Kinetics and mechanisms of peptide aggregation from molecular dynamics simulations

Beata Szała-Mendyk

beata.szala@amu.edu.pl

A doctoral thesis submitted to
the Faculty of Chemistry
of Adam Mickiewicz University

Supervisor: Prof. dr hab. Andrzej Molski

02.11.2021

Acknowledgements

I would like to thank those, who have been with me and helped me all these years.

I especially thank Professor Andrzej Molski for being my supervisor. This thesis would not have been possible without His patience, counsel and support throughout all this time. I am thankful for the inspiring discussions which teach and motivate me for further effort. The work on this thesis has been a pleasure thanks to His extraordinary sense of humor making every moment a great memory.

I would like to thank Professor Alberta Ferrarini and Professor Birgit Strodel, who hosted me during the internships. The studies performed under their supervision broadened my knowledge and horizons.

Also, I thank my family, my Parents, Sister and Husband, who endured this long process with me, lent me a hand in difficult moments and motivated to overcome challenges.

List of publications

This thesis is based on publications, according to Article 187 of The Law on Higher Education and Science, the Act of 20 July 2018.

The research discussed in this thesis was presented in the following publications:

P1 Szała B., Molski A., Chiral structure fluctuations predicted by a coarse-grained model of peptide aggregation *Soft Matter*, **2020**, 16:5071-5080.

P2 Szała-Mendyk B., Andrzej Molski A., Clustering and Fibril Formation during GNNQQNY Aggregation: A Molecular Dynamics Study *Biomolecules*, **2020**, 10:1362.

P3 Szała-Mendyk B., Molski A., Diverse Aggregation Kinetics Predicted by a Coarse-Grained Peptide Model *J. Phys. Chem. B*, **2021**, 125:7587-7597.

Abstract

Protein and peptide aggregation has been intensely studied due to its biological importance as well as potential technological applications. Despite much effort, the molecular mechanism of aggregation is still not completely understood. In particular, the initial aggregation steps are difficult to investigate experimentally. To fill this gap, the molecular dynamics simulations have been employed as a method which can give insight into the molecular details of biologically relevant aggregation.

This thesis focuses on the equilibrium properties of mature aggregates and on the kinetics and molecular mechanism of transient oligomers formation. Those transient oligomers are potentially more toxic than the mature aggregates. In this thesis the peptide aggregation is studied as a function of the peptide sequence and the chain properties. Our results give information about the molecular origin of aggregation diversity and fibril polymorphism.

Due to the computational cost of all-atom simulations, coarse-grained models have been applied allowing for larger systems sizes. We used two coarse-grained models. The first model, developed in our research group, treats each amino acid residue as one super-atom. Such approach allows for investigating how the generic peptide properties connected with the peptide backbone influence aggregation. This model reproduces the diversity of aggregates structures and kinetics observed in experiments as functions of just two parameters, the chain stiffness and the strength of non-bonded interactions. Moreover, this simple model reveals a new feature, the chiral fluctuations of small oligomers. Despite the lack of intrinsic chirality of individual peptide chains, oligomers undergo spontaneous collective twists. The overall handedness of aggregates fluctuates between left and right. The life-times of the twisted states and their amplitudes depend on the cluster mass. The fluctuations, found by our model, shed light on the origin of the chiral polymorphism observed in mature fibrils.

In the second part of this thesis, the coarse grained Bereau and Deserno (BD) model is used. The presence of the side chains allows for investigating the sequence-dependence of peptide self-assembly. This part focuses on two heptapeptides: a short

fragment from yeast prion protein, GNNQQNY, and its point mutation, GNNQQNA. Both peptides aggregate via the two-step mechanism leading to fibril-like structures. The simulations show that the aggregation slows down when the tyrosine residue is replaced by alanine. Moreover, this sequence mutation causes differences in the morphology of the final aggregates and the intermediate oligomers.

Our research indicates that the aggregation kinetics and the aggregate morphology depend on the properties of the backbone as modulated by the side chains. Notably, our simulations show that the structural diversity of final aggregates may not be reflected by the diversity of monomer and cluster kinetics. Nevertheless, the internal organization of the initial oligomers parallels the structures of the mature aggregates, which implies the relevance of early aggregation steps.

Streszczenie

Zjawisko agregacji białek i peptydów cieszy się dużym zainteresowaniem naukowców ze względu na swoje biologiczne znaczenie, jak również potencjalne zastosowanie w technologii. Mimo licznych wysiłków, mechanizm molekularny tego procesu nie został dotychczas całkowicie wyjaśniony. Szczególnie trudne dla badań eksperymentalnych jest prowadzenie obserwacji początkowych etapów agregacji. W tym miejscu z pomocą przychodzą symulacje dynamiki molekularnej, które umożliwiają przyglądanie się procesom biologicznym na poziomie pojedynczych cząsteczek.

W przeciwieństwie do często podejmowanych badań równowagowych, niniejsza praca koncentruje się dodatkowo na kinetyce i mechanizmie formowania początkowych oligomerów, które podejrzewane są o większą nawet toksyczność niż dojrzałe agregaty. Podobieństwa i różnice procesu agregacji zbadane zostały w zależności od sekwencji aminokwasowej i właściwości łańcucha peptydowego. Wyniki dały wgląd w molekularne przyczyny obserwowanego polimorfizmu fibryli jak i różnice w przebiegu samego procesu.

W niniejszej pracy zastosowane zostały modele gruboziarniste, aby zredukować koszt obliczeniowy i dzięki temu zwiększyć rozmiar badanych systemów w stosunku do ujęcia atomistycznego. Pierwszym użytym modelem jest rozwijany w naszej grupie badawczej Model Minimalny, który przedstawia peptydy z dokładnością jednego super-atomu na aminokwas, co odpowiada reprezentacji łańcucha głównego z pominięciem łańcuchów bocznych. Takie podejście umożliwia prześledzenie wpływu jaki na agregację wywierają ogólne właściwości peptydów, tzn. takie, które wynikają ze specyfiki łańcucha głównego. Model ten pozwolił na odtworzenie strukturalnej oraz kinetycznej różnorodności obserwowanej w eksperymencie jako funkcji dwóch parametrów, sztywności łańcucha oraz siły oddziaływań niewiążących. Ponadto, model ten ujawnił występowanie chiralnych fluktuacji dla oligomerów. Pomimo braku wewnętrznej chiralności, peptydy skręcają się spontanicznie w tym samym kierunku, przy czym wypadkowa chiralność całego agregatu oscyluje pomiędzy prawo- i lewoskrętnością w trakcie symulacji. Czas życia stanów chiralnych, oraz amplituda skrętu, zależą od masy klastru. Fluktuacje te rzucają nowe światło na pochodzenie polimor-

fizmu dojrzałych fibryli, obserwowanego zarówno *in vitro*, jak i *in vivo*.

W drugiej części pracy wykorzystany został model opracowany przez Bereau i Deserno (BD). Obecność dodatkowych super-atomów reprezentujących łańcuchy boczne umożliwia zbadanie wpływu structury I-szo rzędowej na proces samoorganizacji peptydów. W tej części skupiono się na dwóch heptapetydach: krótkim fragmencie białka prionowego drożdży o sekwencji GNNQQNY, oraz jego jednopunktowej mutacji, GNNQQNA. Obie cząsteczki agregują według dwu-etapowego mechanizmu z utworzeniem fibrylarnych agregatów. Pomimo ogólnych podobieństw, szybkość agregacji maleje, gdy reszta tyrozynowa zastąpiona zostaje przez alaninę. Ponadto, jedno-aminokwasowa mutacja powoduje zmianę w strukturze zarówno początkowych oligomerów jak i finalnych agregatów.

Wyniki prezentowane w rozprawie wskazują, iż kinetyka procesu i morfologia agregatów zależą od właściwości łańcucha głównego jak i łańcuchów bocznych. Zmiany w strukturze i kinetyce, wywołane modyfikacjami cząsteczek peptydów, nie zachodzą jednocześnie, różnica w strukturze niekoniecznie wiąże się z innym zachowaniem kinetycznym. Niemniej jednak, heterogeniczność finalnych aggregatów znajduje odzwierciedlenie w morfologii początkowych oligomerów, co wskazuje na duże znaczenie wczesnych etapów agregacji dla obserwowanego polimorfizmu amyloidów.

List of abbreviations

AA	All-Atom models
b	asphericity, eq. 2.6
BD	Bereau and Deserno peptide model
C_{n}	end-to-end Correlation parameter, eq 2.7
CG	Coarse-Grained models
GNN	heptapeptide from yeast prion protein, Sup35, with sequence: GNNQQNY
c_0	initial peptide Concentration
c_{SA}	Super-Atom Concentration
CP	Change Point analysis
MFPT	Mean First-Passage Time analysis
H	fibril Helicity, eq 2.9
L	Length of the simulation box edge
M, M_{max}, $\overline{M}_{\mathrm{max}}$	Mass (size) of an aggregate, the Mass (size) of the largest (max) aggregate, and the Mass (size) of the largest (max) aggregate averaged over all simulation repeats, defined as the number of peptides in the aggregate
m^*	nucleus Mass (size) defined as the number of peptides in the nucleus
MD	Molecular Dynamics
N_0	Number of peptides in the system equal to the initial number of free monomers
N_{m}, $\overline{N}_{\mathrm{m}}$, $\overline{N}_{\mathrm{m,sc}}$	Number of free Monomers, Number of free Monomers averaged over all simulation repeats, and the SCaled average number of free Monomers, eq. 2.4, respectively
$\overline{N}_{\mathrm{m,lag}}$, $\overline{N}_{\mathrm{m,eq}}$	average Number of free Monomers in the Lag phase and the EQuilibrium phase, respectively
N_{c}, $\overline{N}_{\mathrm{c}}$	Number of Clusters and the Number of Clusters averaged over all simulation repeats

q	cluster twist descriptor, eq. 2.8
R_{g}	Radius of Gyration
SA	number of Super-Atoms per peptide
TPM	Transition Probability Matrix
τ^*	average nucleation time
Y7A	heptapeptide with amino acid sequence: GNNQQNA

Contents

Chapter 1

Introduction

This chapter introduces the problem of protein and peptide aggregation and briefly reviews the state of the art of this research topic. First, the importance of aggregation is discussed as the motivation for this thesis. Then, the structural properties of aggregates and aggregation kinetics are reviewed. As an example, the aggregation of the yeast prion protein, Sup35, and its heptapeptide fragment, GNNQQNY, is presented. Next, the molecular dynamic simulation is introduced as the working method in this thesis. Several aspects of peptide aggregation have remained unexplained so far. With this in mind, the last section discusses the general aims and specific objectives of this thesis.

1.1 Protein and peptide aggregation

It is well known that the formation of protein or peptide aggregates is connected with various diseases collectively called amyloidosis [1]. At least 50 human pathologies have been linked with the intra- or extra-cellular deposition of amyloid fibrils, including such neurodegenerative disorders as Alzheimer's, Parkinson's or Huntington's diseases.

However, protein and peptide fibrils might also play important role in the proper cell activities [2]. For instance, protein fibrils can assist the protein storage in inert form as in the case of peptide hormones or toxic protein in white blood cells. Self-templating properties make the fibrils of functional prion proteins able to save epigenetic memory. Due to the structural strength, stability and adhesive surface, fibrils are involved in the biofilm formation in bacteria or in the oocyte protection in fish mammals. Those and other examples of the positive effects of fibrils have been reviewed in the literature [2, 3].

Recently, amyloid fibrils have gained interest in nanotechnology and tissue en-

gineering due to their exceptional mechanical features such as the strength that is comparable to that of steel and the stability in broad range of environmental conditions [4]. For that reason, fibrils can be potentially used as scaffolds for cell-specific growth. The hydrogels made up of amyloids are also considered for drug delivery and slow drug release [5].

Despite numerous studies, the mechanism of protein and peptide aggregation is still not completely understood. Especially, the initial aggregation steps, including transient oligomers formation, remain elusive [6]. For example, it has been suggested that the transient oligomers can be more toxic than the mature fibrils [6]. Therefore, information on the formation and stability of transient oligomers can be more important than that about final fibrils. Moreover, possible applications of amyloid fibrils also require an understanding of the self-assembly process to control and modify the final products.

The biological and technological significance on the one hand and the gaps in our knowledge on the other, make the protein and peptide aggregation an important issue in bio-molecular science.

1.1.1 Morphology of protein and peptide aggregates

Although high aggregation propensity is a common feature of numerous peptides and proteins, the final aggregates vary from amorphous to highly-ordered fibrils or crystalline structures [7]. Interestingly, various aggregate structures can be observed for one sequence depending on the environmental conditions [8].

Amyloid fibrils with the common cross-β pattern are often formed by various proteins and peptides. Such fibrils usually consist of several protofibrils and have up to several micromiters in length with 6-12 nanometers in width [6].

Despite the general similarities, mature fibrils differ in mesoscopic details such as the number of protofilaments, the helix pitch for twisted fibrils and also the handedness [6]. Several main topologies have been specified, such as nanotubes, helical and twisted ribbons, achiral flat ribbons and crystals. Transitions between these forms can occur with the change in number of protofilaments. As an example, the HEWL lysozyme initially forms twisted ribbons which are changing into helical ribbons when the number of protofilaments exceeds 4. Moreover, different twists are observed for the fibrils formed by amyloid-β peptide when the number of protofilaments change [8]. Even if the number of protofilaments does not change other kind of fibril polymorphism might be observed. This fibril polymorphism has two origins: the internal structure of single protofilements is different, as it has been discovered for α-synuclein, or the protofilaments are packing in a different way, as for the tau fibrils [8].

