table 5.2.

The two classes of residues identified in both the kinetic and thermodynamic analyses can be delineated also in the Φ-values. I29A and I53A show a high Φ-value (>0.6) in both cases taken in consideration (kinetic data from 5 and 20 µM injections), while I22A has a high Φ-value when measured at 5 µM, while the value at 20 µM is just slightly higher than I82A. The three slowest and least stable mutants are all located in the core of the fibril, with the side chain facing the inside of the fibril, therefore a higher contribution to the transition state was to be expected (Fig. 5.4c). The drastic effect on the elongation barrier is common to all three mutants ($\Delta\Delta G_{\ddagger-U} > 4$ kJ mol^{-1}), highlighting the fact that even such conservative mutations can significantly alter the energetic landscape of the elongation reaction. Hence, these Φ-values suggest a significant role in forming the core of the interaction that is the transition state for these three residues.

I77A and I82A are both located outside of the core, and show the lowest Φ-values. Differently from the three slowest mutants, the Φ-values are consistent in both the 5 and 20 µM measurements. In both I77A and I82A the increased elongation rates suggest a lower elongation barrier compared to the WT. The close similarity of both the elongation rate and thermodynamic stability of I77A to the WT delivers the lowest Φ-value, -0.12, which hints towards an absence of contacts for Ile77 in the transition state. The negative value arise from the opposite effect on the two

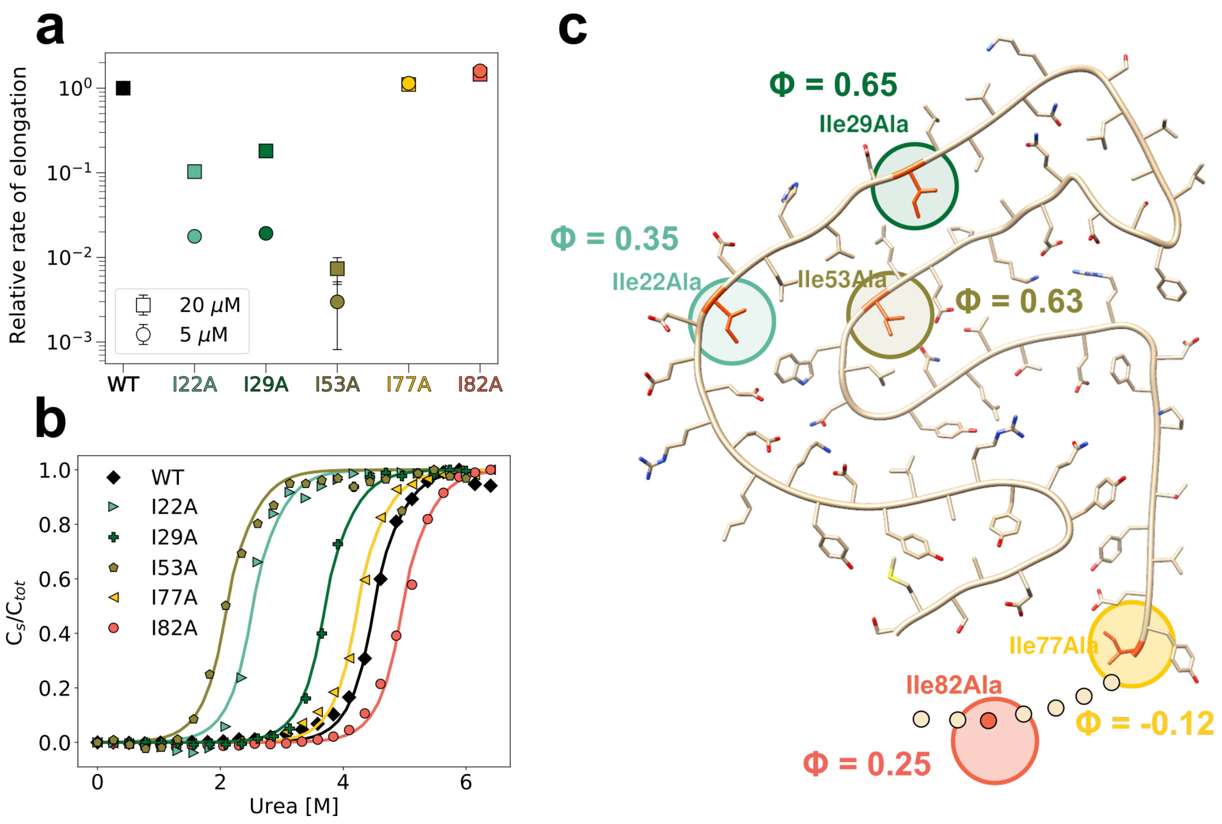

Figure 5.4: Ensemble view of relative rates, thermodynamic stability and Φ-values of the different PI3K-SH3 mutants. **a**) Rate of elongation of the mutants normalized on the elongation rate of PI3K-SH3 WT. The error bar show the standard deviation on three different measures. **b**) Denaturation profile of PI3K-SH3 fibrils in presence of urea. **c**) Amyloid conformation of a PI3K-SH3 WT monomer inside the amyloid fibril. The mutated residues are highlighted in red and marked with the corresponding color. The Φ-values corresponding to the kinetic measurements done at 20 μM for every mutant are also reported.

different measurements: while the kinetic barrier seems to be lower, the final amyloid state seems destabilised. The I82A Φ-value is closer to the I22A Φ-value at 20 μM than I22A Φ-value is to the other two slowest mutants, but its effect is diametrically opposed on both the elongation and thermodynamic behaviour. A stabilization of the transition state, identified through a lower energetic barrier, is proved above experimental error for both the concentration used in the kinetic experiment. Still, the values of $\Delta\Delta G_{\ddagger-U}$ and $\Delta\Delta G_{A-U}$ are closer to I77A, suggesting a lack of a well defined role in the formation of contacts both in the transition state and in the final amyloid conformation. The higher Φ-value can therefore be addressed as noise deriving from the small values of both $\Delta\Delta G_{\ddagger-U}$ and $\Delta\Delta G_{A-U}$.

5.4. Discussion

This first study on a small cohort of mutants introduced the methodology necessary to approach the study of the transition state of the amyloid elongation reaction. Through the use of five chemically identical mutations spacing the PI3K-SH3 sequence, a first glimpse of the spatial role of the residues in the formation of the transition state was possible.

The kinetic analysis proved to be a robust method to probe the elongation barrier, with errors lower than 15% of the measure for all the mutants except I53A. I53A is placed at the exact center

Table 5.2: Elongation barrier and amyloid stability parameters of PI3K-SH3 fibril elongation.

	$\Delta\Delta G_{\ddagger-U}$ 5 µM kJ mol^{-1}	$\Delta\Delta G_{\ddagger-U}$ 20 µM kJ mol^{-1}	ΔG_{A-U} kJ mol^{-1}	$\Delta\Delta G_{A-U}$ kJ mol^{-1}	Φ 5µM	Φ 20µM
I22A	-10.1	-5.7	-47.2	-16.1	+0.62	+0.35
I29A	-9.9	-4.3	-56.8	-6.5	+1.51	+0.65
I53A	-14.5	-12.3	-43.8	-19.6	+0.74	+0.63
I77A	+0.3	+0.3	-61.2	-2.2	-0.15	-0.12
I82A	+1.2	+0.9	-67.1	+3.8	0.31	+0.25

of the amyloid core structure, and the simple "non-disruptive" mutation seems to contradict its own nature, causing a disruptive effect on the amyloid stability. Even though the chemical difference between the two residues is minimal, it must be noted that the concept of what is a disruptive mutation in the amyloid elongation reaction has to be redefined. In the classical folding Φ-value analysis, a protein carrying a non-disruptive mutation is expected to acquire the WT fold as the mutation should not interfere with the overall pattern of interactions. Instead, in the amyloid elongation reaction, the templating effect of the existing fibril is inducing the correct amyloid fold on the mutated monomer [194, 216, 217]. This means that mutants that cannot acquire the correct conformation will not be able to elongate the pre-existing WT seeds. However, this will result in a lack of adaptability of the system, that will not allow elongation of mutants which do not form the essential number of contacts to sustain further elongation. The mutation of the Ile53 to alanine could fall indeed in this category of mutants that would not affect drastically the native fold, but that instead rises very significantly the barrier of the elongation reaction. While this new limit requires revision of the concept of non-disruptive mutation, overall this property of the elongation reaction leads to an increased accuracy of the methodology through an intrinsic selection of the usable mutants.

Another novelty emerges from the concentration dependence of the Φ-value. The apparent discrepancy of the Φ-value upon the use of different concentrations is intrinsic to the bimolecular nature of the elongation reaction. The effect of the overall concentration plays a role also in the folding of the monomer in the native structure, yet it does not affect directly the mechanism of the reaction [218]. The elongation of the amyloid fibril however can saturate upon the increase of the monomeric concentration as the templating effect of the fibril ends requires the same ends to be available for new bindings [34]. The linear relation $k_+ [M] [P]$ holds its linearity until the saturation of the fibril ends become significant. The saturating behaviour depends on the speed of the elongation reaction: the longer a monomer takes to rearrange on the fibril end, the lower will be the saturating concentration, as the fibril ends will be occupied for a longer time. The elongation kinetics of PI3K-SH3 WT were shown to be fast [187], with a linear dependency of the rate on the concentration that continues above 100 µM [162]. In this work, PI3K-SH3 WT and mutant elongation kinetics were analysed at 5 and 20 µM. It is fundamental that the comparison between rates is done at a concentration where both the WT and the mutants are in a linear regime, so that the ratio of rates is only dependent on the intrinsic difference of k_+. By keeping the concentration low, the linear regime of WT is established, while linearity is expected for the mutants, as they elongate slower (or just slightly faster) than the WT.

