Dissertation

submitted to the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Engineering Sciences

Presented by

M.Sc. Matthias Julian Brosz

born in: Stuttgart, Germany

Oral examination: 29.11.2023

Coarse-grained Molecular Dynamics of semi-flexible (bio-)polymers under Force

Referees: Prof. Dr. Frauke Gräter

Prof. Dr. Uwe Bunz

GESAMTFAKULTÄT FÜR MATHEMATIK,
INGENIEUR- UND NATURWISSENSCHAFTEN

COMBINED FACULTY OF MATHEMATICS,
ENGINEERING AND NATURAL SCIENCES

RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

Eidesstattliche Versicherung gemäß § 8 der Promotionsordnung für die Gesamtfakultät für Mathematik, Ingenieur- und Naturwissenschaften der Universität Heidelberg / Sworn Affidavit according to § 8 of the doctoral degree regulations of the Combined Faculty of Mathematics, Engineering and Natural Sciences at the Heidelberg University

1. Bei der eingereichten Dissertation zu dem Thema / The thesis I have submitted entitled

   Coarse-grained Molecular Dynamics of semi-flexible (bio-)polymers under Force

   handelt es sich um meine eigenständig erbrachte Leistung / is my own work.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht. / I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

3. Die Arbeit oder Teile davon habe ich ~~wie folgt~~/bislang nicht[1)] an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt. / I have not yet/~~have already~~[1)] presented this thesis or parts thereof to a university as part of an examination or degree.

   Titel der Arbeit / Title of the thesis:……

   Hochschule und Jahr / University and year:……

   Art der Prüfungs- oder Qualifikationsleistung / Type of examination or degree:……

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich. / I confirm that the declarations made above are correct.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. / I am aware of the importance of a sworn affidavit and the criminal prosecution in case of a false or incomplete affidavit

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe. / I affirm that the above is the absolute truth to the best of my knowledge and that I have not concealed anything.

Heidelberg, den 28.09.2023 ……

Ort und Datum / Place and date

…… Unterschrift / Signature

[1)] Nicht Zutreffendes streichen. Bei Bejahung sind anzugeben: der Titel der andernorts vorgelegten Arbeit, die Hochschule, das Jahr der Vorlage und die Art der Prüfungs- oder Qualifikationsleistung. / Please cross out what is not applicable. If applicable, please provide: the title of the thesis that was presented elsewhere, the name of the university, the year of presentation and the type of examination or degree.

The German text is legally binding.

Zusammenfassung

Lebende Organismen bestehen aus semi-flexiblen Biopolymeren, die Zellen strukturelle Integrität und Funktionalität verleihen. Semi-flexible (Bio-)polymere assemblieren in hierarchischen Netzwerken, die durch das Zusammenspiel von entropischen und enthalpischen Effekten bestimmt werden. Solch assemblierte Netzwerke weisen bei mechanischer Beanspruchung ein nicht lineares Verhalten auf, wie zum Beispiel Versteifung bei Dehnung und Erweichung durch Kompression. Diese Nichtlinearität ist auf die Vielkörpernatur auf der Mikroskala zurückzuführen, die das Verhalten auf der Mesoskala maßgeblich beeinflusst. Aufgrund des Mangels an nicht-generischen skalen-übergreifenden Modellen ist das Verhalten semi-flexibler Polymernetzwerke bei mechanischer Last noch nicht vollständig verstanden. Demnach ist das Ziel dieser Arbeit, die wichtigsten molekularen Deformationsmechanismen in semi-flexiblen (Bio-)polymernetzwerken durch großskalige und chemisch informierte Molekulardynamik-Simulationen zu untersuchen.

Wir haben grob-körnige Modelle für zwei semi-flexible Polymere mit ähnlichen Persistenzlängen, nämlich Poly(*para*-phenylene ethynylene)s (PPEs) und Kollagen, mittels des Martini 3 Kraftfeldes entwickelt, um Molekulardynamik-Simulationen unter Kraft durchzuführen und Bereiche mit hoher Kraftkonzentration und bruchgefährdeten Bindungen zu identifizieren. Unsere Martini 3 Modelle erfassen die strukturellen, mechanischen und thermodynamischen Observablen aus atomistischen Simulationen und Experimenten aus der Literatur weitestgehend, einschließlich der Packung zwischen den Ketten, der mechanischen Biegesteifigkeit sowie den Lösungseigenschaften.

Wir zeigen, dass die Verschränkung von PPEs in großskaligen Netzwerken mit der Kettenlänge zunimmt. Zudem stellen wir fest, dass langkettigen PPEs unter Scherbeanspruchung Scherbänder mit extremen Scherraten im Schnellband ausbilden, in dem die Bruchkräfte am höchsten sind und Bindungen am ehesten versagen. Außerdem haben wir ein atomistisches Strukturmodell für die Kollagenmikrofibrille mit einstellbarer Crosslinkdichte erstellt und das Martini 3 Kraftfeld mit Gō-Potenzialen kombiniert, um eine Zunahme der Mikrofibril-

lenstreckung bei abnehmender Crosslinkdichte zu beobachten. Unser Martini 3 Kollagenmodel eignet sich dazu, die Mikrofibrillenstreckung von atomistischen Simulationen unter Kraft zu erfassen, die in Kollaboration mit dem Riken Institut in Kobe durchgeführt wurden.

Die beiden hier entwickelten grob-körnigen Modelle für das semi-flexible PPE und Kollagen ergänzen experimentelle Untersuchungen durch Vorhersage von Bindungsbrüchen in assemblierten Polymernetzwerken. Sie verschieben die Grenze von Molekulardynamik-Simulationen näher an die Realität, das heißt an ihre tatsächlichen biologischen oder synthetischen Pendants und werden in Zukunft die Untersuchung von mikrometergroßen Systemen mit verschiedenen Strukturkonfigurationen ermöglichen.

Abstract

Living organisms build upon semi-flexible biopolymers to confer structural integrity and functionality to cells. Semi-flexible (bio-)polymers assemble into hierarchical networks governed by an interplay between entropic and enthalpic effects. The assembled network features a non-linear response to mechanical load, like strain-stiffening and compression-softening. This non-linearity stems from the many-body nature at the microscale, which significantly influences the behaviour at the mesoscale. Due to the lack of non-generic scale-bridging models, the response of semi-flexible (bio-)polymer networks to mechanical stress is not yet fully understood. The aim of this thesis is thus to explore major molecular deformation mechanisms of semi-flexible (bio-)polymer networks by large-scale yet chemically informed molecular dynamics simulations.

We developed coarse-grained models for two semi-flexible (bio-)polymers with similar persistence lengths, namely poly(*para*-phenylene ethynylene)s (PPEs) and collagen, using the Martini 3 force field to perform molecular dynamics simulations under force and to identify locations of high-force concentration with bonds being prone to rupture. Our Martini 3 models largely capture key structural, mechanical and thermodynamic observables from atomistic simulations and experiments from the literature, including interchain packing, mechanical bending stiffness and solvation properties.

We show that the entanglement of PPEs in large-scale bulk assemblies increases with polymer chain length. We further observe that long-chain PPE networks under shear-flow form shear bands with extreme shear rates in the fast band, that is, where rupture forces are highest and bonds are likely to fail. Also, we built atomistic structural models for collagen microfibrils with a tuneable crosslink density and combined Martini 3 with Gō-like potentials to find an increase in microfibrillar stretching with decreasing number of crosslinks. Our Martini 3 collagen model is suited to capture the force-stretching of collagen microfibrils from all-atom simulations, performed in collaboration with the Riken institute in Kobe.

The two newly developed coarse-grained models for the semi-flexible PPE and collagen complement experiments by predicting bond rupture events in the large-scale assembled polymer networks. They push the frontier of molecular dynamics simulations more close to realism, that is, to their actual biological or synthetic counterparts, and will in future allow probing micrometer sized systems of various structural configurations.

Publications

Within the scope of this thesis, we published or prepared the following articles for submission into peer reviewed journals:

- Brosz M, Michelarakis N, Bunz U, Aponte-Santamaría C, Gräter F: Martini 3 coarse-grained force field for poly(*para*-phenylene ethynylene)s. *Physical Chemistry Chemical Physics*, **24**, 17, 9998-10010, 2022, *In press*.

- Diedrich L, Brosz M, Abele T, Steinke S, Gräter F, Göpfrich K, Aponte-Santamaría C: Energetics and kinetics of membrane permeation of photoresists for bioprinting. *Advanced Functional Materials*, 2023, *In revision*.

The second publication was performed in collaboration with two students, namely Lucas Diedrich and Salome Steinke, and is not scope of this thesis. Note that the first, second and third author contributed equally to this publication. The next journal article is based on chapters 4 and 5 of this thesis, however, still remains to be published.

- Brosz M, Buck J, Aponte-Santamaría C, Gräter F: Martini 3 coarse-grained force field for collagen microfibrils. *Journal of Chemical Theory and Computation*, *In preparation*.

Last but not least, the author of this thesis contributed with sections 3.3.8 and 3.3.9 to a joint publication under the lead of Maximilian Elter from the group of Uwe Bunz.

- Elter M, Brosz M, Kurth M, Gräter F and Bunz U: Breaking extremely strong polymer bonds: poly(*para*-phenylene ethynylene)s under mechanical stress. *In preparation*.

Acknowledgement

Over the past four years, many people have supported and encouraged me in my work on this thesis. First and foremost, I would like to express my deepest gratitude to Prof. Dr. Frauke Gräter for offering me this great opportunity, for her confidence in my abilities and guidance along the way. Her ideas, suggestions and advice on scientific topics and beyond, combined with her trust to let me explore freely inspired and enriched me a lot, for which I am very thankful. Additionally, I would like to thank Prof. Dr. Uwe Bunz for his role as secondary supervisor, TAC-committee member and referee, in particular, for his valuable comments and support during my work. Furthermore, I would like acknowledge the financial support from the HeiKa Graduate School on Functional Materials and the Klaus Tschira Foundation.

Special thanks is due to former and current members of the molecular biomechanics group who created a trustworthy and pleasant atmosphere at work. Svenja, Benedikt, Kai and Saber, with whom I shared both happy and painful moments at work. Salome, Lucas, Denis and Johanna for enduring and shaping me as a supervisor during our joint research projects. Further, I would like to express my deepest gratitude to Nick and Camilo, for not only providing post-doctoral wisdom and a lot of guidance along this way, but also for competitive soccer games. Also, I thank Debora for her patient support towards the end.

Besides, I would like to express my gratitude to the Heidelberger Institute for Theoretical Studies. This remarkable space for science and creativity was a great environment to foster my thoughts. Therefore, I will always remember the fellow scientists, administration, IT, kitchen and cleaning staff who created this wonderful place. I am very grateful to Prof. Yuji Sugita and Dr. Jaewoon Jung for providing me with this great opportunity to visit the Riken institute in Kobe and experience three amazing weeks in Japan.

Last but not least, I thank my family and friends who supported me over all those years. Particularly, my partner Lena, for believing in me, for the emotional support along the way and for all these wonderful memories that we collected over the past years.

Contents

List of Figures

List of Tables

1 Introduction

Over millions of years, polymer-based biological matter has continuously adapted to their environment and responded to mechanical stimuli to meet the divers structural and functional needs of living organisms. This evolution has equipped these materials to support processes, like cell growth and division, intracellular transport, and signaling [1, 2]. Living systems comprise cells that are responsible for the production of biopolymers, for instance, polynucleotides like DNA, and polypeptides like collagen and actin, which serve as the major constituents of the extracellular matrix and cytoskeleton, respectively, and provide mechanical stability and structural integrity to cells [3, 4]. For example in the cytoskeleton, stiff actin filaments are crosslinked by soft actin-binding proteins to form higher-order structures, like filopedia bundles and cortex networks [5, 6]. Such actin-based cytoskeletons provide cells with structural scaffolds and function as mechanosensing force transmitters to facilitate cellular processes, such as cell shape regulation and cell migration [7–9].

In direct proximity to the cytoskeleton, separated by a biological membrane, is the extracellular matrix comprising a hierarchical network of collagen fibers, consisting of fibrils with covalently crosslinked triple helices [10, 11]. Collagen-based fibrillar structures can either serve as force transmitting proteins, for example, between muscles and bones, or as structural support to the connective tissue by bearing high mechanical loads, e.g., up to $90\,\mathrm{MPa}$ in stretched achilles tendon [12–14]. Due to its high abundance in the human body, collagen is a commonly used biomaterial in tissue engineering to treat burns and wounds, to confer as a biological scaffold for bone and cartilage regeneration, and to promote cell integration and proliferation in cardiovascular applications [15]. Accordingly, collagen-based scaffolds are tailored to meet the divers mechanical and structural needs of the native tissue, which is achieved by altering, e.g., its composition and crosslinking conditions [16–18].

Biological scaffolds thus rely on polymers with sufficient bending rigidity to maintain a straight conformation, yet still exhibiting thermal shape fluctuations, to preserve structural integrity and facilitate post-functionalization [4, 19, 20]. This competition between bending energy and entropic effects is fundamental for the class of semi-flexible polymers and is characterized by the persistence length, which describes the length scale over which the

bond orientation persists along the chain [3, 21]. Figuratively speaking, the rigidity of a semi-flexible polymer lies in between a stiff and flexible chain, similar to an *al-dente* cooked spaghetti, which stiffness is somewhere between that of an uncooked and cooked noodle.

Beside living matter, synthetic semi-flexible polymers have been designed to capture similar mechanical properties and include, among others, supramolecular assemblies of carbon nanotubes and dendronized polymers [22, 23]. Of particular interest are highly conjugated and conductive semi-flexible polymers, like poly(*para*-phenylene)s or poly(*para*-phenylene vinylene)s [24, 25]. For instance, poly(*para*-phenylene ethynylene)s, or short PPEs, show superior optical and optoelectronic properties, due to their conjugated backbone, and can be straightforwardly functionalized with various substituents [26]. Hence, PPEs are highly attractive for biological or chemical sensor applications, and the printing of organic electronic devices, e.g., light emitting diodes or photovoltaic cells [27–29]. In stark contrast to conventional polymers with saturated backbone bonds, and thus rather low persistence lengths, these PPEs feature a persistence length approaching that of their assembled biological counterparts, i.e., collagen or actin, and form hierarchical networks with a complex many-body nature [25]. Such supramolecular assemblies exhibit a non-linear viscoelastic response to mechanical stress, such as reversible stiffening under shear, softening under compression, and, more generally, shape recovering from deformation [4, 19].

This non-linear response of networks comprising crosslinked or entangled semi-flexible (bio-)polymers, like collagen or PPEs, to mechanical perturbation is governed by processes on the microscopic to the mesoscopic scale, and, due to the lack of scale-bridging models, is not yet fully understood. So far, modeling semi-flexible polymers mainly focused on combining statistical physics-based continuum approaches, such as worm-like chain theory, with scaling laws, and hence considering the microscopic polymer dynamics only implicitly [30].

1.1 Modeling semi-flexible Polymers under Force

Polymers are long-chain molecules consisting of many repeating segments, the so-called monomers, and can be expressed mathematically using models from statistical physics. For example, the freely-jointed-chain model describes a polymer as a chain of uncorrelated linked segments of fixed length, and therefore as a random walk [31]. Besides, the freely-rotating-chain model also fixes the bond angle between adjacent segments to introduce orientational correlations along the polymer chain [32]. Godt et al., for instance, combined this model with electron paramagnetic resonance spectroscopy experiments and computational simulations to capture the persistence length of a single semi-flexible PPE chain [33].

By considering the rigid limit of this freely-rotating-chain model, i.e., bond length and bond angle approach zero, the most commonly used statistical physics-based model for semi-flexible polymers is derived, namely the worm-like chain model [32]. According to this model, semi-flexible polymers are represented in the continuum framework as inextensible elastic strands with a finite bending resistance and are expressed mathematically through the bending Hamiltonian (eq. 2.28) [30]. Broedersz and MacKintosh reviewed advances in the field, such as extensions of the classical worm-like chain (WLC) model to conditions out-of-equilibrium [19]. Among others, Marko and Siggia used a stretched WLC model to describe the response of a single DNA molecule to longitudinal stress [34–36]. Frey et al. proposed the worm-like bundle model to consider crosslinked fibrillar structures under shear force [37, 38]. Kroy et al. coupled a rate equation to the glassy WLC model, that embeds single polymers in glassy environments using a mean-field approximation, to monitor the competition between bond cleavage and bond formation in polymer networks under force [39, 40]. Alternatively, effective medium theory was introduced by Morse as an elastic continuum-based approximation of semi-flexible polymer networks [41]. Levine et al. relied on effective medium theory to predict crossovers between affine and non-affine strain regimes, and MacKintosh et al. to capture non-affine deformations with thermal fluctuations and non-linear elastic effects in fiber networks, such as strain-stiffening [42, 43].

Despite the far-reaching success of theoretical continuum models for inferring scaling laws, microscopic contributions due to the many-body nature of semi-flexible polymer networks are largely neglected. For this purpose, mesoscopic particle-based simulation techniques, such as multi-particle collision dynamics, dissipative particle dynamics or brownian dynamics, are often combined with generic bead-spring models to evolve many-body systems, subject to interparticle conservative, dissipative and random forces, in time. For example, multi-particle collision dynamics simulations of semi-flexible polymers in solution were performed by Nikoubashman et al. to find an increase in viscosity with persistence length under steady shear-flow, and by Gompper et al. to describe U-shape conformations under Poiseulle-flow [44, 45]. In addition, Frey et al. used Brownian dynamics to characterize such tumbling behaviour of semi-flexible polymers in shear-flow [46]. Recently, Winkler et al. incorporated hydrodynamic interactions into Brownian dynamics to perform Brownian multi-particle collision dynamics simulations of semi-flexible polymers tethered at impenetrable walls under an oscillatory shear-flow in order to confirm a transition from U-shape to stretched conformation with increasing shear rate [47, 48].

Hence, mesoscopic particle-based simulation methods utilize bead-spring models that, due to their generic nature, lack molecular interactions arising from the structure and chemistry of the semi-flexible polymer in question, which account for the enhanced chain stiffness. Furthermore, these models may not adequately capture the thermodynamic behaviour within highly entangled semi-flexible polymer networks, since the interchain packing in supramolecular assemblies relates to the chemical specificity of the polymer chain. As a consequence of this, chemical and thermodynamic characterization is important to determine the response of large-scale semi-flexible polymer networks to mechanical perturbation, in particular, to resolve the actual deformation across different chemical bonds and structural units in order to identify the exact locations of bonds being prone to rupture (Fig. 1.1).

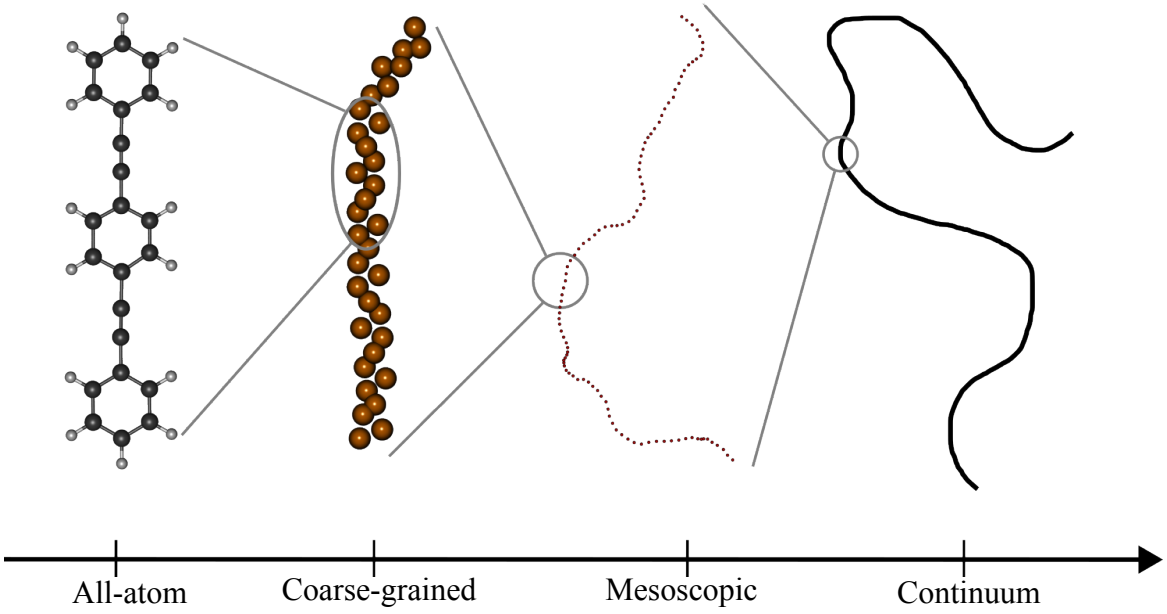

Figure 1.1. Multiscale modeling of semi-flexible polymers. Continuum and mesoscopic methods are suitable to simulate large networks of semi-flexible polymers, like poly(*para*-phenylene ethynylene)s shown here, at the expense of structural detail. All-atom molecular dynamics resolves structural details, however, is limited to mid-size networks. Coarse-grained modeling with Martini 3 not only captures sufficient atomistic details, but also enables molecular dynamics simulations of large systems close to the mesoscale.

On the microscopic scale, molecular dynamics simulations are a method of choice to determine the temporal and spatial evolution of a molecular system composed of many atoms or beads. All-atom molecular dynamics (MD) provides detailed information about interatomic forces and microscopic interactions, however, is limited to mid-size polymer networks of a few thousand chains on short timescales, due to large amount of computa-

tional resources required. For instance, McClintock et al. performed all-atom simulations of semi-flexible polymer blends to shed light on the microscopic diffusion dynamics and their strengthening upon welding. However, these atomistic simulations were limited to system sizes of roughly $62,000$ atoms and time-scales below $1.25\,\mu$s [49].

Alternatively, coarse-grained MD simulations cover larger spatio-temporal scales than all-atom ones by reducing the atomistic degrees of freedom, while still taking sufficient fine-grained information from the underlying atomistic structure into account. Coarse-grained modeling assumes that the dynamics of supramolecular assemblies can be captured without resolving every atomistic degree of freedom, and therefore aims to simplify a given atomistic model by grouping several atoms together and replacing them with coarse-grained interaction sites, so-called beads. Expanding on this, coarse-grained (CG) modeling averages over fast bond vibrations to smooth the free energy landscape of the atomistic model in order to enable the simulation of larger molecular complexes on longer timescales. Therefore, CG-MD simulations are suitable to determine the response of large networks comprising many semi-flexible (bio-)polymers to mechanical perturbation, while still maintaining a close to atomistic resolution to resolve rupture-prone bonds between pairwise segments.

1.2 Martini 3 force field for semi-flexible Polymers

The Martini force field is one of the most commonly applied CG simulation methods in the field of biomolecular and material sciences [50, 51]. It evolved over the past two decades comprising many molecular systems, like lipid-membranes, proteins and sugars [52–54], as well as polymers, ions and surfactants [55–57]. The success of the Martini force field can be attributed to its generic building block principle, that enables the modeling of complex molecular systems by coupling individual building blocks representing different chemical moieties [58]. According to the building block principle, an average of four heavy atoms are commonly replaced by one regular bead, although, higher resolution mappings are possible to consider structural details of complex molecules. For instance, the conjugated PPE backbone comprises aromatic rings and triple bonds, and is thus mapped to triangles and single beads using a two-to-one mapping with tiny beads (Fig. 1.2, *left, orange*).

Additionally, the Martini 3 force field combines this building block principle with a mixed atomistic- and experimental-based parameterisation and validation strategy to derive molecule specific force field parameters in an iterative process [59]. Most importantly, modeling semi-flexible polymers with Martini 3 enables simulations of large polymer networks close to the mesoscale, while still profiting from the underlying atomistic resolution [60].

This scale-bridging property of the Martini 3 force field arises from its parameterisation procedure, that focuses on reproducing semi-flexible polymer type specific mechanical, structural and thermodynamic properties, such as the mechanical bending stiffness of a single chain, the intermolecular packing and the oil-/water partitioning behaviour. In contrast to mesoscopic methods (see section 1.1), the Martini 3 force field relies on non-generic models, hence each parameterisation is directly related to a certain type of molecule. Besides, recent progress focused on improving the Martini 3 force field by rescaling protein-water interactions [61–63], targeting large-scale simulations on the whole cell or the SARS-CoV-2 envelope level [64, 65], and incorporating chemical reactions [66, 67]. For now, we refer the interested reader to section 2.2.1, and more importantly to the publication of Souza et al. to understand the Martini 3 coarse-grained force field in greater detail [59].

Thus, the Martini 3 force field is the simulation method of choice to capture the non-linear response of large-scale semi-flexible polymer networks to mechanical load, and moreover to resolve the stretching across divers chemical units to reveal rupture-prone bonds. Due to their similar persistence length, we selected the conjugated PPE (*left*) and collagen (*right*) as paradigmatic semi-flexible (bio-)polymers to be coarse-grained with the Martini 3 force field, and outline their chemical specifications in more detail below (Fig. 1.2).

Poly(*para*-phenylene ethynylene)s

PPEs are a class of semi-conductive polymers with a strongly delocalized π-electron system extending over the entire backbone [68]. The main chain of PPEs consists of alternating aromatic rings and triple bonds, that are linked by single carbon bonds (Fig. 1.2, *left*) [69]. Due to the alternation of single and multiple carbon bonds, the orbitals of the carbon atoms overlap, such that the PPE backbone is π-conjugated, and hence features a linear shape with an enhanced mechanical bending stiffness [70]. The persistence length of PPEs was estimated by Cotts et al. through a combination of light scattering experiments and worm-like chain theory (13.5 -16 nm), and by Godt et al. using computational methods and electron paramagnetic resonance spectroscopy (14.3 -19.1 nm), as mentioned in section 1.1 [25, 33].

Even though single-chains or mixtures of few PPE chains have been studied in detail, the dynamics of large-scale assemblies under equilibrium and out-of-equilibrium remains to be explored fully [71, 72]. So far, CG modeling of PPEs has been restricted to the freely-rotating-chain model proposed by Godt et al. (see section 1.1) [33]. Consequently, a model that inporporates the atomistic structure and experimentally known chain properties is yet to be developed, but is required to capture the many-body nature of entangled networks comprising semi-flexible PPEs under equilibrium conditions and under shear-flow.