Another aspect of fibril diversity is the handedness of twisted fibrils. Although classical β-sheets show the left-handed twist, right-handed amyloid fibrils are also observed [6]. For instance, murine and human serum amyloid A proteins, SAA, form fibrils composed of two protofilaments which are twisted around the fibril axis. Despite the 78% sequence identity, the formed fibrils have the opposite chirality, the right-handed in the case of human SAA and the left-handed in the murine fibril [9]. The sequence dependence of fibril twist is also visible for short fragments from N-terminus of human SAA: SAA_{1-12}, SAA_{1-11}, and SAA_{2-9}. The longest fragment, SAA_{1-12} aggregates into right-handed fibrils, the one-residue shorter peptide, SAA_{1-11}, aggregates into left-handed fibrils, whereas the shortest one, SAA_{2-9}, forms a racemate [10]. Moreover, the left- and right-handed fibrils can be formed depending on the environmental conditions, such as pH [11], temperature [12] and seeding [13].

The origin of fibril variability is unclear but seems to be important as various fibril structures formed by the same protein or peptide can be connected with different diseases [8].

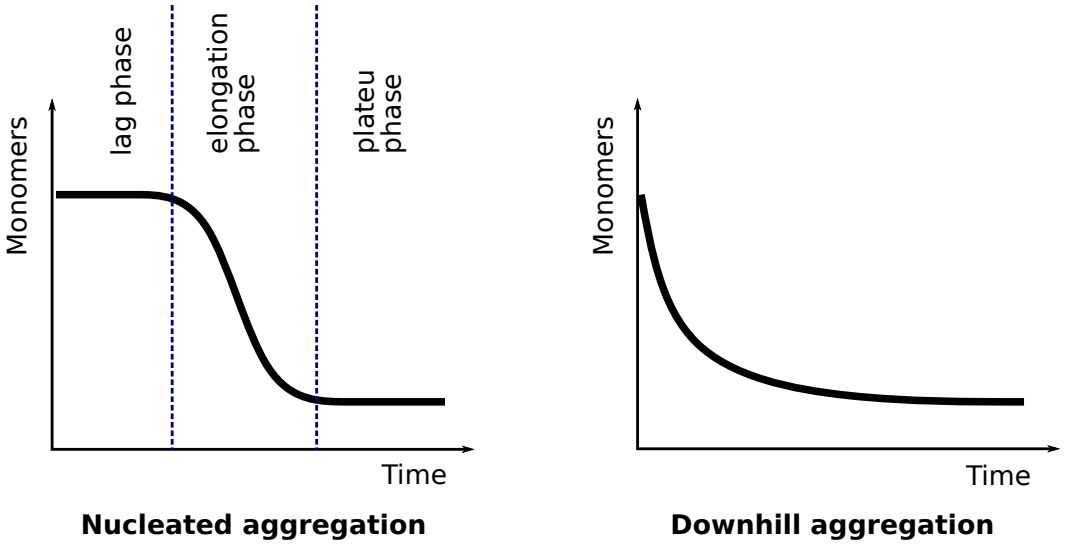

Figure 1.1: Schematic plots for the nucleated (left) and downhill (right) aggregation kinetic modes.

Beside highly-ordered fibrils, amorphous aggregates are also observed in both living cells and *in vitro* experiments. For instance, such disordered structures are observed for amyloid-β peptides in acidic solution, pH $\approx$ 5.5 [14, 15]. The significance of amorphous aggregates in fibril formation is unclear and they are considered as both off- or on-pathway individua [7]. Regardless of their relevance for fibril formation, they are postulated as even more toxic than the mature fibrils [14, 15].

1.1.2 Peptide aggregation kinetics

Various aggregation kinetics have been observed in the experiments. Two main types of aggregation kinetics can be distinguished: nucleated and downhill, see the schematic plots in fig. 1.1. In nucleated aggregation, a nucleus is initially formed. This process is visible as the lag phase in a kinetic plot. In the next phase, called a downhill or elongation phase, monomers decay quickly as they aggregate with the growing nucleus. The last stage is called the plateau or equilibrium phase, when the monomer concentration does not change, although the aggregates can undergo an internal reorganization or fragmentation. The downhill aggregation does not have a visible lag phase as the monomers decay quickly right from the beginning.

Different nucleation behaviors can be defined in protein and peptide aggregation:

1. Primary nucleation where an initial nucleus is formed directly from the solution [1]. This process is critical for the aggregation initiation. Based on the nucleus properties one can distinguish two types of primary nucleation:

    - Nucleated polymerization or one-step nucleation where the nucleaus is

defined only by the critical aggregate mass [1, 16].

- Nucleated conformational conversion, where the nucleus formed in the first step undergoes further structural transformations into the structure of the mature fibril. This mechanism is also known as two-step nucleation [1, 16].

Both one- and two-step nucleation mechanisms are observed in protein and peptide aggregation depending on the experimental conditions such as the concentration and the temperature. Specifically, one-step nucleation at low peptide concentrations, whereas two-step nucleation occurs when the monomer concentration is higher than the oligomer solubility at given temperature [16].

2. Secondary nucleation which occurs in a later aggregation stage. The nucleaus formation is accelaterated by the existing fibrils. Secondary nucleation includes the fibril-catalyzed nucleation and the nucleation by fibril fragmentation [1].

1.1.3 Aggregation of Sup35 protein and its fragment

Fibril formation was intensively studied for the yeast prion protein, Sup35, which is related not only to its biological relevance but also to the accessibility and relative simplicity of the yeast organisms [17].

Aggregation of the whole Sup35 protein is an interesting example of changes in the aggregation mechanism following the experimental conditions. Serio at al. tested various aggregation models and established that at concentrations in the range 0.1-50 μM the fibrillation process occurs by the nucleated conformational conversion with a micelle-like intermediate oligomers [18]. Ohhashi et al. found that the fibril formation for Sup35 may occur in two different ways depending on the temperature [19]. At low temperatures, 4 ^{o}C, the nucleated conformational conversion mechanism is observed with oligomer reorganization as the rate-determining step. However, at high temperatures, 37 ^{o}C, a monomer or a small oligomer is the nucleus and reorganization events are not observed [19]. Such one-step mechanism for the Sup35 aggregation at high temperatures was also confirmed by recent experimental studies [20]. In this case, the formed oligomers have been determined as off-pathway species for fibril formation.

It was found, that the short fragment of Sup35 protein can aggregate into fibrils which are similar to those formed by whole protein [21]. This fragment is a heptapeptide with the sequence GNNQQNY, abbreviated here as GNN. Short peptide fragments are often used to study the aggregation behavior of the whole protein.

Experimental conditions needed to study model peptides can be less demanding than those for the whole protein [21].

Recent studies show the one-step fibrillation mechanism for GNN heptapetide at concentrations in the range ≈ 0.7-1.5 mM [21]. The nucleus size decreases slightly with decreasing temperature from 7 peptides at 37 °C to 6 peptides at 23 °C. For the lowest temperature, 4 °C, aggregation occurs rapidly and the lag-phase was not observed [21].

Fibrilar polymorphism is observed for both the Sup35 protein and its fragment GNN. In the case of whole protein, Sup35, two types of fibrils are observed following two aggregation mechanisms described above [19]. These two fibrils differ in the physical properties such as the melting temperature [22].

Even more aggregate structures are observed in the case of the GNN peptides, including two crystalline forms. The structure of GNN assembles depends on the aggregation conditions such as peptide concentration [17] or stirring [23]. At low concentration, ≈ 2.4 mM, orthorhombic nanocrystals are observed. At the concentration in the range 2.4-12 mM monoclinic crystals are the most favorable species. At the concentration higher than 12 mM, GNN can form fibrils beside monoclinic crystals. When the concentration is higher than ≈ 24 mM only fibrils are present in solution after aggregation.

Additionally, attempts have been made to determine the internal structure of GNN fibrils. Eisenberg and co-workers proposed a fibril model based on the similarities between GNN fibrils and microcrystals [24]. In this model, GNN peptides are arranged in parralell and in-register, forming single β-sheet. Such β-sheets form antiparallel pairs, with dry interface between them. This interface is called a steric-zipper and is proposed as the fundamental property of fibrils [24, 25]. On the other hand, Marshall et al. tested the properties of GNN fibrils and crystals and found some dissimilarities, which suggest different internal structures. Based on these results, they hypothesize that the dry steric-zipper may appear only in GNN crystals whereas GNN fibrils are more hydrated [26]. Also the more recent work shows that the β-sheet zipper in fibrils differ from this found in crystals [27]. Langkilde and co-workers proposed a different packing model, where a single unit is formed by four rather than two β-sheets. These β-sheets forms two parallel pairs with face-to-face zipper. Two pairs are arranged in such a way, that they contact by tyrosine residues [27].

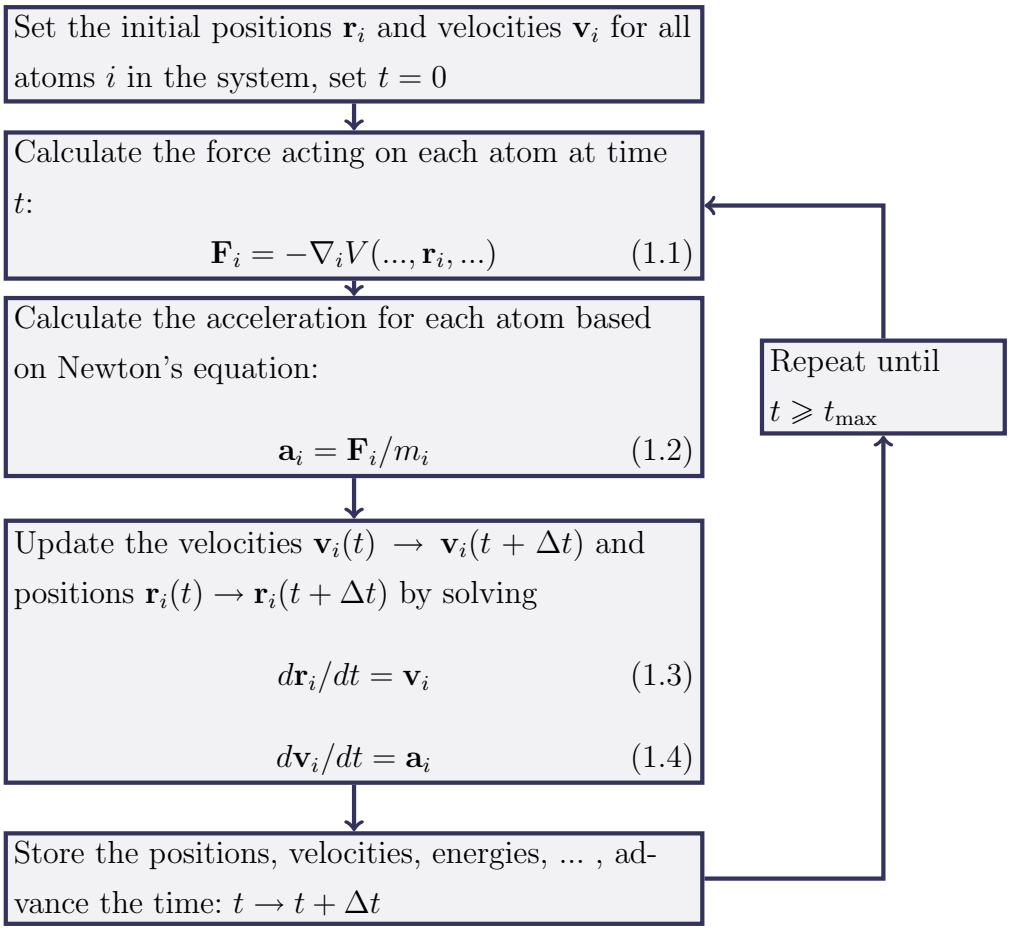

Figure 1.2: Schematic illustration of molecular dynamics simulation procedure.

1.2 Molecular dynamic simulations

Molecular dynamic (MD) simulations explore biological processes at the molecular level. MD studies are commonly used to investigate such processes as: drug targeting and docking, neuronal signaling, protein folding, conformational dynamics of biomacromolecules, and proteins and peptides self-assembly [28, 29]. Figure 1.2 shows the scheme for a basic molecular dynamic simulation algorithm. The time evolution is calculated based on Newton's laws of motion [29]. At the beginning, the initial positions and velocities of all molecules are set. In the next step the forces are calculated for each atom based on the interaction potential, eq. 1.1. When the forces are known, the acceleration can be calculated for each atom, eq. 1.2. And then, the new positions and velocities are calculated from the acceleration, eqs 1.3 and 1.4 respectively. The whole procedure is then repeated.

An important problem of molecular dynamics is the definition of potentials between atoms [30]. The most detailed are all-atom (AA) force fields, which describe molecules with atomistic resolution. Because the calculations for AA models take o lot of time, AA simulations are limited to small systems and short time-scales.

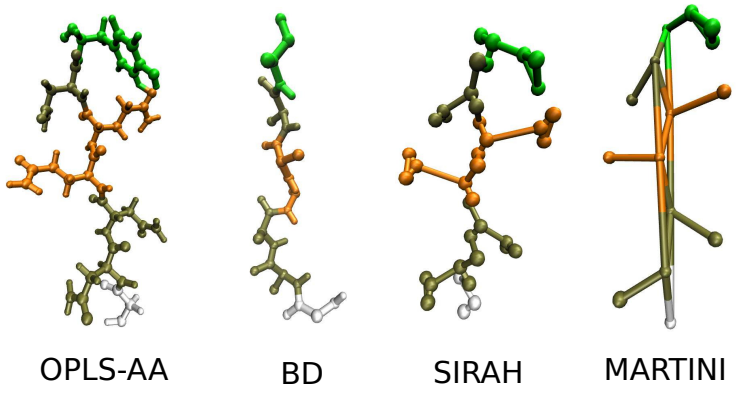

Figure 1.3: GNNQQNY peptide presented for four different models (force fields): one all-atom - OPLS-AA [33] and three coarse-grained: Bereau-Deserno abbreviated to BD [34], SIRAH [35] and MARTINI [36] models. Colors represent the residues, white - glycine (G), brown - asparagine (N), orange - glutamine (Q) and green - tyrosine (Y).