The apparent change of Φ-value is most probably related to the fact that in very low concentrations (namely 5 µM), a deviation from the simple linearity is the result of the stability of the fibrils themselves. As the concentration goes toward 0, the least stable fibrils will approach the rate of zero elongation faster than the most stable ones, causing the increased Φ-values that can be observed in this work for I22A, I29A and I53A at low concentration (5 μM). Likewise, the fastest and most stable mutants, I77A and I82A, do not depend on the concentration, as for both 5 and 20 μM they are far from the rate of zero elongation, like the WT. From the above it is possible to conclude that the values at 20 μM are more reliable, and that slow mutants should always be tested at higher concentrations if the rate is close to zero.

Another point that was widely discussed for the classical Φ-value analysis of protein folding is the significance of the Φ-values themselves. Different opinions about the threshold of significance were exposed and rebutted, but intervals of confidence for both kinetic and thermodynamic measurements were set over the years [55, 219–223]. All of the above agreed upon the fact that more conservative mutations are more easily interpretable, that the solution conditions should be well defined and that very high or very low changes of $\Delta\Delta G_{N-U}$ should be excluded [224]. More insights in the reliability of the measurements is conferred by multiple substitutions in the same position, which should reflect the bulkiness of the mutation in the magnitude of the Φ-value [223]. Φ-values below 0.3 are considered to correspond to residues not involved in the formation of any contact in the transition state or the average result of too drastic mutations, while values above 0.6 can be safely linked to a role in the transition state, as well as very high values of $\Delta\Delta G_{\ddagger-U}$ [220, 221]. The kinetic aspects were taken in consideration, with emphasis on the danger of extrapolating the measured rates to the absence of denaturant. The proposed solution was to use the same parameters of extrapolation of the WT (namely the m-value) or defining the Φ-value in presence of denaturant [222]. Finally, the thermodynamic aspects were the most criticized, and common agreement was reached in defining a limit of confidence for the $\Delta\Delta G_{N-U}$ values to ca. 3 kJ mol^{-1} [55, 221, 222].

A similar comparison cannot yet be drawn with the Φ-value measured of amyloid elongation, as to date only one attempt was made to apply the Φ-value analysis to protein aggregation [225]. In the pioneering work of Wang and Ferhst, mutants of p53 were compared in an effort to characterize simultaneously different rate constants, with a focus on the conformational rearrangement that the monomer should undergo before being incorporated in the fibril. The approach presented in this thesis is instead focused at the description of the transition state of the elongation reaction only, with the sound assumption of an unstructured monomeric starting state [39, 105, 106]. The details unveiled about the transition state of the reaction will hopefully build the basis for a much greater study on more amyloid system. Until then a wide comparison of significant values will be difficult to make. In light of this, some first observation can be drawn together with a comparison with the classical Φ-value analysis.

As in the classical Φ-value analysis, also in its application to the amyloid elongation reaction the final Φ-value is the result of a ratio of differences, which amplifies the different experimental errors [222]. At the same time, in the work here presented the error on the kinetic measurement is reduced, as no denaturant and therefore no extrapolation is required for both the WT and mutants. On a similar note, the value of $\Delta\Delta G_{A-U}$ is extrapolated using a global fit to share the same m-value among all the different mutants, allowing only the stability parameter to define the differences among the different mutants.

This, together with the considerations above mentioned, allows us to define the three slowest mutants, I22A, I29A and I53A as part of the transition state ensemble of interactions, with a contribution of I22A that seems to be limited if compared to the other two. Also comparing the values here measured with the limitations previously set for the classical folding reaction, all the three mutants show a significant Φ-value above the threshold of 0.3, with both $\Delta\Delta G_{\ddagger-U}$ and

$\Delta\Delta G_{A-U}$ above the values indicative for participation in the transition state. The higher Φ-values of I29A and I53A suggest a key role for the bent β-arch in the elongation reaction.

The fastest mutants, I77A and I82A, have instead a very limited contribution to the reaction barrier. This, together with the small $\Delta\Delta G_{A-U}$, suggests the lack of a role in the formation of the transition state core. I77A Φ-value is circa zero, which supports the lack of contacts in the transition state, but at the same time is obtained from both $\Delta\Delta G_{\ddagger-U}$ and $\Delta\Delta G_{A-U}$ lower than the significant threshold. I82A also shows a similar behaviour, and even though the Φ-value of ca. 0.25 may suggests a different role compared to I77A, the lack of a resolved atomic structure hints towards an experimental artefact due to the very low $\Delta\Delta G_{\ddagger-U}$.

A further remark should be addressed toward polymorphism, one of the most interesting properties of the amyloid fibrils [226, 227, 229]. The polymorphism of the amyloid fold seems to stem from the absence of an evolutionary drive, opposite to the conserved functional native fold. This means that from a single sequence, the WT, a single native functional state can be reached through the folding reaction. Contrarily, numerous different amyloid folds can be achieved even under the same solution conditions. This factor has to be kept in mind as in the classical Φ-value analysis, multiple folding pathways seem to be accessible and mutations could cause the polypetide chain to prefer one route instead of the other [219]. The eventuality of multiple folding pathways, multiplied by the different kind of amyloid states deriving from the polymorphism, would make a Φ-value analysis of the elongation reaction virtually impossible. In order to avoid such a complexity, both experimental methods and theoretical framework were adequately adjusted. First, to prevent a high degree of polymorphism, a highly pure preparation of fibrils was prepared and used for further rounds of seeding. This lead to the isolation of a single filament polymorph, previously characterized in the second part of this work. Second, the elongation reaction is a bimolecular reaction, where the fibril end, with its rigid structure, guides the soluble monomer to acquire the correct amyloid fold, as it has been demonstrated in this work and in literature [194, 216, 217]. Third, no evolutionary pressure means that there is no push to reach the same misfolded state through different routes. This, although it leads to polymorphism, probably does not allow different folding routes for the same final state.

One last observation concerns the concept of stability and amyloid fibrils. Since the definition of fibrils as the most stable thermodynamic state of proteins [30] until the recent finding that classifies crystals of amyloid peptides as the most stable state [228], the difference with the concept of stability of the native fold became more clear. On one side, the kinetically driven and functionally stable native fold, and on the other the thermodynamically driven and more stable amyloid fold [214]. This dichotomy led to a direct comparison of the stability of the two folds and to the conclusion that life happens in a regime of metastability [74]. In the work presented here, the information related to the transition state of the elongation reaction depict how a soluble unfolded monomer acquires this highly stable conformation. The same concept of metastability should be applied here. The time limitation of the experiment allows us to probe a particular amyloid fold, the one presented in the structure previously described. The effective minimum of conformational energy may be also in this case represented by another amyloid fold, leading to the fact that also this analysis is done in a metastable regime. This metastability does not hinder the final aim of the analysis, as it still holds the aim of characterizing the transition state of the single elongation reaction, regardless of the polymorph studied.

5.5. Conclusion

Through the Φ-value analysis applied to the amyloid context, it is possible to discern which residues are involved and which are not in the formation of the transition state of PI3K-SH3 elongation reaction. Three residues showed a significantly higher Φ-value, corresponding to a key position both in the final structure and in the transition state, while the two residues with lower Φ-value influence the reaction to a lower extent and are placed outside, at the edge of the fibrillar core.

Applying the Φ-value analysis approach from the classical folding to the amyloid elongation reaction, with the appropriate considerations, opens up the possibility to study the elongation reaction at an atomic level of resolution. As more data are going to be collected on new mutants and new proteins, a wider picture of the elongation state reaction is going to be drawn. More rigid standards to select meaningfully the residues and to interpret the data will be reached as it happened for protein folding, and new insights in the reaction will be obtained.

5.6. Materials and methods

5.6.1 Proteins

PI3K-SH3 WT and the mutants were expressed, produced and purified according to the protocol described in section 4.4.1. The lyophilized protein was then dissolved in the required buffer.

5.6.2 Atomic force microscopy imaging

Atomic force microscopy imaging was performed as described in the sections 4.4.3.

5.6.3 Fibril preparation

PI3K-SH3 WT fibrils were formed from protein solutions in 10 mM glycine hydrochloride pH 2 at a final concentration of 200 µM. The solutions were stirred at 42°C overnight and the presence of presence was verified through atomic force microscopy. These fibrils were then sonicated with a VialTweeter-sonotrode (Hielscher, Teltow, Germany) for 10 s at 100% amplitude in a total volume of 800 µl. The so obtained fibrils were used to seed, at 5% (equivalent monomer mass), a new solution of PI3K-SH3 WT whose concentration ranges between 200 and 300 µM. The solution was left overnight at room temperature, and investigated the day after for fibrils through atomic force microscopy and supernatant absorbance measurements. The fibrils so prepared were used both for attachment to QCM sensors and for further seeding.

PI3K-SH3 fibrils for the equilibrium denaturation experiments were prepared starting from solutions of WT and mutants at 200 µM. Identical volumes of these solutions were mixed with an equal volume of fibrils prepared as above described, to reach a final concentration of 5% in equivalent monomer mass. The fibrils were checked by atomic force microscopy.

