

Dissertation  
submitted to the  
Combined Faculties of the Natural Sciences and Mathematics  
of the Ruperto-Carola-University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

Put forward by  
Ana María Herrera Rodríguez  
born in: Bucaramanga (Colombia)  
Oral examination: 08 January 2020



THE ROLE OF FLOW IN THE  
SELF-ASSEMBLY OF DRAGLINE SPIDER  
SILK PROTEINS

Referees:

Prof. Dr. Ulrich Schwarz

Prof. Dr. Joachim Spatz



## Zusammenfassung

Seidenfasern sind aufgrund der kontrollierten Selbstassemblierung ihrer Proteinbausteine, den Spidroins, ein außergewöhnlich widerstandsfähiges Biomaterial. Die Kombination aus Dehnbarkeit und hoher Reißfestigkeit hängt von der mikroskopischen Zusammensetzung innerhalb der Faser ab: kleine und starke Beta-Faltblatt-Kristalle, welche hauptsächlich aus Polyalaninen bestehen, sind in eine flexible, amorphe Matrix mit hohem Glycerinanteil eingeschlossen. Die innere Molekülstruktur der Seidenproteine und die daraus resultierende hohe Empfindlichkeit gegenüber Dehnströmungen sind von grundlegender Bedeutung für das Spinnen von Seidenfasern. Der Mechanismus von strömungsinduzierter Selbstassemblierung sowie relevante Dynamiken einzelner Seidenproteine unter Strömungseinfluss sind jedoch weitgehend unbekannt.

Um eben diese Dynamiken sowie die Selbstassemblierung von Spinnenseidenproteinen unter gleichförmiger Strömung zu untersuchen, wurde in der vorliegenden Arbeit ein Bottom-up Ansatz verwendet. Hierfür wurden Nicht-Gleichgewichts-Molekulardynamik-Simulationen (MD Simulationen) angewandt, um diese Prozesse auf zwei Skalen zu eruieren: Zum einen einem atomistischen Modell mit expliziter Beschreibung des Wassers sowie zum anderen ein grobkörniges Modell, welches Hydrodynamiken durch Multi-Partikel-Kollisionen integriert. Eine vorherige Implementierung und systematische Untersuchung von MD-Simulationen mit gleichförmiger Strömung wurde auf Basis der GROMACS MD-Software durchgeführt, um die Bedeutung von Strömung auf Spinnenseidenmoleküle atomistisch zu analysieren.

Wird ein gebundenes einzelnes Seidenpeptid gleichmäßiger Strömung ausgesetzt, führt dies zu einem Übergang von aufgewickelterm zu gestrecktem Peptid, wobei dieses eine Vielzahl von Zwischenzuständen durchläuft. Dieser Veränderungsprozess ist abhängig von der mittleren Strömungsgeschwindigkeit. Strömungsinduzierte Strukturveränderungen einzelner Spidroins weisen eine auffällige Tendenz von Alaninketten in Beta-Faltblatt-Konformationen auf. Atomistische Simulationen des Assemblierungsprozesses bei niedrigen Strömungsraten zeigten, dass sich Kontakte innerhalb der Ketten hauptsächlich in den Polyalanin-Wiederholungen bilden. Dies ist eine geeignete Voraussetzung für die Kristallbildung sowie das Fibrillieren. Des Weiteren konnten Beta-Faltblatt-Formationen bei geringen Strömungen festgestellt werden, was bestätigt, dass Strömung die Kristallbildung fördert. Die Erkenntnisse wurden mit den grobkörnigen Hydrodynamiksimulationen in Aminosäurelösungen komplementiert, bei welchen die Seidenproteine als halbflexible Blockcopolymer betrachtet wurden. Hierbei zeigte die Beobachtung der Oligomerbildung, dass Spidroins schneller aggregieren, wenn sie weniger ausgedehnt sind. Bei mittleren Peptidextensionen (ca. 60-70%) nimmt die Spidroinausrichtung zu, während sich ihre Assemblierung aufgrund der reduzierten Fluktuation orthogonal zur Strömungsrichtung verlangsamt.

Das mikroskopische Verständnis von Spidroindynamiken, welches in dieser Arbeit vorgelegt wird, ist voraussichtlich für andere strömungsabhängige Proteine relevant.

## Abstract

Silk fibers are outstandingly tough biomaterials, a result of the controlled self-assembly of their protein building blocks, spidroins. The combination of extensibility and high tensile strength relies on the microscopic composition within the fiber: small and strong beta-sheet crystals formed mainly by poly-alanine repeats enclosed into a flexible amorphous matrix of glycine-rich repeats. The internal molecular structure of silk proteins makes them sensitive to an elongational flow, which is a crucial factor for spider silk fiber spinning. However, the mechanism of flow-induced silk self-assembly, as well as the relevant dynamics of single silk proteins under flow remain largely unknown.

In the present work, a bottom-up approach was used to study the dynamics and self-assembly of spider silk proteins under uniform flow conditions. We used non-equilibrium molecular dynamics (MD) simulations to study these processes at two scales, an atomistic model with explicit water, and a coarse-grained model with hydrodynamics incorporated by multi-particle collision dynamics. To be able to analyze the role of the flow on spider silk molecules atomistically, a prior implementation and systematic study of uniform flow MD simulations were carried out based on the GROMACS MD software. Subjecting a tethered single silk peptide to uniform flow leads to a coiled-to-stretch transition involving a multitude of intermediates states, the process of which depends on the mean flow velocity. The flow-induced structural changes of single spidroins exhibit a prominent tendency of alanine residues to be in beta-sheet conformation. All-atom simulations of the assembly process at low flow regimes revealed that the interchain contacts happen primarily in the poly-alanine repeats, which is a suitable condition for crystal formation and fibrillation.

We also found beta-sheets formation at low flow regimes, confirming that flow promotes crystal formation. We complemented these findings to the more coarse-grained hydrodynamic simulations at aminoacid resolution, treating the silk proteins as semi-flexible block copolymers. We observed that the spidroins aggregate faster when they are less extended by monitoring oligomer formation in time. At medium peptide extensions (around 60-70%), the spidroin alignment increases, while their assembly slows down because of the reduced fluctuations orthogonal to the flow direction. The microscopic understanding of the spidroin dynamics provided in this work is likely relevant for other flow-dependent proteins.

## Acknowledgements

At the final stage of this experience full of unforgettable and rewarding moments, I am immensely grateful for those, who were present during this time of my life. I would like to express my gratitude to my thesis advisors Prof. Dr. Ulrich SCHWARZ and Prof. Dr. Frauke GRÄTER. Without their guidance it would have not been possible. Frauke, my day-to-day advisor, I would like to deeply thank her for giving me the great opportunity to work in a very passionate research. I am truly grateful for her sincere dedication and motivation. She impressed me with her ideas towards the understanding of diverse scientific topics. In my research, these ideas were very helpful, especially when I was facing problems. Furthermore, I want to thank all current and old members of the MBM group, who helped me to get started in the research topics that were new for me- in particular, Dr. Camilo APONTE and Dr. Katra KOLSEK for their advice and support during my period of the Ph.D. My sincere gratitude to Dr. Anil K. DASSANA for his valuable support and help. The discussions with him were indispensable during my Ph.D.

On a more personal note, I want to thank my friends: Martin, Sebastian, Nohora, Diego, Julieth, y Miguel. They have been essential during these years, fulfilling my life outside of the Ph.D. with wonderful moments.

A profound gratitude is owed to my mom Ana Herrera and my whole family, who have always been present, supporting me from the distance and giving me strength and unconditional love.

Thanks from my heart to my love, Ale. Part of the happiness I experience right now is related to her.

*Sincerely,*

Ana María Herrera



# Contents

<b>List of Figures</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Spider Silk . . . . .	1
1.2 Motivation and Objectives . . . . .	8
1.3 Outline of the Thesis . . . . .	9
<b>2 Simulation Methods</b>	<b>11</b>
2.1 Molecular Dynamics simulations . . . . .	11
2.1.1 Force Fields . . . . .	15
2.1.2 Multiparticle Collision Dynamics . . . . .	19
2.1.3 Equations of Motion . . . . .	20
<b>3 MD Simulations of Molecules in Uniform Flow</b>	<b>23</b>
3.1 Introduction . . . . .	23
3.2 Methods . . . . .	25
3.3 Results and Discussion . . . . .	29
3.4 Conclusion . . . . .	36
<b>4 Dynamics of a Single Silk Peptide under Uniform Flow</b>	<b>39</b>
4.1 Introduction . . . . .	39
4.2 Methods . . . . .	41
4.3 Results and Discussion . . . . .	44
4.4 Conclusion . . . . .	54

## CONTENTS

---

<b>5</b>	<b>Nanoscale Self-Assembly of Dragline Spidroins</b>	<b>57</b>
5.1	Introduction . . . . .	57
5.2	Methods . . . . .	59
5.2.1	Simulation Setup MD Simulations . . . . .	59
5.3	Results and Discussion . . . . .	61
5.4	Conclusion . . . . .	67
<b>6</b>	<b>Summary and Conclusions</b>	<b>69</b>
	<b>Bibliography</b>	<b>75</b>

# List of Figures

1.1	Schematic overview of the seven spider silk types and their applications. . .	2
1.2	Nano-scale structure of dragline spider silk. . . . .	4
1.3	Native spidroin. . . . .	4
1.4	MA spinning gland. . . . .	5
1.5	Viscosity dependence of silk feedstocks on the shear rate. . . . .	7
2.1	Modelling of a silk peptide. . . . .	15
2.2	All-atom force field . . . . .	16
2.3	Coarse Grained force field . . . . .	18
3.1	Example setup for flow MD simulations. . . . .	29
3.2	Water density varies along the flow direction. . . . .	30
3.3	Flow MD produces uniform flow in the m/s regime. . . . .	31
3.4	Uniform flow velocities under different forces applied in the slice. . . . .	31
3.5	Mean velocity over average force per molecule in the pull slice. . . . .	32
3.6	Number density of water molecules along the flow orientation without smoothening the applied force inside the slice. . . . .	34
3.7	Number density as a function of $x$ at flow velocity of 1.3 m/s for three different pull slice sizes. . . . .	34
3.8	Flow-induced unfolding is illustrated for the 40 AA intrinsically disorder linker of A1 and A2 domains of vWF. . . . .	35
3.9	(a) Flow velocity for a pulling slice force of 1.5 pN as a function of the coupling time of the thermostat. b) Comparison of the flow velocity of a pure water system for SPC and Tip4p-D water models. . . . .	36
4.1	Modelling of a silk peptide under uniform flow . . . . .	45

## LIST OF FIGURES

---

4.2	Peptide extension over time for every flow rate. . . . .	46
4.3	Fractional extension as a function of flow velocity. . . . .	46
4.4	Extension of the spidroin fragment increases with flow velocity and drag force. . . . .	48
4.5	Worm-like Chain model. . . . .	49
4.6	Normalized probability distribution of $C_\alpha$ atoms along $x$ . . . . .	50
4.7	Number of contacts of Ala residues. . . . .	51
4.8	Total number of contacts and amorphous region contacts. . . . .	52
4.9	Flow populates $\beta$ -sheet and PPII conformations. . . . .	54
5.1	Initial peptide setup MD simulations under flow. . . . .	59
5.2	AA simulations description. . . . .	60
5.3	CG simulation description. . . . .	60
5.4	Friend of Friends algorithm . . . . .	61
5.5	Monitoring in time of every silk peptide of one replicate depending on the size of the oligomer they form. . . . .	63
5.6	Oligomerization in all-atom simulations. . . . .	64
5.7	All mean contacts formation between the spidroins. . . . .	64
5.8	$\beta$ -sheet content for all-atom simulations . . . . .	65
5.9	Assembly in CG simulations. . . . .	66
5.10	Contact formation in CG simulations . . . . .	67

## Abbreviations

AA	All-Atom
ALA	Alanines
BB	Backbone-Backbone
BD	Brownian Dynamics
BS	Backbone-Sidechain
CFD	Computational Fluid Dynamics
CG	Coarse-Grained
COM	Center-Of-Mass
CSVR	Canonical Sampling Velocity Rescaling
CT	C-Terminal Domain
DPD	Dissipative Particle Dynamics
FENE	Finitely Extensible Nonlinear Elastic
FFT	Fast Fourier Transform
FJC	Freely Joint Chain
FOF	Friend of Friends
GROMACS	Groningen Machine for Chemical Simulations
HI	Hydrodynamic Interactions
LJ	Lennard-Jones
MA	Major Ampullate
MD	Molecular Dynamics
MPCD	Multiparticle Collision Dynamics
NT	N-Terminal Domain
PBC	Periodic Boundary Conditions
PME	Particel Mesh Ewald
PPII	Poly-Proline II
SOP-SC	Self Organized Polymer Model with Side Chains
SS	Sidechain-Sidechain
vdW	van der Waals
vWF	von Willebrand factor
WLC	Worm-Like Chain



# Introduction

## 1.1 Spider Silk

Spider silks are biomaterials made of proteins called spidroins, which self-assemble into a variety of fibers. Orb-weaving spiders have evolved to spin up to seven types of silk fibers for different biological functions, which vary in mechanical properties, structure and constituent proteins [1, 2]. A special combination of low density, high tensile strength and large extensibility renders some of these fibers remarkably tough.

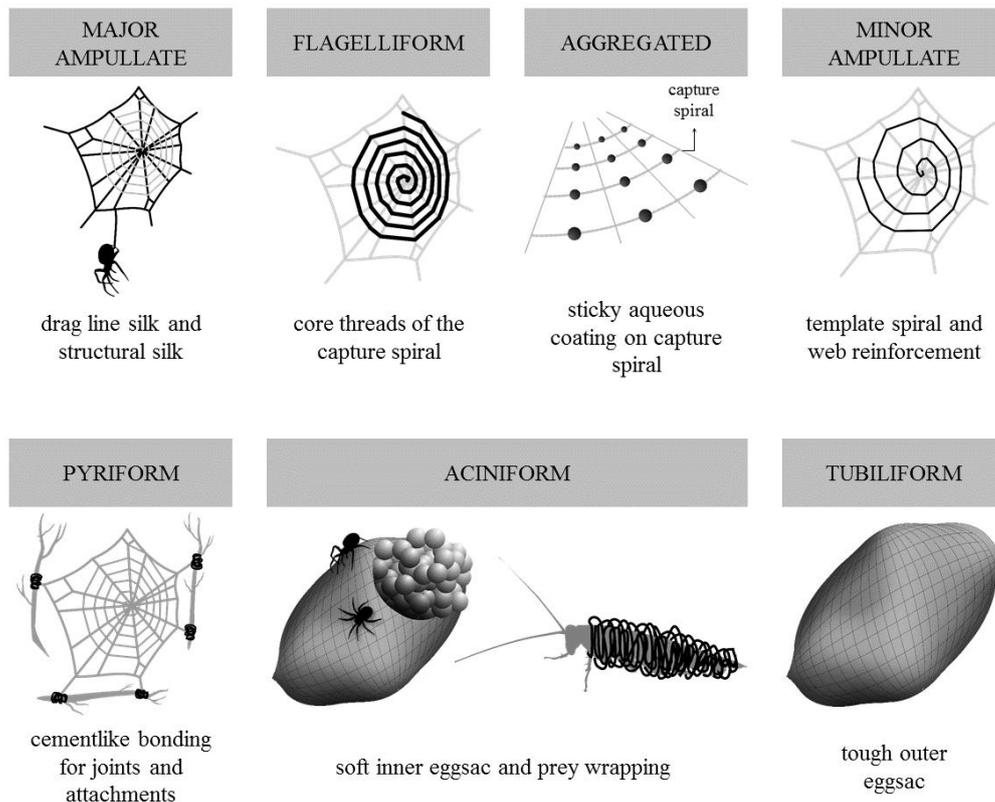
The mechanical properties of a specific silk type depend on their molecular structure and their manufacturing process. Unraveling the process of spider silk formation is an important step towards fabricating bioinspired fibers. This allows to take advantage of the fact that silks are produced under environmentally friendly conditions: a highly energy efficient process that takes place at ambient temperatures and involves renewable resources. In contrast to silkworms, spiders are difficult to domesticate and under these circumstances produce a small amount of silk [3]. Therefore, a large-scale production of spider silk fibers requires expression and purification of silk proteins using recombinant DNA technologies and artificial spinning. The latter is a key factor to obtain a fiber with the desired mechanical performance. It has not yet been possible to synthetically manufacture a fiber with the same characteristics of the natural one, reflecting the precise control involved in the assembly pathway of spidroins.

# 1. INTRODUCTION

---

A broad knowledge of the biosynthesis of spider silk would permit to build tailored biomaterials with a variety of functionalities, since different silk properties like biocompatibility, biodegradability and mechanical performance are encoded in the protein sequence. Currently, silk proteins are being used as prototypes to synthesize tunable polymeric materials [4, 5, 6]. However, the underlying self-assembly mechanism remains to be fully understood.

Orb-weaving spiders are masters in making and spinning tunable biopolymers. They are spun in separated spinning glands inside the spider's abdomen, whereby the name of each one depends on the gland where it is formed [1] (see Figure 1.1):



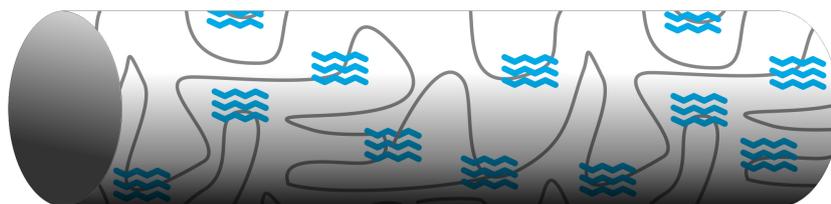
**Figure 1.1:** Schematic overview of the seven spider silk types and their applications. Own picture adapted from [1].

- *Major Ampullate (MA) silk*: MA silk is mostly known as dragline silk. It forms part of the orb-web frame and the lifeline for the spider hanging.
- *Flagelliform silk*: Sticky capture-spiral silk. The energy of an insect impact on the spider web is dissipated in the stretching of the elastic capture spiral.
- *Aggregate silk*: Glue coats in the capture-spiral. Aggregate silk forms droplets along the flagelliform silk fibers. As soon as an insect is on the capture spiral, the aggregate silk droplets behave as viscoelastic solids that adhere to the prey and elongate to maintain the prey attached to the web.
- *Minor Ampullate silk*: Temporary scaffolds during the process of web construction. These scaffolds give reinforcement to the spider web.
- *Pyriform silk*: Pyriform silk agglutinates dragline silk threads to substrates. It behaves like a strong adhesive, even on slippery surfaces like Teflon.
- *Aciniform silk*: It is the toughest spider silk type and used to wrap prey and cover the inner part of the eggsac.
- *Tubiform silk*: It is the stiffest silk fiber, and covers the outer part of the eggsac.

Among the different spider silk types, dragline silk is the one most extensively studied. It comprises major and minor ampullate silks, whereby experiments often refer to MA silk. It has a high capacity to absorb energy, due to the balance between ductility and tensile strength. This special combination makes dragline silk outperform most materials in terms of toughness ( $354 \pm 93 \text{ MJ m}^{-3}$ ) and extensibility of up to 91% of its length [7]. Dragline silk typically measures 2-5  $\mu\text{m}$  in diameter and exhibits a nano-scale structure: nanocrystals with a size on the order of  $2 \times 5 \times 7 \text{ nm}^3$  [8] surrounded by an amorphous region. The former provides strength and the latter provides flexibility to the fiber (Figure 1.2), according to a generally accepted assumption, which among others is also a result of a simple spider silk elastic network model [9]. This model describes dragline silk as a regular 2D lattice, whose nodes have different stress-strain relationships depending on the amorphous and crystalline regions. The microscopic structure of spider silk depends not only on its formation process, but also on its molecular components, which are the spidroins MaSp1 and MaSp2 [10, 11] with molecular weights of 250-400 kDa per monomer [12]. A spidroin monomer consists of pH dependent and conserved N and C terminal (NT and CT) regions, which enclose a

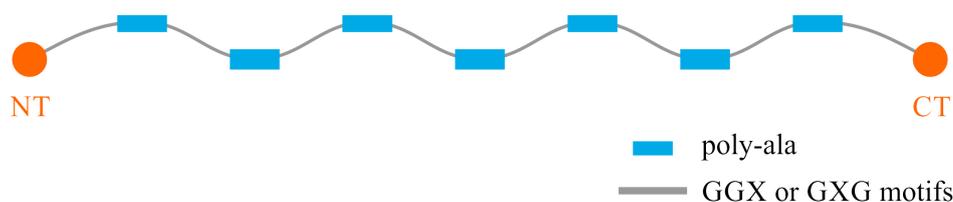
## 1. INTRODUCTION

---



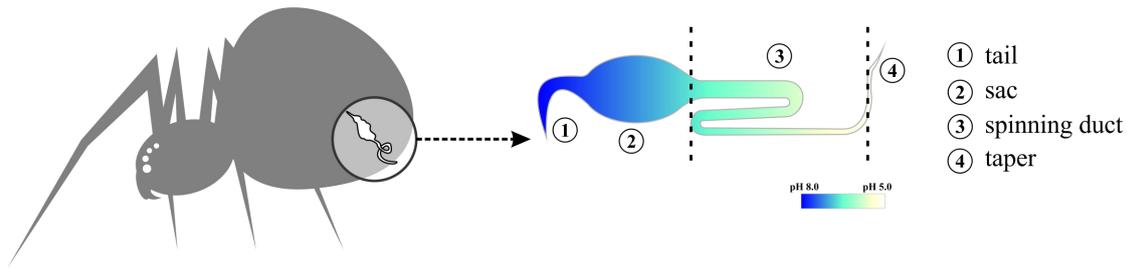
**Figure 1.2:** Nano-scale structure of dragline spider silk. Tough crystals embedded in a flexible amorphous matrix of GGX motifs.

highly repetitive core rich in poly-alanine (poly-Ala) and poly-glycine-alanine (poly-Gly-Ala) repeats and intrinsically disordered fragments rich in glycine motifs (GGX, GXG and GXX, where X stands for Y=tyrosine, L=leucine or Q = glutamine). The lengths of the terminal domains are 80-140 amino acids, while for the repetitive part, every block contains about 40-200 amino acids and it repeated up to 20 times in one monomer (Figure 1.3). During spider silk formation, each of these region plays a special role in order to build an outstanding natural fiber.



**Figure 1.3:** Native spidroin. A long repetitive region of poly-Ala repeats and intrinsically disordered regions are flanked by conserved NT and CT domains.

The production and directed self-assembly of dragline silk proteins happen in the MA spinning gland (s-gland) (Figure 1.4), in which the spidroins are stored as a concentrated aqueous solution (from 20% to 50% w/v ) named spinning dope [13]. The s-gland is divided into four regions: in the tail (1) the proteins are produced and secreted to the sac (2), where they are stored without aggregating. This silk solution flows towards the spinning duct (s-duct) (3), where an elongational flow, a decreasing pH gradient and ions exchange take place to self-assemble the spidroins under controlled conditions. After assembly, the fiber is extracted by the spider from the taper (4). There is still no consensus on the assembly mechanism of the spidroins inside the s-gland, in accordance with which two models have been proposed [14, 15]. The first model suggests either disulfide-linked multimers or monomers pass to the s-duct at very high concentrations ( $\sim 50\%$ ), where the proteins align



**Figure 1.4:** MA spinning gland. The proteins are produced in the tail (1) and stored into the sac (2) under pH 8.0 to keep them soluble. After storing, they flow through the s-duct (3) to self-assemble under a decreasing pH gradient down to 5.0 and elongational flow conditions along the duct. Once the protein assembly occurs, the nano and micro fibrils are extracted from the taper (4) by a pulling out made by the spider's leg.

and crystallize in the elongational flow [15]. The alignment and extension in flow of the high concentrated dope induces a liquid-crystalline state, in which the intermolecular van der Waals interactions promote hydrogen bond formation. Once a dehydration process occurs, the fiber is formed. In the second model, since the spidroins feature an amphiphilic behavior due to their molecular structure, they self-assemble into small micelles of diameter on the order of 100-200 nm inside the sac [16]. A large group of micelles form larger globules that are elongated under shear forces in the s-duct, leading to fiber formation. Recently, NMR experiments reported hierarchical micelles with flake-like sub-domains networks in the spinning dope. Upon physical shearing, the subdomains become elongated and remain assembled as entangled networks in the parent micelle [17]. The two assembly models do not exclude each other, and micelle formation and flow induced elongation likely both play a role.