Moreover, simulations at high resolution level give a lot information which can hinder understanding of interesting process or system properties [30]. Coarse-grained (CG) models are commonly used when larger system and longer times are needed. CG approach induces smothing of free energy landscape, which protects systems from thermodynamical traps [31]. In CG models, several atoms or whole molecules are replaced by one super-atom with average potential. Figure 1.3 presents the GNNQQNY peptide in four representation, an all-atom and three coarse-grained models. A disadvantage of CG models is that important information can be lost. Indeed, the CG models are not able to capture the process undergoing below the size of the average radius of super-atoms [30]. Additionally, the coarse-grained degrees of freedom may be affected by the fine-grained degrees of freedom, which are present in real systems and omitted in CG approach [32]. These limitations require carefully consideration of the structural levels which are involved in process under investigation. However, CG models might give the reliable outcomes for the size and time scales unavailable for all-atom simulations [32], thus they provide an unique opportunity to study biological systems and processes.

1.3 The aim of the research

The research presented in this thesis deals with protein and peptide aggregation. Specifically, this thesis focuses on short peptides and aims at defining the key features responsible for aggregation propensity. Equilibrium studies give important information about the thermodynamic properties, such as the stability and structure of final aggregates. This thesis includes not only the structural analysis of the final and intermediate species but addresses also the aggregation kinetics to unravel the underlying molecular mechanisms. The main questions addressed here focus on the factors controlling aggregation:

- Are there any generic properties of peptide aggregation? Can we define the behaviors common for broad classes of peptides and point out the features which are sequence sensitive?
- What does control the aggregation rate? Can the backbone or side chains properties accelerate or inhibit the self-assembly?
- Which peptide attributes are the main determinants of the kinetic mode? Are the backbone or side chains more important from the kinetic point of view?

Moreover, this work tries to correlate the kinetics with molecular events and addresses the question of the nature of the oligomeric intermediates, which are currently under intense investigation due to their potential toxicity [14, 15].

All of these issues contribute to the question about the origin of the self-assembly diversity. Especially, we search for molecular events in the initial aggregation steps leading to the structural polymorphism and kinetic multiplicity observed *in vivo* and *in vitro*.

These goals are realized using molecular dynamics simulations with two different coarse-grained models: Minimal Model developed in our group and the sequence-specific model proposed by Bereau and Deserno [34]. Accordingly, the results are presented in two parts.

The first part focuses on the non-specific Minimal Model, described in our publications, P1 and P3. This simple model has only one superatom per residue, implying the backbone controlled aggregation. Such approach is based on the suggestion, that the ability to form amyloids is a generic feature, i. e. connected with backbone rather than side chains interactions [7]. Aggregation of a simple chain is studied as a function of the chain parameters: the chain stiffness and the inter-molecular interaction strength. These features determine the aggregate morphology and aggregation kinetics. Minimal Model addresses the question: is it possible to reproduce the variety

of aggregates structures and aggregation kinetics observed in experiments only by modifying the backbone properties? If so, which of the chain parameters is a key for the aggregation mechanism?

In the second part, referring to publication P2, a residue-specific model is used to study effect of residue substitutions on aggregation. Because the amino acids residues differ in side chains, these studies focus on the role of the side chain groups in peptide aggregation. The peptide chosen to investigate fibril formation is a heptapeptide fragment from the yeast prion protein Sup35 [21]. The aggregation kinetics and final aggregates are compared with the results obtained for the single-residue mutant of this heptapetide to elucidate the side chain effect on peptide aggregation. The heptapeptide aggregates in experiments but the internal structure of the fibrils is not entirely resolved [27]. Molecular dynamics simulations of the early aggregation steps may indicate which of the proposed fibril models is most realistic. Nevertheless,

Chapter 2

Methodology

This part presents essential technical details of the simulations and data analysis in our publications, P1-P3. In the first section, the two coarse-grained models used in this thesis are introduced: Minimal Model and the Bereau-Deserno peptide model. The simulations protocol and systems details are also presented. In the second section, the data analysis methods are listed and the relevant descriptors are defined.

2.1 Molecular dynamic simulations

2.1.1 Minimal Model

Model details

An implicit solvent model called Minimal Model was described in detail in publication P1. This section presents only the main features of Minimal Model.

Minimal Model has only one super-atom per residue, see fig 2.1, which are bound via the harmonic bond potential:

$$V_{\mathrm{b}}\left(r_{ij}\right) = \frac{1}{2} k_b \left(r_{ij} - r_0\right)^2 \tag{2.1}$$

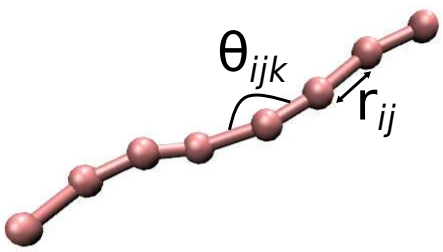

Figure 2.1: A homo-octapeptide in Minimal Model. The symbols r_{ij} and θ_{ijk} represent bond lengths and angles, respectively.

where i, j are two consecutive super-atoms, k_b is the force constants equal to 1250 kJ mol^{-1} nm^{-2}, and r_0 is the equilibrium bond length equal to 0.35 nm.

The second bonded interaction is the cosine based angle potential defined for three consecutive super-atoms, i, j, k:

$$V_\alpha\left(\theta_{ijk}\right) = \frac{1}{2} k_\theta \left[\cos\left(\theta_{ikj}\right) - \cos\left(\theta_0\right)\right]^2 \tag{2.2}$$

where θ_0 is the equilibrium bond angle equal to 180°. The angle force constant k_θ is variable and takes the values: 10, 100, 200, 400, 800, and 1000 kJ/mol.

The non-bonded potential includes only the Lennard-Jones (LJ) interaction potential:

$$V_{\mathrm{LJ}}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{2.3}$$

where r is the distance between two non-bonded super-atoms, ε is the depth of the potential well, and σ is the distance at which the LJ potential is equal to zero. In our simulations σ is fixed at 0.47 nm whereas ε is variable and takes the values: 1.0, 1.3, 1.4, 1.5, 1.6, 1.7, 2.0, 3.0, 4.0 kJ/mol.

Simulation protocol

GROMACS 4.6.5 [37] package was used for the simulations with the leap-frog stochastic dynamics (sd) integrator. The time step was 25 fs. The simulations were performed in the canonical ensemble with the temperature kept at 303 K using the GROMACS sd integrator as the thermostat. Due to the absence of explicit solvent molecules, an inverse friction coefficient of 0.17 ps was applied to mimic the friction effect of solvent. The simulation times were between 4-10 μs.

The simulations were initially conducted for octapeptides, i.e. for the model peptides with 8 super-atoms per chain, SA = 8. The base initial monomer concentration, c_0, was equal to 2.8 mM, corresponding to the super-atoms concentration $c_{\mathrm{SA}} = 22.3$ mM. The simulated systems contained 72 peptides, $N_0 = 72$, in a cubic box with edge $L = 35$ nm. A detailed simulation list is presented in the table 2.1. Initially, all simulations were 4 μs long. For better estimation of the kinetic parameters, some simulations were extended to 10 μs and the simulations were repeated 5 times. Additional repeats (up to 25) were performed for systems with nucleated aggregation.

Additional simulations were performed to investigate the impacts of concentration and of chain length on aggregation. The concentration effect was studied for an octapeptide, SA=8, with the interaction strength $\varepsilon = 1.6$ kJ/mol and the chain stiffness $k_\theta = 200$ kJ/mol for the base concentration $c_0 = 2.8$ mM and two lower concentrations $c_0 = 1.4$ and $c_0 = 0.7$ mM.

Table 2.1: Details of the Minimal Model simulations for octapeptides, SA=8, at the base monomers concentration, $c_0 = 2.8$ mM, corresponding to $N_0 = 72$ peptides and the box edge $L = 35$ nm.

ε	k_θ [kJ/mol]					
[kJ/mol]	10	100	200	400	800	1000
1.0	4 μs	4 μs	4μs	4μs	4μs	4μs
1.3	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s
1.4	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$25 \times 10\mu$s	$5 \times 10\mu$s	$25 \times 10\mu$s
1.5	$5 \times 10\mu$s	$25 \times 10\mu$s	$5 \times 10\mu$s	$25 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s
1.6	$25 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s
1.7	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s
2.0	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$5 \times 10\mu$s	$25 \times 10\mu$s
3.0	4μs	4μs	4μs	4μs	4μs	4μs
4.0	4μs	4μs	4μs	4μs	4μs	4μs

The influence of the chain length was investigated for two longer peptides: SA=16 and SA=20. The super-atom concentration was kept constant, $c_{\mathrm{SA}} = 22.3$ mM. The details about systems sizes and values of ε and k_θ are presented in table 2.2.

Table 2.2: Details of the Minimal Model simulations for longer peptide chains.

SA	L [nm]	N_0	ε [kJ/mol]	k_θ [kJ/mol]	Time and repeats μs
16	44	72	1.0	1000	5×10 μs
16	44	72	1.4	1000	5×10 μs
20	40	45	2.0	1000	4 μs

For octapeptide with $\varepsilon = 2.0$ and $k_\theta = 1000$ kJ/mol additional simulations were performed for various system sizes: $N_0 = 9, 18, 45, 72, 100, 200$. Size of simulation box was setting in such a way, that the peptide concentration was constant, $c_0 = 2.8$ mM. The simulations time is 4 μs and each simulation was repeated three times.

2.1.2 Bereau-Deserno peptide model

Model description

The second coarse-grained model used in this work, developed by Bereau and Deserno (BD) [34], is more detailed than Minimal Model. This implicit solvent model maps a single amino acid by three or four super-atoms: three for the backbone and

one for the side chain except for glycine that is mapped only by three backbone beads.

The bonded interactions are defined by the bond and angle harmonic potentials with large spring constants to keep the coordinates close to the potential minimum. The bonded interactions include also proper and improper dihedral potentials, which mimic the chirality of C_α carbon. The backbone dihedrals, ϕ and ψ, are additionally modified to include the dipole interaction between carbonyl and amide backbone groups.

The non-bonded interactions are defined by three potentials, the Weeks-Chandler-Andersen potential for the local excluded volume of backbone super-atoms, the Lennard-Jones potential with additional attractive part and the Lennard-Jones potential modified with an angular term to mimic hydrogen bonds.

Simulation protocol

Simulations were performed in the Espresso 3.3.1 package [38] using Langevin dynamics. The temperature was kept at 300 K. The time step was 0.01 τ and the friction coefficient was τ^{-1}, where the derived time unit $\tau \approx 0.16$ ns. The simulation length was 10^9 steps and each simulation was repeated four times.

This model was used to study aggregation of the heptapeptide fragment from yest prion protein, GNNQQNY abbreviated as GNN, and its point mutation, GNNQQNA abbreviated as Y7A. The simulations were conducted for systems with 72 peptide molecules initially randomly placed in a cubic box with edge length equal to 20 nm, giving the initial monomers concentration equal to 15 mM.

Table 2.3: Details of the BD simulations for the GNN peptide.

c_0 [mM]	$N_0 = 20$ L [nm]	$N_0 = 72$ L [nm]
4	19.57	30.00
8	16.31	25.00
15	13.05	20.00
35	9.79	15.00

Additionally, the concentration effect was investigated for the wild peptide, GNN. The concentration was changed in the range 4-35 mM, with a constant number of peptides in system, $N_0 = 72$, and a variable lengths of box edge, as it is presented in the third column of table 2.3.

The simulations of GNN aggregation were also conducted in a smaller system,

$N_0 = 20$, at the same concentration range, $4 \leqslant c_0 \leqslant 35$ mM. The simulation box sizes, L, for $N_0 = 20$ and $N_0 = 72$ are presented in table 2.3.

2.2 Analysis methods

2.2.1 Descriptors

The peptide aggregation was characterized by several descriptors.

1. The instantaneous number of free monomers, N_{m}, the average number of free monomers, $\overline{N}_{\mathrm{m}}$, calculated over all simulation repeats, and the scaled number of free monomers, $\overline{N}_{\mathrm{m,sc}}$, defined as:

$$\overline{N}_{\mathrm{m,sc}} = \frac{\overline{N}_{\mathrm{m}} - \overline{N}_{\mathrm{m,eq}}}{\overline{N}_{\mathrm{m,lag}} - \overline{N}_{\mathrm{m,eq}}} \tag{2.4}$$

   where $\overline{N}_{\mathrm{m,eq}}$ and $\overline{N}_{\mathrm{m,lag}}$ are the average numbers of monomers in the equilibrium phase and in the lag phase, respectively.

2. The instantaneous number of clusters, N_{c}, and the average number of clusters, $\overline{N}_{\mathrm{c}}$, calculated over all simulation repeats.

3. The size of the largest cluster in the system, M_{max}, defined as a number of peptides in this cluster, the average size of the largest cluster in the system, $\overline{M}_{\mathrm{max}}$, calculated over all simulation repeats, and the relative size of the largest cluster in the system, $M_{\mathrm{rel}} = M/N_0$ where N_0, is the total number of peptides in system.

4. The average nucleation time, τ^*, was estimated by two methods:

   - Change point (CP) analysis, τ^*_{CP}. The linear regression was used for monomer kinetic curves and a large change of the slope point out the end of the nucleation phase.
   - The mean first-passage times (MFPT) analysis proposed by Wedekind et al. [39] and extended by Yi et al. [40]. In this method the average first-arrival time τ of cluster with size m is estimated as:

$$\begin{aligned} \tau(m) = 0.5\tau^*_{\mathrm{MFPT}} \left[1 + \mathrm{erf}\left(Z\sqrt{\pi}\,(m - m^*_{\mathrm{MFPT}})\right)\right] \\ +0.5G^{-1}\,(m - m^*_{\mathrm{MFPT}})\,[1 + \mathrm{erf}\,(C\,(m - m^*_{\mathrm{MFPT}}))] \end{aligned} \tag{2.5}$$

     where m^*_{MFPT} is the critical size of nucleus, Z is the Zeldovich factor, and G is the growth rate.