5.6.4 Ultra-centrifugation

Sonicated fibrillar samples were spun in triplicate at 54 rpm using a TLA55 rotor for 45 mins in an Optima MAX-XP Ultracentrifuge (Beckman Coulter, Brea, CA, USA). 20 µl of supernatant were removed and its absorbance was measured using a Nanodrop instrument (Thermofisher, Waltham MA, USA). The concentration of the protein was determined at 280 nm using the extinction coefficient ε_{280}=15930 $M^{-1}cm^{-1}$.

5.6.5 Depolymerization experiments

The preparation of the samples was carried out according to the same protocol as described in Section 2.5.2. The WT and the mutants were measured with a final concentration of 40 µM in a range of denaturant of 0 to 6 M in 24 samples for the mutants I22A, I29A and I53A, and 0 to 6.4 M in 26 samples for I77A, I82A and the WT. The data of the WT and the mutants are fitted globally with the parameters m and σ, while the stability parameter ΔG_0 is fitted locally for every mutant.

5.6.6 Fluorescence measurements

The samples were measured as described in Section 2.5.5.

5.6.7 Quartz crystal microbalance kinetic measurements

Fibrils attachment to the gold sensors and the measurement technique was performed as described in section 4.4.4.

To increase the sensitivity of the measurements to the mutants with slower elongation rates, two different modifications to the protocol were made: 1) the concentration of the fibril solution which was left to incubate on the gold sensor was increased to 200 µM; 2) the concentration of the protein injected in the measurement flow was increased to 20 µM.

5.6.8 Φ-value calculation

To determine the Φ-value of a mutant, the values of $\Delta\Delta G_{A-U}$ and $\Delta\Delta G_{\ddagger-U}$ have to be calculated.

$\Delta\Delta G_{A-U}$ is defined as the difference between the stability of amyloid fold of the PI3K-SH3 WT and the mutant. Therefore for every mutant it can be defined as follows:

$$\Delta\Delta G_{A-U} = \Delta G^{WT}_{A-U} - \Delta G^{mut}_{A-U}$$

where the values of ΔG_{A-U} are extrapolated from the equilibrium denaturation analysis through the use of the cooperative model.

$\Delta\Delta G_{\ddagger-U}$ is as well defined as the difference between the energetic barrier $\Delta G_{\ddagger-U}$ of PI3K-SH3 WT and the one of the mutant. $\Delta\Delta G_{\ddagger-U}$ for every mutant is defined as follows:

$$\Delta\Delta G_{\ddagger-U} = \Delta G^{WT}_{\ddagger-U} - \Delta G^{mut}_{\ddagger-U} = -RT\,\ln\frac{k_{WT}}{k_i}$$

This expression of the formula allows avoiding the complicated estimation of the kinetic prefactor Γ that is necessary to estimate the absolute value of $\Delta G_{\ddagger-U}$.

The final Φ-value is measured for every mutant as the ratio $\Phi = \frac{\Delta\Delta G_{\ddagger-U}}{\Delta\Delta G_{A-U}}$.

5.7. Appendix

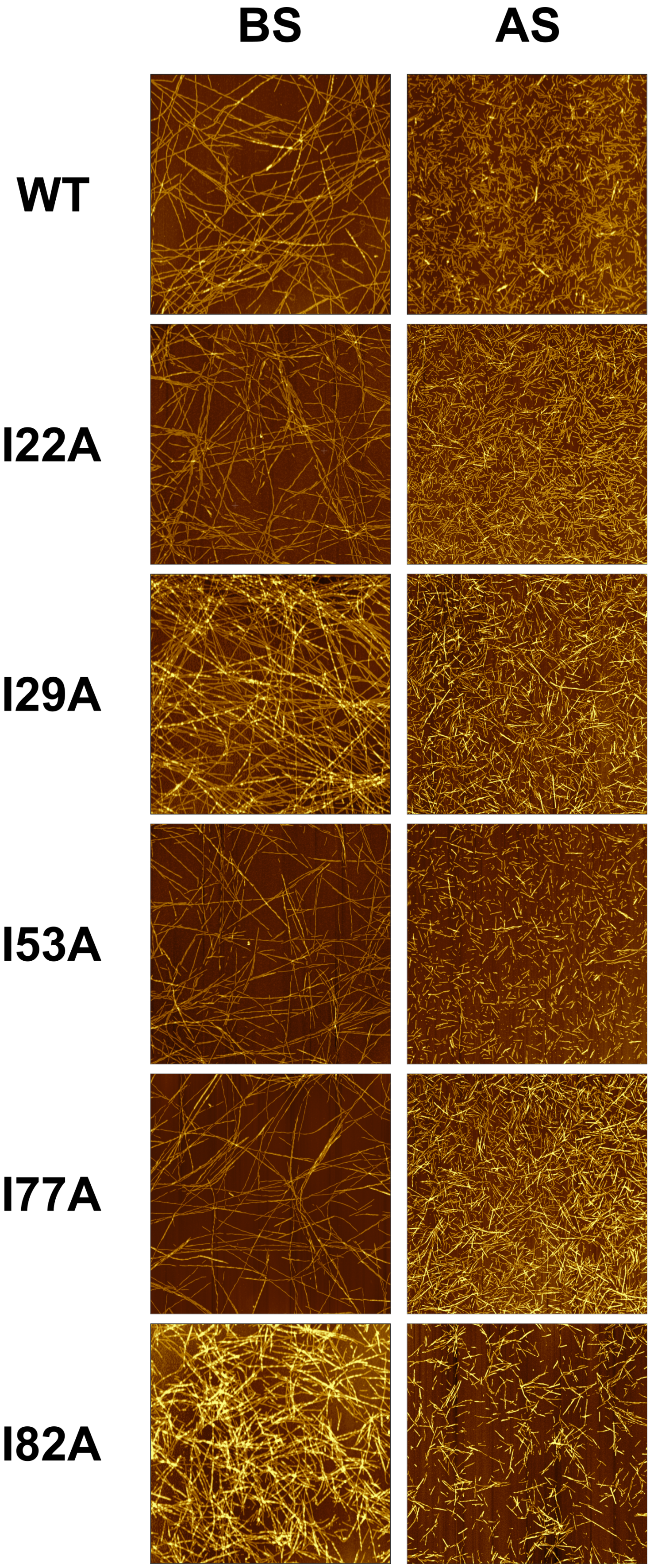

Figure 5.5: Atomic force microscopy of PI3K-SH3 WT fibrils elongated with WT and mutant monomers. The left column shows the fibrils after three days of incubation, while the right column shows the fibrils after the sonication protocol.

6

CONCLUSION AND FUTURE PERSPECTIVES

So far aymloid fibril stability has been analysed only in the framework of the isodesmic model [73–75]. This model represents the simplest description of linear polymerization and it is known to fail to depict on a theoretical level the complexity of amyloid polymerization. In Chapter 2 I describe how this became evident also on an experimental level when it failed to model the concentration dependence of the denaturation curves (Section 2.2.2). Introducing the cooperative model, which includes the concept of a nucleus, the fitting of the data improved considerably, even though the model still represents an approximation of the polymerization behaviour (Section 2.2.3). Through the comparison of the salt dependence of the stability and of the energetic barrier of PI3K-SH3 elongation, this new framework provided further evidence of the coordinated transition state between monomer and fibril end (Section 2.3.4).

Establishing the correct model to analyse the stability of the amyloid state was possible through the use of denaturants. In Chapter 3 I explored the temperature dimension to study the elongation reaction. In the first part I propose a model in which a combination of denaturant and temperature is used to extract an approximation of the stability parameter of glucagon fibrils throughout the two dimensions, while keeping the system out of equilibrium (Section 3.2.2). In the second part I model the change of the rate of growth of PI3K-SH3 fibrils with temperature as the result of the competition between the elongation and depolymerization reactions. By establishing this framework, I am able to extract an approximation of the stability parameter throughout the whole temperature range (Section 3.2.3).

In Chapter 4 the structure of PI3K-SH3 amyloid fibrils is presented. Amyloid fibril structures started just recently to be resolved and this work represents a strong contribution in the structural understanding of these assemblies. Moreover the PI3K-SH3 fibril structure allows to rationalize the biggest part of the literature results and constitutes a solid ground for further studies of fibrillation of this model protein.

Chapter 5 lays the foundations for the Φ-value analysis of the amyloid fibril elongation reaction. Through the characterization of five different mutants I obtained information about how the different residues are involved in the transition state. It was possible to define two classes of residues based on their Φ-values and on their position in the amyloid structure. Three residues show a clear role in the formation of the transition state of the reaction, while the other two do not show a significant contribution.

All together this work represents a first attempt towards a detailed description at atomic level of the energy landscape of the elongation reaction of PI3K-SH3 amyloid fibrils. A solid methodological framework has been established to study the stability of the amyloid state, while a first successful attempt was made to describe part of its transition state. Combined with the existing knowledge on the soluble state, the elongation reaction of PI3K-SH3 fibrils is closer to become the first described all along its reaction coordinate. Further efforts will be directed towards probing the role of different residues to obtain a comprehensive picture of the transition state. Finally, this framework can be applied to all amyloid system characterized by an unstructured soluble state, becoming a potentially interesting tool for disease-related proteins such as Aβ, α-synuclein or IAPP.

A

DECLARATION OF CONTRIBUTIONS

CHAPTER ONE

Article: Thermodynamics of amyloid fibril formation from chemical depolymerization

Authors: Nicola Vettore and Alexander K. Buell

Details: Phys. Chem. Chem. Phys., 2019,21, 26184-26194 doi:10.1039/C9CP04524D

Contribution: I performed the experimental work and the data analysis. The research design and the paper writing was done in collaboration with Prof. Alexander K. Büll.