Regarding the spidroins, the role of the NT and CT domains in the spinning process has been investigated in depth. Both terminal regions possess five helices that change their conformation upon pH change, change in ionic strength and flow. They are mainly in charge of controlling the solubility and the assembly of the native spidroins by: preventing fatal aggregation of the protein monomers in the storage state, and promoting the primary interconnections in the s-duct, called *the lock and trigger mechanism hypothesis* [2]. From pH 7.0 to 8.0 in the sac, the NT region is monomeric. Once the spidroins arrive to the s-duct, the NT monomers experience anti-parallel dimerization at low pH conditions ('lock'). In the case of the CT domain, it dimerizes in the tail by aligning two pairs of oppositely charged residues and remains associated along the whole gland. Furthermore, the CT region exper-

## 1. INTRODUCTION

---

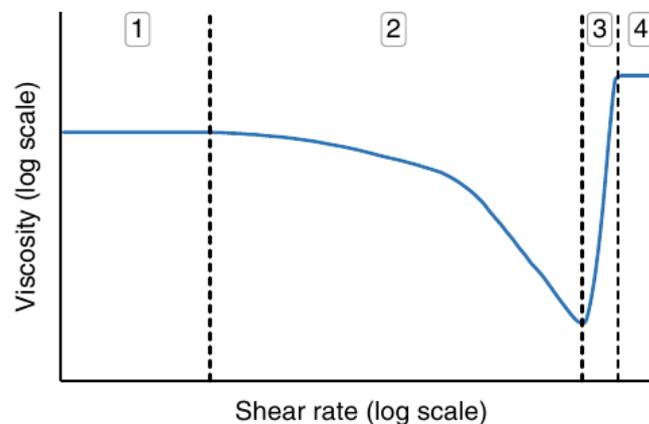
iences conformational changes ('trigger') along the s-duct behaving as a switch from the storage to the assembly stage [18].

In contrast to the N and C terminus structures, the repetitive region is insensitive to pH and strongly affected by flow. Under shear forces and with the help of the NT and CT domains, the repetitive part of the spidroins aligns and assembles by forming the crystalline and amorphous phases of the fiber (Figure 1.2). The crystalline phase comprises  $\beta$ -sheets crystals in the poly-Ala and poly-Gly-Ala regions of about 4-15 residues oriented mainly parallel to the fiber axis [19, 20]. Finite element modeling based on MD simulations predicts higher toughness if the crystallites arrange in a lamellar fashion. Furthermore, long  $\beta$ -sheet crystals increase the toughness, elastic modulus, and rupture stress of the fiber [21, 22]. The amorphous phase corresponds to the intrinsically disorder repeats comprising GGX and GPGGX, which do not have a well defined structure.  $3_1$ -helical structures are found in GGX motifs, while GPGGX and GQGX motifs form type II  $\beta$ -turns [23, 24]. These  $\beta$ -turns are associated with the elasticity properties of the fiber.

Assembling the spidroins successfully requires to extent, align and pack them in the elongational flow produced in the s-duct. However, the role of flow in spider silk formation is fairly unknown. Microfluidic experiments have shown how elongational flow enhances beta-sheet secondary structure formation [25], which has been also confirmed by MD simulations [6, 26]. Nevertheless, the mechanism behind flow-induced  $\beta$ -sheet formation remains elusive. One of the main difficulties towards this aim is the high hydrophobicity of the silk dope, requiring a careful control of the solvent during the assembly pathway. Additionally, simulations have shown that small anti-parallel  $\beta$ -sheet arrangements are favorable to increase fiber strength [27, 28, 29], but still fail trying to explain how this arrangement is controlled during the assembly process. The elongational flow inside of the s-duct has been mostly characterized by rheological experiments of the millimeter and micrometer size silk feedstocks at different concentrations that are under shear and elongational stress. These studies explain the complex viscoelastic rheology inside the s-duct [15, 30, 31, 32]. Before spinning, the silk dope has to pass through the narrow s-duct around 10 times less in diameter than the sac diameter. This poses a challenge for a spider, since there is flow resistance of the liquid crystalline solutions. According to the Hagen-Poiseuille equation for laminar-Newtonian fluids;

$$\Delta P = \frac{8Q}{\pi} \mu L R^4, \quad (1.1)$$

where  $\Delta P$  is the pressure change,  $Q$  is the flow rate,  $R$  the s-duct radius and  $\mu$  is the fluid viscosity. The pressure to push the silk solution into the s-duct is 10.000 times higher than inside the sac. To spin a fiber under newtonian conditions requires high pulling forces that can not be generated by the spider's walking (from 20 mm/s to 800 mm/s) [15]. However, the silk dope behaves like a non-newtonian fluid, in which the viscosity depends on the shear rate [15, 30, 31, 32]. It allows low cost metabolic consumption by the spider to spin the fiber, since the pressure drop needed to extrude the silk dope is less. A number of studies are in agreement with a non-newtonian behavior of silk, which has also been shown for other non-biological polymer solutions. However, the quantitative description of the shear-dependent viscosity varies [30, 32, 33]. Four regimes have been described for the response of silk feedstocks under an increasing shear rate [32], see Figure 1.5: (1) A pseudo-newtonian flow at low shear rates, (2) a shear thinning behavior that facilitates protein alignment with a subsequent fibrillation (3), followed by a second plateau (4) where a viscoelastic solid is formed.



**Figure 1.5:** Viscosity dependence of silk feedstocks on the shear rate. (1) Pseudo-newtonian fluid, (2) shear thinning behavior, (3) fibrillation (4) and viscoelastic solid. Figure taken from [32].

An established model that describes the spinning in the later part of the gland [30](Figure 1.4) still remains to be confirmed: instead of spinning the silk thread inside the gland, the spider squirts out silk solution, and with the help of her legs and her body weight she elongates the liquid crystal phase inducing a phase separation in the gel to form the fiber.

### 1.2 Motivation and Objectives

Considering the complex processes involved in spider silk formation, to date it has not been possible to understand the role of the hydrodynamics flow in the assembly of spidroins. The experimental challenge lies in the high propensity of the silk proteins to aggregate and in the nanometer scale structuring, which is difficult to monitor in real time.

Given the large number and size of the proteins involved, computational techniques have so far not been applied to address the assembly but instead have focused on examining the mechanical properties of silk fibers.

The motivation of this work arises from the interest of revealing the molecular mechanisms at play during the assembly under flow.

We aim to study in this work the influence of flow on silk proteins in the microscopic regime by atomistic and coarse-grained simulations. We focus on monitoring assembly in idealized conditions, in which flow is uniform and proteins are tethered to each other through the terminal domains, which renders the problem tractable by molecular simulations. The specific objectives of this thesis are:

- to analyze the flow-dependent dynamics of individual silk proteins in flow. We will study the transition from collapsed to stretched conformations of tethered silk proteins by uniform flow at both atomistic and coarse-grained simulations, and will compare these results to polymers in flow.
- to understand how these flow-dependent dynamics promote assembly of multiple spidroin molecules into structures present in silk fibers, again using both atomistic and coarse-grained simulations.

With this work, we aim to contribute to the understanding of the role of flow in the formation and internal arrangement of spidroin nano-assemblies or other flow-induced protein aggregation processes.

To study the dynamics of silk molecules in flow, molecular dynamics (MD) technique is used in this work. The proteins are modeled in amino acid (or coarse-grained, (CG)) and all-atom (AA) resolution. Both models are in the nano-scale, with the CG validated by and supplementing the AA model in order to achieve longer time scales. The CG model is coupled with a mesoscopic technique called *Multi-Particle Collision Dynamics* (MPCD) [34] to model solvent-peptide interactions and the correct hydrodynamics of the

system. The simulations are in non-equilibrium, since we introduce uniform flow to monitor the dynamics and aggregation of the silk proteins. The basic principles and concepts of the simulations at the two scales used in this thesis are explained in the Chapter 2.

### 1.3 Outline of the Thesis

The thesis is structured in 6 chapters: Chapter 1 introduces the topics and aims. Chapter 2 describes the physical principles and concepts of the computational techniques used in this work: the all-atom and coarse-grained (CG) MD simulations. In Chapter 3, we describe the implementation of uniform flow at microscopic scales by using MD simulations, a step prior to the subsequent model of silk proteins under flow. Chapter 4 investigates the dynamics of single dragline silk peptides under constant flow using multi-scale MD simulations. The directed self-assembly by flow is analyzed in Chapter 5. We summarize in the Chapter 6, in which we conclude the findings of Chapters 3, 4 and 5 and outline the future perspectives of this study.



# Simulation Methods

## 2.1 Molecular Dynamics simulations

MD is a computational technique for studying the dynamics of molecular systems from the micro-scale to the meso-scale [35]. This method permits to compute not only equilibrium but also transport properties of these systems. For highly dynamic molecules such as proteins and their complexes, conformational changes on time and space encompass several orders of magnitude that determine the behavior of these biomolecules. A detailed knowledge of the dynamics of biomacromolecular systems is difficult to obtain from experiments directly. As a consequence, MD simulations have been widely used to complement biophysical techniques, as they provide the hidden atomistic or molecular details behind the dynamics and thermodynamics of proteins [36]. In essence, MD simulations study the time evolution of many-body systems numerically, from which we get average observables of interest over an adequately long time. The comparison of MD simulations with experiments is based on foundations of statistical mechanics, which predict the macroscopic behavior of a system taking into account the microscopic properties of the particles that conform the system. In this chapter, we cover briefly the basic concepts before introducing the algorithm of classical MD simulations.

## 2. SIMULATION METHODS

---

### Statistical ensembles and averages

In general, a statistical ensemble is a hypothetical set of a large number of  $N_s$  identical systems, where every element of the set represents a possible state in which macroscopic system might be in. A real physical system in the macro-scale performs fast and stochastic transitions between its quantum states. Any macroscopic measurement detects only an average of the properties of the very large number of quantum states, it leads to a fundamental postulate: *a closed macroscopic system samples every allowed quantum state with equal probability.*

A micro-state of the system can be defined by the atomic positions  $\vec{r}$  and momenta  $\vec{p}$ , which correspond to the coordinates in a  $6N$  dimension space called phase space. A point in the phase space constitutes one state of the system. The system is often defined macroscopically in terms of parameters such as number of particles  $N$ , energy  $E$ , temperature  $T$ , pressure  $P$  and volume  $V$ . The collection of points in the phase space that satisfies a particular state of the system is a thermodynamic ensemble.

We can define a *microcanonical* or NVE *ensemble* based on the fundamental postulate described above and considering that a closed system does not exchange energy and particles with its environment: a statistical set of identical isolated systems with constant number of particles, volume and energy. Frequently, the theoretical background in statistical mechanics is based on this ensemble.

In an experiment, usually a series of measurements of an observable are performed in a macroscopic sample over a period of time, and the sample contains an enormous amount of atoms or molecules. Hence, an experimental measure is the result of sampling in an extremely large number of possible conformations from which we get an average value. Theoretically, ensemble averages of an observable  $A(p^N, r^N)$  correspond to the averages in experiments, and they are described by

$$\langle A \rangle = \int \int dp^N dr^N A(p^N, r^N) \rho(p^N, r^N), \quad (2.1)$$

where  $\rho(p^N, r^N)$  is the probability density function.

The ensemble average is highly difficult to calculate, since one must include in the calculation all possible states of the system. In the case of MD simulations, the states or points in the phase space of the ensemble are calculated consecutively in time, and time-averages

of an observable are computed during a simulation time  $t$ :

$$\bar{A} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t A(p^N(t'), r^N(t')) dt' \approx \frac{1}{T_0} \sum_{i=1}^{T_0} A(p_i^N, r_i^N), \quad (2.2)$$

where  $T_0$  is the total number of time steps of the simulation, and  $A(p_i^N, r_i^N)$  is the instantaneous value of the observable. The system will pass eventually through all its possible states if we let it evolve indefinitely in time. The time average is equal to the ensemble average, which is the statement of the *ergodic hypothesis*:

$$\bar{A} = \langle A \rangle. \quad (2.3)$$

The hypothesis means all accessible micro-states that satisfy a macro-state have the same probability of being chosen, so that the system will be the same period of time in each of these microscopic states. To study system of molecules by MD simulations we assume the ergodic hypothesis. Hence, if enough representative conformations are generated through MD simulations in an NVE ensemble to satisfy the ergodic hypothesis, the time-average can be used to compute the observables of the simulated system. In order to compute the ensemble average from MD simulations, one must assure that the system goes through all accessible states. However, sampling the whole conformational space through MD simulations is a challenge, in particular for macromolecules. In practice, however, it is complicated to find out the energy for a specific thermodynamic macro-state of a system. In experiments, systems under constant temperature are more accessible than systems under constant energy, these are the cases of NVT and NpT ensembles. An NVT ensemble corresponds to the *canonical ensemble*, in which the probability density of the possible micro-states is described by

$$\rho(p^N, r^N) = \frac{1}{Q} e^{-H(p^N, r^N)/k_B T}, \quad (2.4)$$

where  $H$  is the Hamiltonian,  $k_B$  the Boltzmann's constant,  $T$  the temperature and  $Q$  the partition function. The partition function measures the volume of an ensemble in the configuration space, and it is given by

$$Q = \int dp^N \int dr^N e^{-H(p^N, r^N)/k_B T}. \quad (2.5)$$

From the probability density and the partition function we can calculate the average value of an observable, i.e.:

$$\langle A \rangle = \frac{1}{Q} \int \int dp^N dr^N A(p^N, r^N) e^{-H(p^N, r^N)/k_B T}, \quad (2.6)$$

## 2. SIMULATION METHODS

---

In the case of an NpT ensemble, the pressure remains constant by adjusting the volume of the system. This ensemble allows to model better real life conditions of proteins, since chemical reactions in condensed phase normally occur under constant pressure, both in vivo and in experiments.

### Temperature Coupling

A *thermostat* or temperature coupling controls the temperature of the system by adjusting particles' velocities. The instantaneous kinetic temperature  $\mathcal{T}$  of an N-body MD system is defined by the equation:

$$\frac{3}{2}k_B\mathcal{T} = \left\langle \frac{mv_i^2}{2} \right\rangle = \frac{1}{N} \sum_{i=1}^N \frac{mv_i^2}{2}, \quad (2.7)$$

averaging  $\mathcal{T}$  over many microscopic states gives the temperature of the system, i.e.

$$T = \frac{1}{N_{steps}} \sum_k \mathcal{T}(\Gamma_k). \quad (2.8)$$

When the instantaneous temperature  $\mathcal{T}$  differs from the target temperature  $T$ , the velocity is rescaled, in the simplest form of a thermostat by:

$$\vec{v}_i' = \sqrt{\frac{T}{\mathcal{T}}} \vec{v}_i. \quad (2.9)$$

However, Equation 2.9 does not reproduce a distribution of the velocities in thermodynamic equilibrium, which leads to residual distortions in the velocity distribution functions and violates the kinetic energy equipartition theorem. Temperature coupling using the velocity rescaling thermostat (CSVr) overcomes this problem [37, 38]. It has been shown that this thermostat is very stable and ergodic, it reproduces the correct canonical ensemble and satisfies the kinetic energy equipartition theorem. The CSVr thermostat mimics weak coupling to an external heat bath. Additionally, a stochastic term ensures a correct kinetic energy distribution, which is done through

$$dK = (K_0 - K) \frac{dt}{T_c} + 2 \sqrt{\frac{K K_0}{N_f}} \frac{dW}{\sqrt{T_c}}, \quad (2.10)$$

where  $N_f$  is the number of degrees of freedom,  $K$  is the kinetic energy,  $K_0$  is the target kinetic energy,  $T_c$  is the temperature coupling time constant,  $dW$  is a Wiener process, and  $dt$  is the time step. While other thermostats have been suggested that also yield a correct canonical ensemble, CSVr proved particularly stable for MD simulations of condensed matter systems and was used in this thesis.

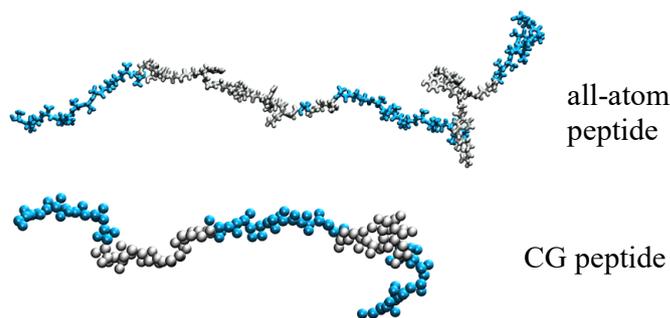
### Pressure Coupling

To simulate an NpT-ensemble, a *barostat* or pressure coupling is needed to keep the pressure of the system constant, by adjusting the volume of the system. In the simplest version of a barostat, coordinates are rescaled, although this does not result in a thermodynamically defined ensemble, violating Newtonian mechanics. The Parrinello-Rahman barostat [39] overcomes this problem by using an extended ensemble method. The system is coupled to an external variable  $V$ , which is the volume of the simulation box. Furthermore, each unit vector of the unit-cell is independent, such that the simulation box can also change its shape. In all NpT simulations of this thesis, the Parrinello-Rahman barostat was used.

### 2.1.1 Force Fields

A force field is an analytical function that models the different interactions of the particles that compose the MD system, in other words, a force field corresponds to the potential energy function of the system. It is parametrized based on the different types of atoms and their local interactions. The parameters of the force field are adjusted to reproduce experimental or theoretical data.

In the context of this thesis, we model a representative fragment of silk spidroin, consisting of three polyalanine repeats separated by two glycine rich regions, and with a total length of  $\sim 80$  aminoacids. Figure 2.1 shows all-atom and CG representations of the repetitive fragment we simulated.



**Figure 2.1:** Modelling of a silk peptide. Representative structures of the atomistic and coarse-grained (CG) models of a fragment of the repetitive region of the silk protein, containing three poly-Ala and two amorphous regions, resulting in  $\sim 80$  aminoacids in total. Spheres represent atoms (for all-atom) or backbone and sidechain beads for each aminoacid (CG).

## 2. SIMULATION METHODS

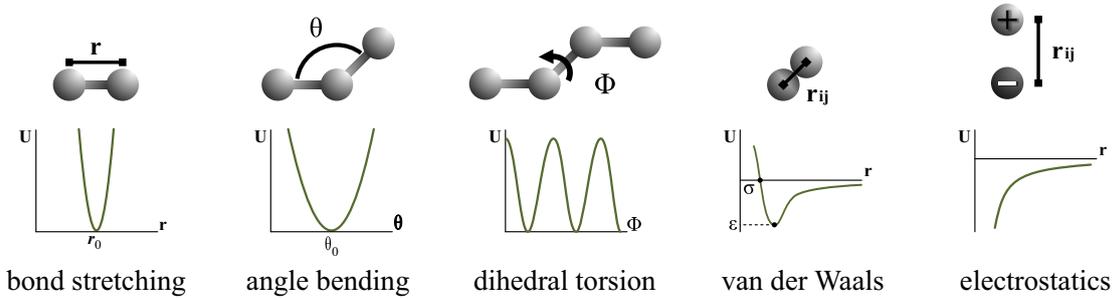
In the all-atom MD model we simulated the silk peptide in explicit water, while in the CG-MD simulations the solvent was modeled mesoscopically. In the next two sub-sections, we describe the two force fields we used in the multi-scale modelling of silk.

### All-atom force field

The potential energy is expressed as a sum of bonded and non-bonded interactions:

$$U(\vec{r}_i) = U_{\text{bonded}}(\vec{r}_i) + U_{\text{non-bonded}}(\vec{r}_i), \quad (2.11)$$

The explicit bonded terms are bond-stretching, angle-bending, and dihedral torsion, while non-bonded interactions comprise a Lennard-Jones or van der Waals part and an electrostatic part, see Figure 2.2. The bonded expressions are described mathematically by the equations:



**Figure 2.2:** All-atom force field. Schematic of the potential energy terms for all-atom MD simulations, which are divided in bonded and non-bonded interaction types. From left to right: bonded-stretching, angle bending and dihedral torsion correspond to equations 2.12, 2.13, and 2.14, while van der Waals and electrostatics correspond to 2.15 and 2.16.

$$U_{\text{bonds}}(r_{ij}) = \frac{1}{2}k_{ij}^r(r_{ij} - r_{ij}^0)^2, \quad (2.12)$$

$$U_{\text{angles}}(\theta_{ijk}) = \frac{1}{2}k_{ijk}^\theta(\theta_{ijk} - \theta_{ijk}^0)^2, \quad (2.13)$$

$$U_{\text{dihedrals}}(\phi_{ijkl}) = k_{ijkl}^\phi \left[ 1 + \cos(n\phi_{ijkl} - \phi_{ijkl}^0) \right]. \quad (2.14)$$

Equations 2.12, 2.13 and 2.14 are 2, 3 and 4-body interactions respectively. The bond lengths  $r_{ij}^0$ , angles  $\theta_{ijk}^0$  and dihedrals  $\phi_{ijkl}^0$  in equilibrium and the spring constants  $k_{ij}^r$ ,  $k_{ijk}^\theta$  and  $k_{ijkl}^\phi$  are obtained from experimental data or *ab initio* calculations.  $U_{\text{bonds}}$  and  $U_{\text{angles}}$  are the energies associated with the internuclear distance between two atoms  $r_{ij} = |\vec{r}_i - \vec{r}_j|$ , and the angle  $\theta_{ijk}$  between three consecutively bonded atoms. Both are described by harmonic

potentials. For a set of four consecutively bonded atoms,  $U_{\text{torsion}}$  represents the angular potential between the planes formed by the first three atoms and the last three atoms, given by the torsional angle  $\phi_{ijk}$

The non-bonded interactions of van der Waals and Coulomb are described by the potential energy functions:

$$U_{\text{LJ}}(r_{ij}) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (2.15)$$

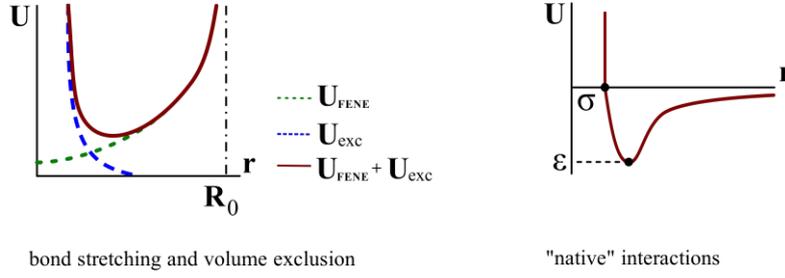
$$U_{\text{Coulomb}} = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q_i q_j}{r_{ij}}. \quad (2.16)$$

The Lennard-Jones term  $U_{\text{LJ}}$  combines short-range Pauli repulsion and attraction of induced dipole - induced dipole. The width of the potential is given by  $\sigma_{ij}$ , the depth by  $\epsilon_{ij}$ . Furthermore, the  $U_{\text{Coulomb}}$  corresponds to the long-range electrostatic interactions: between two charges  $q_i$  and  $q_j$ , where  $\epsilon_0$  and  $\epsilon_r$  are the vacuum permittivity and the relative permittivity respectively. The cutoff used for the non-bonded potentials is typically between 1.0 and 1.2 nm. However, as Coulombic interactions are comparably long-range, an approximation is used to calculate  $U_{\text{Coulomb}}$  for  $r_{ij}$  larger than the cut-off, i.e. by particle-mesh methods, which take advantage of fast Fourier-transforms. A particle-mesh method was also used in the all-atom MD of the thesis.

### CG force field

At coarse-grained (CG) level, we described the protein at the resolution of two beads per aminoacid (backbone bead and sidechain bead) rather than atoms. The two beads are located on the  $C_\alpha$  and  $C_\beta$  atoms of the protein. The peptide is simplified as a block copolymer composed of two types of aminoacids: alanines (Ala) and a non-specific aminoacid type for the disordered region 2.1. Since the primary intermolecular interaction within silk fibers are  $\beta$ -sheets formed from the poly-Ala regions, we considered attractive (native) interactions only between Ala residues in our model, more specifically between pairs of alanine backbone beads (representing the backbone hydrogen bonding within  $\beta$ -sheets) and between pairs of alanine sidechain beads (representing hydrophobic packing between methyl groups across  $\beta$ -sheets) [40]. These 'native' interactions were modelled as attractive Lennard-Jones interactions (Equation 2.15), all other bead-bead interactions as purely repulsive terms. We now describe all interactions of our spidroin mesoscopic SOP model including sidechains (SOP-SC) [41] in detail. The total potential energy  $U_{\text{total}}$  to describe the interaction between these

## 2. SIMULATION METHODS



**Figure 2.3:** CG force field. Schematic of the potential energy terms for CG MD simulations. From left to right: FENE and volume exclusion potentials (Equations 2.18 and 2.21) and the native Lennard-Jones potential (Equation 2.19).

monomers is based on the SOP-SC polymer model [41]. It is defined as (Figure 2.3)

$$U_{\text{total}} = U_{\text{FENE}} + U_{\text{nat}} + U_{\text{exc}}. \quad (2.17)$$

The first term  $U_{\text{FENE}}$  is a modified form of a Finitely Extensible Nonlinear Elastic (FENE) potential for the bonded interactions, i.e. beads are connected by nonlinear springs. The summation runs over all bonded backbone-backbone (BB) and backbone-sidechain (BS) beads. The potential is:

$$U_{\text{FENE}} = - \sum_i^N \frac{k}{2} R_0^2 \log \left( 1 - \frac{(r_i - r_0)^2}{R_0^2} \right) \quad (2.18)$$

where  $r_0$  is the equilibrium length of the bond,  $R_0$  is the maximum extent of the bond and  $k$  is the spring constant. The second term in equation (1) represents native interactions between the alanine pairs  $i,j$  such that  $|i - j| > 2$ . It is modeled with Lennard-Jones (LJ) potential with equilibrium distance  $\sigma$  and energy  $\epsilon$ :

$$U_{\text{nat}} = \sum_{i,j}^{\text{Ala-pairs}} U_{\text{LJ}}^{i,j} \quad (2.19)$$

with

$$U_{\text{LJ}} = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]. \quad (2.20)$$

The indices  $i$  and  $j$  of the Equation 2.19 runs over all Ala-pairs. The distance of the energy minimum between BB and sidechain-sidechain (SS) beads is 0.5 nm, corresponding to the distance when there is  $\beta$ -sheet secondary structure formation in silk. i.e.  $\sigma = 0.5$  nm. BS

interactions within alanines are not treated as attractive interactions, but are included in the repulsive term (see below).