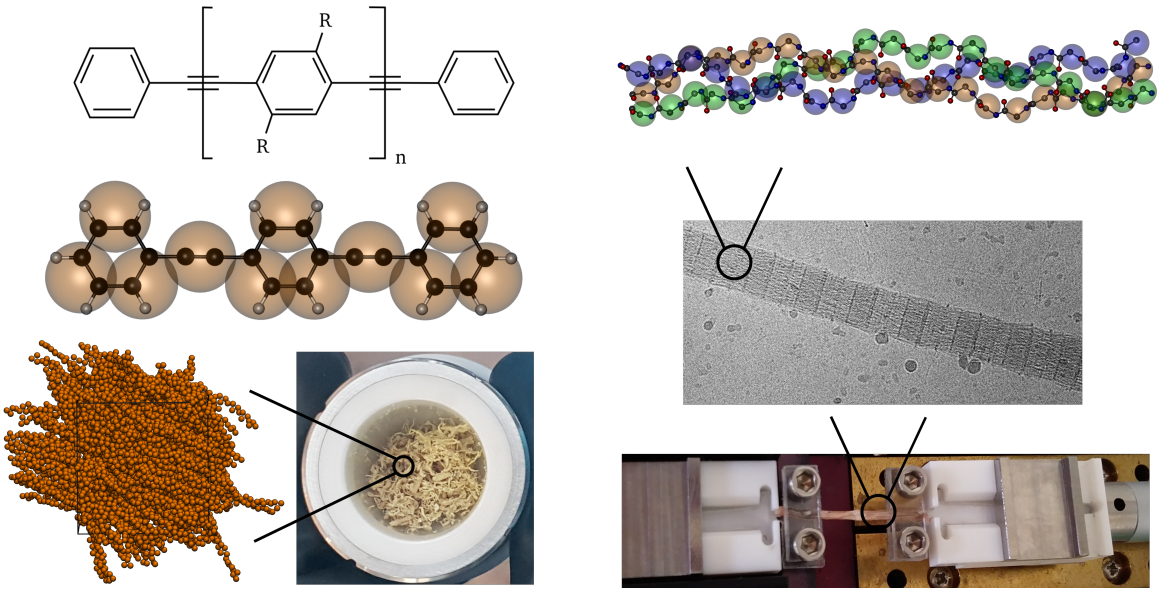

Figure 1.2. Modeling semi-flexible polymers under force. The PPE backbone comprises aromatic rings and stiff triple bonds (*left*) and features a persistence length close to that of the collagen triple helix (*right*). Due to this, both PPE and collagen form hierarchical networks, that determine their response to mechanical load. Of note, experimental graphics were kindly provided by Dr. Markus Kurth, Maximilian Elter and Aysecan Ünal.[1]

Collagen

Collagen type I features a hierarchical structure of fibers, fibrils and crosslinked triple helices with a 67 nm wide pattern of gap and overlap regions, i.e., called D-band (Fig. 1.2, *right*) [73]. Collagen forms stiff triple helices with persistence lengths between 11 -15 nm or 65 -180 nm obtained by experiments, including atomic force microscopy imaging or dynamic light scattering [74, 75]. Until now, Buehler et al. relied on a mesoscopic model for the collagen microfibril to characterize the influence of crosslinks on the response of the fibril to mechanical force [76]. Gautieri et al., however, used force-extension curves of small peptides to parameterise the Martini 2 force field, such as the force-extension of a glycine-proline fragment to derive the force constant of the CG bond length potential [77].

Despite such advances, the hierarchical structure of collagen, in particular, the triple helical shape and crosslinks, is not adequately described by the latter, thus providing grounds for reparameterising the collagen model with Martini 3 (Fig. 1.2, *right*). More importantly, collagen fibrils were still limited to the size of one D-band, and thereby preventing the sim-

[1] Cryo-milling PPEs from Maximilian Elter (see section 3.4), rat tail tendon under force from Dr. Markus Kurth, and Cryogenic electron microscopy image of collagen microfibril from Aysecan Ünal (see section 4.1).

ulation of 300 nm-long microfibrils under force, which is essential to elucidate the response of this crosslinked biopolymer network to mechanical perturbation on the microscale.

1.3 Aims and Scope

The aim of this doctoral thesis is to further our understanding of the structure and mechanics of semi-flexible (bio-)polymers with large persistence lengths in the range of the contour length. We focus on developing CG models for PPEs and collagen using the Martini 3 force field to enable simulations of multi-million particle systems, far beyond the capabilities of AA simulations, on customized high-performance computing clusters. By reducing the atomistic degrees of freedom, we aim to determine the structural alignment and force distribution within large-scale semi-flexible (bio-)polymer networks on long timescales to identify locations of high-force concentration, in particular, bonds being prone to rupture.

Coarse-grained poly(*para*-phenylene ethynylene)s networks under shear

Chapter 3 aims to elucidate the response of large-scale bulk systems composed of semi-flexible PPEs under shear. We develop a Martini 3 model for PPEs to characterize the structural alignment under equilibrium as well as the distribution of forces under non-equilibrium conditions, more precisely under shear-flow. We therefore seek to perform shearing simulations of highly entangled networks of semi-flexible PPEs on the Martini 3 level of resolution to predict the distribution of forces within the bulk system and reveal locations of high-force concentration, indicating bonds being prone to rupture. Besides, we collaborate with Maximilian Elter from the group of Uwe Bunz at the Heidelberg University to back up our findings, obtained form large shearing simulations, with cryo-milling experiments.

Structure generation of collagen microfibrils

We aim to determine the response of collagen microfibrils to mechanical force on the mesoscopic scale, in particular, the influence of crosslinking on the overall force-stretching. Therefore, we seek to build collagen fibrils close to the micrometer scale with tuneable crosslink configurations in chapter 4, and combine them with our Martini 3 force field model from chapter 5, to explore the relationship between crosslink density and fibrillar response to mechanical load. Also, we collaborate with Jung from the Theoretical Molecular Science Laboratory at the Riken institute in Kobe to validate our CG simulations with AA ones (43 million atoms), that were performed on the second largest supercomputer today.

Coarse-grained collagen microfibril under force

Chapter 5 mainly focuses on combining the collagen microfibrils from chapter 4 with the Martini 3 coarse-grained force field. We aim to identify a Martini 3 force field parameter set for the collagen triple helix, for one divalent and one trivalent crosslink type, that enables us to perform large-scale CG-MD simulations of differently crosslinked collagen microfibrils under force. With our Martini 3 force field parameter set at hand, we seek to elucidate the influence of various crosslinking configurations on the force-stretching of the collagen microfibril, in particular, to locate regions of high-force concentration within the microfibrillar network, where bonds between structural units are likely to fail.

1.4 Outline

In order to achieve these aims, we start with introducing the basic theoretical and methodological concepts about, e.g., molecular dynamics simulations, the Martini 3 coarse-grained force field, numerical integrators and applied boundary conditions, in chapter 2 of this thesis. Subsequently, we address the specific aims outlined above, i.e., PPE networks under shear-flow, the structure generation of differently crosslinked collagen microfibrils, and coarse-grained collagen microfibrils under force, in chapters 3, 4 and 5 of this thesis, respectively.

Finally, we conclude with summarizing our key findings and provide an outlook about modeling semi-flexible polymers with the Martini 3 force field and beyond. Specifically, we comment on future modeling approaches to elucidate the response of semi-flexible (bio-)polymer networks to mechanical stress in order to locate rupture-prone bonds, and on possible implications for technological applications, such as tissue engineering.

2 Theory and Methods

2.1 Molecular Dynamics Simulations

Molecular dynamics is a numerical simulation method to evolve a microscopic system of one or more molecules in time. Each molecule consists of several interacting particles, such as atoms or coarse-grained beads, which are modeled as point masses connected by harmonic springs, forming a classical N-body problem. Each particle possesses an instantaneous velocity and position in the three-dimensional Euclidean space, together defining the energetic state of the system. To propagate the many-particle system in time, Newton's equations of motion are integrated numerically using customized simulation softwares, so called molecular dynamics engines, on high-performance computing clusters.

Pioneering work in molecular dynamics simulations was performed by Alder and Wainwright, who solved the many-body problem for a system of several hard-spheres numerically using computational simulations [78, 79]. Later in the 1960s, Rahman introduced the, nowadays most commonly used, Lennard-Jones potential to model pair wise interactions between liquid argon atoms [80]. Due to the growing complexity of biomolecular or polymer systems, molecular dynamics simulations are usually performed on supercomputers or high-performance computing clusters with thousands of cores enabling the simulation of multi-million particle systems, such as large protein aggregates or protein-membrane complexes solvated in explicit water. Until now, the largest molecular dynamics simulation, a crowded 1.6 billion atom system, was performed in 2020 by Jung combining a specially designed extreme-scale molecular dynamics engine GENESIS, short for Generalized-Ensemble Simulation System, with more than $100,000$ CPU cores on the Fugaku supercomputer [81].

In the text that follows, a general overview of the basic principles of molecular dynamics simulation is provided, mainly based on the *GROningen MAchine for Chemical Simulations* (GROMACS) manual [82], the PhD theses of Florian Franz and Bogdan Costescu, former members of the Molecular Biomechanics group [83, 84], and more [85, 86].

2.1.1 Interparticle Interactions and Force Field

In molecular dynamics (MD) simulations, the motion of N individual particles in a system of molecules subject to internal forces acting between them, is determined by integrating Newton's equations of motion numerically over time. For this, the total energy of the molecular system is expressed mathematically with the Hamiltonian $H(\boldsymbol{r}, \boldsymbol{p})$, combining the kinetic energy $K = \sum_i^N \boldsymbol{p_i}/2m_i$ with a set of empirical functions representing the interparticle potential $V(\boldsymbol{r}_1, ..., \boldsymbol{r}_N)$ of the system as a function of the particle's positions $\boldsymbol{r}_1, ..., \boldsymbol{r}_N$,

$$H(\boldsymbol{r}, \boldsymbol{p}) = \sum_{i=1}^{N} \frac{\boldsymbol{p_i}}{2m_i} + V(\boldsymbol{r}_1, ..., \boldsymbol{r}_N), \tag{2.1}$$

where $\boldsymbol{p_1}, ..., \boldsymbol{p_N}$ are the momenta $\boldsymbol{p_i} = m_i \boldsymbol{v}_i$ of the particles with masses m_i and instantaneous velocities $\boldsymbol{v}_i$ for $i = 1, ..., N$ particles of the many-body system.

The interparticle potential $V(\boldsymbol{r}_1, ..., \boldsymbol{r}_N)$, often referred to as *force field*, contains detailed information about each particle type, such as charge or mass. Moreover, it comprises both the functional forms and an empirical parameter set, which together represent the interactions within the many-body system. The potential energy for a system of molecules originates from both intramolecular interactions between covalently bonded particles and those between non-bonded ones. The former type of interaction is modeled with 2-, 3- or 4-body interaction, such as a bond stretching, bond angle or a dihedral angle spring-like potential, respectively. In addition, non-bonded interactions between particles are modeled through electrostatic and van-der-Waals interactions. Taken together, force fields provide the essential information to describe the internal forces between particles within a molecular system, such that they are the basis for each MD simulation. The potential energy for covalent bonds between pairs of particles i and j with position vectors $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ at a given distance $r_{ij} = |\boldsymbol{r}_j - \boldsymbol{r}_i|$ is defined by,

$$V_b(r_{ij}) = \sum_{ij} \frac{1}{2} k_{ij}^b (r_{ij} - r_{ij}^0), \tag{2.2}$$

where k_{ij}^b is the force constant representing the interaction strength, and r_{ij}^0 is the reference distance of a specific particle pair ij with $(i, j) = 1, ..., N$ and $j \neq i$.

The 3-body interaction between triplets of particles i, j and k is defined as a harmonic angular restriction $V_a(\theta_{ijk})$ with an equilibrium value θ_{ijk}^0 and a force constant k_{ijk}^θ,

$$V_a(\theta_{ijk}) = \sum_{ijk} \frac{1}{2} k_{ijk}^\theta (\theta_{ijk} - \theta_{ijk}^0). \tag{2.3}$$

A dihedral angle between four particles i,j,k and l represents the angle between two intersecting planes, each spanned by two triplets i,j,k and j,k,l. In the simplest case, like for the improper dihedral, the potential is also modeled with a harmonic approximation. However, for simulations of proteins or membrane lipids, more complex functional forms are recommended. More precisely, a sum of cosine functions is used to allow the existence of multiple minima, each representing a different conformation, such as trans- or gauche.

$$V_d(\phi_{ijk}) = \sum_{ijkl} \frac{1}{2} k^{\phi}_{ijkl}(1 + \cos(n\phi_{ijkl} - \phi^0_{ijkl}), \tag{2.4}$$

where k^{ϕ}_{ijkl}, ϕ^0_{ijkl} and n are the force constant, reference torsion angle and multiplicity, that sets the number of minima, e.g. conformations, in the torsion potential, respectively.

Different molecules, like alkanes or polymers, require a more sophisticated function type to properly describe the 4-body interaction, namely the Ryckaert-Bellemans potential, commonly used for the Optimized Potential for Liquid Simulation (OPLS) force field [87]:

$$V_{rb}(\phi_{ijk}) = \sum_{n=0}^{5} C_n \cos(\phi_{ijkl} - 180°)^n. \tag{2.5}$$

Here, C_n are the Ryckaert-Bellemans coefficients and ϕ_{ijkl} the reference torsion angle. Note that the Ryckaert-Bellemans potential is suitable for maintaining the aromatic rings in a planar orientation and was used to model the π-conjugated backbone of PPEs [70].

Non-bonded interactions in MD simulations are grouped into two contributions. On the one hand, interparticle pairwise interactions are modeled with the 6-12 Lennard-Jones potential $V_{LJ}(r_{ij})$ and on the other, electrostatic interactions are expressed by using Coulomb's law $V_{Coulomb}(r_{ij})$. The functional form of the 6-12 Lennard-Jones potential combines a soft repulsive term for the overlap of electron clouds at short distances with an attractive term at longer ones r_{ij} representing electron correlations.

$$V_{LJ}(r_{ij}) = \sum_{ij} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \tag{2.6}$$

where ϵ_{ij} represents the well and σ_{ij} zero of the Lennard-Jones (LJ) potential.

Electrostatic interactions between particles i an j separated by a distance r_{ij} and carrying charges q_i and q_j, are described by incorporating the dielectric conversion factor $1/(4\pi\epsilon_r)$ and dielectric constant ϵ_r through the following mathematical expression:

$$V_{Coulomb}(r_{ij}) = \sum_{ij} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}. \tag{2.7}$$

Altogether, the total potential energy of the molecular system is given by summation of all bonded (eq. 2.2-2.4) and non-bonded contributions (eq. 2.6-2.7):

$$V(\boldsymbol{r}_1, ..., \boldsymbol{r}_N) = V_b(r_{ij}) + V_a(\theta_{ijk}) + V_d(\phi_{ijk}) + V_{LJ}(r_{ij}) + V_{Coulomb}(r_{ij}). \tag{2.8}$$

The potential energy function contains detailed information about the energetic state of each particle in the system as a function of its position in the Euclidean space. To describe the motion of a particle i in time, the internal forces exerted by the surrounding particles are determined by partial differentiation of $V(\boldsymbol{r}_1, ..., \boldsymbol{r}_N)$ with respect to $\boldsymbol{r}_i$. Due to the harmonic- or cosine-like function types in the total potential energy (eq. 2.8), analytic derivation is straightforward in order to obtain the total force $\boldsymbol{F}_i$ acting on a particle i:

$$\boldsymbol{F}_i(r_i) = -\frac{\partial V(\boldsymbol{r}_1, ..., \boldsymbol{r}_N)}{\partial \boldsymbol{r}_i}. \tag{2.9}$$

2.1.2 Time-Evolution and Numerical Integrators

The Hamiltonian $H(\boldsymbol{r}, \boldsymbol{p})$, defined by eq. 2.1, expresses the energetic state of a molecular system of N interacting particles as a single point $(\boldsymbol{p}_1, ..., \boldsymbol{p}_N, \boldsymbol{r}_1, ..., \boldsymbol{r}_N)$ in the phase space, that is a $6N$-dimensional space representing all possible states of the system. To evolve the molecular system in time, the total differential of the Hamiltonian $H(\boldsymbol{r}, \boldsymbol{p})$ is required [85]:

$$\frac{\mathrm{d}\boldsymbol{p}_i}{\mathrm{d}t} = -\frac{\partial H}{\partial \boldsymbol{r}_i} = -\frac{\partial V(\boldsymbol{r}_1, ..., \boldsymbol{r}_N)}{\partial \boldsymbol{r}_i}, \text{and} \tag{2.10}$$

$$\frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} = -\frac{\partial H}{\partial \boldsymbol{p}_i} = -\frac{\boldsymbol{p}_i}{m_i}, \tag{2.11}$$

where Hamilton's equations of motion (eq. 2.10 and eq. 2.11) describe the behaviour of a molecular system as a trajectory in phase space. For practical purposes, however, Hamilton's equations of motion are transformed by taking the time-derivative of eq. 2.11 and substituting it into eq. 2.10 to derive Newton's equations of motion,

$$\boldsymbol{F}_i(r_i) = m_i \frac{\mathrm{d}^2\boldsymbol{r}_i}{\mathrm{d}t^2} = m_i \frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t}. \tag{2.12}$$

Starting from Newton's equations of motion, a system of $i = 1, ..., N$ interacting particles, each with mass m_i, we obtain a set of $3N$ coupled second-order linear differential equations, which can be reformulated as a set of $2 \times 3N$ coupled first-order differential equations. A general solution of the $6N$ first-order differential equations requires a set of initial

conditions, given by the particles positions $\boldsymbol{r}_1(0), ..., \boldsymbol{r}_N(0)$ and velocities $\boldsymbol{v}_1(0), ..., \boldsymbol{v}_N(0)$. To numerically integrate the $6N$ first-order differential equations, a second-order energy-conserving leapfrog integrator is applied to propagate the system in time. The leapfrog algorithm is comparable to the velocity Verlet scheme [88, 89] and based on Taylor-expanding Newton's equations of motion (eq. 2.12) for $i = 1, ..., N$ particles:

$$\boldsymbol{r}_i(t + \Delta t) = \boldsymbol{r}_i(t) + \boldsymbol{v}_i(t + \Delta t/2) + \mathcal{O}(\Delta t^4), \tag{2.13}$$

with the discretisation error in Landau-notation $\mathcal{O}(\Delta t^4)$ and for the particle velocities [90],

$$\boldsymbol{v}_i(t + \Delta t/2) = \boldsymbol{v}_i(t - \Delta t/2) + \frac{\boldsymbol{F}_i(t)}{m_i}\Delta t + \mathcal{O}(\Delta t^2). \tag{2.14}$$

Starting from the initial conditions, the integrator first computes the interparticle potential together with the respective forces acting on each individual particle (eq. 2.8-2.9). Next, the position and velocity vector of each particle is updated to evolve the N-body system in time by applying eqs. 2.13 and 2.14. It is important to comment on the time increment Δt, which defines the accessible timescales of a given molecular system. In order to reach long timescales at low computational costs, the time increment should be as large as possible, although limited upwards by the fastest bond vibrations in the system. In case of atomistic MD simulations, constraint algorithms, such as LINCS or SETTLE [91–93], were developed to eliminate high-frequency bond vibrations between heavy atoms and hydrogens making the use of integration time steps of $\Delta t = 2$ fs possible. For coarse-grained simulations, like the ones presented in chapter 3 and 5, high-frequency bond vibrations are smeared out, due to the reduced degrees of freedom, enabling larger time increments of $\Delta t = 20 - 30$ fs, as commonly used for the Martini force field.

2.1.3 Short- and Long-Range Non-Bonded Interactions

A microscopic system consisting of N particles, communicating through non-bonded interactions, yields $N(N-1)/2$ pairwise interactions. Due to the quadratic increase in computational cost with increasing number of particles $[\mathcal{O}(\Delta N^2)]$, simulations of large molecular complexes consisting of many particles are computationally demanding. To deal with these scenarios, special MD algorithms were developed that truncate non-bonded interaction potentials after a certain cut-off distance, and therefore only consider the local environment around each particle. According to eqs. 2.7-2.6 each potential is defined as the inverse function of the interparticle distance $V(r_{ij}) \propto 1/r_{ij}^n$. This implies that both potentials, along with

their respective forces, approach zero as the distance between particles increases. More precisely, the LJ potential decays with $n = 6$, which is faster than the electrostatic potential that decays with $n = 1$. As a result, at a certain distance r_{ij}, LJ interactions approach zero much faster, while the electrostatic interactions remain significant. Therefore, errors originating from truncating pairwise LJ interactions beyond a cut-off of $1.0 - 1.2$ nm are negligible. Such truncations reduce the number of short-range interactions to be evaluated, and thus boost computational performance. In practice, all particles in the local environment of a reference particle, defined by a cut-off distance r_c plus some additional buffer r_v, are stored in a Verlet neighbor list. While this list is used to calculate the total force on the reference particle, it is updated less frequently to enhance computational performance [88].

For slowly decaying long-range electrostatic interactions on the other hand, the application of the above-mentioned cut-off scheme results in unfavorable simulation artifacts and is not sufficient for truncation purposes. Therefore, the Particle Mesh Ewald method is used to determine the long-range electrostatic interactions between particles [94, 95]. Here, the charges of all particles are mapped onto a discrete grid, and the electrostatic forces between grid points are concurrently evaluated in Fourier space. Next, the grid point specific force is distributed to all particles within the cell.

In case of coarse-grained simulations, an alternative method for the treatment of long-range electrostatics is commonly applied, namely the reaction field method. The reaction field method computes short-range electrostatic interactions for particles within a certain cut-off distance to a reference particle $r < r_c$ explicitly, while long-range interactions $r > r_c$ are treated with a mean-field approximation [96, 97]. Moreover, for long-range electrostatics, the surrounding medium is modeled as a continuum with the dielectric constant ϵ_{rf}:

$$V_{crf}(r_{ij}) = \sum_{ij} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \left[1 + \frac{\epsilon_{rf} - \epsilon_r}{2\epsilon_{rf} - \epsilon_r} \frac{r_{ij}^3}{r_c}\right] - \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_c} \frac{3\epsilon_{rf}}{2\epsilon_{rf} + \epsilon_r}. \tag{2.15}$$

The electrostatic contribution of the mean-field approximation is given by the constant term in eq. 2.15 to ensure that the potential approaches zero at the cut-off distance. As an example, for Martini 3 the dielectic constant of the reaction field is $\epsilon_{rf} = 15$.

2.1.4 Periodic Boundary Condition

In MD simulations, the molecular system of interest, such as a polymer chain or a protein complex, is surrounded by explicit solvent molecules, for example water or toluene. Particles in the molecular system are able to move in a predefined simulation box or unit cell

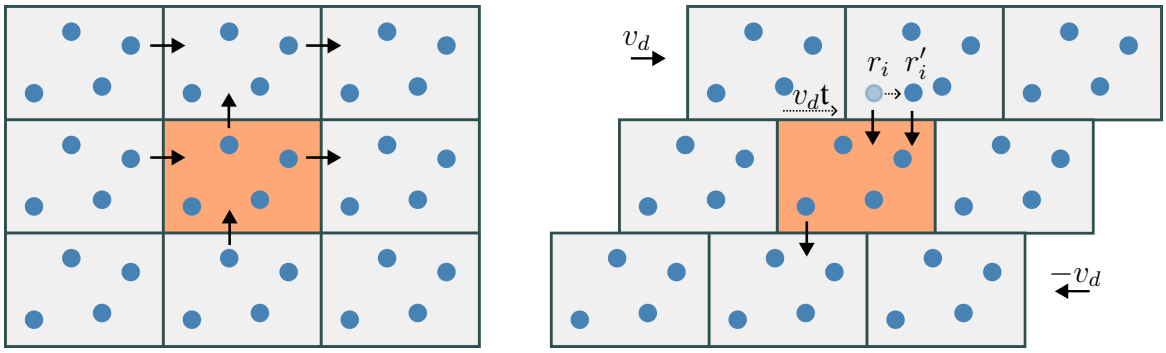

Figure 2.1. Periodic and Lee-Edwards boundary conditions. A Periodic boundary conditions are commonly used in MD simulations to approximate an infinite system with a finite one. Particles leaving one side of the unit cell (*orange*) reenter on the opposite. **B** The Lee-Edwards boundary condition combines the periodic boundary condition with a shear-like box-deformation. The upper and lower periodic images are deformed at constant velocity $\boldsymbol{v}_d$ and particles traversing the boundary are updated, i.e., r_i to r_i'.

and are subject to boundary conditions at the edges of the simulation box. To reduce edge effects, due to the finite size of the unit cell, MD engines apply periodic boundary conditions to approximate a large, infinite system by translating copies of the unit cell to the surroundings (Fig. 2.1A). More precisely, particles with a given velocity, crossing one edge of the unit cell, are replaced by particles with the exact same velocity entering from the opposite site. Thereby the number of particles in the unit cell is kept constant.

Besides, to prevent artifacts due to the periodicity of the molecular system, the size of the unit cell should be chosen to be large enough to prevent unwanted interactions between the molecule and its neighboring periodic image. For this purpose, modern MD algorithms rely on the combination of periodic boundary conditions and the minimum image convention algorithm. In particular, for the calculation of short-range interactions between particles, the minimum image convention ensures that only the closest image of each particle is considered for the pairwise interactions. The logical consequence of the minimum image convention is that the smallest possible box size of any molecular system is at least twice the interparticle cut-off distance of the non-bonded interactions.

2.1.5 Temperature and Pressure Coupling

Uncoupled MD simulations generate a microcanonical ensemble, in which each state is equally probable, by keeping the particle number, unit cell volume and total energy of the molecular system constant. Unfortunately, real-world experiments are usually performed

under isothermal-isobaric conditions, and therefore a pressure and temperature coupling are required to perform simulations in the canonical NpT (short for constant particle number, pressure and temperature) ensemble. In order to control both pressure and temperature in the simulation of a many-body system, a connection between macroscopic properties and microscropic particles is required. The equipartition theorem states that the kinetic energy is equally partitioned over the number of particles, such that the temperature relates to the kinetic energy of the microscopic system, given by the particle's velocities,

$$\frac{3}{2}Nk_BT = \left\langle \sum_{i=1}^{N} \frac{m_i v_i^2}{2} \right\rangle . \tag{2.16}$$

Here k_B is the Boltzmann constant and $v_i^2 = \boldsymbol{v}_i \cdot \boldsymbol{v}_i$ the dot product of the particle velocity. Obviously, the macroscopic temperature of the system is controlled by scaling the microscopic particle velocities $\boldsymbol{v}_i$. Similar to the temperature coupling, the virial theorem relates the potential energy of a molecular system, given by the sum of intermolecular pairwise interactions, to the macroscopic pressure according to [98, 99]

$$p = \frac{1}{3V}\left(\left\langle \sum_{i=1}^{N} m_i v_i^2 \right\rangle + \sum_{i=1}^{N} \langle \boldsymbol{F}_i \cdot \boldsymbol{r}_i \rangle \right), \tag{2.17}$$

where $\boldsymbol{r}_i$ is the position vector for and $\boldsymbol{F}_i$ the force acting on particle i. To match the pressure of the molecular system to the target value, the position of each particle is rescaled.

In general, a simple coupling scheme of temperature T or pressure p to an external heat or pressure bath was introduced by Berendsen using a first-order kinetic approach with relaxation time constant τ, that approximates the target value exponentially:

$$\frac{\mathrm{d}\phi(t)}{\mathrm{d}t} = \frac{\phi_0 - \phi(t)}{\tau}. \tag{2.18}$$

Here $\phi_0 = \{T_0, p_0\}$ represents the target temperature or pressure. Note that the Berendsen algorithm suppresses thermal fluctations of the kinetic energy, and therefore is not suited to generate a proper canonical ensemble. For this reason, the Bussi-Donadio-Parrinello temperature coupling was developed to enforce a canoncial distribution of the kinetic energy K by adding a stochastic term, namely a Wiener process [100].

$$\mathrm{d}K = (K_0 - K)\frac{\mathrm{d}t}{\tau} + 2\sqrt{\frac{KK_0}{N_f}}\frac{\mathrm{d}W}{\sqrt{\tau}}, \tag{2.19}$$

where N_f are the number of degrees of freedom, and W represents the Wiener noise. The kinetic energy is drawn from a Boltzmann distribution $P(K) \propto e^{K/k_BT}$. Hence, the velocity-rescaling factor, given by $\sqrt{K/K_0}$, is applied to each particle [101]. Accordingly, for the

Berendsen pressure coupling a stochastic term was added to generate a proper canonical ensemble. The stochastic cell rescaling barostat modifies the simulation box volume V, and thereby rescales the particle positions isotropically with $1/\sqrt[3]{V}$ [102].