CHAPTER THREE

Article: Atomic Structure of PI3-Kinase SH3 Amyloid Fibrils by Cryo-Electron Microscopy

Authors: Christine Röder, Nicola Vettore, Lena N. Mangels, Lothar Gremer, Raimond B. G. Ravelli, Dieter Willbold, Wolfgang Hoyer, Alexander K. Buell and Gunnar F. Schröder

Details: Nat Commun 10, 3754 (2019) doi:10.1038/s41467-019-11320-8

Contribution: I performed the samples preparation, the imaging through AFM, part of the biochemical assays and the whole analysis of them. I collaborated in the drafting and writing of the manuscript.

B

List of abbreviations

AFM Atomic force microscopy

ANS 8-Anilinonaphthalene-1-sulfonic acid

AS After sonication

Aβ Amyloid β

BS Before sonication

BSE Bovine spongiform encephalopathy

CD Circular dichroism

Cryo-EM Cryogenic electron microscopy

DF Double filament

ESI-MS Electrospray ionization mass spectroscopy

GdnCl Guanidine chloride

GdnSCN Guanidine tiocyanate

GlyHCl Glycine hydrochloride

HCl Hydrochloric acid

IAPP Islet amyloid polypeptide

IDP Intrinsically disordered protein

NMR Nuclear magnetic resonance

NaCl Sodium chloride

PI3K-SH3 Phosphoinositide 3-kinase SRC homology 3 domain

QCM Quarty crystal microbalance

SF Single filament

SPC-SH3 Spectrin SRC homology 3 domain

ssNMR Solid state nuclear magnetic resonance

ThT Thioflavin T

Trp Tryptophan

WT Wild type

‡ Transition state

Φ Phi-value

α**-syn** α-synuclein

BIBLIOGRAPHY

1. De Greef, T. F. A. *et al.* Supramolecular Polymerization. *Chem. Rev.* **109,** 5687–5754 (2009).
2. Sijbesma, R. P. *et al.* Reversible Polymers Formed from Self-Complementary Monomers Using Quadruple Hydrogen Bonding. *Science* **278,** 1601–1604 (1997).
3. Sevim, S. *et al.* Self-Assembled Materials and Supramolecular Chemistry within Microfluidic Environments: From Common Thermodynamic States to Non-Equilibrium Structures. *Chem. Soc. Rev.* **47,** 3788–3803 (2018).
4. Asakura, S. A Kinetic Study of in Vitro Polymerization of Flagellin. *Journal of Molecular Biology* **35,** 237–239 (1968).
5. Lodish, H. *et al.* The Dynamics of Actin Assembly. *Molecular Cell Biology. 4th edition* (2000).
6. Mitchison, T. & Kirschner, M. Dynamic Instability of Microtubule Growth. *Nature* **312,** 237–242 (1984).
7. Hoeben, F. J. M., Jonkheijm, P., Meijer, E. W. & Schenning, A. P. H. J. About Supramolecular Assemblies of π-Conjugated Systems. *Chem. Rev.* **105,** 1491–1546 (2005).
8. Aprelev, A., Liu, Z. & Ferrone, F. A. The Growth of Sickle Hemoglobin Polymers. *Biophysical Journal* **101,** 885–891 (2011).
9. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86,** 27–68 (2017).
10. Astbury, W. T. & Woods, H. J. The X-Ray Interpretation of the Structure and Elastic Properties of Hair Keratin. *Nature* **126,** 913–914 (1930).
11. Straub, F. *Actin. "Studies from the Institute of Medical Chemistry, University Szeged. Vol II." ed. A. Szent-Gyorgyi* 1942.
12. Kasai, M., Asakura, S. & Oosawa, F. The Cooperative Nature of G-F Transformation of Actin. *Biochimica et Biophysica Acta* **57,** 22–31 (1962).
13. Lowy, J. & Mcdonough, M. W. Structure of Filaments Produced By Re-Aggregation of Salmonella Flagellin. *Nature* **204,** 125–127 (1964).
14. Gerber, B. R., Asakura, S. & Oosawa, F. Effect of Temperature on the in Vitro Assembly of Bacterial Flagella. *Journal of Molecular Biology* **74,** 467–487 (1973).
15. Frieden, C. & Goddette, D. W. Polymerization of Actin and Actin-like Systems: Evaluation of the Time Course of Polymerization in Relation to the Mechanism. *Biochemistry* **22,** 5836–5843 (1983).
16. Flory, P. J. *Principles of Polymer Chemistry* 696 pp. (Cornell University Press, 1953).
17. Oosawa, F. & Asakura, S. *Thermodynamics of the Polymerization of Protein* 224 pp. (Academic Press, 1975).
18. Wagstaff, J. & Löwe, J. Prokaryotic Cytoskeletons: Protein Filaments Organizing Small Cells. *Nature Reviews Microbiology* **16,** 187–201 (2018).
19. Pollard, T. D., Blanchoin, L. & Mullins, R. D. Molecular Mechanisms Controlling Actin Filament Dynamics in Nonmuscle Cells. *Annual Review of Biophysics and Biomolecular Structure* **29,** 545–576 (2000).

20. Pollard, T. D. Rate Constants for the Reactions of ATP- and ADP-Actin with the Ends of Actin Filaments. *The Journal of Cell Biology* **103,** 2747–2754 (1986).
21. Vulevic, B. & Correia, J. J. Thermodynamic and Structural Analysis of Microtubule Assembly: The Role of GTP Hydrolysis. *Biophysical Journal* **72,** 1357–1375 (1997).
22. Izdebska, M., Zielińska, W., Grzanka, D. & Gagat, M. The Role of Actin Dynamics and Actin-Binding Proteins Expression in Epithelial-to-Mesenchymal Transition and Its Association with Cancer Progression and Evaluation of Possible Therapeutic Targets. *Biomed Res Int* **2018** (2018).
23. Parker, A. L., Kavallaris, M. & McCarroll, J. A. Microtubules and Their Role in Cellular Stress in Cancer. *Front Oncol* **4** (2014).
24. Smulders, M. M. J. *et al.* How to Distinguish Isodesmic from Cooperative Supramolecular Polymerisation. *Chemistry – A European Journal* **16,** 362–367 (2010).
25. Korevaar, P. A., Schaefer, C., de Greef, T. F. A. & Meijer, E. W. Controlling Chemical Self-Assembly by Solvent-Dependent Dynamics. *J. Am. Chem. Soc.* **134,** 13482–13491 (2012).
26. Kawai, S., Kuni, M. & Sugiyasu, K. Regression Analysis for Nucleation-Elongation Model of Supramolecular Assembly: How to Determine Nucleus Size. *J. Phys. Chem. B* (2018).
27. Maji, S. K. *et al.* Functional Amyloids as Natural Storage of Peptide Hormones in Pituitary Secretory Granules. *Science* **325,** 328–332 (2009).
28. Romero, D. & Kolter, R. Functional Amyloids in Bacteria. *Int. Microbiol.* **17,** 65–73 (2014).
29. Iadanza, M. G., Jackson, M. P., Hewitt, E. W., Ranson, N. A. & Radford, S. E. A New Era for Understanding Amyloid Structures and Disease. *Nature Reviews Molecular Cell Biology* **19,** 755–773 (2018).
30. Dobson, C. M. Protein Folding and Misfolding. *Nature* **426,** 884–890 (2003).
31. Buell, A. K. The Growth of Amyloid Fibrils: Rates and Mechanisms. *Biochem J* **476,** 2677–2703 (2019).
32. DePace, A. H. & Weissman, J. S. Origins and Kinetic Consequences of Diversity in Sup35 Yeast Prion Fibers. *Nat Struct Mol Biol* **9,** 389–396 (2002).
33. J. Young, L., Schierle, G. S. K. & F. Kaminski, C. Imaging Aβ(1–42) Fibril Elongation Reveals Strongly Polarised Growth and Growth Incompetent States. *Physical Chemistry Chemical Physics* **19,** 27987–27996 (2017).
34. Buell, A. K. *et al.* Frequency Factors in a Landscape Model of Filamentous Protein Aggregation. *Phys. Rev. Lett.* **104,** 228101 (2010).
35. Honda, R. P. & Kuwata, K. The Native State of Prion Protein (PrP) Directly Inhibits Formation of PrP-Amyloid Fibrils in Vitro. *Sci Rep* **7,** 1–11 (2017).
36. Milto, K., Botyriute, A. & Smirnovas, V. Amyloid-Like Fibril Elongation Follows Michaelis-Menten Kinetics. *PLOS ONE* **8,** e68684 (10-lug-2013).
37. Buell, A. K. *et al.* Electrostatic Effects in Filamentous Protein Aggregation. *Biophysical Journal* **104,** 1116–1126 (2013).
38. Knowles, T. P. J. *et al.* Kinetics and Thermodynamics of Amyloid Formation from Direct Measurements of Fluctuations in Fibril Mass. *PNAS* **104,** 10016–10021 (2007).
39. Zurdo, J., Guijarro, J. I., Jiménez, J. L., Saibil, H. R. & Dobson, C. M. Dependence on Solution Conditions of Aggregation and Amyloid Formation by an SH3 Domain. *Journal of Molecular Biology* **311,** 325–340 (2001).
40. Buell, A. K. *et al.* Population of Nonnative States of Lysozyme Variants Drives Amyloid Fibril Formation. *J. Am. Chem. Soc.* **133,** 7737–7743 (2011).