5. The critical size of nucleus, m^*, defined by the number of peptides in the nucleus and estimated by two methods:

   - The mean first-passage times (MFPT), see eq. 2.5, where m^*_{MFPT} indicates the critical size of nucleus.
   - The transition probability matrix (TPM), estimating the probabilities of changes of the largest cluster size, $m = M_{\max}$. The transition between states are counted and normalized. The matrix allows for calculating of the growth probability, $P_{\mathrm{growth}}(m)$, defined as $P_{\mathrm{growth}}(m) = P_{\mathrm{f}}(m) - P_{\mathrm{b}}(m)$, where $P_{\mathrm{f}}(m)$ is the forward transition probability, i.e. $m \to m+1$, and $P_{\mathrm{b}}(m)$ is the backward transition probability, i.e. $m \to m-1$. The critical nucleus size from TPM, m^*_{TPM}, is taken as the smallest aggregate size with $P_{\mathrm{growth}}(m) \geqslant 0$.

6. The asphericity, b, defined as

$$b = \lambda_z^2 - (\lambda_x^2 + \lambda_y^2)/2 \tag{2.6}$$

   where $\lambda_x, \lambda_y, \lambda_z$ are the principal moments of the gyration tensor and the axes are chosen such that $\lambda_x^2 \leqslant \lambda_y^2 \leqslant \lambda_z^2$. The average asphericity $\overline{b}$ is a function of the cluster mass M, $\overline{b}(M)$, and is calculated over all frames and all simulation repeats, where an aggregate of size M is formed.

7. The radius of gyration of a cluster, R_{g}, and the average radius of gyration as a function of cluster mass, $\overline{R}_{\mathrm{g}}(M)$, calculated over all frames and all simulation repeats, where an aggregate of size M was formed.

8. The end-to-end correlation parameter introduced by Zierenberg and co-workers [41]:

$$C_{\mathrm{n}} = \frac{2}{M(M-1)} \sum_{i<j} (\mathbf{n}_i \cdot \mathbf{n}_j)^2 \tag{2.7}$$

   where the unit vector $\mathbf{n}_i$ is the normalized end-to-end vector for the peptide i backbone atoms. C_{n} is normalized in such a way that for uncorrelated $\mathbf{n}_i$ vectors the parameter assumes the value 1/3 whereas for correlated $\mathbf{n}_i$-s it tends to 1.

9. The cluster twist, q, defined as the average peptide chain twist in a cluster, $\cos(\tau_i)$:

$$q = \overline{\cos(\tau_i)} = \frac{1}{M} \sum_i (\mathbf{t}_i \times \mathbf{b}_i) \cdot \mathbf{A} \tag{2.8}$$

where $\mathbf{A}$ is vector along the cluster axis, $\mathbf{t}_i$ and $\mathbf{b}_i$ are the unit vectors for top and bottom ends of chain i, respectively, M is the cluster mass equal to the number of peptides in the cluster. This descriptor was introduced in publication P1.

10. The fibril helicity, H, which is the average ribbons helicity H_{rib} over all ribbons:

$$H = \frac{1}{n_{\text{rib}}} \sum_{\text{ribbons}} H_{\text{rib}} \tag{2.9}$$

where n_{rib} is the number of ribbons, and the helicity of single ribbon with size M_{rib} is calculated as an average:

$$H_{\text{r}} = \frac{1}{M_{\text{rib}} - 1} \sum_{\text{ribbon}} h_{i,i+1} \tag{2.10}$$

of the twists between pairs of consecutive peptides, $i, i+1$, along the ribbon:

$$h_{i,i+1} = \hat{\mathbf{g}}_i \cdot \hat{\mathbf{e}}_{i+1} \tag{2.11}$$

The unit vectors, $\hat{\mathbf{g}}_i$ and $\hat{\mathbf{e}}_j$ are calculated as:

$$\hat{\mathbf{g}}_i = \frac{\vec{\mathbf{c}}_i - \vec{\mathbf{x}}_i}{|\vec{\mathbf{c}}_i - \vec{\mathbf{x}}_i|} \tag{2.12}$$

$$\hat{\mathbf{e}}_j = \frac{\vec{\mathbf{c}}_j \times \hat{\mathbf{A}}}{|\vec{\mathbf{c}}_j \times \hat{\mathbf{A}}|} \tag{2.13}$$

where $\vec{\mathbf{c}}_i$ and $\vec{\mathbf{c}}_j$ are vectors pointing to the peptide mass center of peptides i, j, $\hat{\mathbf{A}}$ is a vector along cluster axis and $\vec{\mathbf{x}}_i$ is a vector pointing to the projection of the mass center of peptide i on the cluster axis.

The helicity H was introduced in publication P2. H is equal to 0 for achiral structures, is less than zero for left-handed fibrils and greater than zero for right-handed fibrils.

2.2.2 Analysis software

The following software was used for data analysis and visualization:

- MDAnalysis python package, version 0.20.1 [42, 43], was used for the analysis of MD simulations, especially the calculation of the structural parameters such as asphericity or radius of gyration.
- In-house python scripts for new analyses methods including identification of aggregates and the proposed chirality and helicity descriptors.
- Grace version 5.1.25 [44], was used for data visualization.
- VMD, versions 1.9.2 and 1.9.3 [45], was used for trajectory visualization.
- GROMACS tools, version 4.6.5 [37], were used in some analysis, for example the calculation of maximal cluster size in time.

All simulations and analysis were run in Linux environments, Ubuntu 16.04 and Ubuntu 18.04.

Chapter 3

Results

This chapter, presenting briefly the results of our aggregation simulations, is divided into two sections. The first section describes the outcomes for Minimal Model developed in our group. Those results were presented in full detail in our two publications, P1 and P3. Both publications, P1 and P3, focus on the changes in aggregation induced by two backbone parameters: the inter-molecular interaction strength, ε, defined by the Lennard-Jones potential, eq. 2.3, and the chain stiffness defined by the angle constant, k_θ, from the cosine angle potential, eq. 2.2.

The second section focuses on the aggregation studies of the GNNQQNY (GNN) heptapeptide from the yeast prion protein Sup35, and its point mutation Y7A, where the terminal tyrosine was replaced by alanine. The simulations were carried out using the coarse-grained model proposed by Bereau and Deserno. Those simulations were described in detail in our publication P2, and only the most interesting issues are brought out in this chapter.

3.1 Minimal Model

The first aspect of aggregation studied by Minimal Model is the variability of aggregate structures, P1 and P3. This problem was discussed in both publications for slightly different ranges of the interaction strength, ε. Figure 3.1 shows a schematic structural phase diagram, basing on data from both, P1 and P3, publications. Publication P1 gives a more general view, as the interaction strength, ε, takes a broader range of values, 1, 2, 3 or 4 kJ/mol. For this range of ε and for the angle constant, k_θ, varying from 10 to 1000 kJ/mol, the structural phase diagram was obtained. For the lowest interaction strength, $\varepsilon = 1$ kJ/mol, peptides do not aggregate for all values of the k_θ. For higher values of ε the aggregates change from amorphous (at low interaction strength and low chain stiffness, left bottom corn on phase dia-

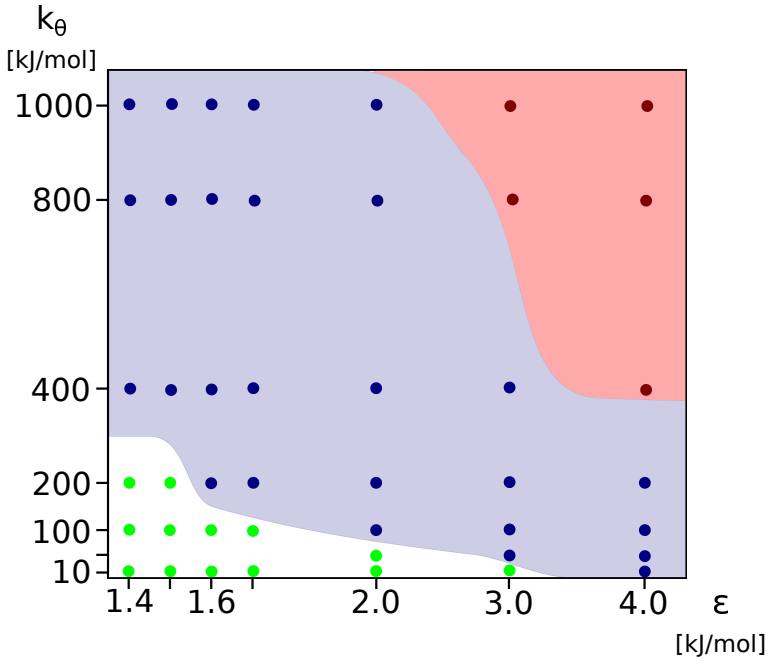

Figure 3.1: Structural phase diagram from Minimal Model simulations.

gram, fig.3.1), to more ordered and then to crystal-like structures observed for the highest values of both parameters, $\varepsilon = 4$ and $k_\theta = 1000$ kJ/mol, the upper right corn on phase diagram, fig.3.1. Publication P3 takes a look at systems with the interaction strength between 1 and 2 kJ/mol, i.e. between no-aggregation and aggregation regimes. These simulations give a more detailed insight into the transition between amorphous and ordered aggregates. However, in this parameter range the crystal-like structures are not observed.

An important result from our Minimal Model simulations is the discovery of fluctuations of aggregate chirality. We found that in ordered aggregates the peptide chains twist collectively around the cluster core. Due to the lack of an intrinsic peptide chirality in Minimal Model, the clusters twist to the right and to the left with the same probability. This phenomenon was carefully investigated for a system with the medium interaction strength, $\varepsilon = 2$ kJ/mol, and high chain stiffness, $k_\theta = 1000$ kJ/mol. For the quantitative description, we proposed a cluster twist parameter, q, eq. 2.8, which measures the average twist of all peptides around the cluster axis. The positive and negative values of q correspond to the left and right twist, respectively. q is equal to zero for a perfectly parallel arrangement of peptides. The example trajectory of twist descriptor is presented in fig. 3.2 with the corresponding aggregate structures.

This descriptor supports quantitatively a qualitative visual observation, that the average life-time of the twisted states decreases with aggregate growth, from $\tau \approx 20$ ns for aggregate with size $M = 20$ to $\tau \approx 2$ ns for $M = 41$ and then

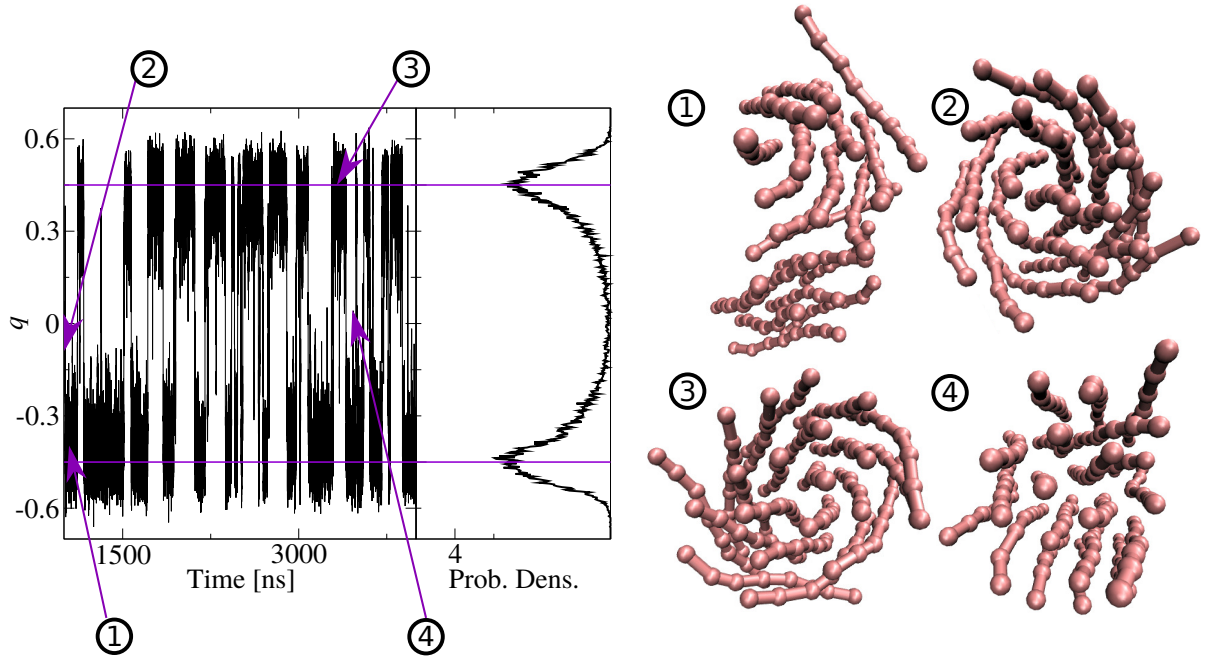

Figure 3.2: Trajectory of the twist descriptor, q, for $\varepsilon = 2$ kJ/mol, $k_\theta = 1000$ kJ/mol, together with the histogram of the cluster twists. The corresponding structures are shown on the right side.

to $\tau \approx 0.7$ ns for $M = 68$. The amplitude of cluster twists also changes with the cluster mass. The amplitude riches a maximum for clusters of size $M \approx 20$ peptides, $q \approx \pm 0.45$, and then decreases. For larger aggregates, $M = 41$, three states are visible: a left- and a right-handed with average twist $q \approx 0.16$ and -0.16, respectively, and also an achiral state with $q \approx 0.0$. With an increase in the aggregate size, the achiral state becomes much more stable than the twisted states and the average twist descriptor fluctuates around zero.

The stability of twisted states increases significantly for longer peptide chains. During aggregation of peptides with 20 super-atoms per chain, SA=20, the aggregates of size $M = 18$ show the twisted states with the life-time in order of hundreds nanosecond, up to 1 μs.