In practice, the Berendsen-type coupling scheme for pressure and temperature is applied during the equilibration phase, where the macroscopic properties reach their predefined reference values. Subsequently, production runs are performed using a Bussi-Donadio-Parrinello temperature and Parinello-Rahman pressure coupling scheme to perform simulations in the canonical NpT ensemble [103]. Note that the Parinello-Rahman barostat not only generates a proper canonical ensemble, but also allows anisotropic reshaping of the unit cell. This is suitable for the equilibration of large asymmetrical boxes, for example, the simulation box of the $335\,\text{nm}$ long collagen microfibril discussed in chapter 5. However, for large molecular complexes the Parinello-Rahman pressure coupling suffers large oscillations. Expanding on this, Martyna, Tobias and Klein introduced chains of subsequent barostats to suppress such large pressure oscillations and to enable molecular dynamics simulations in highly asymmetrical boxes, like the collagen microfibrils in chapter 4 [104]. To be as close as possible to experiments, production runs are performed under isothermal-isobaric conditions to obtain a trajectory of the molecular system evolving over time.

2.1.6 Lee-Edwards Boundary Condition

A shear-like deformation of a molecular system is commonly required to determine the shear-induced self-assembly of single chains, viscosity of solvated polymers or the shear stress in a semi-flexible polymer network, such as PPEs (see section 3.3.8). To induce a shear-flow, various approaches were developed from accelerations groups combined with mass-weighted forces to the introduction of walls deforming at a different velocity. In 1972, Lee and Edwards combined a laminar flow with periodic boundary conditions to derive the Lee-Edwards boundary condition to impose shearing forces on a semi-infinite system [105].

A molecular system in a predefined unit cell is surrounded by periodic images of itself (Fig. 2.1A), each deformed at constant velocity v_d, defining the shear rate $\dot{\gamma} = v_D/L_C$, where L_C is the box length orthogonal to the shear flow with $L_C = L_z$. Accordingly, the periodic images above and below are deformed at constant velocity $\pm v_D$ (Fig. 2.1B), imposing a midpoint centered linear velocity profile on the particles in the flow direction,

$$v_i(z) = v_{i,Th} + v_D(z/L_C - 0.5). \tag{2.20}$$

where $v_{i,Th} = \sqrt{k_B T/m}$ is the thermal velocity of particle i sampled from a Maxwell–

Boltzmann distribution. To propagate a N-body system under shear in time, the shear rate in the flow direction $\dot{\gamma}\boldsymbol{e}_z$ is incorporated into Newton's equations of motion using [106],

$$\frac{\mathrm{d}^2\boldsymbol{r}_i}{\mathrm{d}t^2} = \frac{\boldsymbol{F}_i}{m_i} + \dot{\gamma}\frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t}\boldsymbol{e}_z. \tag{2.21}$$

Besides, particles crossing the bottom edge of the unit cell are reintroduced at the top with an updated position $r'_{i,z} = r_{i,z} + v_D t$ and velocity $v'_{i,z} = v_{i,z} + v_D$ in the flow direction, adjusted to their neighboring cell image. Note that combining the Bussi-Donadio-Parrinello thermostat with the Lee-Edwards boundary condition restricts the velocity-rescaling of each particle to its thermal component $v_{i,Th}$, resulting in a linear velocity profile (eq. 2.20).

2.1.7 Molecular Dynamics Simulations under Force

Besides the Lee-Edwards boundary condition, modern MD engines offer the possibility to apply external forces to the molecular systems, and thereby drive it out-of-equilibrium. Force-probe MD simulations are suitable to determine the distribution of forces within a large and highly hierarchical network of semi-flexible polymers, such as the collagen microfibril consisting of hundreds of intertwined molecules introduced in chapter 4 and 5.

Biasing MD simulations of collagen with an external force provides information, not only about the force-extension dynamics of a single molecule, but also about the distribution of forces within the microfibril. For this reason, an external pulling force is applied to the center-of-mass of a group of particles, e.g., the caps at the protein backbone, and distributed over the individual particles belonging to that pull group. Moreover, within the scope of this work, a constant force acting on each pull group was used (constant force pulling), resulting in a fast extension of the molecular system. Within a certain time, the distribution of forces in the system reaches a stationary state, from which the analysis can take place.

2.2 Coarse-Grained Molecular Dynamics Simulations

Biological or chemical processes, involving the communication of large molecular complexes, such as proteins, lipids or polymers, usually occur on long time scales beyond what is feasible with atomistic MD simulations performed on high-performance computing clusters or supercomputers. Due to the vast amount of fine-grained details within a large macromolecular complex, all-atom (AA) MD simulations reach a well-defined spatio-temporal limit, defined by the number of particles and the available computational resources. In order to bridge the gap between spatial and temporal scales of the molecular process and the

available computing hardware, several techniques were developed, among others, enhanced sampling methods, parallel computing architectures and coarse-grained modeling.

Coarse-grained modeling, based on the idea that not each atomistic degree of freedom is required to describe the overall dynamics of large molecular complexes, aims to simplify a given atomistic structure by grouping several atoms together and replacing them with coarse-grained interaction sites, so-called beads, pseudo- or super-atoms. Specifically, a molecular system on atomistic resolution consisting of N_{AA} atoms is mapped to coarse-grained (CG) representation with fewer degrees of freedom $N_{CG} < N_{AA}$ using a predefined mapping scheme, as illustrated for benzene in Fig. 2.2. This CG model thus, not only requires less evaluations of Newton's equations of motion per integration step, but also enables larger time increments, due to softer potentials between CG beads. In addition, CG modeling smooths the free energy landscape of the atomistic structure by averaging over fast bond vibrations to reduce molecular friction and accelerate the overall dynamics of the system.

In general, CG models for MD simulations are developed with an iterative procedure given by first identifying a mapping scheme between fine- and coarse-grained resolution, second parameterizing an interparticle potential, and third reproducing target properties from both higher-resolution and experimental data for validation purposes. Moreover, interparticle CG potentials are either parameterized though a top-down approach, reproducing experimental key observables, or with a bottom-up scheme using fine-grained reference simulations. Further distinction is made between compound specific or transferable CG models. While the former focuses on reproducing properties of certain molecules (e.g. DNA), the latter aims to map the entire chemical compound space from fine- to coarse-grained resolution in order to obtain a transferable, general purpose CG force field, like the Martini 3 CG force field used throughout this work [59].

Of note, Levitt and Warshel were the first to develop a simplified representation of the bovine pancreatic trypsin inhibitor by collapsing both the C_α and side-chain atoms into one single CG interaction site each in 1970s [107]. The parameters for their interparticle potentials were derived by time-averaging effective forces between residue specific CG beads. With the CG model at hand, key driving factors for protein folding were identified by searching the configurational space for stable low-energy conformations [108]. In the following sections a brief introduction for the Martini 3 coarse-grained force field together with Gō-like potentials is given. For a more in depth explanation about the Martini 3 coarse-grained force field the main publication with the supplementary information is recommended [59], as well as for the Gō-Martini 3 extension [109, 110].

2.2.1 Martini 3 Coarse-Grained Force Field

The coarse-grained Martini force field is one of the most commonly applied CG simulation methods in the fields of biomolecular or material sciences [50, 52, 111]. In general, the Martini force field is based on a generic building block principle [51]. Each building block is parameterized separately to cover a broad spectrum of diverse chemical moieties. Hence, the modeling of complex macromolecular systems is facilitated by coupling several building blocks together to new or larger molecules [112]. The parameterisation of the Martini 3 building blocks relies on reproducing experimental thermodynamic properties, like the partitioning free energies for a broad range of chemical substances between multiple polar and apolar solvents, the free energy of mixing in binary systems composed of different compounds, or the bulk densities of organic fluids [60].

For Martini 3, an average of four heavy atoms plus the associated hydrogens are replaced by one CG interaction site. When taking more complex structures into account (e.g. ring-like compounds), a higher resolution mapping is recommended, where three or two heavy atoms plus the associated hydrogens are substituted by one small or tiny bead type [50, 113]. Reparameterised small and tiny beads with improved solute-solvent interactions better reproduce packing properties of the system, such as oil-/ water partitioning behaviour and bulk densities. The Martini 3 bead types are classified as polar (P), nonpolar (N), apolar (C), halo-compounds (X), monovalent ions (Q), divalent ions (D) and water (W). Each of the P-, N-, C-, X- and Q-bead type is divided into subgroups describing the degree of polarity from 0 to a maximum of 6, like for the P-, N-, and C-bead type. The resulting interaction matrix consists of four blocks (organic, ion, others, water) with 22 levels. In addition, the Martini 3 force field introduces labels to fine-tune interactions between the CG sites by dividing each bead type into either chemical (e.g. hydrogen bonding, electron polarizability) or generic subgroups (e.g. self-interactions or partial charges).

The optimization of bonded interactions between Martini 3 beads relies on both mapping AA structures to CG resolution with the center-of-geometry scheme and matching the probability densities describing the bonded interactions between CG beads to those of the mapped trajectory. Moreover, the obtained bond lengths are refined by taking the shape and symmetry of the underlying atomistic structure into account, e.g. through comparison of the solvent accessible surface area or the Connolly's surface with AA data [59, 114]. Developing CG models within the framework of the Martini 3 force field is an iterative process consisting of three major steps: mapping, parameterisation and validation (Fig. 2.2).

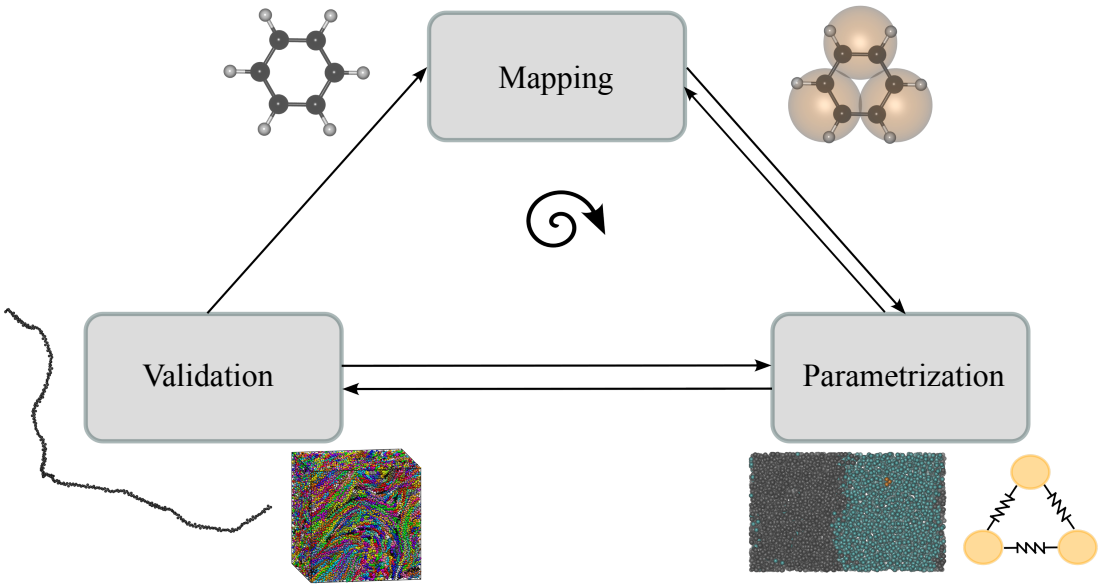

Figure 2.2. Iterative coarse-grained modeling with Martini 3. Mapping an AA structure, e.g., benzene, to CG resolution with the building block principle. Parameterising bonded and non-bonded interactions with a bottom-up and top-down approach by matching bonded terms probability densities to AA simulation and the oil-/ water partitioning to experimental reference data. For the validation, structural (e.g. bulk density), mechanical (e.g. persistence length) or thermodynamic properties are often chosen.

For the mapping step, a given AA structure is transformed to CG resolution by following the building block principle (e.g a benzene ring is represented by three tiny beads). For the parameterisation of bonded and non-bonded interactions between CG beads, a mixed top-down and bottom-up approach is applied. On one hand, the selection of chemical bead types is based on a top-down approach by finding similar fragments in the Martini 3 data base and comparing the partitioning free energy between hydrophilic or hydrophobic solvents to experimental measurements. On the other hand, parameterisation of bonded terms is performed using a bottom-up scheme by matching probability densities to a center-of-geometry mapped atomistic reference simulations, as outlined above. Finally, for validation purposes, the developed CG model should be able to reproduce key structural or thermodynamic observables, like chain stiffness or interchain packing in a bulk system.

Note that higher order structures, like the DNA double helix or the collagen triple helix, are not reproducible with Martini 3, due to collapsing hydrogen and heavy atoms into single interaction sites. Nonetheless, to maintain higher order protein structures, various structure-based CG techniques, such as the Gō-model from section 2.2.2, were combined with the Martini 3 force field, enabling the simulations of large protein complexes.

More specific, an elastic network model uses harmonic bond potentials between backbone beads to preserve the native molecular structure, however, prevents large conformational changes, such as folding processes [115]. For this reason, Gō-like contact maps were combined with the Martini CG force field to not only preserve the higher-order protein structures, but also to allow simulations of large conformational transitions using a LJ potential for the Gō-like interactions [110].

2.2.2 Structure-Based Gō-like Potentials

Structure-based coarse-grained modeling aims to simplify the underlying atomistic structure of a biomolecule by collapsing each residue into one interaction center to yield a bead chain model representing the amino acid sequence of complex proteins. Due to the strong emphasis on the native structure, such CG models are suitable for studying large conformational changes, like folding processes and the kinetics of proteins [109]. Regarding Martini 3, structure-based Gō-models are commonly used to maintain tertiary structures of proteins, e.g., the collagen triple helix, and are not required for polymers like PPEs (see section 2.2.3).

The structure-based Gō-model was introduced as a lattice model in 1975 to describe folding processes of proteins using Monte-Carlo simulations [116]. Proteins were mapped to a two dimensional square lattice by representing an amino acid sequence as grid points, such that a protein is conceptualized as a self-avoiding polymeric model. In addition, the amino acid sequence was taken into account by introducing both local bonds and non-local interactions between neighboring and more distant lattice points within a certain contact distance in the underlying native structure, respectively, together forming a Gō-like contact map. Hence, in Gō-like models each protein conformation is associated with a well-defined energetic state, such that transition probabilities between conformations can be estimated using the Metropolis algorithm to determine the protein folding kinetics [117].

Obviously, conformational transitions obtained from structure-based CG simulations strongly depend on the native structure of the protein complex, and therefore on the contact map used to define the local and non-local interactions between grid points. These contact maps are constructed either by geometric consideration based on overlaps of enlarged Van-der-Waals spheres, or by taking chemical information about atom contacts into account. The latter not only enables differentiation between repulsive and attractive contacts, but also considers the hydrophilic and aromatic properties [109].

2.2.3 Modeling with Gō-Martini 3

Incorporating the structure-based Gō-like approach into the physics-based Martini 3 force field aims to preserve the tertiary structure of the protein and to enable the simulations of conformational changes [110]. Hence, as outlined in the previous section 2.2.2, semi-flexible polymers with a linear shape (e.g. PPEs) can be simulated without Gō-like potentials, however, more complex molecules (e.g. collagen) require Gō-like potentials to preserve the triple helical shape under equilibrium conditions (see chapter 5.3.5). Gō-Martini 3 uses the atomistic structure of a molecular system to identify a Gō-like contact map by using an overlap criterion, besides balancing attractive and repulsive contacts [118]. Based on the derived Gō-like contact map, additional Gō-beads are positioned at the C_α atoms of the underlying atomistic model and introduced as virtual sites, linked to the corresponding backbone bead of the CG model, as shown later at the top of Fig. 5.4 in chapter 5.

Pairwise interactions between Gō-beads within a certain contact distance are expressed mathematically with a classical 6-12 LJ potential (eq. 2.6). Because of the introduced Gō-like potentials, regular non-bonded interactions between Martini 3 beads are excluded to ensure that the tertiary structure of the protein is well preserved. Hence, interactions between Gō-beads through the 6-12 LJ potential are transferred to the Martini 3 model by updating the positions of the backbone beads, that are linked through virtual sites. More precisely, for the standard Gō-Martini 3 model, the interaction strength between Gō-beads is set to $\epsilon_{G\bar{o}} = 9.414\,\mathrm{kJ\,mol^{-1}}$, i.e., 1.5 times the strength of hydrogen bonds, and the lower and upper cut-off distance to $0.3\,\mathrm{nm}$ and $1.1\,\mathrm{nm}$, respectively.

2.3 Free Energy Calculations

Besides the aforementioned technique of CG-MD simulations, free energy calculations are commonly used to bridge the gap between computational limitations and required timescales for biological process, like protein-ligand binding or the partitioning of small molecules across lipid bilayers [119]. To better understand the thermodynamics behind such biological processes, the underlying free energy landscape is calculated, specifically the difference in free energy between an initial state A and target B state is estimated [120].

In general, the Helmholtz free energy F, a thermodynamic potential for the NVT ensemble, is linked to the canonical partition function from statistical mechanics Z according to $F = -\beta^{-1} \ln Z(N, V, T)$ with the thermodynamic $\beta = 1/k_B T$. Hence, to determine the free energy for an arbitrary state, the partition function Z needs to be evaluated by integration over the whole phase space, which is impossible to perform in computer simulations. To overcome this limitation, free energy calculations aim to estimate the free energy differences between states A and B instead of an absolute value, using the following expression,

$$\Delta F = -\beta^{-1} \ln Z_B / Z_A. \tag{2.22}$$

Note that the approximation $\Delta F \approx \Delta G$ is valid for soft condensed matter systems with low compressibility [121]. Accordingly, the Gibbs free energy difference ΔG in the NpT ensemble represents a transition between states A and B along a predefined reaction coordinate or order parameter. Moreover, to obtain reliable estimates of the free energy of transferring a system from state A to state B, commonly used methods, like umbrella sampling, free energy perturbation or thermodynamic integration, are based on sampling multiple states along this transition path. For a more in depth explanation, the interested reader is referred to the literature [119, 121]. Finally, a very important relationship between the free energy difference ΔG and the partition coefficient $\log P$, which quantifies the distribution of a solute in a biphasic oil-/ water system, is mentioned,

$$\log P = \beta \cdot \frac{\Delta G_{\emptyset \to A} - \Delta G_{\emptyset \to B}}{N_A \cdot \log(e)}, \tag{2.23}$$

where N_A is Avogado's constant, e the Euler number and $\Delta G_{\emptyset \to A}$ the free energy of transferring a molecule from a reference state in vacuum to A. Since $\Delta G_{\emptyset \to A} - \Delta G_{\emptyset \to B}$ describes the free energy of partitioning a molecule in a binary oil-/ water system, eq. 2.23 enables the comparison of partitioning data for small molecules obtained from MD simulations with experimental measurements. In other words, the free energy of partitioning is directly related to experiments, and therefore the major validation target for selecting bead types, specifically determining non-bonded interactions for new molecules with the Martini 3 force field, such as the divalent and trivalent collagen crosslink (see section 5.3.1).

2.3.1 Umbrella Sampling

Umbrella sampling MD simulations are commonly used to derive a free energy profile along a predefined reaction coordinate termed potential of mean forces (PMF). The PMF connects two thermodynamic state points, usually separated by a free energy barrier. To overcome such a high energetic barrier and thereby transition from one state to the other, bias is added to the simulation, enabling proper sampling along the reaction coordinate ξ. For this reason, the transition path is discretized in multiple windows ξ_i for $i = 1, ..., n_w$, in each of which a harmonic bias potential $V(\xi_i)$ is applied to keep the molecule in a region close to ξ_i and sample the molecular state independent of the barrier height. Additional fine tuning of the harmonic bias potential is possible by adjusting the width and number of windows to ensure homogeneous sampling along the transition path, in particular in the vicinity of peaks in the PMF. To recover an unbiased PMF $\Delta G_i^u(\xi)$ from a biased probability distribution $P_i^b(\xi_i) = \exp\left(-\beta \Delta G_i^b(\xi)\right)$ sampled with biased MD simulations the added biasing potential $V(\xi_i)$ is subtracted in each window using $G(\xi_0)$ as a reference [119, 121, 122],

$$\Delta G_i^u(\xi) = -\beta^{-1} \ln P_i^b(\xi) - V(\xi) + G(\xi_0). \tag{2.24}$$

Advanced analysis methods used today, like the weighted-histogram-analysis method (WHAM) [122, 123], are suitable to correct biased probability distributions obtained from MD simulations in multiple windows at once using an iterative procedure. Here fore, WHAM assigns a weight $p_i(\xi)$ to each biased probability distributions $P_i^b(\xi)$ to minimize the uncertainty of the global unbiased probability distribution. Therefore, $\nabla_{p_i} \sigma^2(P_u) = 0$ with $\sum p_i = 1$. Finally, a reliable unbiased PMF and its statistical error are estimated [121, 124],

$$P^u(\xi) = \sum_i^{n_w} p_i(\xi) P_i^b(\xi). \tag{2.25}$$

Within this thesis, we performed umbrella sampling MD simulations compare the octanol-/ water partition coefficient of di(*para*-phenylene ethynylene) to experiments from the literature to select appropriate bead types for the Martini 3 mode of PPEs (see section 3.2.3).

2.3.2 Alchemical Transformations

Inspired by *alchemists*, former times scientists concerned with transmuting matter to gold, alchemical transformation is a free energy calculation method to transition a system from an initial state A to an end state B, passing through several unphysical intermediate states [119].

Alchemical transitions rely on both a thermodynamic cycle and an order parameter λ to enable the estimation of free energy differences between states, which are not directly accessible through computer simulations. In detail, a thermodynamic cycle is a set of reversible transitions between states, recovering the initial state at the end, and therefore conserving the free energy along its transition path [119]. Accordingly, the free energy difference associated to transition from states A to B is obtained through $\Delta\Delta G_{A\to B} = \Delta G_{\emptyset\to B} - \Delta G_{\emptyset\to A}$, where $\emptyset$ represents a reference state for both the initial and end state (Fig. 2.3).

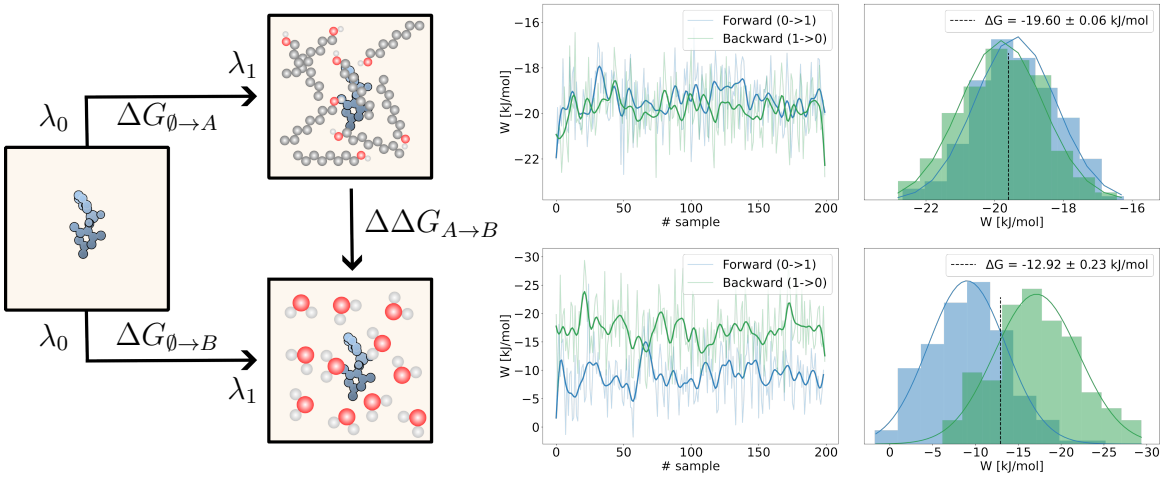

Figure 2.3. Alchemical transformations. The free energy of transferring a molecule from a hydrophobic to hydrophilic solvent $\Delta\Delta G_{A\to B}$ is estimated by combining a thermodynamic cycle with non-equilibrium MD-based free energy calculations. For each transition path $\emptyset \to A$ or $\emptyset \to B$, multiple thermodynamic integration replicas are performed to obtain a well-sampled work distribution for both forward and backward transition. According to the thermodynamic cycle, these two free energy estimates are sufficient to obtain the transfer free energy $\Delta\Delta G_{A\to B} = \Delta G_{\emptyset\to B} - \Delta G_{\emptyset\to A}$ [Figure adapted from [125, 126]].

The free energy difference to transition from a reference state to each target state $\Delta G_{\emptyset\to A}$ and $\Delta G_{\emptyset\to B}$ is estimated by thermodynamic integration [120]. To facilitate this, a coupling parameter λ is introduced along the Hamiltonian. The N-body system $H(\boldsymbol{r}, \boldsymbol{p})$ (eq. 2.1) can be integrated from the initial state at $\lambda = 0$ to the end state with $\lambda = 1$. Due to the reversibility of the thermodynamic cycle, alchemical transitions are usually computed both ways by gradually switching the solute-solvent interactions on for the forward ($0 \to 1$), and off for the backward ($1 \to 0$) transition. The non-equilibrium work associated to the forward $W_{\emptyset\to S} = \int_0^1 d\lambda \partial_\lambda H(\boldsymbol{r}, \boldsymbol{p})$ and backward $W_{S\to\emptyset} = \int_1^0 d\lambda \partial_\lambda H(\boldsymbol{r}, \boldsymbol{p})$ transitions is obtained by numerically integrating the instantaneous forces exerted on the system, where $S = \{A, B\}$ represents each target state [126]. Combining both forward $P_{\emptyset\to S}(W)$ and

backward $P_{S\to\emptyset}(W)$ work distributions with Crooks fluctuation theorem enables the estimation of the free energy difference associated to transitioning from A to B [127],

$$\frac{P_{\emptyset\to S}(W)}{P_{\emptyset\to S}(-W)} = \exp\left(\beta\left(W-\Delta G\right)\right). \tag{2.26}$$

The difference in free energy ΔG, as defined in Crooks fluctuation theorem, is estimated numerically with Bennett's acceptance ratio method by minimizing the uncertainty of the statistical estimate, similar to WHAM [128]. More precisely, starting from both forward and backward work distributions, Bennett's method represents a maximum likelihood estimator in order to obtain an unbiased free energy difference with minimum variance [129],

$$\left\langle \frac{1}{1+\exp\left(\beta\left(W-\Delta G\right)\right)} \right\rangle_{\emptyset\to S} = \left\langle \frac{1}{\left(1+\exp\left(-\beta\left(W-\Delta G\right)\right)\right)} \right\rangle_{S\to\emptyset} . \tag{2.27}$$

Lastly, it should be noted that free energy calculations were performed to validate the Martini 3 models for the collagen crosslinks with AA simulations (see section 5.3.1).

2.4 Semi-flexible Polymer Physics

As mentioned in section 1.1, polymers are long macromolecules comprising many repeating units, so-called monomers, and can be classified, depending on their mechanical bending stifness, as flexible, semi-flexible or rigid polymers. Semi-flexible polymers, as the name suggests, are softer than rigid polymers and stiffer than soft polymers, and are characterized by their persistence length L_P, that is within orders of magnitude of the polymer length L_0. Since the dynamics of such semi-flexible polymers is governed by a competition between entropic and energetic contributions, the worm-like chain model and its bending Hamiltonian was introduced by Kratky and Porod in 1949 [30, 130]:

$$H = \frac{\kappa}{2}\int_0^{L_0} ds \left(\frac{\mathrm{d}^2\boldsymbol{r}(s)}{\mathrm{d}s^2}\right)^2 . \tag{2.28}$$

where $\kappa = k_B T L_P$ is the bending rigidity, $\boldsymbol{r}(s)$ the three-dimensional space curve of the polymer chain, and $\boldsymbol{t}(s) = \partial\boldsymbol{r}(s)/\partial s$ the tangent vector describing the local orientation of the polymer. Accordingly, the persistence length is defined as the ratio of the thermal and bending energy of a single worm-like chain, that is expressed through $L_P = \kappa/k_B T$.