41. Kusumoto, Y., Lomakin, A., Teplow, D. B. & Benedek, G. B. Temperature Dependence of Amyloid β-Protein Fibrillization. *PNAS* **95,** 12277–12282 (1998).
42. Hasegawa, K., Ono, K., Yamada, M. & Naiki, H. Kinetic Modeling and Determination of Reaction Constants of Alzheimer's β-Amyloid Fibril Extension and Dissociation Using Surface Plasmon Resonance. *Biochemistry* **41,** 13489–13498 (2002).
43. Kardos, J., Yamamoto, K., Hasegawa, K., Naiki, H. & Goto, Y. Direct Measurement of the Thermodynamic Parameters of Amyloid Formation by Isothermal Titration Calorimetry. *J. Biol. Chem.* **279,** 55308–55314 (2004).
44. Carrotta, R., Manno, M., Bulone, D., Martorana, V. & Biagio, P. L. S. Protofibril Formation of Amyloid β-Protein at Low pH via a Non-Cooperative Elongation Mechanism. *J. Biol. Chem.* **280,** 30001–30008 (2005).
45. Pedersen, J. S. *et al.* The Changing Face of Glucagon Fibrillation: Structural Polymorphism and Conformational Imprinting. *Journal of Molecular Biology* **355,** 501–523 (2006).
46. Rezaei-Ghaleh, N., Ebrahim-Habibi, A., Moosavi-Movahedi, A. A. & Nemat-Gorgani, M. Role of Electrostatic Interactions in 2,2,2-Trifluoroethanol-Induced Structural Changes and Aggregation of α-Chymotrypsin. *Archives of Biochemistry and Biophysics* **457,** 160–169 (2007).
47. Škerget, K. *et al.* The Mechanism of Amyloid-Fibril Formation by Stefin B: Temperature and Protein Concentration Dependence of the Rates. *Proteins: Structure, Function, and Bioinformatics* **74,** 425–436 (2009).
48. Sabaté, R., Castillo, V., Espargaró, A., Saupe, S. J. & Ventura, S. Energy Barriers for HET-s Prion Forming Domain Amyloid Formation. *The FEBS Journal* **276,** 5053–5064 (2009).
49. Buell, A. K. *et al.* Detailed Analysis of the Energy Barriers for Amyloid Fibril Growth. *Angew. Chem. Int. Ed.* **51,** 5247–5251 (2012).
50. Milto, K., Michailova, K. & Smirnovas, V. Elongation of Mouse Prion Protein Amyloid-Like Fibrils: Effect of Temperature and Denaturant Concentration. *PLOS ONE* **9,** e94469 (18-apr-2014).
51. Cohen, S. I. A. *et al.* Distinct Thermodynamic Signatures of Oligomer Generation in the Aggregation of the Amyloid-β Peptide. *Nature Chemistry* **10,** 523–531 (2018).
52. Rodriguez, R. A., Chen, L. Y., Plascencia-Villa, G. & Perry, G. Thermodynamics of Amyloid-β Fibril Elongation: Atomistic Details of the Transition State. *ACS Chem. Neurosci.* (2017).
53. Liu, D. *et al.* Inhibitor Discovery Targeting the Intermediate Structure of β-Amyloid Peptide on the Conformational Transition Pathway: Implications in the Aggregation Mechanism of β-Amyloid Peptide. *Biochemistry* **45,** 10963–10972 (2006).
54. Wang, L., Zeng, R., Pang, X., Gu, Q. & Tan, W. The Mechanisms of Flavonoids Inhibiting Conformational Transition of Amyloid-B42 Monomer: A Comparative Molecular Dynamics Simulation Study. *RSC Adv.* **5,** 66391–66402 (2015).
55. Fersht, A. R. & Sato, S. Φ-Value Analysis and the Nature of Protein-Folding Transition States. *PNAS* **101,** 7976–7981 (2004).
56. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* 660 pp. (W. H. Freeman, 1999).
57. Baldwin, R. L. Energetics of Protein Folding. *Journal of Molecular Biology* **371,** 283–301 (2007).

58. Walker, R. A. *et al.* Dynamic Instability of Individual Microtubules Analyzed by Video Light Microscopy: Rate Constants and Transition Frequencies. *The Journal of Cell Biology* **107,** 1437–1448 (1988).
59. Ooi, T., Mihashi, K. & Kobayashi, H. On the Polymerization of Tropomyosin. *Archives of Biochemistry and Biophysics* **98,** 1–11 (1962).
60. Kasai, M., Asakura, S. & Oosawa, F. The G - F Equilibrium in Actin Solutions under Various Conditions. *Biochimica et Biophysica Acta* **57,** 13–21 (1962).
61. Oosawa, F. & Kasai, M. A Theory of Linear and Helical Aggregations of Macromolecules. *Journal of Molecular Biology* **4,** 10–21 (1962).
62. Johnson, K. A. Thermodynamics of Microtubule Assembly. *Biophysical Journal* **32,** 443–445 (1980).
63. Huecas, S. *et al.* Energetics and Geometry of FtsZ Polymers: Nucleated Self-Assembly of Single Protofilaments. *Biophysical Journal* **94,** 1796–1806 (2008).
64. Brunsveld, L., Folmer, B. J. B., Meijer, E. W. & Sijbesma, R. P. Supramolecular Polymers. *Chem. Rev.* **101,** 4071–4098 (2001).
65. Zhao, D. & S. Moore, J. Nucleation–Elongation: A Mechanism for Cooperative Supramolecular Polymerization. *Organic & Biomolecular Chemistry* **1,** 3471–3491 (2003).
66. Casellas, N. M. *et al.* From Isodesmic to Highly Cooperative: Reverting the Supramolecular Polymerization Mechanism in Water by Fine Monomer Design. *Chem. Commun.* **54,** 4112–4115 (2018).
67. Jonkheijm, P., van der Schoot, P., Schenning, A. P. H. J. & Meijer, E. W. Probing the Solvent-Assisted Nucleation Pathway in Chemical Self-Assembly. *Science* **313,** 80–83 (2006).
68. Hasegawa, K., Ono, K., Yamada, M. & Naiki, H. Kinetic Modeling and Determination of Reaction Constants of Alzheimer's Beta-Amyloid Fibril Extension and Dissociation Using Surface Plasmon Resonance. *Biochemistry* **41,** 13489–13498 (2002).
69. Williams, A. D., Shivaprasad, S. & Wetzel, R. Alanine Scanning Mutagenesis of Abeta(1-40) Amyloid Fibril Stability. *J. Mol. Biol.* **357,** 1283–1294 (2006).
70. Brännström, K., Ohman, A. & Olofsson, A. Aβ Peptide Fibrillar Architectures Controlled by Conformational Constraints of the Monomer. *PLoS ONE* **6,** e25157 (2011).
71. Xu, Y. *et al.* Steady, Symmetric, and Reversible Growth and Dissolution of Individual Amyloid-β Fibrils. *ACS Chem Neurosci* **10,** 2967–2976 (2019).
72. Xue, W.-F., Homans, S. W. & Radford, S. E. Systematic Analysis of Nucleation-Dependent Polymerization Reveals New Insights into the Mechanism of Amyloid Self-Assembly. *PNAS* **105,** 8926–8931 (2008).
73. Narimoto, T. *et al.* Conformational Stability of Amyloid Fibrils of B2-Microglobulin Probed by Guanidine-Hydrochloride-Induced Unfolding. *FEBS Letters* **576,** 313–319 (2004).
74. Baldwin, A. J. *et al.* Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *J. Am. Chem. Soc.* **133,** 14160–14163 (2011).
75. Porcari, R. *et al.* The H50Q mutation induces a 10-fold decrease in the solubility of α-synuclein. eng. *J Biol Chem* **290,** 2395–2404 (2015).
76. England, J. L. & Haran, G. Role of Solvation Effects in Protein Denaturation: From Thermodynamics to Single Molecules and Back. *Annu. Rev. Phys. Chem.* **62,** 257–277 (2011).

