Modelling the Kinetics of Amyloid Fibril Nucleation

Raffaela Cabriolu

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Physics and Astronomy

September 2012

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the others authors to this work has been explicitly indicated below.

Chapter 3 of my thesis is based on the following publication:
Atomistic theory of amyloid fibril nucleation. R. Cabriolu, D. Kashchiev and S. Auer, J. Chem. Phys. 133, 225101 (2010)
In this work I wrote a computer program to perform all the numerical calculations, I prepared all the figures in the article and I have contributed to the development of the theory.
S. Auer supervised the work, contributed to the development of the theory and wrote the article.
D. Kashchiev supervised the work, developed the theory and wrote the article.

Chapter 4 is based on the following publication:
Amyloid Fibrillation Kinetics: Insight from Atomistic Nucleation Theory. R. Cabriolu and S. Auer, J. Mol. Biol. 411, 275 (2011)
In this work I performed all the numerical calculations and prepared all the figures in the article.
S. Auer supervised the work and wrote the article.

Chapter 5 is based on the following publication:
Size Distribution of Amyloid Nanofibrils. R. Cabriolu and, D. Kashchiev and S. Auer, J. Biophysical Journal, 101, 2232 (2011)
In this work I wrote a computer program to perform all the numerical calculations and I prepared all the figures in the article.
S. Auer has supervised the work and wrote the article.
D. Kashchiev has supervised the work and wrote the article.

Chapter 6 of my thesis is based on the following publication:
Breakdown of nucleation theory for crystals with strongly anisotropic interactions between molecules. R. Cabriolu, D. Kashchiev and S. Auer, Submitted.

In this work I developed and wrote the Kinetic Monte Carlo code, I performed all the simulations, prepared all figures and wrote the article.
S. Auer has supervised the work and wrote the article.
D. Kashchiev has supervised the work and wrote the article.

Acknowledgements

First af all, I would like to thank my main supervisor Dr. Stefan Auer for having given me the opportunity to study this exciting and successful subject, and also for his support during the entire duration of my PhD. A particular thanks goes to my external supervisor Prof. Dimo Kashchiev who introduced me to the wonderful world of the nucleation theory and who had the patience to constantly guide me despite the distance. Dimo and his wife Mimi, moreover, played a really important personal role in my PhD experience. A warm thanks to every friend that has contributed to make my time in Leeds wonderful and fully enjoyable. Many thanks to all my PhD colleagues in the Soft Matter group, to Dr. Sarah Harris, Dr. Emanuele Paci and Dr. Piero Ricchiuto for having created a warm and friendly environment around me. A big thanks to Prof. Peter Olmsted, for all its personal and scientific support. Thanks to Glenys who with her sympathy and her efficient work made the bureaucratic aspects of my PhD easy and trouble-free. A particular thanks goes to my family and to Jason, because they always supported me with their love and understanding in every choice. A particular thanks goes to Dr. Elia Ibba for his relentless support. I would like also to thank the former supervisor of my Master Thesis, Dr. Pietro Ballone, who has always been willing to help and to advise me. A particular thought goes to Sardinia and all my friends there of which I have missed every feature.

Abstract

A kinetic theory has been developed within the framework of pre-existing nucleation theory and applied, for the first time, to investigate the one-step formation of amyloid fibrils. Atomistic Nucleation Theory (ANT) for fibrils, in particular, has been successfully applied to model real peptides and proteins, in order to investigate at the molecular level the nucleation of amyloid fibrils from a homogeneous solution. Kinetic parameters predicted by the theory, such as the nucleation rates, have been compared successfully to the results of experiments. The present theoretical study has shown that variations in solubility are the primary origin of the changes in the nucleation rates between a protein and its point-mutations. The same ANT approach allows the analysis of the fibril size distribution, whose results, once again, are consistent with experimental observations. In the last stage of the investigation, computer simulations have been carried out to test selected assumptions underlying the theory. For the first time, the nucleation of strongly anisotropic systems has been investigated using kinetic Monte Carlo (KMC) simulations. Novel and unexpected features, never discussed before in either experiments or simulations studies, have been revealed by the simulations. Although obtained within the study of amyloid fibrils nucleation, these last results are of general validity, providing useful insight on the nucleation of all systems whose molecules interact via strongly anisotropic forces.

Contents

List of Figures

List of Tables

Symbols and abbreviations

C_1	Concentration of β-strands
C_e	Equilibrium concentration of β-strands
C_n	Equilibrium fibril size distribution
E	Strong binding energy
E_h	Weak (Hydrophobic) binding energy
f_n	Attachment frequency of a β-strand to a n-sized fibril
J	Stationary nucleation rate
k_B	Boltzmann's constant
k_BT	Thermal energy
M	Size of the a large supernucleus fibril
n	Number of molecules in the fibril
n^*	Number of molecules in the nucleus
s	Dimensionless supersaturation
T	Absolute Temperature
t_l	lag-time of nucleation
V	Volume of the solution
W	Work of fibril formation
w	Dimensionless work of fibril formation
W^*	Work of fibril nucleus formation
w^*	Dimensionless work of fibril nucleus formation
X_n	Stationary nanofibril size distribution
z	Zeldovich factor
$\Delta\mu$	Supersaturation
ε	Hydrogen bond energy unit
ε_h	Hydrophobicity mediated aminoacid bond
σ	Specific surface energy perpendicular to the i direction
σ_h	Specific surface energy perpendicular to the m direction
ψ	Dimensionless strong broken bond energy
ψ_h	Dimensionless weak broken bond energy
$k^+(j)$	Attachment frequency
$k^-(j)$	Detachment frequency
ω	Isotropic molecular energy within a cluster
P_2	Dimer growth probability
CNT	Classical Nucleation Theory
ANT	Atomistic Nucleation Theory
KMC	Kinetic Monte Carlo

Chapter 1

Introduction

Proteins seem to have an intrinsic tendency to assemble into highly ordered fibrillar aggregates (28), known as amyloid fibrils. The self-assembly of proteins into amyloid fibrils is an important phenomenon with wide implications ranging from human disease to nanoscience. Amyloid fibril formation is in fact associated with a significant number of neurological and systemic diseases (22) including Alzheimer, Huntington and Parkinson's diseases, and the application of amyloid fibrils as molecular building blocks in bio-sensors, tissue engineering and antibacterial agents has been demonstrated (43). In addition, different studies (e.g. Ref. (22; 36)) have identified living systems such as bacteria, fungi and insects which exploit the extraordinary properties of amyloid fibers for functional roles. Escherichia coli and Salmonella, for instance, employ amyloid fibers to create a matrix that modulates surface adhesion to support their colony formation (22; 36). Fungi use amyloid fibers to lower the water surface tension to enable the spore formation (22). Moths and spiders exploit the strength of the fibers to protects their egg shells and to build nest and webs (36). In mammals the amyloid fiber is used as a template for the synthesis of melanin (11).

An essential point in the current understanding of amyloids is that different peptides and proteins, unrelated in sequence and structure, have been shown to convert into amyloid fibrils (22). The main insight into the amyloid structures formed by proteins comes from structural biology experiments, whose results have been confirmed and complemented by computer simulations (4; 7; 65; 102; 109; 110; 124; 136; 145; 152) (see Section 1.2). Proteins self-aggregate into these highly ordered fibrillar structures which may assume different conformations depending on detailed amino-acid sequence, and on a variety

of environmental factors that, however, are still largely unknown. Despite the variety of conformations, fibrillar aggregates are classified as amyloids if they share a common cross-β structure formed by intertwined layers of β-sheets as shown by Fig. 1.1 from Ref. (69). Within each β-sheet the proteins or peptides (coloured arrows in the picture) are in an extended conformation that allows them to form a strong hydrogen bonding network (4; 65; 102; 109; 110; 111; 124; 136; 145; 152) with neighbouring proteins or peptides. The successive layering of β-sheets is caused by much weaker hydrophobicity-mediated interactions between the proteins or peptides. In amyloid fibrils the β-sheets are aligned parallel to the fibril elongation axis, and the proteins or peptides are perpendicular to it (4; 109; 110; 111; 136).

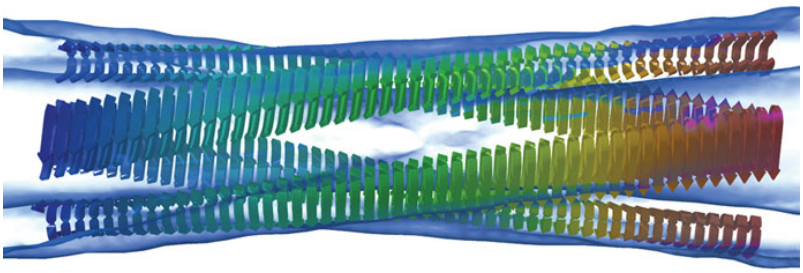

Figure 1.1: Amyloid fibril structure obtained by cryo-electron microscopy with a cross-β structure modelled into the electron density map based on the work of J.L. Jimenez et al. (69). Each arrow in the picture represents a peptide. The peptides are linked to their neighbours by hydrogen bonds to form a β-sheet that stack on top of on other β-sheet. The interwining of the saked β-sheets give the amyloid fibril.

It is now well established (4; 5; 31; 32; 33; 38; 40; 41; 54; 60; 61; 67; 98; 100; 101; 109; 110; 111; 128; 157; 163) that fibrillar protein aggregates form through a nucleation mechanism. The amyloid fibril formation, in particular, is considered a nucleation-mediated process because of the activated and reversible character of the process itself as will be explained later in Section 1.2. Nucleation, in turn, is a phenomenon characteristic of first order phase transitions. In the present case, in particular, the phase transition connects a homogeneous solution of individual peptides, with a population of fibrils solvated in water. The study of the amyloid formation process, therefore, is a complex and multidisciplinary subject, whose comprehensive description requires notions on the structure of peptides,

proteins and fibres, on their mutual interactions in solution, together with theoretical concepts concerning nucleation theory as well as a working knowledge of statistical mechanics, simulation, and computational methods. To set the stage for the discussion that follows, I briefly introduce here the major ingredients that I will use in the development of my investigations.

1.1 Nucleation Theory

Understanding phase transitions is extremely important not only for technology and industrial applications, but also for the interpretation of a wide variety of natural processes, whose detailed knowledge could open the way to their control and exploitation. Nucleation is a fundamental aspect of every first order phase transition, and nucleation theory gives an exhaustive theoretical description of the transformation kinetics. Early versions of nucleation theory have been primarily developed to model the formation of liquid droplets from a vapour phase. Later versions covered the case of crystal nucleation from the liquid phase, or the formation of an ordered ad-layer on top of a crystalline substrate.

The self-assembly of peptides and proteins into particular structures called amyloid fibrils represents a transition from a fluid phase to a new phase characterised by a substantial amount of ordering, and by negligible intra-fibril diffusion. In many respects, the transition is the analogue of a fluid to solid phase change, and experiments show that the transition is first order. For this reason, in my thesis the formation of amyloid fibrils is investigated within the framework of nucleation theory.

At equilibrium every system exhibits the phase at which its Gibbs free energy is minimal, and, in such a case, the system is in a stable thermodynamic state. The identity of the stable phase can change upon changing parameters such as the system temperature and pressure, and, to maintain its equilibrium character, the system undergoes a phase transition. There are different types of phase transition, depending on the order of discontinuity that appears in some thermodynamic variables. A first order phase transition is revealed by a discontinuity in the first derivative of the free energy with respect to some thermodynamic variable. Fig. 1.2 shows a plot of the van der Waals equation that provides a good model to describe the gas-liquid phase transition ending in a critical point at T_c.

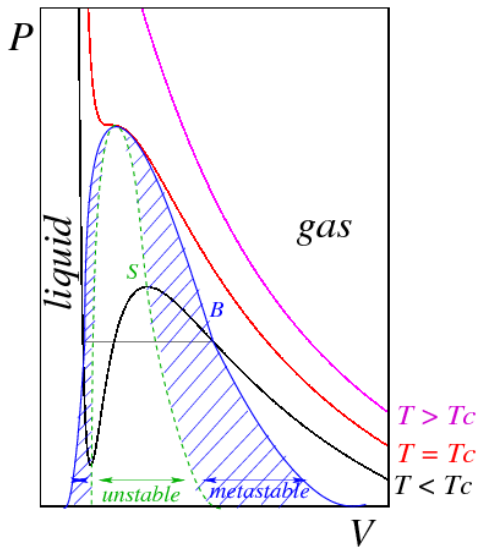

Figure 1.2: Volume dependence of the pressure for a van der Waals fluid.

As can be seen from the graph, the van der Waals (vdW) equation shows no phase transition for $T > T_c$ (magenta line) because the system is in a thermodynamically and mechanically stable gas phase, while at $T = T_c$ (red line) the liquid and gas phase are indistinguishable from each other. At $T < T_c$ (black line), the liquid and vapour phases coexist, separated by an anomalous region corresponding to a density discontinuity, pointing to a first order phase transition. The vdW isotherm under T_c develops a loop, with regions that are thermodynamically metastable, where nucleation processes take place (shaded regions in the graph), and regions in which the system is mechanically and thermodynamically unstable, where the phase separation occurs by spinodal decomposition (white region in the graph). Since the part of the isotherm in the white region of the graph is unstable and thus forbidden, the system must find some other way to get from the right side of the graph to the left not following the vdW isotherm. It does so by means of a discontinuity in the volume (or, equivalently, density) following the straight black line given by the so called "Maxwell construction"; in short, a first order phase transition is taking place (77).

As mentioned, the shaded region in the vdW graph identifies the portion of the phase space in which the system is metastable, and the first order transition occurs through nucleation and growth of the stable phase. During nucleation, the system lowers its free energy and moves towards a stable state corresponding

to the new liquid phase on the left side of the barrier in the graph 1.2, upon crossing a free energy barrier schematically represented in Fig. 1.3

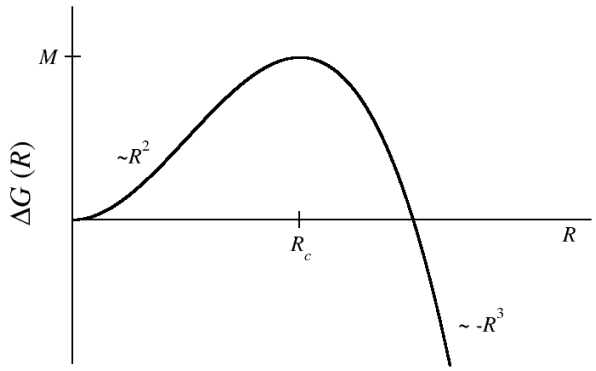

Figure 1.3: Free energy ΔG barrier for a spherical droplet of radius R.

The nucleation process can be understood qualitatively using the framework of classical nucleation theory (CNT). For the sake of simplicity, I first consider the case of a spherical liquid droplet nucleating in an otherwise homogeneous vapour. It is worth remembering that for a transition to happen, the free energy of the new phase must be lower than the free energy of the old phase. The transition occurs through nucleation and growth whenever the two phases are separated by a thermodynamic barrier. According to the classical nucleation theory (CNT), the difference (ΔG) between the free energies of the system in the new (liquid) ($G(R)$) and old phase (gas) ($G(1)$) is given as a function of the droplet radius by the two terms in Eq. 1.1:

$$\Delta G = G(R) - G(1) = -\frac{4}{3}\pi R^3 \rho_l \Delta\mu + 4\pi R^2 \sigma, \tag{1.1}$$

In Eq. 1.1, R is the radius of the formed droplet, ρ_l is the number density of the liquid, $\Delta\mu$ is the difference in the chemical potential between the gas phase and the liquid phase, σ is the surface tension or the liquid-gas surface free energy density. The free energy $4\pi R^2\sigma$ required to create the liquid/gas interface of the new cluster represents the obstacle that in the metastable region of the vdW graph prevents the formation of droplets and it is given by the second term in

Eq. 1.1 (surface term). The first term, known as a "volume" or bulk term, accounts for the free energy advantage of the liquid phase with respect to the vapour phase.

Fig. 1.3 shows ΔG as a function of the droplet radius R. From the figure the competing effect of the two terms can be seen. For small R (small droplets) the surface term dominates and the free energy increases with increasing R. Thus, the gas phase will persist until thermal fluctuations pay the free energy cost to reach the top of the barrier at the critical size R_c. Beyond that dimension, the free energy decreases with increasing radius and the phase transition will continue spontaneously, driven by a dominant volume/bulk term.

The first theoretical explanation of the vapour-liquid transition phase at equilibrium was developed by Gibbs in the 19th century (44). The case of interest here, however, is the nucleation of a solid-like phase out of a solution. This problem is somewhat more difficult than the vapour-liquid nucleation, and for this reason its theoretical analysis took longer to develop. In 1934 Stranski and Kaischew wrote the first papers (73; 74; 75; 134; 135) referring to a classical nucleation theory (CNT) of crystals and crystal monolayers. A comprehensive description of the nucleation process including the kinetics aspects was given one year later by Becker and Döring (10) based on the experiments of Volmer, Weber and Farkas (29; 150). Since that time, the CNT for crystal nucleation has been corrected and modified. Good reviews of the evolution of the theory through the years can be found in different exhaustive texts (77; 125). The model used by Stranski and Kaishew to investigate the crystal nucleation is the Kossel crystal that has a one to one correspondence with the Ising model of ferromagnets (64). The Kossel crystal is a fairly basic model that allows to measure important information into the crystallisation of the nucleation of different systems (141) by means of theoretical studies and simulations.

A wide range of experiments have been able to investigate the nucleation phenomena in different fields giving an estimation of parameters including the nucleation rate J (number of clusters per unit time and volume). The range of the experimentally predicted nucleation rate spans more than 20 orders of magnitude (from 10^{-3} to 10^{20} s^{-1}/cm^3), depending on the technique used to measure it (153).

Unfortunately, theoretical predictions of J by CNT do not agree with experimental results except in some rare case (56; 59; 159; 162). For example the CNT

predictions for the nucleation rate of argon deviates from experimental data by 16 to 26 order of magnitude while for nitrogen the discrepancy ranges from 10 to 20 order of magnitude (153) depending on temperature. This disagreement is surprising, considering that a substance such as argon behaves almost ideally. The deviations of the CNT from the experimental results can be associated with the fact that this theory is "classical" and neither quantum mechanical corrections nor the atomistic structure of the nucleus have been taken into account. However, it is still apparent that CNT can qualitatively explain the process. Furthermore, in crystal nucleation the enormous discrepancy between CNT, numerical and experimental results for both 2D and 3D Kossel crystal nucleation motivated Walton to put forward in 1962 an atomistic nucleation theory (ANT) which represents the process of crystal nucleation in terms of microscopic or atomistic parameters (151). The important difference between CNT and ANT is in the shape that the cluster assumes during its evolution. While CNT allows only the same equilibrium shape of the nucleus independently of the size of the cluster, ANT takes into account all of the possible irregular shapes and considers the lowest in surface energy. The ANT theory has been further developed by Shneidman and Nita (130; 131; 132; 133).

With the advent of the computer the nucleation process has been increasingly investigated by computational approaches and simulation results have been compared with experimental findings and theoretical predictions (45; 82; 147; 154). There have been numerous examples of computer simulations of nucleation using both the Monte Carlo (MC) and Molecular Dynamics (MD) technique. The MC technique is a stochastic method capable of giving a description of the nucleation process at equilibrium, while MD is more suitable to obtain information about the dynamics of the process (37). However, already in the early stages of computer simulation, the MC method has been generalised to the so-called kinetic Monte Carlo technique, which can give information on relaxation processes and on the real-time evolution of nucleating clusters (13).

Simulations of nucleation processes can be extremely time-consuming because nucleation is a rare event and a realistic description might require a large simulation sample. The Kossel crystal nucleation for example, has been extensively investigated by Monte Carlo simulations that continue to be carried out nowadays (15; 16; 70; 71; 94; 119; 129; 140; 141; 156) because they are usually less time-consuming than the respective molecular dynamic simulations. Back

in 1970 Gilmer and Bennema studied by MC simulations the 2D Kossel crystal nucleation and growth on a substrate with and without surface diffusion showing that their simulation results are independent of the amount of surface diffusion (45). In 1977 Van der Erden et al. studied the migration of crystallites with a MC technique that introduces time as a continuous parameter (146) while Kashchiev et al. investigated the thin film deposition on a Kossel crystal face as a substrate (82). Additional early examples of simulation studies of nucleation include those by Swope and Anderson, who have studied the formation of crystals from a Lennard-Jones fluid in 1990 (137). In 1998 ten Wolde and Frenkel carried out extensive MC simulations of the condensation of a Lennard-Jones fluid (139), and shortly afterwards Yasuoka and Matsumoto employed MD simulations to study the same system (158). Since then, a variety of computer experiments have been performed to study homogeneous/heterogeneous nucleation and crystal growth of 2D and 3D Kossel crystal with kinetic Monte Carlo. Recent works from ter Horst and from Kashchiev et al. calculated the rate and growth probability of isotropic 2D and 3D Kossel crystals, showing that the rate prediction from CNT overestimates the simulation results while ANT is closer to the simulation and experimental results (140; 141; 142). The correction to the classical nucleation theory (CCNT) takes into account the above overestimation and allows for a reliable analytical estimation of the nucleation rate that is closer to the ANT predictions and to simulation results (80).

It is worth pointing out that, almost without exceptions, all these studies considered the simplest case of nucleating solids, in which the system was made of cubic-like particles. Anisotropic cases have received only little attention so far. Moreover, the few available papers on anisotropic systems (5; 27; 57; 68; 120; 121) have been restricted to the study of phase diagrams or geometrical aspects of the self-assembly. My work, instead, is specifically devoted to the study of highly anisotropic systems, such as amyloid fibrils, and to the study of the effect of the anysotropic interactions on the nucleation kinetics, never approached before in terms of nucleation theory.

Understanding the nucleation process is a fundamental topic with application in many fields. The brief introduction given above concerned homogeneous phase transitions, but in nature the heterogenous nucleation is the most common process. To study the kinetics of a first order transition, therefore, it is important to distinguish if it happens under heterogeneous or homogeneous conditions. In

heterogenous nucleation the phase transition occurs in the presence of alien surfaces or particles present in the old phase. Water, for example, would freeze well below the well known freezing point in the absence of particles that are acting as nucleation centres in the solution. In heterogeneous nucleation the transition process could be triggered or prevented by the external body. Despite the fact that in experiments and in nature most of the transition phases are heterogeneous, a complete understanding of the homogeneous nucleation is a necessary pre-requisite to obtain a full description of the nucleation process. The amyloid fibril formation from a solution of proteins or peptides is an example of a nucleation mediated process. The aim of this study is to apply nucleation theory to describe the homogeneous amyloid fibril nucleation.

1.2 Amyloid fibrils

According to basic principles of statistical mechanics and thermodynamics, and consistently with many experimental and computational works (1; 12; 54; 66; 76; 108; 113), amyloid fibrils start to form out of the homogeneous solution whenever the peptide or protein concentration raises above an equilibrium value C_e, measuring their solubility limit for any given thermodynamic condition. The equilibrium concentration for the fibril condensation is usually measured experimentally by determining when the fibril formation and the fibril dissolution rates are equal (1; 54; 66; 108). Measurements are carried out using the turbidity method, and sometimes the fibril phase is heated to speed up the formation and solvation kinetics (1; 54; 66). Even above C_e, i.e., in the stability range of amyloids, fibrils may not form instantaneously, since the process is usually controlled by important kinetic effects limiting the nucleation process. According to the elementary picture of nucleation, schematically illustrated in Fig. 1.3, the homogeneous solution and the fibrillar state are separated by a free energy barrier ($\Delta G(R)$ in Fig. 1.3), that is determined by the competition between the free energy advantage of the condensed fibril phase, and the free energy cost of forming an interface between the two phases. Only at peptide concentrations far above C_e, the activation barrier for the fibril's formation vanishes, and the phase change occurs instantaneously through spinodal decomposition, driven by the large free energy advantage of fibrils with respect to the homogeneous peptide solution. These considerations set the stage for the essential role of kinetics and

of the nucleation stage in the formation of amyloid fibrils, which is in fact the central topic of my thesis.

According to the nucleated-polymerisation model (54; 66) the fibril formation occurs via the random aggregation of monomers into small clusters that are thermodynamically unstable because of the barrier. The process proceeds slowly through its initial stages (nucleation stage in Fig. 1.4) since it is driven only by thermodynamic fluctuations. However if a cluster reaches the size corresponding to the top of the barrier in Fig. (1.3), it becomes a critical nucleus and it has a 50% probability of growing into a mature fibrils. At this stage the process is relatively fast (elongation stage in Fig. 1.4) because the attachment of further monomers to the cluster is thermodynamically favourable, i.e. there is no barrier (right side of Fig.1.3).

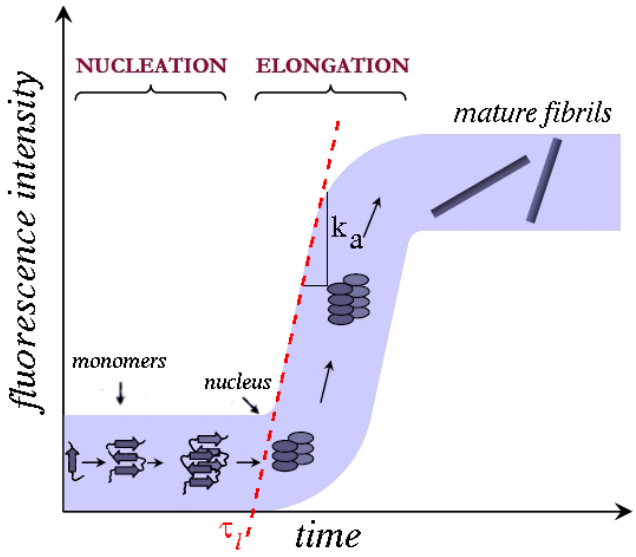

Figure 1.4: Schematic representation of the amyloid fibril nucleation process as revealed by fluorescence signal. By convention the maximal nucleation rate, k_a, is the slope of the tangent in the picture and the lag time, τ_l, is the intersection of the tangent line with the time axis.

Kinetic information about the amyloid fibril nucleation can be obtained from studies of the overall process of protein aggregation that is characterised by an initial lag time τ_l during which no protein aggregates can be detected (slow process) and by a maximum rate, k_a, of the protein conversion into aggregates (quick process) (6; 24; 28; 33; 34; 51; 58; 61; 62; 87; 90; 92; 112; 116; 122; 157; 164).

Both the aggregation lag-time and the maximum rate may depend on the overall fibril nucleation rate and are obtainable from time-resolved optical experiments such as those that measure the fluorescence signal arising from dye molecules bound to protein aggregates (Fig. 1.4). Dye molecules suitable for these measurements include the Congo red and thioflavin, which bind to mature fibrils, giving their typical fluorescence signal, the intensity of which is proportional to the amount of amyloid fibrils in the sample. As shown in the schematic picture of Fig. 1.4, by convention the maximal rate of aggregation, k_a, is defined as the slope of the maximal tangent to the curve, while the τ_l is the intersection of the tangent to the time axis.

High resolution structural details can be revealed by X-ray crystallography, solid state nuclear magnetic resonance (SSNMR), cryo-electron microscopy (cryoEM) and circular dichroism (22; 97; 136). A variety of fibrillar morphologies have been observed, with diameters of 7 - 15 nm and a wide range of lengths, which can reach several μm or even more. As mentioned previously, the amyloid fibril is a fibrillar structure whose core has the typical diffraction pattern of cross-β structure (see Fig. 1.5) that shows a distance between the proteins or peptides in the β-sheet of $\sim$ 4.7 Å and a distance between two different sheets of $\sim$ 6 - 11 Å.

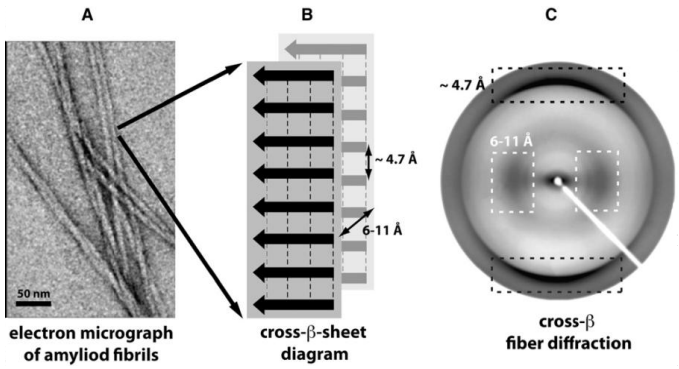

Figure 1.5: Structure of Amyloid fibrils from (50). A) Mature Amyloid fibrils from negatively stained transmission electron microscopy. B) Schematic diagram of the β-cross structure and C) typical cross-beta diffraction pattern (12).

However there are different configurations and architectures that these units can assume into the amyloid fibril. Due to the lack of long range order, X-ray diffraction alone is unable to provide an atom-by-atom view of amyloids'

structure. For these reasons, atomistic-scale models play an invaluable role in understanding the chemical principles of the amyloid structure (22). The picture in Fig. 1.1, for example, is a model based on a cryo-electron microscopy map representing the amyloid fibril formed by SH3 polypeptide chains (69). From these models, it has been found that the cohesive portion of the fibril does not involve the entire protein, but it is made of only some amino-acidic sequences giving rise to the β-sheet in the fibril (22; 107). Furthermore, within each β-sheet two or more extended protein or peptide segments can be found in a parallel or antiparallel (Fig. 1.6a) arrangement, in such a way that the carbonyl group of one fragment is linked by hydrogen bonds to the complementary group of the other peptide. The parallel or antiparallel arrangement can be in-register or out-of-register if the amino-acids of the first peptide are paired to the respective amino-acids of the adjacent identical peptide. The majority of amyloids studied appear to be in a parallel in-register arrangement (22; 144), even though some parallel out-of-register and some antiparallel arrangements are also observed (22; 107). The β-sheet can then stack on each other in different final conformations, a few of them being shown in the picture.

It is known that a polypeptide chain can assume a multitude of conformational states and in solutions it converts from one state to another one. Fig. 1.7 illustrates a "hypothetical" conformational pathway starting from the biosynthesis of the protein on a ribosome, from which the protein is initially unfolded. From this starting point, the protein can fold and unfold through different intermediate conformational states or oligomers (clusters of a small numbers of protein or peptide molecules without a fibrillar structure) and eventually it might "misfold", giving rise to amyloid fibrils. The study of the intermediate states are becoming more intriguing since it has been found that the oligomeric states are toxic and they play a critical role in the pathologies related to amyloids (22).

Experiments (24; 28; 51; 58; 62; 87; 92; 112; 116; 122; 157; 164) have been performed to reveal the relationship between the physicochemical properties of the natural amino acids forming the peptide and the kinetics of amyloid formation and to validate phenomenological models (23; 30; 138; 144; 161) able to predict the changes in the propensity of proteins to aggregate upon mutation as well as to identify amino acid sequences of proteins that are likely to belong to the fibril core. Among these phenomenological models, PASTA (144) is a representative example of a numerical tool that can predict the regions in the

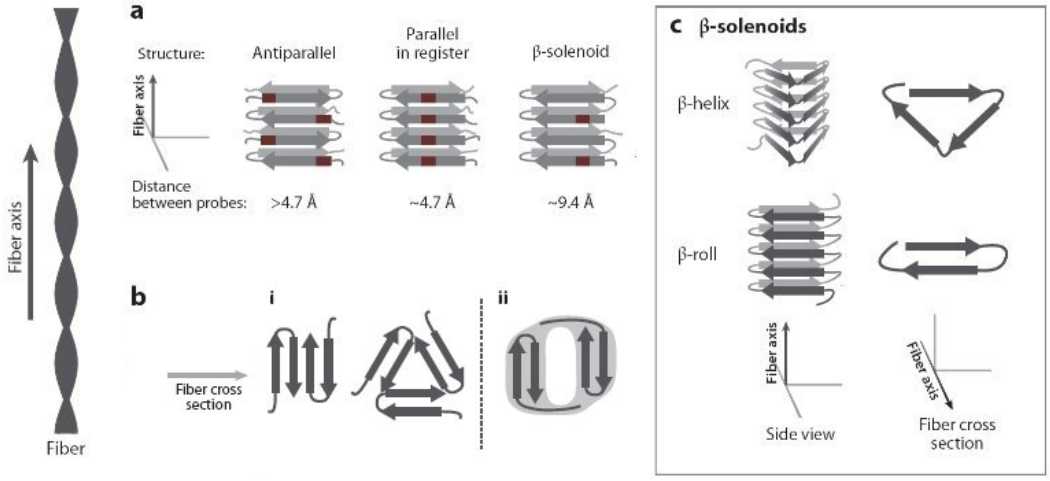

Figure 1.6: Most common peptide patterns in amyloid fibrils from Ref. (14) The arrows represent peptides. (A) Peptides in antiparallel arrangement and in parallel in-register, in the β-solenoid structure where a single monomers forms two layers of the fiber structure. (B) Conformational diversity of fibers according to several models of Aβ fiber structures based on SSNMR, and CryoEM ((14)). (C) Types of β-solenoids, Two types of solenoid include β-helix (top) (3β-sheet form a triangular interface) and β-roll, where the two sheets form an in interface akin to the β-sandwich.

peptide chain that are more likely to be part of the cross-β structure of the amyloid fibril. Furthermore, this approach can predict if the peptide in the β-sheet is in a parallel or antiparallel, in-register or out-of-register arrangement. The calibration of PASTA is based on the frequency at which a given amino-acid pair is found in globular proteins, as reported in standard data bases. A reference PASTA implementation is hosted on a freely accessible server that scans a given sequence and returns the pairing energies for all the likely combinations.

Also, the statistical mechanics algorithm Tango (30) predicts the β-sheet aggregation propensity for any given amino-acid sequence. In particular, it estimates the pairing free energy using statistical and empirical considerations, taking into account hydrophobicity, electrostatic interactions and hydrogen bonding between the two amino-acids being considered, as well as physiological conditions such as pH, buffer and concentration of the initial protein solution. It was the first algorithm to predict the effect of a given protein mutation on the aggregation properties.

To gain further insight into the nucleation mechanism and to determine im-

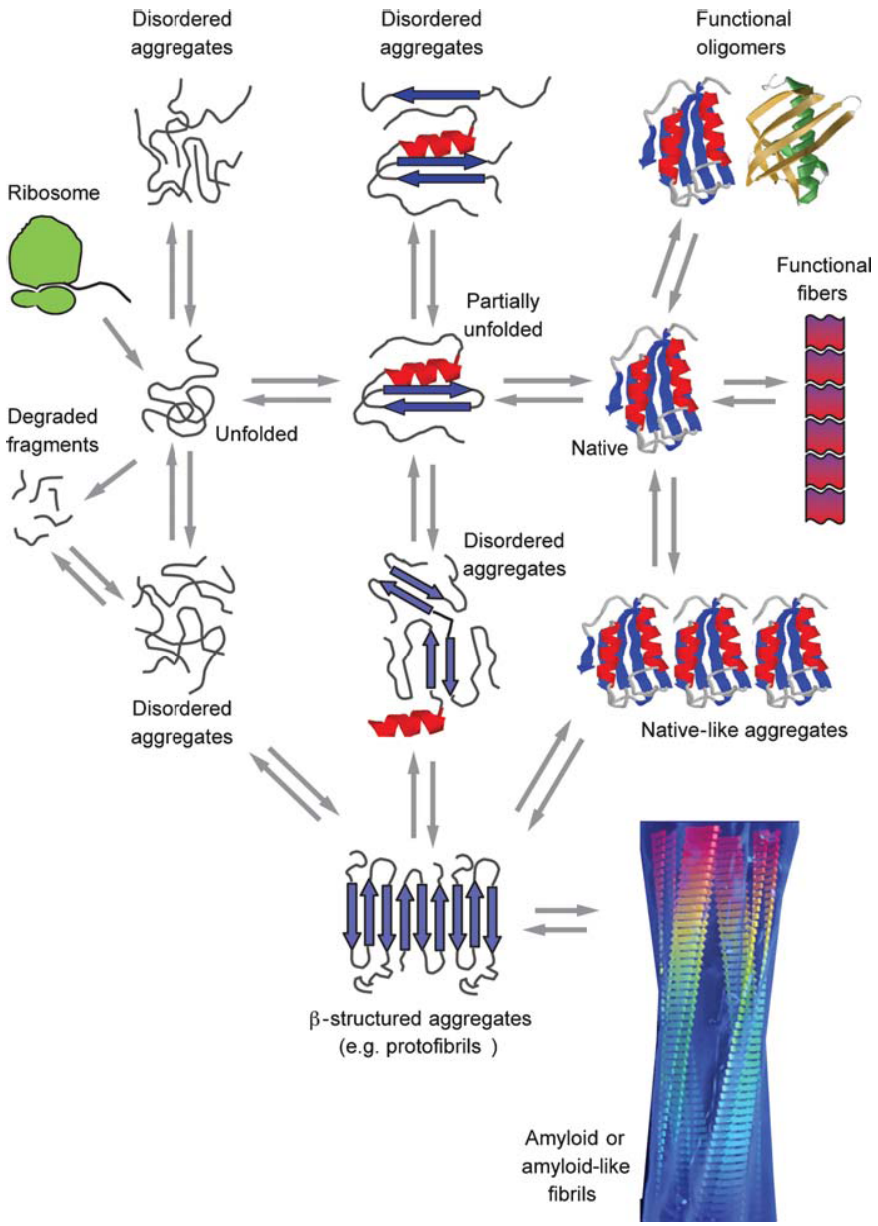

Figure 1.7: Diagram of different conformational states that can be assumed by a polypeptide chains, eventually ending into an amyloid fiber conformation. The scheme has been taken from Ref. (22).

portant nucleation parameters such as the fibril nucleus size and nucleation rate, small-angle X-ray scattering (114; 149) and fluorescence correlation spectroscopy experiments (23; 24; 28; 30; 41; 42; 51; 58; 62; 87; 92; 112; 114; 116; 122; 138; 144; 149; 161; 163; 164) have been used to monitor the fibril population density and to identify changes in the size and structure of protein aggregates in

solution. In particular Garai et al. (41) have carried out experiments using fluorescence correlation spectroscopy (FCS) (97) to estimates the size distribution of a solution of amyloidβ peptides nucleating into amyloid fibrils. From the size distribution of the aggregates they were able to provide an estimation of the surface energy of the peptide and of the barrier for the nucleation process using the framework of the CNT.

Despite the many experimental and computational studies briefly reviewed previously, and despite the fundamental role of the amyloid formation process, it is fair to say that the nucleation of nanofibrils is still poorly understood. The reason is that experiments characterising the early stages leading to the formation of these aggregates are difficult mainly because of the transient nature and structural heterogeneity of the amyloid precursors. Furthermore the complexity and size of the protein systems has made it difficult to simulate the fibril nucleation events.

To understand and interpret experiments on the early stages of amyloid fibril formation it is necessary to develop a theoretical model of fibril nucleation. Models based on protein physico-chemical properties have been proposed and used for predicting changes in the aggregation propensities (20; 23; 30; 138; 144; 161). Although such models have been able to correctly predict changes due to mutation, as well as to identify protein sequences that are likely to be part of the amyloid fibril core, they do not provide information about the fibril nucleation barrier and rate. Models at the molecular level have been put forward for describing amyloid fibrillation (90; 101; 157; 163) but they do not fully exploit existing general theories of nucleation for treating the fibril nucleation process.

Zhang et al. (163), for instance, performed a series of MC simulations to study the growth and the kinetics of amyloid fibril formation from a solution. Using a coarse grained lattice model, important information such as lag-time, fibril mass and morphology, critical concentrations depending on temperature and molecular interactions have been obtained. Lomakin et al. (101) have studied the fibrillogenesis of amyloidβ protein using quasi-elastic light scattering spectroscopy and they have examined the structure of the amyloid fibrils with circular dichroism spectroscopy and electron spectroscopy. Their esperimental findings are consistent with a model for the fibrillogenesis where the peptide micelles form above a critical concentration, the fibrils nucleate on heterogeneous nuclei or micelle and then grow by binding monomers to its end.

In summary, there are a multitude of theoretical, experimental and simulation studies on amyloid fibril formation, but up to now none of them have fully exploited the pre-existing atomistic nucleation theory.

A topical question, in the field of amyloid fibril self-assembly, is whether the amyloid fibrils nucleate in one-step, where the unfolded monomeric protein/peptide in solutions polymerises directly into fibrils, or they follow a two-step path, where metastable oligomers first appear in solution (like droplet condensation in supersaturated vapour) which convert into fibrils at a later stage (reordering step). Two step nucleation has been observed in many peptide and protein amyloid systems (8; 9; 39; 58).

This thesis is specifically devoted to the one-step nucleation case, and assumes that before fibrillation the building blocks (peptides or proteins) are already in the unfolded state most suitable for forming amyloid fibrils. The analysis will reveal the general principles underlying the nucleation kinetics of amyloid fibrillation, showing that amyloid fibril formation can be treated in the framework of existing general theories of nucleation of new phases.

1.3 Aim and results of this thesis

The task initially set for my work has been to develop and apply a nucleation theory able to describe at the molecular level the one-step homogeneous nucleation of amyloid fibrils from monomers.

For the first time, the atomistic nucleation theory (ANT) has been adapted and applied to study the assembly of proteins into amyloid fibrils, providing important results that open the way to further developments and applications. My work has been carried out within the framework of pre-existing nucleation theories. As a result of my investigations, the kinetic and thermodynamics of the fibril formation have been predicted and compared with experimental results, giving confidence to the validity of the theory.

In particular, the ANT prediction of the nucleation rate of fibrils has been applied to real protein models, adopting parameters for the amino acid-amino acid interaction proposed in recent theoretical works. Using this approach, the study of the effect of point-mutations lead to the important conclusion that the solubility is primarily responsible for the changes in the nucleation rate. The theoretical curves are in very good agreement with the experimental results,

despite of the fairly idealised character of the model. The study has shown that the theory is a powerful tool, not only because it allows for theoretically understanding of the qualitative features of amyloid fibril nucleation, but also because it provides semi-quantitative data to support experiments. The analysis, in fact, is able to reproduce the effect of mutations, and to identify the physical conditions that can trigger, slow down, accelerate or prevent the amyloid fibril nucleation. For these reasons the theory can have a strong impact in the research field devoted to the analysis of amyloid fibrils formation.

Additionally, the application of ANT to fibril nucleation allowed the estimation of the size distribution of the aggregates. This time a direct comparison of the analytical calculation with experimental results has been impossible for lack of experimental data measured at conditions comparable to those used in my calculations. However, qualitative features in the size distribution, such as the presence of special ("magic") sizes, are apparent both in the analytical results and in the experimental data, providing indirect support for the validity of my results. Hopefully, this study will stimulate new experiments at the right physical conditions able to confirm directly the predicted size distribution, and to provide further insight on the important estimation of the fibril sizes.