However, by integrating the orientational correlations $\langle \boldsymbol{t}(s) \cdot \boldsymbol{t}(s') \rangle$ along the arc length of the polymer s, the statistical physics interpretation of the persistence length is derived [3],

$$\langle \boldsymbol{t}(s) \cdot \boldsymbol{t}(s') \rangle = \exp\left(-|s - s'|/L_P\right). \tag{2.29}$$

where the persistence length describes the exponential decay of orientation correlations between tangent vectors along the polymer chain. In the continuum limit, by integrating eq. 2.29 along s, we obtain the mean-squared end-to-end distance [32, 130],

$$\langle R_e^2 \rangle = 2L_P^2 \left(\exp\left(-L_0/L_P\right) + L_0/L_P - 1\right), \tag{2.30}$$

which reduces to $\langle R_e^2 \rangle = L_0^2$ in the rigid rod limit, $L_0/L_P \to \infty$, and to $\langle R_e^2 \rangle = 2L_P L_0$ in the random coil limit, $L_0/L_P \to 0$. As a consequence of this, the behaviour of semi-flexible polymers is determined by the polymer's contour length, resulting in long semi-flexible chains, that behave like a flexible polymer chain with segments of size L_P [3].

In 1995, Bustamante, Marko and Siggia extended this worm-like chain model towards stretched polymer chains to describe the out-of-equilibrium behaviour under force [36]. Here fore, a linear force-extension relation $-F \cdot z$ was added to the Hamiltonian from eq. 2.28 to reduce the entropic energy, namely the amount of possible chain configurations, and moreover enable the persistence length estimation under force, such as commonly applied for optical tweezers or atomic force microscopy experiments [34, 35],

$$F_z = \frac{k_B \cdot T}{4 \cdot L_P} \cdot \left(\left(1 - \frac{z}{L_0}\right)^{-2} + 4 \cdot \frac{z}{L_0} - 1 \right), \tag{2.31}$$

where F_z is the force in the pull direction z, that is the end-to-end distance of the polymer chain projected on the z-dimension. It is important to note that eq. 2.31 is an interpolation formula, that captures the overall force-stretching of worm-like polymers well, however, is not an analytical solution, and thus possesses an error of around 10% [35].

Although, more complex models for the force-stretching of single worm-like chain polymers exist [131, 132], the approximation of Marko et al. from eq. 2.31 suffices our requirements, and was used to compare the persistence length of a single collagen molecule with atomistic simulations and experiments reported in the literature (see section 5.3.5).

3 Coarse-grained Modeling of Poly(*para*-phenylene ethynylene)s for Molecular Dynamics under Shear Flow

This chapter is based on our previous publication about the modeling of poly(*para*-phenylene ethynylene)s with the Martini 3 force field [133]. We used our CG model to shed light on the nematic ordering of large assemblies in PPE bulk systems. Beyond, we performed non-equilibrium CG-MD simulations under shear for large polymer bulks at different shear rates to analyze the concentration of rupture forces within semi-flexible polymer networks.

3.1 Introduction

As mentioned in chapter 1.2, PPEs belong to the class of semi-conductive and conjugated polymers with a strong delocalized π-electron system extending from one end of the chain to the other [134]. Due to the π-conjugated carbon atoms, PPEs exhibit extraordinary opto-electronic properties and have a wide range of applications, from the printing of electronic devices, e.g., organic transistors, to the fabrication of biological or chemical sensors [135–139]. The PPE backbone consists of aromatic rings and stiff triple bonds linked by single carbon bonds (Fig. 3.2A). Based on the alternation of single and multiple bonds, the π-orbitals of the carbon atoms overlap and the electrons are delocalized over the entire backbone. This is the reason for the planar structure of the aromatic rings, forming a linear backbone with an increased bending stiffness [140]. Thus, PPEs are classified as semi-flexible polymers and their bending stiffness is characterized by their persistence length [19]. Specifically, the persistence length of PPEs was determined experimentally using light scattering experiments and electron paramagnetic resonance spectroscopy combined with computational methods. Cotts et al. estimated a persistence length of $13.5\,\text{nm}$-$16\,\text{nm}$ and Godt et al. of $14.3\,\text{nm}$-$19.1\,\text{nm}$ [25, 33]. Accordingly, PPEs are approximately 10 to 50-fold stiffer than most conventional polymers without a π-conjugated backbone, such as polyethylene ($0.45 \pm 0.1\,\text{nm}$) or polymethyl methacrylate ($1.2 \pm 0.1\,\text{nm}$) [141–144].

The molecular origin of the high mechanical bending stiffness for PPEs has been addressed with high-level-of-accuracy quantum mechanical calculations. In fact, both the excited states and the energy barrier of the torsion angle for the backbone rotation of di(*para*-phenylene ethynylene)s or bis(phenyl ethynyl) benzenes were determined by density functional theory (DFT) calculations [70, 71]. Bagheri et al. combined MD with DFT simulations to improve the OPLS force field by keeping the PPE backbone planar [70, 140]. Additional studies used AA-MD simulations based on the developed OPLS force field to simulate the conformation of single PPEs and examine the rigidity and side-chain aggregation in different solvents [72, 140, 145]. Nevertheless, DFT calculations and AA simulations are limited to a few PPEs in solution, since both are computationally costly [52]. Alternatively, CG-MD is a well-established approach to cover larger spatio-temporal scales compared to AA-MD, what is achieved by reducing the atomistic degrees of freedom [50]. Godt et al. developed a simplified freely-rotating-chain model for PPEs by substituting the phenylene ring, the bond between phenylene and ethynylene unit and the ethynylene bond with one single bead each (see section 1.1) [33]. Although, this model reproduces the stiff nature of PPEs on a coarser level, it lacks the chemical and thermodynamic distinction between aromatic rings and triple bonds, giving rise to the π-stacking. To our knowledge, there exists no CG model that considers atomic structure and enables the simulation of large PPE networks on long timescales, properly capturing their thermodynamic properties.

To address this issue, we developed a CG model for PPEs based on the Martini 3 force field [59]. Martini 3 is suitable to model long-chain PPEs, while still taking the essential chemical details of the aromatic rings and the triple bond into account, giving rise to the enhanced chain stiffness. For PPEs, we combined a geometrical model to consider the shape of the conjugated backbone with an additional bending potential to fine-tune its mechanical bending stiffness. Also, we analyzed the structural organization by characterizing the nematic ordering of chains in a bulk material to find that the degree of alignment decreases with chain length. This shows how the dynamics of long PPEs is increasingly driven by entropic effects. Moreover, we performed shearing simulations to determine the response of large-scale PPE bulk systems to various shear rates by applying the Lee-Edwards boundary condition (see section 2.1.6). We observed that entangled networks of intermediate- and long-chain PPEs form shear bands with extreme shear rates in the fast band, i.e., where rupture forces are highest and bonds are prone to fail. Overall, our Martini 3 PPE model adequately reproduces the transition from rigid rod to random-coil like dynamics of networks comprising semi-flexible polymers under equilibrium and non-equilibrium conditions.

3.2 Methods

For Martini 3, on average, four heavy atoms and their respective hydrogens are replaced by one CG bead (see section 2.2.1) [52]. To consider more complex atomistic structures, like aromatic rings as in the case of PPE, a higher resolution mapping is recommended, where three or two heavy atoms plus the associated hydrogens are collapsed into one small or tiny bead [50, 111]. We followed the bottom-up and top-down Martini procedure and defined four tiny beads as our main building block (Fig. 3.2B). Next, we compared the octanol-/water partitioning free energy of di(*para*-phenylene ethynylene) to experiments from the literature to select the respective bead type of each building block [146, 147]. To optimize the bonded terms, we relied on a bottom-up strategy by matching the distribution functions for the bonded terms from CG to AA simulations [59]. Both AA- and CG-MD simulations were performed with the GROMACS 2020 MD simulation engine [148, 149]. A protocol regarding the simulation setup is described below.

3.2.1 All-Atom Simulation Protocol

MD simulations for PPEs on the atomistic level of detail were run with a modified version of the OPLS force field. The modification is implemented by converting parameters from the Polymer Consistent Force Field to OPLS and adding a Rickaert-Bellemanns function type for the dihedral potential to ensure a planar backbone (see section 2.1.1) [70]. Bagheri et al. performed DFT calculations to obtain an estimate of $4\,\mathrm{kJ\,mol^{-1}}$ for the energetic barrier of the rotation of aromatic rings, slightly above experimentally derived values of around $2\,\mathrm{kJ\,mol^{-1}}$ to $2.5\,\mathrm{kJ\,mol^{-1}}$ obtained from fluorescence spectroscopy [150–152].

All-Atom Single and Multiple Chains

For the parameterisation of the CG model, a single chain of PPE, consisting of four monomers, solvated with the SPC2016 water model in a cubic simulation box was selected as a reference [153]. We used the steepest descent algorithm to energy minimize the solvated system by gradually switching off backbone restraints. Next, we increased the temperature of the system to $300\,\mathrm{K}$ by applying the velocity rescaling thermostat (eq. 2.19) for $500\,\mathrm{ps}$ with $\tau_T = 0.1\,\mathrm{ps}$ [101]. For the NpT equilibration, we kept the pressure at $1\,\mathrm{bar}$ using the Parinello-Rahman barostat for $500\,\mathrm{ps}$ with $\tau_P = 2\,\mathrm{ps}$ and a compressibility of $4.5 \cdot 10^{-4}\,\mathrm{bar^{-1}}$ [103]. During the equilibration phase, we applied the Verlet-scheme to cut off short-range electrostatics and Van-der-Waals interactions at $1.0\,\mathrm{nm}$ [154]. Long-range

electrostatic contributions were treated with the Particle-Mesh-Ewald scheme [94, 95]. Each solvated PPE chain was simulated for 100 ns. In case of multiple chains, we randomly distributed ten chains of the same length in a $10 \times 10 \times 10\,\mathrm{nm}^3$ simulation box, and solvated the chains in water. We then applied the same simulation setup used for a single chain and performed a 100 ns production run to sample the chain packing within the polymer bundle.

All-Atom Bulk System

To setup of a PPE bulk system, GROMACS was used to randomly insert 300 chains, consisting of four monomers each, into a box of $80 \times 80 \times 80\,\mathrm{nm}^3$ Next, a three-step energy minimization was performed and a stochastic dynamics integrator applied to equilibrate the system for 1 ns at 300 K in the canoncial NVT ensemble with $\tau_T = 0.1$ ps [155]. Also, we gradually deformed the box volume for 100 ns to give the polymer chains enough time to assemble into a dense structure, and used the Parinello-Rahman barostat with $\tau_P = 2$ ps and a 1 fs time increment to pressurize the system and prevent numerical instabilities [104]. Finally, we simulated the assembled bulk for a $1\ \mu$s to enable the alignment of polymer chains, forming a densely packed network of semi-flexible PPEs.

3.2.2 Coarse-Grained Simulation Protocol

For developing a PPE CG model, we chose the Martini 3 force field [59]. We used the polyply package to generate the structures and topology files for PPEs with various chain lengths [112]. To calculate the neighbor list and to consider the electrostatics or Van-der-Waals interactions, the commonly used Martini 3 simulation parameters were applied [59]. The Verlet neighbor search algorithm updated the list every 20 to 40 steps, with a buffer tolerance of $5\,\mathrm{J\,mol^{-1}\,ps^{-1}}$. For electrostatics, the reaction-field cut off the Coulomb potential at $r_c = 1.1$ nm with a relative permittivity of $\epsilon_r = 15$, according to eq. 2.15. We modeled the Van-der-Waals interactions with the Lennard-Jones potential, a cut-off at $r_c = 1.1$ nm and the potential-shift with the Verlet scheme [55, 156]. We selected the common Martini 3 parameter set for CG-MD simulations, provided on https://www.cgmartini.nl for both the equilibration and production phases [59, 157].

Coarse-Grained Single and Multiple Chains

Parameterisation of bonded terms was obtained from single PPE chain in solution simulations. Specifically, we performed an energy minimization of a single chain of four monomers in vacuum with the steepest descent algorithm, and solvated the polymer chain

in a cubic box with either Martini 3 water beads and a Van-der-Waals radius of $0.21\,\mathrm{nm}$ or toluene [158]. The solvated system was energy minimized, while position restraining the backbone beads. For the NVT ensemble, the velocity rescaling thermostat stabilized the temperature at $300\,\mathrm{K}$ for $5\,\mathrm{ns}$ with a time constant of $\tau_T = 1.0\,\mathrm{ps}$. Next, the Parinello-Rahman barostat stabilized the pressure at $1\,\mathrm{bar}$ with $\tau_P = 12\,\mathrm{ps}$ and a compressibility of $3 \cdot 10^{-4}\,\mathrm{bar}^{-1}$. Finally, we performed production simulations for $100\,\mathrm{ns}$ with a $20\,\mathrm{fs}$ time increment for both the single chain and the ten chains in solution setups.

Coarse-Grained Bulk System

The bulk system for the Martini 3 model consisted of hundreds of PPE chains of varying polymer length. To get the initial position of the chains, we used GROMACS to randomly insert between 300 and 800 PPEs, each with a length of 20, 40 and 60 monomers, to a large $1000 \times 1000 \times 1000\,\mathrm{nm}^3$ box of vacuum. For the energy minimization of the bulk system, we selected the steepest descent algorithm. Next, the stochastic dynamics integrator was chosen to ensure a stochastic temperature coupling during the equilibration steps. The stochastic integrator stabilized the temperature for 5 -$8\,\mathrm{ns}$ at $300\,\mathrm{K}$ with a time constant of $\tau_T = 1.0\,\mathrm{ps}$. Moreover, we shrunk the simulation box for $1\,\mu\mathrm{s}$ to give the PPE chains sufficient time to assemble into a dense bulk structure. To this end, we deformed the simulation box constantly with a velocity of $-1.2\,\mathrm{m\,s}^{-1}$ to $-0.8\,\mathrm{m\,s}^{-1}$ and a reduced time increment of $5\,\mathrm{fs}$ to ensure numerical stability. Starting from the deformed bulk system, we applied the Parinello-Rahman barostat with $\tau_P = 12\,\mathrm{ps}$ and a slightly reduced timestep of $2\,\mathrm{fs}$ to pressurize the system at $1\,\mathrm{bar}$ for 2 -$10\,\mathrm{ns}$. Finally, we performed a $5\,\mu\mathrm{s}$ production run with a timestep of $20\,\mathrm{fs}$ to sufficiently sample the free energy landscape of the bulk system.

Besides, to sample energetic favourable states of the polymer networks, simulated annealing simulations were performed, according to the Martini setup for polymers [159–161]. The system was heated up to $400\,\mathrm{K}$ for $20\,\mathrm{ns}$, $500\,\mathrm{K}$ for $20\,\mathrm{ns}$, $600\,\mathrm{K}$ for $160\,\mathrm{ns}$ and finally to $800\,\mathrm{K}$ for $1.8\,\mu\mathrm{s}$ using a time increment of $10\,\mathrm{fs}$. Next, we cooled the system down to $300\,\mathrm{K}$ during a $400\,\mathrm{ns}$ equilibration stage giving the polymer chains sufficient time to self-organize. Compared to experimental annealing cycles at $\sim 360\,\mathrm{K}$, we not only annealed at higher temperatures, but also on a shorter timescales [162].

3.2.3 Free Energy Calculations

The free energy of transferring tolane, short for di(*para*-phenylene ethynylene), from octanol to water is the basis for the bead type selection with Martini 3. For this purpose, we performed biased MD simulations by umbrella sampling of tolane in an octanol-/water box to estimate the free energy of partitioning, as well as the PMF and the partition coefficient (eq. 2.23), quantifying the distribution of solute between the two immiscible solvents [163]. For the umbrella sampling simulations, we defined a reaction coordinate ξ pointing from the center of water towards the center of octanol with a 0.05 nm window spacing. Next, we restrained the solute for 100 ps with a force constant of $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$, energy minimized each window with the steepest descent algorithm and equilibrated the system for 3 ns with a 10 fs timestep. We applied the velocity rescaling thermostat to increase the temperature to 300 K with $\tau_T = 1$ ps and the Parrinello-Rahman barostat to stabilize the pressure at 1 bar with $\tau_P = 12$ ps and a compressibility of $3 \cdot 10^{-4}\,\mathrm{bar^{-1}}$. To ensure proper sampling of the underlying free energy landscape, each window was simulated for 500 ns with a 25 fs timestep. To estimate the PMF according to eq. 2.25, we applied the WHAM, as implemented in GROMACS, and compared the partition coefficient (eq. 2.23) with experiments [122, 124]. We finally checked for PMF convergence by backward block averaging the free energy barrier for tolane in the biphasic octanol-/water system using a 5 ns step size and performed a Bayesian bootstrap analysis of complete histograms to quantify the statistical uncertainty of the estimated free energy profile [124, 164].

3.2.4 Methods for Validation

The validation of the CG force field parameters followed a mixed bottom-up and top-down approach by comparing properties of interest with experimental or AA simulation data. More precisely, we focused on the mechanical bending stiffness of PPEs by estimating the persistence length and on packing properties, such as the bulk density and the orientation of packed chains, as validation targets. Therefore, we imported the trajectory into Python using the Molecular Dynamics Analysis software package (MDAnalysis) and calculated the properties with the Numerical Python package (NumPy) [165, 166]. To analyze the packing, we estimated the orthogonal distances between aligned chains, and projected the center-of-mass (COM) of each chain onto the main COM axis of the bundle.

Persistence Length

According to section 2.4, the persistence length describes the mechanical bending stiffness of the polymer, thereby characterizing its local intrinsic flexibility, and is defined as the rate of the exponentially decaying orientation correlations along the polymer chain (eq. 2.29) [167, 168]. To estimate the persistence length from MD simulations of single PPE chains, we imported the CG trajectory into Python and applied the scalar product function of NumPy to determine the angle between pairwise segments along the backbone θ_{ij}. More specifically, we defined the CG beads representing the triple bonds as the segments of the polymer chain, linked adjacent segments by a vector, and then normalized each vector to obtain the unit tangent vector. Hence, we obtained a chain of unit tangent vectors following the backbone of the polymer. From this chain, we calculated the pairwise scalar product $\boldsymbol{t}_i \cdot \boldsymbol{t}_j$ to determine the cosines of the angles $\cos\theta_{ij}$ enclosed by two segments i and j. Next, we averaged over the number of segments $N_M = |j - i|$, before averaging over time.

We also applied the worm-like chain theory to estimate the persistence length according to eq. 2.30 [169]. For this, we used the trajectory from the single chain in solution protocol and calculated the end-to-end distance between pairwise segments along the polymer chain. Subsequently, we squared and averaged the end-to-end distances over the count of segments N_M, before averaging over time, as described above. After preparing the CG data, we plotted both the averaged cosine angles $\langle\cos\theta_{ij}\rangle$ and squared end-to-end distances $\langle R_e^2\rangle$ as a function of the count of segments N_M. To estimate the persistence length of semi-flexible PPEs, we fitted eq. 2.29 and eq. 2.30 to the trajectory obtained from CG simulations using the curve fitting function from the Scientific Python package [170].

Packing Properties

The packing properties are characterized by the assembly of multiple PPEs into a parallel, aligned bundle. We analyzed the packing of the Martini 3 model by comparing the axial and radial displacement of single chains within a polymer bundle to experimental and AA simulation data. For the radial displacement, we determined the π-stacking by calculating the orthogonal distance between chains of the bundle. For the axial displacement, we projected the COM of each polymer onto the main COM axis of the bundle. More specifically, to estimate the orthogonal distance between adjacent polymers d_π, we defined the end-to-end vector for a single chain of the bundle $\boldsymbol{R}_e$, and linked the COM of all the other polymers to one end of the end-to-end vector $\boldsymbol{R}_p$. From the norm of the vector product $|\boldsymbol{R}_e \times \boldsymbol{R}_p|$, we computed the area content spanned by both vectors and divided the normalized vector

product by the length of $|\boldsymbol{R}_e|$ to estimate the orthogonal distance between chains:

$$d_\pi = \frac{|\boldsymbol{R}_p \times \boldsymbol{R}_e|}{|\boldsymbol{R}_e|}. \tag{3.1}$$

In addition, we calculated the axial displacement describing the movement of aligned polymers. Here fore, we projected the COM for each polymer orthogonally onto the main COM axis of the bundle to obtain an estimate for the axial displacement. More precisely, we defined the main COM axis as the vector between the COMs at both ends of the bundle $\boldsymbol{R}_b$, linked one end of the COM axis to the COM of any other polymer $\boldsymbol{R}_k$ and projected the COM of each polymer onto the main COM axis, according to [171],

$$\boldsymbol{P}_{\boldsymbol{R}_b}(\boldsymbol{R}_k) = \frac{\boldsymbol{R}_k \cdot \boldsymbol{R}_b}{\boldsymbol{R}_b \cdot \boldsymbol{R}_b} \cdot \boldsymbol{R}_b. \tag{3.2}$$

Here, $\boldsymbol{P}_{\boldsymbol{R}_b}(\boldsymbol{R}_k)$ denotes the vector projection of the COM of the polymer onto the COM axis of the bundle. We normalized the projected vector $|\boldsymbol{P}_{\boldsymbol{R}_b}(\boldsymbol{R}_k)|$ and subtracted half the length of the main COM axis $|\boldsymbol{R}_b|$ to obtain the axial displacement for the COM of each polymer with regard to the COM of the bundle. In the end, the respective probability distributions were obtained with the kernel density estimator from the Seaborn package [172].

Density of the Bulk System

For the analysis of the bulk system, we focused on both macroscopic and microscopic properties that characterize the packing of multiple PPEs within a network. On the macroscopic scale, we calculated the bulk density of PPEs from both AA and CG simulations by performing a NpT equilibration and computing the property of interest from the energy file with GROMACS. In addition, we rescaled the Martini 3 density with the ratio of molar masses from the AA and CG representation to enable a comparison with AA simulations.

Radial Distribution Function

On the microscopic scale, we analyzed the spatial correlation by computing the radial distribution function (RDF). To enable a comparison between Martini 3 and AA simulations, we first mapped both the AA and CG trajectories to a resolution of one particle per aromatic ring. We then imported the mapped trajectory into MDAnalysis, excluded the intrachain particles and used the InterRDF command to determine the RDF between N_P polymers consisting of N_M beads each. The RDF $g_{ij}(r)$ quantifies the probability of finding bead i from polymer I at a distance r from bead j belonging to polymer $J > I$ [173]:

$$g_{ij}(r) = (N_P N_M g_0)^{-1} \cdot \sum_{I=1}^{N_P} \sum_{J>I}^{N_P} \sum_{\substack{i=1 \\ i\in I}}^{N_M} \sum_{\substack{j=1 \\ j\in J}}^{N_M} \delta\left(|\boldsymbol{r}_i - \boldsymbol{r}_j| - r\right), \tag{3.3}$$

where $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ denote the position of particle i and j from polymer I and J with $i \in mI$ and $j \in J$, respectively, r is the distance from a reference bead i, δ is the Dirac delta function and $g_0 = 4\rho\pi r^2 dr$ a normalization factor, assuming a uniform particle distribution.

Nematic Correlation Function

In addition to the pairwise spatial correlation, we estimated the nematic correlation function (NCF) to characterize the orientation of different polymer chains relative to each other [174]. The NCF $\Pi_{ij}(r)$ quantifies the alignment of PPE chains within a bundled network by combining the orientation correlations with the RDF. The NCF therefore weights the probability of finding two beads i and j at a distance $r = |\boldsymbol{r}_i - \boldsymbol{r}_j|$ with the absolute value of the dot product of their tangent vectors $\boldsymbol{t}_i$ and $\boldsymbol{t}_j$, as given in eq. 3.4:

$$\Pi_{ij}(r) = \frac{\sum_I \sum_{J>I} \sum_{i\in I} \sum_{j\in J} |\boldsymbol{t}_i \cdot \boldsymbol{t}_j| \, \delta\left(|\boldsymbol{r}_i - \boldsymbol{r}_j| - r\right)}{\sum_I \sum_{J>I} \sum_{i\in I} \sum_{j\in J} \delta\left(|\boldsymbol{r}_i - \boldsymbol{r}_j| - r\right)}. \tag{3.4}$$

As before, i and j denote the beads of polymer I and J. Hence, the NCF describes the decay of orientation correlations with increasing distance between particles r. We notice that the upper and lower limit of eq. 3.4 are given as a step function with height one for parallel aligned and one half for randomly oriented chains [174]. To estimate the NCF, we extended the InterRDF command by multiplying the dot product of the unit tangent vectors with the probability of finding bead i and j at a distance $r = |\boldsymbol{r}_i - \boldsymbol{r}_j|$.

Furthermore, we extended the NCF (eq. 3.4) to quantify the nematic alignment under shear-flow, for example, shearing the xz-plane of an entangled polymer network with constant velocity $v_{D,x}$. The corresponding NCF $\Pi_{ij}(y)$ projects the orientational correlations between pairs of particles onto the axis orthogonal to the flow, i.e., the y-dimension:

$$\Pi_{ij}(y) = \frac{\sum_I \sum_{J>I} \sum_{i\in I} \sum_{j\in J} |\boldsymbol{t}_i \cdot \boldsymbol{t}_j| \, \delta\left(\boldsymbol{y}_i - y\right) \delta\left(\boldsymbol{y}_j - y\right)}{\sum_I \sum_{J>I} \sum_{i\in I} \sum_{j\in J} \delta\left(\boldsymbol{y}_i - y\right) \delta\left(\boldsymbol{y}_j - y\right)}, \tag{3.5}$$

where $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ is the y-position of particle i and j, respectively. More precisely, we binned the particles' positions in flow direction (e.g. $\boldsymbol{y}_i$) to calculate the pairwise dot products between all pairs of particles within each bin to derive the NCF profile. Lastly, we notice that the upper and lower limit of the NCF profile can be derived by bin-wise integration of eq. 3.4 along r and is given by 1 and 0.75, respectively.

3.3 Results

The development of the Martini 3 CG model included the mapping of the underlying AA structure to CG resolution, the identification of bead types for the non-bonded interactions and the parameterisation of bonded terms, along with an experimental or AA based validation using an iterative process. Specifically, we validated our parameterisation against experimental measurements from the literature, such as the mechanical bending stiffness, the partitioning free energies, the π-stacking and the bulk density. In this section, a summarizing overview about the properties used for validation purposes (Table 3.1) and the steps performed for the force field parameterisation (Fig. 3.1) are provided, taken from [133].

Table 3.1. Validation and tuning of the CG PPE model. To optimize the Martini 3 force field, we selected the partition coefficient, persistence length and solvent accessible surface area (SASA) as single chain properties. For validation, we compared the packing in bundles and mid-size bulk systems to experiments or AA simulations. For the former, we focused on the radial and axial displacement, like the π-stacking, within a polymer bundle, and for the latter on the bulk density and spatial correlations of mid-size polymer networks.