Our simulations for systems with the interaction strength between $1.3 \leqslant \varepsilon \leqslant 2$ kJ/mol and the chain stiffness, $10 \leqslant k_\theta \leqslant 1000$ kJ/mol revealed three types of aggregation kinetics. These three kinetic modes can be presented on a kinetic phase diagram as a function of model parameters, ε and k_θ, see fig. 3.3. The corner with the low values of both parameters is occupied by the no-aggregation mode, the opposite corner relates to the downhill aggregation and the narrow region between them represents the nucleated aggregation.

Two parameters were measured to characterize nucleation aggregation: the nucleus size, m^* and the average nucleation time, τ^*. Both descriptors were defined by

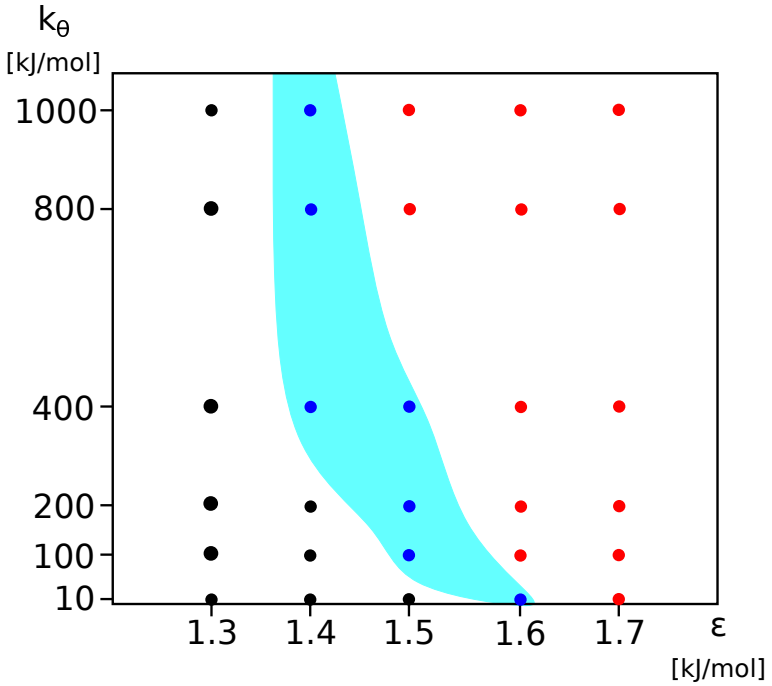

Figure 3.3: Kinetic phase diagram for Minimal Model simulations.

the MFPT, CP and TPM methods described in Methodology.

Table 3.1: Average nucleation times τ and nucleus sizes, m^* for systems with various ε [kJ/mol], and k_θ [kJ/mol]. The subscripts indicate analysis methods: CP - the change point analysis, MFPT - the mean first-passage time method, and TPM - the transition probability matrix. If aggregation did not occur in all simulations, the subscript for the average nucleation time from CP analysis, τ^*_{CP}, includes also information about the fraction of repeats with aggregation.

	$\varepsilon = 1.4$ $k_\theta = 400$	$\varepsilon = 1.4$ $k_\theta = 1000$	$\varepsilon = 1.5$ $k_\theta = 100$	$\varepsilon = 1.5$ $k_\theta = 400$	$\varepsilon = 1.6$ $k_\theta = 10$
m^*_{TMP}	12	8	7	8	10
m^*_{MFPT}	7.84	8.12	6.87	5.74	9.72
τ^*_{CP} [ns]	$\tau^*_{\mathrm{CP},10/25} =$ 3500 ± 3100	2700 ± 2500	640 ± 540	250 ± 270	$\tau^*_{\mathrm{CP},20/25} =$ 3300 ± 2800
τ^*_{MFPT} [ns]	4200	2700	710	270	3500

The nucleation time differs significantly for various simulation repeats of the same system, which is seen as a high standard deviation for the estimated average nucleation time, see table 3.1. Nevertheless, the average nucleation time shows clear trends when the model parameters are changed. For instance, for systems with interaction strength $\varepsilon = 1.4$ kJ/mol, the average nucleation time decreases with increasing chain stiffness. Also a comparison of systems with the same chain stiffness

shows that the average nucleation time decreases for higher interaction strength. Such trends are not observed for the nucleus size, regardless of the method used for its estimation.

Despite differences in the length of the nucleation phase and the size of nucleus, the overall kinetics is the same for all systems with nucleated aggregation. The kinetic curves for the scaled monomer decay overlap. Similarly, the kinetic curves for the maximal cluster size overlap. This suggests the same aggregation mechanism. Indeed, for all nucleated systems, the aggregate growth undergoes mainly by addition of single monomers, independently of the aggregate structure.

On the other hand, the scaled kinetic curves for downhill systems do not overlap. Aggregation undergoes mainly by monomer addition for systems close to the nucleation regime. When the interaction strength or chain stiffness increases, the monomers decay faster to form several stable oligomers in system. Then, the further growth of the largest aggregate undergoes by the cluster-cluster coalescence.

Experimental studies show, that the transition between downhill and nucleated aggregation may be caused by a decrease in concentration. Our model reproduces this behavior. The concentration effect was investigated for system with $\varepsilon = 1.6$ kJ/mol, and $k_\theta = 200$ kJ/mol which shows downhill aggregation at our base concentration, $c_0 = 2.8$ mM. The two-fold reduction in concentration, $c_0 = 1.4$ mM, causes the transition from downhill to nucleated aggregation. Further decrease in concentration to $c_0 = 0.7$ mM, leads to elongation of the nucleation phase, and aggregation does not occur in all repeats suggesting that this concentration is close to the no-aggregation mode. The corresponding times of nucleation are equal to $\tau^*_{\mathrm{CP}} = 360 \pm 320$ ns and $\tau^*_{\mathrm{CP},3/5} = 2500 \pm 1700$ ns for concentrations $c_0 = 1.4$ and $c_0 = 0.7$ mM respectively.

The aggregation propensity and the stability of aggregates increase for longer peptide chains. This effect was studied by changing the peptide chain from 8 super-atoms, SA=8, to 16 super-atoms per chain, SA=16. The super-atom concentration was constant, $c_{\mathrm{SA}} = 22.3$ mM. The scaled monomer and cluster kinetic curves do not overlap with the curves for shorter peptides with the same interaction strength, $\varepsilon = 1.4$ kJ/mol, and chain stiffness, $k_\theta = 1000$ kJ/mol. The nucleation phase, observed for aggregation of the shorter peptide, disappears for longer peptide which aggregates with downhill kinetics. Moreover, the longer peptides form initially more oligomers and the further aggregates growth undergoes mainly by cluster-cluster coalescence. This behavior is similar to the short peptide downhill systems with higher interaction strength, $\varepsilon = 2.0$ kJ/mol.

Another difference in long and short peptides aggregation is observed in the

final monomer concentration. The super-atom concentration of free monomers at equilibrium, for systems with the same interaction strengths, $\varepsilon = 1.4$ kJ/mol, and chain stiffness, $k_\theta = 1000$ kJ/mol, is much lower for the long peptides, $c_{\mathrm{SA,m,eq}} = 1.8 \cdot 10^{-5}$ mM than for shorter peptides, $c_{\mathrm{SA,m,eq}} = 1.2$ mM. Even the short peptides with higher interaction strength, $\varepsilon = 2.0$ kJ/mol, show a higher equilibrium concentration of free monomers, $c_{\mathrm{SA,m,eq}} = 4.3 \cdot 10^{-4}$ mM than that for long peptide system.

However, long peptides can also show nucleated aggregation if the interaction strength is sufficiently low. For example, the nucleation aggregation is observed for long peptides, SA=16, with $k_\theta = 1000$ kJ/mol and $\varepsilon = 1.0$ kJ/mol, so the parameters values for which short peptides, SA=8, does not aggregate at all. The nucleation time for this long peptide system is equal to $\tau^*_{\mathrm{CP}} = 5400 \pm 3200$ ns and the nucleus size defined from MFPT method is $m^*_{\mathrm{MFPT}} = 9.05$ so it is in range of nucleus sizes observed for short peptides, $5.74 \leqslant m^*_{\mathrm{MFPT}} \leqslant 9.72$. This indicates, that the aggregation is faster for longer peptides, but the general mechanism does not change.

3.2 Bereau-Deserno model

In our simulations, both peptides, GNN and Y7A, show ability to form fibrils. Mature fibrils of the wild sequence, GNN, are helical and consist of 5-6 ribbons twisting around the core, see the structures in the fig. 3.4. The fibril core is formed by the tyrosine residues. In a single ribbon, the peptides are arranged parallel to each other and perpendicular to fibril axis. Replacement of tyrosine by alanine leads to the disappearance of the fibril core. Y7A peptides are oriented in parallel in a single sheets, see the structure in the fig. 3.5. Mature fibrils consists of a few sheets which are arranged in parallel or anti-parallel to each others.

Because of these structural differences in mature fibrils, several descriptors were used to measure internal reorganization for GNN and Y7A fibrils. Due to the visible chirality of GNN fibrils, the helicity descriptor, H, eq. 2.9, was developed in our group to describe the structural changes of GNN aggregates, fig. 3.4. On the other hand, almost all Y7A peptides are parallel or anti-parallel, thus the end-to-end correlation parameter, C_{n}, eq. 2.7, captures changes in the internal order, fig.3.5.

The trajectories of these two structural parameters for the largest cluster were compared with trajectories of the largest cluster growth. At the beginning, when only small oligomers exist, the structural parameters fluctuate around the values indication lack of structural order: $H \approx 0$ implies no helicity and $C_{\mathrm{n}} \approx 0.33$ corresponds to a random peptide arrangement. Despite the growth of clusters mass,

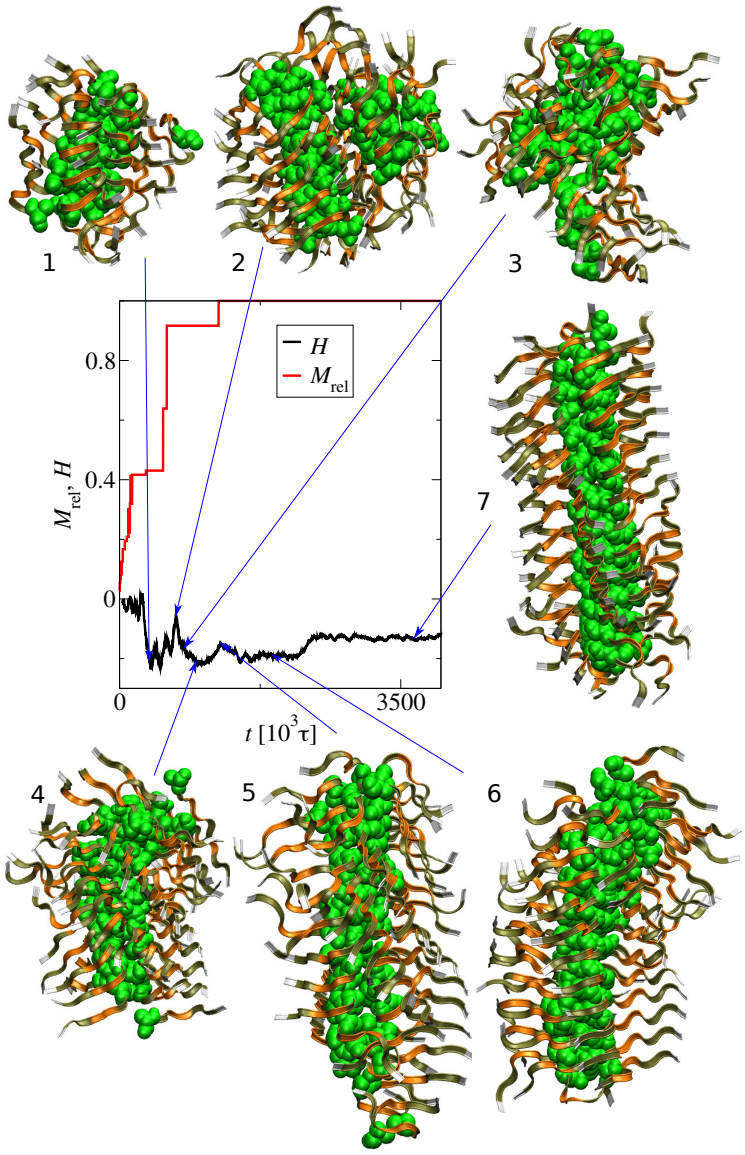

Figure 3.4: Helicity trajectory H, for largest aggregate of the wild peptide, GNN, (black line). The relative size of the largest aggregate, M_{rel}, is presented as a red line. The corresponding structures illustrate the internal reorganization of GNN aggregate. Peptide chains are presented as ribbons except the terminal tyrosine which is marked by green beads.