Finally a computer simulation study has been carried out to test selected theoretical assumptions used in the development of the theory discussed in the first five chapters. My results for isotropic systems are in excellent agreements with previous simulation and theoretical study, thus validating my implementation of a kinetic Monte Carlo (KMC) simulation. More importantly, for the first time, the nucleation of strongly anisotropic systems has been investigated using kinetic Monte Carlo simulations. This study revealed novel and surprising results, never obtained previously in either experiments or simulations studies. Because of the universal character of the model, the results can give important information about any nucleating system whose building blocks interact among themselves with strong anisotropic forces.

1.4 Thesis structure

The thesis is organised as follow:

In chapter 2 the classical nucleation theory (CNT) and a model for amyloid fibril nucleation will be given. I will show how the fibrillation process can

be qualitatively described by CNT and how a corrected version of the classical nucleation theory (CCNT) can provide more accurate parameters for this process. Furthermore a theory-independent estimation of the fibril dimension will be given.

Chapter 3 discusses the nucleation of amyloid fibrils at the molecular level. The atomistic nucleation theory (ANT), specifically developed for amyloid fibril, will be applied to an idealised peptide system and compared with CNT and CCNT.

ANT for amyloid fibrils will be applied to a real protein model in chapter 4. The application of ANT to the proteins β_2-microglobuline and Amyloid β fibrils will allow for the predictions of recent kinetic experiments.

In chapter 5, ANT is applied to the nucleation of Aβ_{40} fibrils, aiming at estimating the size of the aggregates under particular physical conditions that are consistent with experimental results.

Chapter 6 presents the results of a computer code that simulates the 2D crystal nucleation with a kinetic Monte Carlo (KMC) method. The idealised character of the model allows the application of the simulation results to highly anisotropic systems, such as amyloid fibrils, and to isotropic Kossel crystal.

Chapter 2

Classical theory of amyloid fibril nucleation[1]

In a recent article (81), the classical nucleation theory (CNT) (77; 84) and the corrected CNT (CCNT) (80) were applied for describing the amyloid fibril nucleation thermodynamics and kinetics when the process occurs by direct polymerisation of β-strands (peptides or protein segments) into β-sheets. This mechanism is operative when no fibril precursor such as droplet-like protein aggregates appears in the solution as a first step in the nucleation process. In this chapter the application of CNT and CCNT for describing the nucleation of amyloid fibril will be discussed. In particular the work to form a nanosized amyloid fibril (protofilament) built up of successively layered β-sheets, the size of the fibril nucleus, the nucleation work and the fibril nucleation rate as explicit functions of the concentration and temperature of the protein solution will be determined. The analysis reveals that application of existing nucleation theories can qualitatively describe the amyloid fibril formation and that the CCNT is more reliable in the prediction of the rate of fibril formation than CNT. The results discussed in this chapter concern homogeneous nucleation, which occurs when the solution is sufficiently pure and/or strongly supersaturated.

2.1 Introduction

Here some important general nucleation theory formulae will be introduced. All the formulae presented here are valid in the scope of the Szilard-Farkas model

[1]This chapter is based on Refs. (81) and (18).

(29) of single-component nucleation according to which each n-sized cluster of the new phase has only one shape and it can change its size by attachment or detachment solely of monomers (single atoms or molecules). Since the theory is being applied to the nucleation of amyloid fibrils, the monomer here are the β-strands, i.e. the peptides or protein segment.

At a given absolute temperature T(K) there exists an equilibrium concentration C_e (m^{-3}) of β-strands at which the fibril neither grows nor dissolves. At this condition the fibril nucleation cannot occur and the solution is saturated. When the actual concentration of β-strands C_1 (m^{-3}) is higher than C_e, fibrils can nucleate and grow to macroscopic sizes, the system is then metastable and the solution is supersaturated. When the actual concentration C_1 is lower than C_e the solution is in a stable state and the nucleation is impossible. In the nucleation theory the fundamental quantity which takes into account the important role of the equilibrium concentration or solubility is the supersaturation $\Delta\mu$ (Joules) and these two quantities are related by the formula:

$$\Delta\mu(C_1) = k_B T \ln(C_1/C_e), \tag{2.1}$$

where k_B is the Boltzmann constant. The supersaturation $\Delta\mu$ is the driving force of the nucleation and its definition contains the difference between the chemical potentials of the β-strands in the solution and in the bulk fibrillar phase as it will be discussed below (77).

The general expression for the work W_n (Joules) to form an n-sized amyloid fibril (n is the number of β-strands building the fibril) is of the form:

$$W_n = -(\mu_s - \mu_f)n + \Phi_n, \tag{2.2}$$

where μ_s (Joules) and μ_f (Joules) are the chemical potentials of a β-strand in the solution and in the macroscopically large fibrillar phase respectively, while Φ_n (Joules) is the fibril excess energy. Physically the first term is the work gained by the system on forming the fibrillar phase while the second term is the work done on creating the fibril/solution interface.

According to the amyloid nucleation process, an n-sized fibril would eventually nucleate in a supersaturated solution by random attachment and detachment of single β-strands to and from smaller fibrils. The size n of the fibril would fluctuate with time until it does grows sufficiently beyond the size $n = n^*$ of a so-called

critical nucleus or fibril nucleus. At $n = n^*$ the work assumes its maximum value and for that reason a fibril nucleus is in labile thermodynamic equilibrium with the remaining β-strands in the solution. The n-sized fibril can grow deterministically up to macroscopic size, only if it is sufficiently bigger than the fibril nucleus. Once n^* has been determined, the work to form the fibril nucleus or nucleation work $W^* \equiv W_{n^*}$ can be obtained and it is of particular interest because beyond this barrier the fibril would evolve spontaneously into super-nucleus fibrils.

The smallest supernucleus fibrils appear then as a result of random attachments of protein monomers to the fibril nuclei. The nucleation rate J ($\mathrm{m}^{-3}\mathrm{s}^{-1}$) is defined as the frequency of appearance of supernucleus fibrils per unit solution volume. The general expression for the stationary, i.e. time-independent, nucleation rate J is of the form (e.g., refs. (77; 84))

$$J = zf^*C^*. \tag{2.3}$$

In amyloid fibril nucleation, $C^* \equiv C(n^*)$ (m^{-3}) is the concentration of fibril nuclei in the solution, and $f^* \equiv f(n^*)$ (s^{-1}) is the frequency of attachment of β-strand monomers to the fibril nucleus. The numerical parameter $z \leq 1$ is the so-called Zeldovich factor which takes into account that C^* is roughly twice the stationary concentration of nuclei and that not every attachment event results in overgrowth of the nucleus to a macroscopic size. The concentration of n-sized fibrils in the solution can be written as a function of the work necessary to form an n-sized cluster:

$$C(n) = C_1 \exp[(W_1 - W_n)/k_BT]. \tag{2.4}$$

In the latter equation $C_1 \equiv C(1)$ (m^{-3}) is the actual concentration of single β-strands in the solution, $W_1 \equiv W(1)$ and $W_n \equiv W(n)$ are, respectively, the work to form a β-strand, which is formally considered as the smallest representative of the nucleating phase, and the work to form an n-sized fibril. Combining Eqs. (2.4) and (2.3) gives the formula :

$$J = zf^*C_1 \exp[(W_1 - W_{n^*})/k_BT]. \tag{2.5}$$

The quantities given in this section are thermodynamically derived and can be analytically calculated having a detailed knowledge of the work $W(n)$ to form the n-sized fibril. Furthermore all the fundamental formulae above can be

applied in the CNT description which provides a clear mathematical explanation of the fibril nucleation process, but, as it has been shown in Ref. (80), this classical theory disregards the work done to attach the first β-strand at the periphery of the CNT fibril nucleus. In order to cure this important inaccuracy the corrected CNT (CCNT) will be applied. While the fibril nucleus size in CCNT is just one β-strand bigger than the one predicted from the classical theory, CNT indeed underestimates considerably the nucleation work and, as a consequence overestimates by many order of magnitude the fibril nucleation rate (81). In this chapter both of these theories will be applied to the assembling of β-strands into fibrils and the main features and differences of the CNT and CCNT approaches will be highlighted.

2.2 Physical model

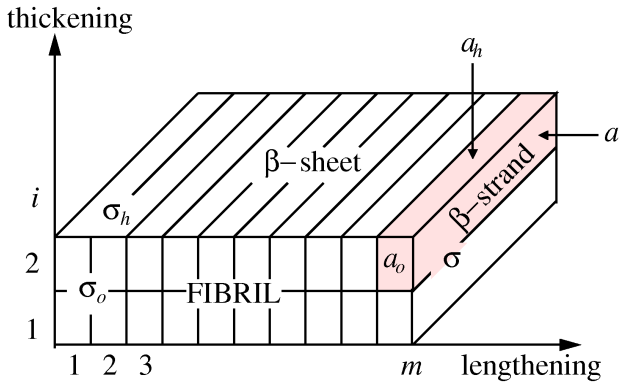

Figure 2.1: Classical model of a nanosized amyoid fibril with thickness of $i = 2$ β-sheets and length of $m = 10$ β-strands. The σ's are the specific surface energies of the three fibril faces, and the a's are the areas of the three peptide faces.

For describing the arrangement of the virtually fully extended β-strands in a nanosized amyloid fibril of successively layered β-sheets (Fig. 2.1) (77), a lattice model which parallels those of Nguyen and Hall (Fig. 8 in Ref. (111)) and of Zhang and Mutukumar (163) has been used. The considerations here pertain to nanosized amyloid fibrils (protofilaments) built up of successively layered β-sheets with fixed width and thickness. Fig. 2.1 schematises such a fibril containing i such β-sheets ($i \geq 1$), each β-sheets having m practically fully extended rod-like β-strands ($m \geq 1$) arranged parallel to each other and perpendicular to

the fibril lengthening axis. The fibril is therefore always prismatically shaped and contains a total of $n = i \times m$ β-strands ($n \geq 1$). The areas occupied by a β-strand at the fibril faces perpendicular to the m and i axes are a (m^2) and a_h (m^2) respectively, and the area occupied by the β-strand in the (m, i)-plane is a_0 (m^2) (Fig. 2.1). These areas are given by $a_0 = d_h d$, $a = d_0 d$, and $a_h = d_0 d_h$, where d_h (m) is the distance between the β-strands in a β-sheet, d (m) is the intersheet distance in the fibril, and d_0 (m) is the β-strand length, i.e. the β-sheet width.

The ontogenesis of the smallest (nanosized) amyloid fibrils in the (m, i)-plane can be described by 2D crystal nucleation and growth theories. Essential parameters in these theories are the specific surface energies of the different crystal faces. As shown by Kashchiev and Auer (81), in amyloid fibril nucleation two of the three fibril specific surface energies (Fig. 2.1) are of immediate importance: the first one is the specific surface energy σ ($\mathrm{J/m}^2$) of the fibril face perpendicular to the m axis, and the second one is the specific surface energy σ_h ($\mathrm{J/m}^2$) of the fibril face perpendicular to the i axis. The third fibril specific surface energy, σ_0 ($\mathrm{J/m}^2$), characterises the fibril face parallel to the (m, i) plane and enters in the description of the fibril energetics only implicitly via the supersaturation (81).

To a first approximation, the surface energy is proportional to the energy of the broken bonds at the respective surface (73; 74; 134) and, for that reason, σ and σ_h are largely determined by the strength of the bonds between the neighbouring β-strands in the fibril. The fibril elongation is primarily driven by the formation of strong hydrogen bonds between the β-strands along the m axis, while the fibril thickening along the i axis is due to the much weaker bonds between the β-strands arising, e.g., from the hydrophobic effect. This allows to write for amyloid fibrils the important inequality $\sigma \gg \sigma_h$. For example, it could be $\sigma_h = \sigma/10$ or $\sigma_h = \sigma/20$ for fibrils with bond energy (per β-strand) between two nearest-neighbours β-sheets that is about 10 or 20 times lower than the hydrogen bond energy between nearest-neighbour β-strands in a β-sheet. In this chapter, in particular, the analytical calculations and results for the nucleation of fibrils characterised by $a = a_h$ and a ratio of the specific surface energies $\sigma/\sigma_h = 10$ will be presented.

2.3 Fibril evolution mechanism

Within CNT all the parameters such as fibril-formation work, nucleus size, nucleation work and nucleation rate can be calculated by analytical formulae that will be presented through this chapter. However it is important to spend some time discussing how the fibril evolves and eventually grows according to the CNT description. A fundamental requirement is that the fibril preserves the equilibrium shape given by the thermodynamic condition of minimal total surface energy and thus of minimal formation work. In CNT the fibril excess energy $\Phi_{i,m}$ is expressed as the fibril total surface energy (77),(84) so that, according to the present model, the work to form an n or i,m-sized fibril can be written from Eq. (2.2) as:

$$W_{i,m} = -(\mu_s - \mu_f)im + 2a_0\sigma_0 im + 2a\sigma i + 2a_h\sigma_h m, \tag{2.6}$$

or

$$W_{i,m} = -im\,\Delta\mu + 2a\sigma i + 2a_h\sigma_h m \tag{2.7}$$

where $\Delta\mu \equiv \mu_s - \mu_f - 2a_0\sigma_0$ is defined as the effective supersaturation specifically for the present 2D model. From the condition for extremum of $W_{i,m}$ from Eq. (2.7), $\partial W_{i,m}/\partial i = \partial W_{i,m}/\partial m = 0$, with i and m treated as continuous variables, it is found that the thickness i and the length m of an equilibrium shaped fibril are not independent of each other but are related by :

$$\frac{m}{i} = \frac{\sigma}{\sigma_h}. \tag{2.8}$$

This important relation is a form of the Gibbs-Curie-Wulff theorem for the equilibrium shape of a crystal (55; 143) as found by Kaischew for rectangularly shaped 2D crystals (72). For the present specific model the ratio in Eq. (2.8) is equal to 10.

With the help of Eqs. (2.2) and (2.8) the evolution of the fibril can be modelled bearing in mind that its shape is always prismatic, so that m and i are not necessarily integers. Furthermore, the equation for the work (Eq. (2.2)) is not valid for $n < m/i$ because, according to Eq. (2.8) and the relation $n = i \times m$, a fibril would have the physically meaningless thickness of $i < 1$. This leads to the conclusion that the fibrils start growing as single β-sheet. The first β-sheet cannot grow unlimitedly as one-dimensional (1D) formation, because its total surface energy may be higher than that of a fibril with the same number

of β-strands arranged in 2β-sheets. Specifically, the fibril would evolve into a 2D formation of 2β-sheets once the size n has reached the transition size $n_t = m/i = 10$. The same mechanism can be extended to describe the further fibril evolution. According to the thermodynamic requirement of minimal surface energy, from $n = n_t$ to $n = 2^2 n_t$ the CNT equilibrium shape requires a fibril thickness between one and two β-sheets, while from $n > 2^2 n_t = 40$ to $n = 3^2 n_t = 90$ a fibril thickness between two and three β-sheets is required, and finally between size $n > 3^2 n_t = 90$ and $n = 4^2 n_t = 160$ a fibril thickness between three and four β-sheets guarantees the CNT equilibrium shape. The fibril evolution would follow the CNT equilibrium shape requirement up to the desired size as explained above, presenting always a geometrically prismatic shape and having at $n = n_t, 2^2 n_t, 3^2 n_t, 4^2 n_t$ the transition to the second, third, forth and fifth β-sheet respectively. It is important to notice that the differently sized fibrils are considered as possessing always the fixed shape of elongated right rectangular prism, consistently with the fact that i and m are not necessary integers. In this way the nanoscale irregularities in the actual fibril shape that arise, e.g., from the presence of incompletely built-up β-sheets on the fibril surface, are totally neglected by CNT.

2.4 Work of fibril formation

As already stated, in CNT the fibril excess energy $\Phi_{i,m}$(J) is considered as equal to the total surface energy of the n-sized fibril ($n = im$):

$$\Phi_{i,m} = 2a_0\sigma_0 n + 2a\sigma i + 2a_h\sigma_h m. \tag{2.9}$$

Up to now CNT neglects the possible dependence of σ_0, σ_h and σ on the fibril size, however the fibrils, unlike the rigid crystallites, are flexible bodies. An additional (i, m)-dependent term that takes into account the bending, twisting and other eventual deformations could be incorporated in $\Phi_{i,m}$ of equation (2.9), without changing the formal expression of the work in the formula (2.2). In the present study, for simplicity, these (i, m)-dependent terms are neglected. The general expression for the dimensionless work $w_{i,m} = W_{i,m}/k_BT$ to form a prismatic nanosized fibril with length of m β-strands, thickness of i β-sheets and size of $n = i \times m \geq 1$ β-strands is obtained by combining Eqs. (2.2) and (2.9) as done

in Ref. (81):

$$w_{i,m} = -sn + 2\psi i + 2\psi_h m. \tag{2.10}$$

Here, $s \equiv (\mu_s - \mu_f - 2a_0\sigma_0)/k_BT \geq 0$ is, for the present model, the effective dimensionless supersaturation or driving force for nucleation of 2D fibrils and their further growth in length and thickness. Furthermore, in the latter formula, the dimensionless surface energies ψ and ψ_h of the fibril faces perpendicular to the m-axis and the i-axis respectively, are given by :

$$\psi = a\sigma/k_BT \tag{2.11}$$

$$\psi_h = a_h\sigma_h/k_BT. \tag{2.12}$$

As shown in Ref. (81), s can be calculated from the equations $s = \ln(C_1/C_e)$ or $s = L\Delta T/k_\beta T_e T$ for solutions in which the supersaturation is experimentally controlled by the concentration $C_1(\mathrm{m}^{-3})$ of monomer proteins at a given T or by T at a given C_1. In these equations $\Delta T \equiv T_e - T$ (K) is the undercooling, T_e (K) is the solution equilibrium temperature, and L(J) is the latent heat (per β-strand) of protein aggregation. In the CNT description, regardless of its size n, the fibril is always prismatically shaped and its size is connected with m and i by the relation $n = i \times m$. Using this relation to eliminate m in Eq. (2.10) it is found that the formation of an n-sized fibril with a fixed number i of β-sheets (the so-called $i\beta$-sheet (7)) requires the work (81)

$$w_{i,n} = -(s - 2\psi_h/i)n + 2\psi i. \tag{2.13}$$

Special attention must be paid to the work

$$w_{1,n} = -(s - 2\psi_h)n + 2\psi \tag{2.14}$$

to form 1β-sheet of n β-strands.

Fig. 2.2 shows the dependence of the work $w_{1,n}$ for different scaled supersaturations $s/\psi_h = 1$, 2 and 3 as straight lines labelled 1β, drawn according to equation (2.14). Equation (2.14) is of great interest because it gives the condition under which 1β-sheet, and thus the fibril formation from a solution of β-strands, can spontaneously occur. In particular, the 1β-sheet forms spontaneously at supersaturations $s > 2\psi_h$ (1β line at $s/\psi_h = 3$ in Fig. 2.2) because

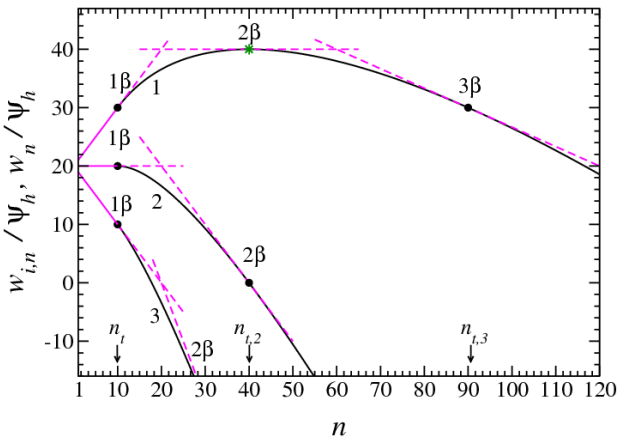

Figure 2.2: Dimensionless work to form an n-sized fibril at different supersaturations $s/\psi_h = 1, 2$ and 3 for $\psi/\psi_h = 10$. The magenta lines, labelled as 1β, 2β and 3β in the graph, illustrate the work to form 1β, 2β and 3β-sheet; Eq. (2.14) for 1β (straight line) and Eq. (2.13) with $i = 2$ and 3 to draw 2β and 3β (dashed lines) have been used. The three black curves represent the work to form a fibril of equilibrium shape given by Eq. (2.18) and are labelled according to the scaled supersaturations. In the graph the transition size n_t and $n_{t,3}$ to the $2^{nd}\beta$-sheet and the $4^{th}\beta$-sheet, respectively, are marked with black points on the curves. $n_{t,2}$, labelled with a green star, represents the transition size to the $4^{th}\beta$-sheet and the fibril nucleus n^* at the same time.

then $w_{1,n}$ diminishes with increasing n. According to CNT, then, the process of fibril formation occurs in a metanucleation regime because it is not impeded by the existence of a nucleation barrier. In contrast, when $s < 2\psi_h$ the 1β-sheet elongation requires work to be done ($w_{1,n}$ increases with n, 1β line at $s/\psi_h = 1$ in Fig. 2.2), which means that fibril formation is then nucleation mediated. The maximum supersaturation at which CNT is applicable is, therefore, the supersaturation $s_{1\beta} = 2\psi_h$ (81). Physically, $s_{1\beta}$ is the supersaturation at which the 1-βsheet is in thermodynamic equilibrium or coexistence with the solution, because at $s = s_{1\beta}$ no work is done on the fibril growth or dissolution (then $w_{1,n}$ is n-independent, 1β line at $s/\psi_h = 2$ in Fig. 2.2). In Fig. 2.2 Eq. (2.13) is also plotted for $i = 2$ and 3 as visualised by the straight dashed lines 2β and 3β, respectively, at scaled supersaturations $s/\psi_h = 1, 2$ and 3 for ratio $\psi/\psi_h = 10$. In particular, Eq. (2.13) for $i = 2$ at $s/\psi_h = 1$ is a constant straight line parallel to the n axis. This means that once the fibril has reached the 2β-sheets configuration, it is in equilibrium or coexistence with the solution of the remaining

β-strands. In fact, no work has to be done to detach or attach β-strands to or from it. Then, for $i = 2$ at $s/\psi_h = 1$ the work $w_{2,n}$ becomes n-independent. In the same way, Eq. (2.13) for $i = 1$ at $s/\psi_h = 2$ (upper supersaturation limit for the nucleation process) shows the coexistence between the 1β-sheet fibril and the solution. Therefore, the supersaturations in which the 3β-sheets and 4β-sheets are in equilibrium or coexistence with the solution can be found by setting to zero the factor in brackets that contains the driving force for the $i\beta$-sheets formation in Eq. (2.13) (81):

$$s_{i,\beta} = 2\psi_h/i. \tag{2.15}$$

Also, the equilibrium concentration $C_{i,\beta}$ of β-strands at which the $i\beta$-sheets are in coexistence with the solution can be obtained using Eq. (2.15) and Eq. (2.1) to give :

$$C_{i,\beta} = C_e \exp(2\psi_h/i). \tag{2.16}$$

Both $s_{i,\beta}$ and $C_{i,\beta}$, as can be seen from Eqs. (2.15) and (2.16), decrease with increasing i thickness of the fibril. In the limit of $i \to \infty$ it is $s_{i,\beta} \to 0$ and $C_{i,\beta} \to C_e$.

Although initially the fibril comes into being as a 1β-sheet, with increasing its size n by lengthening along the m-axis, its total surface energy becomes greater than that of a 2β-sheets with the same size n. For that reason, as already discussed in the previous section, after a certain time a second β-sheet starts forming on top of the original 1β-sheet. The transition size n_t, given in (81) can now be re-written as $n_t = \psi/\psi_h = 10$. n_t marks the fibril transition from 1D evolution (growth in length only) into 2D evolution (growth in both length and thickness) and it is indicated with a black point in Fig. 2.2. Therefore, the condition that characterises the fibril equilibrium shape as a function of the dimensionless surface energies (81) becomes:

$$\frac{i}{m} = \frac{\psi_h}{\psi}. \tag{2.17}$$

Following the latter condition and considering that $n = i \times m$, the fibril will eventually start to form a third β-sheet for $n > n_{t,2} = 2^2 n_t = 40$. Similarly the transition between the 3β-sheets and the 4β-sheets configuration will appear at $n > n_{t,3} = 3^2 n_t = 90$, and the fifth β-sheet will start to be formed at $n > n_{t,4} =$

$4^2 n_t = 160$ and so on. The circles in Fig. 2.2 mark the different transitions at n_t $n_{t,2}$ and $n_{t,3}$.

Using $n = i \times m$ and Eq. (2.17) for elimination of i and m in Eq. (2.10) yields the dimensionless work $w_n \equiv W_n/k_BT$ to form an n-sized fibril with equilibrium shape at a given supersaturation $s \leq s_{1\beta} = 2\psi_h$:

$$w_n = -sn + 4(\psi\psi_h)^{1/2} n^{1/2}. \tag{2.18}$$

It is important to note that Eq. (2.18) is not valid for $n < n_t = \psi/\psi_h$, because according to it and the relation $n = i \times m$, such small fibrils would have the physically meaningless thickness $i < 1$. Hence, the dependence of the fibril formation work is given by Eq. (2.14) in the region $1 \leq n \leq n_t$ and by Eq. (2.18) in the region $n > n_t$.

In Fig. 2.2 the curves w_n according to Eq. (2.18) at different scaled supersaturations $s/\psi_h = 1, 2$ and 3 for ratio $\psi/\psi_h = 10$ are shown. From Fig. 2.2 it can be seen that all three curves are below the lines 1β 2β and 3β which are drawn with the help of equation (2.13) at the correspondent supersaturation. In fact, the three portions of straight lines 1β, 2β and 3β envelop the corresponding w_n curve. From the same figure it can clearly be seen how the barrier w_n changes with respect to the supersaturation. At $s/\psi_h = 3$ the fibril will form spontaneously, since the work decreases with increasing n for any thickness i. At $s/\psi_h = 2$, the curve is characterised by the equilibrium shape of 1β-sheet up to $n = 40$, after which the fibril will pass to a 2D evolution characterised by spontaneous growth since the barrier w_n is decreasing with the increasing of n. Finally, at $s/\psi_h = 1$, the system encounters a barrier to complete its first layer and even after the transition point $n = n_t = 10$ the curve w_n will increase with the fibril size n. The size $n = n_{t,2} = n^*$ marks the fibril transition from a 2β-sheets to a 3β-sheets configuration, and, importantly, determines the end of the increasing trend of w_n as a function of n. Beyond $n = n^*$ the second layer of the fibril will be completed without spending any work, and at $n = n^*$ the fibril is in coexistence with the solution. Subsequently a third β-sheet will be spontaneously build, since w_n is decreasing with increasing n.

2.5 Nucleus size and nucleation work

The fibril of size $n = n^*$ that corresponds to the top of the curve in Fig. 2.3 is called the fibril nucleus or critical nucleus and it is defined as the maximum of the w_n function.

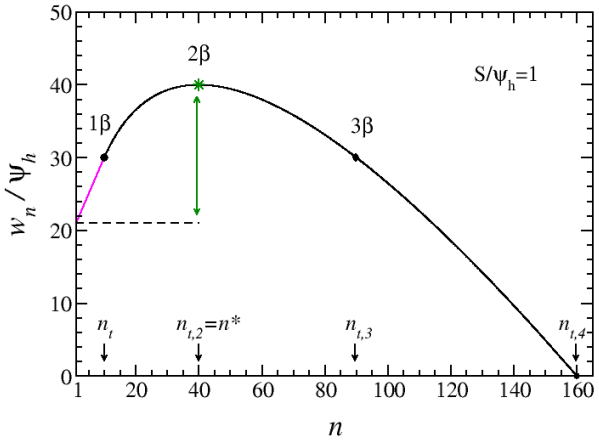

Figure 2.3: Dimensionless work w_n to form an n-sized fibril at $s/\psi_h = 1$ for $\psi/\psi_h = 10$ according to Eq. (2.14) for $1 \leq n \leq n_t$ and Eq. (2.18) for $n > n_t$. Black points indicate n_t, $n_{t,3}$ and $n_{t,4}$. $n_{t,2}$ is marked with a green star because it is a transition point and fibril nucleus at the same time (n^* in the graph). The green double arrow illustrates the magnitude of the energy barrier that the fibril has to overcome to assume the size $n = n^*$

The condition for maximum, $(dw_n/dn)_{n=n^*} = 0$, employed to Eq. (2.18) leads to the following simple analytical formulae for the supersaturation dependence of the size n^* of the fibril nucleus and the dimensionless nucleation work $w^* \equiv w_{n^*}$ $(0 \leq s \leq s_{1\beta} = 2\psi_h)$ (81):

$$n^* = \frac{4\psi\psi_h}{s^2} \tag{2.19}$$

$$w^* = \frac{4\psi\psi_h}{s}. \tag{2.20}$$

The formula (2.19) is the Gibbs-Thomson equation for nucleation of 2D condensed phases (e.g., Ref. (77)) and furthermore, w^* and n^* are related by the formula $w^* = sn^*$ and by $dw^*/ds = -n^*$ known as the nucleation theorem in $\Delta\mu$ form (77; 78; 83). These quantities are of great interest because they determine

the energy barrier, $(w^* - w_{1,1})k_BT$, that the 1β-sheet fibril has to overcome to be able to grow up to macroscopic sizes. In Fig. 2.3 the barrier height for $\psi/\psi_h = 10$ at $s/\psi_h = 1$ is shown.

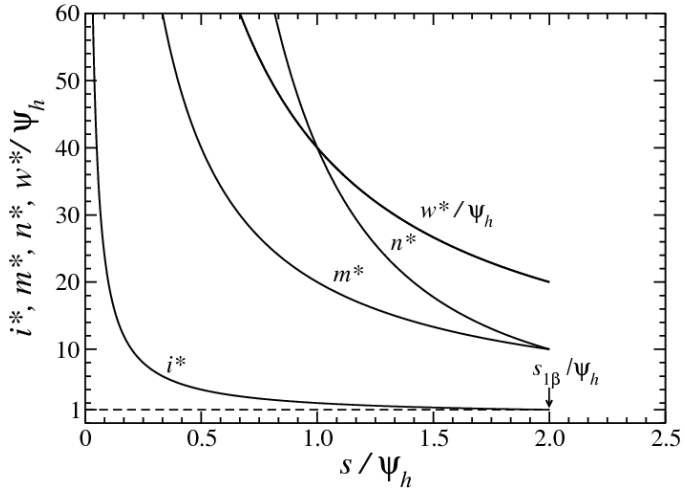

Figure 2.4: Supersaturation dependence of n^*, i^*, m^* and w^*/ψ_h, as labelled, are shown according to Eqs. (2.19)-(2.22) for $\psi/\psi_h = 10$. In the graph the arrow points to the supersaturation $s_{1,\beta}$, which is the limit for the CNT applicability.

Combining Eq. (2.17), Eq. (2.19) and the relation $n^* = i^* \times m^*$, leads to analytical formulae expressing the nucleus thickness i^* and the nucleus length m^* $(0 \leq s \leq s_{1\beta} = 2\psi_h)$:

$$i^* = \frac{2\psi_h}{s}, \tag{2.21}$$

$$m^* = \frac{2\psi}{s}. \tag{2.22}$$

In Fig. 2.4 the supersaturation dependence of w^*/ψ_h, i^*, m^* and n^* are plotted according to Eqs. (2.19)-(2.22) for $\psi/\psi_h = 10$. All four curves decrease as the supersaturation increases, in particular at $s = s_{1,\beta} = 2\psi_h$, $i^* = 1$, $m^* = n_t = 10$ and then $n^* = 10$, which means that the fibril can grow barrierlessly since $w^* = w_{1,1}$. At this supersaturation, in fact, Fig. 2.2 shows that a single β-strand can attach to the $(n_t - 1)\beta$-strands to complete the 1β-sheet fibril without any work needed to be done. As already stated, $s_{1,\beta}$ is considered as the upper limit of s for the CNT applicability. All the formulae described in this section are not valid for $s > s_{1,\beta}$, because for higher supersaturations the fibril formation

would be spontaneous and barrierless and i^* would be smaller than 1, which is physically meaningless.

2.6 Nucleation rate

If the fibril size n is sufficiently larger than the nucleus size n^*, the fibril can grow deterministically up to a macroscopically large size. In particular it suffices that the size M of this supernucleus is $M > 2n^*$ (79), hence from this size up, the fibril will irreversibly both lengthen and thicken as long as $C_1 > C_e$. The stationary rate J of appearance of supernucleus fibrils per unit solution volume, at constant supersaturation and temperature, is given by the general formulae for J, (2.3) and (2.4). In order to calculate J, an explicit expressions for the Zeldovich factor, z, the frequency $f^* = f(n^*)$ and the equilibrium nucleus concentration C^* are needed. In the scope of the classical Szilard-Farkas model of single-component nucleation (e.g. Refs. (77; 84)), for 2D nucleation, according to CNT in its self-consistent formulation, z and C^* are expressed as (e.g. Ref. (77))

$$z = \left(\frac{s}{4\pi n^*}\right)^{1/2} \tag{2.23}$$

and

$$C^* = C_1 \exp(w_{1,1} - w^*). \tag{2.24}$$

In Eq. (2.24) $w_{1,1}$ is the dimensionless work to form a β-strand monomer, which is formally considered as the smallest representative of the nucleating phase, and C_1 (m^{-3}) is the concentration of monomeric β-strands. Eq. (2.24) is often used in the equivalent form $C^* = C_0 \exp(-w^*)$, because C_1, $w_{1,1}$ and the concentration C_0 (m^{-3}) of nucleation sites in the solution are related by the expression (77) $C_1 = C_0 \exp(-w_{1,1})$. As to f^*, it depends on the particular mechanism of β-strand attachment to the nucleus fibril. Under the assumption that the β-strand diffusion to the fibril nucleus is so fast that there are always β-strand monomers ready for attaching to the fibril ends, the frequency of β-strand attachment to the fibril nucleus can be written down as (81)

$$f^* = f_1 i^*. \tag{2.25}$$

Here f_1 (s^{-1}) is the s-dependent frequency of β-strand attachment to the two ends of a 1β-sheet, and i^* is the thickness of the fibril nucleus ,i.e. the number of β-sheets in the fibril nucleus. When T is fixed and s is controlled by means of the concentration C_1 of monomer β-strand, the $f_1(s)$ dependence is given by (81) $f_1 = f_{1,e}C_1/C_e = f_{1,e}e^s$, where $f_{1,e}$ (s^{-1}) is the value of f_1 at the equilibrium concentration C_e, i.e. at $s = 0$. Substituting Eqs. (2.23) - (2.25) into Eq. (2.3), using Eq. (2.14) to determine $w_{1,1}$, taking into account Eq. (2.19) and the relation (81) $i^* = 2\psi_h/s$ yields the CNT formula for the supersaturation dependence of the rate of amyloid fibril nucleation (81) ($0 \leq s \leq s_{1,\beta} = 2\psi_h$)

$$J = \left(\frac{\psi_h s}{4\pi\psi}\right)^{1/2} \times f_1(s)C_1 \left[2(\psi + \psi_h) - s - \frac{4\psi\psi_h}{s}\right]. \tag{2.26}$$

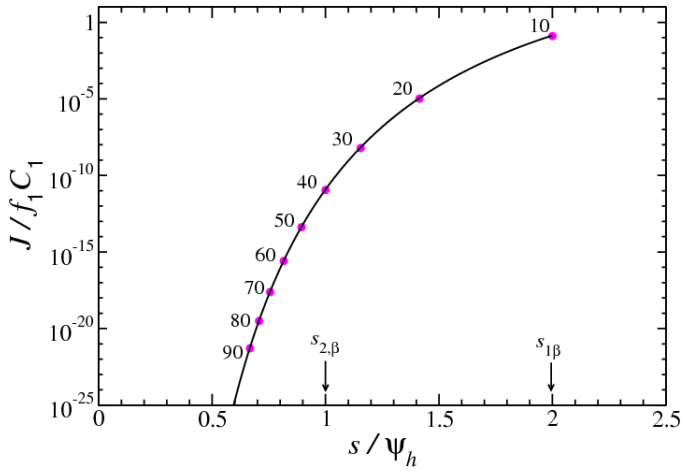

Figure 2.5: Supersaturation dependence of the nucleation rate J according to Eq. (2.26) at $\psi = 12.0$ and $\psi/\psi_h = 10$. In the graph nucleus sizes n^* according to Eq. (2.19) are shown in magenta points. The arrows indicate the supersaturation limit for the CNT applicability, $s_{1,\beta}$, and the transition supersaturation $s_{2,\beta}$ at which the fibril assume a 2β-sheet configuration.

The line in Fig. 2.5 illustrates the supersaturation dependence of the nucleation rate J in the CNT description according to Eq. (2.26) for $\psi/\psi_h = 10$. In fact the graph is plotted using $T = 300$ K, $a = 5$ nm^2 and $\sigma = 10$ mJ/m^2 according with a specific surface energy in the range of 0.1-30 mJ/m^2 as reported by experimental works for protein crystals in aqueous solutions (41). These parameters lead to the values $\psi = a\sigma/k_BT = 12.0$ and $\psi_h = a\sigma_h/k_BT = 1.2$, at $T = 300$ K. The nucleation rate J is a smooth monotonic function that increases

with increasing supersaturation, however its rate of increase is higher at lower supersaturations and tends to get lower and lower as $s_{1,\beta}$ is approached. This trend can be interpreted considering that with increasing supersaturation the nucleation work w^* decreases as does the critical nucleus size n^*. The numbers labelling the line in Fig. 2.5 represent the nucleus sizes at the corresponding supersaturation. The supersaturation limit for the CNT applicability $s_{1,\beta}$ and the transition supersaturation to the 2β-sheet configuration, $s_{2,\beta}$, are also highlighted.

2.7 Corrected classical nucleation theory

CCNT does not provide a correction of the CNT Eq. (2.18) for the work w_n to form an n-sized fibril. It only corrects the CNT Eqs. (2.19), (2.20) and (2.26) for the nucleus size n^*, the nucleation work w^* and the nucleation rate J, respectively. This correction is necessary, because CNT disregards the work $(-s+2\psi)k_BT$ done in attaching side-wise the first β-strand to the surface of the CNT nucleus. This β-strand increases the size of the CNT nucleus by unity and triggers the barrierless build-up of a new β-sheet on the CNT nucleus, a β-sheet needed for the nucleus to thicken and, thereby, to preserve the CNT equilibrium shape. Fig. 2.6 illustrates the geometrical difference between the CCNT and the CNT fibril nuclei at the same supersaturation s and same ψ/ψ_h ratio.

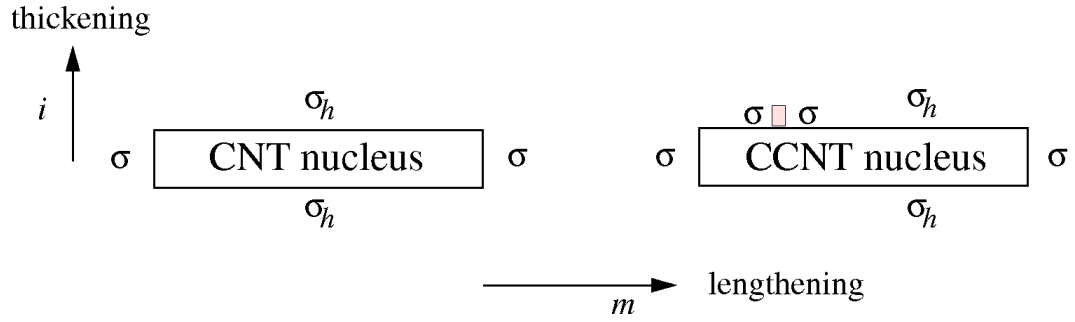

Figure 2.6: Geometrical difference between the CNT and CCNT nucleus as labelled in the picture. The pink rectangle on one of the σ_h faces, represents the β-strand that triggers the barrierless growth of a new β-sheet.

Accounting for the above, CCNT provides the equations (81) ($0 \le s \le 2\psi_h$)

$$n^* = \frac{4\psi\psi_h}{s^2} + 1 \tag{2.27}$$

$$w^* = \frac{4\psi\psi_h}{s} - s + 2\psi, \tag{2.28}$$

which correct the CNT Eqs. (2.19) and (2.20) for n^* and w^*. The applicability of these formulae is restricted to the CNT supersaturation range $[0,s_{1,\beta}]$, therefore Eq. (2.27) is known as the corrected Gibbs-Thomson equation. It can also be shown that Eqs. (2.27) and (2.28) satisfy the nucleation theorem in the form $dw^*/ds = -n^*$ (77; 78; 83) and are related by the formula $w^* = (n^* - 2)s + 2\psi$. While the correction of the CNT nucleus size is small (one more β-strand in the nucleus), the correction of the CNT nucleation work is of major importance, because the energy term $-s + 2\psi$ is usually much greater than unity (81).

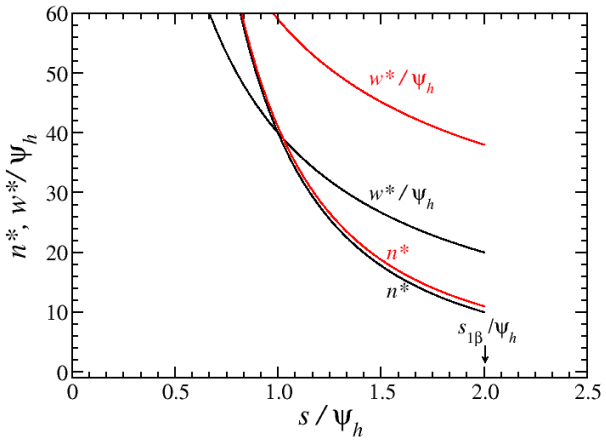

Figure 2.7: Supersaturation dependence of n^* and w^*/ψ_h at ratio $\psi/\psi_h = 10$ for CNT (black lines Eq. (2.19) and (2.20)) and CCNT (red lines Eq. (2.27) and (2.28)) descriptions.

These differences are clear in Fig. 2.7 that shows n^* and w^*/ψ_h against s/ψ_h in black lines for CNT and in red lines for CCNT. The n^* and w^*/ψ_h lines are drawn using Eqs. (2.19) and (2.20) of CNT and Eqs. (2.27) and (2.28) of CCNT at ratio $\psi/\psi_h = 10$. As in the CNT description, in the CCNT one n^* and w^* decrease with increasing supersaturation. Fig. 2.7 clearly confirms that

while the differences in the CNT and CCNT predictions for n^* are negligible, the differences in the predictions for the nucleation work w^*/ψ_h are considerable and are increasing with the supersaturation up to a value of 18 units at s/ψ_h=2.0.

Interestingly, at $s = s_{1,\beta}$ the CCNT nucleation barrier is still different from zero ($(w^* - w_{1,1})k_BT = 2(\psi - \psi_h)k_BT$), unlike the CNT barrier which is nil at the same supersaturation. Hence, CCNT predicts that the fibril formation is a nucleation mediated process even beyond the CNT supersaturation upper limit $s_{1,\beta}$ after which the fibril growth will proceed barrierlessly in the so-called metanucleation region (80).