77. Ahmad, F. & Bigelow, C. C. Estimation of the Free Energy of Stabilization of Ribonuclease A, Lysozyme, Alpha-Lactalbumin, and Myoglobin. *J. Biol. Chem.* **257,** 12935–12938 (1982).
78. Herskovits, T. T., Jaillet, H. & Gadegbeku, B. On the Structural Stability and Solvent Denaturation of Proteins II. DENATURATION BY THE UREAS. *J. Biol. Chem.* **245,** 4544–4550 (1970).
79. Mirsky, A. E. & Pauling, L. On the Structure of Native, Denatured, and Coagulated Proteins. *PNAS* **22,** 439–447 (1936).
80. Aune, K. C. & Tanford, C. Thermodynamics of the Denaturation of Lysozyme by Guanidine Hydrochloride. I. Dependence on pH at 25°. *Biochemistry* **8,** 4579–4585 (1969).
81. Aune, K. C. & Tanford, C. Thermodynamics of the Denaturation of Lysozyme by Guanidine Hydrochloride. II. Dependence on Denaturant Concentration at 25°. *Biochemistry* **8,** 4586–4590 (1969).
82. Salahuddin, A. & Tanford, C. Thermodynamics of the Denaturation of Ribonuclease by Guanidine Hydrochloride. *Biochemistry* **9,** 1342–1347 (1970).
83. Tanford, C. in *Advances in Protein Chemistry* (eds Anfinsen, C. B., Edsall, J. T. & Richards, F. M.) 1–95 (Academic Press, 1970).
84. Greene, R. F. & Pace, C. N. Urea and Guanidine Hydrochloride Denaturation of Ribonuclease, Lysozyme, α-Chymotrypsin, and β-Lactoglobulin. *J. Biol. Chem.* **249,** 5388–5393 (1974).
85. Wingfield, P. T. Use of Protein Folding Reagents. *Current Protocols in Protein Science* **00,** A.3A.1–A.3A.4 (1995).
86. Salvi, G., Rios, P. D. L. & Vendruscolo, M. Effective Interactions between Chaotropic Agents and Proteins. *Proteins: Structure, Function, and Bioinformatics* **61,** 492–499 (2005).
87. Tateishi, J., Tashima, T. & Kitamoto, T. Practical Methods for Chemical Inactivation of Creutzfeldt-Jakob Disease Pathogen. *Microbiol. Immunol.* **35,** 163–166 (1991).
88. Safar, J., Roller, P. P., Gajdusek, D. C. & Gibbs, C. J. Conformational Transitions, Dissociation, and Unfolding of Scrapie Amyloid (Prion) Protein. *J. Biol. Chem.* **268,** 20276–20284 (1993).
89. Peretz, D. *et al.* Strain-Specified Relative Conformational Stability of the Scrapie Prion Protein. *Protein Sci* **10,** 854–863 (2001).
90. Pirisinu, L. *et al.* A New Method for the Characterization of Strain-Specific Conformational Stability of Protease-Sensitive and Protease-Resistant PrPSc. *PLoS One* **5** (2010).
91. Levitsky, D. I., Pivovarova, A. V., Mikhailova, V. V. & Nikolaeva, O. P. Thermal Unfolding and Aggregation of Actin. *The FEBS Journal* **275,** 4280–4295 (2008).
92. Vernaglia, B. A., Huang, J. & Clark, E. D. Guanidine Hydrochloride Can Induce Amyloid Fibril Formation from Hen Egg-White Lysozyme. *Biomacromolecules* **5,** 1362–1370 (2004).
93. Sasahara, K., Yagi, H., Naiki, H. & Goto, Y. Heat-Induced Conversion of B2-Microglobulin and Hen Egg-White Lysozyme into Amyloid Fibrils. *Journal of Molecular Biology* **372,** 981–991 (2007).
94. Bolisetty, S., Harnau, L., Jung, J.-M. & Mezzenga, R. Gelation, Phase Behavior, and Dynamics of β-Lactoglobulin Amyloid Fibrils at Varying Concentrations and Ionic Strengths. *Biomacromolecules* **13,** 3241–3252 (2012).
95. Pedersen, J. S. The Nature of Amyloid-like Glucagon Fibrils. *Journal of Diabetes Science and Technology SYMPOSIUM Volume 4, Issue 6, November 2010* (2010).

96. Goto, Y., Adachi, M., Muta, H. & So, M. Salt-Induced Formations of Partially Folded Intermediates and Amyloid Fibrils Suggests a Common Underlying Mechanism. *Biophys Rev,* 1–10 (2017).
97. Wymann, M. P. & Pirola, L. Structure and Function of Phosphoinositide 3-Kinases. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1436,** 127–150 (1998).
98. Fruman, D. A. *et al.* The PI3K Pathway in Human Disease. *Cell* **170,** 605–635 (2017).
99. Kurochkina, N. & Guha, U. SH3 Domains: Modules of Protein–Protein Interactions. *Biophys Rev* **5,** 29–39 (2012).
100. Koyama, S. *et al.* Structure of the Pl3K SH3 Domain and Analysis of the SH3 Family. *Cell* **72,** 945–952 (1993).
101. Weng, Z. *et al.* Structure-Function Analysis of SH3 Domains: SH3 Binding Specificity Altered by Single Amino Acid Substitutions. *Molecular and Cellular Biology* **15,** 5627–5634 (1995).
102. Batra-Safferling, R., Granzin, J., Mödder, S., Hoffmann, S. & Willbold, D. Structural Studies of the Phosphatidylinositol 3-Kinase (PI3K) SH3 Domain in Complex with a Peptide Ligand: Role of the Anchor Residue in Ligand Binding. *Biological Chemistry* **391,** 33–42 (2009).
103. Guijarro, J. I., Morton, C. J., Plaxco, K. W., Campbell, I. D. & Dobson, C. M. Folding Kinetics of the SH3 Domain of PI3 Kinase by Real-Time NMR Combined with Optical Spectroscopy. *Journal of Molecular Biology* **276,** 657–667 (1998).
104. Matsumura, Y. *et al.* Transient Helical Structure during PI3K and Fyn SH3 Domain Folding. *J. Phys. Chem. B* **117,** 4836–4843 (2013).
105. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. Amyloid Fibril Formation by an SH3 Domain. *Proc Natl Acad Sci U S A* **95,** 4224–4228 (1998).
106. Polverino de Laureto, P. *et al.* Protein Aggregation and Amyloid Fibril Formation by an SH3 Domain Probed by Limited Proteolysis. *Journal of Molecular Biology* **334,** 129–141 (2003).
107. Graña-Montes, R. *et al.* Contribution of Disulfide Bonds to Stability, Folding, and Amyloid Fibril Formation: The PI3-SH3 Domain Case. *Antioxidants & Redox Signaling* **16,** 1–15 (2011).
108. Ahn, H.-C. *et al.* NMR Characterizations of an Amyloidogenic Conformational Ensemble of the PI3K SH3 Domain. *Protein Science* **15,** 2552–2557 (2006).
109. Lim, K. H., Nagchowdhuri, P., Rathinavelan, T. & Im, W. NMR Characterization of Hydrophobic Collapses in Amyloidogenic Unfolded States and Their Implications for Amyloid Formation. *Biochemical and Biophysical Research Communications* **396,** 800–805 (2010).
110. Müller, T. D., Finan, B., Clemmensen, C., DiMarchi, R. D. & Tschöp, M. H. The New Biology and Pharmacology of Glucagon. *Physiological Reviews* **97,** 721–766 (2017).
111. Sasaki, K., Dockerill, S., Adamiak, D. A., Tickle, I. J. & Blundell, T. X-Ray Analysis of Glucagon and Its Relationship to Receptor Binding. *Nature* **257,** 751–757 (1975).
112. Gratzer, W. B., Beven, G. H., Rrattle, H. W. E. & Bradbury, E. M. A Conformational Study of Glucagon. *European Journal of Biochemistry* **3,** 276–283 (1968).
113. Swann, J. C. & Hammes, G. G. Self-Association of Glucagon, Equilibrium Studies. *Biochemistry* **8,** 1–7 (1969).
114. Onoue, S. *et al.* Novel Dry Powder Inhaler Formulation of Glucagon with Addition of Citric Acid for Enhanced Pulmonary Delivery. *International Journal of Pharmaceutics* **382,** 144–150 (2009).

115. Andersen, C. B. *et al.* Glucagon Fibril Polymorphism Reflects Differences in Protofilament Backbone Structure. *Journal of Molecular Biology* **397,** 932–946 (2010).
116. Gelenter, M. D. *et al.* The Peptide Hormone Glucagon Forms Amyloid Fibrils with Two Coexisting β-Strand Conformations. *Nature Structural & Molecular Biology* **26,** 592 (2019).
117. Pedersen, J. S., Flink, J. M., Dikov, D. & Otzen, D. E. Sulfates Dramatically Stabilize a Salt-Dependent Type of Glucagon Fibrils. *Biophysical Journal* **90,** 4181–4194 (2006).
118. Oosawa, F., Asakura, S., Hotta, K., Imai, N. & Ooi, T. G-F Transformation of Actin as a Fibrous Condensation. *Journal of Polymer Science* **37,** 323–336 (1959).
119. Borisy, G. G. & Olmsted, J. B. Nucleated assembly of microtubules in porcine brain extracts. *Science* **177,** 1196–1197 (1972).
120. Hofrichter, J., Ross, P. D. & Eaton, W. A. Kinetics and mechanism of deoxyhemoglobin S gelation: a new approach to understanding sickle cell disease. eng. *Proc Natl Acad Sci U S A* **71,** 4864–4868 (1974).
121. Oda, T., Iwasa, M., Aihara, T., Maéda, Y. & Narita, A. The nature of the globular- to fibrous-actin transition. *Nature* **457,** 441–445 (2009).
122. Magdoff-Fairchild, B. & Chiu, C. C. X-ray diffraction studies of fibers and crystals of deoxygenated sickle cell hemoglobin. *PNAS* **76,** 223–226 (1979).
123. Sunde, M. *et al.* Common Core Structure of Amyloid Fibrils by Synchrotron X-Ray Diffraction. *Journal of Molecular Biology* **273,** 729–739 (1997).
124. Greenwald, J. & Riek, R. Biology of Amyloid: Structure, Function, and Regulation. *Structure* **18,** 1244–1260 (2010).
125. Conway, K. A., Harper, J. D. & Lansbury, P. T. Fibrils Formed in Vitro from α-Synuclein and Two Mutant Forms Linked to Parkinson's Disease are Typical Amyloid. *Biochemistry* **39,** 2552–2563 (2000).
126. Gosal, W. S., Clark, A. H. & Ross-Murphy, S. B. Fibrillar beta-lactoglobulin gels: Part 1. Fibril formation and structure. *Biomacromolecules* **5,** 2408–2419 (2004).
127. Lomakin, A., Teplow, D. B., Kirschner, D. A. & Benedek, G. B. Kinetic theory of fibrillogenesis of amyloid beta-protein. eng. *Proc Natl Acad Sci U S A* **94,** 7942–7947 (1997).
128. TPJ, K. *et al.* An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science* **326,** 1533–37 (December 2009).
129. Cohen, S. I. A. *et al.* Nucleated Polymerization with Secondary Pathways. I. Time Evolution of the Principal Moments. *J. Chem. Phys.* **135,** 065105 (2011).
130. Buell, A. K. *et al.* Solution Conditions Determine the Relative Importance of Nucleation and Growth Processes in α-Synuclein Aggregation. *PNAS* **111,** 7671–7676 (2014).
131. O'Brien, E. P., Okamoto, Y., Straub, J. E., Brooks, B. R. & Thirumalai, D. Thermodynamic Perspective on the Dock- Lock Growth Mechanism of Amyloid Fibrils. *The Journal of Physical Chemistry B* **113,** 14421–14430 (2009).
132. Bard, F. *et al.* Peripherally administered antibodies against amyloid β-peptide enter the central nervous system and reducepathology in a mouse model of Alzheimer disease. *Nature Medicine* **6,** 916–919 (2000).
133. Gao, X. *et al.* Human Hsp70 Disaggregase Reverses Parkinson's-Linked α-Synuclein Amyloid Fibrils. *Molecular cell* **59,** 781–793 (5 2015).
134. Hellstrand, E., Boland, B., Walsh, D. M. & Linse, S. Amyloid beta-Protein Aggregation Produces Highly Reproducible Kinetic Data and Occurs by a Two-Phase Process. *ACS Chem. Neurosc.* **1,** 13–18 (2010).