Property	Martini 3	OPLS-AA	Experimental
Properties to tune the model			
Log P_{Oct-H_2O} [-]	4.82 ± 0.02	-	4.78[1]
Persistence length L_P [nm]	14.7 ± 1.3	-	$13.5 - 16.0$[2]
SASA [nm^2]	13.20	13.22	-
Properties to validate the model			
Bulk density [kg m^{-3}]	1098	1092	-
π-stacking [nm]	~ 0.5	~ 0.5	~ 0.4[3,4]
Side-by-side sliding peak [nm]	0	± 0.13	-
Radial distribution function [nm]	0.57	0.58	-

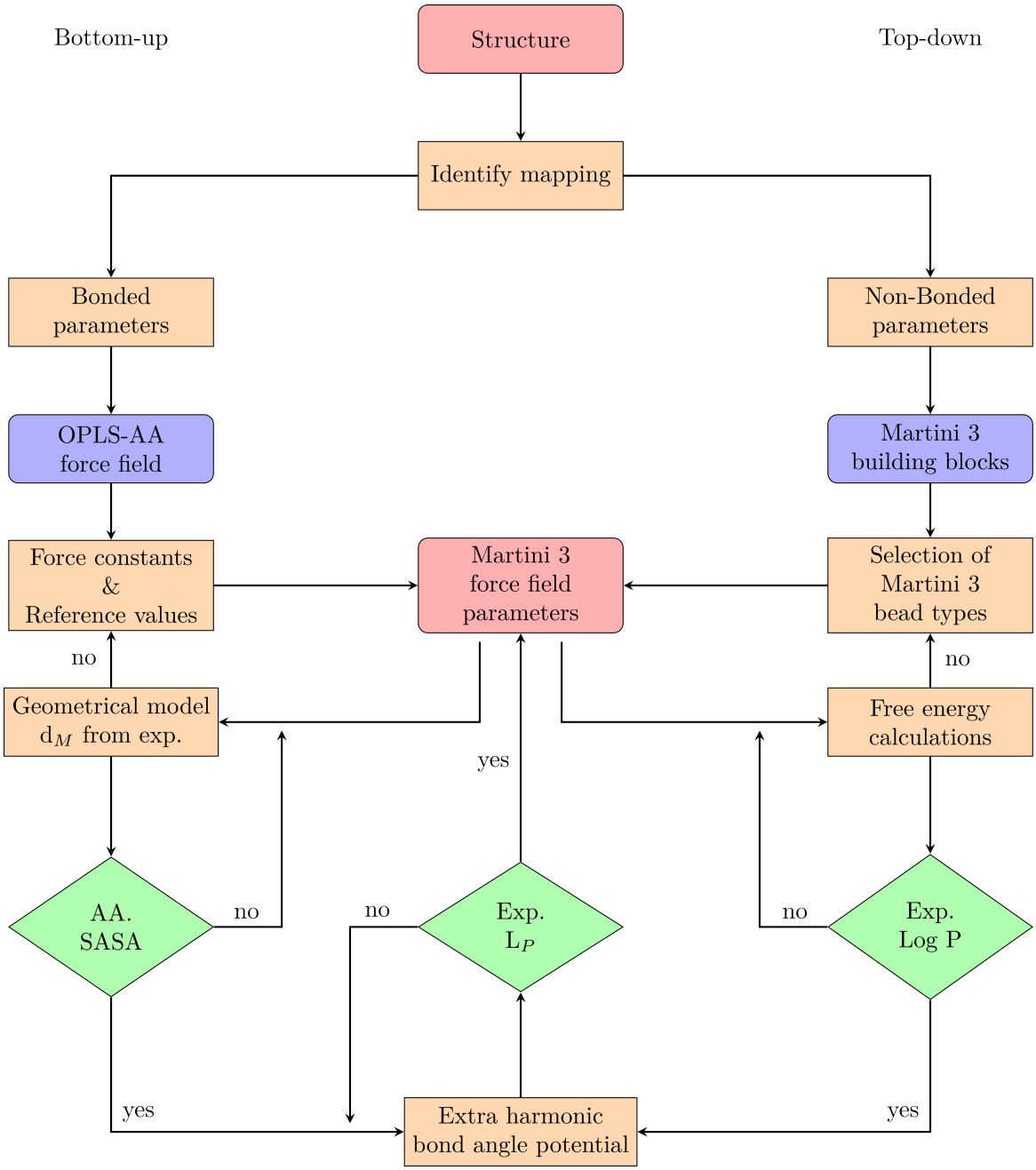

Figure 3.1. Overview for modeling PPEs with Martini 3. Purple boxes provide external information from the OPLS-AA force field or Martini 3 building block data base. Orange boxes are force field parameterisation steps. Green rhombuses represent decision steps for validation or tuning purposes. Bonded terms (*left*) are optimized with a geometrical model, and non-bonded ones (*right*) due to the partitioning behaviour of tolane. As soon as both properties agree, an extra bond angle potential is introduced to tailor the mechanical bending stiffness of a single chain to experiments by matching the persistence length L_P [25, 33].

3.3.1 Mapping Scheme for PPE

Based on the Martini 3 building block principle, we identified the mapping scheme for PPE that properly resembles its chemistry and shape of its π-conjugated backbone [59, 168]. The PPE backbone consists of alternating phenyl rings and triple bonds, linked by single carbon bonds as shown in Fig. 3.2A. Due to the alternation of single and multiple bonds, the π-orbitals of the carbon atoms overlap, and the electrons are delocalized from one end of the backbone to the other. To adequately represent the π-conjugated backbone of PPEs, we split the monomer into two building blocks, using different bead types for each (Fig. 3.2B). More precisely, we substituted the phenyl ring with three apolar *TC5* beads forming a triangle with constrained bond lengths to mimic the rigidity of the aromatic ring. For the triple bond between the phenyl rings, we selected one slightly less polar *TC4* bead.

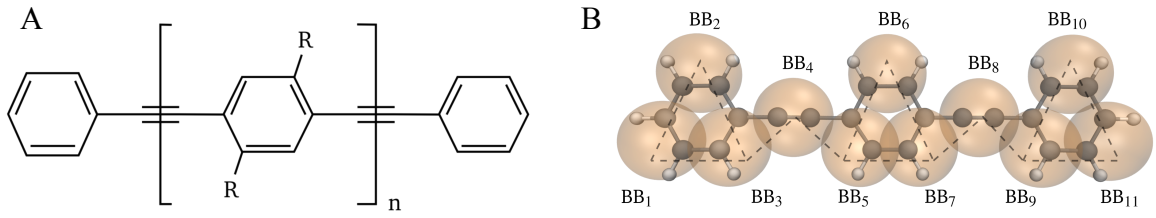

Figure 3.2. Chemical structure of PPEs and mapping. **A** The backbone of Poly(*para*-phenylene ethynylene)s consists of alternating phenylene rings and stiff triple bonds linked by carbon bonds. For side-chains we assume R = H. **B** Mapping from AA (*black*) to CG (*orange*) resolution. The dotted lines represent the bonded potentials.

3.3.2 Parameterisation of the Bonded Potentials for PPE

The parameterisation of the bonded potentials followed a bottom-up strategy based on the identification of bonded terms between Martini 3 beads and the fine-tuning through comparison to AA simulations, considering the molecular shape [59]. This methodology has been previously outlined for modeling non-complex molecular structures with Martini and partly automated for high-throughput applications [161, 175–177]. We started with mapping the AA trajectory to CG resolution by placing each CG bead at the center-of-geometry of the underlying structure (Fig. 3.2A). From the mapped trajectory, we computed the probability densities for the bonded terms and performed CG simulations to determine the exact same distributions. Next, we iterated over the force constants and equilibrium values of the CG force field to match the resultant distributions to the mapped AA ones (Fig. A1).

Bonds, Angles and Torsions from All-Atom model

We started the parameterisation of the PPE backbone with the identification of bond potentials between the CG beads. As shown in Fig. 3.2B, one monomer of PPE consists of three Martini beads forming a triangular structure [BB_1-BB_2-BB_3] and a single bead BB_4 at the left and right edge of each triangle. In particular, we used three bond constraints to model the edges of each triangle [BB_1-BB_2-BB_3]. For the bond between the triangle and the neighboring single beads BB_3 or BB_5 to BB_4, we applied a bond length potential without constraints to provide a certain degree of flexibility along the backbone. To stabilize the linkage between mapped fragments, we introduced two bond angle potentials. The first one, we applied between the base of the triangle and the neighboring ethyne bead [BB_1-BB_3-BB_4], and the second one between each side of the triangle and the adjacent single bead [BB_2-BB_3-BB_4]. Torsion potentials enable us to keep the aromatic rings along the backbone planar, but also might lead to smaller time steps due to numerical instability. Thus, we reduced the amount of dihedral angles to one improper dihedral potential between each triangle and its two adjacent single beads [BB_4-BB_2-BB_1-BB_3] or [BB_4-BB_6-BB_7-BB_5] to prevent the out-of-plane bending and numerical instabilities. The resulting Martini 3-based parameters therefore allow for stable simulations with at least a 20 fs timestep. However, due to the planar structure of the monomers, we underestimated the volume of the aromatic rings and thereby overestimated the PPE bulk density compared to the AA one by approximately 5% (Fig. 3.3C). Accordingly, our method required further optimization to consider the shape and volume of the π-conjugated backbone in more detail.

Geometrical Modeling based Optimization

To reduce the density mismatch and better capture the shape of PPEs, we refined the bond lengths and angles between the CG interaction sites. For this reason, we approximated the polymer backbone as a chain of benzene molecules: According to Martini 3, as a concatenation of regular triangles with side length d_B and spacing d_M (Fig. 3.3A). We assumed that the triple bond bead is positioned in the geometric center of two adjacent phenyl rings at a distance d_{BT}. We set the orthogonal distance between the base of the triangle and the triple bond bead to half the altitude of the regular triangle $h_T = \sqrt{3}/4 \cdot d_B$. According to this, we finally derived the following equations for the unknown bond length d_{BT} and bond angles θ_{An} with $n = (1, 2)$, $\theta_{01} = 180°$ and $\theta_{02} = 120°$:

$$d_{BT} = \sqrt{h_T^2 + \left(\frac{d_M - d_B}{2}\right)^2} \quad \text{, and} \tag{3.6}$$

$$\theta_{An} = \theta_{0n} - \arcsin\left(\frac{h_T}{d_{BT}}\right). \tag{3.7}$$

As shown in Fig. 3.3A, the equations for the equilibrium bond length (eq. 3.6) and bond angles (eq. 3.7) only depend on the side length d_B and spacing d_M of the triangles. To identify each unknown, we set the distance between monomers d_M to 0.7 nm, in agreement with AA simulations (0.7 nm) and experiments (0.69 nm) [33, 158]. Besides, we determined the side length of the regular triangle d_B by matching the solvent accessible surface area (SASA) of PPEs solvated in water from Martini 3 to AA simulations [59, 114]. We retrieved the SASA from fine-grained simulations of one small four-monomer long PPE chain ($A_{SASA} = 13.22\,\text{nm}^2$). By setting the side length of the triangle to $d_B = 0.325$ nm, the SASA with our Martini 3 model was $A_{SASA} = 13.20\,\text{nm}^2$ (Fig. 3.3B), a value which is in good agreement with the AA prediction. The equilibrium values obtained from this optimized geometric modeling approach are summarized in Table 3.2.

Table 3.2. Martini 3 force field parameters for PPEs. Force constants were derived from AA data and equilibrium values from the geometrical model.

Beads selected for bonded terms	Reference value	Force constant
BB_1-BB_2	0.325 nm	constraint
BB_1-BB_3	0.325 nm	constraint
BB_2-BB_3	0.325 nm	constraint
BB_3-BB_4	0.325 nm	$9000\,\text{kJ mol}^{-1}\,\text{nm}^{-2}$
BB_1-BB_3-BB_4	$143\,°$	$550\,\text{kJ mol}^{-1}$
BB_2-BB_3-BB_4	$83\,°$	$650\,\text{kJ mol}^{-1}$
BB_4-BB_8-BB_{12} [a]	$180\,°$	$50\,\text{kJ mol}^{-1}$
BB_1-BB_4-BB_8 [a,b]	$165\,°$	$50\,\text{kJ mol}^{-1}$
BB_4-BB_8-BB_{11} [a,b]	$165\,°$	$50\,\text{kJ mol}^{-1}$
BB_4-BB_2-BB_1-BB_3	$0\,°$	$50\,\text{kJ mol}^{-1}$

[a] Extra harmonic bond angle potential for the π-conjugated backbone. Force constant is set by matching the persistence length from Martini 3 to experiments [25, 33].

[b] Extra harmonic bond angle potential extends over both polymer ends.

On the microscale, we evaluated the solvent-excluded area by comparing the Connolly surface for one four-monomer long PPE between AA (*gray*) and CG (*blue*) resolution (Fig. 3.3E). Accordingly, both molecular surfaces coincide well with slight differences due to the mapping from fine- to coarse-grained resolution, i.e., an aromatic ring to a triangle.

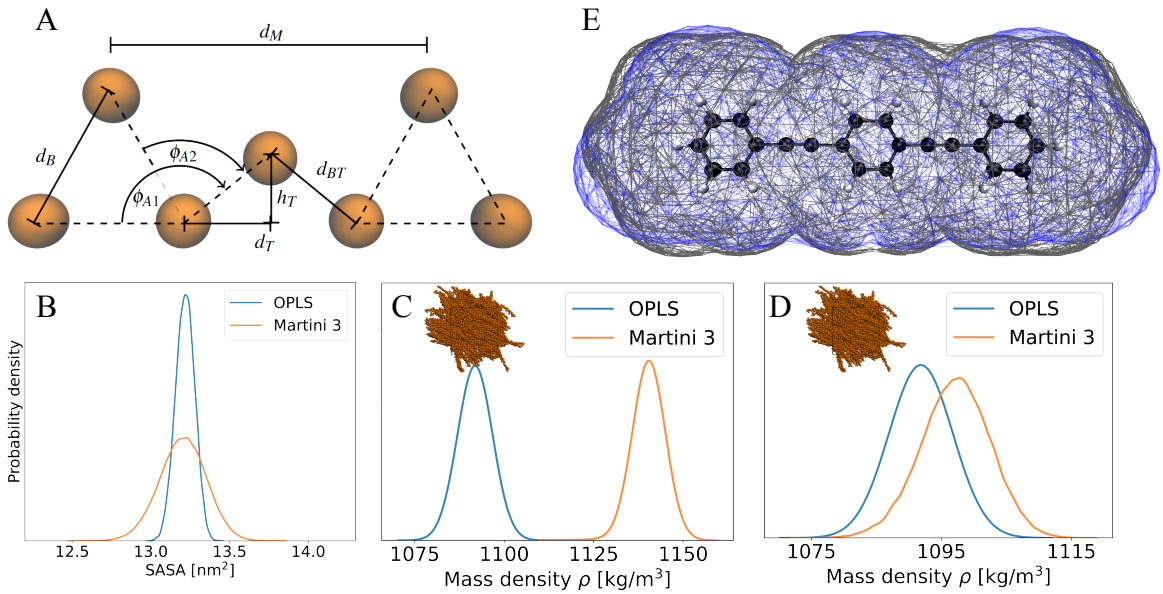

Figure 3.3. Geometrical model to coarse-grain PPEs with Martini 3. **A** The geometrical model takes the shape of the π-conjugated backbone into account. PPEs are represented as a linear chain of regular triangles with side length d_B and spacing d_M. The remaining equilibrium values are calculated by trigonometrical relations (eq. 3.6- 3.8). **B** We tuned the geometrical model, in particular the side length of the regular triangles d_{BT}, through matching the solvent accessible surface area (SASA) of a single four-monomer long PPE chain from Martini 3 to AA simulations. **C** Mass density for a PPE bulk system containing 300 chains with length of four monomers each. Mass density calculated from the center-of-geometry based mapping scheme. **D** Mass density obtained from geometrical modeling reduces the bulk density difference between AA and Martini 3 simulations to less than 1%. **E** The Connolly surface indicates minor differences between AA (*gray*) to CG (*blue*) resolution due to mapping an aromatic ring with six heavy atoms to a triangle.

Modeling the π-conjugated Backbone of PPE

We also focused on modeling the π-conjugated backbone of PPEs, which spans from one end of the polymer to the other, and is the reason for its planar shape, backbone linearity and increased mechanical bending stiffness. We introduced an additional harmonic bond angle potential, which extends over three neighboring *TC4* beads [BB_4-BB_8-BB_{12}], and has an equilibrium value of $180°$ to keep the backbone linear. Besides, we also applied a harmonic bond angle potential at both ends of the polymer chain to prevent the last monomer at both ends from bending over. The bond angle potential at each end of the chain included the outermost bead BB_1 or BB_{11} and its two closest triple bond beads [BB_4-BB_8] (Fig. 3.2B).

To this end, we derived an expression for the reference bond angle describing the π-conjugation at both ends of the polymer by using the geometric modeling approach:

$$\theta_{\pi} = \theta_{0\pi} - \arctan\left(\frac{h_T}{d_B + d_T}\right), \tag{3.8}$$

where θ_{π} is the equilibrium angle at both ends, and $\theta_{0\pi} = \theta_{01}$ and $d_B + d_T$ is the distance between the outermost bead and closest triple bond bead [BB_1-BB_4] or [BB_8-BB_{11}]. The newly introduced harmonic bond angle potential reduces the bending motion of the backbone as shown by the comparison of the probability densities of three kinds of dihedral angles between the CG and mapped AA simulations (Fig. A1-A2). We defined the first torsion angle between two aromatic rings by connecting the bases of adjacent triangles [BB_1-BB_3-BB_5-BB_7], and the second one by linking the base of the triangle to the two neighbouring triple bond beads [BB_4-BB_5-BB_7-BB_8]. We further specified a dihedral angle from the apex of the triangle, over the triple bond bead to the base of the adjacent triangle [BB_1-BB_3-BB_4-BB_6] or [BB_2-BB_4-BB_5-BB_7]. We observed good agreement between the dihedral angle probability densities obtained from the mapped AA and CG trajectory (Fig. A2). In particular, we noticed that the π-conjugation potential prevents the sampling of unwanted dihedral configurations and keeps the aromatic rings along the backbone planar.

3.3.3 Partitioning Free Energy for Tolane

The chosen Martini 3 bead types, namely *TC5* for the rings and *TC4* for the triple bonds, determine the non-bonded interactions. We evaluated the non-bonded interactions by comparing the solubility of such-chosen CG beads obtained in simulations with that from experiments. We estimated the partitioning free energy in a biphasic octanol-/water system to compute the distribution of solute between the hydrophilic and hydrophobic solvent (Fig. 3.4). Due to the absence of experimental data regarding the solubility of PPEs, we selected tolane, short for di(*para*-phenylene ethynylene), as a reference for validation purposes. We computed the free energy of transferring tolane from water towards octanol and validated the partition coefficient with experiments reported in the literature [146, 147].

For this reason, we calculated the PMF, using the WHAM, and the resultant partitioning free energy for tolane $\Delta\Delta G_{Oct-H_2O} = 27.69\,\mathrm{kJ\,mol^{-1}}$ in the biphasic octanol-/water mixture. Based on a cumulative simulation of $6\,\mu$s per umbrella, we performed Bayesian bootstrapping to obtain an absolute error below $0.10\,\mathrm{kJ\,mol^{-1}}$, and backward block averaging to show convergence within $200\,$ns (Fig. 3.4B) [164].

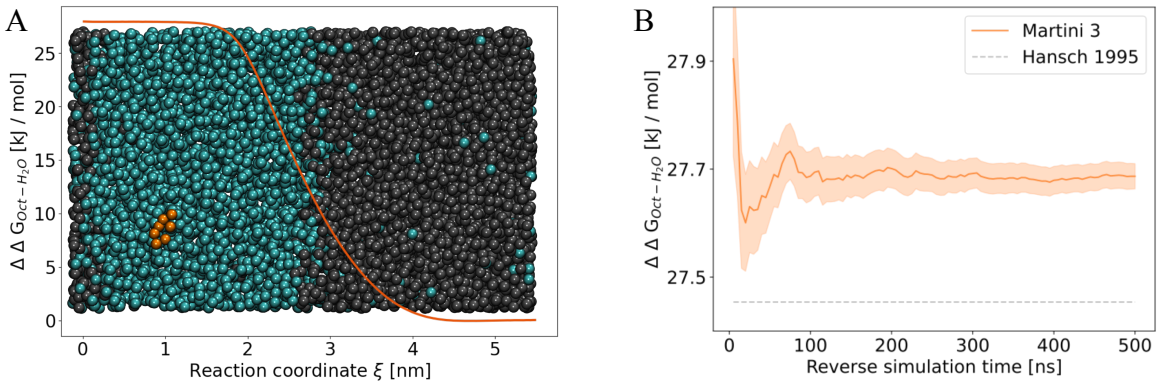

Figure 3.4. Convergence analysis for the partitioning free energy for tolane A The free energy of transferring tolane (*orange*) from water (*blue*) to octanol (*gray*) was estimated by umbrella sampling MD simulations. The reaction coordinate ξ points from the center of the hydrophilic to the hydrophic solvent, and is split into equally spaced umbrellas. **B** The backward block average and the absolute error of tolane's free energy of transfer in an octanol-/water system. For the convergence analysis, we performed backward block averaging with a 5 ns step size and Bayesian bootstrapping to estimate the associated error. We obtained a partitioning free energy $\Delta\Delta G_{Oct\text{-}H_2O} = 27.69\,\mathrm{kJ\,mol^{-1}}$ with an error below $0.10\,\mathrm{kJ\,mol^{-1}}$, i.e., a partition coefficient of $\log P_{Oct\text{-}H_2O} = 4.82\pm0.02$, less than $0.25\,\mathrm{kJ\,mol^{-1}}$ above experiments reported in the literature [147].

For the partition coefficient of tolane $\log P_{Oct\text{-}H_2O} = 4.82 \pm 0.02$, calculated with eq. 2.23, we observe good agreement with experiments from the literature (4.78) [147]. Further, we suggest that similar chemical moieties, such as bibenzyl or trans-stilbene, could be equally considered with our Martini 3 model due to a comparable partition coefficient of 4.70 ± 0.20 and 4.81 ± 0.40, respectively [146]. Table 3.3 provides experimentally measured partition coefficients for tolane, bibenzyl and trans-stilbene from the literature [146, 147].

Table 3.3. Partition coefficients from experiments and Martini 3. Partition coefficient for tolane, bibenzyl and trans-stilbene in a biphasic octanol-/water system [146, 147]

Chemical substance	Experiments	Martini 3
Tolane[147]	4.78	4.82 ± 0.02
Bibenzyl[146]	4.70 ± 0.20	
Trans-stilbene[146]	4.81 ± 0.40	

3.3.4 Mechanical Bending Stiffness for PPE

We further refined our CG model by fine-tuning the mechanical bending stiffness of a single chain such that the experimental persistence length is reproduced. The π-conjugated backbone of PPEs is known to induce an increased bending stiffness for a single chain which classifies PPEs as semi-flexible polymers. To estimate the persistence length L_P for the PPE backbone, we performed simulations of a single chain in solution, and fitted both the unit tangent vector auto-correlation between pairwise segments (eq. 2.29) and the squared end-to-end distance (eq. 2.30) to the Martini 3 trajectory, as two independent measures for L_P. We simulated single PPEs of various chain lengths ranging from 20 to 100 monomers with an increment of 10 in water and toluene. For each chain length, we performed 10 replicas to sample the free energy landscape adequately and obtain a significant estimate for the persistence length. To model the π-conjugated backbone and tune the chain stiffness, we introduced an extra harmonic angle potential between neighboring *TC4* beads and tuned the force constant to match experimental persistence lengths (see section 3.3.2) [25, 33].

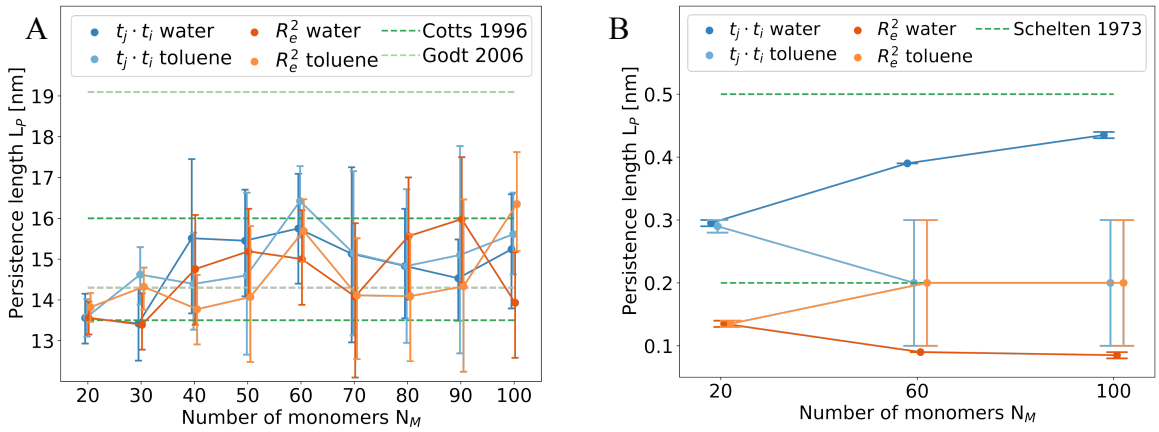

Figure 3.5. Persistence length of PPE and PS from Martini 3. Average and standard error for the persistence length of a single PPE (**A**) and Polystyrene (PS) (**B**) chain as function of the polymerization degree was computed in water and toluene, and estimated with the unit tangent vector auto-correlation $\vec{t}_j \cdot \vec{t}_i$ (*blue*) and squared end-to-end distance R_e^2 (*orange*). Experimental estimates for PPEs by Cotts, using light scattering experiments (L_P = 13.5-16 nm) [25], and by Godt, performing electron paramagnetic resonance spectroscopy (L_P = 14.3-19.1 nm) [33], as well as for PS by Schelten, using low-angle neutron scattering experiments (L_P = 0.5-0.6 nm) [178], and by Pital, performing atomic force microscopy (L_P = 0.2 nm) [179], are shown in *green*. Accordingly, semi-flexible PPEs are 30 to 50 times stiffer than the flexible PS, both in experiments and Martini 3.

Fig. 3.5A shows the persistence length for a single PPE chain for chains of size ranging from 20 to 100 monomers, obtained when choosing a spring constant of $50\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ for the additional bond angle potential of the CG model. The persistence length obtained from the unit tangent vector auto-correlation ($14.8\,\mathrm{nm}$) and from the squared end-to-end distance ($14.6\,\mathrm{nm}$) were comparable. A detailed summary of the persistence lengths for each replica is given in Fig. A3-A4 for water and in Fig. A5-A6 for toluene. The persistence lengths are largely independent of the polymer length and solely depend on the chemical nature of the monomer [33]. The covered range, as well as the obtained average from both estimation methods and solvents of $14.7\,\mathrm{nm}$ (with a standard error of $1.3\,\mathrm{nm}$) is in very good agreement with the experimental values reported from Cotts et al. ($13.5\,\mathrm{nm}$-$16\,\mathrm{nm}$).

To emphasize the distinct nature of PPEs, we performed single polystyrene (PS) in solution simulations for chains consisting of 20, 60 and 100 monomers with the given Martini 3 force field parameter set [55] (Fig. 3.5B). We calculated the unit tangent vector auto-correlation ($0.37\,\mathrm{nm}$) and squared end-to-end distance ($0.11\,\mathrm{nm}$) to derive the persistence length of this flexible polymer. By comparing the persistence lengths of PS and PPE, we observed that semi-flexible PPEs are approximately 30 to 50 fold stiffer than flexible PS, which is the reason for the outstanding mechanical properties of PPEs.