To correct the CNT nucleation rate J from Eq. (2.26), CCNT uses again the general Eq. (2.3) in combination with C^* from Eq. (2.24) in which $w_{1,1}$ is determined according to Eq. (2.14), but w^* is taken from Eq. (2.28). In addition, CCNT employs the approximation $z = 1/2$, because only about half of the equilibrium nucleus concentration C^* is effective in stationary nucleation (77) and because after attaching a protein monomer, the CCNT nucleus virtually always grows to a macroscopically large size. Also, in accordance with Eq. (2.25), in CCNT f^* is approximated by $f^* = 2f_1$, since in most cases the CCNT nucleus is expected to be two β-sheets thick (one β-sheet plus one β-strand on the sheet). As a result, the CCNT nucleation rate J is given by the expression (81) ($0 \leq s \leq s_{1\beta} = 2\psi_h$)

$$J = f_1(s)C_1 \exp\left[2\psi_h - \frac{4\psi\psi_h}{s}\right]. \tag{2.29}$$

Comparison of Eqs. (2.26) and (2.29) shows that the CCNT nucleation rate is much lower than that of CNT because the exponential function in Eq. (2.29) does not contain the term $(2\psi - s)$ which, as already noted, is usually much greater than unity. The absence of this term cannot be compensated by the absence in Eq. (2.29) of the factor $(\psi_h s/4\pi\psi)^{1/2}$, which is typically a number between 0.01 and 0.1. Like Eq. (2.26), Eq. (2.29) is applicable in the CNT supersaturation range [0,$s_{1\beta}$].

Fig. 2.8 shows the supersaturation dependence of J according to the CNT description (Eq. (2.26)) and the CCNT one (Eq. (2.29)). The two curves show the same trend but the CCNT nucleation rate J is generally slower and it becomes slower and slower with increasing supersaturation, due to the factor $(2\psi - s)$. In

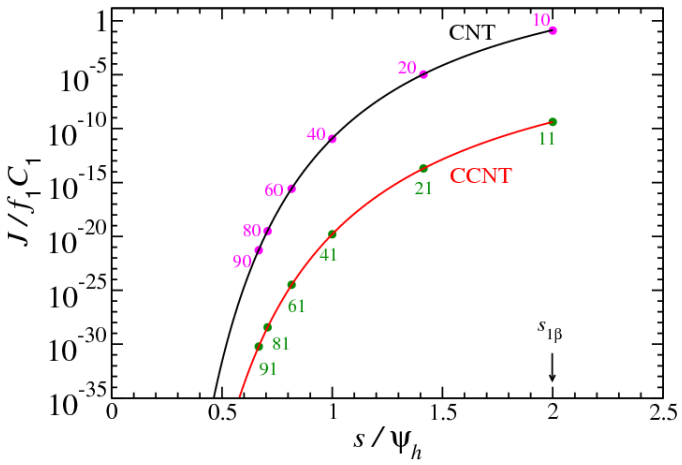

Figure 2.8: Supersaturation dependence of J at $\psi = 12.0$ and ratio $\psi/\psi_h = 10$ for CNT (black line Eq. (2.26)) and CCNT (red line Eq. (2.29)) descriptions. In the graph the nucleus sizes at the corresponding supersaturations are marked on the J curves with magenta (obtained with Eqs. (2.19)) and green points (obtained with Eqs. (2.27)) for CNT and for CCNT, respectively.

the J curves are highlighted for certain supersaturations the nucleus sizes (magenta and green points on the CNT and CCNT J curves respectively) calculated according to formulae (2.19) and (2.27) of CNT and CCNT respectively.

When T is fixed and s is controlled by means of the concentration C_1 of β-strands in the solution, the nucleation rate J from (2.29) can be expressed as $J(C_1)$ ($1 \leq C_1/C_e \leq \exp(2\psi_h)$):

$$J = \underbrace{f_{1,e} C_e \exp(2\psi_h)}_{A} \left(\frac{C_1}{C_e}\right)^2 \exp[-\underbrace{4\psi\psi_h}_{B}/ln(C_1/C_e)]. \tag{2.30}$$

The latter is an important formula that can be related to experimental conditions. In Eq. (2.30) A ($\mathrm{m^{-3}s^{-1}}$) is known as the kinetic factor and B is the dimensionless thermodynamic parameter. Furthermore, differentiating with respect to C_1 the logarithm of both sides of Eq. (2.30) and using Eq. (2.27) the important relation can be obtained:

$$n^*(C_1) = \frac{d(\ln J)}{d(\ln C_1)} - 1. \tag{2.31}$$

The last formula is of fundamental applicability. It is in fact valid in the CNT approach as well, as can easily be obtained from n^* of Eq. (2.19) and expressing

J of Eq. (2.26) as a function of C_1. Eq. (2.31) is in fact a general result that comes from the application of the nucleation theorem to isothermal nucleation of condensed single-component phases (77; 78; 81; 83). Then, if isothermal $J(C_1)$ data are available, Eq. (2.31) gives a theory independent estimation of the fibril nucleus size as a function of the concentration of the β-strands in the solution measuring the slope of $\ln(J)$ versus $\ln(C_1)$. Furthermore, keeping in mind that $s = \ln(C_1/C_e)$, Eq. (2.31) can be approximately written as a function of the supersaturation in the form:

$$n^*(s) \approx \frac{d(\ln J)}{d(s)} - 1, \tag{2.32}$$

Eq. (2.32) gives an estimation of the nucleus size n^* in term of the parameter s easily controlled during experiments.

2.8 Conclusion

The aim of this chapter was to study amyloid fibril nucleation using standard nucleation theory in the form of CNT and CCNT. A general introduction gave a universal nucleation description that is valid for both of these theories. Then the amyloid fibril nucleation has been studied with the CNT and CCNT description highlighting the conceptual differences in the results. In particular the work to form an n-sized fibril has been obtained, and from it, the fundamental nucleation quantities such as the nucleus size n^*, nucleation work w^* and nucleation rate J of amyloid fibrils have been derived.

In the CNT description, the fibril formation is described using a 2D lattice model where the single β-strand is represented as a right prism with surface energies that are fundamental parameters for the protein aggregation into amyloid fibrils. The nano-sized fibril would eventually form because of random attachment and detachment of β-strands. If the process climbs the thermodynamic barrier up to the nucleation work w^* the fibril will become a nucleus of size $n = n^*$ and from then on, it could grow to macroscopic sizes. The rate J of fibril nucleation can be analytically calculated taking into account the barrier that the process has to overcome to form nucleus fibrils.

Although CNT could give a good qualitative explanation of the process, it has been shown that it strongly underestimates the nucleation work. CCNT corrects

this result by adding one unit to the nucleus size formula and introducing the necessary work to add that unit to the nucleation work formula. In the CCNT description, upon reaching the nucleus size, the fibril could grow barrierlessly a new β-sheet that allows the fibril to thicken and reach the equilibrium prismatic shape given by the CNT description. CCNT provides a new formula for the nucleation rate which is also expressed in terms of the nucleation work. It has been shown that the CNT formula for J is highly inaccurate for crystals with dimensionless specific energy $\psi > 1$ and it should not be used. A reliable formula for J is given by the CCNT equation Eq. (2.30). Importantly, it has been discussed that the nucleation rate can be related to experimental isothermal data and to the protein solubility. Furthermore the estimation of the nucleus size from $J(s)$ data is theory-independent, because it is a general result based on the nucleation theorem.

To summarise, in this chapter, kinetic and thermodynamic aspects of the amyloid fibril nucleation have been studied from the classical nucleation point of view and it has been shown that standard nucleation theories can give a general qualitative explanation of the process. The analysis offers answers to common questions in the intriguing phenomenon of amyloid fibril nucleation such as numerical estimation of the fibril size, rate and barrier of the assembly process.

Chapter 3

Atomistic theory of amyloid fibril nucleation[1]

Also in this chapter the focus of the discussion is the nucleation of amyloid fibrils at the molecular level, when the process takes place by a direct polymerisation of peptides or protein segments into β-sheets. Here, in particular, employing the atomistic nucleation theory (ANT), a general expression for the work to form a nanosized amyloid fibril (protofilament) composed of successively layered β-sheets will be derived. The application of this expression to a recently studied peptide system will allow to determine the size of the fibril nucleus, the fibril nucleation work and the fibril nucleation rate as functions of the supersaturation of the protein solution. The analysis illustrates the unique feature of ANT that the size of the fibril nucleus is a constant integer in a given supersaturation range. Furthermore the ANT nucleation rate will be calculated and compared with the rates determined in the previous chapter in the scope of the classical nucleation theory (CNT) and the corrected classical nucleation theory (CCNT). The main result is that while the CNT nucleation rate is orders of magnitude greater than the ANT one, the CCNT and ANT nucleation rates are in very good quantitative agreement. The results obtained in the chapter are applicable to homogeneous nucleation which occurs when the protein solution is sufficiently pure and/or strongly supersaturated.

[1]This chapter is based on Refs. (18) and (81).

3.1 Introduction

In the previous chapter, the classical nucleation theory (CNT) (e.g., Refs. (77; 84)) and the corrected CNT (CCNT) (80) were applied for describing the amyloid fibril nucleation thermodynamics and kinetics when the process occurs by direct polymerisation of peptides or protein segments into β-sheets. In these theories the fibril energetics is described in terms of the fibril dimensionless surface energies ψ and ψ_h which are macroscopic characteristics of the fibril/solution interface. Also, the differently sized fibrils are considered as possessing the fixed shape of an elongated right rectangular prism and in this way the nanoscale irregularities in the actual fibril shape that arise, e.g., from the presence of incompletely built-up β-sheets on the fibril surface, are neglected. It is, however, of interest to analyse the fibril nucleation thermodynamics and kinetics at the molecular level, expressing ψ and ψ_h in terms of the broken bond energies of a β-strand (peptide or protein segment) constituting the fibril and then obtaining information about the actual fibril shape. The atomistic nucleation theory (ANT) (e.g., Refs. (77; 80; 96)) provides a framework for treating nucleation at the molecular level. In the present chapter ANT will be applied to the nucleation of amyloid fibrils by the mechanism of direct polymerisation, according to which the smallest fibrils appear as a result of a series of random attachments and detachments of monomer peptides or protein segments. This mechanism is operative under conditions when no fibril precursors such as droplet-like protein aggregates are formed in the solution as a first step in the fibril nucleation process.

The dimensionless work to form a n-sized fibril, oriented along i and m axes of a 2D lattice, can be obtained by considering the general Eq. (2.2) and replacing the appropriate excess energy into it, giving:

$$w_n = -sn + l_n\psi + l_{h,n}\psi_h. \tag{3.1}$$

The details of the last formula will be discussed later in the chapter. For now it is important to note that it is applicable in the framework of both ANT and CNT. However, in CNT, $n = i \times m$, $l_n = 2i$ and $l_{h,n} = 2m$ (as in Eq. (2.10)) and because of the assumed prismatic shape of all fibrils regardless of their size n the building β-sheets are equally long. In ANT, instead, l_n and $l_{h,n}$ are the numbers of all broken bonds parallel to the fibril m and i axes, n has only integer values

and the β-sheets can have different length, accounting for the molecular level shapes. For all these reasons the ANT fibril shape can depart from the prismatic shape postulated by CNT.

Given Eq. (3.1), expressions for the nucleus size n^*, nucleation work w^* and nucleation rate J can be obtained. In particular, following the same procedure already adopted in CNT, from the global maximum of w_n from Eq. (3.1), n^* and then w^* will be calculated for ANT, as described later in the chapter.

For the exact determination of the stationary nucleation rate J the following general exact Becker-Döring formula (10), (77) will be used :

$$J = \frac{1}{\sum_{n=1}^{M-1} \left(\frac{1}{f_n C_n} \right)}. \tag{3.2}$$

Here M is the size of a large enough supernucleus fibril, C_n (m^{-3}) is the equilibrium concentration of n-sized fibrils, and f_n (s^{-1}) is the frequency of attachment of monomer β-strands to an n-sized fibril. Eq. (2.3), given in the introduction of Chapter 2, is a good approximation to the exact formula (3.2). Furthermore, once the appropriate expression for C_n and f_n of the classical and atomistic description are introduced, Eqs. (2.3) and (3.2) are applicable to both CNT and ANT.

ANT originated in the pioneering papers on crystal nucleation and growth (73; 74; 75; 134; 135) and was later applied to other cases of nucleation (77; 95; 96; 104; 105; 106; 151). Recently (80), it has been shown that ANT describes much more accurately than CNT the nucleation rate of model two-dimensional (2D) and three-dimensional (3D) crystals because of its more realistic accounting for the actual shape of the smallest crystalline clusters. ANT was found to be superior also to CCNT despite that both theories were in good quantitative agreement with simulation and numerical data for the crystal nucleation rate. It is therefore important to verify whether ANT is successful in predicting the nucleation rate of amyloid fibrils when these are modelled at the molecular level.

In particular, the objective of the present chapter is (i) to employ ANT for describing the work to form the nanosized amyloid fibrils (protofilaments) and for determining the size of the fibril nucleus, the fibril nucleation work and the fibril nucleation rate, and (ii) to compare ANT and CCNT predictions for the fibril nucleus size, nucleation work and nucleation rate.

3.2 Physical model

In the following, both the ANT and the classical descriptions will be applied to a model peptide system studied previously (4; 5; 7; 8). In this system the peptides in the amyloid fibril represent the β-strands of the general description given in the previous chapter and the model in Fig 2.1 will be adapted to an ANT description.

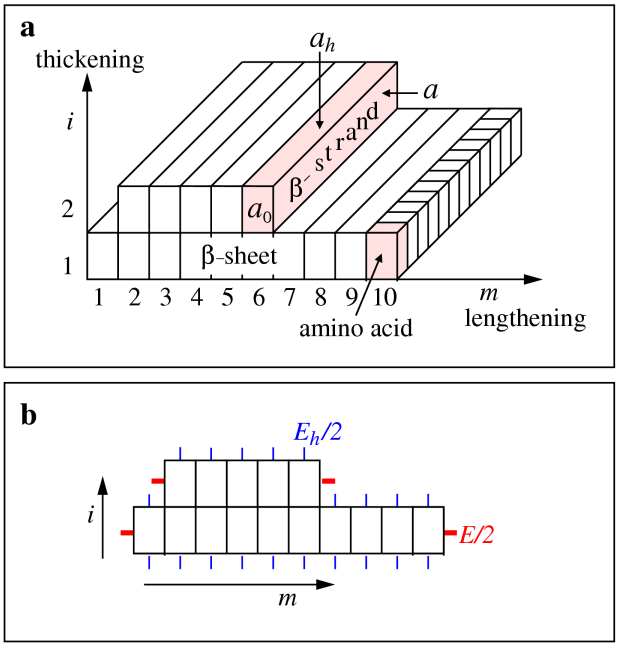

Figure 3.1: (a)Atomistic model of a nanosized amyloid fibril consisting of $n = 15$ β-strands arranged in two β-sheets, the first of them having ten β-strands, and the second of them having five β-strands. In this example each β-strand is composed of 12 amino acids which are schematised only in the rightmost β-strand. The a's are the areas of the three β-strand faces. (b) Corresponding fibril cross section in the (m, i) plane. The thick and the thin lines represent, respectively, the strong and the weak broken bonds of the β-strands at the periphery of the fibril cross section. In this example there are $l_n = 4$ strong broken bonds with energy $E/2$ per bond and $l_{h,n} = 20$ weak broken bonds with energy $E_h/2$ per bond.

In the ANT for amyloid fibrils, each amino acid in a peptide is represented by a right rectangular prism (twelve such amino acids are illustrated in Fig. 3.1a by the prisms in the rightmost peptide). The amino acids in the rod-like peptide are in a row and the peptide is thus also a right rectangular prism (Fig. 3.1a). In the

model, due to their strong hydrogen bonds, the peptide can arrange themselves laterally into β-sheets. The sheets consist of a different number m of peptides ($m = 1, 2, 3, ...$) and are parallel to the fibril lengthening axis (the m axis in Fig. 3.1a). Along its thickening axis (the i axis in Fig. 3.1a), the fibril is made of i β-sheets ($i = 1, 2, 3, ...$) which are held together by relatively weak bonds, such as hydrophobic-mediated bonds between the peptides. Since the fibril width is fixed and equal to the peptide length, the fibril can be considered as a 2D aggregate in the (m, i) plane, with building blocks (the peptides) arranged on a 2D lattice with simple rectangular symmetry (Fig. 3.1b). The application is valid under the assumption that each amino acid in a peptide can form hydrogen bonds solely along the fibril m axis, and that the hydrophobic-mediated bonds can be formed along both the m axis and the i axis of the fibril. Assuming further that the strength of these two kinds of bonds are the same for all amino acids and that bonds can only be formed between nearest-neighbour amino acids, the inter-peptide binding energy denoted by E (Joules) along the m axis can be written as

$$E = q\varepsilon + q_h\varepsilon_h. \tag{3.3}$$

Similarly, the inter-peptide binding energy E_h (Joules) along the i axis is given by

$$E_h = q_h\varepsilon_h. \tag{3.4}$$

In Eqs. (3.3) and (3.4) ε (Joules) is the energy of the hydrogen bond between two nearest-neighbour amino acids, ε_h (Joules) is the hydrophobicity-mediated bond energy of such amino acids, and q and q_h are the numbers of hydrogen and hydrophobic bonds between two nearest-neighbour peptides, respectively. For the model system considered here $a = a_h$ and $q = q_h = 10$ have been set. The latter differs somewhat from $q = 10$ and $q_h = 12$ in Refs. (4; 5; 7; 8), because in the model studied in those references, the first and the last of the 12 amino acids in a peptide did not form hydrogen bonds with nearest-neighbour amino acids. Also, for the ratio between the energies of the hydrogen and the hydrophobicity-mediated bonds it has been used $\varepsilon/\varepsilon_h = 9$, a value about twice as small as the one employed in Refs (4; 5; 7; 8). The interactions are then confined solely between nearest-neighbour β-strands in the fibril and the interstrand binding energies are $E = (100/9) \times \varepsilon$ and $E_h = (10/9) \times \varepsilon$. The corresponding broken-bond energy of a β-strand at the fibril σ or σ_h face is then $E/2$ or $E_h/2$, respectively (Fig. 3.1b).

Owing to the rectangular symmetry of the fibril lattice in the $(m,\ i)$ plane, to first approximation (73; 74; 80; 95; 134), σ and σ_h are related to E and E_h by $\sigma = E/2a$ and $\sigma_h = E_h/2a_h$. Taking into account Eqs. (2.11) and (2.12), ψ and ψ_h are thus

$$\psi = \frac{E}{2k_BT} \tag{3.5}$$

and

$$\psi_h = \frac{E_h}{2k_BT}. \tag{3.6}$$

As seen, physically, according to this picture, in ANT ψ or ψ_h is merely the dimensionless energy of the broken bond of a β-strand at the fibril σ or σ_h face, respectively. In accordance with Eqs. (2.11), (2.12), (3.5) and (3.6) and the last approximations, the $\varepsilon/\varepsilon_h$ value results in $\psi/\psi_h = 10$ for both the ANT and the classical descriptions.

3.3 ANT fibril evolution mechanism

In ANT, the determination of w_n and the ensuing calculation of n^*, w^* and J requires specification of the sequence of shapes that an aggregate takes during its evolution from the monomer size $n = 1$ to the considered size $n > 1$. In reality, there are many such sequences, but similar to CNT, ANT is restricted to considering a single sequence of shapes for which the aggregate total surface energy is as low as possible. A reference low-energy shape can be the CNT equilibrium shape, because the latter corresponds to the minimal total surface energy. This argument will be tested by kinetic Monte Carlo simulations of the growth of the nanosized fibrils in Chapter 6, since the choice of a fibril shape sequence for theoretical analysis can affect the results obtained.

Fig. 3.2 exemplifies fibril shapes in the (m, i) plane from a sequence which models amyloid fibril nucleation in the case of $\psi/\psi_h = n_t = 10$ and which will be used in the considerations to follow in order to obtain general results for any $\psi/\psi_h \geq 1$. In figure 3.2, from $n = 1$ to the transition size $n_t = 10$, the fibril grows as a single β-sheet, i.e. it exhibits 1D growth in order to keep minimal its total surface energy or, equivalently, the total energy of its broken bonds l_n and $l_{h,n}$. According to CNT, transition from 1D to 2D growth occurs at n_t so that the $(n_t + 1)$-sized fibril should already be more than one β-sheet thick. This

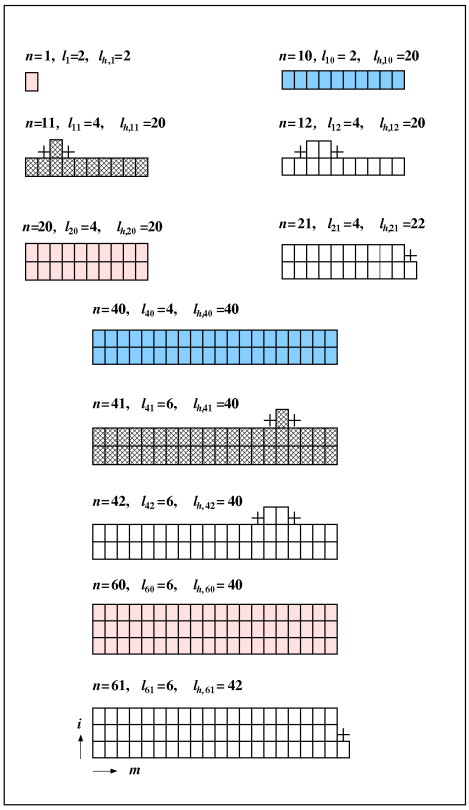

Figure 3.2: Fibril cross sections in the (m, i) plane which illustrate partially the sequence of fibril shapes during the evolution of an amyloid fibril characterised by $\psi/\psi_h = 10$. From the fibril nuclei, hatched sizes), the β-strand attachment occurs at no surface-energy cost up to the pink nanofibril; from that size on, the fibril growth corresponds to jagged lines in the barrier up to the blue fibril sizes that are followed by fibril nuclei (hatched). The plus signes indicate the kink sites.

fibril (the one of size $n = 11$ in Fig. 3.2) can therefore be assumed to have the shape of 1β-sheet of length n_t with one peptide monomer adsorbed side-wise. Importantly, this monomer gives birth to the fibril second β-sheet and creates two kink sites at the fibril surface (the sites are marked by plus signs in Fig. 3.2). Further peptide attachment to a kink site does not increase the fibril total surface energy, because fibrils of size $n = n_t + 1, n_t + 2, ..., 2n_t$ have the same number of broken bonds l_n and $l_{h,n}$. In Fig. 3.2 these are the fibrils with n =11, 12, 20. Thus, the growth of the second β-sheet continues at no surface-energy cost until $n = 2n_t = 20$. Since from $n > n_t = 10$ to $n = 2 \times 2n_t = 40$ the CNT equilibrium shape requires fibril thickness between one and two β-sheets, the attachment of

the next peptide monomer to the $2n_t$-sized fibril can be assumed to occur at one of the two fibril ends. This lengthwise monomer attachment creates a kink site at the end of the fibril of size $n = 2n_t + 1 = 21$ (see Fig. 3.2) and the next monomer is attached to this site with no surface-energy cost. Thus, further fibril growth occurs by alternating creation and annihilation of kink sites at the fibril ends until the fibril reaches the size $n = 2 \times 2n_t = 40$.

At this size, the attachment of the next peptide monomer can be assumed to occur again to the surface of one of the two β-sheets of the fibril, because the CNT equilibrium shape requires a fibril thickness between two and three β-sheets for fibril sizes from $n > 2 \times 2n_t = 40$ to $n = 3 \times 3n_t = 90$. The fibril of size $n = 2 \times 2n_t + 1$, the one with $n = 41$ in Fig. 3.2, will therefore have the shape of a 2β-sheet of length $2n_t$ with one peptide monomer adsorbed on one of the β-sheet surfaces rather than on one of the β-sheet ends. This monomer triggers the growth of the fibril third β-sheet, because it generates two kink sites at which subsequent peptides can attach themselves without changing the number of the fibril broken bonds. In Fig. 3.2, the fibrils with $n = 41$ and 42 illustrate this situation.

When the third β-sheet is completed at size $3 \times 2n_t = 60$, the fibril growth continues by lengthwise attachment of the next peptide monomer (see the fibrils with n =60 and 61 in Fig. 3.2). Subsequent lengthwise attachment of peptides to the fibril is thermodynamically favoured only up to size $n = 3 \times 3n_t = 90$, since between $n = 3 \times 3n_t + 1$ and the size $n = 4 \times 4n_t = 160$ the fibril has to increase its thickness by one more β-sheet in order to keep its shape close to the CNT equilibrium shape. Thus, it can again be assumed that the fibril of size $n = 3 \times 3n_t + 1 = 91$ will have the shape of a 3β-sheet of length $3n_t$ with one peptide adsorbed on the surface of the fibril β-sheets.

Further growth of the fibril can be expected to occur in the above manner provided it is not disturbed by other processes, e.g., by fibril flocculation. Using the sequence of fibril shapes described above and partially illustrated in Fig. 3.2 for the model peptide system considered, which is characterised by the value of $\psi/\psi_h = 10$, the non-analytical n-dependence of the numbers l_n and $l_{h,n}$ of the fibril strong and weak broken bonds will be determined. This determination will make it possible to calculate the ANT w_n, n^*, $l^* = l_{n^*}$, $l_h^* = l_{h,n^*}$, w^* and J.

3.4 Work of fibril formation

According to ANT (80), for any n-sized fibril the role of the part $2a\sigma i+2a_h\sigma_h m$ of the fibril excess energy $\Phi_{i,m}$ from Eq. (2.9) is played by the sum $l_nE/2+l_{h,n}E_h/2$ of all broken-bond energies at the periphery of the fibril cross-section with the (m,i) plane, i.e.

$$2a\sigma i + 2a_h\sigma_h m = \frac{l_nE}{2} + \frac{l_{h,n}E_h}{2}, \tag{3.7}$$

where l_n or $l_{h,n}$, as already stated, is the total number of the n-sized fibril strong and weak broken bonds that are parallel to the fibril m or i axis with corresponding energy $E/2$ or $E_h/2$, respectively (see Fig. 3.1). Combining Eq. (2.2), (2.9), (3.5)-(3.7) and recalling that $s \equiv (\mu_s-\mu_f-2a_0\sigma_0)/k_BT$ leads to Eq. (3.1) for the dimensionless work w_n to form an n-sized fibril ($n = 1, 2, 3, ...$). In general, the differently determined fibril shape in CNT and ANT leads to a quantitative difference in the prediction of the work w_n to form an n-sized fibril. When the fibrils have the CNT equilibrium shape, i and m are not independent variables, but are connected by Eq. (2.17). Using this equation and the relation $n = i \times m$ yields $i = (\psi_h/\psi)^{1/2}n^{1/2}$ and $m = (\psi/\psi_h)^{1/2}n^{1/2}$ so that in CNT ($n \geq n_t = \psi/\psi_h$),

$$l_n = 2\left(\frac{\psi_h}{\psi}\right)^{1/2} n^{1/2} \tag{3.8}$$

and

$$l_{h,n} = 2\left(\frac{\psi}{\psi_h}\right)^{1/2} n^{1/2}. \tag{3.9}$$

Substitution of l_n and $l_{h,n}$ from these formulae into Eq. (3.1) transforms this equation into Eq. (2.18). This means that the CNT and ANT descriptions yield the same w_n values, but only for those n-sized fibrils that happen to have the CNT equilibrium shape. Regrettably, in ANT the numbers l_n and $l_{h,n}$ of fibril broken bonds are not analytical functions of the fibril size n. They are obtainable solely by considerations at the molecular level. Hence, given the n value, the ANT l_n and $l_{h,n}$ have to be calculated with the aid of, e.g. a computer program for all possible fibril shapes corresponding to this n value. For the peptide model considered here it has been selected a particular sequence of shapes, having as a reference the CNT evolution mechanism, which specifies for each n the numbers of broken bonds l_n and $l_{h,n}$ as in Fig. 3.2. By directly counting the number of broken bonds at the fibril periphery in the (m,i) plane for nucleation of fibrils

with the shape sequence illustrated partially in Fig. 3.2, it has been found that l_n and $l_{h,n}$ have the values given in Table 3.1 for $n = 1, 2, 3, ..., 300$.

Using the l_n and $l_{h,n}$ values of Table 3.1 in Eq. (3.1) yields the ANT dependence of the fibril excess energy Φ_n and then of the work w_n for formation of fibrils with the shape sequence illustrated in Fig. 3.2.

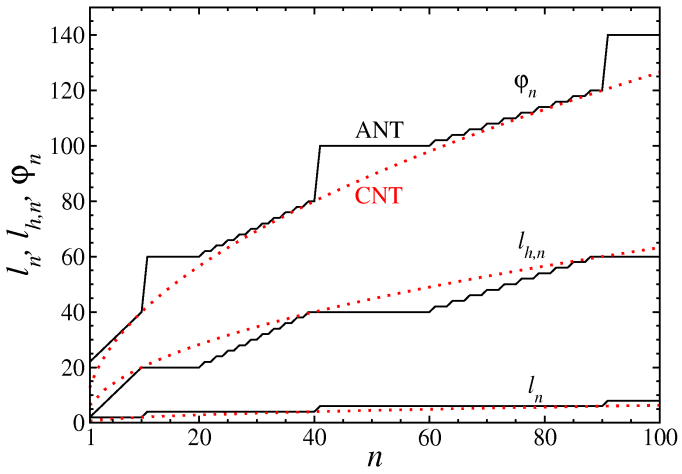

Figure 3.3: Dependence of $l_n, l_{h,n}$ and φ_n on the fibril size at $\psi/\psi_h = 10$: solid lines ANT $l_n, l_{h,n}$ from Table 3.1 and ANT φ_n from the equation $\varphi_n \equiv l_n\psi_n + l_{h,n}\psi_{h,n}$; dotted lines CNT Eqs. (3.8), (3.9) and $\varphi_n = 4(\psi\psi_n)^{1/2}n^{1/2}$.

The ANT dependence of l_n and $l_{h,n}$ and of the dimensionless fibril excess energy $\Phi_n/k_BT = \varphi_n \equiv l_n\psi_n + l_{h,n}\psi_{h,n}$ of the fibril total surface energy is illustrated in Fig. 3.3 by the solid lines. For comparison, the dotted lines in the figure show the CNT $\Phi_n/k_BT = \varphi_n \equiv 4(\psi\psi_h)^{1/2}n^{1/2}$, l_n and $l_{h,n}$ from Eqs. (2.18), (3.8) and (3.9). From the figure, it can be seen that CNT agrees with ANT only for those fibril sizes (n = 10, 40, and 90) for which the ANT and CNT fibril shapes are identical. Also, Fig. 3.3 shows that CNT underestimates l_n and overestimates $l_{h,n}$ for all other n values. While in ANT the initiation of a new β-sheet occurs stepwise at sizes n = 11, 41, and 91 at which the first peptide monomer of this β-sheet is adsorbed on a fibril σ_h face, in CNT at these sizes the new β-sheet starts to build up continuously on such a face. Thus, in ANT the first adsorbed peptide on the σ_h face creates two new strong broken bonds and afterwards l_n remains unchanged until the fibril reaches the next size at which the ANT and CNT shapes become identical. In contrast, the adsorption of the first peptide

n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$
1	2	2	61	6	42	121	8	62	181	10	80	241	10	98
2	2	4	62	6	42	122	8	62	182	10	80	242	10	98
3	2	6	63	6	42	123	8	62	183	10	80	243	10	98
4	2	8	64	6	44	124	8	62	184	10	80	244	10	98
5	2	10	65	6	44	125	8	64	185	10	80	245	10	98
6	2	12	66	6	44	126	8	64	186	10	80	246	10	100
7	2	14	67	6	46	127	8	64	187	10	80	247	10	100
8	2	16	68	6	46	128	8	64	188	10	80	248	10	100
9	2	18	69	6	46	129	8	66	189	10	80	249	10	100
10	2	20	70	6	48	130	8	66	190	10	80	250	10	100
11	4	20	71	6	48	131	8	66	191	10	80	251	12	100
12	4	20	72	6	48	132	8	66	192	10	80	252	12	100
13	4	20	73	6	50	133	8	68	193	10	80	253	12	100
14	4	20	74	6	50	134	8	68	194	10	80	254	12	100
15	4	20	75	6	50	135	8	68	195	10	80	255	12	100
16	4	20	76	6	52	136	8	68	196	10	80	256	12	100
17	4	20	77	6	52	137	8	70	197	10	80	257	12	100
18	4	20	78	6	52	138	8	70	198	10	80	258	12	100
19	4	20	59	6	54	139	8	70	199	10	80	259	12	100
20	4	20	80	6	54	140	8	70	200	10	80	260	12	100
21	4	22	81	6	54	141	8	72	201	10	82	261	12	100
22	4	22	82	6	56	142	8	72	202	10	82	262	12	100
23	4	24	83	6	56	143	8	72	203	10	82	263	12	100
24	4	24	84	6	56	144	8	72	204	10	82	264	12	100
25	4	26	85	6	58	145	8	74	205	10	82	265	12	100
26	4	26	86	6	58	146	8	74	206	10	84	266	12	100
27	4	28	87	6	58	147	8	74	207	10	84	267	12	100
28	4	28	88	6	60	148	8	74	208	10	84	268	12	100
29	4	30	89	6	60	149	8	76	209	10	84	269	12	100
30	4	30	90	6	60	150	8	76	210	10	84	270	12	100
31	4	32	91	8	60	151	8	76	211	10	86	271	12	100
32	4	32	92	8	60	152	8	76	212	10	86	272	12	100
33	4	34	93	8	60	153	8	78	213	10	86	273	12	100
34	4	34	94	8	60	154	8	78	214	10	86	274	12	100
35	4	36	95	8	60	155	8	78	215	10	86	275	12	100
36	4	36	96	8	60	156	8	78	216	10	88	276	12	100
37	4	38	97	8	60	157	8	80	217	10	88	277	10	100
38	4	38	98	8	60	158	8	80	218	10	88	278	10	100
39	4	40	99	8	60	159	8	80	219	10	88	279	12	100
40	4	40	100	8	60	160	8	80	220	10	88	280	12	100
41	6	40	101	8	60	161	10	80	221	10	90	281	12	100
42	6	40	102	8	60	162	10	80	222	10	90	282	12	100
43	6	40	103	8	60	163	10	80	223	10	90	283	12	100
44	6	40	104	8	60	164	10	80	224	10	90	284	12	100
45	6	40	105	8	60	165	10	80	225	10	90	285	12	100
46	6	40	106	8	60	166	10	80	226	10	92	286	12	100
47	6	40	107	8	60	167	10	80	227	10	92	287	12	100
48	6	40	108	8	60	168	10	80	228	10	92	288	12	100
49	6	40	109	8	60	169	10	80	229	10	92	289	12	100
50	6	40	110	8	60	170	10	80	230	10	92	290	12	100
51	6	40	111	8	60	171	10	80	231	10	94	291	12	100
52	6	40	112	8	60	172	10	80	232	10	94	292	12	100
53	6	40	113	8	60	173	10	80	233	10	94	293	12	100
54	6	40	114	8	60	174	10	80	234	10	94	294	12	100
55	6	40	115	8	60	175	10	80	235	10	94	295	12	100
56	6	40	116	8	60	176	10	80	236	10	96	296	12	100
57	6	40	117	8	60	177	10	80	237	10	96	297	12	100
58	6	40	118	8	60	178	10	80	238	10	96	298	12	100
59	6	40	119	8	60	179	10	80	239	10	96	299	12	100
60	6	40	120	8	60	180	10	80	240	10	96	300	12	100

Table 3.1: ANT numbers l_n and $l_{h,n}$ of, respectively, the strong and weak broken bonds of an n-sized amyloid fibril at the periphery of the fibril cross section in the (m, i) plane ($\psi/\psi_h = n_t = 10$).

monomer leaves $l_{h,n}$ unchanged for all subsequent fibril sizes until the new β-sheet is fully formed (Fig. 3.3). When the fibril grows further, $l_{h,n}$ increases stepwise (Fig. 3.3) due to the alternating creation and annihilation of kink sites at the fibril ends until the fibril assumes the next size at which the ANT and CNT shapes become identical. Importantly, CNT underestimates the part φ_n of fibril total surface energy for all n values for which the ANT and CNT shapes differ. This underestimation turns out to be the main reason for which CNT predicts a much lower nucleation rate than the one obtained using Eq. (3.2) for ANT.

The w_n plots at $s/\psi_h = 1$, 4/3, 2 and 3 are seen in Fig. 3.4. The jaggedness

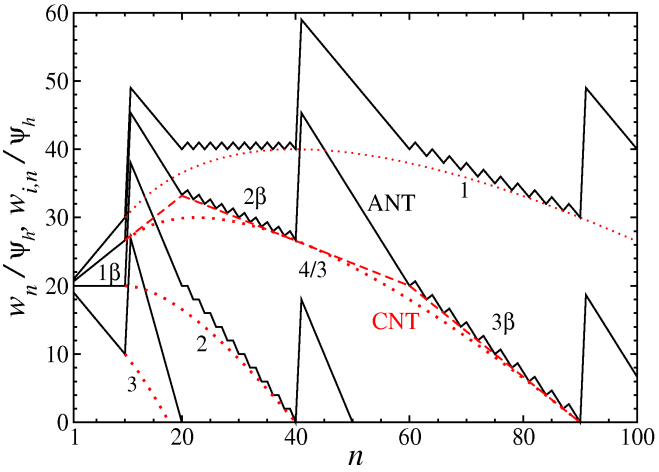

Figure 3.4: Dependence of w_n and $w_{i,n}$ on the fibril size at ψ/ψ_h= 10 at scaled supersaturation s/ψ_h=1, 4/3, 2 or 3 (as indicated): solid lines ANT Eq. (3.1) with l_n and $l_{h,n}$ from Table 3.1; dotted lines CNT Eq. (2.18); dashed lines 1β, 2β and 3β CNT Eq. (2.13) at i = 1, 2 or 3 for 1β-sheet, 2β-sheet or 3β-sheet, respectively.

of the dependences is inherited from that of the l_n and $l_{h,n}$ dependences in Fig. 3.4 and reflects the fibril nucleation mechanism. The sharp peak of w_n at fibril size n = 11, 41, or 91 corresponds to the adsorption of the first peptide monomer on a fibril outer β-sheet (the ascending straight line) and the ensuing completion of the newly born β-sheet (the descending straight line). The jagged line after the peak reflects the consequent elongation of the fibril by alternating creation and annihilation of kink sites at the fibril ends until the fibril reaches the size corresponding to the next peak. For comparison, by dashed and dotted lines, respectively, in Fig. 3.4 also the CNT $w_{i,n}$ from Eq. (2.13) (at $s/\psi_h = 4/3$ and

$i = 1$, 2 and 3) and CNT w_n from Eq. (2.18) (at $s/\psi_h = 1$, 4/3, 2 and 3) are displayed. In conformity with the l_n and $l_{h,n}$ dependences, CNT and ANT w_n agree only for $n = 10$, 40, and 90, and CNT from Eq. (2.13) provides a tangent to the ANT w_n function on the side of the minima. As seen in Fig. 3.4, the CNT $w_{i,n}$ lines for successive $i\beta$-sheets form a broken straight line enveloping the CNT w_n line. The reason for which the former is above the latter is that w_n is derived under the condition of equilibrium shape of the fibrils. On the other hand, the CNT $w_{i,n}$ lines connect the minima of the respective ANT w_n line because $w_{i,n}$ corresponds to fibrils with strictly rectangular shape, and does not account for the atomistic irregularities in the actual fibril shape. As already noted, in CNT the supersaturation $s_{1,\beta} = 2\psi_h$ is the divider between the nucleation ($s \leq 2\psi_h$) and metanucleation ($s > 2\psi_h$) regimes of fibril formation. In the latter regime, each monomer peptide is a nucleus and the fibrillation process occurs spontaneously. In contrast, in ANT the nucleus size is still $n^* = 11$ when $s > 2\psi_h$. Although the work is gained on building up the 1β-sheet up to $n = 10$ (see line 3 in Fig. 3.4), the adsorption of the first peptide monomer on the β-sheet surface requires the creation of two kink sites and costs a high additional energy. This energy represents a barrier to the fibril overgrowth and for that reason, as predicted by CCNT (81), for $s > 2\psi_h$ fibril formation is still nucleation-mediated, but in a non-classical way. Only at a supersaturation sufficiently higher than $s_{1\beta}$ does this barrier vanish and the fibrils then form in the metanucleation regime.

3.5 Nucleus size and nucleation work

The fibril shape characterised by the lowest total surface energy has been used for determining the ANT w_n dependence from Eq. (3.1). The calculation of the respective nucleus size n^*, nucleation work w^* can be done as explained in the previous sections. Given the supersaturation s, in ANT the nucleus size n^* can be determined only numerically from the global maximum of w_n from Eq. (3.1). Importantly, ANT predicts that n^* has different s-independent integer values in successive s ranges, which means that n^* is a stepwise function of s. For that reason, the $n^*(s)$ dependence which corresponds to those given by the classical and corrected Gibbs-Thomson Eq. (2.19) and (2.27) cannot be represented analytically. Nonetheless, using the obtained numerical n^* values for calculating the numbers $l^* \equiv l_{n^*}$ and $l_h^* \equiv l_{h,n^*}$ of strong and weak broken-bonds of the fibril

nucleus allows numerical determination of the supersaturation dependence of the dimensionless nucleation work w^* with the help of the expression ($s \geq 0$)

$$w^* = -sn^* + l^*\psi + l_h^*\psi_h, \tag{3.10}$$

which follows from Eq. (3.1) and in which all n^*, l^* and l_h^* are s-independent positive integers in a given supersaturation range. For that reason, the ANT nucleation work from Eq. (3.10) obeys the nucleation theorem in the form (77), (83), (78) $dw^*/ds = -n^*$. Also, at $\psi/\psi_h = 1$, Eq. (3.10) passes into that in Ref. (80) for 2D nucleation of crystals.

The supersaturation at which the ANT nucleus changes its size stepwise from n_j^* to n_{j+1}^* ($j = 1, 2, 3, ...$) is the so-called transition supersaturation $s_{t,j}$. This quantity can be determined from the crossing of two successive $w^*(s)$ straight lines, i.e. from the definition equation $w^*(s_{t,j}, n_j^*) = w^*(s_{t,j}, n_{j+1}^*)$ (106). Using in this equation w^* from Eq. (3.10) leads to the general expression ($j = 1, 2, 3, ...$)

$$s_{t,j} = [(l_{j+1}^* - l_j^*)\psi + (l_{h,j+1}^* - l_{h,j}^*)\psi_h]/(n_{j+1}^* - n_j^*). \tag{3.11}$$

For the model peptide system considered, the ANT lines in Fig. 3.4 show that the nucleus size is $n_1^* = 11$ for $s/\psi_h > 4/3$ and jumps to $n_2^* = 41$ at the transition supersaturation $s_{t,1} = 4/3$. The fibril nuclei of these two sizes are shown shaded in Fig. 3.2. The values of the first five nucleus sizes and transition supersaturations, listed in Table 3.2, are obtained from w_n plots as those in Fig. 3.4 but for sufficiently wide n and s ranges.

Table 3.2 confirms that an outstanding characteristic feature of ANT for amyloid nucleation is that the size of fibril nucleus is a constant integer in a given supersaturation range.