The last term represents exclusion interactions:

$$U_{\text{exc}} = \sum_{i,j \notin \text{Ala-pairs}} U_{\text{LJ,rep}}^{i,j} \quad (2.21)$$

and

$$U_{\text{LJ,rep}} = \begin{cases} U_{\text{LJ}}, r < r_{\text{cut}}, \\ 0, \text{ otherwise} \end{cases} \quad (2.22)$$

where  $r_{\text{cut}} = r_m$  is the distance at which the potential has a minimum value. The summation in Equation 2.21 runs over all possible pairs of beads, except from native interaction pairs.

### 2.1.2 Multiparticle Collision Dynamics

To model solvent-peptide interactions in the CG model, the simulations are combined with Multiparticle Collision Dynamics (MPCD). It is a mesoscopic hydrodynamics technique [34]. It involves two steps:

(i) a streaming step where the solvent particles of mass  $m$  move ballistically during a collision time step  $\Delta t_c$ . Particle positions at every timestep are described by

$$\vec{r}_i(t + \Delta t_c) = \vec{r}_i(t) + \Delta t_c \vec{v}_i(t), \quad (2.23)$$

where  $\vec{r}_i$  and  $\vec{v}_i$  are the positions and velocities of every particle. Position update follows thus a standard integration of Newton's equation of motion (Section 2.1.3)

(ii) The collision step in which the particles are assigned to cells of size  $a$  in a lattice, the center-of-mass velocity of the cells  $\vec{u}_{\text{com}}$  is computed, and subsequently the relative velocities of the particles with respect to  $\vec{u}_{\text{com}}$  are rotated around a random axis by an angle  $\alpha$ , which ensures linear momentum conservation and thermal fluctuations. The velocity evolution for every particle in time is given by

$$\vec{v}_i(t + \Delta t_c) = \vec{u}_{\text{com},j}(t) + \vec{\Omega}_j(\alpha)[\vec{v}_i(t) - \vec{u}_{\text{com},j}(t)]. \quad (2.24)$$

$\vec{\Omega}$  and  $j$  correspond to the rotation matrix in spherical coordinates and the cell index respectively.

## 2. SIMULATION METHODS

---

The monomers of the peptide are included in the MPC collision step to take into account solvent-peptide interactions. Therefore, integration time step  $\Delta t$  for MD (see Section 2.1.3) is chosen to be smaller than the collision time step  $\Delta t_c$ . The simulation parameters are in terms of the basic units  $k_B T = 1$ ,  $a = 1$  and  $m = 1$ , where  $k_B$  is the Boltzmann constant and  $T$  is the temperature.

### CG simulation parameters

The parameters associated with the potential energy are based on the SOP-SC model proposed previously for proteins [41]:

- $U_{FENE}$  potential: bond length  $r_0 = 1.0 [a]$ ,  $k = 500 [k_B T/a^2]$  and  $R_0 = 0.5 [a]$ .
- $U_{ALA}$  potential:  $\epsilon_{ALA} = 0.25 [k_B T]$  and  $\sigma_{ALA} = 2.0 [a]$  of the LJ potential for both BB and SS interactions.
- $U_{exc}$  potential: the parameters have the same value for BB, SS and BS interactions, which are  $\epsilon_{rep} = 1.0 [k_B T]$  and  $\sigma_{rep} = 1.0 [a]$ .

### 2.1.3 Equations of Motion

Both for all-atom and coarse-grained simulations (with the additional MPCD procedure), MD simulations are employed. The MD simulation method is based on Newton's second law for AA and CG simulations:  $\vec{F} = m\vec{a}$ , where  $\vec{F}$  is the force exerted on the particle,  $m$  is its mass, and  $a$  is its acceleration. The force acting on every atom is defined by the gradient of the potential energy function  $U(\vec{r}_i)$  with respect to each atom's position:

$$\vec{F}_i(\vec{r}_i) = -\frac{\partial U(\vec{r}_i)}{\partial \vec{r}_i}. \quad (2.25)$$

Once the individual forces are computed through the Equation 2.25, all forces exerted in a certain atom  $i$  by another atom  $j$  are summed up:

$$\vec{F}_i = \sum_j \vec{F}_{ij} \quad (2.26)$$

This sum 2.26 is placed into Newton's equation of motion:

$$\frac{d^2 \vec{r}}{dt^2} = \frac{\vec{F}_i(\vec{r}_i)}{m_i}, \quad (2.27)$$

$$\frac{d\vec{r}}{dt} = \vec{v}_i. \quad (2.28)$$

Having the initial coordinates and velocities of the system's atoms, we can solve the Newton's equation of motion 2.27 and 2.28 numerically.

Time integration of the equations of motion yields the trajectory of the system, which describes the positions, velocities and accelerations of the particles. Considering the ergodic hypothesis (Equation 2.3), the observables of the system can be determined from the trajectory. The integration of the equations of motion is done in this thesis through the Verlet algorithm [42], which calculates the positions  $\vec{r}_i$  and velocities  $\vec{v}_i$  at the times  $t$  and  $t - \frac{\Delta t}{2}$  respectively:

$$\vec{v}\left(t + \frac{\Delta t}{2}\right) = \vec{v}\left(t - \frac{\Delta t}{2}\right) + \frac{\vec{F}(t)}{m}\Delta t, \quad (2.29)$$

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}\left(t + \frac{\Delta t}{2}\right)\Delta t. \quad (2.30)$$

The time steps  $\Delta t$  are chosen in a way that they are smaller than the fastest motions of the simulated system, which are the covalent bond vibrations. The vibrational frequencies of a covalent bond can be expressed through:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}, \quad (2.31)$$

where  $\mu$  is the reduced mass of a bond and  $k$  is the spring constant, i.e. the faster the atom, the higher the vibrations. MD simulations of biomolecules with chemical bonds to hydrogens are stable considering only time steps of 1 fs or lower. Usually, the dynamics of atoms relative to one another are not of interest. Hence, constraints of high frequency motions like hydrogen bond vibrations allow to increase the time step to  $\sim 2$  fs. Bond-distances constraints are introduced with algorithms such as LINCS [43, 44].

### Boundary Conditions

The number of particles contained in molecular systems is significantly smaller than any macroscopic system. Periodic Boundary Conditions (PBC) are adopted to avoid surface effects and have been employed in all-atom MD simulations presented herein. PBC conditions consist of system replications along all three Cartesian coordinates. Particles that exceed the limits of the simulation box on one side are moved to the other side of the simulated system. The box dimensions should be a multiple of  $\sigma$  for Lennard-Jones fluids. For protein simulations, such as those performed here, a distance to the periodic image larger than the

## 2. SIMULATION METHODS

---

non-bonded cut-off distance is used. In this way, the interaction of a certain particle with its periodic image is negligible.

# MD Simulations of Molecules in Uniform Flow

This chapter has been published in: Herrera-Rodriguez A. M., Vedran M., Aponte-Santamaria C. and Gräter F., "Molecular Dynamics simulations of molecules in uniform flow", *Biophysical Journal*, vol. 116, pp.1579-1585, 2019.

## 3.1 Introduction

Many biological processes such as hemostasis to stop bleeding, leukocyte adhesion at sites of inflammation, or spider silk spinning occur under flow conditions [25, 45, 46]. Flow, being it elongational or shear flow, is an inherent property of these systems. The proteins involved in the flow-dependent processes undergo specific conformational changes triggered or influenced by the flow. Examples of this are the von Willebrand factor (vWF) in the blood or spider silk proteins. vWF plays a crucial role in sensing blood flow to promote platelet adhesion at wound sites for hemostasis[46, 47, 48, 49]. It undergoes a flow-induced globule-to-stretch transition which activates its adhesive function. Spider silk proteins self-assemble into fibers of outstanding mechanical properties[13], a transition which is triggered by acidification, ion exchange and elongational flow through a precise mechanism that remains to be fully unravelled [25]. The key role of flow in silk fiber spring has been described in depth in section 1.1.

### 3. MD SIMULATIONS OF MOLECULES IN UNIFORM FLOW

---

The molecular details of flow-induced protein dynamics are often unknown. However, experiments along with computer simulations have been able to increasingly shed light on these details at mesoscopic scale. Computational methods that have proven highly successful to study flow-dependent processes include multi-particle collision dynamics (MPCD) [34], lattice Boltzmann (LB) [50] methods and dissipative particle dynamics (DPD) [51, 52]. Those computational fluid dynamics (CFD) methods are used to model a fluid by either a discretization of the Boltzmann equation (LB) [50] or a solution of the Navier Stokes equations. An example for the latter is the MPCD method which uses collision models, where the fluid particles are under streaming and collision processes that reproduce the dynamics of the flow (for details see section 2.1.2). For the DPD method, the particles interact through three types of forces: a conservative force arising from a potential energy function, a friction force to reduce the relative velocities between the particles and a stochastic force accounting for the random thermal collisions with the solvent molecules [51].

The main advantage of CFD methods is the accessibility of large computational time and length scales relevant for the biological process under study. However, important molecular aspects, such as hydrophobicity, hydrogen bonding, and entropic effects due to the nature of the solvent molecules are not taken into account, since molecules are represented in a highly coarse-grained manner. Mostly, bead-spring models are used for the proteins, with one bead representing full proteins, protein domains, or aminoacids. As the detailed atomistic structure of the proteins critically influences their molecular properties and conformations under flow, an atomistic simulation scheme with explicit molecular water under flow is required. A very versatile way to induce explicit uniform (or shear) flow is to apply an additional constant force to a subset of wall molecules, which is the method of choice here. Such implementation has been made available for the NAMD package [53, 54, 55], one of the major Molecular Dynamics suites, where the flow is implemented as a Tcl script that can be modified by the user. However, the validity range of the flow under different parameter sets has not been assessed in this implementation. Gaussian dynamics [56], a method for non-equilibrium MD simulations was recently developed, to model uniform flow. Based on the Gauss Principle of least constraints [57]: the constraint keeps the total mass flux constant by an external force that drives the system to satisfy this constraint. Also, an appropriate thermostat was designed in order to maintain the steady state and dissipate the heat of the system. Shear and uniform flow has been implemented previously to CHARMM [58] and employed to investigate the dynamical behavior of free and short tethered chains of ss/ds DNA molecules in shear flow [59]. For ubiquitin unfolding in uniform flow [60], flow has

been implemented through the boundary driven method [61] by translating surfaces either of gold or frozen water in opposite directions (shear flow) and in the same direction (uniform flow). They reported high shear rates  $0.5 - 48 \times 10^{10} \text{ s}^{-1}$  and uniform flow velocities up to 180 m/s, which take place in particular cases where the fluid properties are not like bulk fluid, for example in polymer thin films of thicknesses of about 3 – 4 molecular diameters [61]. Explicit shear flow, in turn, can also be produced by defining two parallel walls where one of which moves with constant velocity [62]. However, such a setup does not offer to easily predefine the shear rate nor to produce uniform flow.

We find very appropriate to implement and systematically test flow MD simulations in one of the most popular MD software packages, the GROMACS package. It offers the possibility to model a high variety of biomolecular systems. Therefore, we present an efficient and easy-to-use implementation of uniform flow in MD simulations with explicit solvent that can be easily extended. Our implementation reuses and extends the pull code available in GROMACS version 2016 [63] and is configured using the same declarative configuration directives as other features in GROMACS. We show that such implementation is very flexible and offers high performance, with a slowdown of the simulation within few percent compared to simulations without flow. By probing a large parameter range, we show that our flow MD scheme is physically realistic, we describe for which flow velocities it fails, and demonstrate its applicability to biological systems by simulations of an unfolded peptide in flow. Our GROMACS extension represents a simple way to study various flow-dependent molecular processes at atomistic scale, and will be used in Chapters 4 and 5 of this thesis.

## 3.2 Methods

### Extension of the center-of-mass pulling code

Constant slice pulling, which is used for introducing flow to the system, was implemented by reusing and extending the center-of-mass (COM) pulling code available in GROMACS<sup>1</sup> [64, 65]. The existing pull code works by applying forces (or constraints) between the centers of mass of one (or more) pairs of groups of atoms. Each pull coordinate operates on two (or more, depending on the coordinate type) pull groups. Each pull group can be utilized by one or more pull coordinates. Furthermore, a coordinate can also utilize one group and an

---

<sup>1</sup>Documentation and software can be found here (sliced pull section): DOI:10.5281/zenodo.2538871

### 3. MD SIMULATIONS OF MOLECULES IN UNIFORM FLOW

---

absolute reference position in space. The pulling can also be done in the direction of a user-defined vector. By default, all atoms are weighted by their mass[38], but an additional user-specified weighting factor can be used to smooth the external potential, it avoids abruptly changes in the external applied force, we call this option smoothed pull slice case.

The extension of this code is done as follows: a new type of a pull group, the sliced pull group, is introduced. The position of the slice is defined by the minimum and maximum value of the  $x$  coordinate of the group; no force will be applied on the atoms in the group positioned outside of the range specified by the minimum and maximum. Note that slicing along the  $x$  coordinate is sufficiently general, as a box used in an MD simulation can be freely rotated. Still, in case an application requires the slicing in multiple coordinates for one or more groups, the present code could be extended to support such use case. Three new variables in the MD parameter file were introduced for the configuration of the constant slice pulling. A boolean variable `pull-group1-sliced` is used to enable or disable the slicing of the group, while the `pull-group1-slice-x-min` and `pull-group1-slice-x-max` control the minimum and maximum value of the  $x$  coordinates which belong to the slice on which the pull force will be applied.

In each step of the MD simulation, the state of all the sliced pull groups is updated (ordinary, i.e. non-sliced pull groups are left unchanged). For each atom in each sliced pull group, the position of the atom is read and compared to the slice boundaries in  $x$  coordinate,  $s_{\min}$  and  $s_{\max}$  (configured by the user for each of the sliced groups by setting the MD parameters). The external potential acting on the group will only be applied to the atoms in the slice, weighted with weighting factor  $w_{\text{sp}}(x)$ , where  $x$  is the  $x$  coordinate of the atom. Coordinate  $x$  of each atom in the slice is first normalized from the interval  $[s_{\min}, s_{\max}]$  to the interval  $[-1, 1]$  using the formula

$$x_{\text{unit}} = \frac{x - \frac{s_{\min} + s_{\max}}{2}}{\frac{s_{\max} - s_{\min}}{2}} \quad (3.1)$$

Then the weighting factor  $w_{\text{sp}}(x)$  is computed by

$$w_{\text{sp}}(x) = x_{\text{unit}}^4 - 2x_{\text{unit}}^2 + 1 \quad (3.2)$$

The weighting factor is set to 0 for all atoms outside the slice. When the pull force is applied during an MD step, it will only be applied to atoms of the sliced pull group that reside in the sliced region and it will be weighted with an additional factor  $w_{\text{sp}}$ . In the special case where no atoms are present in the slice, the weighting would set  $w_{\text{sp}} = 0$  for all atoms. This

would result in the application of an infinite force and a subsequent crash of the simulation. Therefore, the number of atoms in the slice is checked before each weighting and, in case it becomes zero, the user is directed to check the simulated system and MD settings. Such "no atoms are present in the slice" situation can occur, for example, when the pull force is high enough to cause the appearance of voids big enough to encompass the entire slice.

### **Simulation setup for a pure water system**

To model the uniform flow by MD simulations we used the Groningen Machine for Chemical Simulations (GROMACS) 4.6 version package, in which the center-of-mass (COM) pulling code was extended by introducing a sliced pull group defined by minimum and maximum values of  $x$  coordinates, the range where the external potential that produces the flow is applied. At every time step of the simulation, the  $x$  positions of the pull group atoms are read, if a pull group atom is inside of the pull region, an external force will be exerted on this atom along the  $x$  coordinate.

The system for simulation is a rectangular box of about 32,000 water molecules and dimensions of  $39 \times 5 \times 5 \text{ nm}^3$ . We used the SPC water model [66]. Before producing the flow, we prepared the system as follows: potential energy minimization to set the system into an energetically favorable state, 5 ns canonical NVT ensemble, an NpT ensemble during 10 ns and production run at constant volume and temperature of 300 ns; an NpT ensemble is not performed in this latter stage, since the barostat would alter the generated flow due to a pressure gradient. The water box was simulated considering the following cases: smooth and not smoothed pull slice of 4nm, smoothed pull-slice of 8 nm and smoothed pull-slice of 24 nm. The type of barostat and thermostat used in the simulations are Parrinello-Rahman [39] and V-rescale[37] respectively, with coupling times 0.1 ps. A 2 fs integration timestep was set to all simulations. We use the Particle Mesh Ewald (PME)[67] technique for the long-range electrostatic interactions, with 4 order cubic interpolation and 0.16 nm grid spacing for FFT. The cutoff radius is 1 nm for electrostatic and Van der Waals interactions, with an update of the neighbor list every 10 fs. In flow simulations, the pull group was defined such that it comprises all oxygen atoms in the system.

### **Simulation setup of a peptide in uniform flow**

We simulated a small intrinsically disordered peptide of 40 amino acids, a linker which connects the vWF A1 and A2 domains [68]. Its initial structure was taken from our previous

### 3. MD SIMULATIONS OF MOLECULES IN UNIFORM FLOW

---

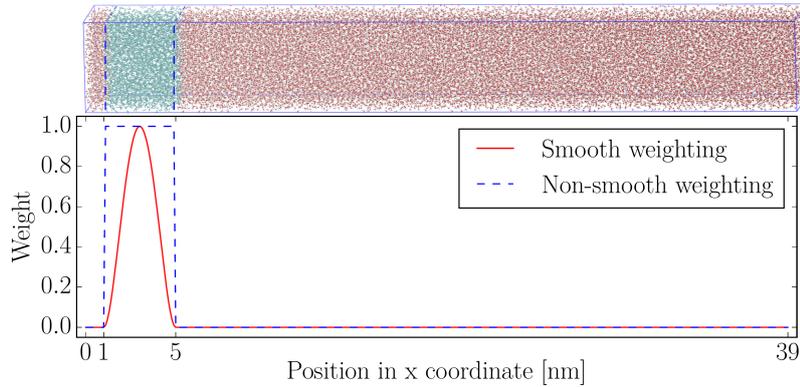
simulation study [69]. The Amber force field amberff99sb-star-ildn [70] in combination with the TIP4P-D water model was used [71]. This combination is appropriate to model accurately the collapse propensity of intrinsically disordered proteins [72]. To neutralize the system we added counter NaCl ions at a concentration of 0.15 mol/L. Periodic boundary conditions in all three dimensions were applied. V-rescale temperature coupling was used during equilibration and production run, and Parrinello-Rahman pressure coupling during system preparation with the LINCS constraints for the protein (see Section 2). Other simulation parameters were the same as those described for water-only system above. The protein was placed in a way that the closest atom to the slice is 4 nm away from the pull region to avoid undesired stretching not associated with the flow, and artifacts due to the pressure drop inside the slice and in the interface. To achieve a uniform flow with the ions included, the water molecules were subjected to a pulling force during 100 ns within a slice of 4 nm located at 1 nm from the origin, while maintaining the protein heavy atoms with position restrained (restraint elastic constant of 1000 kJ/mol/nm<sup>2</sup>), the solvent dynamics with ions is slower because NaCl ions are heavier than water requiring long equilibration times. This simulation was conducted maintaining the volume of the simulation box fixed (NVT conditions). For the production run, the protein was released and the first alpha-carbon of the peptide was position-restrained with a force constant of 1000 kJ/mol/nm<sup>2</sup>, otherwise the peptide would be dragged through the water box without being extended.

A uniform flow even for high wall shear rates of about  $10^2 - 10^8 \text{s}^{-1}$  is a reasonable approach for bulk fluids in the nanometer scale. As an example, the estimated flow velocity change within a distance of 100 nm for a wall shear rate of  $10^5 \text{s}^{-1}$  is 0.01 nm/ns. This is a negligible value, being valid to consider a constant flow

### 3.3 Results and Discussion

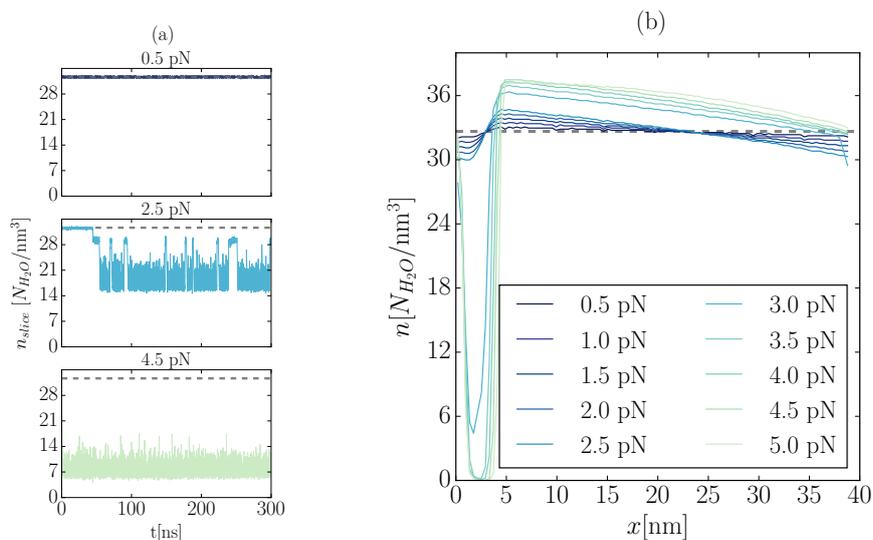
We first demonstrate the simple principle of flow MD simulations by showing results for a water-only system. We introduce flow at constant velocity by adding a pre-defined constant force acting along the flow direction to all water molecules present at any given time  $t$  within a vertical slice of the simulation system (Figure 3.1). The position and width of the slice, which we denote pull-slice in the following, are pre-defined by the user. Adding the force to only a subset of water molecules within a slice has the advantage of not directly perturb the surrounding of a macromolecule, such that it does not penetrate the slice where the force is applied.

Our extension of the pulling code in GROMACS was tested with a  $39 \times 5 \times 5 \text{ nm}^3$  rectangular box (Figure 3.1, top). Force is applied to water molecules within the slice boundaries after application of a weighting scheme. A smoothed distribution of external force along the flow direction is applied (see in Figure 3.1, bottom). At each MD step, a dynamic reassignment of the molecules belonging to the slice is performed, and the weighting process is repeated according to the number of water molecules found in this region.



**Figure 3.1:** Example setup for flow MD simulations. a) Rectangular water box (dimensions  $39 \times 5 \times 5 \text{ nm}^3$ ) with approximately 32,000 water molecules. Water molecules within the slice are highlighted in cyan. b) Slice weighting function considering smooth (red) and non-smooth (blue) options (equations 3.1 and 3.2, respectively, in Methods section) plotted along the  $x$  coordinate of the simulation box. The smoothing allows to avoid abrupt discontinuities in the applied force at the slice interfaces.