3.3.5 Packing within a Bundle of PPEs

We next aimed to evaluate the packing properties of systems containing multiple parallel aligned PPEs. For validating the Martini 3 parameterisation presented here, we analyzed the local structural organization of the semi-flexible chains, and compared it to AA-MD simulations and experiments from the literature [180, 181]. Finally, we constructed and equilibrated a finite bundle of short-chain PPE (10 chains, each 4 monomers in length), and monitored two major degrees of freedom, namely the radial arrangement of aligned polymers, and their axial displacement within the assembled bundle. These observables jointly tested both the bonded and non-bonded interactions between CG beads, i.e., defined in the many-body potential of the Martini 3 force field $V^{CG}(\boldsymbol{r}_1, ..., \boldsymbol{r}_N)$.

First, we quantified the radial displacement between PPE chains by computing the orthogonal distances between parallel aligned chains (eq. 3.1). We determined the probability distribution for the π-stacking (Fig. 3.6A). The π-stacking obtained from the AA and CG trajectories exhibits a similar behaviour. In particular, we monitored three distinct peaks of the distribution, at $\sim 0.5\,\mathrm{nm}$, 0.8-$0.9\,\mathrm{nm}$ and 1.2-$1.3\,\mathrm{nm}$. Both the AA and CG simulations show good agreement with the experimentally observed interchain distances [180, 181].

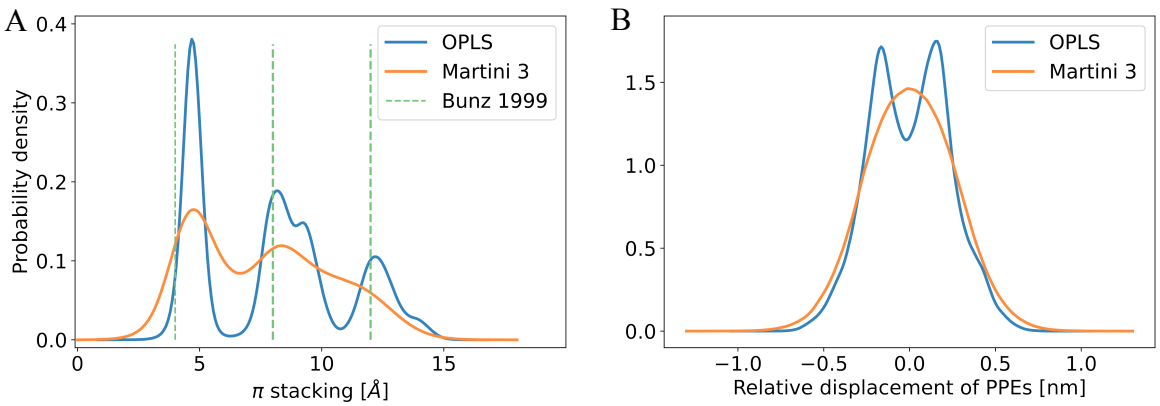

Figure 3.6. Packing within a bundle of aligned PPEs. **A** Radial orientation of aligned chains in a bundle of PPEs estimated with the π-stacking distance. Experimental measurements reported in literature used for comparison [180, 181]. **B** Axial displacement or side-by-side sliding of single chains within the polymer bundle. Both packing properties are validated by comparing the probability distributions with reference to OPLS-AA data [70].

However, both models slightly overestimated the π-stacking distance with reference to the experimental value (~ 0.4 nm) by 15% for the first and 5% for the second and third peak. In contrast to the AA case, for Martini 3, the probability distribution did not drop to zero between the peaks, the second and third peaks were hardly separated, and finer structural features seen in the AA model were lost. This reflects an overall less ordered packing of the polymer chains, which is expected, given the loss of resolution upon coarse-graining. To resolve the second and third peak in more detail, a finer resolution, e.g., mapping a ring of six heavy atoms to a rhombus of four beads, would be suitable at the expense of performance. Still, the CG simulations capture the major features of radial packing of chains within a bundle of semi-flexible polymer and, thus, yield the expected local structure of PPE chains.

Second, we characterized the axial degree of freedom by calculating the horizontal displacement of each polymer relative to the center of the bundle. Finally, we projected the COM of each polymer onto the main COM axis of the bundle (eq. 3.2). We compared the probability distributions from both AA- and CG-MD simulations (Fig. 3.6B). Both the CG and AA models show distributions around zero displacement with highly similar width. They both strongly disfavor chains to displace relative to each other by more than 0.5 nm. Again, the CG model fails to reproduce a finer subnanometer-scale structure of the PPEs, i.e., a bimodal distribution with peaks at ± 0.13 nm. Aromatic rings slightly prefer an axial shift relative to one another by this length, a low-energy π-stacking mode also known for benzene that the AA force field is able to capture [111, 182–184]. In contrast, the probability density from Martini 3 is unimodal and centered around mode zero.

This deviation is due to the flattened free energy surface resulting from mapping the AA structure to CG resolution, that is, a ring of six heavy atoms to a triangle of three beads. Yet, on the coarser scale beyond 1-2 Å, the CG data closely follows the AA data. Within the expected level of resolution of Martini 3, we conclude that our CG model accurately represents both the axial and radial degree of freedom within a bundle of aligned PPEs.

3.3.6 Structural Organization in PPE Bulk Systems

We next moved from finite PPE systems to a mid-size bulk system with full periodicity of the polymer material and absence of a solvent. The system size was chosen such that simulations at the atomistic scale are still feasible, and comprised 300 chains of 4 monomers each. To describe the structural organization within such a PPE network, we focused on properties at both the macroscopic and microscopic length scales, namely global packing by measuring the densities and local packing by short-range structural parameters. First, we evaluated the bulk densities from CG with AA simulations and obtained a bulk density of $\rho_{AA} = 1092\,\mathrm{kg\,m^{-3}}$ and $\rho_{CG} = 1098\,\mathrm{kg\,m^{-3}}$ for the AA and CG models, respectively (Fig. 3.3D). This agreement is satisfying and suggests that the CG model largely reproduces the inter-PPE distances. It is worth noting that experimental densities for solid-state PPEs, between $997\,\mathrm{kg\,m^{-3}}$ and $1118\,\mathrm{kg\,m^{-3}}$, obtained from X-ray powder diffraction, depend strongly on the side-chain concentration, and are not suited for validation [69, 185].

On the microscopic scale, we characterized the local organization within the bulk system by computing the radial distribution function (RDF) for the previously mentioned PPE network. Figure 3.7 shows a snapshot of the PPE bulk obtained from an AA system (*left*) and from a backmapped CG one (*right*,A) with their respective RDFs (B-D). For direct comparison, we mapped the AA and CG trajectory to a one-bead-per aromatic ring resolution (B), a one-bead-per triple bond resolution (C) as well as a one-bead-per aromatic ring and triple bond resolution (D). Next, we computed the RDF for the mapped trajectories to analyze the packing of 300 PPEs. Due to the symmetry along the backbone, the RDF between the triple bonds (B) and the one between the aromatic and the ethynylene groups (C) are very similar. Specifically, the RDF of the one-bead-per aromatic ring mapped AA trajectory exhibits a first peak around $0.57\,\mathrm{nm}$, a minimum at $0.73\,\mathrm{nm}$, and a second and third peak at $0.92\,\mathrm{nm}$ and $1.13\,\mathrm{nm}$, respectively. Hence, the PPE assembly, obtained from AA simulations, possesses a high degree of local order as shown in the *right* panel of Fig. 3.7A. The RDF computed from the same mapped Martini 3 trajectory exhibits a first peak at $0.58\,\mathrm{nm}$, in close agreement with the AA force field. Its second maximum at $0.74\,\mathrm{nm}$ is broader and covers the second and third peaks of the RDF from the mapped AA trajectory.

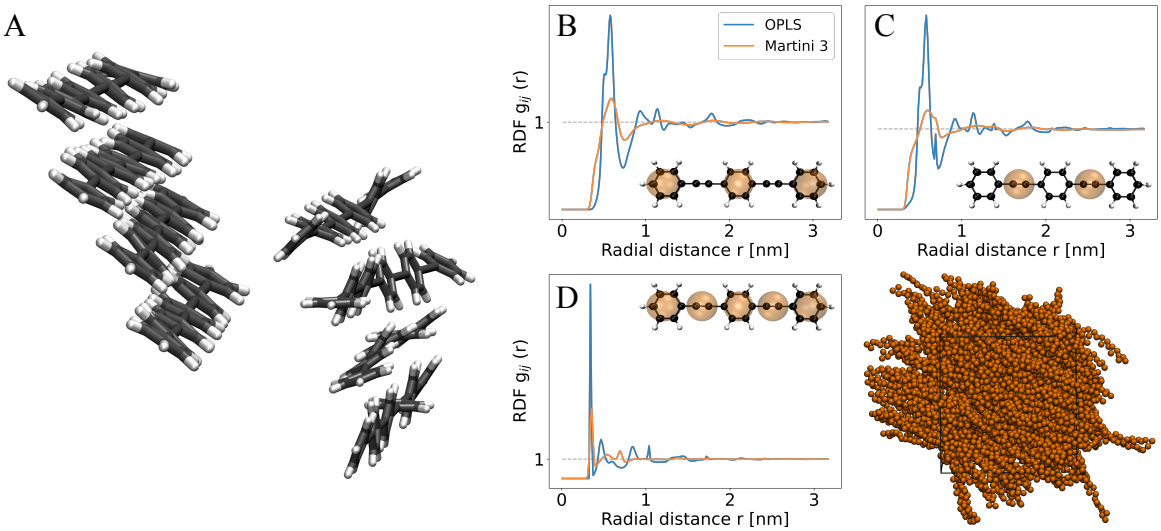

Figure 3.7. Packing in mid-size bulk systems of PPEs. A Snapshot of the AA trajectory indicates a high degree of local order (*left*). Backmapping from coarse- to fine-grained resolution reveals a decrease in order (*right*). **B-D** Comparing the RDF from mapped AA and CG simulations using a one-bead-per aromatic ring (**B**), one-bear-per triple bond (**C**) and one-bead-per aromatic ring and triple bond resolution (**D**). Independent of the mapping scheme, Martini 3 simulations exhibit a less distinct pattern with lower order compared to mapped AA ones. However, local ordering of the nearest neighboring chains, given by the RDF's first two maxima, is captured by Martini 3.

Thus, the Martini 3 model broadly shows a similar packing geometry also at such larger distances. As expected and as also seen for finite bundles (Fig. 3.6), it is evident that the PPE assembly obtained on the CG scale is less locally ordered (Fig. 3.7A, *left*). For this reason, the CG model for PPEs did not resolve features beyond the second peak, which is independent on the applied mapping scheme. Still, overall, the packing of the first and second neighbors, i.e., the first and second peaks in the RDF, is captured by Martini 3.

3.3.7 Nematic Alignment in large Bulk Systems

The Martin 3 model now opens the route towards analyzing PPE assemblies at a larger length and time scale. We set out to construct simulation systems of 300 to 800 chains, with 20 to 60 monomers, resulting in box sizes of $10 \times 10 \times 10\,\mathrm{nm}^3$ to $20 \times 20 \times 20\,\mathrm{nm}^3$. In the largest bulk system mentioned here, around $200,000$ CG beads represent roughly $600,000$ atoms, and have been simulated for about $5\,\mu$s. It is important to note that these spatial and temporal scales are already difficult to assess at all-atom level of resolution.

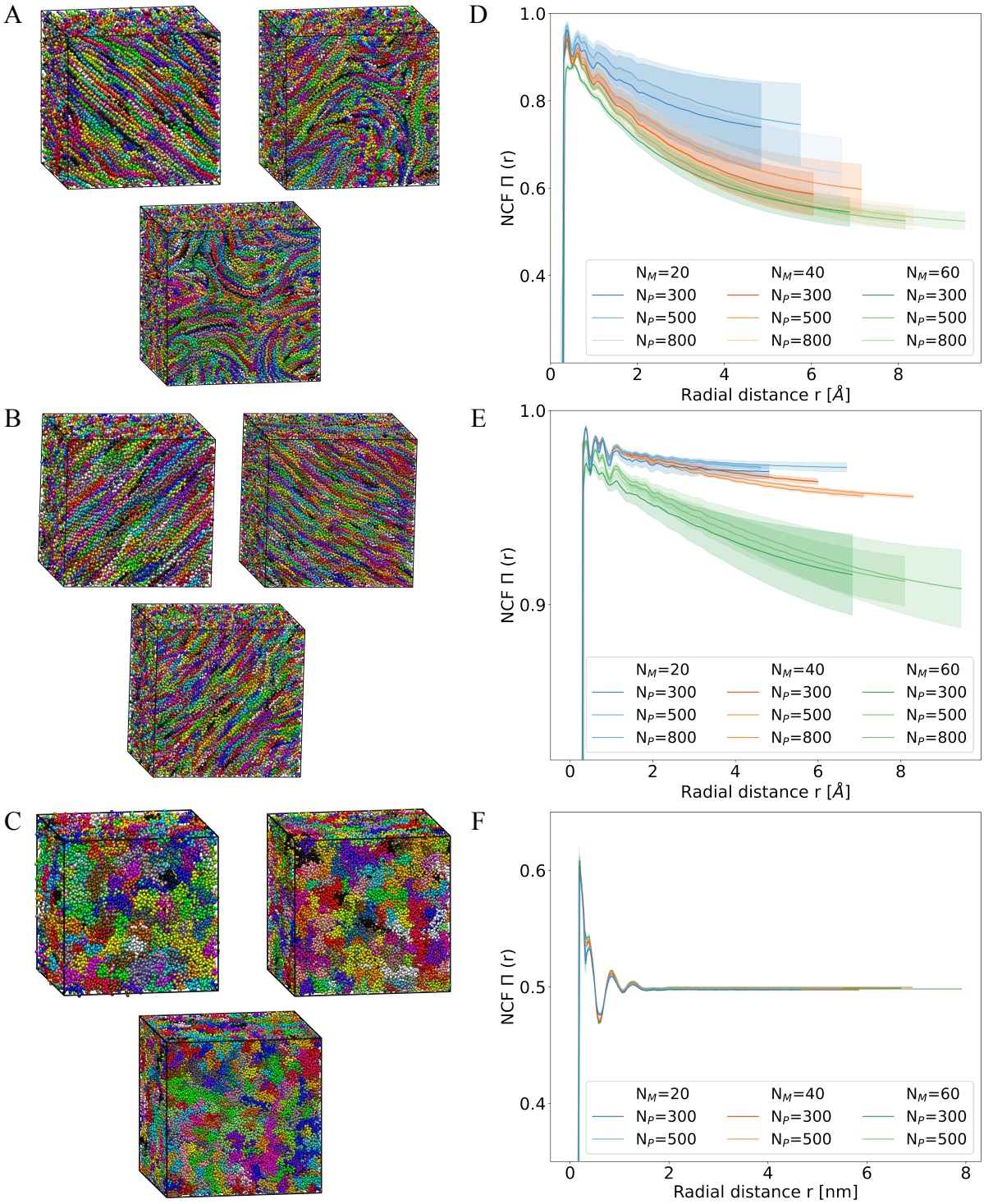

Figure 3.8. Alignment in large-scale PPE bulk systems. A-C Snapshots for bulks of 500 PPEs before **A** and after **B** annealing or 500 PS **C** with lengths of 20, 40 or 60 monomers. **D-F** Comparing the NCF for each bulk system with various number of monomers N_M and polymers N_P. Global ordering of PPEs decreases with increasing polymer chain length.

Fig. 3.8A shows snapshots for a bulk containing 500 PPEs of 20, 40 and 60 monomers each. It is evident that short-chains mostly align in parallel, while this alignment is lost across larger distances when the chain length grows. To quantify this nanometer-scale chain-alignment, we calculated the nematic correlation function (NCF). The NCF describes the decay in structural order with increasing radial distance from a reference monomer (eq. 3.4). The lower limit of the NCF (with 0.5) defines randomly oriented polymers, and the upper limit (with 1.0) parallel oriented ones. Fig. 3.8D shows the NCF analysis for various chain lengths ranging from 20 to 60 monomers. We computed the average and standard deviation for the NCF from 8 replicas. The averaged NCF increases sharply at the beginning and decreases with increasing distance until the curve asymptotically approaches the lower limit of 0.5. As reflected by the standard deviations, the variations in structural order reached after independent assembly simulations varied strongly for a given chain length and system size, as the ordering extends across length scales comparable to the system sizes.

Notwithstanding these fluctuations, the extensive sampling aided by the efficient Martini model allowed us to reveal significant differences when modifying the PPE chain length. Most importantly, we observe a steady decrease in ordering with increasing polymer chain length, across the whole range of radial distances. PPEs with a chain length in the range of their persistence length (20 monomers) show a higher degree of alignment in comparison to longer-chain PPEs (40 or 60 monomers). The latter are less parallel and rather assemble through entanglement as opposed to alignment. In this case, parallel alignment is maintained on a length scale of up to $\sim$ 5 nm, beyond which a nano-domain forms comprising PPE chains aligned along a different direction. The decrease in alignment with increase in chain length can be attributed to the competition between entropic and enthalpic effects. The influence of the entropic effects on the polymer dynamics increases with the polymer length, thus long-chain PPEs exhibit a random coil-like behaviour with less ordering.

To exclude effects from the limited system size, we also examined the influence of the number of polymer chains on the NCF (Fig. 3.8D, lighter colors). We compared networks consisting of 300, 500 and 800 chains (*e.g.* system sizes of $12.1 \times 12.1 \times 12.1\,\text{nm}^3$, $14.3 \times 14.3 \times 14.3\,\text{nm}^3$ and $16.9 \times 16.9 \times 16.9\,\text{nm}^3$ for a 40 monomers long chain), and found that the alignment of PPEs within a bulk system is constant with increasing number of polymers, and hence independent of the box size. Starting from the equilibrated bulk structures, we confirmed our observation regarding the decrease in order with increasing chain length and distance through annealing simulations of PPE networks (Fig. 3.8B,E). Here, we observe an overall very high degree of order across different chain lengths, and independent of the box size. To put our observations for semi-flexible PPEs into perspective, we constructed

bulk systems of flexible PS containing 300 and 500 chains with 20, 40 and 60 monomers each to characterize their chain alignment with the NCF (Fig. 3.8C,F). Specifically, PS bulk systems show an overall very low degree of structural order with a NCF below 0.6, this indicates random oriented chains with no preferred direction of alignment, in contrast to semi-flexible PPEs, showing chain length dependent ordering.

Taken together, these results show a decay in nematic alignment of PPE chains on the nanometer length scale, resulting in the formation of nano-crystalline domains. The loss of interchain alignment is more pronounced for long-chain PPEs which form nano-domains on the $\sim 5\,\text{nm}$ scale, independent from the system size. Longer chains show more chain fluctuations and thus entanglement, which prevents the preferred interchain alignment observed for PPEs with shorter chain lengths, i.e., lengths in the range of their persistence length.

3.3.8 Nematic Alignment in large Bulk Systems under Shear

We extended our CG-MD simulations to non-equilibrium conditions by performing shearing simulations of large bulk systems containing thousands of PPEs, consisting of 20, 60 or 120 monomers each, to monitor the shear-induced structural changes, to analyze the force concentration within the network and to predict bond rupture events. For each system, we ensured that the characteristic box length is longer than the contour of the polymer $L_C > L_0$ to exclude periodic boundary effects. Hence, the largest bulk system probed here comprised around 12 million CG beads ($25,000$ PPEs with 120 monomers each), i.e., far beyond the capabilities of AA simulations. Note that shearing simulations were performed with the GROMACS 2023 developer version fc221d406a[1] in the NVT ensemble, since the Lee-Edwards boundary condition (see section 2.1.6) was not yet implemented for NpT conditions. Thus, we recommend to interpret the following results with caution, although, shearing simulations from the literature have been performed under NVT conditions before [186, 187].

We constructed PPE bulk systems composed of chains with 20, 60 and 120 monomers and applied shear rates of $\dot{\gamma} = 0.01\,\text{ns}^{-1}$, $\dot{\gamma} = 0.1\,\text{ns}^{-1}$ and $\dot{\gamma} = 1.0\,\text{ns}^{-1}$ to drive the system out-of-equilibrium. Fig. 3.9A shows snapshots of the sheared polymer networks composed of $1,500$; $11,000$ or $25,000$ chains consisting of 20, 60 or 120 monomers, respectively, at a shear rate of $\dot{\gamma} = 1.0\,\text{ns}^{-1}$. It is obvious that, at such a high shear rate, short chains (20 monomers) mostly align parallel to the imposed shear flow, while this alignment is partly lost for networks of intermediate (60 monomers) or long (120 monomers) semi-flexible polymers, instead, suggesting the formation of shear bands parallel to the flow direction.

[1] Issue #4607 from GitLab repository of GROMACS 2023 processed by Berk Hess (06.10.2022) [82].

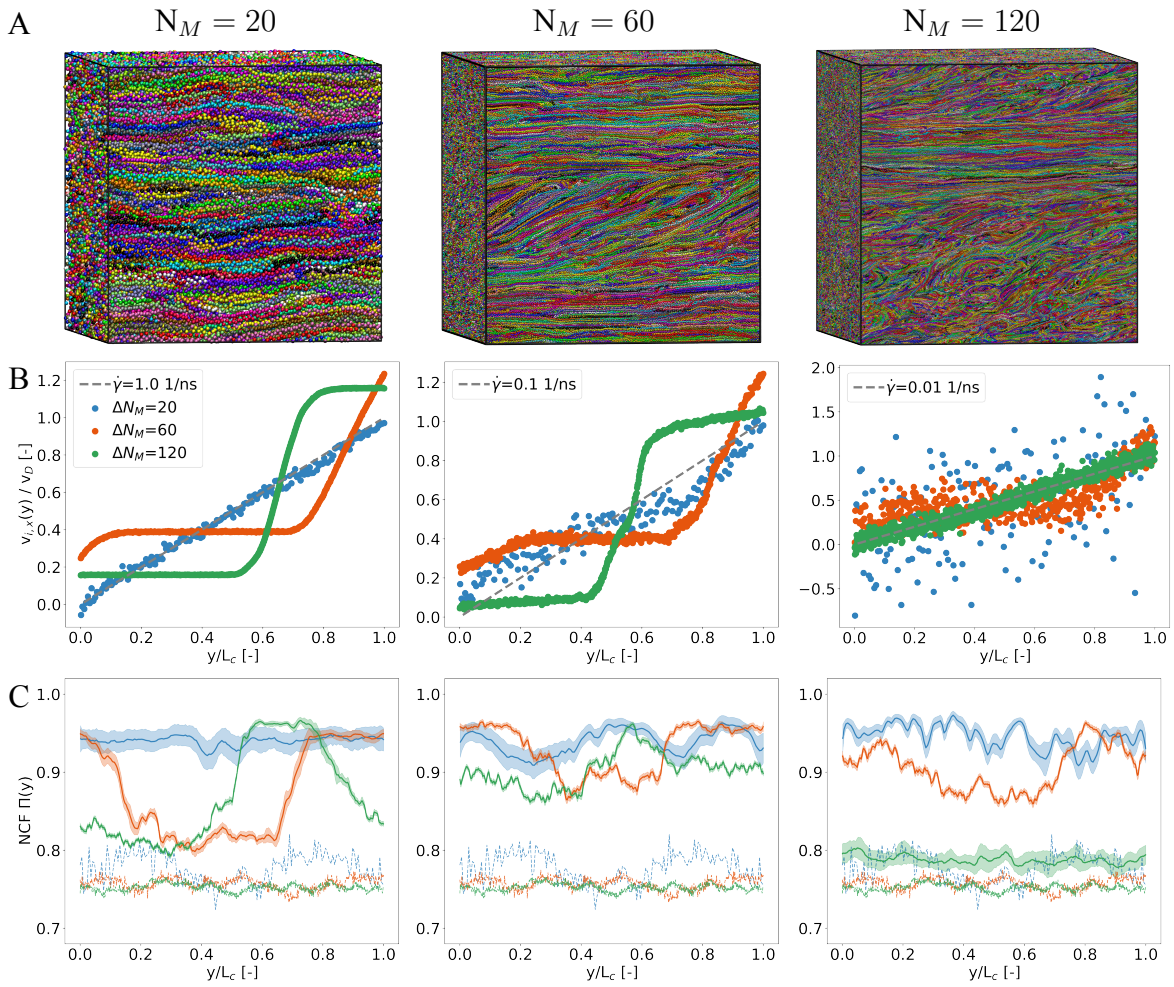

Figure 3.9. Large-scale PPE bulk systems under shear-flow. **A** Snapshots of bulk systems composed of $1,500$; $11,000$ or $25,000$ PPEs at an imposed shear rate of $\dot{\gamma} = 1.0\,\mathrm{ns}^{-1}$ (*left* to *right*). **B** Particles' velocity profile $v_x(y)$ for networks composed of PPEs of length 20 (*blue*), 60 (*orange*) and 120 (*green*) monomers at shear rates of $\dot{\gamma} = 1.0\,\mathrm{ns}^{-1}$, $\dot{\gamma} = 0.1\,\mathrm{ns}^{-1}$ and $\dot{\gamma} = 0.01\,\mathrm{ns}^{-1}$ (*left* to *right*). **C** NCF profile for each bulk system at the starting configuration (*dashed*) and under shear (*solid*). Nematic alignment increases with shear rate and drops sharply at the interface between shear bands.

To show the formation of shear banding, we calculated the velocity profile of each bulk system by binning the x-component of the instantaneous particle velocities $v_{i,x}$ orthogonal to the shear in y-direction $v_{i,x}(y)$. We further normalized $v_{i,x}$ to the imposed deformation velocity $v_{D,x}$ to obtain a non-dimensional velocity profile, and thereby enable the comparison across different shear rates. Of note, a velocity profile with a linear shape represents a homogeneous deformation with a constant shear rate, however, a non-linear velocity profile suggests the formation of shear bands with different shear rates [188, 189].

Fig. 3.9B presents the velocity profiles for bulk systems composed of 20 (*blue*), 60 (*orange*) and 120 (*green*) monomers-long PPEs and the analytical solution $v_x(y) = \dot{\gamma} \cdot y$ (*gray*) according to eq. 2.20 (see section 2.1.6). We observed a tendency for networks of short- and long-chain PPEs to develop a linear velocity profile at high and low shear rates, respectively. In contrast, networks comprising PPEs beyond 20 monomers reveal non-linear velocity profiles at an intermediate and high imposed shear rate combined with the formation of shear bands, resulting in extreme shear rates in the fast band. Most importantly, shear rates of bulk systems composed of intermediate- (60 monomers) and long-chain PPEs (120 monomers) are two to three and three to four times higher, respectively, than the imposed one. As reflected by the spread of the velocity profile, small scale motions of particles due to thermal fluctuations are mostly seen at medium and low imposed shear rates across all chain lengths, although, this effect decreases with increasing chain length of PPEs.

In order to combine the shear-induced strand alignment with the non-linear velocity profiles of the bulk systems, we extended the NCF towards shearing simulations by projecting the particles' velocities onto the axis orthogonal to the shear-flow (eq. 3.5), such that a NCF profile with an upper limit for parallel chains (with 1.0) and a lower limit for random oriented ones (with 0.75) is obtained (Fig. 3.9C). Note that the NCF profile of each starting configuration is shown (*dashed*) to emphasize the increase in nematic order upon shearing.

The NCF profiles show differences in alignment depending on the polymer length and shear rate. While bulk networks composed of short-chain PPEs (20 monomers) mostly align in parallel, independent of the shear rate, the nematic order of bulk systems containing intermediate- and long-chain PPEs increases with shear rate. Particularly, at the interface between the fast (high shear rate) and slow band (low shear rate), the nematic alignment in bulk systems of intermediate- (*orange*) and long-chain PPEs (*green*) rises by nearly 50% for the high (*left*) and by 25% for the intermediate imposed shear rate (*center*).