Inspection of Table 3.2 shows that for the model peptide system considered l_j^*, $l_{h,j}^*$, n_j^* and $s_{t,j}/\psi_h$ are related by the simple expressions ($j = 1, 2, 3, ...$)

$$l_j^* = 2(j+1), \tag{3.12}$$

$$l_{h,j}^* = 2jn_t, \tag{3.13}$$

$$n_j^* = 1 + j^2 n_t, \tag{3.14}$$

$$\frac{s_{t,j}}{\psi_h} = \frac{4}{2j+1}. \tag{3.15}$$

j	$s_{t,j}/\psi_h$	s/ψ_h range	n^*	$m^*_{j,1}$	$m^*_{j,2}$	$m^*_{j,3}$	$m^*_{j,4}$	$m^*_{j,5}$	$m^*_{j,6}$	l^*_j	$l^*_{h,j}$
1	4/3	$> 4/3$	11	10	1	0	0	0	0	4	20
2	4/5	4/3-4/5	41	20	20	1	0	0	0	6	40
3	4/7	4/5-4/7	91	30	30	30	1	0	0	8	60
4	4/9	4/7-4/9	161	40	40	40	40	1	0	10	80
5	4/11	4/9-4/11	251	50	50	50	50	50	1	12	100

Table 3.2: ANT transition supersaturations $s_{t,j}$ and supersaturation ranges in which the fibril nucleus has constant size of n^*_j peptides ($j = 1, 2, 3,...$). The j^{th} nucleus has $m^*_{j,p}$ peptides in its p^{th} β-sheet ($p = 1, 2, 3,...$) so that $n^*_j = m^*_{j,1} + m^*_{j,2} + m^*_{j,3} + ...$ The last two columns list the numbers l^*_j and $l^*_{h,j}$ of the strong and weak broken bonds at the periphery of the nucleus cross section in the (m, i) plane. The values are relevant for $\psi/\psi_h = n_t = 10$.

When $n_t = \psi/\psi_h = 10$, these equations say, for example, that the first ANT nucleus ($j = 1$) has four strong and 20 weak broken bonds, that it contains 11 peptides, and the first transition supersaturation $s_{t,1}$ equals $4\psi_h/3$. It can be noted as well that Eq. (3.15) follows from Eq. (3.11) in the particular case of l^*_j, $l^*_{h,j}$, and n^*_j given by Eqs. (3.12)-(3.14).

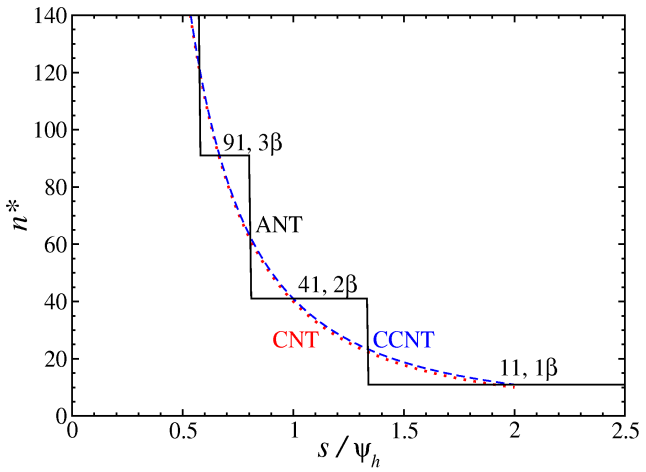

Figure 3.5: Supersaturation dependence of the nucleus size at $\psi/\psi_h = 10$: red dotted line CNT Eq. (2.19); blue dashed line CCNT Eq. (2.27); black solid line ANT Eq. (3.14).

The ANT $n^*(s)$ dependence presented in Table 3.2 is depicted by the stepped line in Fig. 3.5. The figure clearly demonstrates that $n^*_j(s)$, from Eq. (3.14),

is a constant integer in a given supersaturation range and exhibits jumps at the transition supersaturations $s_{t,j}$, from Eq. (3.15). In CNT and CCNT the $n^*(s)$ dependence is described by the Gibbs-Thomson Eqs. (2.19) and (2.27), both of which approximate relatively well the size even of the smallest ANT nucleus. In Fig. 3.6 the ANT dependence of $w^*(s)$ dependence obtained from Eq. (3.10) with the aid of the n^*, l^* and l_h^* values from Table 3.2 is illustrated.

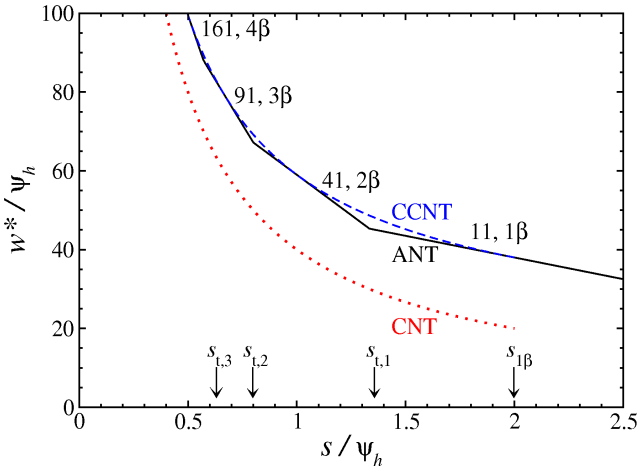

Figure 3.6: Supersaturation dependence of the nucleation work w^* at ψ/ψ_h = 10: red dotted line CNT Eq. (2.20); blue dashed line CCNT Eq. (2.28); black solid line ANT Eq. (3.16). The arrows point at the first three transition supersaturations and at the supersaturation $s_{1,\beta}$ which limits the CNT and CCNT applicability. The numbers indicate the constant ANT nucleus size in the corresponding s range.

It can be observed that the ANT nucleation work is a broken linear function of s and the slopes of the linear portions determine the corresponding n^* values. This is so, because according to Eq. (3.10), the ANT $w^*(s)$ is a linear function of s if the nucleus size n^* and the numbers l_n and $l_{h,n}$ of the nucleus strong and weak broken bonds are constant. The s dependence of the j^{th} nucleation work w_j^* is given by the formula ($j = 1, 2, 3...$):

$$w_j^* = -(1 + j^2 n_t)s + 2(2j + 1)\psi, \tag{3.16}$$

which follows from Eqs. (3.10) and (3.12)-(3.14). For comparison, by dotted and dashed lines, in Fig. 3.6 is also shown the respective CNT and CCNT $w^*(s)$ dependences from Eqs. (2.20) and (2.28). It is seen that while CNT

greatly underestimates the ANT nucleation work, there is a good quantitative agreement between the CCNT and ANT description except around the first transition supersaturation $s_{t,1}$.

3.6 Nucleation rate

For the determination of the ANT nucleation rate J the exact formula given in the introduction Eq. (3.2) has been used. In the formula (3.2) the frequency of attachment can be approximated as $f_n = f_1 l_n/2$, where the proportionality of f_n to l_n reflects the fact that the β-strands are attached predominantly to the ends of the fibril β-sheets (the fibril hydrogen broken bonds are at these ends), and the divisor 2 takes into account that the monomer β-strand has two hydrogen broken bonds. With the aid of C_n from the general formula (77), $C_n = C_1 \exp(w_{1,1} - w_n)$ (cf. Eq. (2.24)),$w_{1,1}$ from Eq. (2.14), w_n from Eq. (3.1), and f_n from the above approximation, Eq. (3.2) leads to the Becker-Döring ANT (BDANT) expression for the fibril nucleation rate ($s \geq 0$):

$$J(s) = \frac{f_1 C_1}{\sum_{n=1}^{M-1} \exp[-s(n-1) + (l_n - 2)\psi + (l_{h,n} - 2)\psi_h]}. \tag{3.17}$$

The above approximation for f_n has a negligible effect on the J values calculated from Eq. (3.17). For that reason, the $J(s)$ dependence from Eq. (3.17) will be used for a reliable quantitative verification not only of the CNT and CCNT $J(s)$ dependences from Eq. (2.26) and (2.29), but also of the ANT $J(s)$ dependence from Eqs. (3.18) obtained below.

As known (e.g., ref. (77)) Eq. (2.3) is a good general approximation to the Becker-Döring Eq. (3.2). For that reason as in CNT and CCNT, an approximate ANT formula for J can be obtained from Eq. (2.3), in combination with Eq. (2.24) for C^* and Eq. (3.10) for the ANT w^*. Also, the CCNT approximations $z = 1/2$ and $f^* = 2f_1$ can be used for the Zeldovich factor z and the attachment frequency f^* in Eq. (2.3). Thus, since in ANT $w_{1,1}$ is again given by $w_{1,1} = -s + 2(\psi + \psi_h)$, using Eq. (2.3) the ANT $J(s)$ dependence can be written in the form ($s \geq 0$)

$$J(s) = f_1(s) C_1 \exp[(n^* - 1)s - (l^* - 2)\psi - (l_h^* - 2)\psi_h]. \tag{3.18}$$

In this expression, ψ and ψ_h are specified by Eqs. (3.5) and (3.6), and n^*, l^* and l_h^* are s-independent positive integers in a given supersaturation range. As shown previously, these three nucleus parameters are obtainable solely by model considerations at the molecular level. Assigning different values of the supersaturation s, from Eqs. (3.17) and (3.18) one can obtain the $J(s)$ dependence for any amyloidogenic protein solution. In the experimentally important case of fibril nucleation at fixed temperature T, the supersaturation s and the product $f_1(s)C_1$ are given by (81) $s = \ln(C_1/C_e)$ and $f_1(s)C_1 = f_{1,e}C_e \exp(2s)$. It thus follows from Eqs. (3.17) and (3.18) that in this case the explicit BDANT and ANT $J(C1)$ dependences read ($C_1 \geq C_e$):

$$J(C_1) = \frac{f_{1,e}C_e}{\sum_{n=1}^{M-1}(2/l_n)(C_1/C_e)^{-(n+1)}\exp[(l_n-2)\psi + (l_{h,n}-2)\psi_h]} \tag{3.19}$$

and

$$J(C_1) = f_{1,e}C_e\left(\frac{C_1}{C_e}\right)^{n^*+1}\exp[-(l^*-2)\psi - (l_h^*-2)\psi_h]. \tag{3.20}$$

respectively. Equation (3.20) parallels the ANT $J(C_1)$ formula for nucleation in solutions (77) and in it n^*, l^* and l_h^* have different C_1-independent integer values in successive C_1 ranges.

The BDANT $J(s)$ dependence was obtained from Eq. (3.17) with the l_n, $l_{h,n}$ values from Table 3.1, $\psi = 12.0$, $\psi_h = \psi/10 = 1.2$, and $M = 300$. This M value was chosen, because larger values had no effect on the J values calculated from Eq. (3.17) at all supersaturation values studied. The resulting BDANT $J(s)$ dependence is illustrated in Fig. 3.7 by circles.

Also shown in Fig. 3.7 is the ANT $J(s)$ dependence (the solid line) obtained from Eq. (3.18) with n^*, l^* and l_h^* from Table 3.2 and with the above ψ and ψ_h values. All n^*, l^* and l_h^* are different s-independent integers in the successive s ranges limited by the respective transition supersaturations $s_{t,j}$, the first two of which are indicated in the figure. For that reason, the ANT $\ln J$ is a broken linear function of s and the slopes of the successive straight lines are determined by the corresponding n^*. As seen, the ANT $\ln J$ has a behaviour similar to that of the BDANT $\ln J$, which exhibits sharp bends followed by practically linear portions. The same similarity between nucleation rates obtained from the Becker-Döring equation (3.2) and that predicted by ANT has also been observed in nucleation of 2D and 3D Kossel crystal (80). The agreement of the ANT nucleation rate

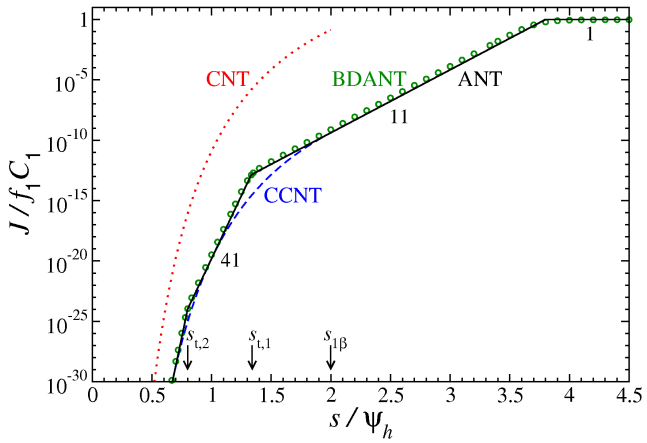

Figure 3.7: Supersaturation dependence of the nucleation rate J at $\psi = 12.0$ and $\psi_h = 1.2$: circles BDANT Eq. (3.17); solid line ANT Eq. (3.18); dashed line CCNT Eq. (2.29); dotted line CNT Eq. (2.26). The arrows point at the first two transition supersaturations and at the supersaturation $s_{1,\beta}$ which limits the CCNT and CNT applicability. The numbers indicate the constant ANT nucleus size in the corresponding s range.

with the BDANT one is remarkable, especially at high supersaturations where the ANT Eq. (3.18) captures the sharp bends of the BDANT $J(s)$ dependence at the consecutive transition supersaturations. Equation (3.18) works well even for supersaturations $s > s_{1\beta} = 2\psi_h$, i.e. in the s range where CNT and CCNT are not applicable. This is due to the fact that in this s range the n_1^*-sized fibril ($n_1^* = 11$) continues to act as nucleus and a work greater than $w_{1,1}$ is still needed for its formation (see solid line 3 in Fig. 3.4). At $s = 3.8\psi_h$, the nucleation work w_1^* becomes equal to $w_{1,1}$ and the nucleation barrier vanishes. At this s value n^* diminishes stepwise from 11 to 1 (i.e. the peptide monomer becomes nucleus) and, accordingly, l^* and l_h^* decrease stepwise from 4 to 2 and from 20 to 2, respectively. Using $n^* = 1$, $l^* = 2$ and $l_h^* = 2$ in the ANT Eq. (3.18) allows determining the nucleation rate for $s > 3.8\ \psi_h$, the result being $J = f_1(s)C_1$ (this relation is represented by the horizontal portion of the solid line in Fig. 3.7). At such high supersaturations, fibril formation for the model peptide system considered occurs in the metanucleation regime.

The CNT and CCNT $J(s)$ dependences from Eq. (2.26) and Eq. (2.29) at the above ψ and ψ_h values are illustrated in Fig. 3.7 by the dotted and dashed lines, respectively. As seen, there is a stark quantitative disagreement (about eight orders of magnitude) between CNT and all BDANT, ANT and CCNT.

The main reason for this is that CNT greatly underestimates the nucleation work w^* (see Fig. 3.6). In contrast, $J(s)$ of CCNT agrees quite well with that of both BDANT and ANT. It deviates most markedly from them only around the first transition supersaturation $s_{t,1}$ because of the greatest difference between the ANT and CCNT nucleation works around this transition supersaturation (Fig. 3.6).

3.7 Conclusion

Application of ANT for describing the nucleation of amyloid fibrils when the process occurs by direct polymerisation of peptides or protein segments into β-sheets leads to general expressions for the work to form a nanosized fibril and for the fibril nucleus size, nucleation work and nucleation rate. The main difference between the ANT description and that of both CNT and CCNT is that while the latter considers all fibrils as having the shape of rectangular prism, the ANT captures the fine details of the fibril shape. It is this explicit consideration of the fibril shape at molecular level that makes the ANT model interesting, allowing it to provide qualitative new results.

Comparison of the ANT expressions for w_n, n^*, w^* and J with the corresponding CNT and CCNT ones illustrates the principal differences between the atomistic and classical descriptions, which are exemplified by the studied model peptide system. The analysis reveals that the CNT work w_n to form a fibril underestimates the ANT one for all fibril sizes, except for those at which the CNT and ANT fibril shapes are identical. A prominent characteristic feature of ANT is that the fibril nucleus size n^* is a constant integer in a given supersaturation range, whereas in CNT and CCNT n^* diminishes continuously with increasing s, as described by the Gibbs-Thomson equation (2.19) and (2.27). The characteristic jumps in the ANT $n^*(s)$ dependence imply that the ANT nucleation work w^* is a linear function of the supersaturation s in a given s range. Comparison of the CNT and ANT Eqs. (2.20) and (3.10) for the fibril nucleation work w^* shows that this quantity is greatly underestimated by CNT. This underestimation is almost completely eliminated by CCNT upon accounting that the CNT nucleation work should include the work to attach that peptide or protein segment to the surface of the CNT nucleus, which generate a new β-sheet on this nucleus. The

resulting CCNT Eq. (2.28) for the nucleation work agrees well with the ANT Eq. (3.10).

The stark quantitative disagreement between CNT and the CCNT and ANT formulae for the nucleation work w^* has a crucial impact on the CNT and the ANT predictions for the fibril nucleation rate J. The CNT-predicted J values are greater than the practically exact BDANT ones by many orders of magnitude. In contrast both the CCNT nucleation rate from Eq. (2.29) and, especially, the ANT one from Eq. (3.18) agree very well with that predicted by the BDANT Eq. (3.17). Remarkably, the ANT $J(s)$ dependence works very well even at supersaturations higher than the maximum supersaturation $s_{1\beta}$ where CNT and CCNT are not applicable. So far, CCNT has been applied only to nucleation of 2D and 3D Kossel crystals, (80), and again has been successful in describing the supersaturation dependence of the nucleation rate. Despite its success however, as the corrections introduced by it are in the scope of CNT, CCNT is limited by the major CNT weakness of not accounting for the possible dependence of the fibril surface energy on the fibril size.

It is worth pointing out that changing the ratio of the β-strand surface energies, $n_t = \psi/\psi_h$, does not change qualitatively the results but only modifies the quantitative details of the curves shown in the previous pages. A change in the ratio n_t implies that a new sequence of optimal shapes has to be determined. According to Eqs. (3.1), (3.10) and (3.14), in particular, increasing the ψ/ψ_h ratio has the effect of increasing the barrier w_n/ψ_h, the nucleus size, n^* and the nucleation work, w^*. Then, as expected, the nucleation rate J will decrease upon increasing the ψ/ψ_h ratio, as apparent from Eq. (3.17).

The major result of the present study is that one should always use CCNT rather than CNT for a quantitatively reliable analytical description of n^*, w^* and J when formation of amyloid fibrils is nucleation-mediated. An even better description of these quantities is provided by ANT, but this theory cannot be solved analitically, since the ANT work w_n to form a fibril does not admit a simple analytical expression. The practically exact $J(s)$ dependence is obtainable numerically from the BDANT Eq. (3.17).

Chapter 4

Amyloid fibrillation kinetics: Insight from atomistic nucleation theory[1]

In this chapter the nucleation of β_2-microglobulin ($\beta_2 m$) and amyloid β_{40} (Aβ_{40}) nanosized fibrils composed of successively layered β-sheets will be discussed at molecular level. As in the previous chapters, the nucleation process considered here takes place by direct polymerisation of protein segments (β-strands) into β-sheets. In particular, employing ANT to amyloid nucleation of β_2m and Aβ_{40} fibrils, predictions of the fibril nucleus size and of the fibril nucleation rate will be determined as functions of the supersaturation and concentration of the protein solution. Then the ANT results will be correlated to recent time-resolved optical experiments, where the effects of the protein concentration and mutations on the initial lag time in the protein solution have been measured. The present analysis reveals the general principles underlying the nucleation kinetics of nanosized amyloid fibrils and indicates that the protein fibrillation can be treated in the framework of existing general theories of the nucleation of new phases.

4.1 Introduction

The assembly of proteins into amyloid fibrils is a widespread and much-studied phenomenon, because it has wide implications ranging from biotechnology to human disease (22; 43). Yet, the nucleation of such nanofibrils is poorly understood.

[1]This chapter is based on Ref. (17).

The reason is that experiments characterising the early stages leading to the formation of these aggregates are notoriously difficult, and the complexity of the protein systems has made it difficult to simulate the fibril nucleation events. It is now well established that fibrillar protein aggregates form through a nucleation mechanism (4; 5; 31; 32; 33; 38; 40; 41; 54; 60; 61; 67; 98; 100; 101; 109; 110; 111; 124; 128; 152; 157; 163). Their formation kinetics in fact is characterised by an initial lag time τ_l(s) during which no aggregates are detected, and by a maximal rate k_a (s^{-1}) of the overall aggregation process. Time-resolved optical experiments that measure the fluorescence signal arising from dye molecules such as thioflavin-T bound to the protein aggregates enable the determination of τ_l and k_a. Numerous experiments (24; 28; 51; 58; 62; 87; 88; 92; 112; 116; 122; 157; 164) have been performed to reveal the relationship between the physicochemical properties of the natural amino acids and the kinetics of amyloid formation. These experiments have in turn been used to substantiate phenomenological models (20; 23; 30; 138; 144; 161) able to predict changes in the propensity of proteins to aggregate upon mutation as well as to identify amino acid sequences of proteins that are likely to belong to the fibril core.

In this chapter the stated problem is approached from a different angle considering that this specificity might be a particular expression of a common fibril nucleation mechanism which could be treated in the framework of existing general theories of nucleation of new phases (e.g., Ref. (77)). Progress to support this view was made in recent works (6; 18; 81), where concepts from the theory of overall crystallisation to describe the kinetics of overall protein aggregation have been used, and classical and atomistic nucleation theories (CNT and ANT, respectively) have been applied to describe the nucleation of amyloid fibrils. In this study the objective is to illustrate the application of ANT to analyse and possibly predict the results of experiments on the fibrillation kinetics of β_2m and Aβ_{40} with some point mutations.

β_2-microglobulin is a 100 amino-acid long protein which can self-assemble into amyloid fibers. It has been shown that the β_2m fibers are associated with numerous hemodialysis human diseases (22; 122). Amyloid β is a 36-42 aminoacid-long peptide formed by cleavage of Amyloid Precursor Protein (APP) and it is the major component of amyloid plaques in the brains of patients affected by Alzheimer's disease (22; 24). The most common peptide found in the amyloid fiber core are Aβ_{40} and Aβ_{42}. Because of health and medical implications there

are a wide range of experimental and theoretical studies related to β_2m and Aβ_{40} peptides (53).

4.2 Modelling the amyloid fibril

In this chapter the fibril model discussed in chapter 3 (Fig. 3.1) is applied to real proteins. The fibril model proposed is based on reported structural and morphological studies of amyloid fibrils and microcrystals (65; 103; 124; 152). It is worth reminding that due to their strong hydrogen bonds, the virtually fully extended β-strands of proteins can arrange along the fibril lengthening axis (the m axis in Fig. 4.1) into β-sheets. The much weaker hydrophobicity-mediated bonds cause the β-strands to arrange along the fibril thickening axis (the i axis in Fig. 4.1) in such a way that a nanosized amyloid fibril composed of successively layered β-sheets can form. Since the fibril width is fixed and equal to the β-strand length, the fibril can be considered as a 2D aggregate in the (m, i) plane, with building blocks (the β-strands) arranged in a 2D lattice of simple rectangular symmetry (Fig. 4.1).

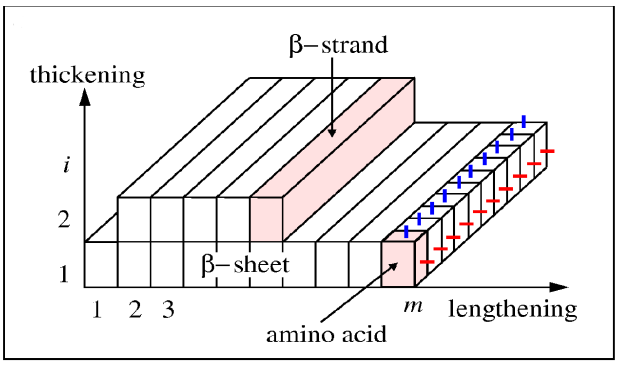

Figure 4.1: Schematic picture of $n = 15$-sized fibril with thickness of $i = 2$ β-sheets of length $m = 10$ β-strands in the first β-sheet and $m = 5$ β-strands in the second β-sheet. Each β-strand is composed of 10 amino acids. The blue and red lines shown for the rightmost β-strand represent the broken bonds of each amino acid in the direction of the i and m axes, respectively. The sum of the broken bond amino acids give the interstrand broken bond energies $E/2$ (Joule) and $E_h/2$ (Joule) according to Eqs. (4.1) and (4.2). Here the total number of $E/2$ and $E_h/2$ sides, parallel to the fibril i and m axis, are given by $l = 4$ and $l_h = 20$, respectively.

To apply the model to a real protein it is necessary to identify the amino acids that represent the β-strand in the fibril model considered. Once the most aggregation-prone regions in the protein have been identified, the binding energies E and E_h need to be calculated.

4.3 Sequence-specific bonding energies

In the atomistic approach to nucleation of amyloid fibrils (17; 18; 80), important parameters are the binding energies of the β-strands within the fibril. As stated in the previous chapter E and E_h are the binding energies between two nearest-neighbour β-strands along the m and i axes in Fig. 4.1, respectively. The excess energy of any n-sized nanofibril equals the fibril total surface energy (81) which, according to ANT (18; 80), is given by the total energy $l_n E/2 + l_{h,n} E_h/2$ of the nearest-neighbour broken-bonds at the periphery of the fibril cross section in the (m, i) plane. In the above sum, n is the number of β-strands in the fibril, and l_n or $l_{h,n}$ is the total number of nearest-neighbour broken bonds that are parallel to the fibril m axis or i axis, respectively.

In order to calculate E and E_h it is assumed that each amino acid in a β-strand (ten such amino acids are illustrated in Fig. 4.1 by the prisms in the rightmost β-strand of the shorter β-sheet) forms hydrogen bonds solely along the m axis, and that hydrophobicity-mediated bonds can be formed along both the m and i axes. Assuming further that bonds can only be formed between nearest-neighbour amino acids, the interstrand binding energies E and E_h are given by summation of the bond energies of such amino acids over all amino acid pairs in neighbouring β-strand:

$$E = \sum_p [\varepsilon_{jq}(p) + \varepsilon_{h,jq}(p)] \tag{4.1}$$

and

$$E_h = \sum_p \varepsilon_{h,jq}(p). \tag{4.2}$$

In Eq. (4.1), the summation goes over all amino acid pairs p in two nearest-neighbour β-strands of a β-sheet, and in Eq. (4.2) over all amino acid pairs p in two nearest-neighbour β-strands of two successive β-sheets. Here the energies of a nearest-neighbour hydrophobicity-mediated amino acid bond and a nearest-neighbour hydrogen bond between two amino acids j and q are denoted by $\varepsilon_{h,jq}$ and ε_{jq}, respectively ($\varepsilon_{jq} >> \varepsilon_{h,jq}$, because the hydrogen bond is much stronger than the hydrophobicity-mediated interaction). The assignment of sequence-specific binding energies to ε_{jq} and $\varepsilon_{h,jq}$ allows the application of the fibril model to specific protein systems. In this study, the sequence-specific hydrogen-bond

energies used are based on a statistical analysis (144) of the frequency that two residue types are found paired in neighbouring strands within β-sheets in globular proteins. The pairing energy depends on the orientation (parallel or antiparallel) of the β-strands. In the model $\varepsilon_{jq}^{\mathrm{P}}$ and $\varepsilon_{jq}^{\mathrm{A}}$ are the dimensionless pairing energies for the parallel and antiparallel orientation, respectively. In particular their values are taken from Table 1 of Ref.(144) to modulate the basic unit ε (Joule) of the hydrogen bond energy in such a way that the values of the binding energy ε_{jq} lie in the interval $[0, 2\varepsilon]$:

$$\varepsilon_{jq} = \frac{2\left(\varepsilon_{\mathrm{PRO,ASP}}^{\mathrm{A}} - \varepsilon_{jq}^{\mathrm{P(A)}}\right)}{\varepsilon_{\mathrm{PRO,ASP}}^{\mathrm{A}} - \varepsilon_{\mathrm{CYS,CYS}}^{\mathrm{A}}}\varepsilon. \tag{4.3}$$

Here, $\varepsilon_{\mathrm{CYS,CYS}}^{\mathrm{A}}$ = -2.57 and $\varepsilon_{\mathrm{PRO,ASP}}^{\mathrm{A}}$ = 3.55 are, respectively, the dimensionless bond energies of a CYS-CYS pair and a PRO-ASP pair in an antiparallel β-sheet, and $\varepsilon_{jq}^{\mathrm{P(A)}}$ equals $\varepsilon_{jq}^{\mathrm{P}}$ or $\varepsilon_{jq}^{\mathrm{A}}$ for the parallel or antiparallel orientation. For the basic unit of the hydrogen bond $\varepsilon = 6.95 \times 10^{-21}$ J (1 kcal/mol) has been used, because this value is in the range of the hydrogen-bond energies measured experimentally (35). For simplicity, the energy of the hydrophobicity-mediated bond is assumed to be the same for all amino acid pairs, so that $\varepsilon_{h,jq} = \varepsilon_h$, and further $\varepsilon_h = \varepsilon/10$ has been set.

4.4 Nanofibril evolution mechanism

The determination of the nucleation rate J from Eq. (3.19) requires specification of the sequence of shapes that the fibril takes during its evolution from the monomer size $n = 1$ to the considered size $n > 1$. In reality, there are many such sequences, as explained in the previous chapter, the sequence of shapes which has the CNT equilibrium shape as a reference has been chosen. This is because the latter corresponds to the minimal total surface energy consistent with the Szilard-Farkas mechanism (81). The same evolution mechanism explained in the previous chapter will be adapted to describe nanofibril growth mechanism for two real proteins, i.e, β_2m and Aβ_{40}.

In the sequence of shapes shown in Fig. 4.2, the fibril initially evolves as a single β-sheet, and according to CNT (81), the transition from one-dimensional

(1D) to two-dimensional (2D) evolution occurs at a transition size $n_t = E/E_h$. This size is an important parameter, because it characterises the fibril equilibrium shape, i.e., the fibril thermodynamically favoured aspect (length/thickness) ratio. In particular the shape sequence in Fig. 4.2 represents the fibril evolution into nanofibrils from β_2m β-strands with $n_t = E/E_h = 12$. Thus, the $(n_t + 1)$-sized

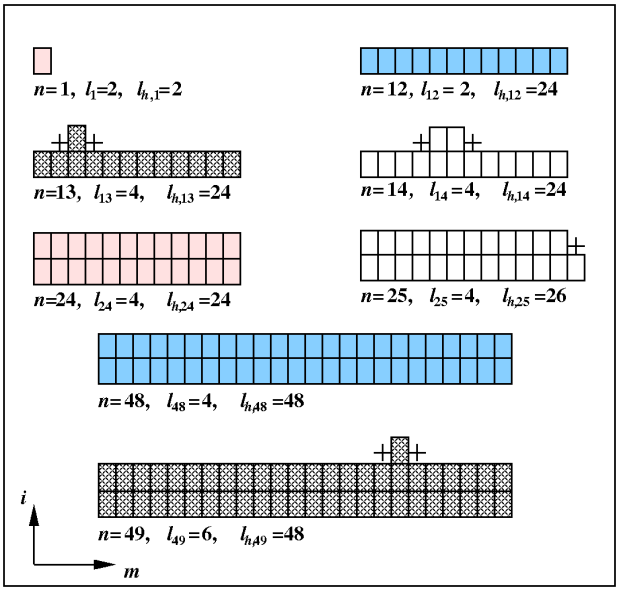

Figure 4.2: Fibril shapes with minimum total surface energy depicting the fibril evolution at $n_t = E/E_h = 12$. The shaded fibrils represent the subsequent nuclei.

fibril (the one with $n = 13$ in Fig. 4.2) has the shape of 1β-sheet of length n_t with one protein monomer (β-strand) adsorbed on the sheet surface. This monomer gives birth to the fibril second β-sheet and creates two kink sites at the fibril surface (marked by plus signs in Fig. 4.2). As shown in Ref. (18), in a certain supersaturation range the $(n_t + 1)$-sized fibril requires maximum work for its formation and is therefore the so called nucleus. Further monomer attachment to a kink site does not increase the fibril total surface energy, because the fibrils of size $n = n_t + 1$, $n_t + 2$,..., $2n_t$ have the same total number of broken bonds (see the fibrils with $n = 13$, 14 and 24 in Fig. 4.2). Since the CNT equilibrium shape requires fibril thickness between one β-sheets and two β-sheets from $n = 2n_t$ to $n = 2 \times 2n_t$, the attachment of the next protein monomer to the $2n_t$-sized fibril can be assumed to occur at one of the two fibril ends. This lengthwise monomer attachment creates a kink site at the end of the fibril of size $n = 2n_t + 1$ (the

fibril with $n = 25$ in Fig. 4.2)) and the next monomer is attached to this site with no surface-energy cost. Thus, further fibril growth occurs by alternating creation and annihilation of kink sites at the fibril ends until the fibril reaches the size $n = 2 \times 2n_t$ (see the fibrils with $n = 25$ and 48 in Fig. 4.2). At this size, the attachment of the next protein monomer can be assumed to occur again to one of the fibril two β-sheets (see the fibril with $n = 49$ in Fig. 4.2), because the CNT equilibrium shape requires fibril thickness between two and three β-sheets for fibril sizes from $n = 2 \times 2n_t + 1$ to $n = 3 \times 3n_t$. Further growth of the fibril can be expected to occur in the above manner. For this shape sequence (Fig. 4.2) it is now possible to determine the nucleation rate J from Eq. (3.19), simply by counting l_n and $l_{h,n}$ for each n. In Appendix A, Table A.1 shows the number of broken bonds l_n and $l_{h,n}$ at the fibril periphery in the (m, i) plane for the shape sequence illustrated in Fig. 4.2 from n =1 to $n = 300$.

4.5 Solubility

As already stated in the previous chapters, for protein solutions in which the supersaturation s is experimentally controlled by the concentration C_1 (m^{-3}) of monomer protein at a fixed temperature T, s can be calculated from the equation (81)

$$s(C_1) = \ln(C_1/C_e). \tag{4.4}$$

The integrated van't Hoff equation relates the equilibrium concentration C_e (m^{-3}) to the latent heat or enthalpy per β-strand L (Joule) of protein aggregation and is given by

$$C_e = C_r \exp(-L/kT), \tag{4.5}$$

where C_r (m^{-3}) is a reference concentration. In analogy to the Haas-Drenth lattice model of protein crystals (52), L is approximately given by half of the average binding energy $< E_b >$ of a β-strand within the bulk fibril phase (81)

$$L =< E_b > /2. \tag{4.6}$$

For the fibril model considered here, $< E_b > = 2(E + E_h)$, where E and E_h are given by Eqs. (4.1) and (4.2). If s_W and s_M are the supersaturations for wild-type and point-mutated proteins respectively, employing Eqs. (4.4) and

(4.5), allows to write the change $\Delta s \equiv s_M - s_W$ in the supersaturation upon a point mutation of the wild-type protein, at the same β-strand concentration C_1 and temperature T, in the form

$$\Delta s \equiv (L_M - L_W)/k_B T. \tag{4.7}$$

The latent heat L_W or L_M of the wild-type protein or the mutant can be calculated from Eq. (4.6) when the corresponding average binding energy $< E_b >$ is known. In the same manner, if $C_{e,W}$ and $C_{e,M}$ are the solubilities of wild-type and point-mutated proteins, respectively, the change in the fibril solubility at the same temperature can be characterised by the ratio

$$C_{e,M}/C_{e,W} = \exp[(L_W - L_M)]/k_B T. \tag{4.8}$$

This formula follows from Eq. (4.5) provided the point mutation has practically no effect on C_r.

4.6 Prediction of experimental results

β_2m and its mutants

At first, it is considered the amyloid fibrillation of β_2m, for which the concentration dependence and the effect of mutations on the overall aggregation has been investigated experimentally (118; 122; 157). In order to apply the present fibril model to describe a β_2m fibril, it is necessary to identify the amino acids of β_2m that represent the β-strand in the fibril model. This can be done unambiguously only when the amyloid fibril structure is known at atomistic resolution. In absence of such detailed structural knowledge of β_2m fibrils, experimental data (118; 122) and theoretical models (30; 144) that identify the most aggregation-prone regions in proteins have been used. Both methods suggest that for β_2m the amino acid sequence 61 to 70 (SFYLLYYTEF) is the most aggregation-prone region (where S = Serine, F = Phenylalanine, Y = Tyrosine, L = Leucine, T = Threonine, E = Glutamic Acid). Thus 10 amino acids form a sequence that is defined to be the β-strand in the fibril model (Fig. 4.1). Furthermore, it is assumed that consecutive β-strands within the same β-sheet are arranged anti-parallel as suggested theoretically (144).

With the set of parameter values given in the previous sections, the inter-strand binding energies E and E_h for the β-strand (SFYLLYYTEF) within antiparallel β_2m fibrils can be calculated from Eqs. (4.1) to (4.3) giving $E = 11.9\ \varepsilon$ and $E_h = 1.0\ \varepsilon$.

The determination of the nucleation rate J from Eqs. (3.17) or (3.19) requires specification of the sequence of shapes that the fibril takes during its evolution from the monomer size $n = 1$ to the considered size $n > 1$. As previously explained in section 4.4, the sequence of shapes (Fig. 4.2) which has the CNT equilibrium shape as a reference will be assumed. By using the numbers of broken bonds l_n and $l_{h,n}$ in Table A.1 of Appendix 1 for the β_2m shape sequences, the J dependence has been determined according to Eqs. (3.17) or (3.19). The broken bond numbers at the fibril periphery in the (m, i) plane for the β_2m shape sequence are partially illustrated in Fig. 4.2.

First the plots as a function of the supersaturation s will be described. The $J(s)$ dependence from Eq. (3.17) in Fig. 4.3 has been plotted using $T = 300$ K, $M = 300$, $C_e = 0.2\ \mu$M (obtained from the comparison with experimental data described below), and $f_{1,e} = 10^{-4} s^{-1}$ inferred from measured elongation

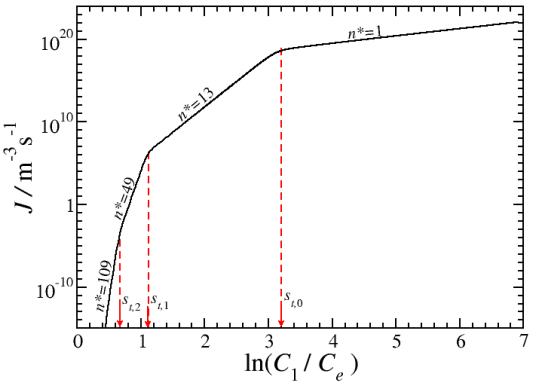

Figure 4.3: Dependence of the fibril nucleation rate J on the supersaturation $s = \ln(C_1/C_e)$ for wild-type β_2m according to Eq. (3.17). Here C_1 is the concentration of monomer protein and C_e is the β_2m fibril solubility. The red arrows in the graph point to the transition supersaturations (from Eq. (3.11)) corresponding at which the J curve bends, decreasing its slope and thus the fibril nucleus size according to Eq. 2.31.

rates (93). The value for M was chosen because larger values had no effect on the J values calculated at all the supersaturations that have been studied. The solid line in Fig. 4.3 illustrates that the dependence exhibits practically linear portions followed by sharp bends that allow the $J(s)$ curve to change slope. In the graph the fibril nuclei size n^* have been highlighted for each supersaturation region and the transition supersaturations $s_{t,0}$, $s_{t,1}$ and $s_{t,2}$, which mark the start and the end of the linear portions, are indicated by red arrows.

The isothermal $J(s)$ dependence, has been used in Eq. (2.32) for the determination of $n^*(s)$. The n^* dependence as a function of the supersaturation s is depicted by the solid line in Fig. 4.4. From the picture it can be seen that within each linear portion the nucleus size is a constant integer which undergoes large changes in the corresponding bending region of $J(s)$. The amyloid fibril formation for β_2m protein is nucleation-mediated within the ANT supersaturation range ($0.37 < s < 3.30$) according to the formula Eq. (3.11). This supersaturation range corresponds to the concentration range $0.29\ \mu\text{M} < C_1 < 4.9\ \mu\text{M}$. Furthermore, the curve in Fig. 4.4 has been calculated using Eq. (2.31) that gives a theory-independent estimation of the nucleus size n^*. Thus, considering the bending parts of the curve, it can be said that the nucleation-mediated region

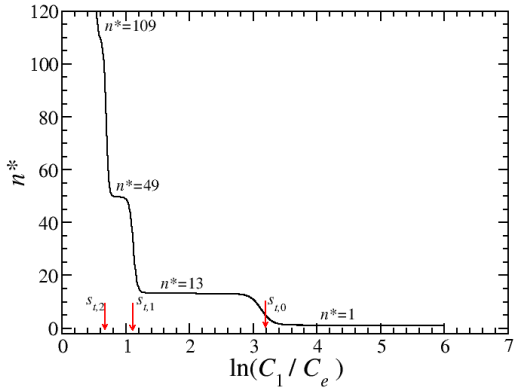

Figure 4.4: Supersaturation dependence of the fibril nucleus size n^* obtained from Eq. (2.31) with the help of the data for J according to 3.17 for wild-type β2m. The red arrows point to the supersaturations value at which the fibril nucleus size jumps to a practically constant portion in a given supersaturation range.

is between $C_1 = 0.2$ μM and $C_1 = 10$ μM. In the nucleation-mediated region the nucleation rate J and n^* depend strongly on the concentration and supersaturation. In Figs. 4.3 and 4.5 at exemplifying protein concentration $C_1 = 1.1$ μM or $s = 1.7$, the nucleation rate is $J = 10^{10}$ m^{-3}s^{-1}, and the fibril nucleus size is $n^* = 13$. In a solution volume of 100 ml typically used in experiments, this nucleation rate corresponds to the formation of 3.6×10^9 supernucleus fibrils in a time period of 1 hour. Diluting the protein solution to $C_1 = 0.45$ μM, which correspond to $s = 0.81$, reduces the nucleation rate by ten orders of magnitude to $J = 1$ m^{-3}s^{-1} and causes the nucleus size to increase to $n^* = 49$. In the same solution volume this nucleation rate corresponds to the formation of 3.6×10^{-1} supernucleus fibrils in the same time period of 1 hour, or in other words to the creation of one supernucleus fibril every 3 hours. This strong concentration or supersaturation dependence of the nucleation rate predicted for the present fibril model shows that the fibrillation kinetics can strongly depend on the solution conditions. The low nucleation rate predicted for nucleus sizes $n^* > 49$ suggests that nuclei of such sizes can hardly be observed experimentally, in agreement with experiments (157) that mainly report smaller nucleus sizes.