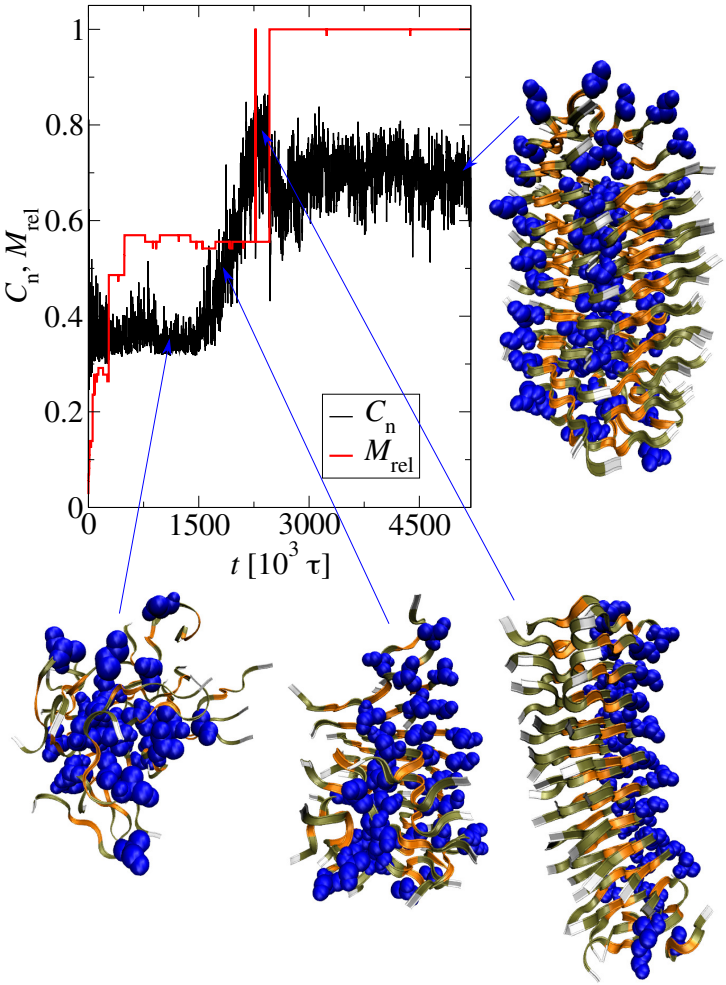

Figure 3.5: Trajectory of the end-to-end correlation parameter, C_n, for the largest aggregate of the mutant peptide, Y7A, (black line). The relative size of the largest aggregate, $M_{rel} = M_{max}/N_0$, is presented as a red line. The corresponding structures illustrate the internal reorganization of Y7A aggregate. Peptide chains are presented as ribbons except the terminal alanine which is marked by blue beads.

the structural parameters stay at their initial values until the aggregate size exceeds the critical value. Then, for sufficiently large aggregates, internal reorganization is observed as a gradual change in structural parameter values. It suggests the same, two-step mechanism for fibril formation for the GNN and Y7A peptides.

The critical aggregate size for fibrillation is confirmed by the histograms of distances between peptide centers of mass in aggregate. For small, disordered oligomers these histograms have only one, wide peak, suggesting no preferred location of amino acid residues within aggregate. The second type of observed histograms has narrow peaks, indicating regular positions of amino acids in the cluster. These histograms are characteristic for large aggregates with fibril-like structures. There is a narrow cluster size interval where we observed two types of histograms, disordered and ordered. For the wild peptide, GNN, this transition region occurs for the cluster sizes between 25-35 peptides, whereas for the mutant, Y7A, it occurs between 32-58 peptides. This is in agreement with the cluster sizes for which the internal reorganization is observed in the structural descriptor plots.

The aggregation of GNN and Y7A peptides differ not only in the size of fibrillation nucleus. The kinetic studies show that the average number of clusters evolve similarly for both sequences, but the monomer kinetic curves are slightly different. The initial decay of monomers is similar for both sequences, however, when the average number of monomers reaches half of the initial value, kinetic curves split up. The monomers of the mutant peptide, Y7A, show slower aggregation and they are not completely aggregated at the end of simulation, whereas the monomer aggregation of the wild peptide is almost total.

On the other hand, the aggregation kinetics for the wild peptide, GNN, does not change noticeably in the studied concentration range, c_0 = 4-35 mM. For all concentrations, the scaled curves for the average number of clusters overlap. The same is true for the scaled curves for monomers. Interestingly, the scaled monomers curves overlap also for different system sizes, $N_0 = 20$ and $N_0 = 72$, at the same concentration. Similarly, the scaled curves for clusters do not vary noticeably between those two system sizes, despite that the final aggregates structures are different. In small systems, $N_0 = 20$, fibrils are not formed at any studied concentration, due to a smaller number of monomers than it is required for fibrillation, $m^* = 25$-35.

Other structural parameters, such as the radius of gyration, R_g, or asphericity, b, when plotted as functions of the aggregate mass, may be also used to define the cluster size regime when the transition between disordered to fibrilar aggregates occurs. An increase of the radius of gyration as a function of aggregate mass shows two slopes according to two types of clusters: small ones which are spherical and

disordered, and the larger, fibrilar structures. For small oligomers the average radius of gyration, $\overline{R}_{\mathrm{g}}$, increases with mass, M, as:

$$\overline{R}_{\mathrm{g}} \propto M^{\alpha} \tag{3.1}$$

where α is a scaling exponent. For larger clusters the average radius of gyration changes as:

$$\overline{R}_{\mathrm{g}} - \overline{R}^{*}_{\mathrm{g}} \propto (M - M^{*})^{\beta} \tag{3.2}$$

Analysis of the radius of gyration plot for GNN gives the scaling exponents $\alpha = 0.30$ and $\beta = 0.88$. These values refer to 3D ($\alpha = 1/3$) and 1D ($\beta = 1$) aggregation mechanisms respectively. Those exponents are slightly smaller than the expected values for 3D and 1D mechanism, indicating that the aggregates grow with increase in density.

Chapter 4

Discussion

This chapter discusses the results of our simulations in the context of the current state of knowledge about peptide aggregation. A direct comparison between simulations and experiments is difficult due to the different time and size scales. Therefore, this discussion focuses on general trends. The review of experimental work indicates gaps in our understanding of peptide aggregation. We addressed some of them in our simulations. Peptide aggregation was previously intensively investigated using molecular dynamics methods with various peptide models. We refer to these studies and point out the differences in our approach as well as the new aspects uncovered by our simulations.

The chapter is divided into two parts according to the peptide models. In the first section Minimal Model is compared to other simple models with only-backbone resolution. Subsequently, we focus on the link between Minimal Model simulations and experiments. New aspects and hypotheses derived from the Minimal Model simulations are also discussed.

The second section focuses mainly on aggregation of the GNN peptide, which has been intensively studied in experiments. In this section, we address the inconsistencies and doubts about the mechanism of GNN aggregation pointed out in previous studies. Additionally, the aggregation diversity of GNN and Y7A peptides is compared with the available all-atom simulations.

4.1 Minimal Model

Minimal Model was used to study aggregation of homopeptides represented by only one super-atom per amino acid. Homopeptides were chosen not only because of their simplicity. Long protein fragments with single amino acid repeats have been found in amyloidogenic proteins, for instance, poly-glutamines fragments are probably re-

sponsible for Huntingtin aggregation in Huntington's disease [46].

Limitation of the peptide representation to only backbone super-atoms was motivated by the hypothesis that the overall aggregation propensity is connected with backbone interactions rather than the side chains, as various peptides may aggregate in similar way [7]. Therefore, our Minimal Model can give insight how the effective backbone properties influence aggregation. Note, that the backbone features in real peptides are modulated by side chains.

There are various coarse-grained models with only-backbone resolution, i.e. with no explicit side chains. The most simple models treat the whole peptide molecule as one entity: a rod [47, 48], a cuboid [49], or a sphere [50]. More flexible models represent peptides as chains of super-atoms, where one super-atom may map several residues [51], or a single amino acid [52, 53, 54, 55].

The common feature of these simple models is an anisotropic non-bonded potential, enabling fibril formation. In contrast, our model does not assume any directional interactions. The potential between non-bonded super-atoms is isotropic, similar to the peptide model proposed by Ranganathan et al. [56]. Our approach avoids predefining a favorable relative orientation of interacting molecules.

These two features of our model, the homo super-atoms and the backbone-only resolution, brings it close to homopeptides, especially poly-glycine, poly-G, and poly-alanine, poly-A. Experimental studies show that both poly-G and poly-A can form fibrils in solution [57, 58]. Also the poly-glycine-alanine, poly-GA, forms fibril, which is probably connected with neurodegenerative diseases, such as amyotrophic lateral sclerosis or frontotemporal dementia [59, 60]. Noteworthy, Bernacki and Murphy [58] showed that the aggregation of poly-A depends on the chain length and temperature. When the temperature increases, the aggregates change from spherical, amorphous oligomers to ordered structures. Minimal Model reproduces this structural transition, from disordered to ordered clusters, although the model uses only an isotropic potential. This transition is also observed for increasing chain stiffness, which corresponds to the persistence length calculated in experiments. The formation of more ordered aggregates for higher k_θ corresponds to the disordered-ordered transition observed between poly-alanine and poly-glutamine accompanied by an increase in the persistence length from 2-3 to 6 Å for poly-A and poly-Q respectively [58]. Similarly, Singh and Lapidus found that the polyglutamine chains are more rigid than other homopeptides and show a tendency to stay in the extended conformation [61]. Our results support their hypothesis that the intrinsic stiffness is responsible for a high tendency to aggregation.

The persistence length of peptide chains can increase also as a result of phospho-

rylation [62]. Kumar et al. show that the phosphorylated Aβ peptides are more prone to aggregation than the unphosphorylated ones [63, 64]. Phosphorylation stabilizes extended conformations, thereby facilitating the nucleus formation and reducing the length of the lag phase. Interestingly, the morphology of the final aggregates is the same for the phosphorylated and unphosphorylated Aβ proteins, despite the various kinetic behavior.

Experimental studies show that the various kinetics may induce formation of different aggregates, from ordered to disordered [65], or may lead to the same aggregate structures, as in the case of phosphorylated/unphosphorylated Aβ peptides [63]. Our simulations with Minimal Model confirm it, as the kinetic and structural diagrams overlap only in part, i.e. the borders between various kinetics modes are not the same as the disordered-nematic-crystal transition lines.

For the nucleated kinetics discussed in the literature, the nucleus size was determined in some cases. For homopeptides, numerous studies focus on poly-glutamine, and indicate, that the nucleus size decreases from ≈ 4 to $\leqslant 0.5$ when the number of glutamine residues increases from 18 to 47 [66, 67, 68]. Coarse-grained simulations gave similar results, as the critical nucleus decreases from 3-4 for Q_{20} to 1 for Q_{30} [69].

Our studies of the nucleating systems give the critical nucleus size in range $m^* =$ 7-12 for octapeptides. The nucleus size does not clearly depend on the interaction strength, chain stiffness or chain length, unlike the nucleation time which decreases with increasing ε and k_θ. The lack of visible trends in the critical nucleus size can be caused by a small number of simulation repeats.

A new aspect revealed by Minimal Model is the chirality fluctuation of peptide aggregates. As it was described in Introduction, chirality of proteins and peptides assembles is not a simple outcome of the amino acids chirality. On the one hand, the origin of the diversity in supramolecular chirality is connected with the higher hierarchical levels, i. e. differences in the fibrils twist are caused by the specific interaction between regular β-sheets or protofibrils in mature fibrils [6]. On the other hand, the fibril twist may result from the interplay between peptide building blocks and, as a consequence, the formation of right-handed, or left-handed β-sheets [8]. The second option has been recently examined by Wang and co-workers [70]. Specifically, they studied the effect of glycine insertion as it is the only achiral molecule among the 20 biological amino acids. They tested the aggregation of a model peptide, Ac-I_3K-NH_2, and its glycine modification, Ac-I_3GK-NH_2, and found that Ac-I_3K-NH_2 forms only the left-handed fibrils. The addition of achiral glycine residues leads to a racemate with an excess of the right-handed fibrils. The experimental studies

were supported by combined MD and QM calculations, indicating that there are two free energy potential minima for Ac-I_3GK-NH_2 β-strands, corresponding to the left- and right-handed arrangements. The difference between these minima is small. These results are very close to our observation from Minimal Model, that the achiral peptides can spontaneously form a racemate of left- and right-twisted structures. In publication P2, we postulated that the chiral fluctuations of initially formed small oligomers can be responsible for the helical diversity observed in mature fibrils.

In the case of Minimal Model, the molecules are completely achiral, whereas in real peptides, except glycine, they have intrinsic chirality leading to at least some chirality preference. Our simulations show, that the frequency of transitions between left- to right-twisted states increases with the cluster mass. This implies for real peptides, that for sufficiently large nucleus size, the fluctuations are rapid, and the thermodynamically more stable conformation can be reached leading to preferred helicity of fibrils. For a small nucleus, the growth may be faster than chirality fluctuations, so the structure can be kinetically trapped in a random state. For many nuclei this leads to a racemate.

The next point is the biological significance of fibril twist. As it was discussed in literature, the fibril twist can control the specific fibril thickness [71]. This is based on theoretical calculations indicating, that the stability of the twisted state dependents on the fibril thickness, and decreases when fibrils become wider. The twist stability decays for both achiral and chiral systems, although for an achiral system the decay is faster. Those theoretical calculations are supported by our Minimal Model simulations, as the amplitude of cluster twist decreases with the cluster mass.

4.2 Bereau-Deserno model

The BD model was used to study aggregation of two peptides whose sequences differ by one amino acid residue, GNNQQNY (GNN) and GNNQQNA (Y7A). The BD model is an implicit solvent model that is more detailed than Minimal Model, as it includes explicit side-chain super-atoms allowing to study the sequence effects. The BD model was previously used to study aggregation of various peptides, such as: polyglutamine [72], fragments of the tau protein [73] and GNN [74, 75].

Our simulations show a two-step aggregation mechanism for both peptides, GNN and Y7A. The mechanism does not change in the studied concentration range, $4 \leqslant c_0 \leqslant 35$ mM for the GNN peptide. Such two-step kinetics has not been reported in the experiments. Experimental studies for GNN at lower concentration range, $\approx$ 0.7-

1.5 mM, point to the one-step mechanism [21]. The one-step mechanism of GNN aggregation was also observed in simulations with BD model but at higher GNN concentrations. Luiken and Bolhuis conducted simulations for small systems at initial monomers concentration equal to $c_0 = 0.25$ M, and they found the temperature-dependent transition of aggregation behaviour which occurs at temperature ≈ 275 K [74]. Below this critical temperature, the formation of β-sheets undergoes by one-step kinetics, whereas above this temperature, aggregation does not occur. They also observed such one-step aggregation mechanism for a slightly lower concentration, $c_0 = 0.07$ M and higher temperature, 285 K and they found the size of critical nucleus $m^* =$ 4-6 peptides [75]. In comparison, our simulations are conducted for lower concentration range, $c_0 =$ 4-35 mM and at higher temperature, 300 K. It suggests, that transition from one to two-step kinetics occurs with decreasing concentration and increasing temperature, what is in contradiction with the hypothesis of Auer and co-workers [16]. On the other hand, this disagreement between Luiken's and our simulations can be caused by the difference in system size, as small systems are more prone to the artificial behavior [76, 77]. This unexpected tendency might be the subject of further research on GNN aggregation.