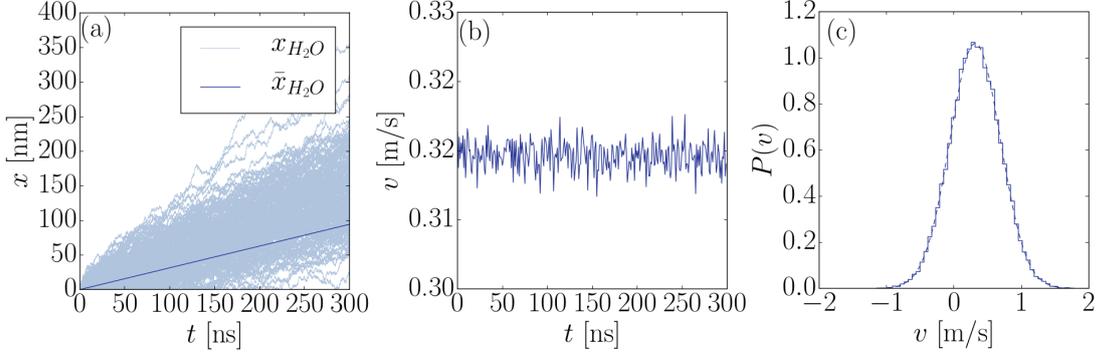
### 3. MD SIMULATIONS OF MOLECULES IN UNIFORM FLOW



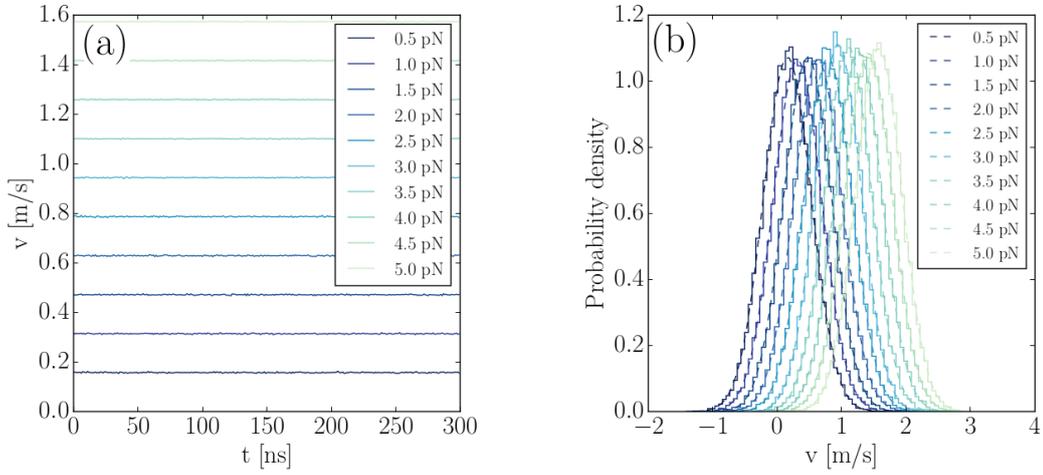
**Figure 3.2:** Water density varies along the flow direction. a) Number density of water molecules inside the slice over time under forces of 0.5 pN, 2.5 pN and 4.5 pN. The size of the slice is 4 nm; and force smoothing is applied. b) Number density of water molecules along the flow orientation. Simulation setup as in a). Up to forces of 2.5 pN, the density does not decrease more than 8.5%. The grey dashed line is the number density at zero force.

Figure 3.2a shows time evolution of the number density of the waters inside the pull-slice under different flow velocities, induced by application of varying external forces on the slice. There is a significant drop of the water density for forces larger than 2.5 pN. This implies simulations with this or higher forces to not result in a stable water flow, but instead to result in cavity formations. The external force applied in the region depends on the number of waters found in it, assuming constant density in the slice. High external forces lead to inhomogeneous densities, which in turn can lead to even higher forces applied to these molecules and eventually cavity formation. For the given water system, applying forces of 2.5 pN or more is thus not recommended. Further below, we will show the force regime for which a stable flow is obtained. We next analyzed the number density along the flow orientation (Figure 3.2b). Up to forces of 2.5 pN the density does not decrease more than 8.5 % compared to the number density when flow is absent. As discussed before, for forces larger than this value the number of water molecules inside the slice is drastically reduced. For all forces, at the slice interface at 5 nm, the water density reaches a maximum which is higher for higher force values. The externally applied force pushes the water molecules to this position, where the molecules, however, are slowed down as adjacent water layers are

not subjected to the force. As expected, the density then linearly decreases along the flow direction and is lowest at the entry zone of the pull-slice.



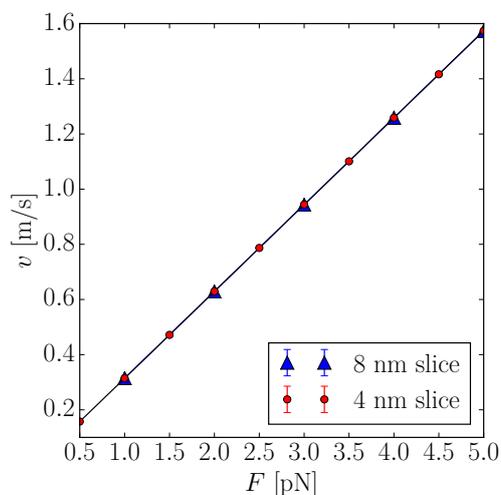
**Figure 3.3:** Flow MD produces uniform flow in the m/s regime. (a) Mean displacement of all water molecules (dark blue) as a function of time for a 1 pN force per molecule applied in the slice of 4 nm. Light-blue curves represent the displacement of single water molecules (shown for  $\sim 200$  molecules). (b) Average velocity over time obtained from the mean displacement (dark-blue in (a)). (c) Velocity probability density function obtained from a linear fit of the displacement curves per molecule (light-blue in (a)), the dashed line is a Gaussian-distribution fit.



**Figure 3.4:** Uniform flow velocities under different forces applied in the slice. (a) mean velocity over time and (b) velocity probability density.

### 3. MD SIMULATIONS OF MOLECULES IN UNIFORM FLOW

---



**Figure 3.5:** Mean velocity over average force per molecule in the pull slice. Slice sizes were 4 nm (circles) and 8 nm (triangles). The black curve is a fit of the data to a line.

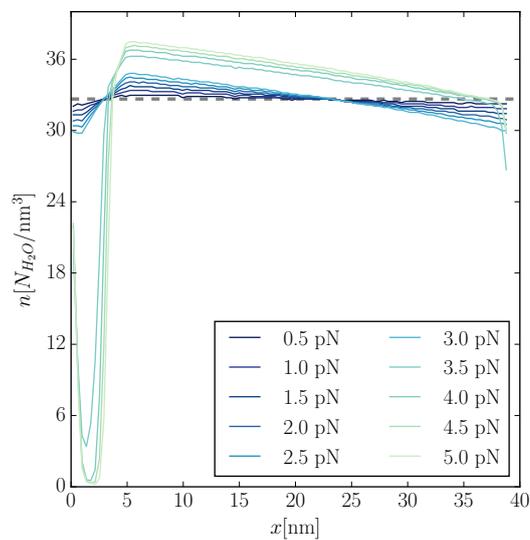
The mean displacement of the water molecules over time for a 1 pN force applied in the slice is shown in Figure 3.3a (solid line). It was defined as the average displacement of all water molecules in the system at every timestep. While individual water molecules show displacements which vary broadly over time, as exemplified by the light-blue curves in Figure 3.3a, the average over all water molecules is linear, implying a constant velocity of the flowing water over time, as aimed for. The mean velocity over time (Figure 3.3b) of the flow is obtained from the mean displacement curve using finite differences, obtaining a uniform flow with a constant velocity of 0.315 m/s for the given system. Figure 3.3c shows the velocity probability density function obtained from fitting the displacement of all individual water molecule displacements to a line from which individual molecular velocities were obtained. The distribution is Gaussian, suggesting that our system remains in a canonical ensemble. Velocity values from the average value of the distribution 0.32 m/s and the mean velocity over time are in good agreement. Higher forces applied to molecules in the slice result in correspondingly higher velocities, as shown in Figures 3.4a and 3.4b for forces from 0.5 to 5.0 pN.

We next asked how the obtained flow velocity changes with the externally applied force for a given size of the slice. Figure 3.5 shows a linear increase of the flow velocity with force, as expected. When only taking systems into account which show a largely homogeneous water density ( $< 2.5$  pN, see Figure 3.2) for a 4 nm pull slice, uniform flow with velocities up to 0.6 m/s can be obtained (data shown in red). However, this range can be increased towards higher velocities if larger pull slices are defined. When using external forces per molecule which are half the size and at the same time doubling the pull slice (8 nm), the same target velocity can be obtained, though at a smaller risk of density fluctuations across the simulation system (data shown in blue). Thus, higher target velocities require larger pull slices, or larger system sizes to prevent undesired drops in water densities along the flow dimension. As a demonstration of the smaller density fluctuations for larger pull slices, Figure 3.7 shows the time-averaged number density as a function of  $x$  for varying sizes of the slice: 4 nm, 8 nm, and 24 nm. All three systems display the same target flow velocity of 1.3 m/s. When the size of the slice is increased there is a reduction of the voids and the number density is closer to the value without flow (grey dashed line). In fact, for slices as small as 4 nm, the pressure gradient is artificially high due to the nano-sized dimensions of our system, but as we see here, this problem ameliorates when increasing the width of the pull size towards 24 nm, which therefore is a recommendation of pressure-sensitive systems. We also simulated flow for a non-smoothened version, i.e. the same force was applied to the water molecules within the slice, instead of using the weighting scheme (Figure 3.1 and Figure 3.6). We did not observe differences, suggesting that the thermostat is able to relax the net velocity of the system even for abrupt applied forces.

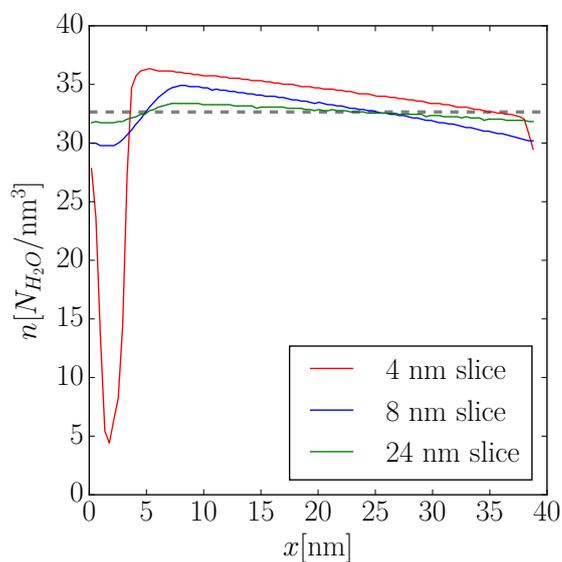
Finally, as proof-of-concept, we performed simulations of a short disordered peptide, an  $\sim 40$  amino-acid fragment of the link between A1 and A2 domains of the vWF protein, in flow. We chose an intrinsically disordered peptide because it adapts more strongly its conformation in response to flow than folded proteins. Our technique can be flexibly used for any type of molecule or material in flow. The peptide was placed away from the pull region (its closest atom to the slice is 4 nm away), and its N-terminal  $C_{\alpha}$ -atom was restrained (details of the simulation setup are given in the Methods section).

### 3. MD SIMULATIONS OF MOLECULES IN UNIFORM FLOW

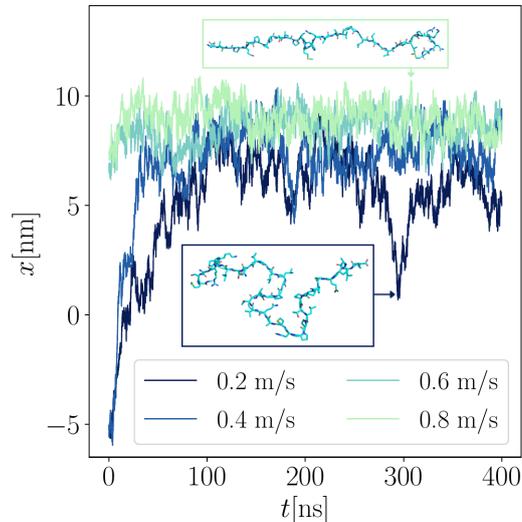
---



**Figure 3.6:** Number density of water molecules along the flow orientation without smoothing the applied force inside the slice.



**Figure 3.7:** Number density as a function of  $x$  at flow velocity of 1.3 m/s for three different pull slice sizes: 4 nm (red line), 8 nm (blue line) and 24 nm (green line). The density drop is reduced with the increase of size of the slice.



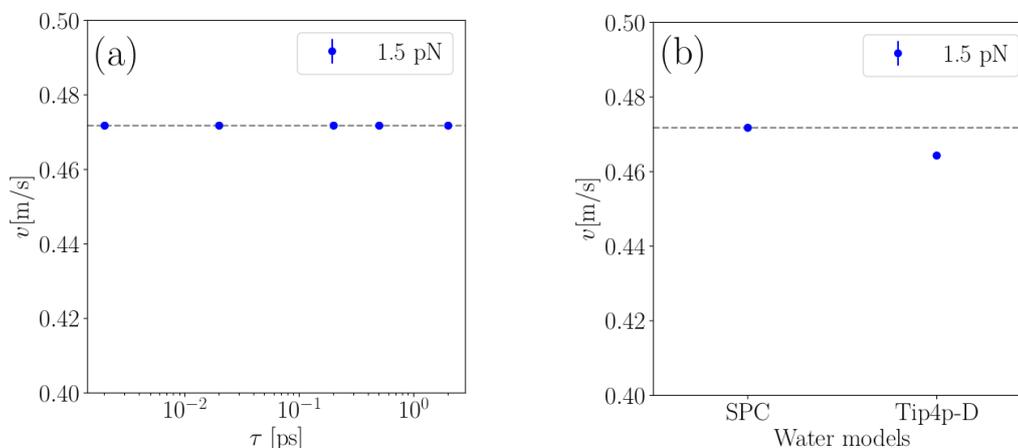
**Figure 3.8:** Flow-induced unfolding is illustrated for the 40 AA intrinsically disorder linker of A1 and A2 domains of vWF. The closest atom of the peptide to the pull slice is 4 nm away and the N-terminus with position restrained. Position along  $x$  (flow direction) of the C-terminus as a function of time is presented, varying applied forces between 0.3 and 1.2 pN. Note that for 0.3 and 0.6 pN, the peptide was initially oriented against the flow direction ( $x$ -position at  $t = 0$ ) and initially rotated. Two representative conformations before and after the flow-induced extension of the peptide are shown.

As shown in Figure 3.8, the peptide features increasing end-to-end distances over time along the direction of the flow when subjected to uniform flow. The peptide length reaches a plateau after  $\sim 100$  ns, around which the value then fluctuates. Fluctuations in length are more pronounced at low flow velocities, while high velocities more strongly straighten the peptide and thereby also reduce the length fluctuations. This behaviour follows the expectations from polymers under flow [73] and confirms the feasibility of the explicit flow MD protocol.

We also tested if the flow velocity and the resulting change in protein dynamics can in principle be affected by the temperature coupling. To this end, we coupled the system with the V-rescale thermostat [37]. It has been shown to be ergodic, stable, reproduce the canonical ensemble and not violate the kinetic energy equipartition theorem [74]. In Figure 3.9a we compare the velocity of the flow under different coupling times of the V-rescale thermostat (0.0002 – 2.0 ps). For a force of 1.5 pN in the pulling slice we obtained the same average velocity of 0.47 m/s. We conclude that our setup is robust with regard to the

### 3. MD SIMULATIONS OF MOLECULES IN UNIFORM FLOW

temperature coupling time.



**Figure 3.9:** (a) Velocity of the flow for a pulling slice force of 1.5 pN as a function of the coupling time of the thermostat. Flow is very stable under different coupling times. Standard errors are smaller than the symbol size and (b) Comparison of the flow velocity of a pure water system for SPC and Tip4p-D water models. The pulling force in the slice is 1.5 pN. The simulation setup is the same as the one described in the Methods section 2.3. Standard errors are smaller than the symbol size.

We note that for the simulations of the peptide under flow, we used TIP4P-D water, which has a lower self-diffusion than SPC water simulations described above ( $3.85 \times 10^{-5} \text{cm}^2/\text{s}$  versus  $2.1 \times 10^{-5} \text{cm}^2/\text{s}$ ). Despite this difference, the two water models give nearly the same average velocity for a given external force (Figure 3.9b, SPC gives 2% higher velocity than TIP4P-D).

## 3.4 Conclusion

Here, we present a protocol to induce a uniform flow in molecular dynamics simulations. This implementation enables a molecular understanding of flow-driven biological processes.

The implementation was tested for pure water systems with varying slices in which water molecules were subjected to different external forces along the desired flow direction. For 4 nm slices, the density remains relatively constant across the box up to 2.5 pN per molecule or 0.6 m/s velocities, while higher forces and velocities lead to instabilities. Larger sizes of the pull slice can reduce the high variance in densities, and allow even larger uniform flow velocities.

We demonstrated the usage of the new flow implementation for the flow-induced globule-to-stretch transition of a 40-residue peptide. The extension of this peptide was proportional to the flow velocity. This example demonstrates the possibility of using GROMACS to analyze the dynamics of proteins under flow at atomistic scale. Application areas range from proteins involved in blood coagulation to (bio)polymer assembly under flow conditions. The atomistic simulations might in future help to parametrize hydrodynamic simulations at a coarse-grained level such as DPD or MPCD. In addition, the minor negative performance impact we measured makes the usage of the implementation generally feasible.

Considering the typical length scales of MD simulations, to have a constant flow is a reasonable approximation, since shear rates can be considered negligible in the nanometer scale for bulk fluids. The uniform flow setup allows us to probe hydrodynamics interactions and to analyze the rheology of proteins in atomistic detail. Further extensions to allow non-uniform flows such as shear flow will be simple to implement and are subject of our future efforts.



# Dynamics of a Single Silk Peptide under Uniform Flow

## 4.1 Introduction

Water plays a major role in the dynamics, structure and stabilization of proteins, mainly via hydrogen bond networking and screening of electrostatic interactions. Hydrodynamic flow can impact protein structure in a very in a complex manner by inducing non-uniform drag force along the protein, which depends on the protein conformation and aminoacid sequence. Flow can mediate a large variety of processes such as material synthesis, blood coagulation and protein misfolding [75]. For instance, flow, ion exchange and pH changes along the *Major Ampullate* (MA) spinning gland drive the assembly of dragline spider silk proteins (spidroins) into a fiber, which outperforms mechanically any biomaterial made by nature [1, 6, 13, 25, 30, 76]. The flow-induced activation of von Willebrand factor (vWf) protein is crucial in hemostasis [47, 49, 77, 78]. Shear and elongational flows can lead to protein unfolding, misfolding and aggregation, and can accelerate fibrillization of amyloid- $\beta$  peptide [79, 80, 81, 82], suggesting a potential role of flow in degenerative diseases [83]. A molecular understanding of how a protein responds to flow, however, remains elusive.

Towards understanding Hydrodynamic Interactions (HI) of biomolecules, single-molecule experiments have been performed, typically of micrometer-size biopolymers like DNA and vWf [49, 73]. Experiments of a double stranded DNA held against uniform flow by optically

#### 4. DYNAMICS OF A SINGLE SILK PEPTIDE UNDER UNIFORM FLOW

---

trapping it on one side with a microsphere show that the DNA shows average extensions in line with a worm-like chain model [84], with chains in a dumbbell shape being more susceptible to shear-induced stretching [73]. vWf when studied in a microfluidic chamber unfolds in shear flow through a reversible globule-stretch transition beyond shear rates of about  $5000 \text{ s}^{-1}$  in the absence of an adsorbing surface [49].

A large body of work explored the dependence of silk fiber assembly under elongational and shear flows using microfluidics [6, 13, 25, 85]. In single molecule force spectroscopy experiments, single silk spidroin molecules largely followed a worm-like chain behavior, but also exhibited unique unfolding steps [86, 87]. Flow-dependent dynamics of single silk spidroin proteins, however, remain unknown.

Polymer theory and simulations have shed further light into the stretching behavior of biopolymers under flow. For the simplest case of biopolymers under uniform flow, Monte Carlo simulations of tethered DNA chains revealed that the nonlinear elastic dumbbell model predicts the deformation of them [88]. The simulations included Brownian motion, entropic elasticity, and variations in the drag coefficient while the chain is deformed. Brownian dynamics simulations of DNA molecules in steady flow showed interesting differences between WLC and Rouse models in a strong steady flow, with convective as opposed to diffusive propagation of tension along the chain, respectively, leading to a distinct time evolution of chain extension [89]. In addition, bead-spring models have proven useful for describing (bio)polymers, including proteins under flow conditions. A prominent example is the coarse-grained (CG) modelling of vWf multimers in shear flow, with one bead representing roughly one monomer and a Lennard-Jones potential to account for average attractive interactions [46, 90]. A variety of methods have been used in such mesoscopic simulations for the hydrodynamic interactions of the polymer with the solvent, based on Brownian Dynamics (BD) [46, 49, 78, 91], multi-particle collision dynamics (MPCD) [34, 92, 93, 94], dissipative particle dynamics (DPD) [51, 90], or lattice-Boltzmann [50].

In contrast to these mesoscale simulations, atomistic MD simulations can give detailed insight into the conformational changes of proteins induced by flow, such as ubiquitin unfolding [60] or secondary structure transitions in the platelet receptor glycoprotein Ib [54]. To model uniform flow, solvent molecules are subjected to a constant force in an NVT ensemble, yielding a constant terminal velocity. The high degree of detail offered by atomistic MD in explicit solvent comes at a high computational cost, resulting in high flow rates used in these studies.

In this study, we examined the extension behavior of spidroin, an intrinsically disordered protein that is highly sensitive to hydrodynamic flow, [6, 25, 30] at molecular detail and at physiologically relevant flow velocities. Major ampullate spidroins consists of pH-dependent folded N-terminal and C-terminal domains, which experience conformational changes from pH 7.0 down to pH 5.0 in the spider spinning gland, thereby initiating assembly [13]. In between these terminal domains, spidroins feature a long repetitive region, which makes up roughly 90% of their total sequence (Figure 4.1a) [95]. This repetitive region of silk proteins comprises alternating blocks of hydrophobic alanines, which form  $\beta$ -sheet stacks in the silk fiber, and hydrophilic glycine-rich repeats, which form the amorphous and extensible matrix of silk [1]. Flow as present during fiber spinning is essential for conformational transitions of the spidroin repetitive region such that assembly into the fiber is promoted.

We present simulations of a single chain of the repetitive region of spidroin in uniform flow, using a combination of mesoscopic and atomistic simulations. We chose an  $\sim 80$  aminoacid-long sequence from the MA *Euprosthenops Australis* dragline spidroin, which consists of three poly-ALA repeats, and two amorphous regions (Figure 4.1a). We modeled this protein at the mesoscopic scale as a simple block-copolymer using a Go-like potential [41] combined with Multiparticle-Collision Dynamics (MPCD) [34] to model a uniform flow. For comparison, we performed atomistic MD simulations of the same spidroin fragment in uniform flow of explicit water [65]. We find the protein to largely follow the expected freely-jointed chain behavior as known from single molecule stretching experiments and simulations, though with interesting differences. Long-range poly-alanine interactions result in a pronounced globule-stretch transition at intermediate flow velocities, and flow promotes  $\beta$ -sheet and poly-proline II conformations, both known to be present in silk fibers, in particular within the poly-alanine repeats. Our results give molecular insight into the drag force and stretching dynamics of disordered and unfolded proteins in flow, with implications for the flow-induced assembly of spider silk and related protein-based materials.

## 4.2 Methods

### All-atom MD simulations in flow

We used the Gromacs 2018 version [63]. The sequence of the peptide originates from the repetitive part of the *Euprosthenops Australis* spidroin at the MA spinning gland and is:

## 4. DYNAMICS OF A SINGLE SILK PEPTIDE UNDER UNIFORM FLOW

---

[A]<sub>13</sub>GQGGQGGYGGLGQGGYGQGAGSS[A]<sub>14</sub>GRGQGGYGQGSNGN[A]<sub>12</sub>. It is organized in alternating poly-Ala repeats and glycine-rich disordered fragments. We used the Amber force field Amberff99sb-star-ILDN in combination with the Tip4pD water model, which previously proved appropriate to model intrinsically disordered proteins such as the one investigated here [70, 71, 72]. Periodic boundary conditions were used in all three dimensions for all simulations. The concatenated simulation time including system preparation and production runs with flow is 18.5  $\mu$ s. Before introducing uniform flow, we prepared and equilibrated the system as follows: An energy minimization with 100000 steps starting from a randomly chosen conformation of the protein was followed by 5 ns of MD simulations in the NVT ensemble and 10 ns of MD in the NpT ensemble with the protein with position restraint. Temperature coupling was done by a V-rescale thermostat during equilibration and production runs [37]. The Parrinello-Rahman barostat [39] was used for pressure coupling in the NpT runs. LINCS constraints were used for all bonds to allow a 2 fs integration step. After preparation, we ran 300 ns in the NpT ensemble to equilibrate the protein without restraining the protein. After protein equilibration, we took 10 different conformations and prepared larger simulation systems along the flow direction. We simulated 10 flows with mean velocities from 0.0 m/s up to 0.48 m/s, each for 600 ns in an NVT ensemble. During the simulation we pulled waters inside a slice of 4 nm size with a constant force, using a modified Gromacs version to introduce uniform flow [65]. The V-rescale thermostat was used, which relaxed the system to achieve a constant velocity flow given by the externally applied velocity. The peptide was placed 4 nm away from the pulling region to avoid undesired stretching not associated with the flow, and artifacts due to the pressure drop inside the slice and at its interface. For the production run, the first alpha carbon of the peptide was position restraint to prevent translation. A timestep of 2 fs and LINCS constraint on h-bonds were used. We simulated for three different replicates for every flow velocity, each starting from a different velocity seed.

### CG MD simulations in flow

We simulated the same system described in the previous section also at CG level, where the amount of Ala residues in every region is the same as in all-atom simulations.