In order to verify the J and n^* predictions, direct measurements for the nucleation rate and nucleus size for amyloid fibril nucleation are needed. At

present, the numerous experimental studies are mainly for the lag time τ_l and the maximal rate k_a of overall aggregation of proteins. Although τ_l may be directly related to the nucleation rate, it is well known that it depends also on solution agitation and post-nucleation events such as fibril growth and fragmentation. Thus, this type of experiments does not allow a reliable determination of J and a reliable verification of the theory. Most generally, it can be expected that, when J increases, τ_l decreases, and for the present analysis the following empirical relation has been used:

$$\tau_l = a/J^x. \tag{4.9}$$

In Eq. (4.9) the constant a (m^{-3x} s^{1-x}) and the exponent x are parameters. A classical example for which the relation between τ_l and J is known comes from the kinetics of overall crystallisation when the phase transformation occurs by simultaneous nucleation and growth of many crystallites. The crystallites are treated as isomorphic and growing irreversibly without fragmentation until the phase transformation is completed. In the special case in which the crystal nucleation and growth rates are time-independent, the relation between τ_l and these rates is known exactly and takes the form of the empirical relation proposed above (77).

The experiments by Routledge et al. (122), measuring lag times τ_l for wild-type β_2m and several of its point mutations at $C_1 = 84$ μM and $T = 300$ K, could be used to test ANT for these fibrils. In order to be able to calculate the $J(C_1)$ dependence for wild-type β_2m and its mutants using Eq. (3.19), the fibril solubility C_e has to be known. The fibril solubility is different for each mutant, because the binding energy of the β-strand within the fibrils is different. Since C_e has not been determined experimentally for any of the mutants, a theoretical estimation of them has been obtained using Eqs. (4.5)-(4.8). This enables the calculation of C_e from the average binding energy of a β-strand within a fibril, yielding $C_e = 0.65$, 0.46, 0.31, 0.31, 0.46, 0.46 and 0.57 μM, for point mutations F62A, Y63A, L64A, L65A, Y66A, Y67A and F70A respectively. The acronym F62A indicates the point mutation obtained replacing the residue F at the 62^{nd} position of the wild-type β_2m with the aminoacid A (A = Alanine). All the other point mutations are named according to the same convention. Following the procedure outlined for wild-type β_2m, the $J(C_1)$ dependence according to Eq. (3.19) for all mutants have been calculated and they are shown in Fig. 4.5.

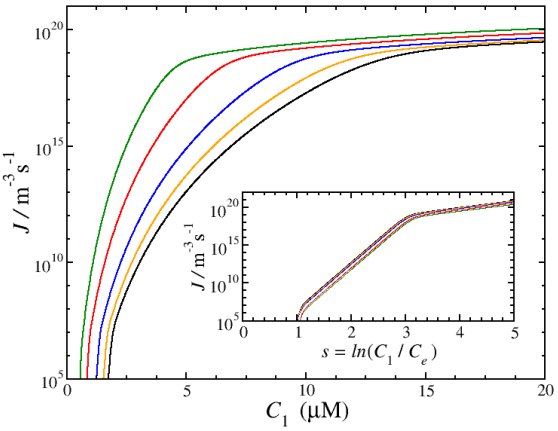

Figure 4.5: Predicted dependence of the nucleation rate J on the concentration C_1 of monomer protein obtained from Eq. (3.19) for wild-type β_2m and a number of its mutants: green line WT; black line F62A; blue line Y63A, Y66A, Y67A; red line L64A, L65A; orange line F70A. The corresponding dependence of J on the supersaturation s calculated according to (3.17) is shown in the inset.

In agreement with Routledge's experiments, all mutations reduce the nucleation rate (or increase the lag time). Furthermore, ANT predicts that the nucleus size for the mutated fibrils is larger than the one calculated for the wild-type. Some of the mutations, such as Y63A, Y66A and Y67A, lead to β-strands that have the same kind of amino acids. Although the specific residue sequence of these β-strands is different, the binding energies of the β-strands within the fibril are identical for parallel β-sheets, and consequently the fibril nucleation rates are the same. In the present case, β-strands within β_2m fibrils are arranged antiparallel, this does not necessarily need to be the case, but for the case above, it does happen. This fact is not a limitation of ANT, but is due to the assumption of the present model that the interaction between amino acids is restricted to the nearest neighbours. As a consequence, the present model does not capture the experimental observation (49) that a specific residue sequence within a β-strand determines fibrillation, rather than amino acid identity alone.

A closer inspection of Fig. 4.5 shows that ANT predicts the correct hierarchy of the effect of a mutation on the nucleation rate, i.e. that the nucleation rate for the wild-type protein is higher than that of the mutant L64A, which is higher than that of mutants Y66A and Y67A, which is higher than that for mutant F70A, which is higher than that of F62A. Not predicted correctly is the effect of

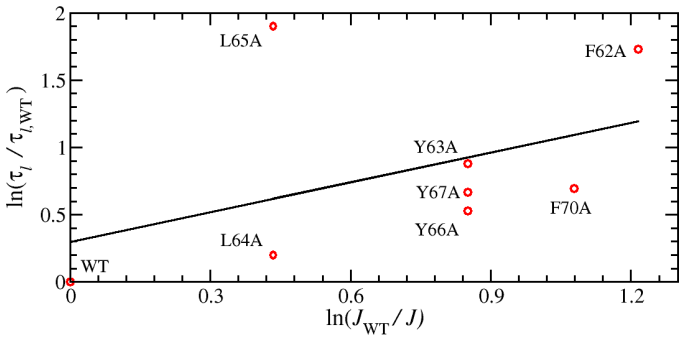

Figure 4.6: Correlation (red circles) between experimentally measured lag times τ_l and predicted nucleation rates $J(C_1)$ obtained from Eq. (3.19) for wild-type β_2m (WT) and several of its mutants, as shown by the labels, at $C_1 = 84$ μM. The linear best fit to the data points (solid line) yields a correlation coefficient of 0.43. The slope of the solid line is the exponent in Eq. (4.9).

the mutations L65A and Y63A. The complete correlation analysis between the experimentally measured lag times τ_l and the nucleation rates $J(C_1)$ predicted by Eq. (3.19) for β_2m and its mutants at $C_1 = 84$ μM is shown in Fig. 4.6 using the relation in Eq. (4.9). The linear fit to the data points, for wild type and mutants as labelled in the graph, yields the exponent $x = 0.74$ of Eq. (4.9) and a correlation coefficient of 0.43. The interpretation of this correlation analysis has to be taken with caution because the correlation is made for data that are obtained in the metanucleation regime. As a consequence, the observed changes upon mutation on τ_l are small (specifically a factor of about 2 in the experiments, and even less in the calculated J). Such small changes in τ_l may be entirely due to the changes in the attachment frequency $f_{1,e}$, because a mutation can lead to an enhanced or reduced ability of lengthwise attachment of β-strands to the fibril ends. For example, the mutation L65A, whose nucleation rate is not correctly predicted, is known to diminish the thermodynamic stability of the native fold of wild-type β_2m, which can enhance the ability of lengthwise attachment of β-strands to the fibril ends. Such changes in $f_{1,e}$ are not explicitly considered in the correlation analysis and a reliable experimental test of the present predicted nucleation rates require quantitative measurements of both τ_l and $f_{1,e}$.

In the insets of Figs. 4.5 and 4.7 the dependencies of J and n^* on the supersaturation $s = \ln(C_1/C_e)$ are shown, respectively. In this cases, all data obtained for wild-type and different mutants fall onto a single line. This implies

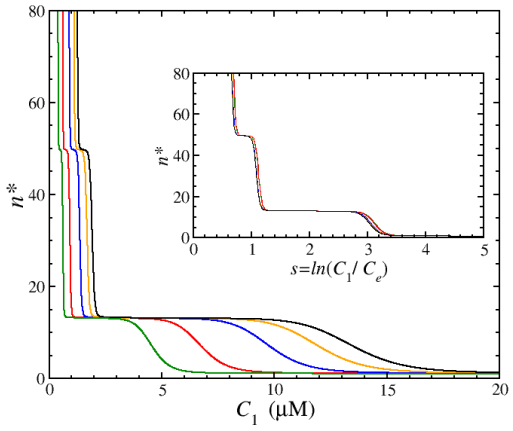

Figure 4.7: Predicted dependence of the nucleus size n^* on the concentration C_1 of monomer protein obtained from Eq. (2.31) for wild-type β_2m and a number of its mutants: green line WT; black line F62A; blue line Y63A, Y66A, Y67A; red line L64A, L65A; orange line F70A. The corresponding s dependence of n^* is shown in the inset.

that the primary effect of a mutation on the nucleation rate and nucleus size is associated to the change in the fibril solubility. It is the explicit consideration of the different solubilities C_e in Eq. (3.19) that leads to changes in $J(C_1)$ for the different mutants at fixed concentration and temperature. Finally, an important general result of this analysis is that a mutation that decreases the thermodynamic stability of a fibril decreases the fibril nucleation rate. For example, since the interstrand binding energy E is reduced from 11.9 ε for the wild-type to 11.2 ε for the point mutation F62A, at $C_1 = 1.5$ μM thereby the fibril nucleation rate J is lowered from 7.9×10^{11} for the wild-type to 6.3×10^{1} m^{-3} s^{-1} for the point mutation F62A.

The exponent $x = 0.74$ in Eq. (4.9) is determined from the correlation analysis between experimentally measured lag times τ_l and the predicted $J(C_1)$ from Eq. (3.19) obtained for wild-type β_2m and several of its mutants as described previously. The correlation analysis allowed to convert the experimentally measured isothermal $\tau_l(C_1)$ dependence by Xue et al.(157) at $T = 300$ K into the corresponding $J(C_1)$ dependence and compare it to the one calculated from Eq. (3.19). The results are shown in Fig. (4.8) by red circles and a best-fit analysis between the experimentally measured and theoretically predicted nucleation rates yields estimates for the constant $a = 10^{17}$ m$^{-2.22}$ s$^{-0.26}$ and the fibril sol-

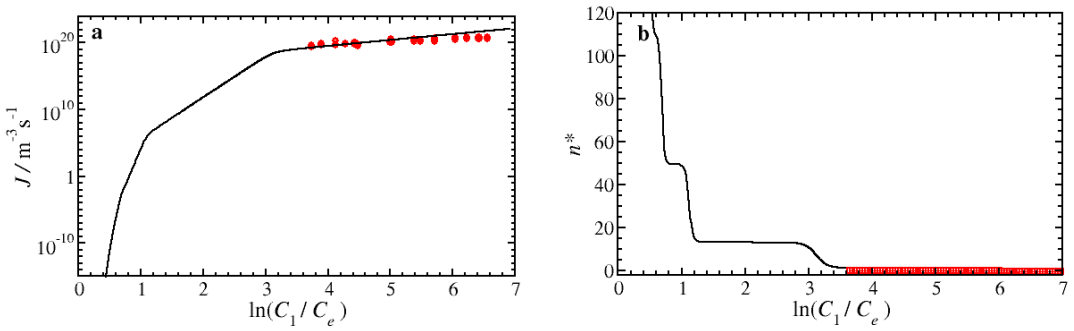

Figure 4.8: (a) Dependence of the fibril nucleation rate J on the supersaturation $s = \ln(C_1/C_e)$ for β_2m (black continuous line) according to Eq. (3.17), the red circles are obtained from experimental data in Ref. (157) in combination with Eqs. (3.17) and (4.9). Here C_1 is the concentration of monomer protein and C_e is the β_2m fibril solubility. (b) Respective supersaturation dependence of the fibril nucleus size n^* obtained from Eq. (2.32) with the help of the data for J shown in (a) for β_2m in black line, the red circles are obtained from experimental data in combination with Eq. (2.32) applied to the data in (a).

ubility $C_e = 0.2$ μM as used in the calculation of $J(C_1)$ from Eq. (3.19). The comparison shows that ANT predicts correctly the increase of J with increasing C_1. The weak $J(C_1)$ dependence obtained from the experiments indicates, however, that they have been performed in the metanucleation regime in which the nucleus is the single β-strand. Indeed, the corresponding nucleus size obtained by using the experimental $J(C_1)$ data in Eq. (2.31) yields $n^* = 1$ over the whole concentration range investigated experimentally. The determination of the transition concentration between the nucleation and the metanucleation regimes requires the determination of the narrow concentration range in which τ_l increases dramatically. Knowledge of this transition concentration would allow a more reliable ANT estimate of the fibril solubility.

AB40 and mutants

As a second application of ANT the fibrillation of Aβ_{40} has been considered. The molecular structure of the Aβ_{40} fibrils is known (117; 123) and the effect of point mutations of the wild-type residue 18 (Valine) on the overall protein aggregation has been investigated experimentally (24). The analysis is based on the molecular structure by Sachse et al. for Aβ_{40} fibrils, in which the β-strand is formed by a long amino acid stretch at the C-terminal tail of the protein monomer (123). As in the case for β_2m fibrils, theoretical models (30; 144) are considered to identify the most aggregation-prone amino acids in the Aβ_{40} protein. The theoretical algorithm by Trovato et al. (144) predicts that the amino acids 12-20 (VHHQKLVFF) and 31-40 (IIGLMVGGVV) are the most aggregation-prone (V = Valine, H = Histidine, Q = Glutamine, K = Lysine, L = Leucine, F = Phenylalanine, I = Isoleucine, G = Glycine, M = Methionine). Joined together, these 19 amino acids form a sequence that defines the β-strand in the present fibril model (Fig. 4.1). The interstrand binding energies E and E_h for the β-strand (VHHQKLVFFIIGLMVGGVV) arranged parallel within wild-type Aβ_{40} fibrils can be calculated from Eqs. (4.1) and (4.2) and are $E = 25.4\ \varepsilon$ and $E_h = 1.9\ \varepsilon$. In the same manner as for β_2m, $J(s)$, $J(C_1)$ and n^* dependencies have been calculated from Eqs. (3.17), (3.19) and (2.31).

For Aβ_{40} however, the depicted fibril evolution needs to be adapted to the case $n_t = E/E_h = 14$. Fig. 4.9 shows the evolution mechanism for Aβ_{40}. For the modified sequence of fibril shapes, the dependence of J and n^* on s for the wild-type Aβ_{40} have been determined with $M = 300$, $C_e = 1\ \mu$M (suggested by few experiments i.e. (115; 155)) and exemplifying $f_{1,e} = 10^{-4}\ \mathrm{s}^{-1}$ (inferred from measured elongation rates (93)).

The $J(s)$ and $n^*(s)$ dependencies are shown as green dashed lines in Fig. 4.10a and 4.10b, respectively. The $J(s)$ dependence exhibits the practically linear portions (Fig. 4.10a) in which the corresponding nucleus size is a constant integer. Due to the larger transition size n_t, the linear portions for Aβ_{40} are more extended than those obtained for β_2m whose corresponding $J(s)$ is shown in black for comparison in Figs. 4.10a and 4.10b. In the nucleation regime, the nucleation rate for Aβ_{40} fibrils is substantially lower than that of β_2m, whereas the fibril nucleus size is always larger (Fig. 4.10b). Interestingly, in the metanucleation regime the nucleation rate predicted for Aβ_{40} fibrils is higher than that

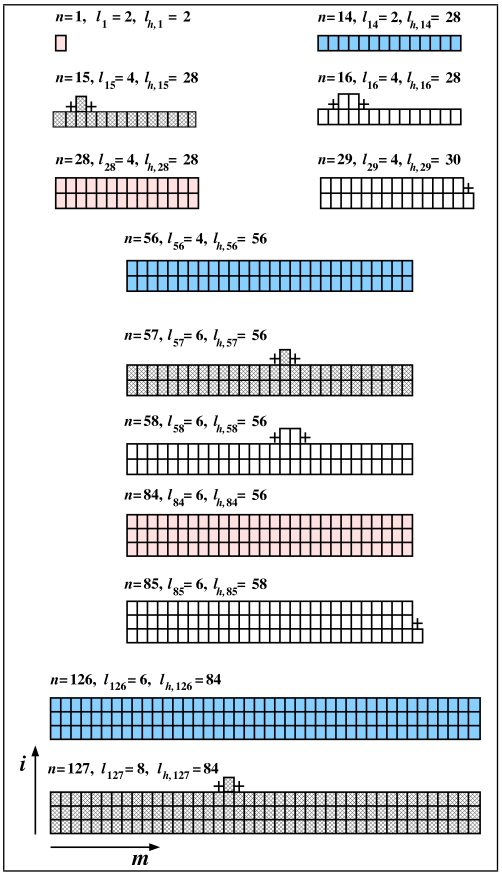

Figure 4.9: Sequence of nanofibril shapes corresponding to minimum total surface energy, which depicts the fibril evolution when the transition from 1D to 2D aggregate is at size $n_t = E/E_h = 14$ as for Aβ_{40}.

of β_2m. This observation is important as most experiments on the fibrillation kinetics of proteins are performed in the metanucleation regime, and a comparison of kinetic data obtained for different proteins at the same solution conditions can lead to the opposite results if the experiments are performed in the nucleation regime.

Finally, the $J(C_1)$ and $n^*(C_1)$ dependencies have been calculated for point mutations of residue 18 (V) with amino acids Alanine (A), Aspartic acid (D), Glutamic acid (E), Phenylalanine (F), Glycine (G), Histidine (H), Isoleucine (I), Lysine (K), Leucine (L), Asparagine (N), Proline (P), Glutamine (Q), Arginine (R), Serine (S), Threonine (T), Tryptophan (W) and Tyrosine (Y). As described for β_2m, the fibril solubilities have been calculated theoretically (see section 4.4) to give for the mutations mentioned above in the same order C_e = 4.3, 7.87,

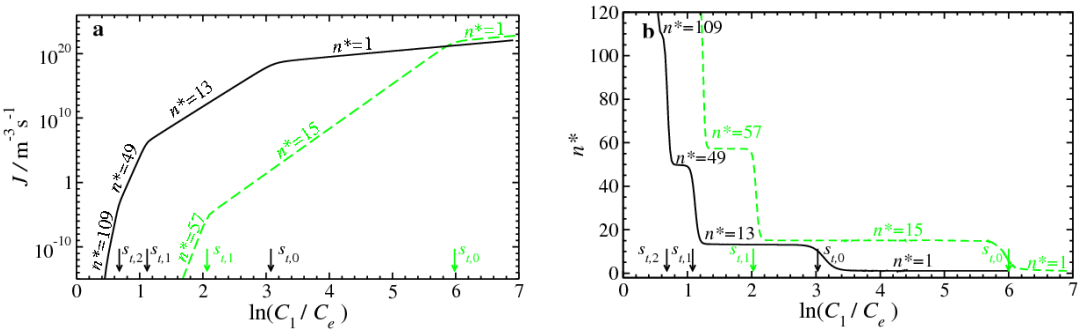

Figure 4.10: (a) Dependence of the fibril nucleation rate J on the supersaturation $s = \ln(C_1/C_e)$ for β_2m (black line) and Aβ_{40} (green line) according to Eq. (3.17). Here C_1 is the concentration of monomer protein and C_e is the β_2m and Aβ_{40} fibril solubility. The black and green arrows in the graph point to the transition supersaturations (obtained using Eqs. (3.11)) at which the J curves bend decreasing its slope. (b) Respective supersaturation dependence of the fibril nucleus size n^* obtained from Eq. (2.32) with the help of the data for J according to 3.17 for both β_2m in black line and Aβ_{40} in dashed green line. The black (β_2m) and green (Aβ_{40}) arrows point to the supersaturations value at which the fibril nucleus size jumps to a practically constant value in a given supersaturation range.

7.78, 1.51, 7.33, 2.26, 1.03, 5.95, 1.99, 2.8, 15.7, 5.66, 7.01, 3.61, 3.63, 2.16, and 2.76 μM, respectively. In agreement with experiments performed at $C_1 = 120$ μM and $T = 300$ K (24), ANT predicts that the mutations with residues A, D, E, F, G, H, K, L, N, P, Q, R, S, T lead to a lower nucleation rate compared to that of the wild-type protein, whereas mutations with residues W, Y and I lead to a higher nucleation rate (see Figs. 4.13 and 4.11). From Fig. 4.11 it can be seen that ANT predicts the correct hierarchy of the effect of a mutation on the nucleation rate, i.e. that the nucleation rate for the wild-type Aβ_{40} is higher than that of the mutant N, which is higher than that of the mutant R, which is higher than that of the mutant P. As stated in Ref. (24), in the experiments it was not possible to observe the formation of fibrils for the mutation with residue P, which is consistent with the present prediction that this mutation has the lowest nucleation rate. In the inset of Fig. 4.11 a plot of the corresponding $J(s)$ dependences is shown. As for the β_2m case, all lines fall on top of each other illustrating that the primary effect of a mutation on the nucleation rate is caused by the change of the fibril solubility. The corresponding $n^*(C_1)$ dependence for Aβ_{40} and its four mutants is depicted in Fig. 4.12 by a continuous curve made of

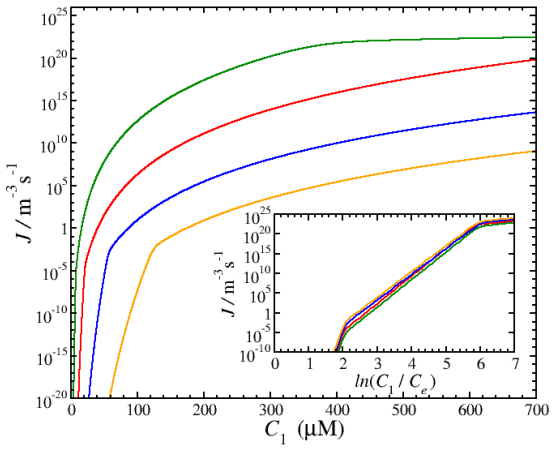

Figure 4.11: Predicted dependence of the nucleation rate J on the concentration C_1 of monomer protein obtained from Eq. (3.19) for wild-type Aβ_{40} and three of its mutants: green line WT; orange line V18P; blue line V18R; red line V18N. The corresponding dependence of J on the supersaturation $s = \ln(C_1/C_e)$ is shown in the inset.

linear portions corresponding to constant integer value for the nucleus size and of bending regions where the nucleus size undergoes large changes. In Fig. 4.12

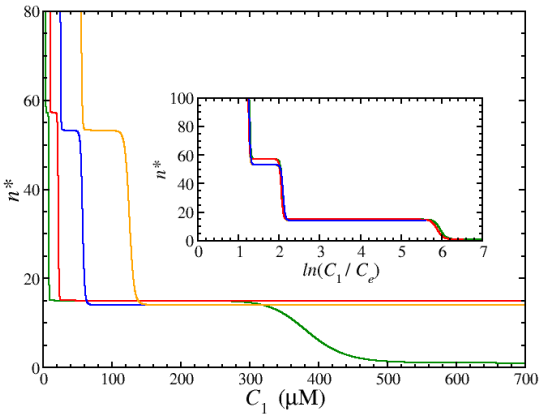

Figure 4.12: Predicted dependence of the nucleus size n^* on the concentration C_1 of monomer protein obtained from Eq. (2.31) for wild-type Aβ_{40} and three of its mutants: green line WT; orange line V18P; blue line V18R; red line V18N. The corresponding dependence of n^* on the supersaturation $s = \ln(C_1/C_e)$ is shown in the inset.

the nucleus size for the P mutant assumes noticeably larger values with respect

to the other mutants and the original protein. In particular, it is worth noting that at the same experimental concentration $C_1 = 120$ μM, the predicted ANT fibril nucleus size is $n^* = 15$ for the wild type protein and the mutations N and R, while for the mutant P the fibril nucleus size is $n^* > 15$ (see Fig. 4.12). This means that the fibril formation is thus nucleation-mediated for all the four cases. Finally, in the inset of Fig. 4.12, the plot of the $n^*(s)$ dependences for wild-type and mutants confirm that the primary effect of a mutation on the nucleus size is given by the change of the fibril solubility, since all the lines in Fig. 4.12 fall on top of each other.

The complete correlation analysis between the experimentally measured lag times τ_l and the nucleation rates J predicted by Eq. (3.19) is shown in Fig. 4.13. A linear fit to the data points shown in Fig. 4.13 yields the exponent $x = 0.05$ and a good correlation coefficient of 0.71. Note that in the correlation

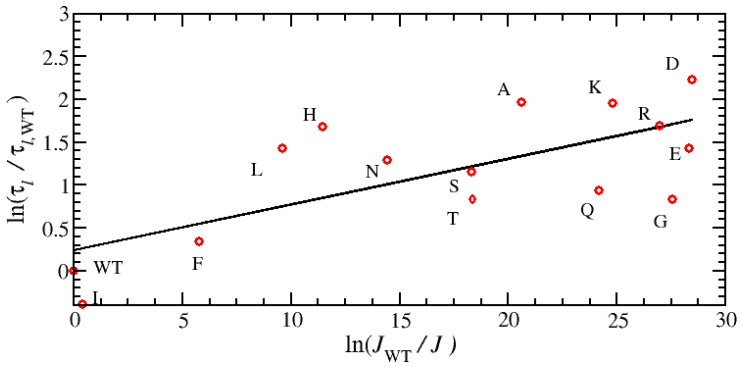

Figure 4.13: Correlation (red circles) between experimentally measured (24) lag times τ_l and the predicted nucleation rates $J(C_1)$ obtained from Eq. (3.19) for wild-type Aβ_{40} and several of its mutants, as shown by the labels, at $C_1 = 120$ μM. The linear best fit to the data points (solid line) yields a correlation coefficient of 0.71. The slope $x = 0.05$ of the solid line is the exponent in Eq. (4.9).

analysis the mutations with residues Y and W, for which ANT predicts higher nucleation rates, are not included. This is because the slope x obtained including these three residues is negative, and a negative value for x leads to nucleation rates that increase with the corresponding lag times invalidating the previous fundamental assumption.

4.7 Conclusion

In this chapter, the ANT predictions of relevant parameters has been discussed for the nucleation of β_2m, Aβ_{40} and their mutants. In particular, the nucleation rate J and the nucleus size n^* as a function of the supersaturation and concentration of the solution have been calculated and correlated to experimental results. The characteristic feature of ANT, according to which the fibril nucleus size n^* is a constant integer in a given supersaturation range, has been confirmed by the results for real proteins. Furthermore, the characteristic jumps in the $n^*(s)$ dependence imply that the nucleation rate $\ln J(s)$ is a linear function of the supersaturation s in successive s ranges.

Remarkably, the analysis of the effect of mutations on the fibril nucleation rate and nucleus size reveals that the primary effect of the predicted changes in J and n^* is due to an alteration of the fibril solubility not considered so far in theoretical and experimental studies. The strength of ANT is, in fact, that it utilises the molecular interactions between the β-strands in the fibrils to predict the supersaturation and concentration dependences of the fibril nucleation rate and nucleus size.

The analysis reported in the chapter is based on a number of assumptions about the fibril structure, the fibril interactions and the fibril nucleation mechanism. Moreover, the correlation between the theoretically predicted nucleation rates and experimentally obtained lag-times is based on Eq. (4.9) which is empirical. For these reasons, the emphasis of the present analysis is not on the quantitative agreement with experimental results, but rather, it is meant to provide an illustration of how amyloid fibril nucleation can be treated in the framework of existing general theories of nucleation of new phases. To validate the application of nucleation theories to treat amyloid fibril formation, experiments that determine the nucleation rate and nucleus size are needed. In absence of such experiments, investigations on the lag time τ_l of overall protein aggregation have been considered in this chapter. However, this type of experiments does not allow a reliable determination of J, because post-nucleation processes such as fragmentation can affect the overall aggregation process. Most generally, one may only expect that τ_l decreases when J increases. Furthermore, the empirical Eq. (4.9) has no theoretical foundation in the field; its form here has been motivated by analogy to a known exact relation between J and τ_l from the kinetics

of overall crystallisation. The presented correlation analysis (Figs. 4.6 and 4.13) between experimentally measured lag times τ_l and the nucleation rates J predicted by Eq. (3.19) for two wild-type proteins and several of their mutants can be used to verify this equation, but the experiments considered have not been designed for this analysis.

Although the obtained correlation coefficients of 0.43 for β_2m and of 0.7 for Aβ_{40} can be considered as reasonably good, the correlation analysis suffers from incomplete knowledge of the important input parameters needed in Eqs. (3.17) and (3.19). In particular, the effect of the mutation on the fibril attachment rate $f_{1,e}$ and the fibril solubility C_e needs to be determined experimentally. In fact, in protein aggregation, $f_{1,e}$ may vary widely from 10^{-4} to 10^5 s^{-1} (81), and, as already noted, the primary effect of mutations on the nucleation rate is due to a change of C_e.

Another important finding of the present analysis is that the relatively weak concentration dependence of J observed experimentally for β_2m (see Fig. 4.8) indicates that the data have been obtained in the metanucleation regime, in which fibril formation is not impeded by the existence of a nucleation barrier and the nucleus is the single β-strand.

The presented theoretical framework is applicable to homogeneous nucleation of amyloid fibrils in sufficiently dilute solutions. In its present version, entropic effects such as the loss of entropy occurring when a β-strand is attached to a fibril or the entropy due to vibrations of a β-strand within the fibril are not considered. Entropic effects can be taken into account by providing expressions for the entropic contribution to the fibril solubility and β-strand binding energy, but this goes beyond the scope of the present thesis. Examples of such expressions can be found in Ref. (31) and they could be the basis of an important extension of the present nanofibril model. It should be noted that the entropic effects on the fibril solubility are automatically accounted for, when experimental data for C_e are used in the ANT Eq. (3.19) for J. Then the missing information is about the entropic contribution to the broken-bond energies, but it could be obtained from full atomistic simulations. It is worth noting as well that in order to apply ANT to two-step nucleation involving formation of oligomeric precursors or to heterogeneous nucleation on nucleation-active sites provided by foreign agents such as nanoparticles, the theory needs appropriate modification. This may be performed following an approach already used in CNT (e.g., Ref. (77)). Since

the present theoretical formalism stems from first principles in nucleation theory, it has the potential not only to become a standard tool in analysing experiments on amyloid fibril nucleation, but also to help designing new experiments aimed at better understanding of how to control the birth of amyloid fibrils.

Chapter 5

Size distribution of amyloid nanofibrils[1]

In this chapter the size distribution of amyloid nanofibrils (protofilaments) in nucleating protein solutions is considered when the nucleation process occurs by the mechanism of direct polymerization of β-strands (extended peptides or protein segments) into β-sheets. Employing atomistic nucleation theory, a general expression for the stationary size distribution of amyloid nanofibrils constituted of successively layered β-sheets has been derived. The application of this expression to amyloidβ_{40} (Aβ_{40}) fibrils allows the prediction of the nanofibril size distribution as a function of the protein concentration and temperature. The most remarkable feature of the distribution is its exhibiting a series of peaks which are positioned at "magic "nanofibril sizes (or lengths) and which are due to points of discontinuity of the work for fibril formation. This finding of magic sizes or lengths is consistent with experimental results for the size distribution of aggregates in solutions of Aβ_{40} proteins. Also, this approach makes it possible to gain insight into the effect of point mutations on the nanofibril size distribution, an effect that may play a role in experimentally observed substantial differences in the fibrillation lag-time of wild-type and point-mutated amyloid-β proteins.

[1]This chapter is based on Ref. (19).

5.1 Introduction

Recently, small-angle X-ray scattering (114; 149) and fluorescence correlation spectroscopy experiments (23; 24; 28; 30; 41; 42; 51; 58; 62; 87; 92; 112; 114; 116; 122; 138; 144; 149; 161; 163; 164) have been used to monitor the population, structural changes and size of protein aggregates in solution. Understanding and interpreting such experiments requires developing a theoretical framework able to describe amyloid fibril nucleation.

In Chapter 2 and Chapter 3, concepts from the theory of overall crystallization to describe the kinetics of protein fibrillation have been used (6); in particular classical and atomistic nucleation theories (CNT and ANT, respectively) have been applied to treat the one-step nucleation of amyloid fibrils (18; 81).

In Chapter 4, the validity of the ANT description has been demonstrated by correlating theoretical fibril nucleation rates with experimental lag times in the fibrillation kinetics of β_2-microglobulin and amyloid β_{40} (Aβ_{40}) proteins (17).

The objective of this chapter is to employ ANT for the determination of the size distribution of amyloid nanofibrils (protofilaments) in protein solutions undergoing polymerization of β-strands into β-sheets. A general expression for the stationary size distribution of amyloid nanofibrils consisting of successively layered β-sheets will be derived and subsequently this expression will be applied to the smallest Aβ_{40} protofilaments. The results are applicable to homogeneous nucleation which takes place when the protein solution is sufficiently pure and/or strongly supersaturated. In the present study, the treatment differed from that in Ref. (127), who considered the case of amyloid fibril nucleation in two steps: single β-strands assemble into disordered oligomers, which then transform into β-sheets.

5.2 Nanofibril size distribution

As known from nucleation theory (e.g., Ref. (77)), the stationary size distribution of the n-sized aggregates of a nucleating single-component phase can be represented exactly as ($n = 1, 2, 3,...$)

$$X_n = C_n \frac{\sum_{k=n}^{M-1} \left(\frac{1}{f_k C_k}\right)}{\sum_{k=1}^{M-1} \left(\frac{1}{f_k C_k}\right)}, \tag{5.1}$$

when nucleation takes place by the Szilard-Farkas mechanism, according to which the aggregates change size by random attachment and detachment solely of monomeric building units. Here n (or k) is the number of building units in an aggregate, X_n (m^{-3}) is the stationary concentration of n-sized aggregates, C_n (m^{-3}) is the corresponding equilibrium concentration of such aggregates, M is the number of building units in a large enough supernucleus aggregate, and f_n (s^{-1}) is the frequency of building-unit attachment to an n-sized aggregate. Equation (5.1) is readily applicable to nucleation of amyloid fibrils in the case when the process occurs by the mechanism of direct polymerization of β-strands into β-sheets. Then n is the number of β-strands in an n-sized amyloid nanofibril and, according to ANT, C_n and f_n are of the form (18)

$$C_n = C_1 \exp[(n-1)s - (l_n - 2)\psi - (l_{h,n} - 2)\psi_h], \tag{5.2}$$

$$f_n = \frac{1}{2} f_1 l_n, \tag{5.3}$$

when only the nearest-neighbour interactions between the β-strands are taken into account. In Eqs. (5.2) and (5.3), C_1 (m^{-3}) is the actual concentration of single β-strands in the protein solution, and f_1 (s^{-1}) is the frequency of attachment of a single β-strand to another single β-strand. Physically, the sum in the exponent of Eq. (5.2), taken with minus sign, is the ratio between the work W_n to form a fibril of size n and the thermal energy $k_B T$ (k_B is the Boltzmann constant, T is the absolute temperature). It is worth reminding that for sufficiently dilute solutions, the dimensionless supersaturation s is given by $s = \ln(C_1/C_e)$, where C_e (m^{-3}) is the equilibrium concentration of single β-strands in the solution (C_e is also known as the protein solubility). Moreover, the dimensionless energies

ψ and ψ_h per strong (mainly hydrogen) and weak (hydrophobicity-mediated) broken bond have been defined in the previous chapter by $\psi = E/(2k_BT)$ and $\psi_h = E_h/(2k_BT)$. Equations (5.1) and (5.2) show that both the respective equilibrium and the stationary concentrations C_n and X_n of amyloid nanofibrils of a given size n increase strongly with increasing the solution supersaturation s, i.e. with a rise of the actual β-strand concentration C_1 and/or a fall of the protein solubility C_e. In Eq. (5.3) the proportionality of f_n to l_n reflects the fact that the β-strands are attached predominantly to the ends of the fibril β-sheets (the fibril strong broken bonds are namely at these ends), and the divisor 2 takes into account that a single β-strand has $l_1 = 2$ strong broken bonds. Combining Eqs. (5.1)-(5.3) with the previous definitions and accounting only for the nearest-neighbour β-strand interactions, the sought stationary fibril size distribution X_n is expressed by the general ANT formula ($n = 1, 2, 3,...$) as

$$X_n = C_n \frac{\sum_{k=n}^{M-1} \left(\frac{1}{l_k}\right) \left(\frac{C_e}{C_1}\right)^k \exp(l_k\psi + l_{h,k}\psi_h)}{\sum_{k=1}^{M-1} \left(\frac{1}{l_k}\right) \left(\frac{C_e}{C_1}\right)^k \exp(l_k\psi + l_{h,k}\psi_h)}, \tag{5.4}$$

in which

$$C_n = C_e \left(\frac{C_1}{C_e}\right)^n \exp[(2 - l_n)\psi + (2 - l_{h,n})\psi_h], \tag{5.5}$$

is the respective equilibrium fibril size distribution.

Equation (5.4) is a central result of this chapter and it is highly reliable, since it is based on the exact Eq. (5.1) and since in it the ratio of the two sums is practically unaffected by the approximate character of Eq. (5.3) for f_n. Importantly, Eq. (5.5) conforms to the *law of mass action* because of the proportionality of C_n to $(C_1)^n$. Also, as a single β-strand has two strong and two weak broken bonds, Eq. (5.5) is self-consistent in the sense that at $n = 1$ it returns the identity $C_1 = C_1$. It should be borne in mind that Eq. (5.4) is applicable solely to a supersaturated protein solution (i.e. when $C_1 > C_e$), since the stationary size distribution X_n can exist in such a solution only. Then, as the solution is metastable with respect to fibril nucleation and growth, X_n replaces the equilibrium size distribution C_n which, albeit mathematically well-defined, has no physical reality because of the solution metastability. When $C_1 = C_e$ or $C_1 < C_e$, however, the solution is saturated or undersaturated, respectively, and

neither nucleation nor growth are possible, for then the solution is thermodynamically stable with respect to these processes. In such a solution, the stationary size distribution X_n is physically irrelevant and the fibril population is described by the equilibrium size distribution C_n from Eq. (5.2) or (5.5).

It should be borne in mind that the determination of the nanofibril size distribution requires specification of the sequence of shapes that the fibril takes during its evolution from the monomer size $n = 1$ to any given size $n > 1$. Once the evolution mechanism is specified l_n and $l_{h,n}$ are known and the summation in Eqs. (5.4) and (5.5) can be done. As already stated in the previous chapter, there are many such sequences, but a sequence of shapes that have the CNT fibril equilibrium shape as a reference low-energy shape has been chosen, because it corresponds to the minimal fibril total surface energy compatible with the Szilard-Farkas mechanism.

The formula in Eq. (5.4) will be applied to calculate the stationary size distribution of Aβ_{40} fibrils under given condition. The nanofibril evolution mechanism used in this chapter for Aβ_{40} fibril nucleation has been described in details in section 4.4 and illustrated in Fig. 4.9.

5.3 Application to Aβ_{40} nanofibrils

As already discussed in the previous chapter, for Aβ_{40} fibril the molecular structure proposed by Sachse et al. (123) is considered. In their model the β-strand is formed by a long amino acid stretch at the C-terminal tail of the protein monomer. The details of the 2D model that will be taken into account for the description of the Aβ_{40} amyloid fibril nucleation process have been illustrated and extensively explained in section 4.2.

As in the previous chapter, the discussion here relies on theoretical models that identify the most aggregation-prone amino acids in proteins. The theoretical algorithm by Trovato et al. (144) predicts that amino acids 12 to 20 (VHHQKLVFF) and 31 to 40 (IIGLMVGGVV) are the most aggregation-prone ones (V = Valine, H = Histidine, Q = Glutamine, K = Lysine, L = Leucine, F = Phenylalanine, I = Isoleucine, G = Glycine, M = Methionine). Joined together, these 19 amino acids form a sequence which defines the β-strand in the previously discussed nanofibril model (Fig. 4.1).

The binding energies E and E_h between two nearest-neighbour β-strands along the m and i axes in Fig. 4.1 are calculated with the help of the summation over all the amino acid pairs in two β-strands of a β-sheet, according to Eqs. (4.1) and (4.2) given in section 4.3. As discussed in the previous chapter, the sequence-specific energy ε_{jq} of the hydrogen bond between amino acids j and q can vary in the interval [0,2ε], depending on the frequency with which two residue types are found paired in neighbouring β-strands within a β-sheet in globular proteins (see Eq. (4.3)). For simplicity, it is assumed that the energy $\varepsilon_{h,jq}$ of a hydrophobicity-mediated bond is the same for all amino acid pairs, so that $\varepsilon_{h,jq} = \varepsilon_h$, and $\varepsilon_h = \varepsilon/10$ has been set. With this set of parameter values, the interstrand binding energies E and E_h for the β-strands (VHHQKLVFFIIGLMVGGVV) arranged parallel within wild-type Aβ_{40} fibrils can be calculated from Eqs. (4.1)-(4.3) and are $E = 25.4\ \varepsilon$ and $E_h = 1.9\ \varepsilon$ (see also Table 5.1).

Aβ_{40}	Wild-type	V18N	V18R	V18P
E (J)	1.76×10^{-19}	1.72×10^{-19}	1.68×10^{-19}	1.65×10^{-19}
E_h	1.32×10^{-20}	1.32×10^{-20}	1.32×10^{-20}	1.32×10^{-20}
ψ	21.3	20.8	20.3	19.9
ψ_h	1.59	1.59	1.59	1.59
n_t	14	14	13	13

Table 5.1: Parameter values related to the size distribution of nanosized fibrils of wild-type and point-mutated Aβ_{40} at $T = 300$ K; the corresponding basic hydrogen-bond energy is $\varepsilon = 6.95 \times 10^{-21}$ J.

The rounded-up ratio between the β-strand strong and weak binding energies is $E/E_h = 14$, and the fibril transition size is thus $n_t = 14$ (see Table 5.1), because the CNT formula (81) gives $n_t = E/E_h$, which characterise the fibril equilibrium shape. It should be born in mind that during nucleation the fibrils are under nonequilibrium conditions: on average, they are growing and thus they assume a kinetically determined growth shape, which may differ considerably from the equilibrium one. Computer simulation data (Fig. 2 in Ref. (163)) indicate that the fibril kinetic aspect (length/thickness) ratio, which quantifies the fibril growth shape, can be several times greater than the thermodynamic aspect ratio E/E_h, i.e., that the fibrils can grow longer than expected on the basis of thermodynamic

considerations. It should be kept in mind therefore that, in reality, the fibril transition size n_t may reach values considerably greater than the value of E/E_h assumed here in the analysis.

Size distribution: supersaturation and concentration dependence

The nanofibril size distribution for wild-type Aβ_{40} solution at $T = 300$ K and supersaturations $s/\psi_h = 0.9$, 1, 1.2, 2 and 2.5 is displayed in Fig. 5.1 by lines 0.9, 1, 1.2, 2 and 2.5, respectively.

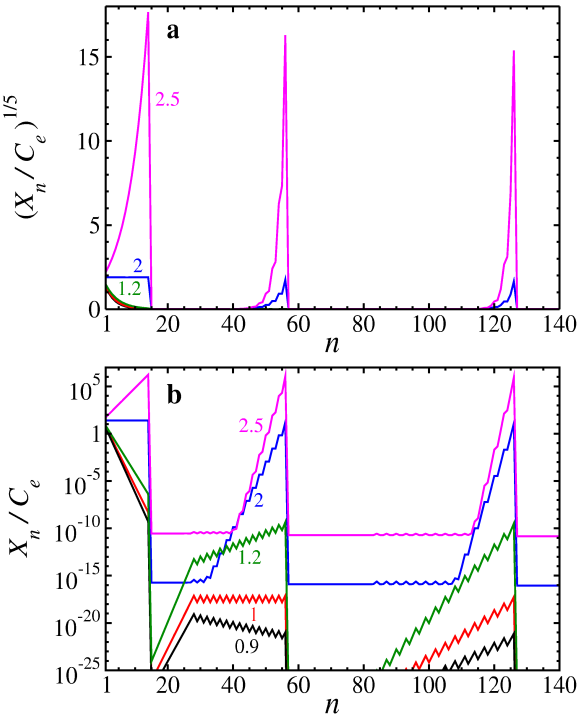

Figure 5.1: Stationary nanofibril size distribution X_n on (a) $X_n^{1/5}$ and (b) log X_n scale according to Eq. (5.4) at $T = 300$ K and supersaturations $s/\psi_h = 0.9$, 1, 1.2, 2 and 2.5 (as indicated) for wild-type Aβ_{40}.