More precisely, as the imposed shear rate increases from $0.1\,\mathrm{ns}^{-1}$ to $1.0\,\mathrm{ns}^{-1}$, the chain alignment in each shear band decreases, due to the presence of small scale thermal fluctuations at medium imposed shear rates, as evident from the velocity profile. Such fluctuations enable the disentanglement of semi-flexible PPEs in highly entangled networks. However, at the lowest imposed shear rate, we observed a linear velocity profile for bulk systems composed of long-chain PPEs with a very low degree of order, suggesting a highly entangled network, that is purely dominated by entropic effects.

Taken together, the non-equilibrium dynamics of entangled PPE networks, i.e., composed of chains beyond 20 monomers, can be attributed to a competition between entropic and enthalpic effects. The dynamics of networks comprising intermediate- and long-chain PPEs, i.e., usually driven by entropic effects under equilibrium (see section 3.3.7), is altered due to the imposed shear-flow, that represents an external force stretching the polymers. This shear-force favours the enthalpic contribution of the Hamiltonian, restricts the conformational space of PPEs and induces strand alignment. Conversely, in bulk systems of long-chain PPEs, excessive shear rates suppress thermal fluctuations orthogonal to the flow in y-direction, thereby preventing chain-disentanglement across the whole system and fostering the formation of shear bands with extreme shear rates in the fast band.

3.3.9 Rupture Forces in large Bulk Systems under Shear

We now turn towards the force distribution analysis within the semi-flexible PPE networks under shear. For this purpose, we combined intermonomer distances with an effective force constant of $k^b_{eff} = \left(k^{-1}_{BT} + k^{-1}_{B} + k^{-1}_{BT}\right)^{-1}$, where k_B is the force constant between two triangle beads and k_{BT} the one between triple bond and adjacent triangle bead (Tab. 3.2), to estimate the rupture force between adjacent segments. Due to the constraints, mimicking the rigidity of the aromatic ring, we obtained an effective force constant of $4,500\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$, assuming $k^{-1}_B \to 0$. This effective force constant is used to estimate the pairwise rupture forces between two monomers i and j with $F^r_{ij} = k^b_{eff}(r_{ij} - r^0_{ij})$. It is worth noting that the reference intermonomer distance of each bulk system was determined from the respective starting configuration and set to $0.6925\,\mathrm{nm}$ as well as $0.6934\,\mathrm{nm}$ for networks comprising short- and intermediate- as well as long-chain PPEs, respectively.

Fig. 3.10A shows the rupture force profile of each bulk system under shear, obtained by projecting pairwise forces onto the axis orthogonal to the flow in y-direction. Specifically, we first calculated pairwise forces between monomers of individual chains, binned them according to their y-position and defined the bin-wise rupture force as the maximum pairwise force. Finally, for each bin, we determined both average and standard deviation of the maximum rupture forces over time. It is evident, that, independent of the shear rates applied here, rupture forces within bulk systems of intermediate and long semi-flexible PPEs are overall higher than those of shorter chains.

Specifically, at the lowest shear rate of $0.01\,\mathrm{ns^{-1}}$ rupture forces are constant and uniformly distributed over the bulk network. However, this uniform distribution is lost at higher shear rates, where non-linear rupture force profiles arise, that coincide with the velocity and NCF profiles from section 3.3.8 (Fig. 3.9B-C). Most importantly, rupture forces for systems

composed of intermediate- and long-chain PPEs are elevated in the fast band, where shear rates can be two to four times larger than the imposed ones (Fig. 3.9A). More quantitatively, networks of intermediate and long PPEs at medium or high shear rates, exhibit rupture forces of 570 ± 55 pN or 590 ± 55 pN and 720 ± 70 pN or 740 ± 70 pN, respectively.

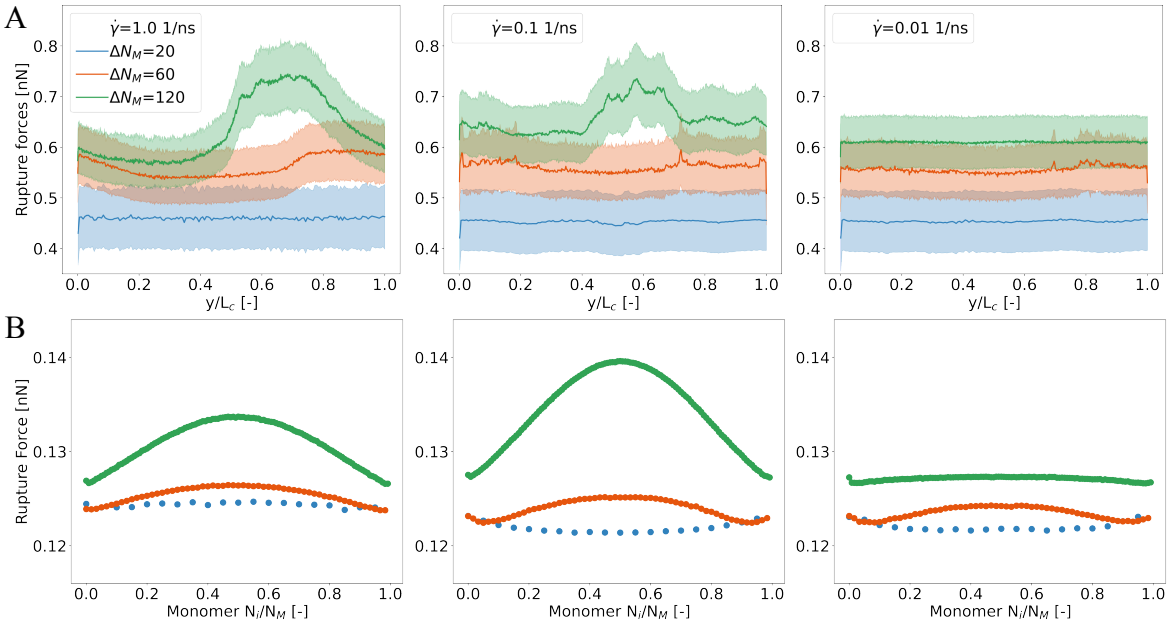

Figure 3.10. upture forces in bulk systems of large-chain PPEs under shear-flow. A Profile of pairwise rupture forces shows an increase with chain length and shear rate. **B** Bond rupture analysis predicts peak of time-averaged pairwise forces at the center of intermediate- (*orange*) and long-chain PPEs (*green*) at medium and high imposed shear rates.

Due to the superior scale-bridging property of Martini 3, that is the ability to combine non-linear mesoscopic effects, such as shear banding, with a fine-grained bond rupture analysis, we could identify a chain length dependence of the pairwise rupture forces along the PPE backbone (Fig. 3.10B). More specifically, while bulk systems composed of short-chain semi-flexible PPEs show a uniform distribution of rupture forces, which is independent of the applied shear rate, bulk systems of intermediate- and long-chain PPEs show a distinct peak, that is located at the center of the polymer at medium and high imposed shear rates.

This suggests an entropically driven bond cleavage process, meaning bond ruptures favour polymer chains of equal length to maximize the overall conformational entropy, as previously reported in the literature [190]. Consequently, semi-flexible polymer networks under shear are subject to continuous bond scissions until PPE reaches the length of its persistence length, where shear bands cease to exist and rupture forces are uniformly distributed along the polymer backbone, hence no preferred bond cleavage location can be identified.

3.4 Conclusions and Discussion

In this chapter, we have developed a Martini 3 model for the π-conjugated backbone of PPEs which we validated against experimental properties for single chains and bulk mixtures, such as the partition behaviour or interchain packing. The CG model reproduced the π-stacking and axial displacement of polymers in a bundle as observed in AA simulations [180, 181]. With the Martini 3 model at hand, we shed light on the structural alignment of PPEs at larger length scales. We found that the overall high nematic order is partly lost for longer chains (Fig. 3.8D,E). In detail, the assembly of semi-flexible polymers with chain lengths in the range of the persistence length exhibits parallel alignment across up to $\sim 6\,\mathrm{nm}$.

Benefiting from the computational boost upon coarse-graining, we combined the scale-bridging properties of Martini 3 with the Lee-Edwards boundary condition to perform shearing simulations of bulk systems containing thousands of semi-flexible PPE chains to elucidate their structural alignment and force distribution. We observed that above a critical shear rate, shear band formation in highly entangled polymer networks arises with extreme shear rates located in the fast band (Fig. 3.9). Such extreme shear rates result in highly parallel aligned chains, which are subject to high rupture forces, eventually leading to bond ruptures, in particular, at the center of the backbone (Fig. 3.10). Accordingly, we predict that shear band formation and force distribution within entangled PPE networks under shear-flow is mainly governed by the imposed shear rate, the persistence and polymer length.

It is important to note that shear banding of entangled polymer solutions was observed experimentally by particle-tracking velocimetry [191]. Specifically, Wang et al. found shear band formation in polybutadiene solutions using a large-amplitude oscillatory shear method [192, 193]. Burroughs et al. used rheomicroscopy to observe shear banding in polystyrene and polybutadiene networks during start-up Taylor-Couette flow [194, 195], and Boukany et al. in entangled DNA solutions using a rotational rheometer [196, 197]. Of note, Wang et al. stressed the importance of combining non-linear polymer rheology with molecular processes to elucidate the response of chain entanglement under shear [188]. Recently, Khomami et al. performed generic model-based dissipative particle dynamics simulations to propose shear-induced chain-disentanglement and stochastic chain dynamics as key factors for shear banding (see section 1.1) [198, 199]. Expanding on this, we observed that stochastic chain dynamics alone not suffices to explain shear banding in entangled networks. We further emphasize that the balance between entropic and enthalpic contributions within supramolecular assemblies, governed by shear rate, persistence and polymer length, might be key to the origin of this non-linear response to a shear-like deformation.

It is thus evident, that bulk systems composed of intermediate- and long-chain PPEs form shear bands with extreme shear rates, causing high rupture forces, preferably at the chain backbone center, and thereby approximately halving the polymer length. Such bond cleavages under shear continue until the length of the PPE chain is reduced to its persistence length, that is 15 nm or roughly 20 monomers, preventing shear band formation and thereby fostering a homogeneous shear deformation. This results in a uniform distribution of rupture forces along the backbone of PPEs, and thus reduces the likelihood of bond rupture events. Beside simulations, time-varying cryo-milling experiments of PPEs were combined with gel permeation chromatography (GPC) measurements by Maximilian Elter to show convergence of mass distributions towards half the initial mass (Fig. 3.11).

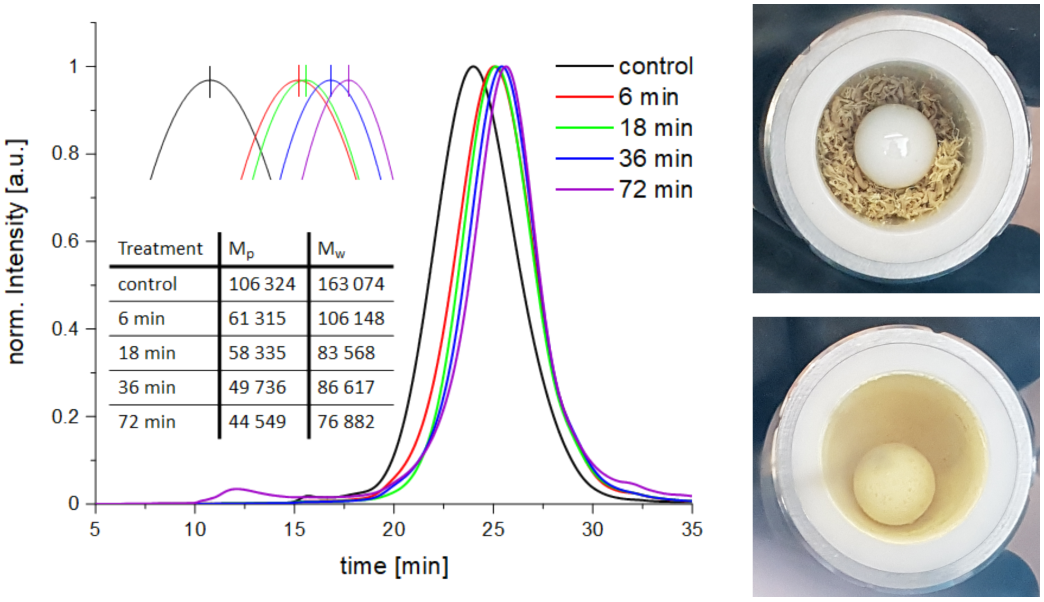

Figure 3.11. Gel permeation chromatography of cryo-milled PPEs. GPC measurements of solid-state PPEs subsequent to time-varying cryo-milling experiments at 30 Hz under liquid nitrogen[2]. Specifically, PPEs were exposed to mechanical stress by the cryo-mill (*right*) and their mass distribution observed by GPC measurements to show convergence towards half the initial mass (*left*). It is important to note that experimental measurements, data analysis and visualization were performed and kindly provided by Maximilian Elter.

[2]Experimental measurements, data analysis and visualization were performed and kindly provided by Maximilian Elter from the group of Uwe Bunz at the Heidelberg University. (15.09.2023)

Specifically, by comparing the time-evolution of the GPC mass distributions, we observed a halving of the initial mass after either 18 min, 36 min or 72 min of cryo-milling time. This suggests that PPE networks under mechanical load roughly halve until a lower chain length is reached, e.g., after 18 min, below which prolonged mechanical stress, i.e., longer cryo-milling times, not automatically leads to shorter polymer lengths. By combining the experiments with our Martini 3 shearing simulations, we can attribute this observation to the fact that shear bands with extreme shear rates, leading to high rupture forces, only form in networks of intermediate- and long-chain PPEs (Fig. 3.9). Short-chain PPE bulks, however, show a uniform distribution of rupture forces, where chemical bonds are less likely to fail, and thus GPC mass distributions stay constant with increasing cryo-milling time.

Beside shear banding, the Martini 3 force field can be straightforwardly extended to comprise side-chains of different chemistry (R in Fig. 3.2A) to analyze the packing of PPEs in greater detail. In addition, backmapping the Martini 3 bulk system to fine-grained resolution allows an advanced analysis on different length scales depending on the required degree of detail. For example, combining our CG model with the backmapping routines or hybrid methods could, e.g., help understand electronic properties, conductance or excited states in the environment found in bulk networks of semi-flexible PPEs.

In particular, combining yet unknown backmapping procedures, based on the symmetry of the PPE backbone, with AA or quantum mechanics calculations is expected to facilitate incorporation of bond rupture events, which are performed on the atomistic level of detail, into coarse-grained modeling. Alternatively, extending hybrid simulations methods, such as Kinetic Monte Carlo/ Molecular Dynamics (KIMMDY) [200], to the Martini 3 force field could be the way to bridge the atomistic scale by the usage of bond dissociation energies from quantum mechanics calculations. However, this poses quiet some challenges, among others, transferring rupture forces to CG resolution, handling bond ruptures inside beads and developing an automated mapping and bead type selection scheme for divers breakages.

Notwithstanding these rather technical challenges, non-linear rheological experiments of entangled PPE networks might complement our findings concerning shear band formation, and jointly guide our efforts in 3D printing such highly conjugated polymers with tuneable electronic and mechanical properties.

4 Structure Generation of Collagen Microfibrils with ColBuilder2

This chapter is concerned with the structure generation of collagen microfibrils within the size range of its molecule. An overview of collagen's hierarchical structure is given, before presenting the lattice-based optimization scheme to generate collagen microfibrils of different crosslink densities with ColBuilder2. This project was performed in collaboration with Benedikt Rennekamp from our group and Jaewoon Jung from the Riken Center for Computational Science in Kobe. Note that all-atom simulations of the collagen microfibrils were performed by Jung using the GENESIS molecular dynamics engine on the Fugaku supercomputer, while the simulation setup was developed jointly with Rennekamp.

4.1 Introduction

As outlined in chapter 1, collagen type I is an important protein in the human body, that, among others, transmits forces, provides structural integrity and renders our tissue mechanically by bearing high mechanical loads. Due to these exceptional requirements, collagen features a hierarchical structure from the atomic- to the mesoscale (Fig. 4.1) [10]. Collagen is a structural protein rich in glycine (GLY), proline and hydroxyproline with a common GLY-X-Y repetition unit, where X and Y are two random amino acids. Around 340 of such amino acids segments form a single collagen strand that further assembles, together with two more strands, into the collagen triple helix. Hence, the collagen molecule consists of three right-handed helices, specifically two α_1 helices and one α_2 helix, with all GLY residues pointing towards the inner core [10]. The collagen molecule extends up to 300 nm in length with 1.5 nm in diameter and is the building block or unit cell of the microfibril.

On a larger scale, many of these building blocks form an intertwined microfibrillar structure of triple helices exhibiting the 67 nm wide pattern of gap and overlap regions [73, 201]. At the interface between gap and overlap, two triple helices are enzymatically crosslinked. The type of crosslink influences the mechanical response of the collagen microfibril and

also changes under mechanical load. More precisely, the divalent (pre-mature) crosslink connects two lysine side-chains from different collagen molecules, and further reacts to a trivalent (mature) crosslink linking three side-chains of the two separate triple helices [10].

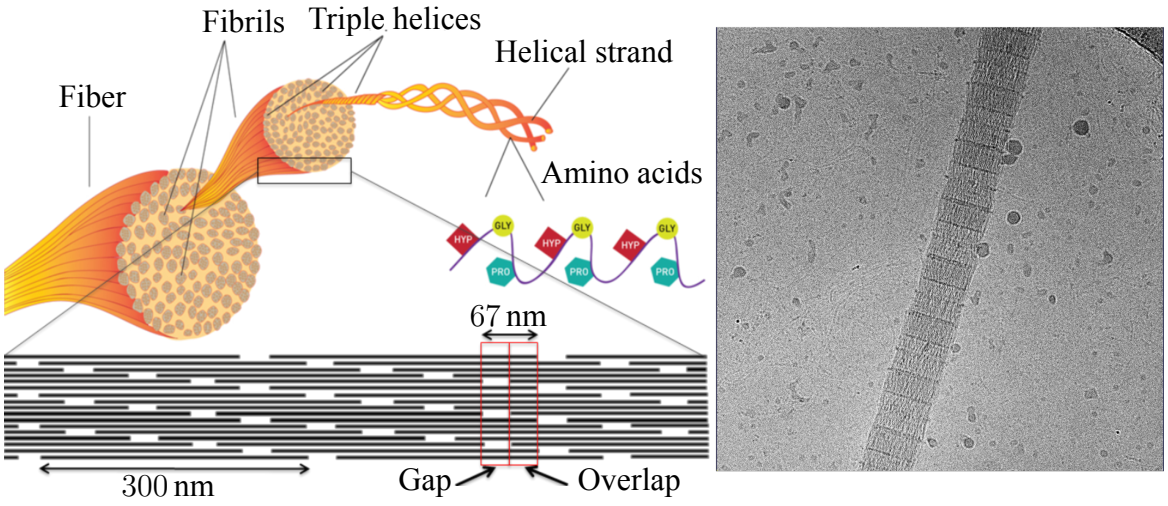

Figure 4.1. Hierarchical structure of collagen. The collagen triple helix consists of three strands, each with around 340 GLY-X-Y segments, where X,Y are mostly proline and hydroxyproline. Many collagen molecules (triple helices) assemble into a higher-order structure with a periodic pattern, i.e., the 67 nm-long D-band with gap and overlap region. Cryogenic electron microscopy image of a collagen microfibril from rat tail tendon was kindly provided by Aysecan Ünal[1]. The D-band pattern is clearly visible. Of note, the artwork of collagen's hierarchical structure (*left*) was provided by Agnieszka Obarska-Kosinska[2] [202].

Up to now, our atomistic models for the collagen fibril were limited to one D-band of 67 nm, or one gap and overlap region, and we are not aware of any larger atomistic models of collagen that have been investigated so far [203]. Hence, many of the long triple helices are only crosslinked at one end, but not connected to other triple helices. Accordingly, we have not yet been able to simulate the full mechanical interplay in the network of fibrils. In addition, we only used one type of crosslink, namely the divalent one, however, since the mechanical response critically depends on the type of crosslinking of individual strands within the microfibrillar structure, we developed a workflow for the generation of 300 nm-long microfibrillar divalent or trivalent crosslinked collagen structures, named ColBuilder2. For this purpose, we combined the *crystal contacts* tool from UCSF Chimera with a structural optimization algorithm on a Bravais lattice to automatize the generation of microfibrils

[1]Modified by Agnieszka Obarska-Kosinska former group member of Frauke Gräter.
[2]Image was taken by Aysecan Ünal group member of Frauke Gräter at Heidelberg University.

with arbitrary packing and shape, such as fibrillar length and diameter [204].

In addition, we realized further specifications, like the ratio of divalent to trivalent crosslinks or the introduction of random crosslink mutations, i.e., replacing randomly selected crosslinks with Lysine residues. Together with Jung from Riken, we performed preliminary AA simulations on Fugaku to validate the structural stability and to determine the influence of crosslinking on the mechanical response of the collagen microfibril.

4.2 Methods

The coordinate file for a single collagen type I triple helix was downloaded from the ColBuilder webserver (https://colbuilder.h-its.org), containing atomic positions, atom types and amino acid residues, and the crystal symmetry information about the unit cell. The *crystal contacts* tool from Chimera was used to generate copies of this unit cell, translate them according to the symmetry information and delete overlapping copies to obtain a higher-order crystal structure of collagen [204]. For optimization purposes, the structure was mapped on a Bravais lattice and optimized regarding structural homogeneity and crosslink connectivity.

4.2.1 Higher-Order Crystal Structure

To generate a higher-order crystal structure from a single coordinate file, the *crystal contacts* command from Chimera was applied [204]. The coordinate file of the collagen molecule and the desired contact distance were provided. The *crystal contacts* command read the crystal information from the coordinate file, more precisely the lattice parameters a, b, c, α, β, γ and space group G_{SP}, to define the unit cell of the molecule. Next, many symmetry copies of this unit cell were generated and positioned in the surrounding environment, each copy with an Euclidean transformation matrix $\boldsymbol{T} = (\boldsymbol{R}, \boldsymbol{t})$ given by a rotational $\boldsymbol{R}$ and translational $\boldsymbol{t}$ contribution. Clashes between overlapping atoms i and j with Van-der-Waals radii $r_{w,i}$ and $r_{w,j}$ from different symmetry copies were identified according to [204],

$$\zeta_{ij} = r_{w,i} + r_{w,j} - r_{ij} - \Delta_{ij}, \tag{4.1}$$

where ζ_{ij} denotes the overlap between two atoms, $r_{ij} = |\boldsymbol{r}_i - \boldsymbol{r}_j|$ is the interatomic distance and Δ_{ij} a buffer set to $0.0\,\text{nm}$. Symmetry copies with $\zeta_{ij} \geq -0.04\,\text{nm}$ were marked as close contacts and removed from the resulting higher-order structure. We obtained a coordinate file of the higher-order collagen structure and the transformation matrix $\boldsymbol{T}$ of each symmetry copy, enabling refinement of the unit cell arrangement on a discrete Bravais lattice.

4.2.2 Bravais Lattice

The Bravais lattice is an abstract concept to map higher-order structures in Cartesian coordinate space $\mathbb{R}^3$ on a lattice of infinite discrete points $\mathbb{Z}^3$. The unit cell of each symmetry copy is represented as a single point on a Bravais lattice. Higher-order crystal structures comprising many unit cells are collapsed into a set of integer points forming a well-defined geometrical arrangement. Therefore, the crystal orientation matrix $\boldsymbol{C}$ was derived from the lattice parameters to link the unit cells in the Cartesian coordinate space to points on the Bravais lattice, and vice versa. For example, the crystal orientation matrix $\boldsymbol{C} = (c_{ij})_{(i,j)=(x,y,z)}$ for a triclinic unit cell with space group 1, cell lengths a, b, c and angles α, β, γ reads [205],

$$\boldsymbol{C} = \begin{pmatrix} a & b \cdot \cos\gamma & c \cdot \cos\beta \\ 0 & b \cdot \sin\gamma & c \cdot \left(\cos\alpha - \frac{\cos\beta \cdot \cos\gamma}{\sin\gamma}\right) \\ 0 & 0 & \sqrt{c^2 - c_{xz}^2 - c_{yz}^2} \end{pmatrix}, \tag{4.2}$$

where the crystal orientation matrix $\boldsymbol{C}$ for the triclinic unit cell is an upper-triangle matrix with all elements below the main diagonal set to zero. Accordingly, a symmetry copy of a unit cell, translated by $\boldsymbol{t}$ from its origin in the Cartesian coordinate space, was mapped to a single point $\boldsymbol{p}$ on a Bravais lattice with eq. 4.3, and backmapped with eq. 4.4:

$$\boldsymbol{p} = \boldsymbol{C}^{-1}\boldsymbol{t}, \tag{4.3}$$

$$\boldsymbol{t} = \boldsymbol{C}\boldsymbol{p}. \tag{4.4}$$

4.2.3 All-Atom Simulation Protocol

While the ColBuilder2 framework primarily serves to generate large collagen microfibrils with various sizes and crosslink configurations at the Martini 3 CG level (see next chapter 5), we also performed all-atom (AA) molecular dynamics (MD) simulations to validate their stability and force-extension behaviour with experiments from the literature. We built the topology of the collagen microfibril with GROMACS 2023 by combining the Amber99sb*-ildnp force field from the ColBuilder webportal with the *pdb2gmx* tool from GROMACS, using ColBuilder2 [82, 203]. Next, we energy minimized the collagen microfibril in a $32 \times 32 \times 455\,\mathrm{nm}^3$ box of vacuum, added TIP3P water molecules, neutralized the system with counter ions and energy minimized the system once more. Since the solvated microfibrillar system contained around 43 million atoms, we relied on the Fugaku supercomputer to

perform AA simulations under force within a reasonable time. As a consequence of this, all subsequent simulations in equilibrium and under force were performed by Jung using the MD engine GENESIS on Fugaku, the second fastest supercomputer in the world. In the following, we outline the steps performed for preparing the solvated collagen microfibril for MD simulations under force. For a more detailed explanation of electrostatics, thermostats and barostats in GENESIS, we refer to the publication by Jung et al. [81].

To obtain a solvated collagen microfibril under isothermal-isobaric conditions, Jung first gradually increased the temperature from 0 to 300 K within 10 ns, and maintained the system for another 10 ns at the target temperature using a velocity rescaling type of coupling [81]. Next, Jung applied the Martyna-Tobias-Klein barostat to keep the pressure at 1 bar, with a 0.2 ps time constant (see section 2.1.5). Then, the time increment was gradually increased during the NpT equilibration from 0.5 fs to 2 fs over a total simulation time of 20 ns. More precisely, Jung started with a 500 ps NpT simulation with a 0.5 fs time step, followed by 500 ps with 1 fs, and ended with a 19 ns-long simulation with the commonly used 2 fs time step. During all equilibration simulations, backbone atoms were position-restrained with a force constant of 1000 kJ mol^{-1} nm^{-2}. Finally, we obtained an equilibrated system for each collagen microfibril, i.e., the pure trivalent, mixed divalent-trivalent and partly trivalent crosslinked microfibril, under isothermal-isobaric conditions.

To simulate the non-equilibrium behaviour of each collagen microfibril and moreover validate the structural stability under an external force, we applied a multi-step constant force simulation setup to arrive at the target pulling force of 1 nN per strand. Specifically, Jung performed five successive MD simulations under constant force, gradually increasing the force from 0 to the target value using an increment of 200 pN. Within 10 ns simulation time, we reached our target force of 1 nN per strand. Jung continued the simulations under force for each type of crosslinked collagen microfibril for another 250 ns to ensure full connectivity and stability. Finally, we calculated the axial distance between adjacent crosslink layers, each located at the gap-overlap transition along the microfibril, to analyze the mechanical response of differently crosslinked collagen microfibrils on the mesoscale.