On the other hand, the nucleated conformational conversion mechanism was observed for the whole yeast prion protein, Sup35 [18], as it was described in Introduction. In this case, the initial step is formation of micelle-like oligomers [18], which correspond to our spherical intermediates.

Our simulations also address the question of the internal structure of GNN fibrils, as described in Introduction, where two competing models were discussed. A common feature of those two fibril models is the parallel, in-register alignment of peptides in single β-sheet [24, 25, 27]. Our simulations confirm that alignment. However, we do not observe anti-parallel pairs of β-sheets interacting via steric-zipper, which was proposed by Eisenberg and co-workers [24]. In our simulations single ribbons are connected to form fibril around tyrosine core, which is more similar to the fibril structure postulated by Langkilde and co-workers [27]. They suggested that a protofilament is formed by two pairs of parallel β-sheets connecting by tyrosine residues. However, such structure is mostly untwisted or has only a small twist. The strongly twisted, fibril-like structures observed in our simulation can be explained as an initial state of fibril formation. In real GNN the twist may disappear during further fibril growth. An important role of tyrosine hydrophobic core for intermediates stability was postulated also by Luiken and Bolhuis [75].

The impact of alanine mutation on aggregation was studied previously by all-atom simulations [78, 79]. Two versions of CHARMM PARAM force field were

used, CHARMM PARAM19 by Caflisch and co-workers [78] and a newer version, CHARMM PARAM22, by Nussinov et al. [79]. Both studies show that the mutant peptides, Y7A, form less stable aggregates than the wild peptide GNN. Due to computational limitations, all-atom studies were performed for small systems, with 3, [78], or up to 8, [79], peptides. The smaller stability of Y7A aggregates is in agreement with our simulations, as in the equilibrium phase deaggregation events are more common for the mutant than for the wild peptide. By increasing the system size, our simulations give more details of aggregation differences between the two sequences, for instance we determined the structural characteristic of fibril-like aggregates. The difference in the mature aggregate morphology between the wild and mutant peptides constitutes a contribution to the fibril polymorphism observed in experiments for wide range of peptides and proteins [6]. The GNN and Y7A fibrils in our simulations differ in the alignment of ribbons, whereas single ribbons are built similarly from parallel, in-register peptide chains. Due to the lack of aromatic residues, Y7A fibrils does not have a well defined core, as is the case for GNN. The Y7A fibrils consist of parallel or antiparallel β-sheets.

Chapter 5

Conclusions and Outlook

In this thesis two coarse-grained models were used to study peptide aggregation. Different levels of resolution of these two models allow for investigating how various structural parameters of peptide chains influence aggregation. The conclusions presented in the next section are general observations regarding the effect of the peptide backbone and side chains on the aggregation kinetics and the structure of aggregates. Our simulations are consistent with the experimental studies showing that numerous proteins and peptides can aggregate into fibrils with similar cross-β pattern. Such observations suggest that the aggregation propensity is connected with intrinsic peptide properties. However, slight differences in the residue sequence can lead to the fibril polymorphism as it was described in Introduction.

After the presentation of general conclusions, the following sections discuss the conclusions specific for Minimal Model and the Bereau-Deserno model.

5.1 General conclusions

This thesis shows that various peptides can aggregate with quite similar kinetics although the final aggregates are different, i.e. the kinetics can be similar for quite different peptides, whereas the structure of mature assembles is strongly dependent on the chain properties determined by the sequence and environmental conditions. With reference to the question about the generic and specific aggregation properties, we hypothesize that in general the aggregation kinetics may be less sequence sensitive than the final aggregate structure.

Both models show that the aggregation rate depends primarily on the intermolecular attraction determined by the backbone properties modulated by the side chains or the presence of highly attractive side-chains. Our research shows also, that the effective backbone properties influence the kinetic mode. Hence, the modification

of the backbone features are sufficient to change the aggregation mechanism. On the other hand, such changes of the mechanism are not observed in the case of a single-residue mutation. Nevertheless, our studies does not rule out the possibility that the sequence modifications may influence the aggregation mechanism.

Another general conclusion is that the structural diversity of the final aggregates mirrors the differences in the oligomer arrangements. It applies to the ordered and disordered clusters in Minimal Model as well as the fibrils observed for GNN and Y7A in the BD model. Notably, the oligomers formed by GNN and Y7A do not reflect exactly the structure of mature fibrils, but some important features may be observed also for the initial clusters allowing for aggregation distinction of these two sequences. It suggests that the early self-assembly steps are important to understand the polymorphism of mature aggregates.

5.2 Minimal Model

The proposed Minimal Model gives insight into the backbone features determining the aggregation propensity. In the simplest version only two backbone parameters were varied, the interaction strength ε and the chain stiffness k_θ. Both parameters influence the peptide aggregation kinetics and the final aggregate structures. The structural phase diagram, obtained for a wide range of these parameters, reproduces the morphological diversity from amorphous to nematic- and crystal-like structures, observed in experiments. The disordered-ordered transition depends on both the peptides attraction and the chain stiffness.

The most interesting structural result is the chiral fluctuations observed for more ordered clusters: the oligomers show collective peptide chain twists around the cluster core. Because Minimal Model does not include intrinsic chirality its behavior corresponds to a homopolymer made up of glycine residues. Thus Minimal Model is consistent with the experimental observation that glycine has a significant impact on the chiral polymorphism of mature fibrils [70]. The twist amplitude and life-time of the twisted states changes with the aggregate size and chain length. They are relatively large for small aggregates and then decrease with cluster size. We hypothesize, that the chiral fluctuations are especially important for oligomers, e.g. for nucleus formation. The dynamics of chiral oscillations indicates that the handedness of mature fibrils can be determined by the kinetics or thermodynamics depending on the nucleus size and length of aggregating peptides. Further studies of chiral fluctuations for small aggregates can give an answer regarding the origin of chirality of mature fibrils observed in experiments. The importance of fibrils twist is connected

also with stability of relatively thin fibrils. Our Minimal Model studies support also this issue, previously investigated in literature by theoretical calculations [71].

Minimal Model also reproduces the kinetic variety observed in experiments. Increasing the interaction strength and chain stiffness cause the transition from no-aggregation to nucleated and then to fast downhill aggregation. The transition between these kinetic modes is primarily determined by the interaction strength, while the chain stiffness is a secondary factor responsible for the aggregation propensity. Nucleated aggregation has the same aggregation mechanism, despite that the length of nucleation phase decreases with increasing values of interaction strength and chain stiffness. The monomer addition is the main path of aggregate growth. The size of nucleus does not change monotonically with interaction strength and chain stiffness. On the other hand, for the downhill kinetics two mechanisms are observed that visible as two scaled curves that do not overlap. For downhill systems with relatively low interaction strength or low chain stiffness, the main path of aggregation growth is the same as for nucleated systems - clusters growth by monomer addition. If the values of model parameters increase, more oligomers are initially formed and further growth undergoes by cluster-cluster coalescence.

Interestingly, the structural and kinetic diagrams do not overlap. This means that the aggregation can be described by the same kinetic mode even if the formed clusters differ significantly in morphology.

Additionally, our studies show the dependence of aggregation on concentration. A decrease in concentration leads from the downhill to nucleated kinetics. Aggregation kinetics change also with the length of peptide chains. Aggregation of longer peptides is faster with fewer deaggregation events, but the general mechanism does not change.

Although the structural and kinetic diversity is reproduced by Minimal Model, fibril formation is not observed. This suggests that fibrillation requires a more complicated model. We hypothesize that anisotropic potentials may be needed.

5.3 Bereau-Deserno model

Our simulations of aggregation for GNN peptide and its point-mutation, Y7A, show that fibril-like structures are formed by both sequences in the studied range of concentrations for sufficiently large system, $N_0 = 72$. The fibril formation occurs through a two-step aggregation mechanism for both sequences. Initially, disordered clusters are formed and then they undergo internal reorganization. This reorganization occurs when the oligomer size exceeds a critical value which is different for each

peptide and is between $M = 25$ and 35 for the wild sequence, GNN, and between $M = 32$ and 58 for the mutant Y7A. These values can be taken as the nucleus size for fibrillization. Despite this similarity, the internal fibrils structure is different, indicates strong dependence of aggregate morphology on the sequence. This agrees with fibril polymorphism observed in experiments and *in vivo*. Interestingly, the structural differences is observed also for the oligomers, implying that the fibril heterogeneity is derived from the initial aggregation steps.

Our aggregation study of GNN indicates a special role of aromatic residues. The 'aromatic core' stabilizes fibrils leading to almost irreversible aggregation. These simulation results give also a contribution to the discussion about the internal structure of the mature fibrils. The observed tyrosine core supports the model proposed by Langkilde and co-workers[27] rather than the crystal-based approach suggested by Eisenberg et al. [24] Noteworthy, the aromatic interactions are also important for spherical and disordered oligomers which are initially formed.

Aggregation of GNN peptides and fibril formation occurs quickly at higher concentration, although, the general two-step mechanism does not change in the studied concentration range. The scaled monomers curves overlap, and so do the clusters kinetic curves, for different system sizes, $N_0 = 20$ and 72, although that the final aggregates structures are different. This suggests that the kinetics similarity can be insufficient to define the molecular mechanism of aggregation. The aggregate growth for GNN differs for disordered and fibrillar structures. The spherical oligomers can growth in all directions, whereas for the fibril-like aggregates the monomers and small cluster addition occurs only at the ends. This implies the 3D or 1D aggregate growth for disordered and fibrillar aggregates, respectively. Although this kinetic difference is not seen in the monomers and clusters kinetics curve, the radius of gyration as a function of aggregate mass shows a change of the slope for the cluster mass close to structural nucleus size. Both types of aggregates, disordered and fibrillar, become more compact during the aggregate growth.

Bibliography

[1] Z. L. Almeida and R. M. M. Brito. Structure and aggregation mechanisms in amyloids. *Molecules*, 25:1195, 2020.

[2] D. Otzen and R. Roland. Functional amyloids. *Cold Spring Harb. Perspect. Biol.*, 11:a033860, 2019.

[3] D. Otzen. Functional amyloid. *Prion*, 4:256–264, 2010.

[4] S. Das, R. S. Jacob, K. Patel, N. Singh, and S. K. Maji. Amyloid fibrils: Versatile biomaterials for cell adhesion and tissue engineering applications. *Biomacromolecules*, 19:1826–1839, 2018.

[5] K. Channon and C. E. MacPhee. Possibilities for 'smart' materials exploiting the self-assembly of polypeptides into fibrils. *Soft Matter*, 4:647–652, 2008.

[6] P. C. Ke, R. Zhou, L. C. Serpell, R. Riek, T. P. J. Knowles, H. A. Lashuel, E. Gazit, I. W. Hamley, T. P. Davis, M. Fändrich, D. E. Otzen, M. R. Chapman, C. M. Dobson, D. S. Eisenberg, and R. Mezzenga. Half a century of amyloids: past, present and future. *Chem. Soc. Rev.*, 49:5473, 2020.

[7] T. P. J. Knowles, M. Vendruscolo, and C. M. Dobson. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Bio.*, 15:384–396, 2014.

[8] R. Gallardo, N. A. Ranson, and S. E. Radford. Amyloid structures: much more than just a cross-β fold. *Curr. Opin. Struct. Biol.*, 60:7–16, 2020.

[9] F. Liberta, S. Loerch, M. Rennegarbe, A. Schierhorn, P. Westermark, G. T. Westermark, B. P. C. Hazenberg, N. Grigorieff, M. Fändrich, and M. Schmidt. Cryo-em fibril structures from systemic aa amyloidosis reveal the species complementarity of pathological amyloids. *Nat. Commun.*, 10:1104, 2019.

[10] N. Rubin, E. Perugia, S. G. Wolf, E. Klein, M. Fridkin, and L. Addadi. Relation between serum amyloid a truncated peptides and their superstructure chirality. *J. Am. Chem. Soc.*, 132:4242–4248, 2010.

[11] D. Kurouski, R. K. Dukor, X. Lu, L. A. Nafie, and I. K. Lednev. Spontaneous inter-conversion of insulin fibril chirality. *Chem. Commun.*, 48:2837–2839, 2012.

[12] A. Loksztejn and W. Dzwolak. Chiral bifurcation in aggregating insulin: an

induced circular dichroism study. *J. Mol. Biol.*, 379:9–16, 2008.

[13] W. Dzwolak, W. Surmacz-Chwedoruk, and V. Babenko. Conformational memory effect reverses chirality of vortex-indeced insulin amyloid superstructures. *Langmuir*, 29:365–370, 2012.

[14] N. Benseny-Cases, O. Klementieva, J. Malý, and J. Cladera. Granular non-fibrillar aggregates and toxicity in alzheimer's disease. *Curr Alzheimer Res.*, 9:962–971, 2020.

[15] O. Klementieva, C. Sandt, I. Martinsson, M. Kansiz, G. K. Gouras, and F. Borondics. Super-resolution infrared imaging of polymorphic amyloid aggregates directly in neurons. *Adv. Sci.*, 7:1903004, 2020.

[16] S. Auer, P. Ricchiuto, and D. Kashchiev. Two-step nucleation of amyloid fibrils: Omnipresent or not? *J. Mol. Biol.*, 422:723–730, 2012.

[17] P. C. A. van der Wel, J. R. Lewandowski, and R. G. Griffin. Solid-state nmr study of amyloid nanocrystals and fibrils formed by the peptide GNNQQNY from yeast prion protein Sup35p. *J. Am. Chem. Soc.*, 129:5117–5130, 2007.

[18] T. Serio, A. G. Cashikar, A. S. Kowal, G. J. Sawicki, J. J. Moslehi, L. Serpell, M. F. Arnsdorf, and S. L. Lindquist. Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science*, 289:1317–1321, 2000.

[19] Y. Ohhashi, K. Ito, B. H. Toyama, J. S. Weissman, and M. Tanaka. Differences in prion strain conformations result from non-native interactions in a nucleus. *Nat. Chem. Biol.*, 6:225–230, 2010.