The lines are obtained from Eq. (5.4) by setting $M = 300$ (as in Ref. (18)) and by using the l_n and $l_{h,n}$ values from Table A.2 in Appendix 1 and the ψ and ψ_h values given in Table 5.1. According to the dimensionless supersaturation $s = \ln(C_1/C_e)$, with ψ_h from Table 5.1, the above supersaturations correspond to the ratios $C_1/C_e = 4.20$, 4.90, 6.78, 24.0 and 53.3, respectively. As it can

be seen in Fig. 5.1, Eq. (5.4) predicts a series of peaks, each at a certain fixed fibril size for high enough supersaturations and at a smaller, but again fixed, size for sufficiently low supersaturations. Due to the choice of nanofibril evolution mechanism illustrated in Fig. 4.9, the first peak is positioned either at fibril size n_1' or at fibril size $n_1'' > n_1'$ which are given by $n_1' = 1$ (for $0 < s < 2\psi_h$) and by $n_1'' = n_t$ (for $s > 2\psi_h$). Here $n_t = 14$ and $2\psi_h = 3.18$ because of the ψ and ψ_h values used in the calculation of X_n in Fig. 5.1. At $s = 2\psi_h$ (line 2), the peak is plateau-like, extending from $n = n_1'$ to $n = n_1''$, i.e. from the monomer size to the fibril size n_t for transition of the fibril from 1D to 2D aggregate. Similarly, the fibril sizes n_j' and n_j'' at which the second, third, etc. peaks are positioned at low or high supersaturations, respectively, can be obtained from the relations ($j = 2, 3, 4, ...$)

$$n_j' = j(j-1)n_t \tag{5.6}$$

(for $0 < s < 2\psi_h/j$) and

$$n_j'' = j^2 n_t \tag{5.7}$$

(for $s > 2\psi_h/j$). At $s = 2\psi_h/j$, the j^{th} peak is a plateau (jagged for j = 2, 3, 4,...) that extends from $n = n_j'$ to $n = n_j''$.

The explicit relation between the evolution mechanism and the size distribution X_n calculated using it are exemplified in Fig. 5.2. The figure shows the partial sequence of shapes for Aβ_{40} (Fig. 5.2a) already fully presented in Fig. 4.9, and the size distribution X_n according to Eq. (5.4) at $s/\psi_h = 1.2$ (Fig. 5.2b). In particular, in Fig. 5.2b some fibril shapes and their distributions are illustrated. Furthermore in Fig. 5.2a, the first three nanofibrils of size n_j' and of size n_j'' are coloured pink and blue, respectively, while the fibril nuclei are hatched.

Fig. 4.9, Eqs. (5.6) and (5.7) show that both the n_j'-sized and n_j''-sized fibrils are constituted of $j\beta$-sheets, but their lengths m_j' and m_j'', respectively, are multiples of n_t (except m_1'):

$$m_j' = (j-1)n_t, \tag{5.8}$$

$$m_j'' = jn_t. \tag{5.9}$$

Eqs. (5.6) and (5.7) show also that for $s > 2\psi_h$, i.e., for high enough supersaturations, all peaks (including the first one) remain fixed at characteristic or

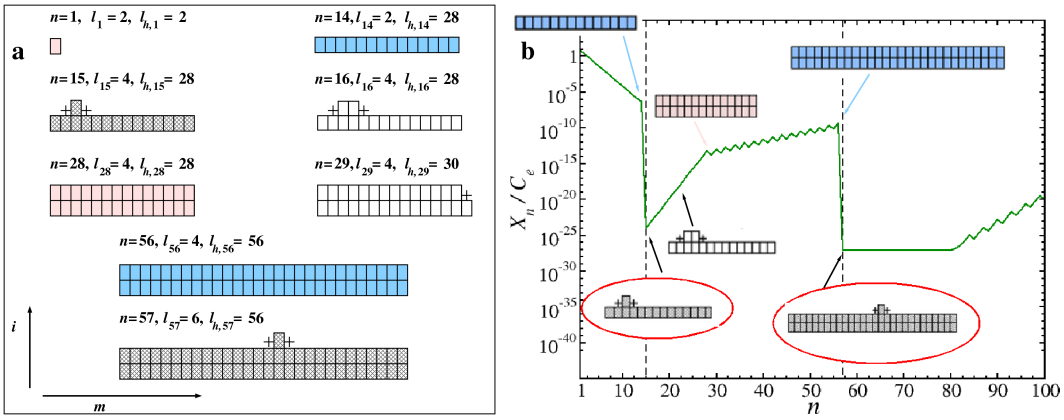

Figure 5.2: (a) Sequence of nanofibril shapes for Aβ_{40} with ratio $E/E_h = 14$. (b) Corresponding stationary nanofibril size distribution, X_n/C_e, according to Eq. (5.4) at $T = 300$ K and $s/\psi_h = 1.2$. In the graph some fibril shapes and their corresponding size distribution are illustrated. The nanofibrils of size n_j', n_j'' and n^* are shown in pink, blue and hatched, respectively, while the pluses indicate the kink sites.

magic fibril sizes n_j'' given by Eq. (5.7). The n_j''-sized fibril itself is $j\beta$-sheets thick and jn_t β-strands long (see the blue fibrils in Fig. 4.9), i.e. its length m_j'' is determined by Eq. (5.9). This equation implies that when $s > 2\psi_h$, if plotted as a function of the nanofibril length m, the stationary size distribution of the nanofibrils will show multiple peaks at magic lengths m_j'', the peaks being at the same distance n_t from their nearest neighbours on the m axis. However, this result regarding equidistant peaks in the nanofibril length distribution, characterises the nanofibril shape sequence illustrated in Fig. 4.9. Deviations from this sequence, in particular regarding the length at which the nanofibril acquires the energetically very costly β-strand that gives birth of its subsequent β-sheet, could lead to different inter-peak distances on the nanofibril length axis.

Combining Eq. (5.7) with the formula $n_j^* = j^2 n_t + 1$ for the successive nucleus sizes n_j^* (Eq. (3.14)), the following equation can be obtained ($j = 1, 2, 3, ...$):

$$n_j^* = n_j'' + 1, \tag{5.10}$$

which says that the j^{th} fibril nucleus is just one β-strand bigger than the fibril corresponding to the j^{th} peak at high enough supersaturations (see Figs. 4.9 and 5.2 in which the fibril nuclei are hatched, and the n_j''-sized fibrils are blue). Equation (5.10) is of particular interest, because it seems to be a result of general

validity when amyloid fibrils are nucleated by the mechanism of direct polymerization of β-strands and when $s > 2\psi_h$. According to this equation, with a negligible difference of one β-strand, the positions $n_j^{''}$ of the peaks in a nanofibril size distribution, that can be obtained in a real or computer experiment at a high enough supersaturation s, merely represent the sizes n_j^* of the successive fibril nuclei.

Table 5.2 lists the first three values of $n_j^{'}, m_j^{'}, n_j^{''}$, and $m_j^{''}$ in the respective s ranges. To get a feeling for these values for the Aβ_{40} protein, it can be useful to note that the molecular weight of the protein monomer is 4.3 kDa (48), and that the distance between β-strands in the Aβ_{40} fibril is about 0.47 nm (85). Hence, from Eqs. (5.7) and (5.9) with $n_t = 14$, the first three peaks of the nanofibril mass or length distribution are at magic masses 4.3 $\times n_j^{''}$, i.e., 60.2, 241 and 542 kDa or at magic lengths 0.47 $\times m_j^{''}$, i.e., 6.58, 13.2 and 19.7 nm respectively, provided the solution is at supersaturation $s > 2\psi_h = 3.18$.

The physical reason for which the size distribution X_n of nanosized fibrils is peaked at "magic" fibril sizes $n_j^{'}$ or $n_j^{''}$ is the presence of points of discontinuity of the work W_n to form a fibril of n β-strands as illustrated in Figs. 5.3a and 5.3b.

The attachment of a β-strand to the end of an $n_j^{'}$-sized fibril or to the surface of an $n_j^{''}$-sized fibril requires considerable work (see Fig. 5.3a) so that, depending on the supersaturation, the $n_j^{'}$- or $n_j^{''}$-sized fibrils become the most numerous in the solution and thus give rise to peaks in the fibril size distribution as illustrated in Fig. 5.3b. The s value at which the j^{th} peak is plateau-like coincides with that at which, according to CNT (Ref. (81), Eq. (13)), a $j\beta$-sheet is in thermodynamic equilibrium with the solution, i.e. it neither grows nor dissolves. Furthermore, regardless of the s value, a given peak concentration $X_{p,j}$ is always higher than the subsequent concentration $X_{p,j+1}$, i.e. the peak concentrations of the nanofibrils obey the ordering $X_{p,1} \geq X_{p,2} \geq X_{p,3} \geq \ldots$.

Clearly, X_n will be closer to a monotonically decreasing function of n when the work $(-s + 2\psi)k_BT$ (see Ref.(81)) of thickness-wise β-strand attachment to the kink-less surface of any of the $n_j^{''}$-sized fibrils (Fig. 4.9) is sufficiently smaller than the thermal energy k_BT. Thus, the condition for relatively low peaks in the stationary nanofibril size distribution reads

$$2\psi < 1 + s. \tag{5.11}$$

Aβ_{40}	Wild-type	V18N	V18R	V18P
n_1' $s/\psi_h < 2$	1	1	1	1
m_1' $s/\psi_h < 2$	1	1	1	1
n_1'' $s/\psi_h > 2$	14	14	13	13
m_1'' $s/\psi_h > 2$	14	14	13	13
n_1^* $s/\psi_h > 4/3$	15	15	14	14
n_2' $s/\psi_h < 1$	28	28	26	26
m_2' $s/\psi_h < 1$	14	14	13	13
n_2'' $s/\psi_h > 1$	56	56	52	52
m_2'' $s/\psi_h > 1$	28	28	26	26
n_2^* $4/3 > s/\psi_h > 4/5$	27	57	53	53
n_3' $s/\psi_h < 2/3$	1	1	1	1
m_3' $s/\psi_h < 2/3$	1	1	1	1
n_3'' $s/\psi_h > 2/3$	14	14	13	13
m_3'' $s/\psi_h > 2/3$	14	14	13	13
n_3^* $4/5 > s/\psi_h > 4/7$	15	15	14	14

Table 5.2: The first three magic sizes n_j', n_j'' and lengths m_j', m_j'' at which the size and length distributions of wild-type and point-mutated Aβ_{40} nanofibrils are peaked. For comparison, the sizes of the three fibril nuclei n_j^* are also given in the respective supersaturation ranges.

Since usually $\psi > 1$, the condition in Eq. (5.11) shows that well pronounced peaks at magic sizes (or lengths) are to be expected in the stationary size (or length) distributions of most amyloid nanofibrils. Moreover, the effect of the supersaturation is quite small, because amyloid fibril nucleation occurs at s values considerably smaller than those of ψ (18; 81). For nanofibrils of shorter β-strands the peaks will be less pronounced, because ψ diminishes with decreasing β-strand length. In principle, any bio-physicochemical factor that lessens the ψ value will also contribute to the suppression of the X_n peaks and, hence, to the possible disappearance of the magic nanofibril sizes or lengths.

Preferred sizes, lengths or thicknesses of Aβ oligomers or fibrils have been recorded in many experiments (e.g. (2; 41; 47; 48; 85; 117)). In this respect, the actual finding of magic sizes is in qualitative agreement with experiment. None of the known experimental data, however, allows their use for a quantitative

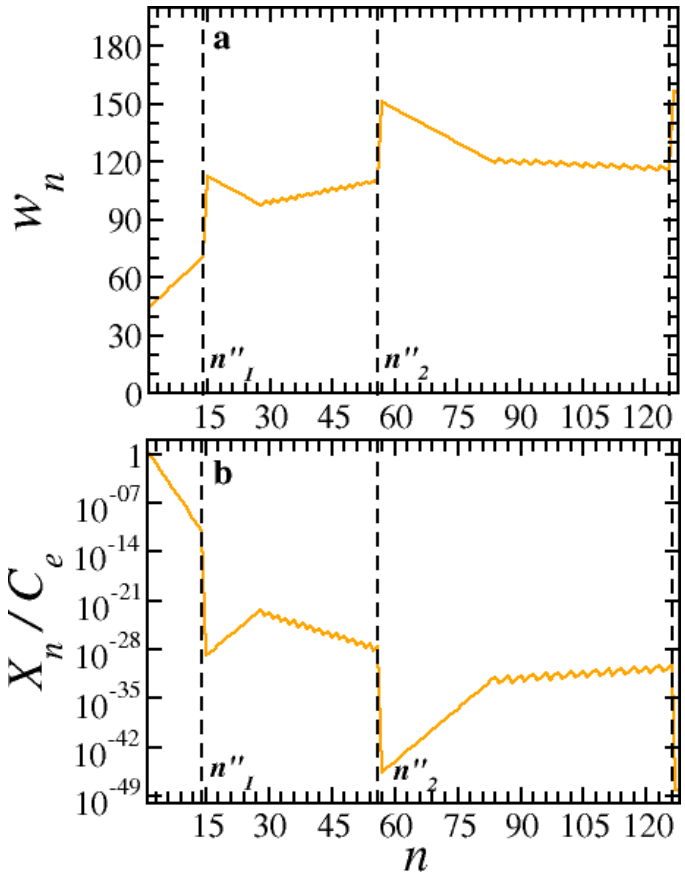

Figure 5.3: (a)Dimensionless work to form a n-sized Aβ_{40} fibril at supersaturation $s/\psi_h = 1.16$ according to the Eq. (3.1). (b) Corresponding size distribution X_n according to Eq. (5.4). The two panels show that the deep local minima for the work n_1'' and n_2'' correspond to peaks in the size distribution function. Moreover, the figure emphasises the close relation in the shape of the two curves.

verification of this finding. Reported histograms of the fibril mass-per-length ratio (2; 21; 46; 47; 48; 91; 117; 123; 126) exhibit one or more peaks, but pertain to size ranges corresponding to rather long fibrils, and there is evidence that these usually contain an integer number of β-sheets. What is needed to reliably quantify the magic nanofibril sizes are histograms of the masses of entire fibrils of nanoscale length. Only Goldsbury et al. (48) have obtained such a histogram (Fig. 4D in (48)), but this refers to pseudo-spherical rather than fibrillar Aβ_{40} aggregates and the authors could fit it by two Gaussian curves with maxima at masses of 311 and 554 kDa corresponding to 72 and 128 protein molecules. The experimental size distributions of Garai et al. (41) for solutions of Aβ_{40} also

exhibit peaks at magic sizes represented by the aggregate hydrodynamic radii. Again, a quantitative comparison between the present theory and Garai et al. experiment (41) is not possible, because the reported size distributions are not stationary, the aggregate morphology is unknown, and the second peak of the size distributions is at radius of approximately 50 nm. This radius corresponds to the rather large number of 1.2×10^5 protein molecules if, as in Garai et al. (41), one uses 4.3 nm^3 for the Aβ_{40} molecular volume and assumes that the aggregates are spherical.

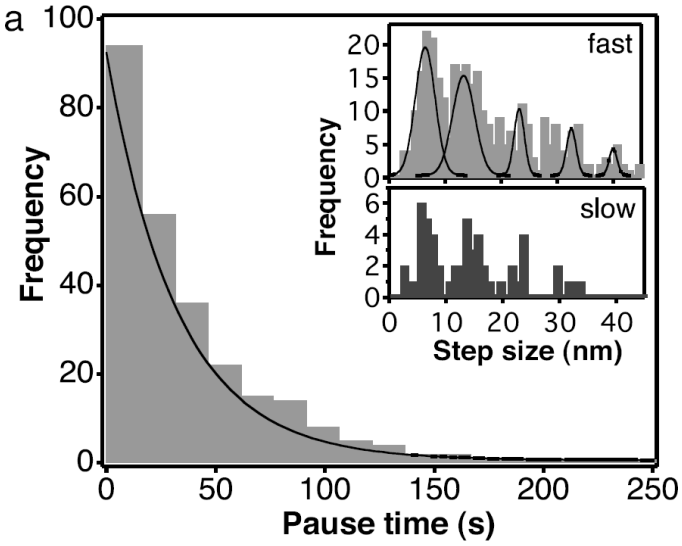

Figure 5.4: From the paper of Kellermayer et al. (85), in the inset a histogram for the increments in the lengths of epitaxially growing Aβ_{25-35} fibrils is shown. As discussed in the text, the histogram exhibits five peaks that are consistent with the results presented in the chapter.

Very interesting as well are the experimental data of Kellermayer et al. (85) which support the finding of magic nanofibril lengths. These authors have presented a histogram of most frequently observed increments of the lengths of epitaxially growing individual Aβ_{25-35} fibrils (the inset in Fig. 2a of Kellermayer et al. (85) is reported in Fig. 5.4). The histogram exhibits five successive peaks at 6.5, 13.3, 23.2, 32.5 and 40 nm. As noted by the Kellermayer et al. (85), the inter-peak distance of nearly 7 nm "corresponds to a stretch of fibril containing $\approx$ 15 peptides along its length ". This is in close quantitative agreement with the present finding for peaks in the Aβ_{40} nanofibril length distribution, which

are equidistant at 14 β-strands or, equivalently, 6.6 nm. However, this agreement is very likely to be fortuitous, because while the fibrils of Kellermayer et al. elongate at a practically constant thickness, according to the present model (Fig. 4.9), every subsequent fibril lengthening of n_t β-strands is preceded by fibril thickening of one β-sheet.

Effect of mutation on the size distribution

The theoretical framework presented in this chapter, enables to study also the effect of point mutations on the nanofibril size distribution. In particular, at fixed T and C_1, how X_n is affected by the point mutations of residue 18 (V) with amino acids Asparagine (N), Argininge (R) and Proline (P) has been examined. The present analysis is compared with experimental results from Christopeit et al. (24) where the effect of these Aβ_{40} mutations on the fibrillation kinetics has been investigated.

In order to calculate X_n from Eq. (5.4) for each mutant, the corresponding fibril solubility C_e and energy parameters ψ and ψ_h for each mutant have to be known. In fact all the mentioned factors have different values for each mutant because the binding energies E and E_h of the β-strands within the fibrils are different (see Table 5.1). For the wild-type Aβ_{40}, the experimentally obtained C_e = 3.9 μM at T = 300 K (41) has been employed. As C_e has not been determined experimentally for any of the mutants, it has been calculated theoretically by making use of Eqs. (4.6)-(4.8) and of the ψ and ψ_h in Table 5.1. The C_e values computed in this way for fibrils with point mutations V18N, V18R, and V18P are given in Table 5.1. Following the procedure outlined above for the wild-type Aβ_{40}, from Eq. (5.4) X_n has been calculated for each mutant by setting $M = 300$ (as in Ref. (18)), by using the ψ and ψ_h values from Table 5.1 and the l_n and $l_{h,n}$ values from Tables A.2 and A.3 in Appendix 1. For mutations V18N, V18R and V18P the transition size $n_t = E/E_h$ takes the values 14, 13 and 13, respectively, and is equal or nearly equal to the transition size n_t for the wild-type protein (Table 5.1).

The nanofibril size distributions obtained at T = 300 K and C_1 = 120 μM are shown in Fig. 5.5. At this concentration, the supersaturation for the wild-type protein and the three mutants V18N, V18R and V18P can be calculated from Eq. (4.4) and is s = 3.43, 2.43, 1.43, and 0.63, respectively. As can be seen in Fig. 5.5,

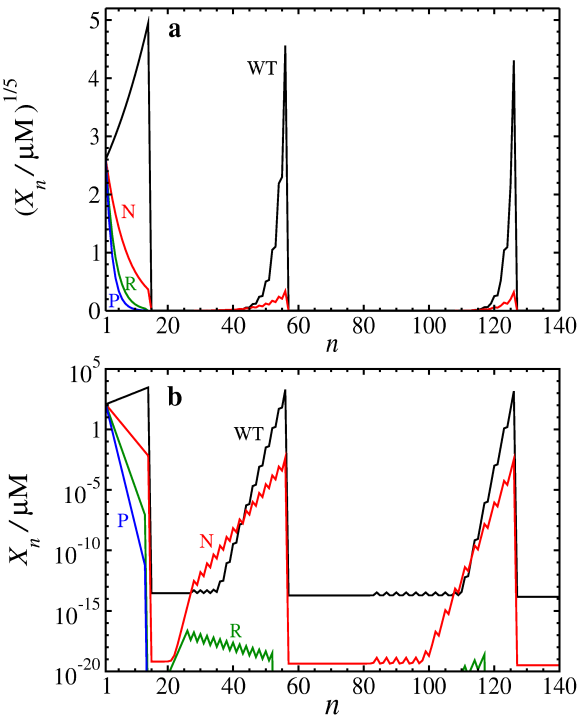

Figure 5.5: Stationary nanofibril size distribution X_n on (a) $X_n^{1/5}$ and (b) log X_n scale according to Eq. (5.4) at $T = 300$ K and monomer concentration $C_1 = 120$ μM for wild-type Aβ_{40} and its mutants V18N, V18R, and V18P: black line WT, wild-type; red line N, V18N; green line R,V18R; and blue line P, V18P.

the main effect of the mutations is that the peaks for all magic fibril sizes n_j' or n_j'' lose height. This is so because by considerably increasing the fibril solubility, the mutations lower the solution supersaturation, which leads to a strong decrease in the entire fibril population and, accordingly, in the peak heights. In contrast, the effect of the mutations on n_j' or n_j'' is minor (because n_t changes only from 14 to 13 for two of the mutants) and, as a result, the peak positions remain essentially the same. For mutant V18R, the peak is at $n_2' = 26$, because the C_e value for this mutant is so high that the corresponding supersaturation ratio C_1/C_e is too low for the peak to be at $n_2'' = 52$. The effect of mutations on the fibril population could be exemplified by considering the number of nanofibrils of 13 β-strands in a solution of volume V = 100 ml, as typically used in experiments. From Eq. (4.4) (see also Fig. 5.5) it has been calculated that while the number of wild-type fibrils of size $n = 13$ is 10^{20}, the V18N and V18R fibrils of the same size are much less numerous: the V18N ones are 9×10^{14}, and the V18R ones are 10^{10}.

Remarkably, the fibrils of the same size for mutant V18P are practically absent from the solution, because their number is 4×10^5. It is worth noting also that the decrease of the peak heights for the different mutants, seen in Fig. 5.5, is consistent with experiments on protein fibrillation kinetics (24), which reveal that the lag-time before detectable fibrillation, is longer for the mutants V18N, V18R than for the wild-type Aβ_{40}; and that the mutant V19P is not detectable at all. This consistency supports the correlation reported in the previous chapter and in Ref. (17) between ANT nucleation rates of amyloid fibrils and experimentally measured fibrillation lag-times of wild-type and point-mutated Aβ proteins.

5.4 Conclusion

The analysis carried out in this chapter illustrates the ability of the atomistic nucleation theory to describe the size distribution of amyloid nanofibrils in protein solutions when nucleation occurs by the mechanism of direct polymerization of β-strands into β-sheets. The atomistic modelling of the amyloid nanofibrils of successively layered β-sheets leads to a general formula Eq. (5.4) for the stationary fibril size distribution as a function of the protein concentration and of the temperature (via C_e, ψ and ψ_h that, given the nanofibril shape sequences, are the only three theoretical parameters). Application of this formula to Aβ_{40} nanofibrils reveals the existence of a series of characteristic peaks in the fibril size distribution, which are positioned at "magic"fibril sizes and which are due to points of discontinuity of the work for fibril formation. This finding of magic sizes or lengths is consistent with experimental results for the size distribution of aggregates in solutions of Aβ_{40} proteins. This analysis provides a new opportunity to obtain information about the sizes of the successive fibril nuclei, because according to Eq. (5.10), these sizes are practically equal to the magic sizes at which the nanofibril size distribution is peaked when the solution supersaturation is sufficiently high. Also, this atomistic approach makes it possible to gain novel insight into the effect of point mutations on the size distribution of amyloid nanofibrils, an effect that may play a role in experimentally observed substantial differences in the fibrillation lag-times of wild-type and point-mutated Aβ proteins.

The results obtained remain without quantitative verification because of the lack of suitable experimental or simulation data for the size distribution of amy-

loid nanofibrils. It is hoped that the present study will inspire new real and computer experiments especially aimed at investigating this distribution and/or the magic nanofibril sizes or lengths.

Chapter 6

Kinetic Monte Carlo simulations of amyloid fibril nucleation[1]

In the previous chapters ANT has been outlined and applied to study amyloid fibril nucleation and it has been validated by comparison with experimental results. The ANT computations relied on a specific growth mechanism that has been chosen using the CNT evolution mechanism as a reference. In this chapter the growth mechanism will be tested by kinetic Monte Carlo (KMC) simulations. The idealised character of the model allows for the simulation results to be applied to any nucleating system with strongly anisotropic interactions between molecules, thus covering the nucleation of amyloid fibrils and also a variety of other organic and inorganic systems. For that reason, all the discussion in this chapter will refer to crystal systems and crystal nucleation in general, rather than being restricted to amyloid fibrils and amyloid fibril nucleation. In particular, the nucleation of two-dimensional model crystals will be studied in order to gain insight into the effect of anisotropic interactions between molecules on the stationary nucleation rate J. With the aid of kinetic Monte Carlo simulations, J has been determined as a function of the supersaturation s. It turned out that with an increasing degree of interaction anisotropy the dependence of $\ln J$ on s becomes step-like, with jumps at certain s values. It will be shown that this $J(s)$ dependence cannot be described by the classical and atomistic nucleation theories. A formula that predicts the identified $J(s)$ behaviour is yet to be derived and verified, and the present study provides the necessary data and understanding for doing that.

[1] This chapter is based on a submitted paper.

6.1 Introduction

Crystal nucleation is the process of random generation of nanoscale crystalline clusters that have the ability of irreversible growth (e.g., Ref. (77)). Unless the size of the smallest clusters exceeds the size n^* of the nucleus, the clusters are more likely to dissolve than to grow (n^* is the number of molecules in the nucleus which is defined as the cluster that requires maximum work W^* for its formation). Since only clusters bigger than the nucleus can grow to macroscopically large crystals, the nucleation rate J is defined as the number of supernuclei that form in a supersaturated old phase per unit volume and time.

Finding J is a central problem in nucleation theory because J has a strong impact on many important quantities of practical significance in crystallization, crystal growth and thin film formation, such as the size distribution and the maximum number of crystals formed in the system, the rate of nucleation-mediated crystal growth, and the induction time in crystallization. Theoretically (see, e.g., Ref. (77)), J is of the form $J = Ae^{-W^*/k_BT}$, where A is a kinetic factor, k_B is the Boltzmann constant, and T is the absolute temperature. Based on this formula, the nucleation theory predicts that in the nucleation of condensed phases at fixed T, J is a concavely increasing function of the supersaturation s.

The various nucleation theories are based on the classical description of the nucleation of phases whose atoms or molecules interact with isotropic potentials, and the effect of anisotropic interactions on the nucleation barrier and/or rate has received only a little attention so far (e.g. Refs. (5; 18; 27; 57; 68; 72; 81)). Previous work has focused mainly on phase diagrams and on geometrical aspects of self-assembly (e.g. Refs. (3; 7; 26; 27; 57; 86; 99; 120; 121; 148; 163)). However, understanding the effect of anisotropic interactions on the nucleation kinetics is important, because condensed phases that form due to directional interactions between molecules are ubiquitous in nature. Prominent examples include the amyloid fibrils whose formation is driven by the strong directional hydrogen bonds between monomeric proteins (22), the liquid crystalline phases that form because of the anisotropic shape of their molecules (25), the close-packed structures and string fluids formed by soft and dipolar colloidal particles (160), and the polymer crystals whose morphology depends on the differently strong interactions between the polymer units (63). Thus, the important question arises: Can existing nucleation theories describe the effect of anisotropic

molecular interactions on the $J(s)$ dependence? The objective of this study is to provide an answer to this question.

6.2 Model and simulation method

According to Stranski and Kaischew (134), pioneers in the classical nucleation theory (CNT), the essential features in the processes of crystal nucleation, growth and dissolution can be seen even in the simplest model. In this respect, the nucleation of two-dimensional (2D) Kossel-Stranski crystals on their own substrate (the (100) face of a three-dimensional (3D) Kossel-Stranski crystal) is a simple prototype that allows for investigation of the effect of different intermolecular bonds on the nucleation behaviour. In line with the Kossel-Stranski model which has a one-to-one correspondence with the Ising model, the molecules (or the building blocks) of the 2D crystal are arranged in a 2D lattice with square symmetry (which represents the substrate in Fig. 6.1). The 2D crystal model re-

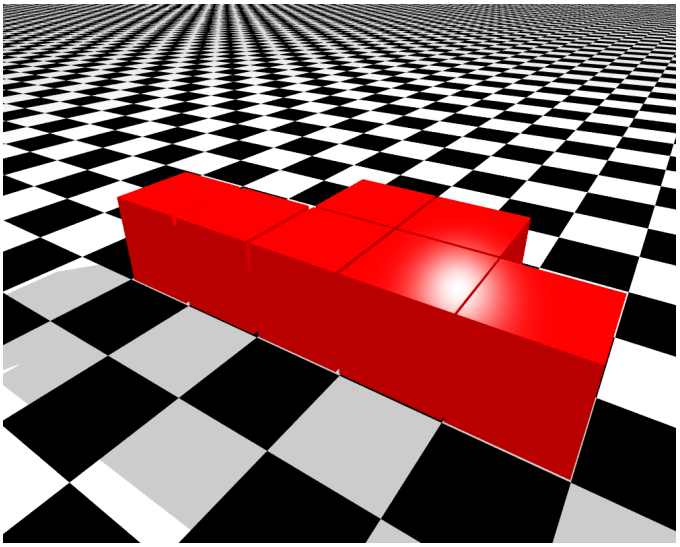

Figure 6.1: Artistic view of the Kossel-Stranski model used to investigate the nucleation of 2D crystal. The cubes schematise molecules with strong and weak interactions along the m and i axes, respectively (the m axis is parallel to the row of five molecules in the anisotropic simulations results).

ferred to in this chapter, is essentially equivalent to the 2D fibril model explained in the previous chapters, except that molecules, nucleating into a crystal, take now the place of the β-strands, nucleating into a nanosized amyloid fibril. In

particular, in this model each monomer is represented by a right rectangular prism with two edges parallel to the m and two edges parallel to the i axis. Furthermore, the monomers are arranged into the nanocrystallite with i rows (i = 1, 2, 3,...) along the i axis and each i-row is constituted of m (m = 1, 2, 3,...) identical monomers (Fig. 6.1).

Because only nearest-neighbour interactions between molecules have been taken into account, intermolecular bonds exist solely along the m and i axes. These are referred to as the 2D crystal lengthening and widening axis, respectively. As for the β-strands within an amyloid fibril, the molecular bond strength within the crystal is characterised by the quantities $\psi = E/2k_BT$ and $\psi_h = E_h/2k_BT$, where E and E_h are the interaction energies between nearest neighbours molecules along the m and i axes, respectively. The parameter $\xi = \psi/\psi_h$ quantifies the anisotropy of the interactions between molecules, $\xi = 1$ meaning isotropic interactions.

To investigate the effect of different interactions on the nucleation rate J, kinetic Monte Carlo (KMC) simulations of the type described in Refs. (13; 141; 142) have been performed. For each molecular interaction, the performed KMC simulations were depicting the nucleation of a 2D cluster on its own substrate (the grid in Fig. 6.1) from monomers at different values of the supersaturation. The method considers only attachment and detachment events of monomers to and from a given n-sized cluster without rejection of any events (13) and periodic boundary conditions included. Furthermore, the algorithm used neglects to simulate the diffusion of molecules along the substrate (the 3D crystal face), because the diffusion does not affect the results (45).

When only nearest-neighbour interactions between monomers are taken into account, the probability p^+ of monomer attachment to a molecular site at the cluster periphery, and the probability p^- of detaching the j^{th} peripheral monomer, are given by the expressions :

$$p^+ = k^+/[l_n^+ k^+ + \sum_{j=1}^{l_n^-} k^-(b^j, b_h^j)] \tag{6.1}$$

and

$$p^- = k^-(b^j, b_h^j)/[l_n^+ k^+ + \sum_{j=1}^{l_n^-} k^-(b^j, b_h^j)]. \tag{6.2}$$

Here k^+ is the frequency of monomer attachment to a molecular site at the cluster periphery, $k^-(b^j, b_h^j)$ is the frequency of detaching the j^{th} peripheral monomer, b^j and b_h^j are the numbers of nearest-neighbour bonds that this monomer has along the m and i axis, respectively, l_n^- is the number of peripheral monomers, l_n^+ is the number of attachment sites at the cluster periphery, and n is the number of monomers constituting the cluster. For the monomer attachment frequency (k^+) the following formula (45; 141; 142) has been used

$$k^+(s) = k_e \exp(s), \tag{6.3}$$

in which the frequency k_e is the value of k^+ at phase equilibrium (when $s = 0$). As seen from the latter formula, k^+ depends on the supersaturation, but it is independent of the number of bonds formed by the attached monomer. The frequency k^+ is, indeed, the same for each site of the cluster to which the monomer could be attached and it is bigger the higher is the supersaturation, since the chance for a monomer to hit a cluster site, and hence attach, increases with increasing s. Physically the attachment event occurs because of a random "rain" of molecules on the cluster.

Conversely to the attachment frequency k^+, the detachment frequency k_j^- of the j^{th} peripheral molecule is independent of the supersaturation and depends on the number of inter-monomer bonds that it is necessary to break in order to detach the monomer. For that reason its formulation depends on the kind of molecular interaction present, i.e. isotropic or anisotropic interactions. The two formulations will be given and described in detail later in the chapter.

To briefly illustrate the KMC algorithm, a cluster with a n free sites as a starting configuration can be considered. As an input for the simulation, the dimensionless inter-molecular broken bond energies ψ and ψ_h, the maximum size of the cluster M, the supersaturation and the position of the monomers in the initial cluster are given. At each Monte Carlo (MC) step, it will be associated to each of the n-sites ($j = 1, 2, 3, ..., n.$) a probability p^+ to attach a new monomer and a p_j^- to detach the monomer occupying the position correspondent to the site, according to the frequency described above. The summation of all the probabilities will be used to normalise each probability and through a random number in the range between [0,1) one event will be chosen, according to the attachment and detachment probabilities. Then the MC step will result in

either a new monomer in the cluster ($(n+1)$-sized cluster) or a missing monomer from the cluster ($(n-1)$-sized cluster). Then a new MC step will follow. The simulation will be terminated when the cluster either decays to a monomer or grows up to a size M, that is the large enough supernucleus ($M > 2n^*$). The M value corresponds to a size at which the cluster does not decay to monomer and furthermore a larger value of M would not have any effect on the nucleation rate J.

The above algorithm has been applied to obtain the nucleation rate J. From a computational point of view, the dimer is the natural initial configuration to obtain the nucleation rate J. Starting from a dimer, in fact, the stationary nucleation rate J can be determined with the aid of the exact general formula (77; 141)

$$J = f_1 C_1 P_2. \tag{6.4}$$

In Eq. (6.4), P_2 is the dimer growth probability (which is defined as the probability that a dimer will grow to a large enough supernucleus size M before decaying to a monomer), f_1 is the monomer-to-monomer attachment frequency, and C_1 is the actual concentration of monomers. Thus, physically, P_2 is merely the nucleation rate in units of the product $f_1 C_1$ in which f_1 and C_1 are known or obtainable by considerations outside nucleation theory. When nucleation occurs at fixed T, C_1 can be used to determine the supersaturation s by the relation $s = \ln(C_1/C_e)$ in which C_e is the equilibrium concentration of monomers (77). In the present case, given the ψ and ψ_h values, each simulation run to determine P_2 at a specified value of s started with a dimer oriented along the m axis and the run terminated when the dimer either decayed to a monomer or grew to a 2D cluster of $M = 200$ molecules, because it was established that clusters of this size do not decay to monomers. The dimer growth probability P_2 was calculated from the relation $P_2 = N^+/(N^+ + N^-)$ where N^+ is the number of "positive runs" (those in which the dimer grew to the size of 200 molecules), and N^- is the number of "negative" runs (those in which the dimer decayed to a monomer).

A crucial ingredient of all Monte Carlo simulations is represented by the random number generator, whose quality becomes particularly important for very long simulations. The random number generator used in the simulation is the "Mersenne Twister pseudorandom number generator (MT19937-64)" because its

reliability is extensively documented in the literature. Importantly, detachments leading to a fragmentation of the clusters in two or more clusters should not be allowed. It has been verified that this is not statistically common and when it does happen, during the following MC step, a monomer would be created in the same site of the previous detachment. Furthermore, the program contained a subroutine to avoid all the cases that would have led to fragmentation of the cluster. The results have shown that P_2 is not considerably affected by the fragmentation mechanism. Then to save computational time the use the subroutine for the rest of the simulations has been avoided.

6.3 Kinetic Monte Carlo simulations for isotropic systems

Previous studies considered the nucleation of crystals formed by monomers with isotropic interactions (140; 141; 142). In the cited works the Bortz Kalos Lebowitz kinetic Monte Carlo (BKL KMC) algorithm (13) has been performed to study the growth of a 2D cluster on the Kossel (001) substrate from a vapour, melt or solution of monomers. In these papers the monomers within the cluster have an idealised cubic shape and in each MC step a monomer can be created on or annihilated from the cluster, simulating a cluster that can grow or shrink by attachment or detachment solely of monomers. In Ref. (140) the algorithm was used to determine the Zeldovich factor z and the nucleus size n^* from the growth probability of the 2D cluster. In Ref. (141) the authors used the BKL KMC method to determine, from the dimer growth probability, the nucleation rate J of a 2D isotropic Kossel crystal nucleating on its own substrate. Finally in Ref. (142) the simulation results for the previous 2D Kossel crystal were used to test the CNT and ANT predictions for the nucleation rate J at different values of the molecular isotropic energy. In all the cases studied the molecules within the cluster are bonded by isotropic interactions so that in the correspondent 2D lattice the dimensionless broken bond energies between nearest neighbours molecules along the m and i axes, ψ and ψ_h respectively, are both equal, i.e., $\psi = \psi_h = \omega$. Then, the anisotropy degree is $\xi = 1$. In this section, the results from the present computational study of the 2D Kossel isotropic crystal nucleation from monomers will be discussed. The implementation of the computer program to simulate isotropic systems nucleation has been the first step to the study of the 2D anisotropic nucleation as will be shown in the next section.

Attachment and detachment frequencies

According to the detailed balance condition, the attachment frequency (or creation rate) of a monomer to an atomic site in the cluster is given by Eq. (6.3), while the detachment frequency (or annihilation rate) of the j^{th}-peripheral atom from the cluster is expressed for a 2D system with isotropic molecular interactions by the formula (140; 141; 142):

$$k^-(b^j, b_h^j) = k_e \exp[2\omega(2 - (b^j + b_h^j))]. \tag{6.5}$$

From the latter formula, it can be seen that, at variance from the attachment frequency k^+, the detachment frequency $k_j^-(b^j, b_h^j)$, depends on the total number of inter-monomer bonds $b^j + b_h^j$ of the j^{th}-peripheral atom which would be detached from the cluster edge. Furthermore, k_j^- is independent of the supersaturation. The expression is consistent with the fact that detaching a monomer bonded to more nearest-neighbours is less likely to happen than detaching a monomer with fewer nearest-neighbours.

Results and comparison with nucleation theory

To investigate the effect of different isotropic interactions between molecules on the nucleation of monomers into a 2D Kossel crystal, a BKL KMC program has been implemented and different simulations have been performed. In particular, in order to study the effect of different ω values on the nucleation rate and to test the ANT and CNT predictions for J, a series of simulations at fixed value of the dimensionless broken bond energy $\omega = 1$, 2, 4, and 5 have been performed. Each ω simulation runs for different supersaturation values in a range that depends from the energy ω. The simulation for $\omega = 1$, for example, runs for the supersaturation values in the range $0.03 \leq s \leq 4.0$ while for the simulation at $\omega = 5$, the supersaturation values used were in the range $3.5 \leq s \leq 11.75$. According to the general algorithm mentioned above, each ω simulation that runs at a given s value started from a dimer initial configuration and ended when the cluster either decayed to a monomer (negative run N^-) or grew to cluster of $M = 200$ monomers (positive run N^+). Then P_2 has been calculated according to Eq. (6.4). To collect enough statistics the simulation at a given ω and s has been performed with different random number sequences for a total number of runs ($N_{tot} = N^+ + N^-$) sufficient to make the error in the measure of P_2 smaller then a pre-selected value. To reach this requirement a number of positive run $N^+ \approx 100$ was generally enough. The total number of runs to obtain P_2 depended on the supersaturation investigated. For example, N_{tot} was between 2×10^9 for the smallest s value ($s = 0.03$ for the simulation at $\omega = 1$) and 2×10^5 for the largest s value ($s = 11.75$ for the simulation at $\omega = 5$).

For kinetic reasons and by convention, (77; 141; 142) the value of 0.5 for the cluster growth probability $P(n)$ is associated to the cluster with $n = n*$. This can be intuitively understood considering that the critical nucleus correspond to the top of the barrier to form an n-sized cluster, according to this interpretation, the condition $P(n^*) = 1/2$ means that the cluster with $n = n^*$ has 50% of chances to decay to a monomer or to grow to macroscopic sizes.