The simulation parameters are expressed in basic units  $k_B T = 1$ ,  $a = 1$  and  $m = 1$  (see sections 2.1.1 and 2.1.2 of the Chapter 2). The box volume is  $V = 100.0 \times 50.0 \times 50.0 a^3$ ,

the total number of solvent particles is  $N_s = 2.5$  million with number of solvent particles per cell  $N_c = 10$ . The mass of every monomer is  $M = N_c [m]$ , the integration time step  $\Delta t$  is  $0.0625\sqrt{ma^2/k_B T}$  and the collision time step is  $\Delta t_c = 0.1\sqrt{ma^2/k_B T}$ .

The simulations consisted of an equilibration of the system during  $12000\sqrt{ma^2/k_B T}$  starting from a fully extended chain, from which its final conformation was taken to simulate under 10 different flows. The force applied to the solvent particles to simulate uniform flow at a range of velocities was in the interval of  $(5 \cdot 10^{-07} - 2.7 \cdot 10^{-4})[(k_B T/a)]$  during  $12000\sqrt{ma^2/k_B T}$ . The simulations were carried out in an NVT ensemble using the cell-level velocity rescaling temperature coupling. We carried out three replicas with different velocity seeds for each flow velocity.

## Persistence and Kuhn lengths

To obtain the Kuhn length of the polymer at AA and CG scale, we fitted an exponential function to the autocorrelation function  $C(s)$  of the silk peptide backbone segments  $s$ , using:

$$C(s) = \exp(-2s/l_k). \quad (4.1)$$

The Kuhn length for AA and CG simulations was 0.2 nm and  $1.6 a$  respectively. With  $l_k = l_p/2$  we obtained the Kuhn length for both models.

## Drag Force Calculation

We calculated the drag force acting on the protein in AA MD simulations from interatomic forces  $F_{ij}$  between protein atom  $i$  and protein atom  $j$ , summed the  $z$ -component of all these forces up (with  $z$  being the flow direction), and averaged over time. We used the same definition of the drag force for the CG simulations, just that in this case, the forces only included the momenta transferred in the collision steps as only local protein-solvent forces are considered in MPCD.

## Secondary structure analysis

For the secondary structure analysis (Figure 5), we defined residues to be in a  $\beta$ -sheet conformation when their dihedrals fall into the intervals of  $90 \leq \psi \leq 180$  and  $-180 \leq \phi \leq -90$  (circles), and in a PPII conformation if  $128 \leq \psi \leq 180$  and  $-90 \leq \phi \leq -58$  (squares).

## 4. DYNAMICS OF A SINGLE SILK PEPTIDE UNDER UNIFORM FLOW

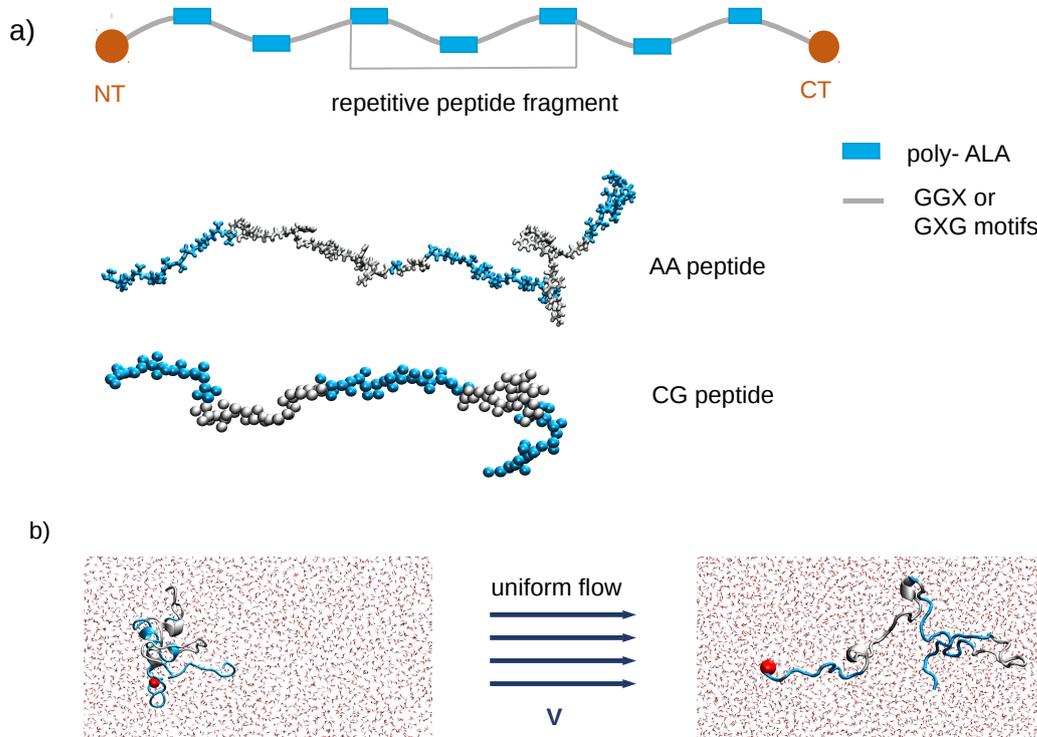
---

In order to observe the relevance of alanines in the formation of  $\beta$ -sheets we show the percentage of Ala and non-Ala residues like:  $\%Ala_{\beta/PPII} = N_{Ala_{\beta/PPII}} * 100 / N_{Ala}$  (blue symbols) and  $\%nonAla_{\beta/PPII} = N_{nonAla_{\beta/PPII}} * 100 / N_{nonAla}$  (gray symbols).  $N_{Ala_{\beta/PPII}}$  and  $N_{ALA}$  correspond to the number of Ala residues in  $\beta$ -sheet or PPII helix and the total number of alanines in the peptide sequence, while  $N_{nonAla_{\beta/PPII}}$  and  $N_{nonAla}$  are the number of residues of the amorphous region in  $\beta$ -sheet or PPII and the total number of residues different from alanines.

### 4.3 Results and Discussion

We chose a representative fragment of silk spidroin, which includes three polyalanine repeats separated by two glycine rich regions, and with a total length of 76 aminoacids is still amenable to atomistic MD simulations. Representative conformations are shown in Figure 4.1a. The spidroin peptide was tethered at its N-terminus by a position restraint and subjected to uniform flow at different velocities (Figure 4.1b). To simulate uniform flow we used similar approaches in all-atom (AA) and CG simulations: we applied a constant force to the solvent particles along the flow direction in an NVT ensemble, with the temperature coupling relaxing the system until a constant terminal velocity is achieved. In all-atom MD simulations, explicit water molecules within one slab of the simulation system were subjected to the constant force. For the CG simulations, we combined MD and MPCD methods, where the peptide is embedded in a mesoscopic solvent, which is represented by point particles that interact through local collisions between themselves and the peptide. See Methods for details.

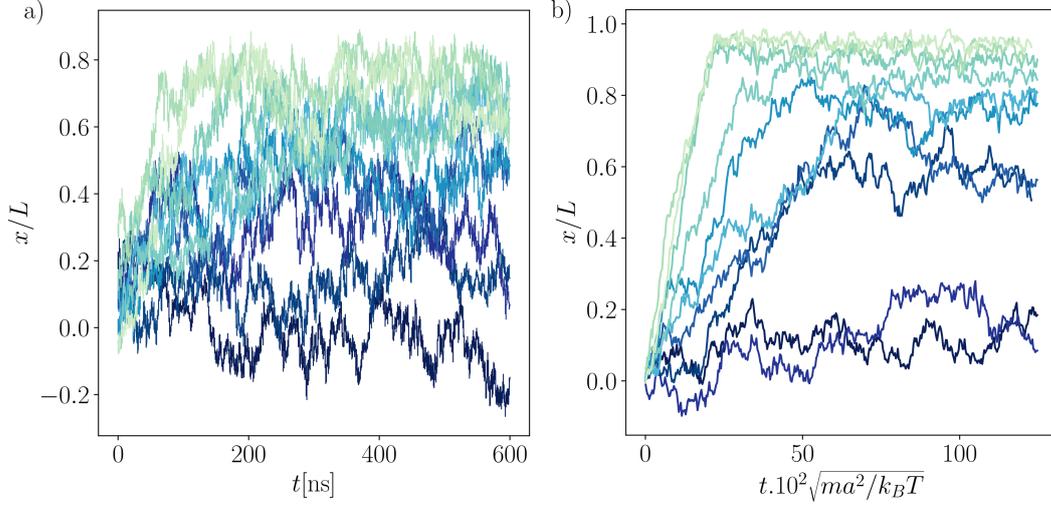
The moment the tethered silk peptide is subjected to flow, it shows increasing extensions (defined as the average signed end-to-end distance along the flow direction of the terminal C- $\alpha$  atoms or beads, respectively) over time along the flow direction (Figures 4.2). Figure 4.2a corresponds to all-atom MD simulations where the peptide extension reaches a plateau after  $\sim 200$  ns, whereby length fluctuations are more pronounced at low and intermediate flow velocities. CG simulations show a highly similar behavior (Figure 4.2b), with extensions reaching a plateau after  $5000 \sqrt{ma^2/k_B T}$ . At the CG level, length fluctuations are more pronounced at low flow velocities, but overall less than in all-atom simulations. This behavior can be attributed to the lack of specificity in the modeling of the CG aminoacids. For example, the CG beads do not take into account the increased flexibility of glycine



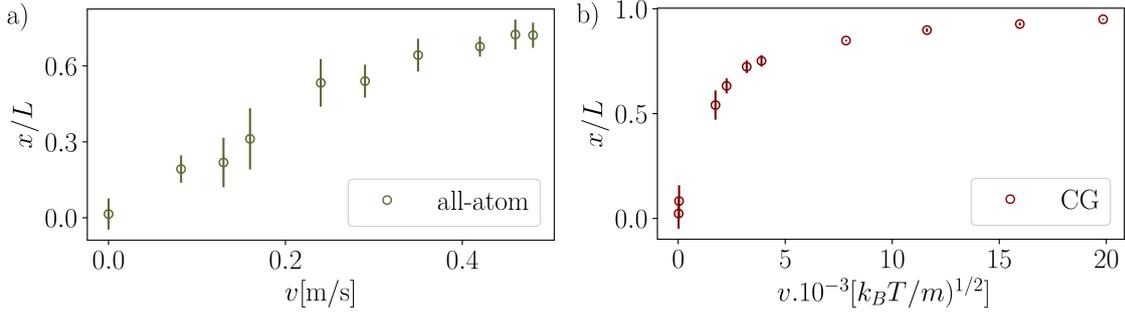
**Figure 4.1:** Modelling of a silk peptide under uniform flow. a) Top: scheme of the *MA Eupros-thenops Australis dragline spidroin* sequence. Its molecular structure comprises three regions: pH dependent N- and C-terminal domains (NT and CT, red), as well as a long repetitive region with alternating poly-alanine (poly-Ala, blue) repeats and glycine-rich motifs (GGX or GXG, with X being another aminoacid, grey). Bottom: Representative structures of the atomistic (AA) and coarse-grained (CG) models of a fragment of the repetitive region, containing three poly-Ala and two amorphous regions, resulting in  $\sim 80$  aminoacids in total. Spheres represent atoms (for AA) or backbone and sidechain beads for each aminoacid (CG) Color coding as in the scheme on top. b) All-atom MD simulation system. The spidroin fragment is tethered at the N-terminus (red sphere), and a uniform flow is produced to stretch the peptide. Color coding as in (a).

residues due to the lack off a sidechain and reversibly forming non-specific attractive interactions within the glycine-rich regions of the peptide. As expected, for both the atomistic and CG scales, we obtain on average larger extensions at larger flow velocities (Figure 4.3). The flow velocities used in the CG simulations cover a wider range and reach the regime of

#### 4. DYNAMICS OF A SINGLE SILK PEPTIDE UNDER UNIFORM FLOW



**Figure 4.2:** Fractional extension along the flow direction over time for a) all-atom and a) CG simulations.



**Figure 4.3:** Fractional extension as a function of flow velocity for a) all-atom MD simulations, b) CG simulations.

largely extended peptide conformations ( $x/L \sim 0.9$ , with  $L$  being the contour length) already from medium flow velocities onwards. The AA simulations cover flow velocities up to 0.5 m/s and reach average extensions up to  $x/L \sim 0.7$ . Higher velocities and extensions could not be reached, as larger forces applied to the water molecules lead to voids [65], and wider pulling slabs and simulations systems would have been needed to overcome this limit. Nevertheless, our results show a good agreement of the CG with the AA simulations, with a velocity of  $5 \cdot 10^{-3} [(k_B T/m)^{1/2}]$  at the CG scale being comparable to  $5 \cdot 10^{-1} m/s$  at the atomistic scale in terms of the resulting average peptide extension.

We next asked how the flow velocities are related to the drag force acting on the peptide.

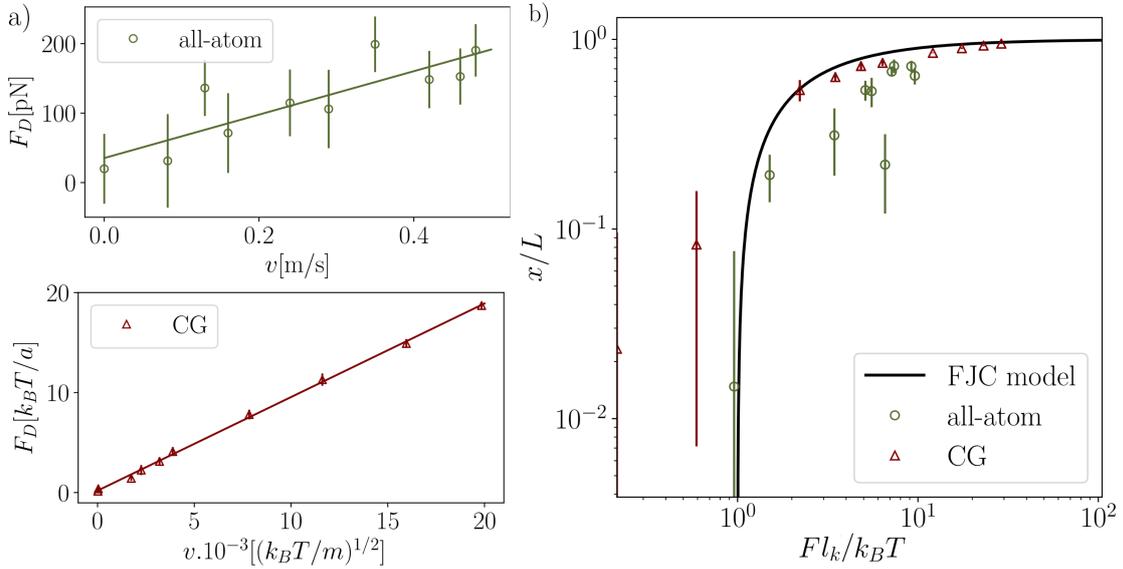
As the drag force changes dynamically with the peptide conformation, a simple analytical model of the drag as a function of velocity might not be appropriate. We instead calculated the drag force directly from the forces between the solvent particles and the protein particles (atoms or beads, respectively). In the all-atom case, we computed the vector-sum of the non-bonded forces between all atoms of the water molecules and of the peptide over time. In the CG case, we obtained the drag force from averaging the collisions of the solvent particles with the beads, as the solvent-protein interactions solely occur through local momentum exchange. In both AA and CG simulations, the force components orthogonal to the flow direction average to zero over time, while the component along the flow yields a non-zero average force, the drag force. We averaged the drag force over the last half of each trajectory and over all three replicates. Interestingly, independent on the resolution of the simulations, i.e. for both AA and CG, we obtain an overall linear increase in drag force with flow velocity (Figure 4.4a). Thus, the drag force effectively follows Stokes' law, with an effective hydrodynamic radius of 2.52 nm (obtained from the linear fit to the AA data of Figure 4.1a). We conclude that our spidroin peptide does not exhibit a pronounced globule stretch transition at a critical flow speed, which would lead to a deviation from a linear drag-flow relationship, but instead steadily extends with increasing flow velocity. This is in sharp contrast to the behavior seen previously for proteins in shear flow, which showed a rather abrupt globule-stretch transition (in case of VWF) [49] or a one-step unfolding (in case of ubiquitin) [92, 93], respectively. However, a gradual stretching through multiple unfolding intermediates were observed for ubiquitin in the case of elongational flow, [92, 93] directly in line with the gradual extension of the spidroin with increasing flow velocities.

To quantitatively compare the AA and CG scales, Figure 4.4b shows for both sets of simulations the dimensionless drag force in the peptide as a function of the normalized extension along the flow.  $L$  and  $l_k$  correspond to the contour length and Kuhn length, which are 24.3 nm and 0.24 nm for all-atom simulations as well as  $80 a$  and  $1.5 a$  for CG simulations. We recover a non-linear force-extension curve typical for polymers when stretched by a pulling force. The all-atom model yields overall lower extensions for a given drag force, but given that they are fully independently parametrized, the agreement is satisfying. That the CG model extends more readily is likely due to our approximation that attractive interactions are only present within alanine residues, thereby neglecting unspecific favorable interactions involving non-alanine residues of the disordered region. The force-extension behavior is reminiscent of the behavior predicted by the freely jointed chain (FJC) [96] or worm-like chain (WLC) model [84] of polymers. Indeed, we find a good agreement of our

#### 4. DYNAMICS OF A SINGLE SILK PEPTIDE UNDER UNIFORM FLOW

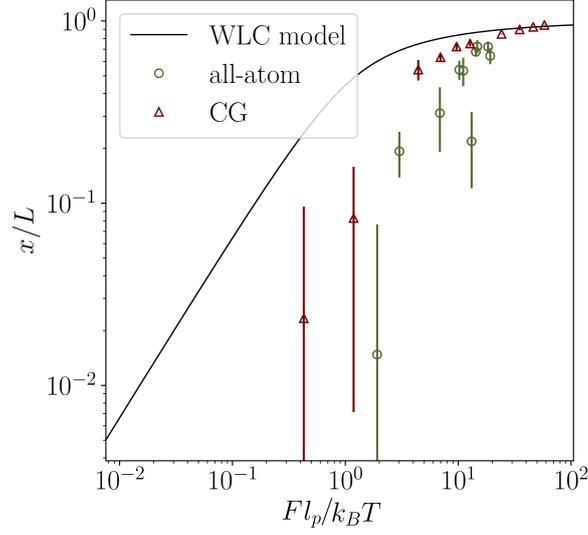
data with the FJC model (Figure 4.4b). Our particle-based simulations agree slightly less well with the WLC model (4.5), which we explain by the nature of the protein backbone the (albeit only partially) rotatable bonds of which are better depicted by a FJC model than by the rod-like WLC with uniform bending stiffness.

Stretching the silk peptide to  $\sim 80\%$  of its contour length ( $x/L = 0.8$ ) requires a drag force from uniform flow in the range of 100-200 pN, according to both AA and CG simulations. Interestingly, this range covers the rupture forces of  $\sim 176 \pm 73$  pN required to stretch single dragline silk molecules by an atomic force microscope [86].



**Figure 4.4:** Extension of the spidroin fragment increases with flow velocity and drag force. a) Averaged peptide drag force over mean flow velocity for all-atom and CG simulations. For all-atom simulations, the drag force in the peptide originates from the non-bonded interactions between the water and protein atoms. In the case of CG simulations, the drag force was computed from the collisions between the solvent particles and the AA of the peptide. The green and maroon solid curves are linear fits. Error bars are standard errors of the mean for nine data points, obtained from time-averages over three 100 ns windows of the second half of each of the three trajectories per flow speed. b) Dimensionless peptide drag force (from (a)) as a function of the normalized peptide extension.  $L$  is the contour length,  $l_k$  is the Kuhn length. Circles and triangles correspond to all-atom and CG simulations, respectively. Black solid curve: normalized theoretical prediction from the FJC model,  $f(x) = 1/(1-x)$ .

In Figures 4.6a and 4.6b, we show the  $C_\alpha$  carbon atoms distribution along the flow direction for all-atom and CG simulations respectively. The probability density of finding

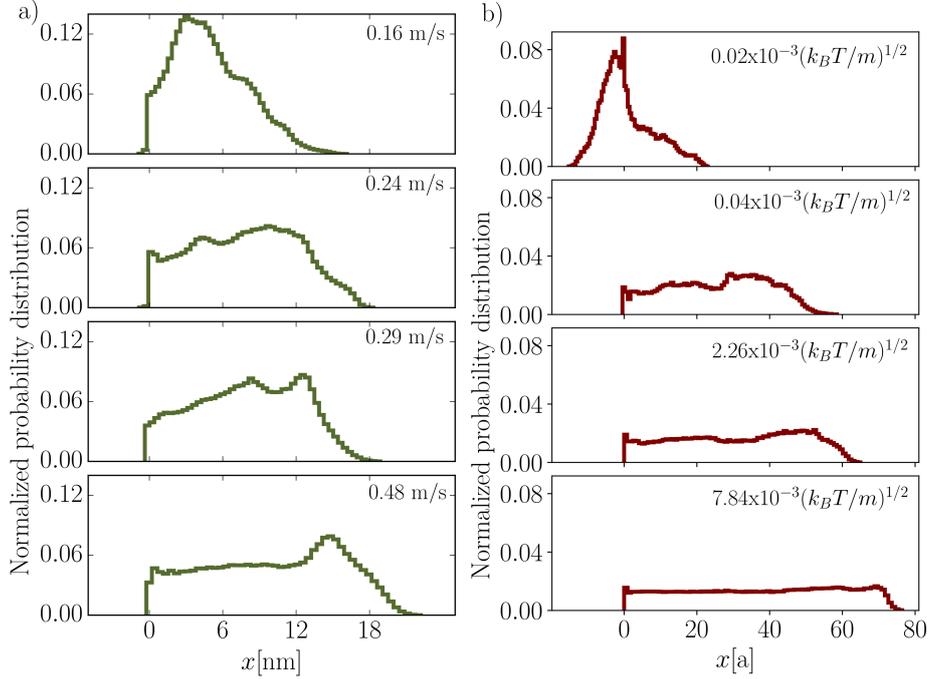


**Figure 4.5:** Dimensionless peptide drag force as a function of the normalized peptide extension along the flow component. Circles and triangles correspond to all-atom and CG simulations respectively. The solid line the theoretical curve of a worm-like chain model.

an atom in a certain position, is measured from the tether point. From intermediate to high velocities the  $C_\alpha$  atoms are more concentrated towards the free end, as the tension is low. Moreover, the chain is more prone to coil in its final part even for high mean flow velocities. This is due to the Ala residues that reside in the last fragment of the peptide who avoid to expose to water. With regard to the CG simulations, the distribution at low and intermediate velocities depends highly on the formation of Ala bonds, i.e. at  $0.02 \times 10^{-3} (k_B T/m)^{1/2}$  and  $2.26 \times 10^{-3} (k_B T/m)^{1/2}$ . Taking this into account, the CG peptide acquires a more stable conformation than the all-atom peptide under certain flow conditions.

The transition of the silk peptide from a coiled to an extended conformation can be influenced by alanines. To analyze the relevance of Ala residues in the silk protein conformational ensembles in flow, we quantified the number of Ala contacts along flow velocities in all-atom (Figure 4.7a) and CG simulations (Figure 4.7b). In the all-atom case, we observe a drop of Ala-Ala contacts between 0.16 and 0.24 m/s of flow rate, This tendency can also be observed if we consider the total number of contacts or only the amorphous region contacts (see Figure 4.8a). Also the fractional extension of the all-atom peptide as a function of the flow rate confirms this behavior (Figure 4.3a): initially the peptide is extended up to about 30% of the contour length from 0.08 m/s up to 0.16 m/s. After this threshold, the peptide is stretched 20% further in the mean flow velocity interval of (0.24 - 0.29) m/s up to around

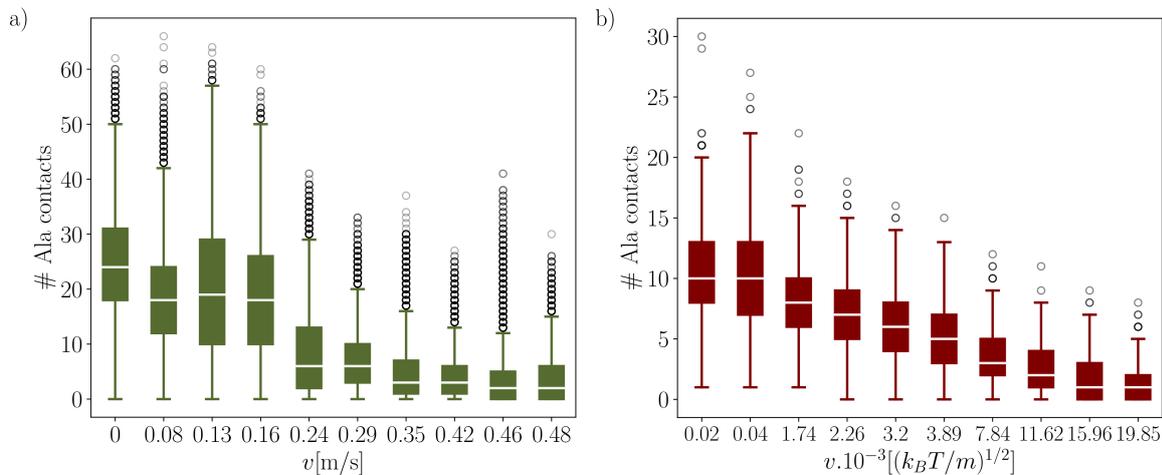
#### 4. DYNAMICS OF A SINGLE SILK PEPTIDE UNDER UNIFORM FLOW



**Figure 4.6:** Normalized probability distribution of  $C_\alpha$  atoms along  $x$ , the distance downstream from the tether point in: a) all-atom MD simulations and b) CG simulations. At lower velocities, the  $C_\alpha$  atoms are concentrated near the tethered residue, i.e. the peptide is largely collapsed. With increasing velocity, the peak in the distribution shifts towards higher  $x$ , i.e. the free end, where the tension due to drag is lowest.