135. J. K. Myers, C. N. P. & Scholtz, J. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Science* **4,** 2138–2148 (1995).
136. Beaven, G. H., Gratzer, W. B. & Davies, H. G. Formation and Structure of Gels and Fibrils from Glucagon. *European Journal of Biochemistry* **11,** 37–42 (1969).
137. Roeder, C. *et al.* Atomic Structure of PI3-Kinase SH3 Amyloid Fibrils by Cryo-Electron Microscopy. *Nature Communications* (2019).
138. Martin J.Parker, J. S. & R.Clarke, A. An Integrated Kinetic Analysis of Intermediates andTransition States in Protein Folding Reactions. *JMB* **253,** 771–786 (1995).
139. Kollipara, L. & Zahedi, R. P. Protein carbamylation: in vivo modification or in vitro artefact? *Proteomics* **13,** 941–944 (2013).
140. Buell, A. K. The Nucleation of Protein Aggregates-From Crystals to Amyloid Fibrils. *International Review of Cell and Molecular Biology* **329,** 187–226 (2017).
141. Korevaar, P. A. *et al.* Pathway Complexity in Supramolecular Polymerization. *Nature* **481,** 492–496 (2012).
142. Boesch, C., Bundi, A., Oppliger, M. & Wüthrich, K. 1H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. *Eur J Biochem* **91,** 209–214 (1978).
143. Shammas, S. L. *et al.* Perturbation of the Stability of Amyloid Fibrils through Alteration of Electrostatic Interactions. *Biophysical Journal* **100,** 2783–2791 (2011).
144. O'Brien, E. P., Dima, R. I., Brooks, B. & Thirumalai, D. Interactions between Hydrophobic and Ionic Solutes inAqueous Guanidinium Chloride and Urea Solutions: Lessons for Protein Denaturation Mechanism. *JACS* **129,** 7346–7353 (2007).
145. Serio, T. R. *et al.* Nucleated Conformational Conversion and the Replication of Conformational Information by a Prion Determinant. *Science* (2000).
146. Cremades, N. *et al.* Direct observation of the interconversion of normal and toxic forms of α-synuclein. eng. *Cell* **149,** 1048–1059 (2012).
147. Garcia, G. A., Cohen, S. I. A., Dobson, C. M. & Knowles, T. P. J. Nucleation-conversion-polymerization reactions of biological macromolecules with prenucleation clusters. *Physical review. E, Statistical, nonlinear, and soft matter physics* **89,** 032712 (3 2014).
148. Gremer, L. *et al.* Fibril Structure of Amyloid-β(1–42) by Cryo–Electron Microscopy. *Science* **358,** 116–119 (2017).
149. Fitzpatrick, A. W. P. *et al.* Cryo-EM Structures of Tau Filaments from Alzheimer's Disease Brain HHS Public Access. *Nature* **547,** 185–190 (2017).
150. Michaels, T. C. *et al.* Reaction rate theory for supramolecular kinetics: application to protein aggregation. *Molecular Physics* **116,** 3055–3065 (2018).
151. Massi, F., Klimov, D., Thirumalai, D. & Straub, J. E. Charge states rather than propensity for β-structure determine enhanced fibrillogenesis in wild-type Alzheimer's β-amyloid peptide compared to E22Q Dutch mutant. *Protein science* **11,** 1639–1647 (2002).
152. Klimov, D. K. & Thirumalai, D. Dissecting the assembly of Aβ16–22 amyloid peptides into antiparallel β sheets. *Structure* **11,** 295–307 (2003).
153. Gunnarsson, G., Joensson, B. & Wennerstroem, H. Surfactant association into micelles. An electrostatic approach. *J. Phys. Chem.* **84,** 3114–3121 (1980).
154. Kegel, W. K. & van der Schoot, P. Competing Hydrophobic and Screened-Coulomb Interactions in Hepatitis B Virus Capsid Assembly. *Biophys. J.* **86,** 3905–3913 (2004).
155. Kar, K., Jayaraman, M., Sahoo, B., Kodali, R. & Wetzel, R. Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. eng. *Nat Struct Mol Biol* **18,** 328–336 (2011).

156. Mirbaha, H. *et al.* Inert and seed-competent tau monomers suggest structural origins of aggregation. *eLife,* e36584 (2018).
157. Debye, P. & Hückel, E. Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Phys Zeitschr* **24,** 185–206 (1923).
158. Asakura, S., Kasai, M. & Oosawa, F. The Effect of Temperature on the Equilibrium State of Actin Solutions. *Journal of Polymer Science* **44,** 35–49 (1960).
159. Ikenoue, T. *et al.* Cold Denaturation of α-Synuclein Amyloid Fibrils. *Angewandte Chemie International Edition* **53,** 7799–7804 (2014).
160. Herkert, L. *et al.* Pathway Control in Cooperative vs. Anti-Cooperative Supramolecular Polymers. *Angewandte Chemie International Edition* **0.**
161. Jeppesen, M. D., Hein, K., Nissen, P., Westh, P. & Otzen, D. E. A Thermodynamic Analysis of Fibrillar Polymorphism. *Biophysical Chemistry* **149,** 40–46 (2010).
162. Coden, A. *Biophysical characterization of amyloid fibril elongation: a surface-based biosensing approach.* 2018.
163. Knowles, T. P. J., Vendruscolo, M. & Dobson, C. M. The Amyloid State and Its Association with Protein Misfolding Diseases. *Nat. Rev. Mol. Cell Biol.* **15,** 384–396 (2014).
164. Kayed, R. *et al.* Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis. *Science* **300,** 486–489 (2003).
165. Cohen, S. I. A. *et al.* Proliferation of Amyloid-B42 Aggregates Occurs through a Secondary Nucleation Mechanism. *PNAS* **110,** 9758–9763 (2013).
166. Pepys, M. B. *et al.* Human Lysozyme Gene Mutations Cause Hereditary Systemic Amyloidosis. *Nature* **362,** 553 (1993).
167. Kheterpal, I., Williams, A., Murphy, C., Bledsoe, B. & Wetzel, R. Structural Features of the Abeta Amyloid Fibril Elucidated by Limited Proteolysis. *Biochemistry* **40,** 11757–11767 (2001).
168. Hoshino, M. *et al.* Mapping the Core of the B$_2$-Microglobulin Amyloid Fibril by H/D Exchange. *Nature Structural & Molecular Biology* **9,** 332–336 (2002).
169. Fitzpatrick, A. W. P. *et al.* Atomic Structure and Hierarchical Assembly of a Cross-β Amyloid Fibril. *PNAS,* 201219476 (2013).
170. Tycko, R. Molecular Structure of Aggregated Amyloid-β: Insights from Solid-State Nuclear Magnetic Resonance. *Cold Spring Harb Perspect Med* **6,** a024083 (2016).
171. Wasmer, C. *et al.* Amyloid Fibrils of the HET-s(218–289) Prion Form a β Solenoid with a Triangular Hydrophobic Core. *Science* **319,** 1523–1526 (2008).
172. Radamaker, L. *et al.* Cryo-EM Structure of a Light Chain-Derived Amyloid Fibril from a Patient with Systemic AL Amyloidosis. *Nat Commun* **10,** 1–8 (2019).
173. Zhang, W. *et al.* Heparin-Induced Tau Filaments Are Polymorphic and Differ from Those in Alzheimer's and Pick's Diseases. *eLife,* 43584 (2019).
174. Falcon, B. *et al.* Structures of Filaments from Pick's Disease Reveal a Novel Tau Protein Fold. *Nature* **561,** 137–140 (2018).
175. Li, B. *et al.* Cryo-EM of Full-Length α-Synuclein Reveals Fibril Polymorphs with a Common Structural Kernel. *Nature Communications* **9,** 3609 (2018).
176. Guerrero-Ferreira, R. *et al.* Cryo-EM Structure of Alpha-Synuclein Fibrils. *eLife* **7,** e360402 (2018).
177. Li, Y. *et al.* Amyloid Fibril Structure of α-Synuclein Determined by Cryo-Electron Microscopy. *Cell Research* **28,** 897–903 (2018).
178. Liberta, F. *et al.* Cryo-EM Structure of an Amyloid Fibril from Systemic Amyloidosis. *Nature communications* **10,** 1104 (2019).