[20] H. Konno, T. Watanabe-Nakayama, T. Uchihashi, M. Okuda, L. Zhu, N. Kodera, Y. Kikuchi, T. Ando, and H. Taguchi. Dynamics of oligomer and amyloid fibril formation by yeast prion Sup35 observed by high-speed atomic force microscopy. *PNAS*, 117:7831–7836, 2020.

[21] G. Burra, M. B. Maina, L. C. Serpell, and A. K. Thakur. Nucleation-dependent aggregation kinetics of yeast Sup35 fragment GNNQQNY. *J. Mol. Biol.*, 433:166732, 2021.

[22] M. Tanaka, P. Chien, N. Naber, R. Cooke, and J. S. Weissman. Conformational variations in an infectious protein determine prion strain differences. *Nature*, 428:323–328, 2004.

[23] M. Haratake, T. Takiguchi, N. Masuda, S. Yoshida, T. Fuchigami, and M. Nakayama. Amyloid formation characteristics of GNNQQNY from yeast prion protein Sup35 and its seeding with heterogeneous polypeptides. *Colloids Surf. B*, 149:72–79, 2017.

[24] R. Nelson, M. R. Sawaya, M. Balbirnie, A. Madsen, C. Riekel, R. Grothe, and

D. Eisenberg. Structure of the cross-β spine of amyloid-like fibrils. *Nature*, 435:773–778, 2005.

[25] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Madsen, C. Riekel, and D. Eisenberg. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature*, 447:453–457, 2007.

[26] K. E. Marshall, M. R. Hicks, T. L. Williams, S. V. Hoffmann, A. Rodger, T. R. Dafforn, and L. C. Serpell. Characterizing the assembly of the Sup35 yeast prion fragment, GNNQQNY: Structural changes accompany a fiber-to-crystal switch. *Biophys. J.*, 98:330–338, 2010.

[27] A. E. Langkilde, K. L. Morris, L. C. Serpell, Svergun D. I., and B. Vestergaard. The architecture of amyloid-like peptide fibrils revealed by x-ray scattering, diffraction and electron microscopy. *Acta Cryst.*, D71:882–895, 2015.

[28] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.*, 8:37–47, 2015.

[29] S. A. Hollingsworth and R. O. Dror. Molecular dynamics simulation for all. *Neuron.*, 99:1129–1143, 2018.

[30] M. Giulini, M. Rigoli, G. Mattiotti, R. Menichetti, T. Tarenzi, R. Fiorentini, and R. Potestio. From system modeling to system analysis: The impact of resolution level and resolution distribution in the computer-aided investigation of biomolecules. *Front. Mol. Biosci.*, 8:676976, 2021.

[31] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski. Coarse-grained protein models and their applications. *Chem. Rev.*, 116:7898–7936, 2016.

[32] A. Liwo, C. Czaplewski, A. K. Sieradzan, S. A. Lipska, A. G. Samsonov, and R. K. Murarka. Theory and practice of coarse-grained molecular dynamics of biologically important systems. *Biomolecules*, 11:1347, 2021.

[33] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.

[34] T. Bereau and M. Deserno. Generic coarse-grained model for protein folding and aggregation. *J. Chem. Phys.*, 130:235106, 2009.

[35] M. R. Machado, E. E. Barrera, F. Klein, M. Sóñora, S. Silva, and S. Pantano. The sirah 2.0 force field: Altius, fortius, citius. *J. Chem. Theory Comput.*, 15:2719–2733, 2019.

[36] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and

Marrink S-J. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory and Comput.*, 4:819–834, 2008.

[37] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible, and free. *J. Comput. Chem.*, 16:1701–1718, 2005.

[38] H.-J. Limbach, A. Arnold, B. A. Mann, and Holm C. ESPResSo – an extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.*, 174:704–727, 2006.

[39] J. Wedekind, R. Strey, and D. Reguera. New method to analyze simulations of activated processes. *J. Chem. Phys.*, 126:134103–7, 2007.

[40] P. Yi, R. Locker, and G. Rutledge. Molecular dynamics simulation of homogeneous crystal nucleation in polyethylene. *Macromolecules*, 46:4723–4733, 2013.

[41] J. Zierenberg and W. Janke. From amorphous aggregates to polymer bundles: The role of stiffness on structural phases in polymer aggregation. *Eur. Phys. Lett*, 109:28002, 2015.

[42] N Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein. Mdanalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32:2319–2327, 2011.

[43] R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, D. L. Dotson, J. Domanski, S. Buchoux, I. M. Kenney, and O. Beckstein. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In Sebastian Benthall and Scott Rostrup, editors, *Proceedings of the 15th Python in Science Conference*, pages 98–105, 2016.

[44] P.J. Turner. Xmgrace, version 5.1.19. *Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR*, 2005.

[45] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

[46] J. Shao and M. I. Diamond. Polyglutamine diseases: emerging concepts in pathogenesis and therapy. *Hum. Mol. Genet.*, 16:R115–R123, 2007.

[47] I. M. Ilie, W. K. den Otter, and W. J. Briels. A coarse grained protein model with internal degrees of freedom. Application to α-synuclein aggregation. *J. Chem. Phys.*, 144:085103, 2016.

[48] N. S. Bieler, T. P. J Knowles, D. Frenkel, and R. Vacha. Connecting macroscopic observables and microscopic assembly events in amyloid formation using coarse grained simulations. *PLoS Cmput. Biol.*, 8:e1002692, 2012.

[49] Zhang J. and M. Muthukumar. Simulations of nucleation and elongation of

amyloid fibrils. *J. Chem. Phys.*, 130:035102, 2009.

[50] S. Khan, J. Haaga, and J. D. Gunton. Kinetics of aggregation of an anisotropic model of self-assembling molecules. *J. Chem. Phys.*, 143:024906, 2015.

[51] R. A. Mansbach and A. L. Ferguson. Patchy particle model of the hierarchical self-assembly of π-conjugated optoelectronic peptides. *J. Phys. Chem. B*, 122:10219–10236, 2018.

[52] T. X. Hoang, A. Trovato, F. Seno, J. R. Banavar, and A. Maritan. Geometry and symmetry presculpt the free-energy landscape of proteins. *PNAS*, 101:7960–7964, 2004.

[53] S. Auer, C. M. Dobson, and M. Vendruscolo. Characterization of the nucleation barriers for protein aggregation and amyloid formation. *HFSP J.*, 1:137–146, 2007.

[54] N. B. Hung, D.-M. Le, and T. X. Hoang. Sequence dependent aggregation of peptides and fibril formation. *J. Chem. Phys.*, 147:105102, 2017.

[55] R. Pellarin and A. Caflisch. Interpreting the aggregation kinetics of amyloid peptides. *J. Mol. Biol.*, 360:882–892, 2006.

[56] S. Ranganathan, S. K. Maji, and R. Padinhateeri. Defining a physical basis for diversity in protein self-assemblies using minimal model. *J. Am. Chem. Soc.*, 138:13911–13922, 2016.

[57] M. Lorusso, A. Pepe, N. Ibris, and B. Bochicchio. Molecular and supramolecular studies on polyglycine and poly-L-proline. *Soft Matter*, 7:6327–6336, 2011.

[58] J. P. Bernacki and R. M. Murphy. Length-dependent aggregation of uninterrupted polyalanine peptides. *Biochemistry*, 50:9200–9211, 2011.

[59] H. Jafarinia, E. van der Giessen, and P. R. Onck. Phase separation of toxic dipeptide repeat proteins related to C9orf72 ALS/FTD. *Biophys. J.*, 119:843–851, 2020.

[60] A. Schmitz, J. P. Marques, I. Oertíg, N. Niran Maharjan, and S. Saxena. Emerging perspectives on dipeptide repeat proteins in C9oRF72 ALS/FTD. *Front. Cell. Neurosci.*, 15:637548, 2021.

[61] V. R. Singh and L. J. Lapidus. The intrinsic stiffness of polyglutamine peptides. *J. Phys. Chem. B*, 112:13172–13176, 2008.

[62] A. F. Chin, D. Toptygin, A. Elam, T. P. Schrank, and V. J. Hilser. Phosphorylation increases persistence length and end-to-end distance of a segment of tau protein. *Biophys. J.*, 110:362–371, 2016.

[63] S. Kumar, N. Rezaei-Ghaleh, D. Terwel, D. R. Thal, M. Richard, M. Hoch, J. M. Mc Donald, U. Wüllner, K. Glebov, M. T. Heneka, et al. Extracellular phosphorylation of the amyloid β-peptide promotes formation of toxic aggre-

gates during the pathogenesis of alzheimer's disease. *EMBO J.*, 30:2255–2265, 2011.

[64] S. Kumar and J. Walter. Phosphorylation of amyloid beta (Aβ) peptides – a trigger for formation of toxic aggregates in alzheimer's disease. *Aging*, 3:803–812, 2011.

[65] Y. Yoshimura, Y. Lin, H. Yagi, Y.-H. Lee, H. Kitayama, K. Sakurai, M. So, H. Ogi, H. Naiki, and Y. Goto. Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. *PNAS*, 109:14446–14451, 2012.

[66] K. Kar, M. Jayaraman, B. Sahoo, R. Kodali, and R. Wetzel. Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat. Struct. Mol. Biol.*, 18:328–337, 2011.

[67] A. M. Bhattacharyya, A. K. Thakur, and R. Wetzel. Polyglutamine aggregation nucleation: Thermodynamics of a highly unfavorable protein folding reaction. *PNAS*, 102:15400–15405, 2005.

[68] E. Landrum and R. Wetzel. Biophysical underpinnings of the repeat length dependence of polyglutamine amyloid formation. *J. Biol. Chem.*, 289:10254–10260, 2014.

[69] M. Chen, M. Tsai, W. Zheng, and P. G. Wolynes. The aggregation free energy landscapes of polyglutamine repeats. *J. Am. Chem. Soc.*, 138:15197–15203, 2016.

[70] M. Wang, Y. Zhao, L. Zhang, J. Deng, K. Qi, P. Zhou, X. Ma, D. Wang, Z. Li, J. Wang, J. Yang, J. R. Lu, J. Zhang, and H. Xu. Unexpected role of achiral glycine in determining the suprastructural handedness of peptide nanofibrils. *ACS Nano*, 15:10328–10341, 2021.

[71] G. M. Grason. Chiral and achiral mechanisms of self-limiting assembly of twisted bundles. *Soft Matter,*, 16:1102–1116, 2020.

[72] J. Haaga, J. D. Gunton, N. Buckles, and J. M. Rickman. Early stage aggregation of a coarse-grained model of polyglutamine. *J. Chem. Phys.*, 148:045106, 2018.

[73] F. X. Smit, J. A. Luiken, and P. G. Bolhuis. Primary fibril nucleation of aggregation prone tau fragments PHF6 and PHF6*. *J. Phys. Chem. B*, 121:3250–3261, 2017.

[74] J. A. Luiken and P. G. Bolhuis. Prediction of a stable associated liquid of short amyloidogenic peptides. *Phys. Chem. Chem. Phys.*, 17:10556–10567, 2015.

[75] J. A. Luiken and P. B. Bolhuis. Primary nucleation kinetics of short fibril-forming amyloidogenic peptides. *J. Phys. Chem. B*, 119:12568–12579, 2015.

[76] A. Pawar and G. Favrin. Finite size effects in simulations of protein aggregation.

PLOS ONE, 3:1–6, 2008.

[77] G. Singh, I. Brovchenko, A. Oleinikova, and R. Winter. Peptide aggregation in finite systems. *Biophys. J.*, 95:3208–3221, 2008.

[78] J. Gsponer, U. Haberth ur, and A. Caflisch. The role of side-chain interactions in the early steps of aggregation: Molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup35. *PNAS*, 100:5154–5159, 2003.

[79] J. Zheng, B. Ma, C.-J. Tsai, and R. Nussinov. Structural stability and dynamics of an amyloid-forming peptide GNNQQNY from the yeast prion Sup35. *Biophys. J.*, 91:824–833, 2006.

Appendix A

Academic achievements

Publications

1. Szała B., Molski A., Aggregation kinetics of short peptides: All-atom and coarse-grained molecular dynamics study *Biophys. Chem.* **2019**, 253:106219.

2. Pannuzzo M., Szała B., Raciti D., Raudino A., Ferrarini A. Helical Inclusions in Phospholipid Membranes: Lipid Adaptation and Chiral Order *J. Phys. Chem. Lett.* **2019**, 10, 18:5629–5633.

3. Szała B., Molski A., Chiral structure fluctuations predicted by a coarse-grained model of peptide aggregation *Soft Matter*, **2020**, 16:5071-5080.

4. Szała-Mendyk B., Andrzej Molski A., Clustering and Fibril Formation during GNNQQNY Aggregation: A Molecular Dynamics Study *Biomolecules*, **2020**, 10:1362.

5. Szała-Mendyk B., Molski A., Diverse Aggregation Kinetics Predicted by a Coarse-Grained Peptide Model *J. Phys. Chem. B*, **2021**, 125:7587-7597.

Fellowships

01-04.2021 Institute of Biological Information Processing - Structural Biochemistry, Julich Research Center, Germany, Fellowship Supervisor: Prof. Birgit Strodel

07-09.2018 Erasmus+ Internship, Chemical Sciences Department, University of Padova, Italy, Fellowship Supervisor: Prof. Alberta Ferrarini

07-09.2017 Erasmus+ Internship, Chemical Sciences Department, University of Padova, Italy, Fellowship Supervisor: Prof. Alberta Ferrarini

Conferences

10.2021 EMBO workshop: Advances and Challenges in Biomolecular Simulations, online

02.2020 ChemCYS 2020, Blankenberge, Belgium,

05.2019 VII Doctoral Symposium on Chemistry, Łódź, Poland,

09.2018 Italian Soft Days, Padova, Italy.

Awards

2020/2021 and 2021/2022 Pro-quality scholarship
2019/2020 and 2020/2021 Rector's scholarship