80% of its contour length at 0.48 m/s.

In the case of the CG peptide we observe a less abrupt transition to a fully extended state without Ala-Ala contacts. The contacts steadily decrease from 10 contacts to nearly 0 contacts between  $0.02$  and  $20 \times 10^{-3} (k_B T / m)^{1/2}$  (Figure 4.7b). The number of total and amorphous contacts over mean flow rate in Figure 4.8b shows the same tendency. With regard to the CG fractional extension (Figure 4.3b) over the mean velocity, Ala contacts are maximal only at 0-10% fractional extensions, and are largely lost only beyond 90% extensions. Thus, the CG model does not reproduce the comparably sudden transition observed at all-atom. In addition, the CG model shows less contacts initially, i.e. at low flow velocities ( $\sim 12$  instead of  $\sim 24$  for AA). We conclude that the CG model reproduces the overall loss of contacts in flow, though it cannot resolve the collapse-stretch transition of AA simulations.

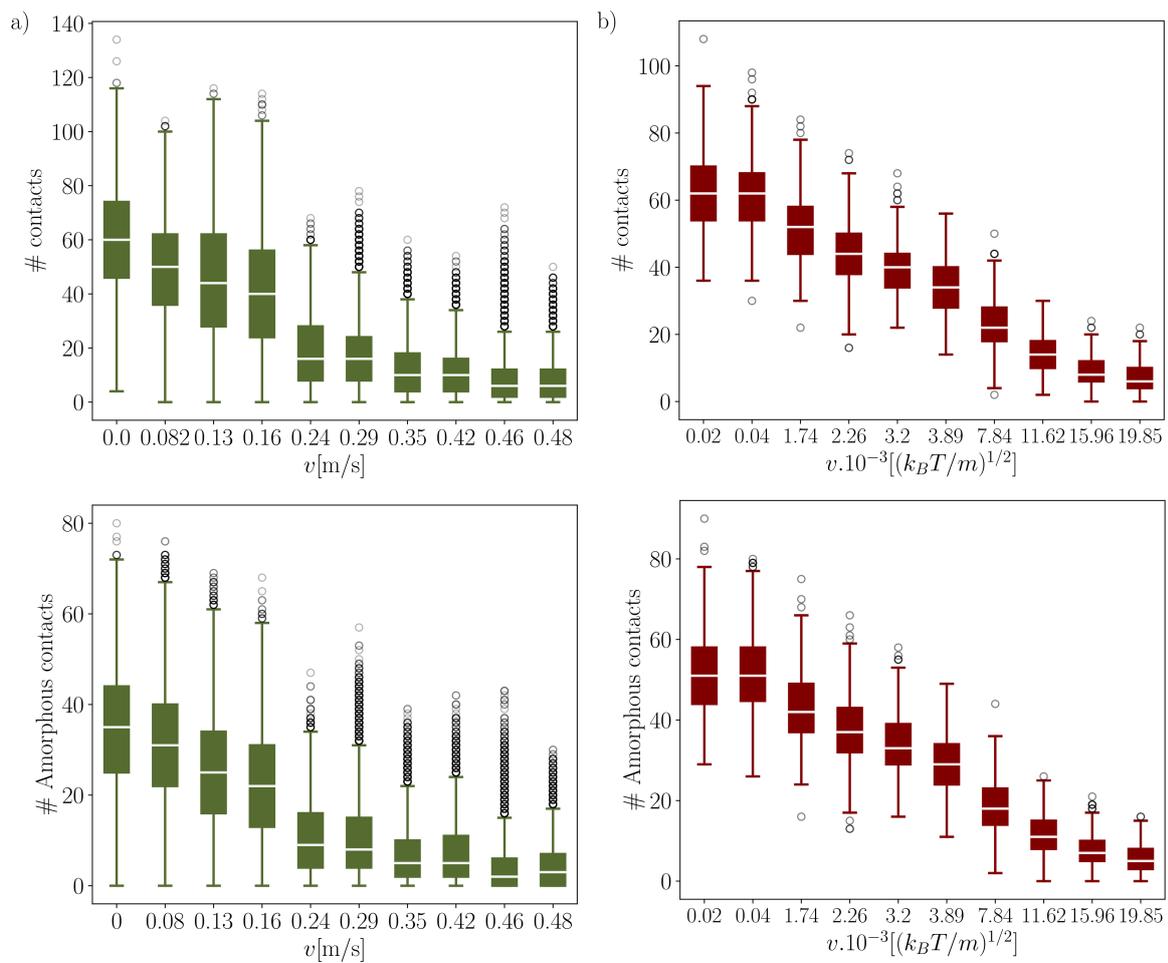


**Figure 4.7:** Number of contacts of Ala residues as a function of time in: a) All-atom MD simulations and b) CG simulations. The cutoff radius within we examine a contact formation is 0.5 nm

The formation of  $\beta$ -sheets is a key factor in the self-assembly of spider silk and requires flow [25]. However, how flow induced  $\beta$ -sheet formation is poorly understood. In recent studies, a significant population, namely  $\sim 24\%$ , of poly-proline II (PPII) helix in the repetitive region of MA spidroin dopes, i.e. prior to flow-induced elongation during spinning, from the *Euprosthenops Australis* spider has been observed [97]. It was suggested that PPII helices can form a rigid structure that can be quickly transformed into  $\beta$ -sheets. We here asked to what extent our spidroin fragment samples  $\beta$ -sheet and PPII conformations and how these backbone propensities are influenced by flow. We defined for each aminoacid the respective backbone configurations based on its two backbone dihedral angles (see Methods for details). We averaged over all alanines and all amorphous-phase residues (non-alanines) and normalized by the total number of alanines and non-alanines, respectively. We note that we analyzed the secondary structure propensities only for the AA simulations, as the CG lacks the required resolution.

Figure 4.9 shows the percentage of aminoacids with dihedral angles either in  $\beta$ -sheet or PPII conformation as a function of mean flow velocity. The percentage of alanines in both,  $\beta$ -sheet and PPII conformations is increasing with the mean flow velocity for both the poly-Ala and amorphous regions. Both,  $\beta$ -sheet and PPII conformations populate the upper left corner of the Ramachandran plot and are close to the most extended backbone configuration with  $\psi = 180$  and  $\phi = -180$ . For this reason, flow is expected to drive the aminoacids into

#### 4. DYNAMICS OF A SINGLE SILK PEPTIDE UNDER UNIFORM FLOW



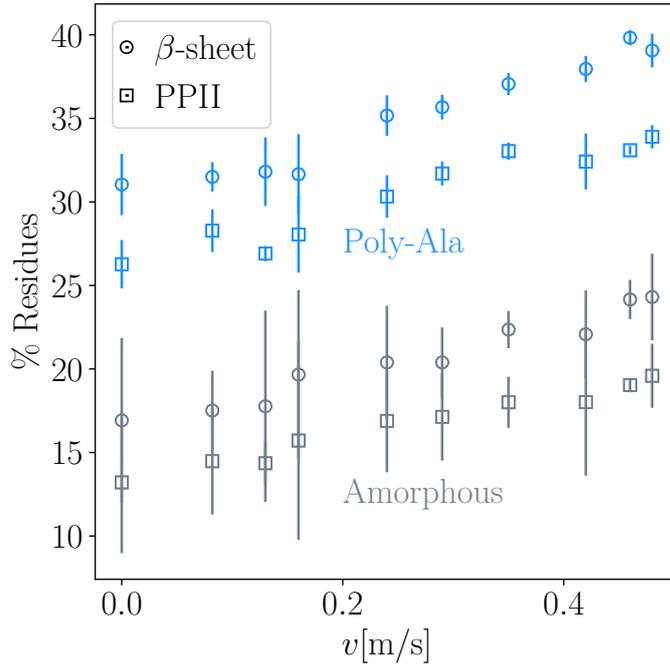
**Figure 4.8:** Total number of contacts and amorphous region contacts for a) all-atom MD and b) CG simulations as a function of flow velocity.

this region of the dihedral space, as previously also observed for disordered proteins under a stretching force [98].

The content of the backbone torsion angles in  $\beta$ -sheet and PPII in the poly-Ala regions remains around 10% higher than in the amorphous regions, largely independent of the flow velocity. This higher ratios are in line with the view that the poly-alanine repeats preferentially drive  $\beta$ -sheet formation, i.e. crystallization, during fiber formation.

In both the alanine and amorphous phase,  $\beta$ -sheet conformations are slightly preferred over the PPII conformation, and they together represent the majority of secondary structure conformations. More specifically, in the poly-alanine regions, at the absence of flow,  $\beta$ -sheet and PPII dihedral conformations are on average sampled by more than 50 % of the residues and reach  $\sim 70$  % at the largest flow rates we have used. Taking into account that the transition from PPII-helix to  $\beta$ -sheet is very likely to happen due to their close proximity in terms of dihedrals, we can conclude that poly-alanine regions in our spidroin are primed for  $\beta$ -sheet formation by uniform flow. The predominantly proportion of alanines of  $\sim 70$  % in  $\beta$ -sheet or close to  $\beta$ -sheet (that is, PPII) conformations is already very close to the 90 % percent  $\beta$ -sheet content by alanines predicted in the model of Alexandra H. Simmons et al. [19], which is based in NMR experimental studies. The occurrence of torsional angles in PPII-helix conformation might result in a prefibrillar form of the repetitive region of the spidroins before forming stronger  $\beta$ -sheet interactions, as proposed by Nur Alia Oktaviani et al. [97].

## 4. DYNAMICS OF A SINGLE SILK PEPTIDE UNDER UNIFORM FLOW



**Figure 4.9:** Flow populates  $\beta$ -sheet and PPII conformations. Percentage of dihedral angles of aminoacids in  $\beta$ -sheet (circles) and PPII conformation (squares) as a function of mean flow velocity, as observed in AA simulations. The data is separated into alanine (blue symbols) and amorphous regions (gray symbols), and normalized by the total number of residues in these two regions. Error bars show the standard error of the mean. See Methods for the definition of the dihedral intervals for both secondary structures.

### 4.4 Conclusion

We here present the globule-stretch transition of a single repetitive spider silk protein under uniform flow, using AA MD and coarse-grained MPCD simulations. We suggest a coarse-grained model with two beads per aminoacid and attractive interactions only between poly-alanine regions, the results of which compare very well to the coarse-grain model.

We obtained hydrodynamic drag forces directly from water-protein forces as present during the simulations, and the resulting drag-force-entention behavior is directly in line with the known force-extension behavior of silk or other disordered proteins from force spectroscopy experiments [86]. We therefore suggest our definition of drag force as a very useful quantity that is straightforward to obtain, independent from the level of resolution, and for both standard MD as well as MPCD methods, and that does not require any addi-

tional assumptions.

We observe a strong tendency of alanine residues to sample dihedral conformations present in  $\beta$ -sheets or PPII helices, which is further increased by a uniform flow to values close to those found for poly-alanines within the silk fiber. This supports the important role of PPII helices in pre-fibrillation and fibrillation steps of silk and the role of flow in stretching and aligning the poly-alanines of spidroins for crystallization into stacks of  $\beta$ -sheets. Our computational approach across scales can be directly used to study simulations of more than one silk peptide, which help towards elucidating the molecular dynamics and hydrodynamic interactions critical for silk fiber assembly under flow. This is the focus of the following chapter.



# Nanoscale Self-Assembly of Dragline Spidroins

## 5.1 Introduction

Self-assembly processes are ubiquitous in proteins. They play a role at different length scales, from the range of nanometers (protein folding) to centimeters (biomaterials) [99, 100]. Focus has been widely set on fibrous protein assemblies (i.e. collagen, silk, hair, and bones) due to their diverse functionalities that can be used for a large variety of applications [14]. A deep understanding of fibrous proteins and how they are built is crucial i.e. for the novo-design of materials, biopharmaceutical processing, and the study of protein misfolding for the treatment of degenerative diseases [5, 6, 101, 102, 103].

Spider silk, one of the well-studied fibrous materials, has been investigated in depth. These studies form a basis for our understanding of the attractive properties of silk: outstanding mechanical properties, biocompatibility and biodegradability. The protein blocks of spider silk have highly repetitive domains of poly-alanine repeats and rich-glycine motifs (Section 1.1). A controlled spinning by spiders with their highly repetitive sequence leads to the formation of  $\beta$ -sheet crystals. The crystals in turn are embedded in a flexible amorphous matrix that is composed of intrinsically disordered fragments. The poly-alanine repeats are mostly involved in the formation of  $\beta$ -sheet crystals.

## 5. NANOSCALE SELF-ASSEMBLY OF DRAGLINE SPIDROINS

---

To date, we lack a complete understanding of the molecular process of silk formation. It occurs mainly inside the s-gland of the spider, where the liquid silk solution is transformed into a fiber through ion exchange, shear forces, and a decreasing pH along the s-duct [1, 2]. The pH sensitive and small protein terminal domains are involved in the solubility control and pre-alignment (or initial interconnections) of the spidroins, while an elongational flow is in charged of the silk protein stretching to promote the fibrillation via  $\beta$ -sheet crystals formation (Figure 1.2 and Figure 1.3 and Figure 1.4) [14, 31, 85]. Unraveling the effects of the elongational flow in the spider silk formation process would bring a broader picture of the assembly pathway of the spidroins. Principally, helping to depict whether the correct spinnability of silk requires intermediate formation of micelles or not [16, 17]. Additionally it helps to describe liquid crystalline flows inside the s-duct [15, 30, 32].

The lack of a more detailed understanding of silk self-assembly is, among others, due to the challenge to experimentally work with the highly aggregation-prone silk proteins. In principle, the proteins should be soluble to prevent fatal aggregation. It is difficult to experimentally manipulate the silk dopes due to their high concentrations  $\sim 50\%$  wt/v and their sensitivity to changes in the microenvironment (pH, ion conditions and shear forces). Moreover, the lack of structuring of the silk monomers in most of its domains limits the use of structural studies.

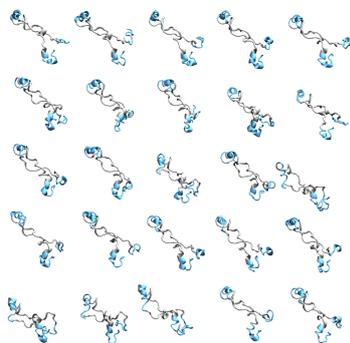
In the case of the assembly of spider silk proteins under flow, MD simulations offer the possibility of analyzing in detail, how flow drives protein interactions and  $\beta$ -sheet formation. We used multi-scale MD simulations to describe the self-assembly process of spider silk under different conditions of concentration and flow velocity. The simulation setup is explained in the next Section 5.2. We only simulated a small fragment of the repetitive part of the silk proteins in order to limit the computational expense. This fragment comprises three poly-Ala repeats and two intrinsically disorder regions. The sequence of the spidroin is described in the previous chapter of this thesis (Chapter 4). An initial pre-alignment of the peptides was performed, mimicking the multimerization at the terminal domains. The pre-alignment consists of tethering the peptides along the flow in the same position along the flow direction, while they can move transversely. Imposing this pre-alignment will enhance an organized packing of the chains (via formation of contacts mostly within the poly-Ala regions). In this way, we monitored assembly in an idealized setting with all chains in phase and subjected to uniform flow. Our study can give insight into flow- induced protein assembly in general, and provide a first glimpse into silk spinning.

We note that the results shown in this chapter are a starting point towards systematic silk assembly simulations, and we describe future directions in the Results and Discussion section 5.3.

## 5.2 Methods

### 5.2.1 Simulation Setup MD Simulations

To monitor the self-assembly of spider silk proteins, we performed simulations of 25 peptides under uniform flow in AA and CG simulations. Every spidroin corresponds to a random conformation taken from the single peptide simulations in a given flow regime (Chapter 4). Like in the single silk spidroins simulations, the silk peptides are tethered as well along the flow direction. Initially, the peptides are placed equi-distant from each other in a matrix of 5x5, where the positions  $(y, z)$  of the tether atoms correspond to the nodes of the matrix. Figure 5.1 shows the initial configuration of the peptides for one replicate of the AA simulations. The peptides are allowed to move transversely (in the  $yz$  plane) and are tethered at the same position in the flow component  $x$ .



**Figure 5.1:** Initial peptide setup MD simulations under flow. Water is not shown for clarity

### All-atom MD simulations

For the system preparation and production run with uniform flow, we used the same protocol and the same parameters as in the single peptide case (Chapter 4, Section 4.2). In total, we performed 3 simulations with 2 replicates each. A description of every simulation can be found in Table 5.2. The initial peptide separation is 6 nm, which corresponds to a concentration of  $\sim 8\%$  wt/v.

## 5. NANOSCALE SELF-ASSEMBLY OF DRAGLINE SPIDROINS

---

Mean flow velocity [cm/s]	Initial mean fractional extension of peptides [L]	Simulation time [ns]	Number of replicants
1	0.0	600	2
3.3	0.2	600	2
3.5	0.3	600	2

**Figure 5.2:** AA simulations description.

### CG MD Simulations

In the CG assembly simulations, we also used the same protocol and simulation parameters as in the single peptide case (Chapter 4, Section 4.2). The initial peptide separation is 8 nm and the system concentration is  $\sim 4\%$  wt/v for the CG system. In total, we performed 6 simulations with 3 replicates each. A description of each simulation is contained in Figure 5.3.

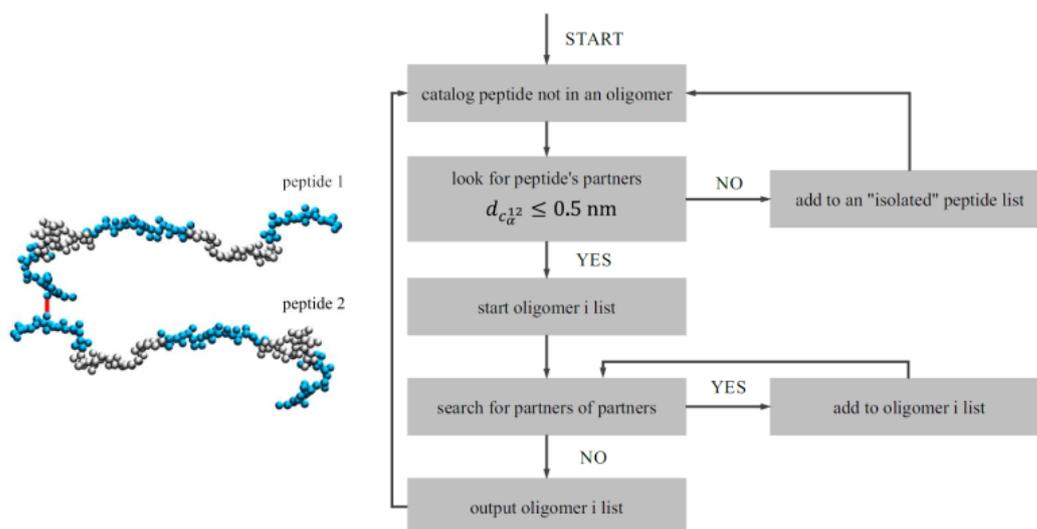
Mean flow velocity [ $10^{-3}(k_B T/m)^{1/2}$ ]	Initial mean fractional extension of peptides [L]	Simulation time [ $10^2 \sqrt{m a^2 / k_B T}$ ]	Number of replicants
0.3	0.0	517	3
0.8	0.5	517	3
1.5	0.6	517	3
2.5	0.7	517	3
8.3	0.8	370	3
10.2	0.9	230	3

**Figure 5.3:** CG simulations description.

### Friend of Friends Algorithm

To monitor whether a silk peptide forms part of an oligomer of certain size or not, we adapted a group selection algorithm called Friend of Friends (FoF) to our case [104]. The flow chart of the algorithm steps is shown in Figure 5.4. First, we chose a peptide, which

has not been previously assigned to an oligomer group  $i$ . Next, we search for partners of this peptide based on the interchain  $C_\alpha$  carbons contact formation, where the condition to have a contact of two atoms between peptide 1 and peptide 2 is defined as  $d_{C_\alpha}^{12} \leq 0.5$  (see the spidroin intercontact sketch at Figure 5.4). In case, the peptide does not have a partner, we assign it to an "isolated" peptide list. When partners are found, all of them are added to the list of  $i$  oligomers. We search repeatedly until no further members can be found. This process runs over all frames of the simulation. Therefore, we can monitor not only the time evolution of individual chains that form oligomers, but also their contact formation with other chains.



**Figure 5.4:** Schematic of the Friends of Friends (FoF) algorithm and two spidroins forming a contact. This algorithm groups the peptides depending on the oligomer size, to which it belongs to. At the left side a sketch of two spidroins forming a contact is shown.

## 5.3 Results and Discussion

In order to analyze the self-assembly of the spidroins, we initially monitored when a certain peptide takes part of an oligomer. This was done through a clustering algorithm called FoF (Section 5.2). The criterion to form an oligomer of certain size is based on the formation of a single interchain  $C_\alpha - C_\alpha$  bond. In Figure 5.5, we visualize the oligomerization process for both AA and CG simulations. Data is shown for one replicate, where the initial mean

## 5. NANOSCALE SELF-ASSEMBLY OF DRAGLINE SPIDROINS

---

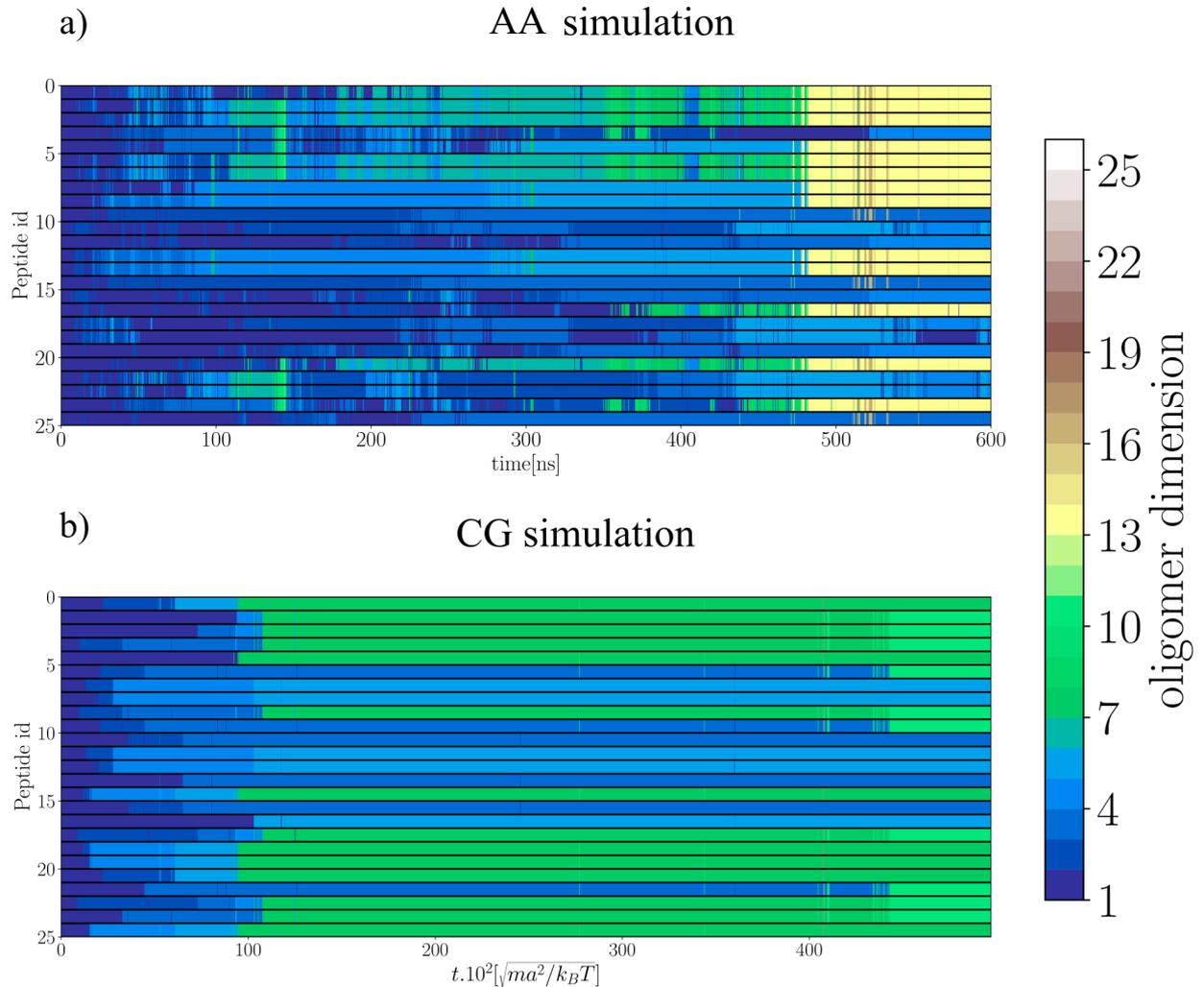
peptide fractional extension is  $0.0L$ . We observe larger fluctuations in the AA simulation, which is associated with the strong response of the intrinsically disordered regions of the spidroins with flow. Furthermore, dimers and trimers are not stable oligomers. They reversibly transform into smaller assemblies, monomers and dimers, or form oligomers of higher dimension. Oligomers larger than a size of 3 are stable and favor the formation of aggregates of a higher dimension. At the end of the simulation, a stable oligomer of size 13 formed from semi-stable pentamers and tetramers. In contrast to the AA simulations, the aggregation of the spidroins in CG simulation show less reversibility (Figure 5.5). The stability of the CG aggregates is due to the model we considered, where attractive interactions only occur between the Ala residues. As a consequence, non-native interactions cannot be formed. We observe a faster aggregation for oligomers of sizes between 2 to 5. In comparison, aggregates of sizes 5 to 7 are more stable and form larger assemblies at longer time scales. We associate this with the growing distances between oligomers during assembly progression. An alternative explanation considers a hydrodynamic effect that leads to shielding. Future studies are required to investigate this.