Fig. 6.2 shows the nucleation rate J calculated according to Eq. (6.4) for each series of simulations at fixed ω. The nucleation rates from the simulations shown

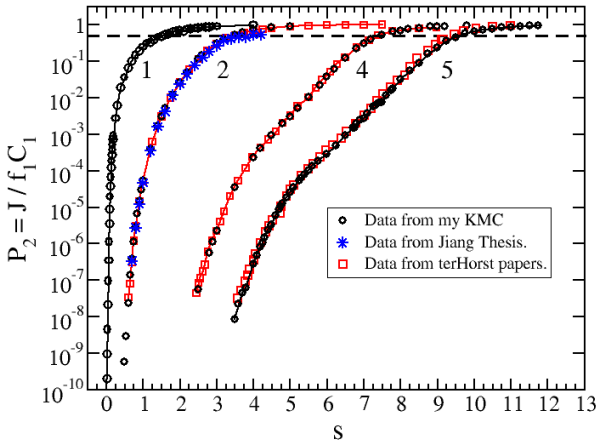

Figure 6.2: Supersaturation dependence of the nucleation rate J by KMC simulation studies at ω= 1, 2, 4 and 5. The present simulation data are shown in black circles, while Jiang's, ter Horst-Kashschiev data (Refs. (68; 141; 141; 142; 154)) are indicated with blue stars and red squares, respectively, as labelled in the graph (the points have been extracted by the thesis and papers (68; 141; 141)).

in the graph display an excellent agreement with the simulation data reported in Refs. (68; 141; 154) for $\omega = 2$, and in Refs. (141; 142) for $\omega = 4$ and 5. In particular, the present simulation results are shown in black circles while the red squares and blue stars represent the data whose origin is given in the legend. For smaller ω values (see curves for $\omega = 1$ and 2 in Fig. 6.2), the $J(s)$ curves from the simulation data are smooth lines representing functions that increase monotonically with increasing supersaturation s. Remarkably, only for higher values of the energy ($\omega = 4$ and 5), linear parts of $\ln J$ are connected by bends as can be seen in Fig. 6.2. The presence of linear parts in $\ln J$ is a typical

feature of the ANT theoretical curves as treated in chapter 3. In the same graph the horizontal dashed line represents the condition $P_2 = 1/2$ that allows for the detection of the supersaturation corresponding to which the dimer is a critical nucleus size. By the graph, the dimer is a critical nucleus size for $\omega = 1$ already at $s = 1.7$, while for $\omega = 2$ at $s = 3.5$, for $\omega = 4$ at $s = 7.5$ and finally for $\omega = 5$ at $s = 9.5$.

The present simulation data were used to test the CCNT and ANT predictions given by Eqs. (2.29) and (3.2), respectively, as shown in Fig. 6.3.

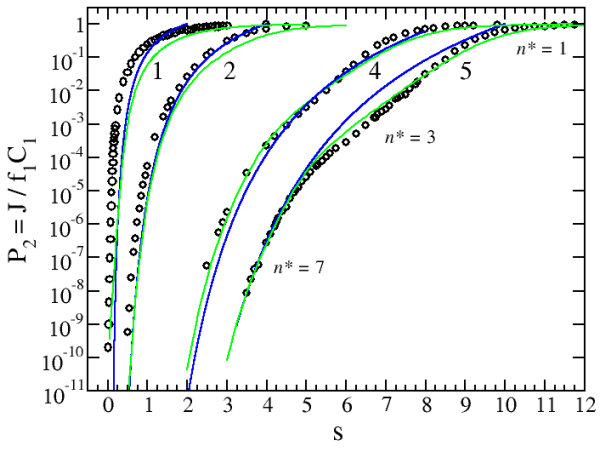

Figure 6.3: Supersaturation dependence of the nucleation rate J at $\omega = 1$, 2, 4 and 5 as indicated in the figure. The open circles represent the present simulation data, while the blue and the green curve are the CCNT and ANT theoretical predictions for $J(s)$ according to Eqs. (2.29) and (3.2) as explained in the text.

In particular, to plot the ANT $J(s)$ for the present isotropic Kossel system, an expression (Eq. (28) of Ref. (80)) of the exact Becker-Döring formula (3.2) in combination with a particular growth mechanism (Table I in Ref. (80)) have been used. The agreement between J from the simulation data and the theoretical predictions for the nucleation rate by ANT and CCNT is generally fairly good. The blue lines show $J(s)$ dependence predicted by CCNT according to Eq. (2.29), suggesting that for small ω values ($\omega = 1$ and 2) CCNT slightly underestimates the rate respect to the simulation data. For $\omega = 4$ the agreement between the KMC data and CCNT line is very good, but for higher values of the energy ($\omega = 5$) CCNT line overestimates the KMC data. On the other hand ANT

underestimates the simulation data for $\omega = 1$ and 2, while its agreement with the KMC data for $\omega = 4$ and 5 is very good, reproducing the broken linear parts by sharp bends reported by the simulation data for $\omega = 4$ and 5.

All of these results are confirmed by the computational and theoretical studies in Refs. (142) and (80). In Ref. (142), KMC simulation data for J at $\omega = 2$, 4 and 5 are compared with the CNT and ANT theoretical predictions of the nucleation rate. The authors in Ref. (142) conclude that ANT nucleation rate describes much better the KMC data for J than CNT does, and that the agreement with the ANT predictions and simulations data improves for higher energy values. In Ref. (80) theoretical studies investigate the ANT, CNT and CCNT predictions for the nucleation rate J. Furthermore KMC simulation data for J are also compared with the theoretical predictions. The paper confirms that the ANT nucleation rate describes much better the KMC data for J compared to CNT, and introduces the corrected classical description. The CCNT agreement with the ANT predictions and the simulation data is good but still CCNT does not reproduce the broken linear parts for higher ω values.

To summarise, ANT and CCNT predict correctly the smoothness of $\ln J$ from the simulations at $\omega = 1$ and 2, but both of the theoretical curves underestimate the KMC data. Only ANT can predict the broken lines given by the simulation results for $\omega = 4$ and 5, but the breaks are more pronounced in the theoretical predictions than in the simulations data. However, in general the theoretical lines are close to the points given by the simulations, according to the results and graphs obtained in Ref. (80).

In general, the differences between the ANT predictions and the circles in the graph 6.3 can be interpreted as suggested in Ref. (142). The bends between the linear part in $\ln J$ are sharp and exaggerated by the ANT predictions for J, and ANT slightly underestimates the simulation data for the lower value of ω because the simulation data are a statistical result to which different shapes contribute. In ANT, instead, the theoretical curves assume that for each n-size of the cluster only one shape contributes to the calculated nucleation rate J, i.e. the one that corresponds to the cluster with lowest surface energy. Always in this prospective, the underestimation of the KMC data for smaller ω values by ANT can be interpreted by considering that the smaller the energy value is, the rougher the cluster periphery will be. The resulting variety of shapes is

considered in the KMC data, while ANT is based on a growth imposition that associates to each size n of the cluster only one shape.

Also, the nucleus sizes at successive s ranges according to Eq. (2.31) have been obtained for $\omega = 4$ and 5. The nucleus size data are in agreement with the ones predicted by ANT with Eq. (2.31) since the slope for all the ω curves is predicted well by the ANT equations. In particular, the graph reports the nucleus sizes for $\omega = 5$.

6.4 Kinetic Monte Carlo simulations for anisotropic systems

In order to understand the effect of anisotropic interactions between monomers on the nucleation of a 2D nanocrystal, the BKL KMC method has been used. The computational method here is an extension of the simulation technique widely used to study isotropic systems as explained in the previous section. In particular, to gain insight into the bond-energy anisotropy effect on the nucleation rate J and on the morphology of the clusters, the $P_2(s)$ dependence has been determined for increasing anisotropic degree $\xi = \psi/\psi_h = 1$, 3, 5, 8, 10 and 14, with a common value of $\psi_h = 1$. Understanding the effect of the anisotropic interactions on the kinetics of the nucleation process is a fundamental problem that finds application in different fields. The subject of this thesis is modelling the kinetics of amyloid fibril nucleation, the fibril formation is driven, in fact, by the strong directional hydrogen bonds between monomeric proteins. Even though the simulations are carried out for a specific aim, the idealised character of the model implies that the results can be applied to any system that is nucleating into a 2D crystal with anisotropic interactions between its molecules.

Attachment and detachment frequencies

In the last section, previous studies on the nucleation of crystals with isotropic interactions have been reviewed. For the same 2D crystal model with anisotropic bond energies, an expression for the detachment frequency $k^-(b^j, b_h^j)$ of a monomer from the cluster periphery has been derived. As known from Boltzmann statistics, the probability for transition from an initial to a final state decreases exponentially with the work spent on the transition. The detachment of the j^{th}

peripheral monomer requires the work $b^j E + b_h^j E_h = 2(b^j \psi + b_h^j \psi_h) k_B T$. Therefore, the frequency $k^-(b^j, b_h^j)$ of monomer detachment from the cluster periphery can be written down as

$$k^-(b^j, b_h^j) = k_{ref} \exp(-2b^j \psi - 2b_h^j \psi_h), \tag{6.6}$$

where k_{ref} is a reference frequency. Elimination of k_{ref} can be achieved with the aid of the relation $k^-(1,1) = k^+(0)$ for the equality between the frequency of monomer attachment at equilibrium and the frequency of monomer detachment from a kink site (for this site $b^j = b_h^j = 1$ because of the square symmetry of our 2D crystal model). Using this relation, Eqs. (6.3) and (6.6) leads to

$$k_{ref} = k_e \exp(2\psi + 2\psi_h), \tag{6.7}$$

so that

$$k^-(b^j, b_h^j) = k_e \exp[2(1 - b^j)\psi + 2(1 - b_h^j)\psi_h]. \tag{6.8}$$

As seen from this equation, in contrast to k^+, k^- depends on the number of intermolecular bonds, but is independent of the supersaturation, exactly as for the isotropic case. It is worth noting as well that, as can be verified, k^+ and k^- from Eqs. (6.3) and (6.8) satisfy the principle of detailed balance. Also, at $\psi = \psi_h$ they correspond to the known attachment and detachment frequencies in the case of 2D crystals with isotropic bond energies as given in the previous section and in different papers (45; 141; 142). Substitution of k^+ and k^- from Eqs. (6.3) and (6.8) into the formulae for p^+ and p^-, Eqs. (6.1) and (6.2), respectively, yields the probabilities used in the KMC simulations.

Results and comparison with nucleation theory

Figure 6.4 shows representative shapes of clusters at successive moments of their evolution during simulations at anisotropy ratio $\xi = 1$, 3 and 8 (the top, middle and bottom clusters in the panels, respectively) and supersaturation $s = 0.9$, 1.5 and 2.5 (the left, middle and right columns of panels, respectively).

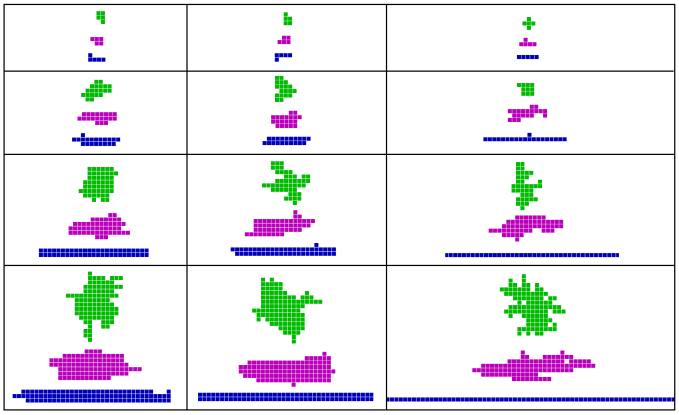

Figure 6.4: Successive snapshots (top view) of 2D Kossel-Stranski crystalline clusters with anisotropy ratio $\xi = 1$, 3 and 8 (the top, the middle and bottom clusters in the panel respectively) which evolve at supersaturation $s = 0.9$, 1,5 and 2.5 (the left, middle and right columns of panels, respectively) The squares represent molecules, and the cluster lengthening is along the m axis.

From Fig. 6.4 it can be seen that at a given s value the cluster grows more elongated when ξ is greater. This effect of the anisotropic interactions on the cluster growth shape was reported also in Ref. (163) and could be anticipated on thermodynamic grounds. The Gibbs-Curie-Wulff theorem for the equilibrium shape of the 2D Kossel-Stranski crystals considered in the previous chapters, in fact, states that the length-to-width ratio of a rectangular crystal is equal to the ratio E/E_h, i.e. to ξ (72; 81; 163). When the anisotropy ratio ξ is high enough, however, a remarkable effect of the supersaturation s on the cluster shape is revealed in Fig. 6.4. Indeed, while at $\xi = 1$ and 3 (isotropic and weakly anisotropic interactions, respectively) the cluster shape is virtually the same at $s = 0.9$, 1.5 and 2.5; at $\xi = 8$ (strongly anisotropic interactions) the clusters mainly elongate (see those at the bottom of the figure panels), but with a different number of rows at the different supersaturations: three rows at $s = 0.9$, two rows at $s = 1.5$, and one row at $s = 2.5$. The roughening of the cluster periphery

at $\xi = 1$ and $\xi = 3$, is a result of the rather small value of the bonding energy along the i axis (as already mentioned, the KMC simulations were performed at $\psi_h = 1$). Increasing ξ suppresses the roughening effect due to the smallness of ψ_h so that the cluster periphery becomes smooth at the molecular scale.

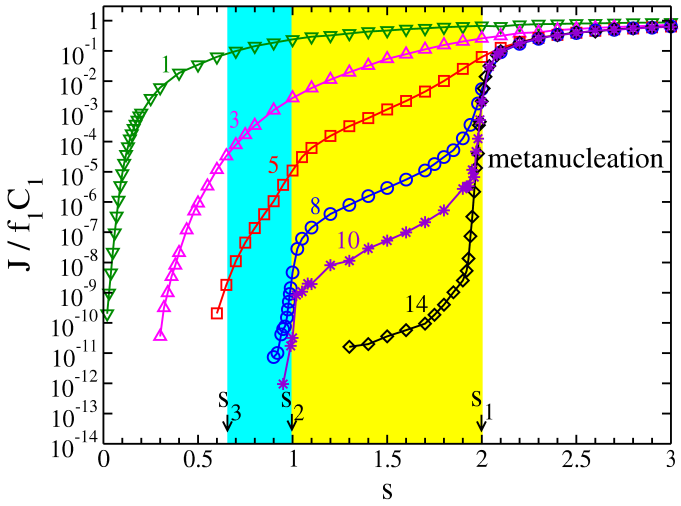

Figure 6.5: Supersaturation dependence of the nucleation rate J at $\xi = 1$, 3, 5, 8, 10 and 14 (as indicated). The threshold supersaturations s_1, s_2 and s_3 are indicated by the arrows, the symbols represent the simulations data, and the lines are only guides to the eye.

The symbols in Fig. 6.5 represent the simulation data for the dependence of P_2 or, equivalently, of J/f_1C_1 on s at $\xi = 1$, 3, 5, 8, 10 and 14 (as indicated). As seen, for ratios $\xi = 1$ and $\xi = 3$ a concave increase of the logarithm of J/f_1C_1 with s is observed, the J/f_1C_1 values spanning up to approximately ten orders of magnitude. This is in agreement with the $J(s)$ dependence reported previously (68; 141; 142) for this system with isotropic ($\xi = 1$) or weakly anisotropic ($\xi = 3$) interactions and, as already mentioned, with the nucleation theory prediction that $\ln J$ increases concavely with s. In contrast, the simulation dependence of the logarithm of J/f_1C_1 on s for anisotropy ratios $\xi = 8$, 10 and 14 is nearly step-like, with jumps that are more pronounced at greater interaction anisotropy. This is seen most clearly at $\xi = 14$, when J/f_1C_1 has a spectacularly sharp jump of approximately seven orders of magnitude. The obtained step-like dependence of $\ln J$ on s is surprising, because it has never been observed in experiments and computer simulations nor predicted theoretically.

In Fig. 6.6 the simulation data for $\xi = 10$ have been compared with the theoretical predictions from CNT and the atomistic nucleation theory (ANT). As the stark discrepancy between simulation and theory persists also for $\xi = 5$,

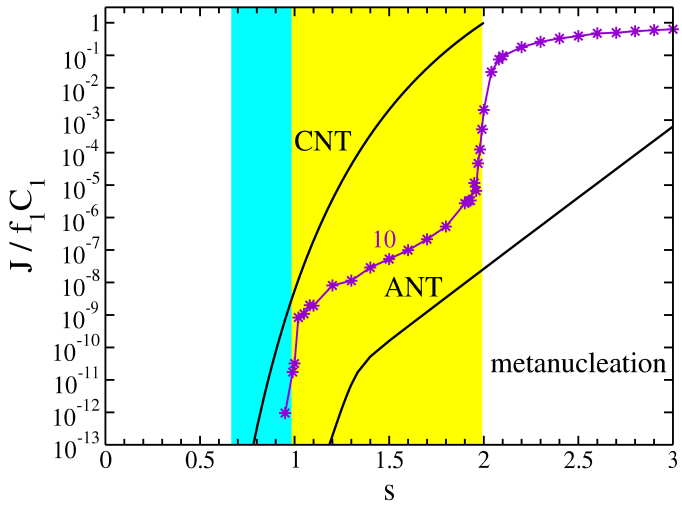

Figure 6.6: Dependence of the nucleation rate J on the supersaturation s at $\xi = 10$: symbols and line 10 represent the simulations data; lines CNT and ANT represent the predictions of CNT (Eq. (2.26)) and ANT (Eq. (3.17)), respectively.

8 and 14 (result not shown), the comparison shows that CNT and ANT are not able to capture the step-like dependence of $\ln J$ on s at high degrees of interaction anisotropy.

To understand the physical reason for the existence of jumps in the s dependence of $\ln J$, crystal nucleation in solutions can be considered. In this case the equilibrium concentration C_e is the crystal solubility, i.e. the solute concentration at which the solution is in thermodynamic equilibrium with a macroscopically large crystal. Theoretical considerations (81) and computer simulations (3; 7) reveal that the solubility of a rectangular crystal (or crystal-like formation) strongly depends on the crystal width, i.e. on the number i of molecular rows that constitute the 2D crystal. At a given temperature T, the equilibrium concentration $C_{e,i}$ at which such a crystal of fixed number i of rows ($i = 1, 2, 3,...$) neither grows nor dissolves is given by (81)

$$C_{e,i} = C_e \exp(2\psi_h/i). \tag{6.9}$$

This expression says that, at a constant T, $C_{e,i}$ decreases with increasing i (i.e. $C_{e,1} > C_{e,2} > C_{e,3} > ...$) and equals C_e in the limit of $i \to \infty$. The threshold supersaturation s_i, defining the lowest s value down to which a 2D crystal of fixed number i of rows can lengthen irreversibly, is readily obtained from Eq. (6.9) by using the formula $s = \ln(C_1/C_e)$. The result is (81):

$$s_i = 2\psi_h/i. \tag{6.10}$$

In the present simulations $\psi_h = 1$ and this formula yields $s_1 = 2$, $s_2 = 1$ and $s_3 = 2/3$. As seen in Fig. 6.5, the first and the second jumps of the nucleation rate are precisely at these s_1 and s_2 values. This can be rationalised by recalling that for supersaturations $s > s_1$ any monomer acts as a nucleus (80; 81) so that even a monomolecular row (i.e. a 2D crystal with the minimum fixed width) can lengthen irreversibly. Then the 2D crystals form in the so-called metanucleation regime (80; 81) which corresponds to the regime of crystal formation by spinodal decomposition. Similarly, when $s > s_2$, the dimers oriented along the i axis act as nuclei of 2D crystals with fixed width of two rows and these crystals are able to irreversible elongate along the m axis (81). In the same way, the 2D crystals of any length and fixed width of three rows can elongate irreversibly when $s > s_3$. Thus, the existence of the threshold supersaturations $s_1 = 2$, $s_2 = 1$ and $s_3 = 2/3$ explains why in Fig. 6.4 the clusters with anisotropy ratio $\xi = 8$ (those at the bottom of the figure panels) lengthen with one row at $s = 2.5$, two rows at $s = 1.5$ and three rows at $s = 0.9$. What is particularly important is that when the nucleation/metanucleation border $s_1 = 2\psi_h$ is crossed by decreasing s, the nucleus changes from a monomer to a single-row nanocrystal plus one monomer on top of it (81). The attachment of this monomer creates two kink sites, each of energy 2ψ, and this leads to a drop in the nucleation rate when ψ is large enough. This explanation is consistent with the observation (see Fig. 6.5) that the drop in J becomes more pronounced with increasing ψ, i.e. ξ, because then the kink-site creation energy is higher. Similarly, when the threshold supersaturation $s_2 = \psi_h$ is crossed by further decreasing s, the nucleus changes from a single-row nanocrystal plus one monomer on top of it, to a two-row nanocrystal plus one monomer on top of it (81). This leads to a second drop in J when ψ, and hence ξ, is sufficiently great (see Fig. 6.5). It is worth noting that at $\xi = 14$ the nucleation rate in the supersaturation range $s_2 < s < s_1$ (the yellow area

in Fig. 6.5) can be more than ten orders of magnitude lower than that in the supersaturation range $s > s_1$ (white area in Fig. 6.5)).

Another reason for which the obtained step-like dependence of the logarithm of J on s is remarkable is that if this dependence is analysed with the help of the nucleation theorem (77; 78; 83), for certain s values it would yield a nucleus size that according to the Eq. (2.31) increases with increasing supersaturation. Indeed, due to this theorem, for nucleation of condensed phases the relation $n^* \approx d(\ln J)/ds - 1$ (Eq. (2.32)) holds at fixed T (77; 78; 83) and hence the slope of the $\ln J(s)$ line gives directly the nucleus size n^*. Thus, with reference to the nucleation theorem, the step-like course of the simulation $J(s)$ dependence would lead to the conclusion that for s tending from below to s_1, s_2, etc., n^* increases with increasing s. The increasing n^* with increasing s is paradoxical if viewed within the scope of CNT and ANT, but it is not when it is realised that it results from applying the nucleation theorem to a process that is not really nucleation when ξ is large enough. Indeed, then the process is more like a highly anisotropic spinodal decomposition in which, depending on the s range, the 2D crystals form at a fixed width of one or two or three, etc. molecular rows (see the bottom clusters in the panels of Fig. 6.4).

The unexpected step-like behaviour of the nucleation rate as a function of supersaturation s, observed at large values of the anisotropy ψ/ψ_h, prompted us to search a sequence of shapes that, once used in the ANT theoretical equation could generate a theoretical line closer to the simulation data. No sequence is really able to reproduce the jumps of $J(s)$. However a sequence of shapes identified by the minimum surface energy turns out to give an ANT curve closer to the simulation data. This result is illustrated in Fig. 6.7, where the red line represents the predictions of equation (3.17) using the new cluster sequence to compute each term.

The sequence of shapes used to draw the red curve is not compatible with the Szilard-Farkas model because, with increasing size n, it presents abrupt changes of cluster shape. For example, the cluster of size $n = 21$ peptides presents a 1-row configuration, while the cluster of $n = 22$ peptides is in a 2-row configuration with 11 peptide in the first row, and 11 in the second row.

Further investigations, currently under way, have been planned to clarify these issues, and have already produced interesting preliminary results, showing

a closer agreement between theoretical and simulation curves. The results will be published soon in a paper now in preparation.

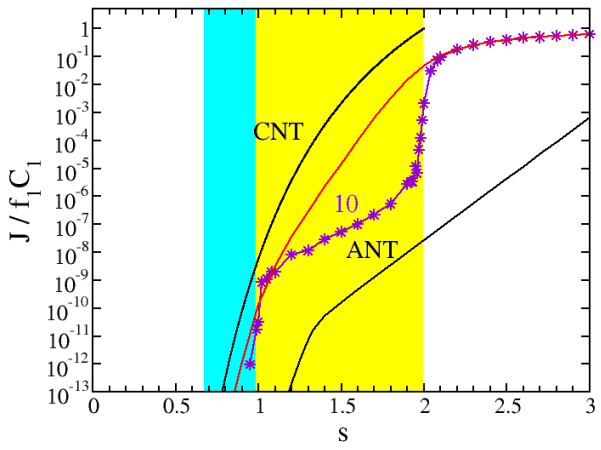

Figure 6.7: Dependence of the nucleation rate J on the supersaturation s at $\xi = 10$: symbols and line 10 represent the simulations data; lines CNT and ANT represent the predictions of CNT (Eq. (2.26)) and ANT (Eq. (3.17)) while the red curve has be drawn with the help of Eq. (3.17) using a sequence of shape not compatible with the Szilard-Farkas model.

6.5 Conclusion

In this chapter the nucleation of a 2D crystal from monomers on its own substrate has been investigated by kinetic Monte Carlo simulations. A brief introduction about the method has been given in section 6.2, then the method has been applied to the case of isotropic inter-monomer interactions within the crystal, and in 6.3 section the method has been applied to the nucleation of a 2D crystal characterised by anisotropic interactions between its monomers.

In particular the effect of the interactions on the nucleation rate J has been investigated and compared with the theoretical ANT and CNT predictions for the isotropic case. For smaller ω values, both CCNT and ANT predict the smooth and monotonic increase of $J(s)$ with increasing supersaturation s given by the simulation data but underestimate the nucleation rate. For higher ω values, instead, ANT reproduces better the features of the J curves from the KMC data. The results shows that ANT reliably predicts the simulation data and the agreement improves at higher ω values. Furthermore, ANT can reproduce the broken linearity of $\ln J(s)$, but exaggerates the bends that are sharp instead smooth as in the simulations data. The underestimation of the KMC data for smaller ω values by ANT can be interpreted with the fact that the smaller the interaction value is, the rougher the cluster periphery will be. The resulting variety of shapes is considered in the KMC data, since the statistical character of KMC accounts for fluctuations in shape of the growing clusters. ANT, instead, is based on a fixed growth sequence that associate to each n-sized cluster only one shape.

The effect of the anisotropic degree on the nucleation rate and on the morphology of the cluster has been discussed in the second part of the chapter. The $J(s)$ dependence obtained in the present study of nucleation of 2D crystals with anisotropic molecular interactions is highly unusual. As the anisotropy of the interactions between molecules increases, there is a transition in the dimensionality of the crystals. Whereas at isotropic or weakly anisotropic interactions the smallest nanocrystals grow in two dimensions by simultaneous lengthening and widening, at strongly anisotropic interactions they evolve virtually in one dimension. More precisely, smallest nanocrystals, characterised by strongly anisotropic interactions within their monomers, mainly lengthen by preserving a fixed width of one or two or three, etc. molecular rows depending on the supersaturation.

As a result, the $J(s)$ dependence of 2D crystals with high enough interaction-anisotropy ratio changes dramatically: $\ln J$ becomes a step-like function of s. This changed dependence can be understood in the CNT-like framework outlined elsewhere (81) and used for obtaining Eq. (6.10), but cannot be described by CNT (81) and ANT (18). A formula that predicts jumps of $\ln J(s)$ at the threshold supersaturations s_i from Eq. (6.10) when the nucleating crystals or crystal-like phases are constituted of molecules with strongly anisotropic interactions is yet to be derived and verified, and the present study provides the necessary data and understanding for doing that.

Although the results here pertain to nucleation of 2D crystals on their own substrate, upon replacing s by an effective supersaturation (see, e.g., Refs. (45; 80)) they are directly applicable to 2D crystals nucleated on a foreign substrate. More generally, the results also apply to nucleation of phases structured in a 2D-like lattice and constituted of molecules with strongly anisotropic interactions. A prominent example in this respect (22) is the nucleation of amyloid fibrils. These protein aggregates are of great interest because of their involvement in various amyloid-related diseases (22) and nanotechnological applications (89). In chapter 4, the nucleation of amyloid-β_{40} fibrils has been studied, and it can be said here that this process corresponds to the nucleation of 2D crystals with anisotropy ratio $\xi = 14$. Thus, the application of our high-anisotropy-ratio results to amyloid fibril nucleation occurring by direct polymerization of monomeric protein is straightforward. It is worth noting as well that the present results are directly applicable also to the 2D square-lattice Ising model with anisotropic nearest-neighbour coupling, provided the broken-bond energies $(\psi/2)k_BT$ and $(\psi_h/2)k_BT$ are replaced by the coupling constants J_x and J_y along the x and y axes, respectively, and the halved supersaturation $(s/2)k_BT$ is replaced by the external field H.

Chapter 7

Summary and Outlook

7.1 Summary and Conclusions

The self-assembly of amyloid fibrils is a widely studied phenomenon because it has important implications ranging from biotechnology to human diseases (Ref. (22)). Amyloid fibrils deposition is in fact associated to many systemic and neurodegenerative diseases. On the other hand the highly ordered structure of this supra-molecular material makes it unique for its mechanical and biological properties (Ref. (22)). Novel materials using amyloid fibers as templates can find applications in food packaging and wrapping, as bio-sensors or in the bio-scaffolding industry, aiming to exploit the biodegradability, strength and bio-compatibility of the amyloid fibers (Ref. (22)). Interestingly living beings such as moths and spiders, synthesise amyloid fibers to build their nest or webs and to protect their shells (Ref. (36)). Fungi and bacteria (Escherichia coli) employ extracellular amyloid (curli) to create a matrix that modulate surface adhesion of biological environments in order to support the colony formation (Ref. (36)). Furthermore amyloid fibrils have functional roles in the human body (Ref. (11; 36)). For instance, a study has shown that the protein Pme17 forms fibrous structures (Mα amyloid) vital for the bio-synthesis of melanin (Ref. (11; 36)).

It is well established that most fibrillar aggregates do form through a nucleation mechanism. Over the last decades there has been a considerable effort in understanding the nucleation mechanism associated with the protein fibrillation, and numerous experiments and theoretical studies have been devoted to the investigation of the self-assembly of protein (Ref. (4; 32; 33; 34; 38; 60; 61; 67; 81; 90; 100)). However, the nucleation of nanofibrils is poorly understood, mainly

because the experiments characterising the early stages of these aggregates are difficult, and furthermore, the complexity of protein systems has made it difficult to simulate the fibril nucleation events.

Understanding and interpreting such experiments requires developing a theoretical framework to describe amyloid fibril nucleation. The aim of this thesis has been to illustrate how the amyloid fibril nucleation can be treated in the framework of existing nucleation theories of nucleation of new phases such as the Classical and Atomistic Nucleation Theory (CNT and ANT respectively). The theoretical formalism developed here (ANT for amyloid fibril nucleation) stems from first principles in nucleation theory and it has not only the potential to become an important tool in analysing experiments in amyloid fibrils but also in designing new experiments aimed at a better understanding of how to control the birth of amyloid fibrils. The results discussed in the thesis are applicable to the one-step homogeneous nucleation which occurs when the solution is sufficiently pure and/or strongly supersaturated.

In chapter 1 a brief introduction illustrated the self-assembly of proteins into amyloid fibrils, giving the experimental and theoretical state of art. In addition to the motivation of this thesis, concepts of nucleation theory have been introduced and the reasons which justify the use of overall crystallization theory have been given.

In chapter 2 CNT (77; 84) and CCNT (80) were applied for describing the thermodynamics and kinetics of amyloid fibril nucleation. In particular, the work to form a nanosized amyloid fibril built up of successively layered β-sheets, the size of the fibril nucleus, the nucleation work and the fibril nucleation rate have been determined as explicit functions of the concentration and temperature of the protein solution. It has been found that although CNT could give a good qualitative explanation of the process, it disregards the work done to attach the first β-strand at the periphery of the CNT fibril nucleus. Because of that, CNT underestimates the fibril nucleus size and, even more, the nucleation work leading to an overestimation of the nucleation rate J by many orders of magnitude. CCNT corrects this important inaccuracy giving a more reliable expression for the calculation of J. Importantly it has been discussed that the CCNT nucleation rate can be related to experimental isothermal data such as the protein solubility. Finally, an important theory-independent estimation of the nucleus size from J data has been given.

In chapter 3 the nucleation of amyloid fibrils has been discussed at the molecular level employing the ANT for fibril. In particular, an expression for the work to form a nanosized amyloid fibril (protofilament) composed of successively layered β-sheets has been derived. With the help of this expression, the nucleation work and the nucleus size have been obtained. Subsequently, ANT has been applied to an idealised peptide system. Also in this case, the size of the fibril nucleus, fibril nucleation work and fibril nucleation rate as functions of supersaturation of the protein solution have been obtained. The analysis illustrates the unique feature of the ANT fibril nucleus size as a constant integer in a given supersaturation range. Furthermore, the comparison between ANT, CNT and CCNT leads to the conclusion that while the CNT nucleation rate is orders of magnitude greater than the ANT one, the CCNT and ANT nucleation rates are in very good quantitative agreement with each other. However, ANT is generally considered more reliable than the previous theories because of its realistic accounting for the actual shape of the clusters. In the field of crystal nucleation, in fact, ANT has been found to be in better quantitative agreement with simulation and numerical data than the CNT and CCNT.

The ANT developed for amyloid fibrils has been applied to a real protein model in chapter 4. In particular, employing ANT to model the amyloid nucleation of β_2m and Aβ_{40} fibrils, predictions of the fibril nucleus size and of the fibril nucleation rate have been determined as functions of the supersaturation and the concentration of the protein solution. Subsequently, these ANT results have been correlated to recent time-resolved optical experiments, where the effects of the protein concentration and mutations on the initial lag time in the protein solution have been measured. These important results lead to the conclusion that the primary effect of the predicted changes in the nucleation rates and nucleus sizes of mutations, is due to the change of the fibril solubility. A correlation analysis between the ANT nucleation rates and the experimentally measured lag times yields a correlation coefficient of 0.43 for β_2m and 0.71 for Aβ_{40}, providing a statistically relevant validation of the present model.

Employing ANT, a general expression for the stationary size distribution of amyloid nanofibrils constituted of successively layered β-sheets has been derived in chapter 5. This expression applied to Aβ_{40} fibrils allowed the prediction of the nanofibril size distribution as a function of the protein concentration and temperature. The most remarkable feature of the distribution is its exhibiting a

series of peaks which are positioned at "magic" nanofibril sizes (or lengths) that are consistent with experimental results for the size distribution of aggregates in solutions of Aβ_{40} proteins. The approach makes it possible to gain insight into the effect of point mutations on the nanofibril size distribution, an effect that might explain the substantial differences experimentally observed in the fibrillation lag-time of wild-type and point-mutated amyloid-β proteins.

In chapter 6, some assumptions of ANT have been tested by kinetic Monte Carlo simulations (KMC). The idealised model in the simulation makes it possible to apply the results to any nucleating system whose molecules interact with strongly anisotropic forces, including the case of amyloid fibrils nucleating from a peptide solution. In particular, the focus of the investigation has been to study the effect of nucleation of isotropic and anisotropic interactions on the stationary nucleation rate J. To this aim, by KMC simulations, J as a function of the supersaturation s has been obtained for a variety of different values of the anisotropic parameter $\xi = \psi/\psi_h$. The results show that with increasing ξ the dependence of $\ln J$ on s becomes step-like, with jumps at certain s values. This $J(s)$ dependence observed in the simulations cannot be described by the classical and atomistic nucleation theories. It represents, in fact, a completely novel result, never reported previously in either experiments or simulations studies.

7.2 Outlook

The developed theory is able to explain the kinetics and thermodynamics of nucleation of amyloid fibrils as one-step process for homogeneous systems, opening the way to new exciting directions. Furthermore, the novel results obtained by KMC simulations extend the study toward a more fundamental problem such as the nucleation of clusters composed of atoms or molecules bonded by anisotropic interactions.

An important extension of the present study would be the inclusion of a specific hydrophobic interaction for each amino acid. In fact, the hydrophobic interaction among proteins is known to play an important role in amyloid fibril formation. However, in the calculations of chapter 4, chapter 5 and in the simulations described in chapter 6, this contribution has been neglected, assuming the same hydrophobicity for each amino acid. The most detailed interaction model that has been used, i.e., the PASTA model, has been derived in Ref.

(144) upon analysing the structure of protein's crystals at negligible or no hydration, and thus, even in this case, hydrophobicity was neglected. Preliminary calculations, not included in the thesis for reasons of length and clarity, have shown that prediction of equilibrium concentrations improves considerably when specific hydrophobic interactions are included in the model.

In the present theory, the model assumes nearest-neighbour interactions among the molecules in the crystal. This simplifying assumption might represent another important limit of the model, possibly affecting the quality of the results. In chapter 4, for instance, the application of the theory to real protein reproduced the effect of some mutations but not of all of them. In the model the effect of replacing amino acid α with amino acid γ within a protein or a peptide does not depend on the position of replacement along the chain, but depends on the identity of the two amino acids α and γ. That means that the nucleation rate for two different mutations has to be the same if both correspond to the change of the same amino acid type, occurring at two different positions along the chain. This fact does not correspond to the experimental reality. The introduction of second neighbour interactions could easily remove this limitation, and improve the results of chapter 4 concerning the nucleation rate of mutants with respect to the wild case.

A further important improvement of the present study would require upgrading the actual model to a more realistic description of polymer peptide system in a biological environment. For example, the charge and polarity of the amino acids have been neglected in the model. Moreover, proteins in biological fluids are influenced by the state of motion of the environment, giving rise to the so called *hydrodynamic* interactions. For example, the fluid's shear can trigger the self-assembly or prevent it depending on the physical condition. Shear, in particular, could contribute significantly to the fragmentation processes that represent an important mechanism of amyloid fibril nucleation.

As a further important aspect, it is worth pointing out that, in its present version, the model does not explicitly include entropic effects such as the loss of entropy occurring when a β-strand is attached to a fibril or the entropy due to vibrations of a β-strand within the fibril. Entropic effects can be taken into account by providing expressions for the entropic contribution to the fibril solubility and β-strand binding energy. Ideas on how to estimate these contributions can be found in Ref. (31) and they could be the basis of an important extension

of the present nanofibril model. It should be noted that the entropic effects on the fibril solubility are automatically accounted for when experimental data for C_e are used in the ANT Eq. (3.19) for J.

Another stimulating new direction could to be the study of the polymorphism phenomenon. Amyloid fibrils from a single polypeptide chain can form polymorphs that can self-propagate in seeded elongation reactions. In particular, the choice of the seed can control the conformation of the final fibril. Furthermore, sequence differences between the seed and monomer can prevent the elongation if the sequences diversity is conspicuous. This phenomenon is known to happen in the context of prion related disease.

It is worth noting as well that the ANT theory requires appropriate modifications to deal with different relevant situations such as: (i) the two-step nucleation case involving formation of oligomeric precursors or (ii) the heterogeneous nucleation on nucleation-active sites provided by foreign agents such as nanoparticles. The process of heterogeneous nucleation can be reformulated in terms of the problem of seeded nucleation which is well related to a real aspect in the self-assembling of protein into amyloid fibrils. It is in fact well known that in the biological systems the aggregation into amyloid fibrils can be triggered by the presence of nanoparticles and lipid bilayers of membranes. The theory could be extended to include in the model and in the actual "effective supersaturation" the presence of a foreign seed or of a seed made from the same protein. The process could be then verified by KMC simulations.

Finally, the computer program written for this thesis, could be easily extended to study the nucleation of crystals in a three-dimensional (3D) space, giving important results applicable to 3D crystal and anisotropic systems in general.

Appendix A

Tables for β_2m, Aβ_{40} and their mutants

The following tables show the ANT numbers l_j^* and $l_{h,j}^*$ of, respectively, the strong and weak broken bonds of an n-sized amyloid fibril at the periphery of the fibril cross section in the m, i plane for:

β_2m and its mutants ($\psi/\psi_h = n_t = 12$) A.1;
Aβ_{40} and V18N ($\psi/\psi_h = n_t = 14$) A.2;
V18R and V18P ($\psi/\psi_h = n_t = 13$) A.3.