179. Swuec, P. *et al.* Cryo-EM Structure of Cardiac Amyloid Fibrils from an Immunoglobulin Light Chain (AL) Amyloidosis Patient. *bioRxiv,* 444901 (2018).
180. Falcon, B. *et al.* Tau Filaments from Multiple Cases of Sporadic and Inherited Alzheimer's Disease Adopt a Common Fold. *Acta Neuropathologica* **136,** 699–708 (2018).
181. Arosio, P. *et al.* Kinetic Analysis Reveals the Diversity of Microscopic Mechanisms through Which Molecular Chaperones Suppress Amyloid Formation. *Nature Communications* **7,** 10948 (2016).
182. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. SH2 and SH3 Domains: Elements That Control Interactions of Cytoplasmic Signaling Proteins. *Science* **252,** 668–674 (1991).
183. Yu, H. *et al.* Structural Basis for the Binding of Proline-Rich Peptides to SH3 Domains. *Cell* **76,** 933–945 (1994).
184. Jiménez, J. L. *et al.* Cryo-electron Microscopy Structure of an SH3 Amyloid Fibril and Model of the Molecular Packing. *The EMBO Journal* **18,** 815–821 (1999).
185. Neudecker, P. *et al.* Structure of an Intermediate State in Protein Folding and Aggregation. *Science* **336,** 362–366 (2012).
186. Bucciantini, M. *et al.* Inherent Toxicity of Aggregates Implies a Common Mechanism for Protein Misfolding Diseases. *Nature* **416,** 507–511 (2002).
187. Carulla, N. *et al.* Molecular Recycling within Amyloid Fibrils. *Nature* **436,** 554 (2005).
188. Orte, A. *et al.* Direct Characterization of Amyloidogenic Oligomers by Single-Molecule Fluorescence. *PNAS* **105,** 14424–14429 (2008).
189. Bayro, M. J. *et al.* High-Resolution MAS NMR Analysis of PI3-SH3 Amyloid Fibrils: Backbone Conformation and Implications for Protofilament Assembly and Structure. *Biochemistry* **49,** 7474–7484 (2010).
190. Scheres, S. H. W. RELION: Implementation of a Bayesian Approach to Cryo-EM Structure Determination. *Journal of Structural Biology* **180,** 519–530 (2012).
191. He, S. & Scheres, S. H. W. Helical Reconstruction in RELION. *Journal of Structural Biology* **198,** 163–176 (2017).
192. Kimanius, D., Forsberg, B. O., Scheres, S. H. W. & Lindahl, E. Accelerated Cryo-EM Structure Determination with Parallelisation Using GPUs in RELION-2. *eLife,* 18722 (2016).
193. Kyte, J. & Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *Journal of Molecular Biology* **157,** 105–132 (1982).
194. Riek, R. & Eisenberg, D. S. The Activities of Amyloids from a Structural Perspective. *Nature* **539,** 227–235 (2016).
195. Zondlo, N. J. Aromatic-Proline Interactions: Electronically Tunable CH/π Interactions. *Accounts of Chemical Research* **36,** 1039–1049 (2013).
196. Unterreitmeier, S. *et al.* Phenylalanine Promotes Interaction of Transmembrane Domains via GxxxG Motifs. *Journal of Molecular Biology* **374,** 705–718 (2007).
197. Carulla, N. *et al.* Experimental Characterization of Disordered and Ordered Aggregates Populated during the Process of Amyloid Fibril Formation. *Proceedings of the National Academy of Sciences* (2009).
198. Ventura, S. *et al.* Short Amino Acid Stretches Can Mediate Amyloid Formation in Globular Proteins: The Src Homology 3 (SH3) Case. *PNAS* **101,** 7258–7263 (2004).
199. Kozma, D., Simon, I. & Tusnády, G. E. CMWeb: An Interactive on-Line Tool for Analysing Residue–Residue Contacts and Contact Prediction Methods. *Nucleic Acids Res* **40,** W329–W333 (2012).

200. Buell, A. K. *et al.* Position-Dependent Electrostatic Protection against Protein Aggregation. *ChemBioChem* **10,** 1309–1312 (2009).
201. Iadanza, M. G. *et al.* The Structure of a Β2-Microglobulin Fibril Suggests a Molecular Basis for Its Amyloid Polymorphism. *Nature communications* **9,** 4517 (2018).
202. Collins, S. R., Douglass, A., Vale, R. D. & Weissman, J. S. Mechanism of Prion Propagation: Amyloid Growth Occurs by Monomer Addition. *PLOS Biology* **2,** e321 (2004).
203. Ventura, S., Lacroix, E. & Serrano, L. Insights into the Origin of the Tendency of the PI3-SH3 Domain to Form Amyloid Fibrils. *Journal of Molecular Biology* **322,** 1147–1158 (2002).
204. Zheng, S. Q., Palovcak, E., Armache, J.-P., Cheng, Y. & Agard, D. A. MotionCor2: Anisotropic Correction of Beam-Induced Motion for Improved Single-Particle Electron Cryo-Microscopy. *Nature Methods* **14,** 331–332 (2017).
205. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and Accurate Defocus Estimation from Electron Micrographs. *Journal of Structural Biology. Recent Advances in Detector Technologies and Applications for Molecular TEM* **192,** 216–221 (2015).
206. Pettersen, E. F. *et al.* UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry* **25,** 1605–1612 (2004).
207. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and Development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66,** 486–501 (2010).
208. Emsley, P. & Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D: Biological Crystallography* **60,** 2126–2132 (2004).
209. Adams, P. D. *et al.* PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallographica Section D: Biological Crystallography* **66,** 213–221 (2010).
210. Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. The Protein Folding Problem. *Annual Review of Biophysics* **37,** 289–316 (2008).
211. Matouschek, A., Kellis, J. T., Serrano, L. & Fersht, A. R. Mapping the Transition State and Pathway of Protein Folding by Protein Engineering. *Nature* **340,** 122 (1989).
212. Fersht, A. R. Nucleation Mechanisms in Protein Folding. *Current Opinion in Structural Biology* **7,** 3–9 (1997).
213. Astbury, W. T., Dickinson, S. & Bailey, K. The X-Ray Interpretation of Denaturation and the Structure of the Seed Globulins. *Biochemical Journal* **29,** 2351–2360.1 (1935).
214. Eichner, T. & Radford, S. E. A Diversity of Assembly Mechanisms of a Generic Amyloid Fold. *Molecular Cell* **43,** 8–18 (2011).
215. Buell, A. K., Dobson, C. M. & Welland, M. E. in *Amyloid Proteins* 101–119 (Humana Press, 2012).
216. Mar, C. D., Greenbaum, E. A., Mayne, L., Englander, S. W. & Woods, V. L. Structure and Properties of α-Synuclein and Other Amyloids Determined at the Amino Acid Level. *PNAS* **102,** 15477–15482 (2005).
217. Eichner, T., Kalverda, A. P., Thompson, G. S., Homans, S. W. & Radford, S. E. Conformational Conversion during Amyloid Formation at Atomic Resolution. *Molecular Cell* **41,** 161–172 (2011).
218. Samiotakis, A., Wittung-Stafshede, P. & Cheung, M. S. Folding, Stability and Shape of Proteins in Crowded Environments: Experimental and Computational Approaches. *International Journal of Molecular Sciences* **10,** 572–588 (2009).
219. Ozkan, S. B., Bahar, I. & Dill, K. A. Transition States and the Meaning of Φ-Values in Protein Folding Kinetics. *Nat Struct Mol Biol* **8,** 765–769 (2001).

220. Northey, J. G. B., Maxwell, K. L. & Davidson, A. R. Protein Folding Kinetics Beyond the Φ Value: Using Multiple Amino Acid Substitutions to Investigate the Structure of the SH3 Domain Folding Transition State. *Journal of Molecular Biology* **320,** 389–402 (2002).
221. Sánchez, I. E. & Kiefhaber, T. Origin of Unusual Φ-Values in Protein Folding: Evidence Against Specific Nucleation Sites. *Journal of Molecular Biology* **334,** 1077–1085 (2003).
222. Rios, M. A. D. L. *et al.* On the Precision of Experimentally Determined Protein Folding Rates and Φ-Values. *Protein Science* **15,** 553–563 (2006).
223. Zarrine-Afsar, A., Lin, S. L. & Neudecker, P. Mutational Investigation of Protein Folding Transition States by Φ-Value Analysis and beyond: Lessons from SH3 Domain Folding. *Biochem. Cell Biol.* **88,** 231–238 (2010).
224. Gianni, S. & Jemth, P. Conserved Nucleation Sites Reinforce the Significance of Phi Value Analysis in Protein-Folding Studies. *IUBMB Life* **66,** 449–452 (2014).
225. Wang, G. & Fersht, A. R. Mechanism of Initiation of Aggregation of P53 Revealed by Φ-Value Analysis. *Proc Natl Acad Sci U S A* **112,** 2437–2442 (2015).
226. Fändrich, M. *et al.* Amyloid Fibril Polymorphism - a Challenge for Molecular Imaging and Therapy. *J Intern Med* **283,** 218–237 (2018).
227. Tycko, R. Amyloid Polymorphism: Structural Basis and Neurobiological Relevance. *Neuron* **86,** 632–645 (2015).
228. Reynolds, N. P. *et al.* Competition between Crystal and Fibril Formation in Molecular Mutations of Amyloidogenic Peptides. *Nat Commun* **8,** 1–10 (2017).
229. Close, W. *et al.* Physical Basis of Amyloid Fibril Polymorphism. *Nature Communications* **9,** 699 (2018).