We quantified the aggregation of the spidroins over all AA simulations, in which the peptides average extension is around  $0.06L$  and the mean flow velocity is  $\sim 2.5$  cm/s (Figure 5.6a). The reasons for the overall low extensions in the simulations are: (i) The forces used to induce flow lead to lower velocities here, compared to previous simulations (Chapter 4), since the 25 peptides increase friction.

(ii) Oligomer formation further reduces the transition to extended states. Future simulations at lower concentration and higher forces are required to study assembly in AA.

We computed the average oligomer size as a function of time per replicate (Figure 5.6a) and the average over the whole replicates (Figure 5.6c). The oligomerization increases monotonically until reaching a mean oligomer size of 10 (Figure 5.6c). As shown in Figure 5.6b, we observe processes all the way up to 10mers or larger. Interestingly, the aggregation process is still reversible even for such large oligomer sizes. As a result it produces a mean oligomerization larger than 10 during the 600 ns.

We next computed the formation of contacts between the protein chains from the AA MD simulations, and averaged over all replicates (Figure 5.7a). As expected, the interchain contacts grow with time, reflecting the progress in assembly. From the mean contact formation we calculated the percentage of Ala and non-Ala residues between the spidroins. Ala residues are involved in the formation of around 57% of the interchain contacts (Figure 5.7b). This supports the relevance of alanine amino acids in the fibrillation process

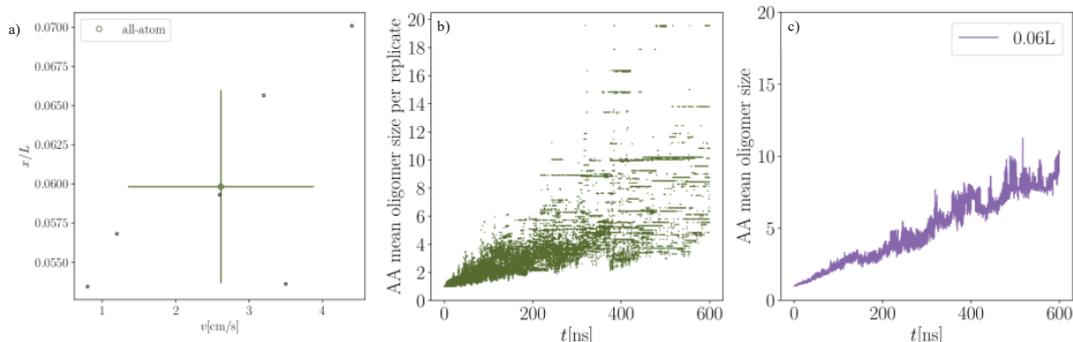


**Figure 5.5:** Monitoring in time of every peptide depending on the size of the oligomer they form. The figures are shown for one replicate in a) all-atom and b) CG MD simulations. We took simulations for initial mean peptide fractional extensions of  $0.0L$  for both systems.

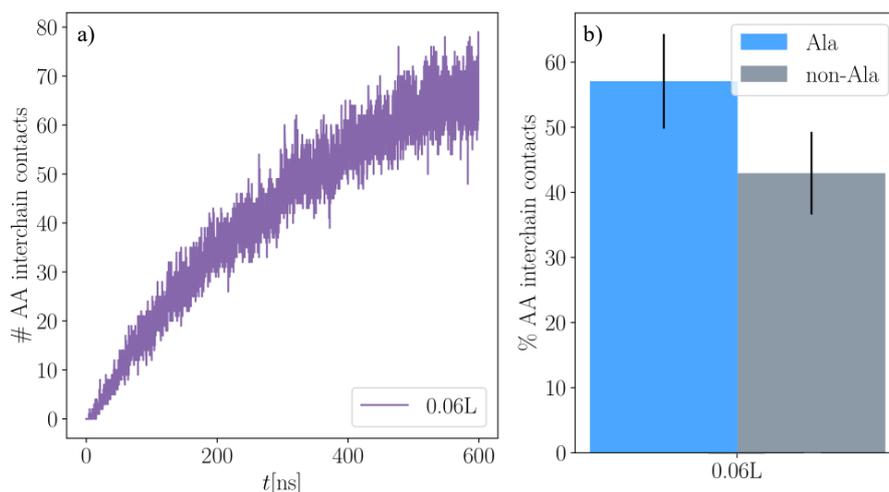
of spidroins. As shown previously,  $\beta$ -sheet crystals are formed mainly in the poly-Ala repeats [1, 19]. Our data suggests that these residues also provide the interactions dominating assembly.

Along these lines, we also calculated the mean number of amino acids involved in the formation of  $\beta$ -sheets (Figure 5.8). The number of  $\beta$ -sheets increases with time as well. In this case, it is related mostly to the hydrogen bond formation during the self-association

## 5. NANOSCALE SELF-ASSEMBLY OF DRAGLINE SPIDROINS

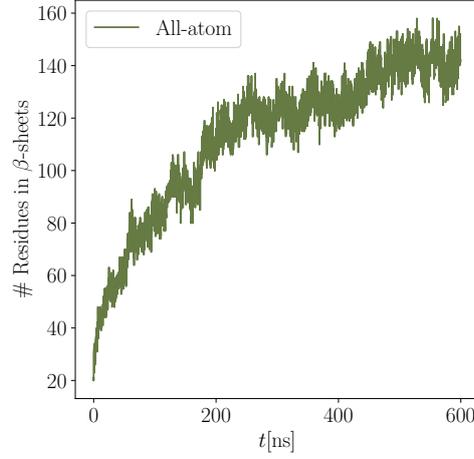


**Figure 5.6:** Oligomerization in all-atom simulations. a) Mean all-atom peptides fractional extension  $x/L$  as a function of the mean flow velocity. Black points correspond to individual replicates, and the green point corresponds to the average value over the replicates. b) All-atom mean oligomer size per replicate as a function of time. c) All-atom mean oligomer size as a function of time. The average is over all AA simulations.



**Figure 5.7:** a) All-atom mean total number of interchain contacts as a function of time, averaged over all replicates per flow regime. b) Percentage of interchain contacts for Ala and non-Ala residues. The average is computed in the last half part of the mean interchain contacts (Figure 5.7a).

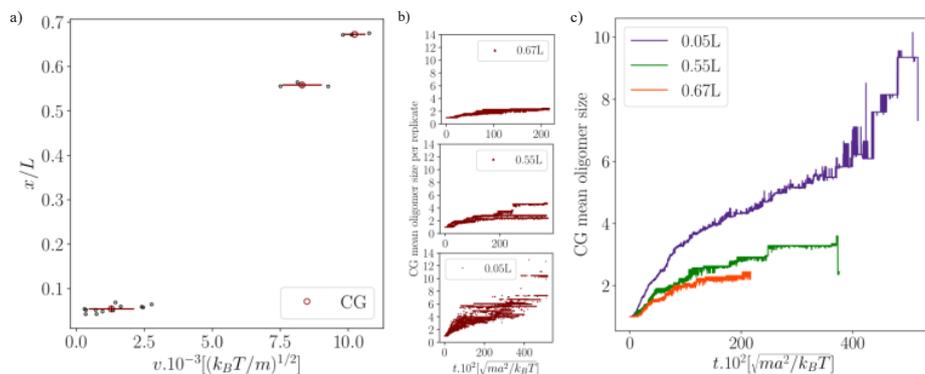
of the spidroins. However, at mean extensions of 0.1L, longer  $\beta$ -sheets cannot be formed extensively. It remains to be explored how higher flows induce a larger proportion of  $\beta$ -sheets, which is fully unknown. The non-equilibrium MD simulations on the nano-scale have the potential to unravel this dependence, which is a future step of this work.



**Figure 5.8:** a) mean number of AA that are in  $\beta$ -sheet as function of time. b) Mean number of AA in  $\beta$ -sheet for every residue type. The average was done over 6 replicates for Figure 5.8a and Figure 5.8

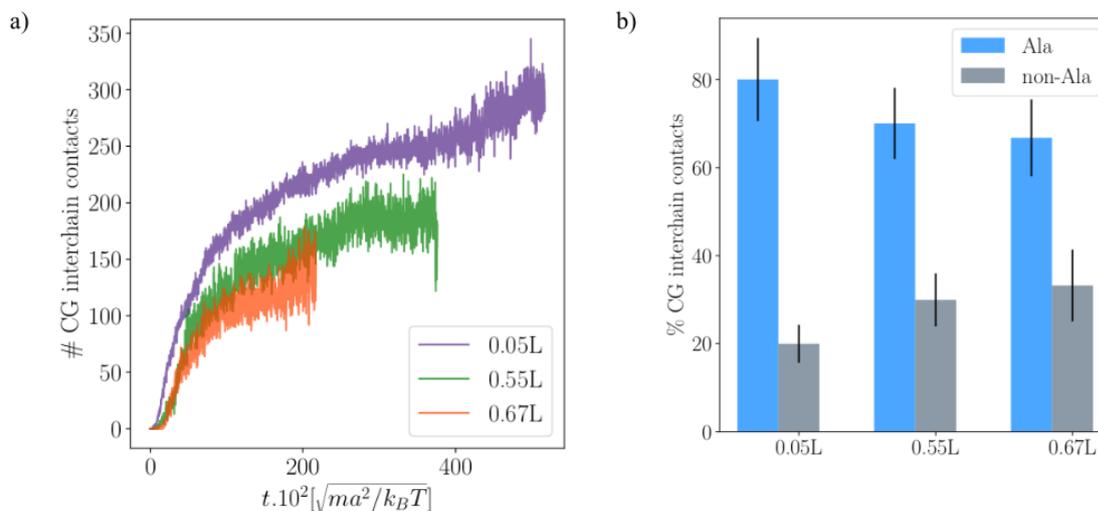
In the CG simulations, we followed the same approach as in the AA simulations. We simulated the systems at lower concentration (4 %wt/v) to observe more easily the possible effects of the flow in the assembly mechanism of the spidroins. This leads to a larger simulation that can be handled at the coarse-grained level. The average over the replicates of the mean peptide fractional extensions as a function of the mean flow velocity is shown in Figure 5.9a. We aimed at covering mean fractional extensions in the range of  $[0.15 - 0.4]L$ . However, spidroins at lower range of flow velocities readily assembled. As soon as two chains aggregate, their extensions are drastically reduced (see outliers at  $2.5 \times 10^3 [\sqrt{k_B T/m}]$ ). We compared the mean oligomer size evolution in time for every flow regime, which is shown by averaging over all replicates in Figures 5.9b and 5.9c, respectively. Higher protein fluctuations are presented in low flow velocities, which results in a higher mean oligomerization at a given time (purple curve in Figure 5.9c) compared to fast flow regimes (green and orange curves in Figure 5.9c). This suggests that flow not only extends the spidroin chains, but in this way also reduces fluctuations orthogonal to the flow direction (Chapter 4, Section 4.3), and thereby the likelihood of diffusional encounters. Thus, random associations are reduced, and controlled assembly into  $\beta$ -sheet enhanced. Overall, self-assembly of silk peptides slows down with increasing flow velocity. The contacts between the spidroins are also highly affected by flow (Figures 5.10a and 5.10b), and show a strong increase with the size and number of oligomers. We observed more inter-

## 5. NANOSCALE SELF-ASSEMBLY OF DRAGLINE SPIDROINS



**Figure 5.9:** Assembly in CG simulations. a) Mean CG peptide fractional extension  $x/L$  as a function of the mean flow velocity. Black points correspond to individual replicates, and maroon points correspond to the average over replicates in every flow regime, considering the last half of every trajectory. b) Mean oligomer size per replicate as a function of time for every mean fractional extension. c) Mean oligomer size as a function of time. For Figures a) and c) the averages are over 12 simulations for initial mean fractional extension of peptides from 0.05L up to 0.7L, and 3 realizations for initial mean fractional extension of peptides of 0.8L and 0.9L.

chain contacts in the CG model (Figure 5.10a) than in the AA model (Figure 5.7a), which is associated mostly with the fact that we considered attractive interactions only between the Ala residues. However, it might be possible that specific interactions of the residues in the intrinsically disordered fragments prevent the formation of interchain contacts in the poly-Ala regions, which remains to be explored and was suggested by A. Simmons et al. [19]. In contrast to the AA simulations (Figure 5.7b), in CG simulations, the percentage of contacts where Ala residues are involved is significantly higher than for the non-Ala residues. This is expected, since non-alanines interact with a purely repulsive potential. Interestingly, higher flow velocities and extensions result in more of those non-alanine contacts, due to the fact that chains are aligned in oligomers. One of these interactions that is relevant in the self-assembly of silk proteins, involve tyrosine residues, as previously shown experimentally [105].



**Figure 5.10:** a) Mean total number of interchain contacts as a function of time, averaged over all replicates per flow regime. b) Percentage of interchain contacts for Ala and non-Ala residues.

## 5.4 Conclusion

In this thesis, we presented multi-scale simulations of the self-assembly of tethered spidroins under uniform flow conditions. We were able to monitor the formation of oligomers, interchain contacts and  $\beta$ -sheets for short and intermediate peptide extension, and in two different concentrations. At all-atom simulations, we restricted to low flow velocities and peptide extensions, as our high peptide concentration impedes flow at high speeds. From this proof of concept simulations, a few preliminary conclusions can be drawn:

- The formation of oligomers is faster, when the peptides are less stretched. Larger protein extension limits their transversal movement, slowing down the process of self-assembly.
- Interchain contacts are larger for low velocity flow regimes, since they depend highly on the oligomerization of the spidroins.
- The Alanines form more interchain contacts than the residues in the amorphous region. This can favor the formation of  $\beta$ -sheet crystals during fibrillation.
- Simulations of the assembly of more extended peptides are required to test if  $\beta$ -sheets formation is enhanced.

## **5. NANOSCALE SELF-ASSEMBLY OF DRAGLINE SPIDROINS**

---

- The limitations of the CG simulations are associated mainly with the non-specificity in the interactions in the amorphous region and the irreversibility of multimerization. Nevertheless, they can elucidate general features of the assembly mechanisms of spidroins. Moreover, they can supplement AA MD simulations by simulating less concentrated systems.

## Summary and Conclusions

In this thesis, we aim to understand the influence of uniform flow on single and multiple dragline spidroins. To achieve this purpose, we performed MD simulations under uniform flow in two scales: atomistic and amino acid resolution. We were inspired by many complex processes involved in the self-assembly of spider silk proteins, from which the effects and possible mechanisms of an elongational flow in the s-duct are predominantly unknown. Based on this, the main motivation that drove our study is the understanding of the dynamics and self-assembly mechanisms of spidroins driven by the flow. To simulate flow at atomistic resolution, we first implemented uniform flow MD simulations in GROMACS and studied the behavior under a constant flow of water molecules. This implementation was later used both in the study of single spidroin stretching and in the self-assembly of spidroins. In the study of single silk peptides under flow, we investigated their coiled-stretch transition and structural dependence on mean uniform flow velocities. At this stage, we proposed a mesoscopic model that integrated hydrodynamics by using multi-particle collision dynamics. The proposed multi-scale model in the single silk peptide simulations is the basis to analyze the pathway formation of spidroin nano-assemblies.

We will discuss the findings of our investigation in the following sections. At the end, we will outlook the future steps of this research.

### MD Simulations of Molecules in Uniform Flow

Flow at the molecular level induces shear-induced unfolding of single proteins and can drive their assembly. In this context, MD simulations is a versatile tool to analyze flow-induced effects at the atomistic level. However, they have been scarcely used under conditions of hydrodynamic flow because of their high computational cost [57, 59, 60], which results in both calculating ensemble properties during short sampling times and using extremely high flows to observe conformational changes [75]. Our implementation of uniform flow in the widely used GROMACS software package takes advantage of the significant improvements in the computing performance of MD simulations [65] of recent years. These improvements are attributed to better hardware and computational parallelization algorithms [63]. Moreover, GROMACS includes several force fields and water models to simulate a large variety of biomolecules. For example, the force field and water model used for simulating spidroins are convenient and generally accepted for intrinsically disordered proteins [71, 72, 106]. In the study of uniform flow MD simulations, we, therefore, extended the "pull" module of GROMACS to force groups of atoms in a defined layer of the simulation box. We tested the extended pull code by the enforcement of explicit water molecules along the flow direction. The temperature coupling is in charge of relaxing the system to achieve a constant terminal velocity, and this velocity is not affected by the coupling time of the thermostat. We simulated our uniform flow MD systems a longer time than the simulations of similar previous studies that have reported flow MD simulations as well [56, 60]. It allowed us to sample in longer times, which results in a reduction of flow velocities until two orders of magnitude less than the flow MD approaches described previously (see Chapter 3 Section 3.2) [54]. Nevertheless, high flow velocities can not be achieved if the size of the pull slice is small, voids appear if the forces applied on the water molecules exceeds a threshold. In the case of slice sizes of 4 nm, this force threshold is 2.5 pN. The voids are reduced with the increase in the size of the pull slice but at the expense of increasing the computational cost of the simulations. In favor of us, mean flow velocities in order of magnitude faster than  $\sim$  cm/s are considered high in biology, i.e. blood flow velocities are around  $\sim$ 50 cm/s.

Classical uniform flow AA MD simulations allow probing hydrodynamic interactions in biomolecules with more accuracy than the CG MD simulations. Nonetheless, despite the fact that efficient atomistic MD suits are available [53, 58, 63], the computational cost of AA MD simulations is still high. Study of large systems (around  $\sim$  3 million atoms) requires not only to use a significant amount of resources but also to shorter simulation times.

---

We simulated uniform flow in contrast to the usual hydrodynamics flows that are modeled (i.e. shear flow, Poiseuille flow, and Couette flow), based on the approach that shear rates are considered imperceptible in the nano-scale. However, further extensions towards simulations of more complex flows can be implemented in all-atom MD systems.

## **Dynamics of a Single Silk Peptide under Uniform Flow**

We proposed multi-scale modeling of spidroins under flow to understand their flow-conformational dependencies at the molecular level. In this way, we included the atomistic details from the AA simulations and covered longer time scales and larger length scales from the CG simulations. At the atomistic modeling, we performed non-equilibrium MD simulations in GROMACS by imposing uniform flow. For the coarse-grained model, we modeled the silk peptide as a block copolymer and combined MD simulations with MPCD to consider hydrodynamics at the mesoscale. The drag force on the peptide was computed explicitly from the interactions between the solvent particles and every amino acid of the spidroin in both models. This computation of drag force allowed us to reproduce the force-extension behavior for silk proteins in atomic force microscopy experiments [86]. The conformation of the silk peptide changes with flow velocity in a similar fashion in both models: a decreasing tension towards the free end, where more stable conformations are present in the CG model for a given flow. In the formation of the contacts as a function of flow velocity is where we observe the major differences between the CG and MD simulations. The CG model is not able to reproduce the acute transition from a coiled to a completely extended state that is present in the AA model, in contrast, it shows a steady decreasing of contacts with the mean flow velocity. We attribute this to the approximation of the CG spidroin to a block copolymer that lacks specificity for the amorphous amino acids, where these non-specific amino acids interact through a LJ repulsive potential. Furthermore, another part that is important to analyze is the mesoscopic solvent of the CG model. MPCD describes the solvent-protein hydrodynamics interactions performed through momentum exchange, but interactions via hydrogen bonds are not included. Hydrogen bonds promoted by water are relevant for example in the formation of  $\beta$ -sheets. In the case of the CG alanine regions, it is considered by the "native" interactions of the CG potential. Nevertheless, within the amorphous region of silk, the interactions associated with hydrogen bonds that lead to the formation of secondary structures are also present and they are not included in the mesoscopic modeling of silk of this work.

## 6. SUMMARY AND CONCLUSIONS

---

With regard to the structural changes of the spidroins with flow velocity, AA MD simulations show the tendency of alanines to be in conformations that favor fibrillation. This behavior can be further explored in simulations of silk proteins with different sequences and larger molecular weights.

Overall, the multi-scale modeling of silk is satisfactory. AA and CG MD simulations complement between each other: AA simulations describe in detail the structural changes of spidroins with the flow, while CG simulations help us to simulate larger systems with a good description of the hydrodynamics interactions.

### **Nanoscale Self-assembly of Dragline Spidroins**

Once the multi-scale model of silk proteins under flow was built, the next step was to apply it in the simulations of the self-assembly of spidroins. The results for the analysis of the silk formation pathway in the nano-scale are still preliminary. However, they have shown us important insight that directs the future work towards bringing a broader explanation of the flow effects in the assembly of the spider silk.

Considering that we do not simulate the terminal domains of the spidroins, we did a pre-alignment of the repetitive silk peptides to have them in phase. The peptides are tethered along the flow, where we can monitor mainly their lateral aggregation. To have the peptides in phase facilitates the assembly via crystal formation in the poly-Ala regions, and allows to focus mostly in the flow effects involved in the spidroins packing. Nevertheless, we are aware of this simplified approximation, which we took based on previous studies: the NT and CT terminal domains experience conformational changes upon assembly inside the s-gland [2]. The CT domain is dimerized from the tail of the gland [76], and the NT domains dimerizes inside the first part of the s-duct after a drop of pH [2]. Therefore, both terminal domains promote the initial interconnections of the proteins (or pre-alignment). Regarding the assembly by the formation of crystals, it has been shown in several studies the high tendency of  $\beta$ -sheet structures between the poly-Ala regions of the spidroins [19, 27, 28].

We could monitor the spidroins oligomerization at different concentrations and flow velocities. Regarding the AA simulation, they are restricted to low extensions because of the high concentration of the system. To simulate more diluted systems at this resolution is limited by the computational cost. CG simulations can supply this limitation of the AA systems, but at expense of less accuracy because of the approximation of the CG spidroins to a block copolymer. From our results, it is reflected in both the irreversibility of oligomer

---

formation and the interchain contacts of the amorphous amino acids. However, they can show clearer effects associated with the flow, for example, the slow down of the assembly with the increasing of peptide extension.

Another important aspect is related to the percentage of interchain contacts where alanine residues are involved. AA MD simulations showed that alanines form more contacts than the residues in the amorphous region. This is a prior condition to promote fibrillation of the spidroins by the formation of  $\beta$ -sheet. Here, it is necessary to simulate the more extended peptides to monitor both the  $\beta$ -sheet arrangement dependence with flow conditions and the enhancement of  $\beta$ -sheet crystals by flow.

Overall, we have investigated by computer simulations the effect of flows on spider silk peptides. Our computational work gives insights into their flow-induced non-equilibrium conformational dynamics and preliminary hints into how flow may influence their self-assembly. We hope this computational approach to be the basis of future studies aiming at studying mechano-activated processes in silk and other biopolymers and the effect of flows on them.

## Outlook

The present work is the basis to continue investigating the effect of flow in the assembly of spider silk proteins. The steps towards this aim include:

- Investigating the formation of  $\beta$ -sheets with the increase of flow velocity. It involves simulating spidroins that are more extended in AA MD simulations.
- Simulation of larger systems with different concentrations and higher flow velocities. It can reveal general effects of the flow in the formation time of oligomers.
- MD simulations of more complex flows, i.e. shear flow, in which the spidroins will not be tethered along the flow direction. It will allow to analyze silk fibrillation along the flow direction.
- Introducing more specificity in the amorphous residues of the CG spidroins. In this way, the irreversibility of the assembly in large scales can be monitored.
- Analyzing the effects of shear in the terminal domains of the spidroins. Atomistically, steered MD simulations can be used to stretch the terminal domains, since the

## 6. SUMMARY AND CONCLUSIONS

---

inclusion of them under flow conditions increases the computational cost drastically. The effects of these terminal domains can be monitored in the assembly by extending CG MD simulations. Adding them to the ends of the repetitive spidroins involves a pre-alignment of the chains.