Table A.1: ANT l and l_h numbers for β_2**m and mutants**.

n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$
1	2	2	61	6	48	121	8	72	181	8	92	241	10	98
2	2	4	62	6	48	122	8	72	182	8	92	242	10	98
3	2	6	63	6	48	123	8	72	183	8	92	243	10	98
4	2	8	64	6	48	124	8	72	184	8	92	244	10	98
5	2	10	65	6	48	125	8	72	185	8	94	245	10	98
6	2	12	66	6	48	126	8	72	186	8	94	246	10	100
7	2	14	67	6	48	127	8	72	187	8	94	247	10	100
8	2	16	68	6	48	128	8	72	188	8	94	248	10	100
9	2	18	69	6	48	129	8	72	189	8	96	249	10	100
10	2	20	70	6	48	130	8	72	190	8	96	250	10	100
11	2	22	71	6	48	131	8	72	191	8	96	251	12	102
12	2	24	72	6	48	132	8	72	192	8	96	252	12	102
13	4	24	73	6	50	133	8	72	193	10	96	253	12	102
14	4	24	74	6	50	134	8	72	194	10	96	254	12	102
15	4	24	75	6	50	135	8	72	195	10	96	255	12	102
16	4	24	76	6	52	136	8	72	196	10	96	256	12	104
17	4	24	77	6	52	137	8	72	197	10	96	257	12	104
18	4	24	78	6	52	138	8	72	198	10	96	258	12	104
19	4	24	59	6	54	139	8	72	199	10	96	259	12	104
20	4	24	80	6	54	140	8	72	200	10	96	260	12	104
21	4	24	81	6	54	141	8	72	201	10	96	261	12	106
22	4	24	82	6	56	142	8	72	202	10	96	262	12	106
23	4	24	83	6	56	143	8	72	203	10	96	263	12	106
24	4	24	84	6	56	144	8	72	204	10	96	264	12	106
25	4	26	85	6	58	145	8	74	205	10	96	265	12	106
26	4	26	86	6	58	146	8	74	206	10	96	266	12	108
27	4	28	87	6	58	147	8	74	207	10	96	267	12	108
28	4	28	88	6	60	148	8	74	208	10	96	268	12	108
29	4	30	89	6	60	149	8	76	209	10	96	269	12	108
30	4	30	90	6	60	150	8	76	210	10	96	270	12	108
31	4	32	91	6	62	151	8	76	211	10	96	271	12	110
32	4	32	92	6	62	152	8	76	212	10	96	272	12	110
33	4	34	93	6	62	153	8	78	213	10	96	273	12	110
34	4	34	94	6	64	154	8	78	214	10	96	274	12	110
35	4	36	95	6	64	155	8	78	215	10	96	275	12	110
36	4	36	96	6	64	156	8	78	216	10	96	276	12	112
37	4	38	97	6	66	157	8	80	217	10	96	277	10	112
38	4	38	98	6	66	158	8	80	218	10	96	278	10	112
39	4	40	99	6	66	159	8	80	219	10	96	279	10	112
40	4	40	100	6	68	160	8	80	220	10	96	280	10	112
41	4	42	101	6	68	161	8	82	221	10	96	281	10	114
42	4	42	102	6	68	162	8	82	222	10	96	282	10	114
43	4	44	103	6	70	163	8	82	223	10	96	283	10	114
44	4	44	104	6	70	164	8	82	224	10	96	284	10	114
45	4	46	105	6	70	165	8	84	225	10	96	285	10	114
46	4	46	106	6	72	166	8	84	226	10	96	286	10	116
47	4	48	107	6	72	167	8	84	227	10	96	287	10	116
48	4	48	108	6	72	168	8	84	228	10	96	288	10	116
49	6	48	109	8	72	169	8	86	229	10	96	289	10	116
50	6	48	110	8	72	170	8	86	230	10	96	290	10	116
51	6	48	111	8	72	171	8	86	231	10	96	291	10	118
52	6	48	112	8	72	172	8	86	232	10	96	292	10	118
53	6	48	113	8	72	173	8	88	233	10	96	293	10	118
54	6	48	114	8	72	174	8	88	234	10	96	294	10	118
55	6	48	115	8	72	175	8	88	235	10	96	295	10	118
56	6	48	116	8	72	176	8	88	236	10	96	296	10	120
57	6	48	117	8	72	177	8	90	237	10	96	297	10	120
58	6	48	118	8	72	178	8	90	238	10	96	298	10	120
59	6	48	119	8	72	179	8	90	239	10	96	299	10	120
60	6	48	120	8	72	180	8	90	240	10	96	300	10	120

Table A.2: ANT l and l_h numbers for **Aβ_{40} wild-type and V18N mutant**.

n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$
1	2	2	61	6	56	121	6	82	181	8	92	241	10	112
2	2	4	62	6	56	122	6	82	182	8	92	242	10	112
3	2	6	63	6	56	123	6	82	183	8	92	243	10	112
4	2	8	64	6	56	124	6	84	184	8	92	244	10	112
5	2	10	65	6	56	125	6	84	185	8	94	245	10	112
6	2	12	66	6	56	126	6	84	186	8	94	246	10	112
7	2	14	67	6	56	127	8	84	187	8	94	247	10	112
8	2	16	68	6	56	128	8	84	188	8	94	248	10	112
9	2	18	69	6	56	129	8	84	189	8	96	249	10	112
10	2	20	70	6	56	130	8	84	190	8	96	250	10	112
11	2	22	71	6	56	131	8	84	191	8	96	251	10	112
12	2	24	72	6	56	132	8	84	192	8	96	252	10	112
13	2	26	73	6	56	133	8	84	193	8	98	253	10	112
14	2	28	74	6	56	134	8	84	194	8	98	254	10	112
15	4	28	75	6	56	135	8	84	195	8	98	255	10	112
16	4	28	76	6	56	136	8	84	196	8	98	256	10	112
17	4	28	77	6	56	137	8	84	197	8	100	257	10	112
18	4	28	78	6	56	138	8	84	198	8	100	258	10	112
19	4	28	59	6	56	139	8	84	199	8	100	259	10	112
20	4	28	80	6	56	140	8	84	200	8	100	260	10	112
21	4	28	81	6	56	141	8	84	201	8	102	261	10	112
22	4	28	82	6	56	142	8	84	202	8	102	262	10	112
23	4	28	83	6	56	143	8	84	203	8	102	263	10	112
24	4	28	84	6	56	144	8	84	204	8	102	264	10	112
25	4	28	85	6	58	145	8	84	205	8	104	265	10	112
26	4	28	86	6	58	146	8	84	206	8	104	266	10	112
27	4	28	87	6	58	147	8	84	207	8	104	267	10	112
28	4	28	88	6	60	148	8	84	208	8	104	268	10	112
29	4	30	89	6	60	149	8	84	209	8	106	269	10	112
30	4	30	90	6	60	150	8	84	210	8	106	270	10	112
31	4	32	91	6	62	151	8	84	211	8	106	271	10	112
32	4	32	92	6	62	152	8	84	212	8	106	272	10	112
33	4	34	93	6	62	153	8	84	213	8	108	273	10	112
34	4	34	94	6	64	154	8	84	214	8	108	274	10	112
35	4	36	95	6	64	155	8	84	215	8	108	275	10	112
36	4	36	96	6	64	156	8	84	216	8	110	276	10	112
37	4	38	97	6	66	157	8	84	217	8	110	277	10	112
38	4	38	98	6	66	158	8	84	218	8	110	278	10	112
39	4	40	99	6	66	159	8	84	219	8	110	279	10	112
40	4	40	100	6	68	160	8	84	220	8	110	280	10	112
41	4	42	101	6	68	161	8	84	221	8	112	281	10	114
42	4	42	102	6	68	162	8	84	222	8	112	282	10	114
43	4	44	103	6	70	163	8	84	223	8	112	283	10	114
44	4	44	104	6	70	164	8	84	224	8	112	284	10	114
45	4	46	105	6	70	165	8	84	225	10	112	285	10	114
46	4	46	106	6	72	166	8	84	226	10	112	286	10	116
47	4	48	107	6	72	167	8	84	227	10	112	287	10	116
48	4	48	108	6	72	168	8	84	228	10	112	288	10	116
49	4	50	109	6	74	169	8	86	229	10	112	289	10	116
50	4	50	110	6	74	170	8	86	230	10	112	290	10	116
51	4	52	111	6	74	171	8	86	231	10	112	291	10	118
52	4	52	112	6	76	172	8	86	232	10	112	292	10	118
53	4	54	113	6	76	173	8	88	233	10	112	293	10	118
54	4	54	114	6	76	174	8	88	234	10	112	294	10	118
55	4	56	115	6	78	175	8	88	235	10	112	295	10	118
56	4	56	116	6	78	176	8	88	236	10	112	296	10	120
57	6	56	117	6	78	177	8	90	237	10	112	297	10	120
58	6	56	118	6	80	178	8	90	238	10	112	298	10	120
59	6	56	119	6	80	179	8	90	239	10	112	299	10	120
60	6	56	120	6	80	180	8	90	240	10	112	300	10	120

Table A.3: ANT l and l_h numbers for **V18R and V18P mutants of A**β_{40}.

n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$	n	l_n	$l_{h,n}$
1	2	2	61	6	52	121	6	78	181	8	92	241	10	104
2	2	4	62	6	52	122	6	78	182	8	92	242	10	104
3	2	6	63	6	52	123	6	78	183	8	92	243	10	104
4	2	8	64	6	52	124	6	78	184	8	92	244	10	104
5	2	10	65	6	52	125	6	78	185	8	94	245	10	104
6	2	12	66	6	52	126	6	78	186	8	94	246	10	104
7	2	14	67	6	52	127	8	78	187	8	94	247	10	104
8	2	16	68	6	52	128	8	78	188	8	94	248	10	104
9	2	18	69	6	52	129	8	78	189	8	96	249	10	104
10	2	20	70	6	52	130	8	78	190	8	96	250	10	104
11	2	22	71	6	52	131	8	78	191	8	96	251	10	104
12	2	24	72	6	52	132	8	78	192	8	96	252	10	104
13	2	26	73	6	52	133	8	78	193	8	98	253	10	104
14	4	26	74	6	52	134	8	78	194	8	98	254	10	104
15	4	26	75	6	52	135	8	78	195	8	98	255	10	104
16	4	26	76	6	52	136	8	78	196	8	98	256	10	104
17	4	26	77	6	52	137	8	78	197	8	100	257	10	104
18	4	26	78	6	52	138	8	78	198	8	100	258	10	104
19	4	26	59	6	54	139	8	78	199	8	100	259	10	104
20	4	26	80	6	54	140	8	78	200	8	100	260	10	104
21	4	26	81	6	54	141	8	78	201	8	102	261	10	106
22	4	26	82	6	56	142	8	78	202	8	102	262	10	106
23	4	26	83	6	56	143	8	78	203	8	102	263	10	106
24	4	26	84	6	56	144	8	78	204	8	102	264	10	106
25	4	26	85	6	58	145	8	78	205	8	104	265	10	106
26	4	26	86	6	58	146	8	78	206	8	104	266	10	108
27	4	28	87	6	58	147	8	78	207	8	104	267	10	108
28	4	28	88	6	60	148	8	78	208	8	104	268	10	108
29	4	30	89	6	60	149	8	78	209	10	104	269	10	108
30	4	30	90	6	60	150	8	78	210	10	104	270	10	108
31	4	32	91	6	62	151	8	78	211	10	104	271	10	110
32	4	32	92	6	62	152	8	78	212	10	104	272	10	110
33	4	34	93	6	62	153	8	78	213	10	104	273	10	110
34	4	34	94	6	64	154	8	78	214	10	104	274	10	110
35	4	36	95	6	64	155	8	78	215	10	104	275	10	110
36	4	36	96	6	64	156	8	78	216	10	104	276	10	112
37	4	38	97	6	66	157	8	80	217	10	104	277	10	112
38	4	38	98	6	66	158	8	80	218	10	104	278	10	112
39	4	40	99	6	66	159	8	80	219	10	104	279	10	112
40	4	40	100	6	68	160	8	80	220	10	104	280	10	112
41	4	42	101	6	68	161	8	82	221	10	104	281	10	114
42	4	42	102	6	68	162	8	82	222	10	104	282	10	114
43	4	44	103	6	70	163	8	82	223	10	104	283	10	114
44	4	44	104	6	70	164	8	82	224	10	104	284	10	114
45	4	46	105	6	70	165	8	84	225	10	104	285	10	114
46	4	46	106	6	72	166	8	84	226	10	104	286	10	116
47	4	48	107	6	72	167	8	84	227	10	104	287	10	116
48	4	48	108	6	72	168	8	84	228	10	104	288	10	116
49	4	50	109	6	74	169	8	86	229	10	104	289	10	116
50	4	50	110	6	74	170	8	86	230	10	104	290	10	116
51	4	52	111	6	74	171	8	86	231	10	104	291	10	118
52	4	52	112	6	76	172	8	86	232	10	104	292	10	118
53	6	52	113	6	76	173	8	88	233	10	104	293	10	118
54	4	52	114	6	76	174	8	88	234	10	104	294	10	118
55	4	52	115	6	78	175	8	88	235	10	104	295	10	118
56	4	52	116	6	78	176	8	88	236	10	104	296	10	120
57	6	52	117	6	78	177	8	90	237	10	104	297	10	120
58	6	52	118	8	78	178	8	90	238	10	104	298	10	120
59	6	52	119	8	78	179	8	90	239	10	104	299	10	120
60	6	52	120	8	78	180	8	90	240	10	104	300	10	120

References

[1] J. M. Andreu, , and S. N. Timasheff. The measurement of cooperative protein slef-assembly by turbidity and other techniques. *Methods in Enzymology*, 130:47, 1986. 9

[2] O.N. Antzutkin, R.D. Leapman, J. J. Balbach, and R. Tycko. Supramolecular structural constraints on Alzheimer's β-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry*, 41:15436, 2002. 96, 97

[3] S. Auer. Phase diagram of polypeptide chains. *J. Chem. Phys.*, 135:175103, 2011. 104, 119

[4] S. Auer, C.M. Dobson, and M. Vendruscolo. Characterization of the nucleation barriers for protein aggregation and amyloid formation. *HFSP J.*, 1:137, 2007. 1, 2, 43, 44, 62, 125

[5] S. Auer, C.M. Dobson, M. Vendruscolo, and A. Maritan. Self-template nucleation in peptide and protein aggregation. *Phys. Rev. Lett.*, 101:258101, 2008. 2, 8, 43, 44, 62, 104

[6] S. Auer and D. Kashchiev. Insight into the correlation between lag time and aggregation rate in the kinetics of protein aggregation. *Proteins*, 78:2412, 2010. 10, 62, 87

[7] S. Auer and D. Kashchiev. Phase diagram of α-helical and β-sheet forming peptides. *Phys. Rev. Lett.*, 104:168105, 2010. 1, 26, 43, 44, 104, 119

[8] S. Auer, F. Meersman, C.M. Dobson, and M. Vendruscolo. A generic mechanism of emergence of amyloid protofilaments from disordered oligomeric aggregates. *PLoS Comput. Biol.*, 4:e1000222, 2008. 16, 43, 44

[9] S. Auer, P. Ricchiuto, and D. Kashchiev. Two-step nucleation of amyloid fibrils: omnipresent or not? *J. Mol. Biol.*, 422:723, 2012. 16

[10] R. Becker and W. Döring. Kinetische Behandlung der Keimbildung in bersttigten Dmpfen. *Ann. Phys. (Leipzig)*, 24:719, 1935. 6, 42

[11] J. Berson, A.C. Theos, D.C. Harper, D. Tenza, G. Raposo, and M.S. Marks. Proprotein convertase cleavage liberates a fibrillogenic fragment of a resident glycoprotein to initiate melanosome biogenesis. *J. Cell. Biol.*, 161:521, 2003. 1, 125

[12] K. J. Binger. *The reversibility of amyloid fibril formation.* 2009. 9, 11

[13] A.B. Bortz, M.H. Kalos, and J.L. Lebowitz. A new algorithm for Monte Carlo simulation of Ising spin systems. *J. Comput. Phys.*, 17:10, 1975. 7, 106, 110

[14] H. T. Brandon and J. S. Weissman. Amyloid structure: Conformational Diversity and Conseguences. *Annu. Rev. Biochem.*, 80:557, 2011. 13

[15] K. Brendel, G. T. Barkema, and H. van Beijeren. Nucleation Time Distribuution in The-Dimensional Ising model With Spin-Flip Dynamics. *AIP Conference Proceedings*, 800:39, 2005. 7

[16] K. Brendel, G. T. Barkema, and H. van Beijeren. Nucleation times in the two-dimensional Ising model. *Phys. Rev. E*, 71:031601, 2005. 7

[17] R. Cabriolu and S. Auer. Amyloid fibrillation kinetics: insight from atomistic nucleation theory. *J. Mol. Biol.*, 411:275, 2011. 61, 65, 87, 101

[18] R. Cabriolu, D. Kashchiev, and S. Auer. Atomistic theory of amyloid fibril nucleation. *J. Chem. Phys.*, 133:225101, 2010. 19, 40, 62, 65, 67, 87, 88, 92, 96, 99, 104, 124

[19] R. Cabriolu, D. Kashchiev, and S. Auer. Size Distribution of Amyloid Nanofibrils. *Biophys. J.*, 101:2232, 2011. 86

[20] A. Caflisch. Computational models for the prediction of polypeptide aggregation propensity. *Curr. Opin. Chem. Biol.*, 10:437, 2008. 15, 62

[21] B. Chen, K. R. Thurber, F. Shewmaker, R.B. Wickner, and R. Tycko. Measurement of amyloid fibril mass-per-length by tilted-beam transmission electron microscopy. *Proc. Natl. Acad. Sci. USA*, 106:14339, 2009. 97

[22] F. Chiti and C. M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.*, 75:333, 2006. 1, 11, 12, 14, 61, 62, 104, 124, 125

[23] F. Chiti, M. Stefani, N. Taddei, G. Ramponi, and C. M. Dobson. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*, 424:805, 2003. 12, 14, 15, 62, 87

[24] T. Christopeit, P. Hortschansky, V. Schroeckh, K. Gührs, G. Zandomeneghi, and M. Fändrich. Mutagenic analysis of the nucleation propensity of oxidized Alzheimer's β-amyloid peptide. *Protein Sci.*, 14:2125, 2005. 10, 12, 14, 62, 78, 80, 82, 87, 99, 101

[25] P.G. de Gennes and J. Prost. *The physics of liquid crystals*. Calendon Press, Oxford, UK, 1995. 104

[26] J.P.K. Doye, A.A. Louis, I.C. Lin, L.R. Allen, E.G. Noya, A.W Wilber, H.C. Kok, and R. Lyus. Controlling crystallization and its absence: proteins, colloids and patchy models. *Phys. Chem. Chem. Phys.*, 9:2197, 2007. 104

[27] N. Duff and B. Peters. Nucleation in a Potts lattice gas model of crystallization from solution. *J. Chem. Phys.*, 131:184101, 2009. 8, 104

[28] M. Fändrich. Absolute correlation between lag time and growth rate in the spontaneous formation of several amyloid-like aggregates and fibrils. *J. Mol. Biol.*, 365:1266, 2007. 1, 10, 12, 14, 62, 87

[29] L. Farkas. Stoechiom. Verwandtschaftsl. *Z. Phys. Chem.*, 125:236, 1927. 6, 20

[30] A. M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, and L. Serrano. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.*, 22:1302, 2004. 12, 13, 14, 15, 62, 70, 78, 87

[31] F. A. Ferrone. Nucleation: the connections between equilibrium and kinetic behaviour. *Methods Enzymol.*, 412:285, 2006. 2, 62, 84, 129

[32] F. A. Ferrone, J. Hofrichter, and W. A. Eaton. Kinetics of sickle haemoglobin polymerization. II. A double nucleation mechanism. *J. Mol. Biol.*, 183:611, 1985. 2, 62, 125

[33] F. A. Ferrone, J. Hofrichter, and W. A. Eaton. Kinetics of sickle hemoglobin polymerization. I.Studies using temperature-jump and laser photolysis techniques. *J. Mol. Biol.*, 183:591, 1985. 2, 10, 62, 125

[34] F.A. Ferrone, J. Hofrichter, H.R. Sunshine, and W.A. Eaton. Kinetic studies on photolysis-induced gelation of sickle cell hemoglobin suggest a new mechanism. *Biophys. J.*, 32:361, 1980. 10, 125

[35] A.R. Fersht, J.P. Shi, J. Knill-Jones, D.M. Lowe, A.J. Wilkinson, D. M. Blow, and et al. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, 314:235, 1985. 66

[36] D.M. Fowler, A.V. Koulov, W.E. Balch, and J.W. Kelly. Functional amyloid-from bacteria to humans. *TRENDS in biochem. Sci.*, 32:217, 2007. 1, 125

[37] D. Frenkel and B. Smit. *Understanding Molecular simulation 2nd.* Academic Press, San Diego, 2002. 7

[38] O. Galkin, R. L. Nagel, and P. G. Vekilov. The kinetics of nucleation and growth of sickle cell haemoglobin fibers. *J. Mol. Biol.*, 365:425, 2007. 2, 62, 125

[39] O. Galkin, W. Pan, L. Filobelo, R.E. Hirsch, R. L. Nagel, and P. G. Vekilov. Two step mechanism of homogeneous nucleation of sickle cell haemoglobin polymers. *Biophys. J.*, 93:902, 2007. 16

[40] O. Galkin and P. G. Vekilov. Mechanism of homogeneous nucleation of polymers of sickle cell anaemia haemoglobin in deoxy state. *J. Mol. Biol.*, 336:43, 2004. 2, 62

[41] K. Garai, B. Sahoo, P. Sengupta, and S. Maiti. Quasihomogeneous nucleation of amyloid beta yields numerical bounds for the critical radius, the surface tension, and the free energy barrier for nucleus formation. *J. Chem. Phys.*, 128:045102, 2008. 2, 14, 15, 33, 62, 87, 96, 97, 98, 99

[42] K. Garai, R. Sureka, and S. Maiti. Detecting amyloid β aggregation with fiber-based fluorescence correlation spectroscopy. *Biophys. J.*, 92:L55, 2007. 14, 87

[43] E. Gazit. Self-assembled peptide nanostructures: the design of molecular building blocks and their technological utilization. *Chem. Soc. Rev.*, 36:1263, 2007. 1, 61

[44] J. W. Gibbs. *The scientific papers of J. Willard Gibbs*. Dover, New York (USA), 1961. 6

[45] G.H. Gilmer and J. Bennema. Simulation of crystal growth with surface diffusion. *J. Appl. Phys.*, 43:1347, 1972. 7, 8, 106, 107, 116, 124

[46] C. Goldsbury, U. Baxa, M.N. Simon, A.C. Steven, A. Engel, J.S. Wall, U. Aebi, and S.A. Muller. Amyloid structure and assembly: insights from scanning transmission electron microscopy. *J. Biol. Chem.*, 173:1, 2011. 97

[47] C. Goldsbury, P. Frey, V. Olivieri, U. Aebi, and S.A.) Muller. Multiple assembly pathways underlie amyloid-β fibril polymorphism. *J. Mol. Biol.*, 352:282, 2005. 96, 97

[48] C.S. Goldsbury, S. Wirtz, S.A. Muller, S. Sunderji, P. Wicki, U. Aebi, and P. Frey. Studies on the in vitro assembly of Aβ_{40}: Implications for the search for a β fibril formation inhibitors. *J. Struct. Biol.*, 130:217, 2000. 95, 96, 97

[49] L. Goldschmidt, P. K. Teng, R. Riek, and D. Eisenberg. Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. USA*, 107:3487, 2008. 74

[50] J. Greenwald and R. Riek. Biology of Amyloid: Structure, Function, and regulation. *Structure*, 18:1244, 2010. 11

[51] S. Grudzielanek, V. Smirnovas, and R. Winter. Solvation-assisted pressure tuning of insulin fibrillation: from novel aggregation pathways to biotechnological applications. *J. Mol. Biol.*, 356:497, 2006. 10, 12, 14, 62, 87

[52] C. Haas and J. Drenth. The interaction energy between two protein molecules related to physical properties of their solution and their crystals and implications for crystal growth. *J. Cryst. Growth*, 154:126, 1995. 68

[53] I.W. Hamley. Review: The Amyloid Beta peptide: a chemist's perspective role in Alzheimer's and fibrillization. *Chem. Rev.*, DOI: 10.1021/cr3000994:A, 2012. 63

[54] J. D. Harper and P. T. Jr Lansbury. Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependence solubility of amyloid proteins. *Annu. Rev. Biochem.*, 66:385, 1997. 2, 9, 10, 62

[55] P. Hartman. *Crystal Growth: An introduction.* P. Hartman (North-Holland, Amsterdam, 1973), p. 358, 1973. 24

[56] R.B. Heady and J.W. Cahn. Experimental test of classical nucleation theory in a liquidliquid miscibility gap system. *J. Chem. Phys.*, 58:896, 1973. 6

[57] L.O. Hedges and S. Whitelam. Limit of validity of Ostwald's rule of stages in a statistical mechanical model of crystallization. *J. Chem. Phys.*, 135:164902, 2011. 8, 104

[58] E. Hellstrand, B. Boland, D. M. Walsh, and S. Linse. Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem. Neurosci.*, 1:13, 2009. 10, 12, 14, 16, 62, 87

[59] A. Heneghan and A.D.J. Haymet. Nucleation of pure and AgI seeded supercooled water using an automated lag time apparatus. *AIP Conference Proceedings*, 534:439, 2000. 6

[60] J. Hofrichter. Kinetics of sickle haemoglobin polymerization III. Nucleation rates determined from stochastic fluctuations in polymerization progress curves. *J. Mol. Biol.*, 189:553, 1986. 2, 62, 125

[61] J. Hofrichter, P. D. Ross, and W. A. Eaton. Kinetics and mechanism of deoxyhemoglobin S gelation: a new approach to understanding sickle cell disease. *Proc. Natl. Acad. Sci. USA*, 71:4864, 1974. 2, 10, 62, 125

[62] P. Hortchansky, V. Schroeckh, T. Christopeit, G. Zandomeneghi, and M. Fändrich. The aggregation kinetics of Alzheimer's β-amyloid peptide is controlled by stochastic nucleation. *Protein Sci.*, 14:1753, 2005. 10, 12, 14, 62, 87

[63] W.B. Hu and D. Frenkel. Polymer crystallization driven by anisotropic interactions. *Adv. Polym. Sci.*, 191:1, 2005. 104

[64] K. Huang. *Statistical Mechanics*. J. Wiley and Sons Inc., New York (USA), 1987. 6

[65] C.P. Jaroniec, C.E. Mac Phee, V.S. Bajaj, M.T. McMahon, C.M. Dobson, and R.G. Griffin. High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. USA*, 101:711, 2004. 1, 2, 64

[66] J. T. Jarrett and P. T. Jr Lansbury. Amyloid fibril formation requires a chemically discriminating nucleation event: studies of an amyloidogenic sequence from the bacterial protein OsmB. *Biochemistry*, 31:12345, 1992. 9, 10

[67] J. T. Jarrett and P. T. Jr Lansbury. Seeding öne-dimensional crystallizationöf amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell*, 73:1055, 1993. 2, 62, 125

[68] S. Jiang. *Crystallization kinetics in polymorphic organic compounds*. Ph. D. Thesis, Delft University of Technology, Delft., 2009. 8, 104, 112, 118

[69] J.L. Jimenez, J.I. Guijarro, E. Orlova, J. Zurdo, C.M. Dobson, M. Sunde, and H.R. Saibil. Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J.*, 18:815, 1999. 2, 12

[70] V. M. Kaganer, W. Braun, and K. K. Sabelfeld. Ostwald ripening of faceted two-dimensional islands. *Phys. Rev. B*, 76:075415, 2007. 7

[71] V. M. Kaganer, K. H. Ploog, and K. K. Sabelfeld. Coarsening of facetted two-dimensional islands by dynamic coalescence. *Phys. Rev. B*, 73:115425, 2006. 7

[72] R. Kaischew. Equilibrium shape and work to form crystal nuclei on foreign substrates. *Bull. Acad. Bulg. Sci. (Phys.)*, 1:100, 1950. 24, 104, 117

[73] R. Kaischew and I.N. Stranski. Über den Mechanismus des Gleichgewichtes kleiner Kriställchen II. *Z. Phys. Chem B*, 26:114, 1934. 6, 23, 42, 45

[74] R. Kaischew and I.N. Stranski. Zur kinetischen Ableitung der Keimbildungsgeschwindigkeit. *Z. Phys. Chem B*, 26:317, 1934. 6, 23, 42, 45

[75] R. Kaischew and I.N. Stranski. Zur Theorie der linearen Kristallisationsgeschwindigkeit. *Z. Phys. Chem A*, 170:295, 1934. 6, 42

[76] J. Kardos, A. Micsonai, H. Pal-Gabor, E. Petrik, L. Graf, J. Kovacs, Y.-H. Lees, H. Naiki, and Y. Goto. Reversible heat-induced dissociation of β_2-Microglobuline Amyloid Fibrils. *Biochemistry*, 50:3211, 2011. 9

[77] D. Kashchiev. *Nucleation: basic theory with applications*. Butterworth-Heinemann, Oxford, UK, 2000. 4, 6, 19, 20, 21, 22, 24, 30, 32, 35, 36, 37, 41, 42, 53, 56, 57, 62, 73, 84, 88, 104, 108, 112, 121, 126

[78] D. Kashchiev. Forms and applications of the nucleation theorem. *J. Chem. Phys.*, 125:014502, 2006. 30, 35, 37, 53, 121

[79] D. Kashchiev. Interrelation between cluster formation time, cluster growth probability, and nucleation rate. *J. Chem. Phys.*, 127:064505, 2007. 32

[80] D. Kashchiev. Toward a better description of the nucleation rate of crystals and crystalline monolayers. *J. Chem. Phys.*, 129:164701, 2008. 8, 19, 22, 36, 41, 42, 45, 48, 53, 57, 60, 65, 113, 114, 120, 124, 126

[81] D. Kashchiev and S. Auer. Nucleation of amyloid fibrils. *J. Chem. Phys.*, 132:215101, 2010. 19, 22, 23, 26, 27, 28, 30, 32, 33, 34, 35, 36, 37, 40, 52, 57, 62, 65, 66, 68, 84, 87, 91, 95, 96, 104, 117, 119, 120, 124, 125

[82] D. Kashchiev, J. Van Der Eerden, and C. van Leeuwen. Transition from island to layer growth of thin films: a Monte carlo simulation. *Journal of Crystal Growth*, 42:31, 1977. 7, 8

[83] D. Kashchiev and G.M. van Rosmalen. On the relation between nucleation work, nucleus size and nucleation rate. *J. Chem. Phys.*, 76:5098, 1982. 30, 35, 37, 53, 121

[84] D. Kashchiev and G.M. van Rosmalen. Review: nucleation in solutions revisited. *Cryst. Res. Technol.*, 38:555, 2003. 19, 21, 24, 32, 41, 126

[85] M.S. Kellermayer, A. Karsai, M. Benke, K. Sooos, and B. Penke. Stepwise dynamics of epitaxially growing single amyloid fibrils. *Proc. Natl. Acad. Sci.*, 105:141, 2008. 95, 96, 98

[86] N. Kern and D. Frenkel. Fluidfluid coexistence in colloidal systems with short-ranged strongly directional attraction. *J. Chem. Phys.*, 118:9882, 2003. 104

[87] Y. S. Kim, S. P. Cape, E. Chi, R. Raffen, P. Wilkins-Stevens, F. J. Stevens, M. C. Manning, T. W. Randolph, A. Solomoni, and J.F. Carpenter. Counteracting effects on amyloid fibril formation by immunoglobulin light chains. *J. Biol. Chem.*, 276:1626, 2001. 10, 12, 14, 62, 87

[88] T. P. Knowels, C. A. Waudby, G. L. Devlin, S. I. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson. An analytical solution to the kinetics of breakable filament assembly. *Science*, 326:1533, 2009. 62

[89] T.P. Knowles and M.J. Buehler. Phase diagram of polypeptide chains. *Nanotechnol.*, 6:469, 2011. 124

[90] T.P.J. Knowles, C.A. Waudby, G.L. Devlin, S.I.A. Cohen, A. Aguzzi, M. Vendruscolo, E.M. Terentjev, M.E. Welland, and C.M. Dobson. An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science*, 326:1533, 2009. 10, 15, 125

[91] H. Komatsu, E. Feingold-Link, K.A. Sharp, T. Rastogi, P.H. Axelsen, and R. Tycko. Intrinsic linear heterogeneity of amyloid β protein fibrils revealed by higher resolution mass-per-length determinations. *J. Biol. Chem.*, 285:41843, 2010. 97

[92] K. C. Kunes, D. L. Cox, and R. R. P. Singh. One-dimensional model of yeast prion aggregation. *Phys. Rev. E*, 72:051915, 2005. 10, 12, 14, 62, 87

[93] Y. Kusumoto, A. Lomakin, D.B. Teplow, and G.B. Benedek. Temperature dependence of amyloid β-protein fibrillation. *Proc. Natl. Acad. Sci. USA*, 95:12277, 1998. 71, 78

[94] P. M. Lam, D. Bayayoko, and X. Y. Hu. Monte Carlo simulation of coarsening in a model of submonolayer epitaxial growth. *Surface Science*, 429:161, 1999. 7

[95] B. Lewis. *Crystal Growth*. B.R. Pamplin (Pergamon,Oxford), 1975. 42, 45

[96] B. Lewis and J.C. Anderson. *Nucleation and growth of thin films*. Academic, New York, 1978. 41, 42

[97] H. Li, F. Rahimi, S. Sinha, P. Maiti, G. Bitan, and K. Murakami. *Amyloids and Protein Aggregation-Analytical Methods*. 2009. 11, 15

[98] S. Linse, C. Cabaleiro-Lago, W. F. Xue, I. Lynch, S. Lindman, and E. Thulin. Nucleation of protein fibrillation by nanoparticles. *Proc. Natl Acad. Sci. USA*, 104:8691, 2007. 2, 62

[99] H.J. Liu, S.K. Kumar, and J.F. Douglas. Self-assembly-induced protein crystallization. *Phys. Rev. Lett.*, 103:018101, 2009. 104

[100] A. Lomakin, D. S. Chung, G. B. Benedek, D. A. Kirschner, and D. B. Teplow. On the nucleation and growth of amyloid β-protein fibrils: detection of nuclei and quantitation of rate constants. *Proc. Natl. Acad. Sci. USA*, 93:1125, 1996. 2, 62, 125

[101] A. Lomakin, D. B. Teplow, D. A. Kirschner, and G. B. Benedek. Kinetic theory of fibrillogenesis of amyloid β-protein. *Proc. Natl. Acad. Sci. USA*, 94:7942, 1997. 2, 15, 62

[102] O.S. Makin, E. Atkins, P. Sikorski, J. Johansson, and C. Serpell. Molecular basis for amyloid fibril formation and stability. *Proc. Natl Acad. Sci. USA*, 102:315, 2005. 1, 2

[103] O.S. Malkin, E. Atkins, P. Sikorski, J. Johansson, and L.C. Serpell. Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci. USA*, 102:315, 2005. 64

[104] A. Milchev, S. Stoyanov, and R. Kaischev. Atomistic theory of electrolytic nucleation: I. *Thin Solid Films*, 22:255, 1974. 42

[105] A. Milchev, S. Stoyanov, and R. Kaischev. Atomistic theory of electrolytic nucleation: II. *Thin Solid Films*, 22:267, 1974. 42

[106] A. Milchev and E. Vassileva. Electrolytic nucleation of silver on a glassy carbon electrode: Part II. Steady-state nucleation rate. *J. Electroanal. Chem.*, 107:337, 1980. 42, 53

[107] Y. Miller and R. Ma, B. Nussinov. Polymorphism in Alzheimer Aβ Amyloid Organization Ceflects Conformational Selection in a Rugged Energy Landscape. *Chemical Reviews*, 110:4820, 2010. 12

[108] H Naiki and K. Nakakuki. First-order kinetic model of Alzheimer's beta-amyloid fibril extension in vitro. *Laboratory investigation*, 74:374, 1996. 9

[109] H. D. Nguyen and C. K. Hall. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl Acad. Sci. USA*, 101:16180, 2004. 1, 2, 62

[110] H. D. Nguyen and C. K. Hall. Spontaneous fibril formation by polyalanines; discontinuous molecular dynamics simulations. *J. Am. Chem. Soc.*, 128:1890, 2006. 1, 2, 62

[111] H.D. Nguyen and C.K. Hall. Kinetics of fibril formation by polyalanine peptides. *J. Biol. Chem.*, 280:9074, 2005. 2, 22, 62

[112] L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V.N. Uversky, and A.L. Fink. Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. *Biochemistry*, 40:6036, 2001. 10, 12, 14, 62, 87

[113] T. N. Niraula, T. Konno, H. Li, H. Yamada, K. Akasaka, and H. Tachibana. Pressure-dissociable reversible assembly of intrinsically denatured lysozyme is a precursor for amyloid fibrils. *Proc. Natl Acad. Sci. USA*, 101:4089, 2004. 9

[114] C.L. Oliveira, M.A. Behrens, J.S. Pedersen, and et al. A SAXS study of glucacon fibrillation. *J. Mol. Biol.*, 387:147, 2009. 14, 87

[115] B. O' Nuallan, S. Shivaprasad, I. Kheterpal, and R. Wetzel. Thermodynamics of A beta(1-40) Amyloid fibril elongation. *Biochemistry*, 44:12709, 2005. 78

[116] J. S. Pedersen, J. M. Flink, D. Dikov, and D. E. Otzen. Sulfates dramatically stabilize a salt-dependent type of glucagon fibrils. *Biophys. J.*, 90:4181, 2006. 10, 12, 14, 62, 87

[117] A.T. Petkova, Y. Ishii, J.J. Balbach, O.N. Antzutkin, R.D. Leapman, F. Delaglio, and R. Tycko. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. USA*, 99:16742, 2002. 78, 96, 97

[118] G. W. Platt, K. E. Routledge, and S. E. Radford. Fibril growth kinetics reveal a region of β_2-microglobulin important for nucleation and elongation of aggregation. *J. Mol. Biol.*, 378:251, 2008. 70

[119] R. A. Ramos, P. A. Rikvold, and M. A. Novotny. Test of the Kolmogorov-Johnson-Mehl-Avrami picture of metastable decay in a model with microscopic dynamics. *Phys. Rev. B*, 59:9053, 1999. 7

[120] F. Romano, E. Sanz, and F. Sciortino. Crystallization of tetrahedral patchy particles in silica. *J. Chem. Phys.*, 134:174502, 2011. 8, 104

[121] F. Romano and F. Sciortino. Two dimensional assembly of triblock Janus particles into crystal phases in the two bond per patch limit. *Soft Matter*, 7:5799, 2011. 8, 104

[122] K. E. Routledge, G. G. Tartaglia, G. W. Platt, M. Vendruscolo, and S. E. Radford. Competition between intramolecular and intermolecular interactions in an amyloid-forming protein. *J. Mol. Biol.*, 389:776, 2009. 10, 12, 14, 62, 70, 73, 87

[123] C. Sachse, M. Fandrich, and N. Griogorieff. Paired β-sheet structure of Aβ(1-40) amyloid fibril revealed by electron microscopy. *Proc. Natl. Acad. Sci. USA*, 105:7462, 2008. 78, 90, 97

[124] M. R. Sawaya, S. Sambashivan, R. Nelson, M.I. Ivanova, S. A. Sievers, M. I. Apostol, and et al. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature*, 447:453, 2007. 1, 2, 62, 64

[125] J. W. P. Schmelzer. *Nucleation Theory and Applications*. WILEY-VCHB Verlag GmbH and Co. KGa Weinheim, 2005. 6

[126] M. Schmidt, C. Sachse, W. Richter, C. Xu, M. Fandrich, and N. Grigorieff. Comparison of Alzheimer Aβ(1-40) and Aβ(1-42) amyloid fibrils reveals similar protofilament structures. *Proc. Natl. Acad. Sci. USA*, 106:19813, 2009. 97

[127] J.D. Schmit, K. Ghosh, and K. Dill. What drives amyloid molecules to assemble into oligomers and fibrils? *Biophys. J.*, 100:450, 2011. 87

[128] T. R. Serio, A. G. Cashikar, A. S. Kowal, G. J. Sawicki, J. J. Moslehi, and L. Serpell. Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science*, 289:1317, 2000. 2, 62

[129] V. A. Shneidman, K. A. Jackson, and K. M. Beatty. Non-equilibrium interface of a two-dimensional low-temperature crystal. *Journal of Crystal Growth*, 212:564, 2000. 7

[130] V. A. Shneidman and G. M. Nita. Modulation of the Nucleation Rate Preexponential in a Low-Temperature Ising System. *Phys. Rev. Lett.*, 89:025701, 2001. 7

[131] V. A. Shneidman and G. M. Nita. Nucleation preexponential in dynamic Ising models at moderately strong fields. *Phys. Rev. E*, 68:021605, 2003. 7

[132] V. A. Shneidman and G. M. Nita. On the critical cluster in the two-dimensional Ising model: Computer-assisted exact results. *J. Chem. Phys.*, 121:11232, 2004. 7

[133] V. A. Shneidman and G. M. Nita. Collapse of Transient Nucleation Fluxes in a Cold Ising Ferromagnet. *Phys. Rev. Lett.*, 97:065703, 2006. 7

[134] I.N. Stranski and R. Kaischew. Über den Mechanismus des Gleichgewichtes kleiner Kriställchen. *Z. Phys. Chem B*, 26:100, 1934. 6, 23, 42, 45, 105

[135] I.N. Stranski and R. Kaischew. Über den Mechanismus des Gleichgewichtes kleiner Kriställchen III. *Z. Phys. Chem B*, 26:312, 1934. 6, 42

[136] M. Sunde and C. Blake. The structure of amyloid fibrils by electron microscopy and x-ray diffraction. *Adv. Protein Chem.*, 50:123, 2008. 1, 2, 11

[137] W. C. Swope and H. C. Andersen. 10^6-particle molecular-dynamics study of homogeneous nucleation of crystals in a supercooled atomic liquid. *Physical Review B*, 10:7042, 1990. 8

[138] G. G. Tartaglia, A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti, and M. Vendruscolo. Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.*, 380:425, 2008. 12, 14, 15, 62, 87

[139] P. R. ten Wolde and D. Frenkel. Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations. *Sciences*, 277:1975, 1997. 8

[140] J.H. ter Horst and P.J. Jansens. Nucleus size and Zeldovich factor in two-dimensional nucleation at the Kossel crystal (001) surface. *Surface Science*, 574:77, 2005. 7, 8, 110

[141] J.H. ter Horst and D. Kashchiev. Determining the nucleation rate from the dimer growth probability. *J. Chem. Phys.*, 123:114507, 2005. 6, 7, 8, 106, 107, 108, 110, 112, 116, 118

[142] J.H. ter Horst and D. Kashchiev. Rate of two-dimensional nucleation: verifying classical and atomistic theories by Monte Carlo simulation. *J. Phys. Chem. B*, 112:8614, 2008. 8, 106, 107, 110, 112, 114, 116, 118

[143] S. Toskev. *Crystal Growth: An introduction.* P. Hartman (North-Holland, Amsterdam, 1973), p.328, 1973. 24

[144] A. Trovato, F. Chiti, A. Maritan, and F. Seno. Insight into the structure of amyloid fibrils from the analysis of globular proteins. *PLoS Comput. Biol.*, 2:e170, 2006. 12, 14, 15, 62, 66, 70, 78, 87, 90, 129

[145] R. Tycko. Molecular structure of amyloid fibrils: insights from solid state NMR. *Q. Rev. Biophys.*, 39:1, 2006. 1, 2

[146] J. Van Der Eerden, D. Kashchiev, and P. Bennema. Surface migration of small crystallites: a Monte carlo simulation with continuous time. *Journal of Crystal Growth*, 42:31, 1977. 8

[147] C. van Leeuwen, R. van Rosmalen, and P. Bennema. Simulation of step motion on crystal surfaces. *Surface Science*, 44:213, 1974. 7

[148] K. Van Vorkum and J.F. Douglas. Symmetry, equivalence and molecular self-assembly. *Phys. Rev. E.*, 73:031502, 2006. 104

[149] B. Vestergaard, M. Groenning, M. Roessie, J. S. Kastrup, M. van de Weert, J.M. Flink, and et al. A helical structural nucleus is the primary elongating unit of insulin amyloid fibrils. *PLos Biol.*, 5:e134, 2007. 14, 87

[150] M. Volmer and A. Weber. Kinetic der Phasenbildung. *Z. Phys. Chem. (Leipzig)*, 119:227, 1926. 6

[151] D. Walton. Nucleation of Vapor Deposits. *J. Chem. Phys.*, 37:2182, 1962. 7, 42

[152] C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, and B. H. Meier. Amyloid fibrils of the HET-s(218-289) prion form a β solenoid with triangular hydrophobic core. *Science*, 319:1523, 2008. 1, 2, 62, 64

[153] Jan Wedekind. *Phd Thesis: Nano-Droplets at Birth, Computer Experiments on Gas Phase Nucleation.* 2006. 6, 7

[154] J.D. Weeks and G.H. Gilmer. Dynamic in crystal growth. *Adv. Chem. Phys.*, 40:157, 1979. 7, 112

[155] A.D. Williams, E. Protelius, I. Kheterpal, J. Guo, K.D. Cook, Y. Xu, and R. Wetzel. Mapping Aβ amyloid fibril secondary structure using scanning proline mutagenesis. *J. Mol. Biol.*, 335:833, 2004. 78

[156] S. Wonczac, R. Stray, and D. Stauffer. Confirmation of classical nucleation theory by Monte Carlo simulations in 3-dimensional Ising model at low-temperature. *J. Chem. Phys.*, 113:1976, 2000. 7

[157] W. F. Xue, S. H. Homans, and S. E. Radford. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc. Natl Acad. Sci. USA*, 105:8926, 2008. 2, 10, 12, 15, 62, 70, 72, 76, 77

[158] K. Yasuoka and M. Matsumoto. Molecular dynamics of homogeneous nucleation in the vapour phase. I. Lennard-Jones fluid. *J. Chem. Phys.*, 109:8451, 1998. 8

[159] S.-T Yau and P. G. Vekilov. Quasi-planar nucleus structure in apoferritin crystallization. *Nature*, 406:494, 2000. 6

[160] A. Yethiraj and A. van Blaaderen. A colloidal model system with an interaction tunable from hard sphere to soft and dipolar. *Nature*, 421:513, 2003. 104

[161] S. Yoon and W. J. Welsh. Detecting hidden sequences propensity for amyloid fibril formation. *Protein Sci.*, 13:2149, 2004. 12, 14, 15, 62, 87

[162] E. D. Zanotto and P. F. James. Experimental tests of the classical nucleation theory for glasses. *J. Non-crystalline Solids*, 74:373, 1985. 6

[163] J. Zhang and M. J. Muthukumar. Simulations of nucleation and elongation of amyloid fibrils. *J. Chem. Phys.*, 130:035102, 2009. 2, 14, 15, 22, 62, 87, 91, 104, 117

[164] L. Zhu, X. J. Zhang, L. Y. Wang, J. M. Zhou, and S. Perret. Relationship between stability of folding intermediates and amyloid formation for the yeast prion Ure2p: a quantitative analysis of the effects of pH and buffer system. *J. Mol. Biol.*, 328:253, 2001. 10, 12, 14, 62, 87