# Bibliography

- [1] E. Lukas, S. Smith, and T. Scheibel, “Decoding the secrets of spider silk,” *materials today*, vol. 14, no. 3, pp. 80–86, 2011. 1, 2, 39, 41, 58, 63
- [2] A. Rising and J. Johansson, “Toward spinning artificial spider silk,” *Nature chemical biology*, vol. 11, pp. 309–315, 05 2015. 1, 5, 58, 72
- [3] S. Winkler and D. L. Kaplan, “Molecular biology of spider silk,” *Reviews in Molecular Biotechnology*, vol. 74, no. 2, pp. 85–93, 2000. 1
- [4] M. J. Buehler, “Materials by design—a perspective from atoms to structures,” *MRS Bulletin*, vol. 38, no. 2, p. 169–176, 2013. 2
- [5] J. Y. Wong, J. McDonald, M. Taylor-Pinney, D. I. Spivak, D. L. Kaplan, and M. J. Buehler, “Materials by design: Merging proteins and music,” *Nano Today*, vol. 7, no. 6, pp. 488–495, 2012. 2, 57
- [6] S. Lin, S. Ryu, O. Tokareva, G. Gronau, M. M. Jacobsen, W. Huang, D. J. Rizzo, D. Li, C. Staii, N. M. Pugno, J. Y. Wong, D. L. Kaplan, and M. J. Buehler, “Predictive modelling-based design and experiments for synthesis and spinning of bioinspired silk fibres,” *Nature communications*, vol. 6, no. 6892, 2015. 2, 6, 39, 40, 41, 57
- [7] I. Agnarsson, M. Kuntner, and T. A. Blackledge, “Bioprospecting finds the toughest biological material: Extraordinary silk from a giant riverine orb spider,” *PLOS ONE*, vol. 5, pp. 1–8, 09 2010. 3
- [8] G. P. Holland, R. V. Lewis, and J. L. Yarger, “Wise nmr characterization of nanoscale heterogeneity and mobility in supercontracted nephila clavipes spider dragline silk,” *Journal of the American Chemical Society*, vol. 126, no. 18, pp. 5867–5872, 2004. 3

## BIBLIOGRAPHY

---

- [9] Y. Termonia, “Molecular modeling of spider silk elasticity,” *Macromolecules*, vol. 27, no. 25, pp. 7378–7381, 1994. 3
- [10] M. Xu and R. V. Lewis, “Structure of a protein superfiber: spider dragline silk,” *Proceedings of the National Academy of Sciences*, vol. 87, no. 18, pp. 7120–7124, 1990. 3
- [11] R. V. Lewis, “Spider silk: the unraveling of a mystery,” *Accounts of Chemical Research*, vol. 25, no. 9, pp. 392–398, 1992. 3
- [12] N. A. Ayoub, J. E. Garb, R. M. Tinghitella, M. A. Collin, and C. Y. Hayashi, “Blueprint for a high-performance biomaterial: Full-length spider dragline silk genes,” *PLOS ONE*, vol. 2, pp. 1–13, 06 2007. 3
- [13] M. Andersson, Q. Jia, A. Abella, X. Y. Lee, M. Landreh, P. Purhonen, H. Hebert, M. Tenje, C. Robinson V, Q. Meng, G. Plaza R, J. Johansson, and A. Rising, “Biomimetic spinning of artificial spider silk from a chimeric minispidroin,” *Nature Chemical Biology*, vol. 13, pp. 262–264, 2017. 4, 23, 39, 40, 41
- [14] P. Scheibel, *Fibrous Proteins*. 10 2008. 4, 57, 58
- [15] F. Vollrath and D. Knight, “Liquid crystalline spinning of spider silk,” *Nature*, vol. 410, pp. 541–8, 04 2001. 4, 5, 6, 7, 58
- [16] H.-J. Jin and D. L. Kaplan, “Mechanism of silk processing in insects and spiders,” *Nature*, vol. 424, pp. 1057–1061, 2003. 5, 58
- [17] L. R. Parent, D. Onofrei, D. Xu, D. Stengel, J. D. Roehling, J. B. Addison, C. Forman, S. A. Amin, B. R. Cherry, J. L. Yarger, N. C. Gianneschi, and G. P. Holland, “Hierarchical spidroin micellar nanoparticles as the fundamental precursors of spider silks,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 45, pp. 11507–11512, 2018. 5, 58
- [18] C. Rat, J. C. Heiby, J. P. Bunz, and H. Neuweiler, “Two-step self-assembly of a spider silk molecular clamp,” *Nature Communications*, vol. 9, no. 4779, 2018. 6
- [19] A. H. Simmons, C. A. Michal, and L. W. Jelinski, “Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk,” *Science*, vol. 271, no. 5245, pp. 84–87, 1996. 6, 53, 63, 66, 72

- [20] J. E. Jenkins, S. Sampath, E. Butler, J. Kim, R. W. Henning, G. P. Holland, and J. L. Yarger, "Characterizing the secondary protein structure of black widow dragline silk using solid-state nmr and x-ray diffraction," *Biomacromolecules*, vol. 14, no. 10, pp. 3472–3483, 2013. 6
- [21] M. Cetinkaya, S. Xiao, B. Markert, W. Stacklies, and F. Gräter, "Silk fiber mechanics from multiscale force distribution analysis," *Biophysical Journal*, vol. 100, no. 5, pp. 1298–1305, 2011. 6
- [22] M. Cetinkaya, S. Xiao, and F. Gräter, "Effects of crystalline subunit size on silk fiber mechanics," *Soft Matter*, vol. 7, pp. 8142–8148, 2011. 6
- [23] G. M. Gray, A. Van der Vaart, C. Guo, J. Jones, D. Onofrei, B. R. Cherry, R. V. Lewis, J. L. Yarger, and G. P. Holland, "Secondary structure adopted by the gly-gly-x repetitive regions of dragline spider silk," *International Journal of Molecular Sciences*, vol. 17, no. 12, 2016. 6
- [24] J. E. Jenkins, M. S. Creager, E. B. Butler, R. V. Lewis, J. L. Yarger, and G. P. Holland, "Solid-state nmr evidence for elastin-like -turn structure in spider dragline silk," *Chem. Commun.*, vol. 46, pp. 6714–6716, 2010. 6
- [25] S. Rammensee, U. Slotta, T. Scheibel, and A. Bausch, "Assembly mechanism of recombinant spider silk proteins," *Proceedings of the National Academy of Sciences*, vol. 105, no. 18, pp. 6590–6595, 2008. 6, 23, 39, 40, 41, 51
- [26] T. Giesa, C. C. Perry, and M. J. Buehler, "Secondary structure transition and critical stress for a model of spider silk assembly," *Biomacromolecules*, vol. 17, no. 2, pp. 427–436, 2016. 6
- [27] S. Keten, Z. Xu, B. Ihle, and M. J. Buehler, "Nanoconfinement controls stiffness, strength and mechanical toughness of  $\beta$ -sheet crystals in silk," *Nature Materials*, vol. 9, pp. 359–367, 2010. 6, 72
- [28] S. Keten and M. J. Buehler, "Nanostructure and molecular mechanics of spider dragline silk protein assemblies," *J.R. Soc. Interface*, vol. 7, no. 53, 2010. 6, 72
- [29] S. Keten and M. J. Buehler, "Atomistic model of the spider silk nanostructure," *Applied Physics Letters*, vol. 96, no. 15, p. 153701, 2010. 6

## BIBLIOGRAPHY

---

- [30] N. Kojic, J. Bico, C. Clasen, and G. H. McKinley, “Ex vivo rheology of spider silk,” *Journal of Experimental Biology*, vol. 209, no. 21, pp. 4355–4362, 2006. 6, 7, 39, 41, 58
- [31] A. Koeppel, P. R. Laity, and C. Holland, “Extensional flow behaviour and spinnability of native silk,” *Soft Matter*, vol. 14, pp. 8838–8845, 2018. 6, 7, 58
- [32] J. Sparkes and C. Holland, “Analysis of the pressure requirements for silk spinning reveals a pultrusion dominated process,” *Nature Communications*, vol. 8, no. 594, 2017. 6, 7, 58
- [33] P. Laity, S. Gilks, and C. Holland, “Rheological behaviour of native silk feedstocks,” *Polymer*, vol. 67, pp. 28–39, 2015. 7
- [34] G. Gompper, T. Ihle, D. Kroll, and R. Winkler, “Multi-particle collision dynamics: a particle-based mesoscale simulation approach to the hydrodynamics of complex fluids,” in *Advanced computer simulation approaches for soft matter sciences III*, pp. 1–87, Springer, 2009. 8, 19, 24, 40, 41
- [35] D. Frenkel and B. Smith, *Understanding molecular simulation: from algorithms to applications*, vol. 1. Academic press, 2001. 11
- [36] S. A. Adcock and J. A. McCammon, “Molecular dynamics: survey of methods for simulating the activity of proteins,” *Chemical Reviews*, vol. 106, no. 5, pp. 1589–1615, 2006. 11
- [37] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *The Journal of Physical Chemistry*, vol. 126, no. 014101, 2007. 14, 27, 35, 42
- [38] T. G. development team, *GROMACS Reference Manual*, version 2016 ed., 2016. 14, 26
- [39] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *Journal of Applied Physics*, vol. 52, pp. 7182–7190, 1981. 15, 27, 42

- [40] S. Xiao, M. Cetinkaya, B. Markert, and F. Gräter, “Mechanical response of silk crystalline units from force-distribution analysis,” *Biophysical Journal*, vol. 96, pp. 3997–4005, 2009. 17
- [41] G. Reddy and D. Thirumalai, “Dissecting ubiquitin folding using the self-organized polymer model,” *The Journal of Physical Chemistry B*, vol. 119, no. 34, pp. 11358–11370, 2015. 17, 18, 20, 41
- [42] R. Hockney, S. Goel, and J. Eastwood, “Quiet high-resolution computer models of a plasma,” *Journal of Computational Physics*, vol. 14, no. 2, pp. 148 – 158, 1974. 21
- [43] W. van Gunsteren and H. Berendsen, “Algorithms for macromolecular dynamics and constraint dynamics,” *Molecular Physics*, vol. 34, no. 5, pp. 1311–1327, 1977. 21
- [44] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, “Lincs: A linear constraint solver for molecular simulations,” *Journal of Computational Chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997. 21
- [45] K. Ley, C. Laudanna, M. I Cybulsky, and S. Nourshargh, “Getting to the site of inflammation: the leukocyte adhesion cascade updated,” *Nature Reviews Immunology*, vol. 7, pp. 678–689, 2007. 23
- [46] A. Alexander-Katz, M. Schneider, S. Schneider, A. Wixforth, and R. Netz, “Shear-flow-induced unfolding of polymeric globules,” *Physical review letters*, vol. 97, no. 13, p. 138101, 2006. 23, 40
- [47] C. Baldauf, R. Schneppenheim, W. Stacklies, T. Obser, A. Pieconka, S. Schneppenheim, U. Budde, J. Zhou, and F. Gräter, “Shear-induced unfolding activates von willebrand factor a2 domain for proteolysis,” *Journal of Thrombosis and Haemostasis*, vol. 7, no. 12, pp. 2096–2105, 2009. 23, 39
- [48] B. Savage, J. J. Sixma, and Z. M. Ruggeri, “Functional self-association of von willebrand factor during platelet adhesion under flow,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 1, pp. 425–430, 2002. 23
- [49] S. Schneider, S. Nuschele, A. Wixforth, C. Gorzelanny, A. Alexander-Katz, R. R. Netz, and M. Schneider, “Shear-induced unfolding triggers adhesion of von willebrand factor fibers,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 19, pp. 7899–7903, 2007. 23, 39, 40, 47

## BIBLIOGRAPHY

---

- [50] T. Krueger, H. Kusumaatmaja, A. Kuzmin, O. Shardt, G. Silva, and E. Viggren, “The lattice boltzmann method principles and practice,” 2017. 24, 40
- [51] P. Hoogerbrugge and J. Koelman, “Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics,” *EPL (Europhysics Letters)*, vol. 19, no. 3, p. 155, 1992. 24, 40
- [52] H. Noguchi, N. Kikuchi, and G. Gompper, “Particle-based mesoscale hydrodynamic techniques,” *EPL (Europhysics Letters)*, vol. 78, no. 1, p. 10005, 2007. 24
- [53] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, “Scalable molecular dynamics with namd,” *Journal of computational chemistry*, vol. 26, no. 16, pp. 1781–1802, 2005. 24, 70
- [54] Z. Chen, J. Lou, C. Zhu, and K. Schulten, “Flow-induced structural transition in the  $\beta$ -switch region of glycoprotein ib,” *Biophysical journal*, vol. 95, no. 3, pp. 1303–1313, 2008. 24, 40, 70
- [55] X. Zou, Y. Liu, Z. Chen, G. I. Cárdenas-Jirón, and S. Klaus, “Flow-induced structural  $\beta$ -hairpin folding of glycoprotein ib  $\alpha$   $\beta$ -switch,” vol. 99, no. 4, pp. 1182–1191, 2010. 24
- [56] S. E. Strong and J. D. Eaves, “The dynamics of water in porous two-dimensional crystals,” *The Journal of Physical Chemistry*, vol. 121, no. 1, pp. 189–207, 2017. 24, 70
- [57] E. Ramm, “Principles of least action and of least constraint,” *GAMM*, vol. 34, no. 2, 2011. 24, 70
- [58] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. a. Swaminathan, and M. Karplus, “Charmm: a program for macromolecular energy, minimization, and dynamics calculations,” *Journal of computational chemistry*, vol. 4, no. 2, pp. 187–217, 1983. 24, 70
- [59] G. M. Wang and W. C. Sandberg, “Non-equilibrium all-atom molecular dynamics simulations of free and tethered dna molecules in nanochannel shear flows,” *Nanotechnology*, vol. 18, no. 135702, 2007. 24, 70

- 
- [60] G. M. Wang and W. C. Sandberg, "Complete all-atom hydrodynamics of protein unfolding in uniform flow," *Nanotechnology*, vol. 21, no. 235101, 2010. 24, 40, 70
- [61] R. Khare, J. J. de Pablo, and A. Yethiraj, "Rheology of confined polymer melts," *Macromolecules*, vol. 29, no. 24, pp. 7910–7918, 1996. 25
- [62] D. Savio, L. Pastewka, and P. Gumbsch, "Boundary lubrication of heterogeneous surfaces and the onset of cavitation in frictional contacts," *Science Advances*, vol. 2, no. 3, 2016. 25
- [63] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1, pp. 19–25, 2015. 25, 41, 70
- [64] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, "Gromacs: fast, flexible, and free," *Journal of computational chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005. 25
- [65] H.-R. A. M. M. Vedran, A.-S. Camilo, and F. Gräter, "Molecular dynamics simulations of molecules in uniform flow," *Biophysical Journal*, vol. 116, pp. 1579–1585, 2019. 25, 41, 42, 46, 70
- [66] H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, and J. Hermans, "Interaction models for water in relation to protein hydration," in *Intermolecular forces*, pp. 331–342, Springer, 1981. 27
- [67] T. Darden, D. York, and L. Pedersen, "Particle mesh ewald: An nlog(n) method for ewald sums in large systems," *The Journal of Chemical Physics*, vol. 98, p. 10089, 1993. 27
- [68] Y. Zhou, E. Eng, J. Zhu, C. Lu, T. Walz, and T. Springer, "Sequence and structure relationships within von willebrand factor," *Blood Journal*, vol. 120, pp. 449–458, 2012. 27
- [69] C. Aponte-Santamaría, V. Huck, S. Posch, A. K. Bronowska, S. Grässle, M. A. Brehm, T. Obser, R. Schneppenheim, P. Hinterdorfer, S. W. Schneider, C. Baldauf, and F. Gräter, "Force-sensitive autoinhibition of the von willebrand factor is mediated by interdomain interactions," *Blood Journal*, vol. 120, pp. 449–458, 2012. 28

## BIBLIOGRAPHY

---

- [70] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, “Improved side-chain torsion potentials for the amber ff99sb protein force field,” *Proteins: Structure, Function, and Bioinformatics*, vol. 78, no. 8, pp. 1950–1958, 2010. 28, 42
- [71] S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw, “Water dispersion interactions strongly influence simulated structural properties of disordered protein states,” *The Journal of Physical Chemistry*, vol. 119, pp. 5113–5123, 2015. 28, 42, 70
- [72] J. a. Henriques and M. Skepö, “Molecular dynamics simulations of intrinsically disordered proteins: On the accuracy of the tip4p-d water model and the representativeness of protein disorder models,” *Journal of Chemical Theory and Computation*, vol. 12, no. 7, p. 3407–3415, 2016. 28, 42, 70
- [73] T. Perkins, D. Smith, R. Larson, and S. Chu, “Stretching of a single tethered polymer in a uniform flow,” *Science*, vol. 268, no. 5207, pp. 83–87, 1995. 35, 39, 40
- [74] E. Braun, S. M. Moosavi, and S. Berend, “Anomalous effects of velocity rescaling algorithms: The flying ice cube effect revisited,” *Journal of Chemical Theory and Computation*, vol. 14, no. 10, pp. 5262–5272, 2018. 35
- [75] P. Szymczak and M. Cieplak, “Hydrodynamics effects in proteins,” *Journal of physics: Condensed Matter*, vol. 23, no. 3, p. 033102, 2010. 39, 70
- [76] C. Rat, J. C. Heiby, J. P. Bunz, and H. Neuweiler, “Two-step self-assembly of a spider silk molecular clamp,” *Nature Communications*, vol. 9, no. 4779, 2018. 39, 72
- [77] H. Chen, M. A. Fallah, V. Huck, J. I. Angerer, A. J. Reininger, S. W. Schneider, M. F. Schneider, and A. Alexander-Katz, “Blood-clotting-inspired reversible polymer-colloid composite assembly in flow,” *Nature Communications*, vol. 4, no. 1333, 2013. 39
- [78] H. Chen, M. A. Fallah, V. Huck, J. I. Angerer, A. J. Reininger, S. W. Shneider, M. F. Schneider, and A. Alexander-Katz, “Blood-clotting-inspired reversible polymer-colloid composite assembly in flow,” *Nature Communications*, vol. 4, no. 1333, 2013. 39, 40

- [79] D. Dunstan, P. Hamilton-Brown, P. Asimakis, and W. Ducker, "Shear flow promotes amyloid- fibrilization," *Protein engineering, design selection : PEDS*, vol. 22, pp. 741–6, 10 2009. 39
- [80] J. Dobson, A. Kumar, L. F. Willis, R. Toma, D. R. Higazi, R. Turner, D. C. Lowe, A. E. Ashcroft, S. E. Radford, N. Kapur, and D. J. Brockwell, "Inducing protein aggregation by extensional flow," *Proceedings of the National Academy of Sciences*, vol. 114, no. 18, pp. 4673–4678, 2017. 39
- [81] A. Kamada, N. Mittal, L. D. Söderberg, T. Ingverud, W. Ohm, S. V. Roth, F. Lundell, and L. Christofer, "Flow assisted assembly of nanostructured protein microfibers," *Proceedings of the National Academy of Sciences*, vol. 114, no. 6, pp. 1232–1237, 2017. 39
- [82] W. Gospodarczyk and M. Kozak, "Microchip circulation drastically accelerates amyloid aggregation of 1–42  $\beta$ -amyloid peptide from felis catus," *ACS Chemical Neuroscience*, vol. 8, no. 11, pp. 2558–2567, 2017. 39
- [83] C. N. Trumbore, "Chapter seven: Shear-induced amyloid formation of idps in the brain," vol. 166, pp. 2558–2567, 2019. 39
- [84] O. Kratky and G. Porod, "Röntgenuntersuchung gelöster fadenmoleküle," *Recueil des Travaux Chimiques des Pays-Bas*, vol. 68, no. 12, pp. 1106–1122, 1949. 40, 47
- [85] D. Li, M. M. Jacobsen, N. G. Rim, D. Backman, D. L. Kaplan, and J. Y. Wong, "Introducing biomimetic shear and ion gradients to microfluidic spinning improves silk fiber strength," *Biofabrication*, vol. 9, no. 2, p. 025025, 2017. 40, 58
- [86] E. Oroudjev, J. Soares, S. Arcidiacono, J. B. Thompson, S. A. Fossey, and H. G. Hansma, "Segmented nanofibers of spider dragline silk: Atomic force microscopy and single-molecule force spectroscopy," *Proceedings of the National Academy of Sciences*, vol. 99, no. suppl 2, pp. 6460–6465, 2002. 40, 48, 54, 71
- [87] M. Geisler, S. Xiao, E. M. Puchner, F. Gräter, and T. Hugel, "Controlling the structure of proteins at surfaces," *Journal of the American Chemical Society*, vol. 132, no. 48, pp. 17277–17281, 2010. 40

## BIBLIOGRAPHY

---

- [88] R. G. Larson, T. T. Perkins, D. E. Smith, and S. Chu, “Hydrodynamics of a dna molecule in a flow field,” *Physical review E*, vol. 55, pp. 1794–1797, 1997. 40
- [89] A. Mohan and P. S. Doyle, “Unraveling of a tethered polymer chain in uniform solvent flow,” *Macromolecules*, vol. 40, pp. 4301–4312, 2007. 40
- [90] K. Rack, V. Huck, H. Masoud, F. D. A. S. W. Schneider, and G. Gompper, “Margination and stretching of von willebrand factor in the blood stream enable adhesion,” *Scientific Reports*, vol. 7, no. 14278, 2017. 40
- [91] C. E. Sing and A. Alexander-Katz, “Globule-stretch transitions of collapsed polymers in elongational flow,” *Macromolecules*, vol. 43, pp. 3532–3541, 2010. 40
- [92] A. S. Lemak, J. R. Lepock, and J. Z. Y. Chen, “Molecular dynamics simulations of a protein model in uniform and elongational flows,” *Proteins: Structure, Function, and Bioinformatics*, vol. 51, no. 2, pp. 224–235, 2003. 40, 47
- [93] A. Lemak, J. Lepock, and J. Chen, “Unfolding proteins in an external field: Can we always observe the intermediate states?,” *Physical review. E*, vol. 67, p. 031910, 2003. 40, 47
- [94] N. Kikuchi, F. J. Ryder, C. M. Pooley, and J. M. Yeomans, “Kinetics of the polymer collapse transition: The role of hydrodynamics,” *Phys. Rev. E*, vol. 71, no. 061804, 2005. 40
- [95] K. N. Ali D. Malay, Kazuharu Arakawa, “Analysis of repetitive amino acid motifs reveals the essential features of spider dragline silk proteins,” *PLoS ONE*, vol. 12, no. 8, p. 025025, 2017. 41
- [96] P. J. Flory and M. Volkenstein, “Statistical mechanics of chain molecules,” *Biopolymers*, vol. 8, no. 5, pp. 699–700, 1969. 47
- [97] N. A. Oktaviani, A. Matsugami, A. D. Malay, F. Hayashi, D. L. Kaplan, and K. Numata, “Conformation and dynamics of soluble repetitive domain elucidates the initial beta-sheet conformation of spider silk,” *Nature communications*, vol. 9, no. 2121, 2018. 51, 53
- [98] S. Cheng, M. Cetinkaya, and F. Gräter, “How sequence determines elasticity of disordered proteins,” *Biophysical journal*, vol. 99, no. 12, pp. 3863–3869, 2010. 53

## BIBLIOGRAPHY

---

- [99] S. B. Ozkan, G. A. Wu, J. D. Chodera, and K. A. Dill, "Protein folding by zipping and assembly," vol. 104, no. 29, pp. 11987–11992, 2007. 57
- [100] G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan, "A backbone-based theory of protein folding," *Proceedings of the National Academy of Sciences*, vol. 103, no. 45, pp. 16623–16633, 2006. 57
- [101] D. Shah and F. Vollrath, *Silk for sustainable composites*, pp. 91–109. Woodhead Publishing, 12 2017. 57
- [102] D. A. Parry and J. M. Squire, "Fibrous proteins: New structural and functional aspects revealed," in *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers*, vol. 70 of *Advances in Protein Chemistry*, pp. 1–10, Academic Press, 2005. 57
- [103] F. Chiti and C. M. Dobson, "Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade," *Annual Review of Biochemistry*, vol. 86, no. 1, pp. 27–68, 2017. 57
- [104] J. P. Huchra and M. J. Geller, "Groups of galaxies. i - nearby groups," *Astrophysical Journal*, vol. 257, no. 1, pp. 224–235, 1982. 60
- [105] B. P. Partlow, M. Bagheri, J. L. Harden, and D. L. Kaplan, "Tyrosine templating in the self-assembly and crystallization of silk fibroin," *Biomacromolecules*, vol. 17, no. 11, pp. 3570–3579, 2016. 66
- [106] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, "Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules," *Computer Physics Communications*, vol. 91, no. 1-3, pp. 1–41, 1995. 70