4.3 Results

The modeling of the collagen microfibril was based on optimizing the higher-order structure from the *crystal contacts* tool with a refinement step performed on a Bravais lattice. For this reason, we mapped the entire structure to a set of points on a Bravais lattice, identified crosslinked molecules and performed an optimization procedure. In detail, we added extra points on the lattice, backmapped them to Cartesian coordinates (eq. 4.4) and translated the crosslinks to identify crosslinked molecules. Finally, an optimized structure of a fully crosslinked collagen microfibril was obtained to be used for production run MD simulations.

4.3.1 Collagen Microfibril on a Bravais Lattice

The atomistic model for the collagen microfibril consists of several symmetry copies of the same collagen triple helix from ColBuilder, aligned in a well-defined manner reflecting its crystal structure. The coordinate file defines each collagen triple helix as a set of atoms $A = \{a_1, ..., a_N\}$ at certain positions $\boldsymbol{Q} = (\boldsymbol{q}_1, ..., \boldsymbol{q}_N)$ together with the crystal information of the unit cell $\boldsymbol{u}$ given by its lattice parameters a, b, c, α, β, γ, space group G_{SP} and crystal orientation matrix $\boldsymbol{C}$ (eq. 4.2). By applying the *crystal contacts* tool, a symmetry copy of the unit cell u' was positioned in direct proximity of the initial unit cell u by first generating a new set of atoms $A' = A$ and transforming them according to [205]:

$$\boldsymbol{T}(\boldsymbol{q}_i) = \boldsymbol{R}\boldsymbol{q}_i + \boldsymbol{t}, \tag{4.5}$$

where $\boldsymbol{T}$ is a Euclidean transformation of $\boldsymbol{q}_i$ with $i = 1, .., N$, $\boldsymbol{R}$ the rotation matrix with $\det(\boldsymbol{R}) = 1$ and $\boldsymbol{t}$ the translation vector of the origin. Each atom of the new set of atoms A' was thus translated by the same vector $\boldsymbol{t}$, such that a translated symmetry copy of the collagen triple helix with positions $\boldsymbol{Q}' = (\boldsymbol{q}'_1, ..., \boldsymbol{q}'_N)$ and $\boldsymbol{q}'_i = \boldsymbol{T}(\boldsymbol{q}_i)$ is obtained. Of note, per definition the initial unit cell u is located at the origin of the coordinate system.

On a Bravais lattice with dimensions $(n_x, n_y, n_z) \in \mathbb{Z}^3$, this setup is represented by two points $\boldsymbol{p}$ and $\boldsymbol{p}'$, one for the unit cell of the collagen triple helix u at the lattice origin and one for its symmetry copy u'. According to eq. 4.3, the translated symmetry copy u' is mapped to the point $\boldsymbol{p}' = \boldsymbol{C}^{-1}\boldsymbol{t}$ on a Bravais lattice. Hence, a collagen microfibril consisting of N_s symmetrical copies with unit cells $u_1, .., u_{N_s}$, each positioned according to the transformations $\boldsymbol{T}_1, .., \boldsymbol{T}_{N_s}$ (eq. 4.5), is mapped to a set of discrete points on a Bravais lattice $\boldsymbol{p}_1, .., \boldsymbol{p}_{N_s}$, forming a geometrical configuration, i.e., dependent on the unit cell.

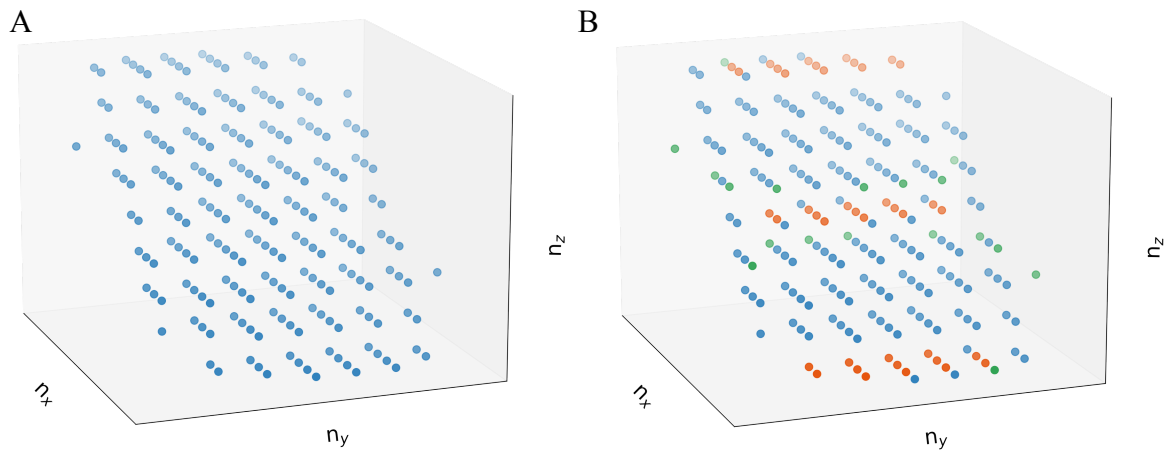

Figure 4.2. Bravais lattice for the collagen microfibril. **A** Bravais lattice for the microfibril obtained from the *crystal contacts* tool with a contact distance of 6 nm. Each point represents one collagen triple helix. A non-homogeneous distribution of points is evident indicating a non-uniform packing in the higher-order crystal structure. **B** Single, pairs and triplets of crosslinked collagen triple helices are colored in *green*, *blue* and *orange*.

Fig. 4.2A shows the Bravais lattice for the higher-order crystal structure of collagen obtained from the *crystal contacts* tool in Chimera [204]. We combined the atomistic coordinate file of *Rattus norvegicus*, which is based on the PDB file 3HR2 and was downloaded from the ColBuilder webportal, with a contact distance of 6 nm to obtain a higher-order crystal structure for the collagen microfibril [73, 203]. Note that the contact distance is the only input parameter for the *crystal contacts* tool, defining the distance inside of which symmetry copies are generated. For the structural analysis of the microfibril, we mapped the unit cell of each collagen triple helix on a Bravais lattice (eq. 4.3) to characterize the arrangement and distribution of the lattice points. As expected by the nature of the triclinic crystal system, the points exhibit a centrosymmetrical pattern with an inversion center at the lattice origin. In addition, more points are in the vicinity of the lattice origin than towards the edges, suggesting a denser packing of collagen triple helices. Hence, the microfibril exhibits a non-homogeneous distribution of unit cells with a looser packing towards the edges.

In addition, regarding the target application, namely performing MD simulations of a crosslinked collagen microfibril under force, we ensured full connectivity between collagen triple helices, preventing them from being pulled out during simulations, by an interatomic distance analysis on the microscopic scale. We calculated intercrosslink distances between different triple helices to identify links within the higher-order crystal structure. We defined pairs and triplets of crosslinked collagen triple helices when the interatomic distance

reached below a 0.3 nm cut-off distance. Fig. 4.2B show the Bravais lattice from A together with a color code related to number of connections; more precisely single, pairs and triplets of connected triple helices are colored in *green*, *blue* and *orange*, respectively. While the majority of collagen triple helices are linked by either one (73%) or on both ends (18%), the non-connected triple helices still account for about (9%). Obviously, this amount of singles, pairs and triplets depends to a large extent on the *crystal contacts* input parameter, that is the contact distance. Moreover, collagen triple helices belonging to triplets are located at the center and the edges of the lattice only, thereby revealing that the crosslink at each end of the biopolymer is linked to another one, and that the size of the microfibril is within the range of the collagen triple helix. Divalently crosslinked triple helices, however, are present over the entire length of the microfibril. Due to the non-uniform distribution of lattice points towards the edges and the amount of non-connected triple helices, the opportunity for structural optimization arises. Specifically, we aim to achieve the most uniform distribution of points possible through a denser molecular packing towards the lattice edges using a computational approach for flexible and optimized structure generation.

4.3.2 Structural Optimization on a Bravais Lattice

The computational workflow for the structural optimization of the collagen microfibril builds directly on the preceding ColBuilder webserver and is based on an iterative addition and crosslinking of new collagen triple helices to already existing ones. Fig. 4.3 provides a flow chart showing the computational workflow for ColBuilder2. The color coding relates to external (*purple*) and internal process steps performed on a Bravais lattice (*gray*) and in the Cartesian coordinate space (*turquoise*).

For algorithmic reasons, it should be noted that in the following we consider a collagen microfibril consisting of several triple helices as a system of models with certain properties. Accordingly, ColBuilder2 combines a model μ_1 from the ColBuilder webserver with the *crystal contacts* tool to generate a system of models $\boldsymbol{\mu} = (\mu_1, ...\mu_M)$, each with individual properties. Each model μ_i contains information about its crosslinks $A_i^c \subset A_i$, Euclidean transformation $\boldsymbol{T}_i$ (eq. 4.5) and Bravais lattice point $\boldsymbol{p}_i$. Both crosslinks at the telopeptide region of the triple helix are further characterized by their types, such as divalent or trivalent, residue id and name, atomic positions $\boldsymbol{Q}_i^c = (\boldsymbol{q}_{i,1}^c, ..., \boldsymbol{q}_{i,K}^c)$ in Cartesian coordinates, state of mutation and their potential contacts to any other model μ_j with $\mu_j \in \boldsymbol{\mu} \wedge \mu_j \neq \mu_i$. As a reminder, a contact between models μ_i and μ_j is established once the interatomic distance $d_{k,l}(\boldsymbol{q}_{i,k}^c, \boldsymbol{q}_{j,l}^c)$ between pairs of atoms $a_{i,k}^c$ and $a_{j,l}^c$ with positions $\boldsymbol{q}_{i,k}^c$ and $\boldsymbol{q}_{j,l}^c$ is below 0.3 nm.

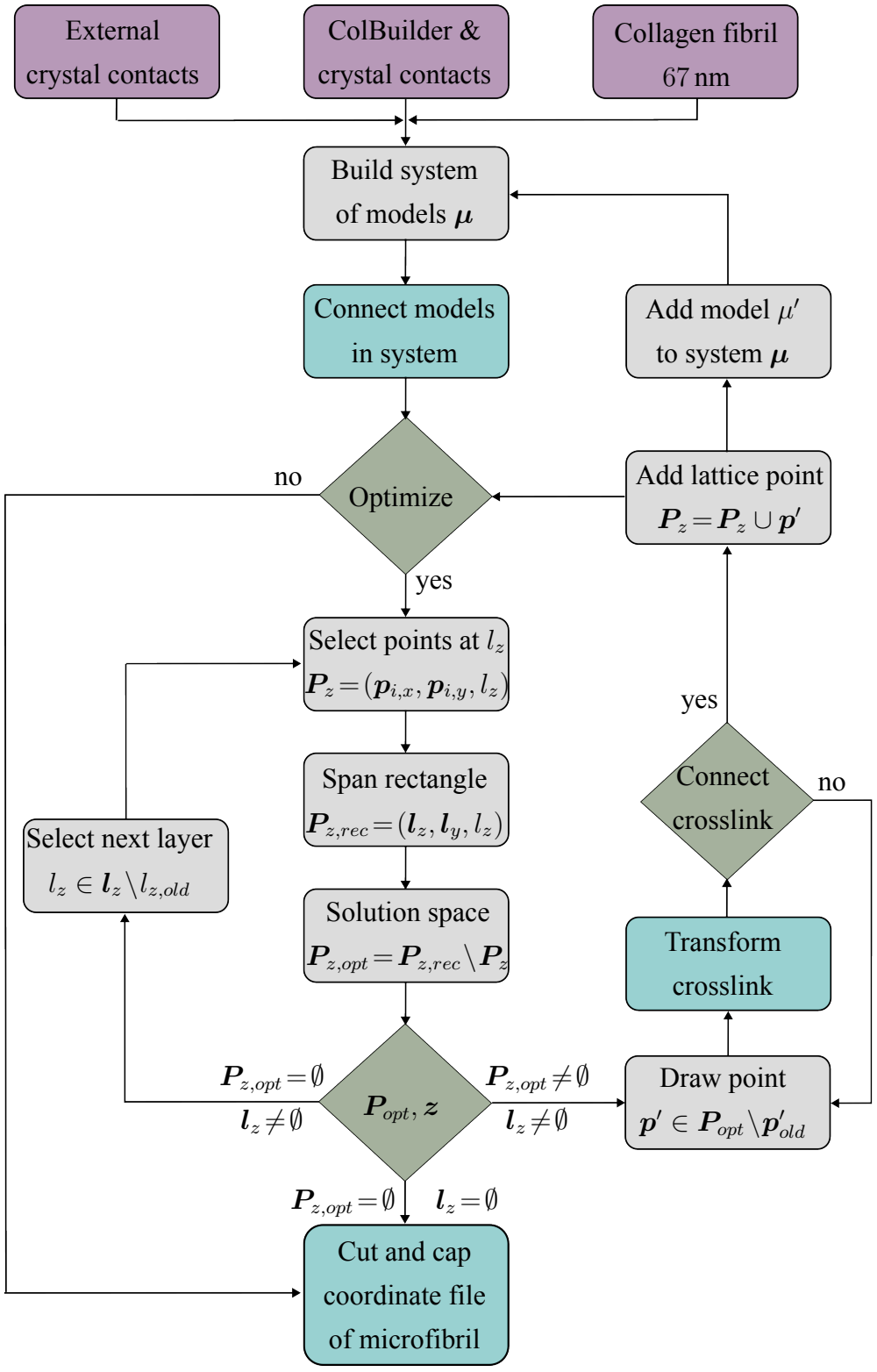

Figure 4.3. Flow chart for ColBuilder2. Steps performed on a Bravais lattice (*gray*) and in Cartesian coordinate space (*turquoise*) to optimize the structure of the collagen microfibril.

A system thus emerges from an ensemble of models, with their individual properties mapping the collagen microfibril as a whole on a Bravais lattice, as a starting point for the structural optimization procedure. Next, we calculated all intercrosslink distances to identify connected models and checked the *optimize* decision in Fig. 4.3. The optimization step aims to maximize the molecular packing within the collagen microfibril under the condition of fully crosslinked collagen triple helices.

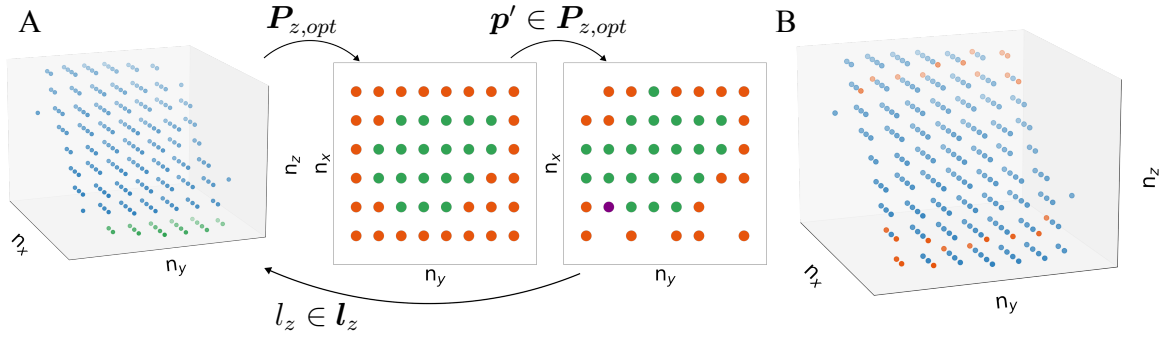

Figure 4.4. Structural optimization of the collagen microfibril on a Bravais lattice. A Structural optimization by selecting a layer (*green*), spanning a solution space (*red*), drawing a random point (*purple*) and identifying crosslinked molecules. **B** The optimized Bravais lattice for the collagen microfibril is centrosymmetric with a uniform point distribution.

We started with defining the solution space of the structural optimization procedure by $\boldsymbol{\delta} = (\delta_x, \delta_y, \delta_z)$ to restrict an infinite set of possible solutions to a finite one. We used the δ_z restriction to select the number of n_x,n_y-layers, counted from both edges in the n_z-dimension. For example, $\delta_z = 2$ binds the solution space to points in the upper and lower two layers $\boldsymbol{l}_z = \{n_{z,min} + \delta_z, n_{z,min} + \delta_z - 1, n_{z,max} - \delta_z, n_{z,max} - \delta_z + 1\}$ (Fig. 4.4A).

Each n_x,n_y-layer at a fixed position $l_z \in \boldsymbol{l}_z$ contains a set of points $\boldsymbol{P}_z = (\boldsymbol{p}_{i,x}, \boldsymbol{p}_{i,y}, l_z)$, with $i = 1, ..., N_z$ (*green*). From this layer, we determined the extreme values in the n_x,n_y-dimensions given by $n_{x,min}$ and $n_{x,max}$ or $n_{y,min}$ and $n_{y,max}$, respectively. By combining these extreme values with the solution space restrictions δ_x and δ_y, we defined two more vectors $\boldsymbol{l}_x = \{n_{x,min} - \delta_x, ..., n_{x,max} + \delta_x\}$ and $\boldsymbol{l}_y = \{n_{y,min} - \delta_y, ..., n_{y,max} + \delta_y\}$ to span a rectangular plane of discrete points $\boldsymbol{P}_{z,rec} = (\boldsymbol{l}_x, \boldsymbol{l}_y, l_z)$. By subtracting the already existing points $\boldsymbol{P}_z$ from $\boldsymbol{P}_{z,rec}$, we derived a set of points $\boldsymbol{P}_{z,opt} = \boldsymbol{P}_z \setminus \boldsymbol{P}_{z,rec}$ representing the solution space for layer l_z of the structural optimization problem (Fig. 4.4A, *red*).

From this solution space, we drew a random point $\boldsymbol{p}' \in \boldsymbol{P}_{z,opt}$, mapped this point to the Cartesian coordinate space with eq. 4.4 to obtain the transformation matrix $\boldsymbol{T}'$ with its translation vector $\boldsymbol{t}'$. In the Cartesian coordinate space, the model at the origin μ_1 is selected and its crosslink positions $\boldsymbol{Q}_1^c$ are translated by $\boldsymbol{t}'$ to a new location $\boldsymbol{Q}_1^{c\prime}$ (eq. 4.5).

Next, we calculated the intercrosslink distances between the potential model μ' and the already existing ones $\boldsymbol{\mu}$ to identify contacts between crosslinks according to our predefined criteria. If the intercrosslink distance was above the $0.3\,\text{nm}$ cutoff, the model μ' was neglected and on the other hand, if the contact criteria was fulfilled, the model was added to the system with $\mu_{M+1} = \mu'$. In both cases, we subtracted the point $\boldsymbol{p}'$ from the solution space $\boldsymbol{P}_{z,opt} = \boldsymbol{P}_{z,opt} \setminus \boldsymbol{p}'$, before the next point was drawn and the procedure started again.

Once the solution space of a certain $(\boldsymbol{l}_x, \boldsymbol{l}_y, l_z)$-layer was empty $\boldsymbol{P}_{z,opt} = \emptyset$, we updated the set of optimization layers $\boldsymbol{l}'_z \in \boldsymbol{l}_z \setminus l_z$ and selected the next $(\boldsymbol{l}_x, \boldsymbol{l}_y, l'_z)$-layer with $l'_z \in \boldsymbol{l}'_z$ to perform the same optimization steps as before. This procedure was repeated until the set of layers was empty $\boldsymbol{l}_z = \emptyset$, and therefore an optimized system of models $\boldsymbol{\mu}'$, as well as their transformation matrices $\boldsymbol{T}'$ were determined. The latter was combined with the *matrixset* command in Chimera to generate an optimized atomistic structure of the collagen microfibril. Finally, we post-processed the fibril by cutting each triple helix to a length of around $335\,\text{nm}$, and capping both termini with neutral Acetyl (ACE) or N-Methyl (NME) residues to enable the generation of topology files for MD simulations.

Before proceeding with a more detailed examination of the crosslinked microfibrils, it is worth mentioning that we performed the structural optimization of the collagen microfibril, used for the AA simulations, manually, since the optimization algorithm, as shown in Fig. 4.3, was developed afterwards. However, a direct comparison of the manually optimized lattice from Fig. 4.5A with the optimized one from ColBuilder2 Fig. 4.4B reveals only slight differences, so that they can be regarded as equivalent. Obviously, by choosing another solution space for the optimization problem $\boldsymbol{\delta}$, such as taking up to the fourth layer in the n_z-dimension into account $\delta_z = 4$, a Bravais lattice with a slightly different geometrical arrangement of points is obtained, in particular, for layers closer to the lattice origin.

4.3.3 Crosslink Specification for Collagen Microfibrils

The microfibril comprises many collagen molecules that are linked through covalent bonds formed in the telopeptide region at both ends of the triple helix. Such intermolecular crosslinks are the product of an enzyme-catalyzed reaction between two or three strands of two different collagen triple helices [10]. Due to the large variety of divalent and trivalent crosslinks, we focused on the divalent keto-amine crosslink, i.e., hydroxylysine-keto-norleucine (HLKNL), and its mature-trivalent form, i.e., hydroxylysyl-pyridinoline (PYD), although, any other type is equally suitable for structure generation with ColBuilder2 [203]. Note that a more in-depth analysis of these two crosslink types is provided later in section 5.3.1 of this thesis.

The crosslink specification for collagen microfibrils, generated with ColBuilder2, is possible by either mixing interconnected collagen triple helices with different crosslink types, or by introducing random mutations to crosslinks within the microfibril, that means replacing a randomly chosen crosslink with a Lysine residue. Therefore, in both cases the crosslink density is altered, affecting the mechanical response of the microfibril both on the macroscopic and microscopic scale. For the mixed crosslinked fibrils, we extended the ColBuilder2 workflow by first generating the whole microfibril for each crosslink type, and second mixing differently crosslinked triple helices together in a random way. By providing crosslink type and ratio within the collagen microfibril as input to ColBuilder2, pairs and triplets of connected triple helices with certain crosslink types are selected randomly to generate a microfibril comprising various crosslinked molecules.

We further introduced random crosslink mutations within the microfibril to reduce the amount of total crosslinks formed. As input a mutation rate is required, that is, the proportion of mutated crosslinks within the fibril. Mutations were performed randomly by calling the *swapaa* command from Chimera, however, are subject to an external constraint [204]. To maintain full connectivity of crosslinked collagen molecules and prevent strands from being pulled out during AA simulations under force, it is important that at least one of the two crosslinks is present in each telopeptide region to form a covalent bond with its matching counterpart. For this reason, we introduced a crosslink specific mutation state for each model that suppressed undesired mutations to prevent mutation rates above 50% (see section 4.3.2).

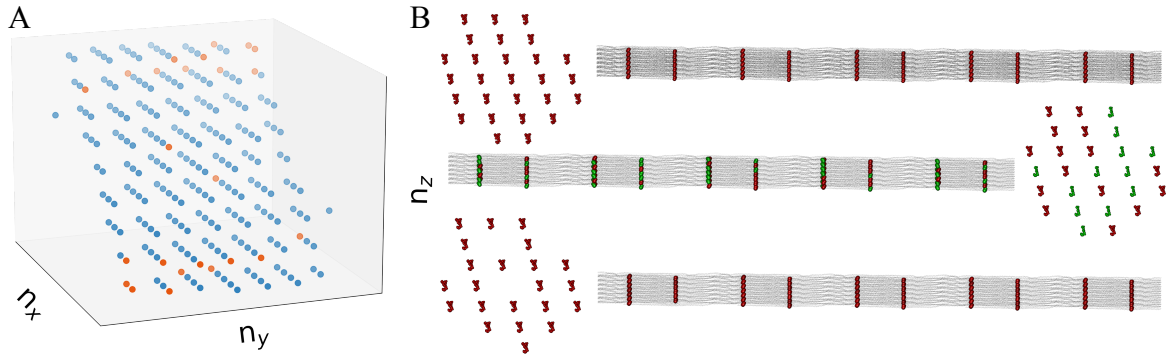

Figure 4.5. Microfibrillar collagen structures for MD simulations. A Bravais lattice of the collagen microfibril used for AA simulations. This lattice was optimized manually using the add-connect principle as explained above, since the automatisation of ColBuilder2 was developed afterwards. **B** Collagen microfibril with crosslink cross-section for three setups, namely a pure trivalent (*red, above*), mix divalent (*green*) and trivalent (*red, center*) or partly mutated trivalent (*red, below*) crosslinked microfibril.

Fig. 4.5B shows the structure of the collagen microfibril, generated based on the Bravais lattice from A, with three different kinds of crosslink setups, namely a pure trivalent (*above*), a mixed divalent-trivalent (*center*) and a trivalent setup with randomly mutated crosslinks (*below*). The pure trivalent crosslinked microfibril consists of five overlap and gap regions and is around $335\,\text{nm}$ in length (see section 4.1). At the transition between gap and overlap, the trivalent PYD crosslink is located (*red*) to link adjacent triple helices. In close proximity to the fibril, we display the cross-section of the microfibril with equally spaced crosslinks forming a rounded shape. For the mixed divalent-trivalent crosslinked microfibril (*center*), we defined four types of crosslinked collagen molecules based on its crosslink at each telopeptide region, namely divalent-divalent, trivalent-divalent, divalent-trivalent and trivalent-trivalent crosslinked collagen molecules. Next, we mixed each of these in equal proportions to generate the mixed crosslinked microfibril containing both divalent HLKNL (*green*) and trivalent PYD (*red*) crosslinks. We also reduced the number of formed PYD crosslinks within the pure trivalent crosslinked microfibril by randomly mutating trivalent crosslinks to Lysine residues with an upper limit of 30%. Accordingly, the trivalent partly mutated crosslinked microfibril consists of 70% pure PYD crosslinks, therefore features a much lower crosslink density compared to the other two microfibrils. In particular, this is evident when comparing cross-sections, as the mutant collagen microfibril reveals distinct gaps in the structure that are neither present in the pure trivalent, nor in the mixed divalent-trivalent crosslinked microfibrillar structure.

4.3.4 Collagen Microfibrils under Force

The collagen microfibrils now enable the analysis of the crosslink density on the mechanical response of the fibril by performing AA simulations under force. Therefore, we first generated the topology files for pairs and triplets of crosslinked collagen triple helices with ColBuilder2 by combining the information about crosslinked triple helices with the *pdb2gmx* command from GROMACS. Hence, we obtained one topology per crosslinked collagen molecules and included each into the main topology of the microfibril to speed up the topology generation process from a couple of weeks to a few hours. Each solvated microfibrillar system comprised around 43 million atoms and, up to now, was simulated for $250\,\text{ns}$ using the GENESIS MD engine on the Fugaku supercomputer at the Riken institute in Kobe.

Fig 4.6 shows a snapshot of the collagen microfibril with partly mutated trivalent PYD crosslinks under a pulling force of $1\,\text{nN}$ per strand. It is obvious that no collagen molecule of the microfibrillar structure is pulled out during the simulation, and that gap and overlap regions differ in terms of molecular packing and stretching under force. While the overlap

region consists of many parallel aligned collagen triple helices, the elongation is rather short compared to the gap region, with more intertwined triple helices. To characterize this difference more quantitatively, we calculated the axial Euclidean distances between crosslink layers at the gap and overlap transitions, as well as the overall end-to-end distance of the collagen microfibril. For the former, we selected the C_α atom of each trivalent and divalent crosslink from the α_1 strand of the triple helix, and used their center-of-mass as reference for the gap and overlap elongation along the fibrillar axis.

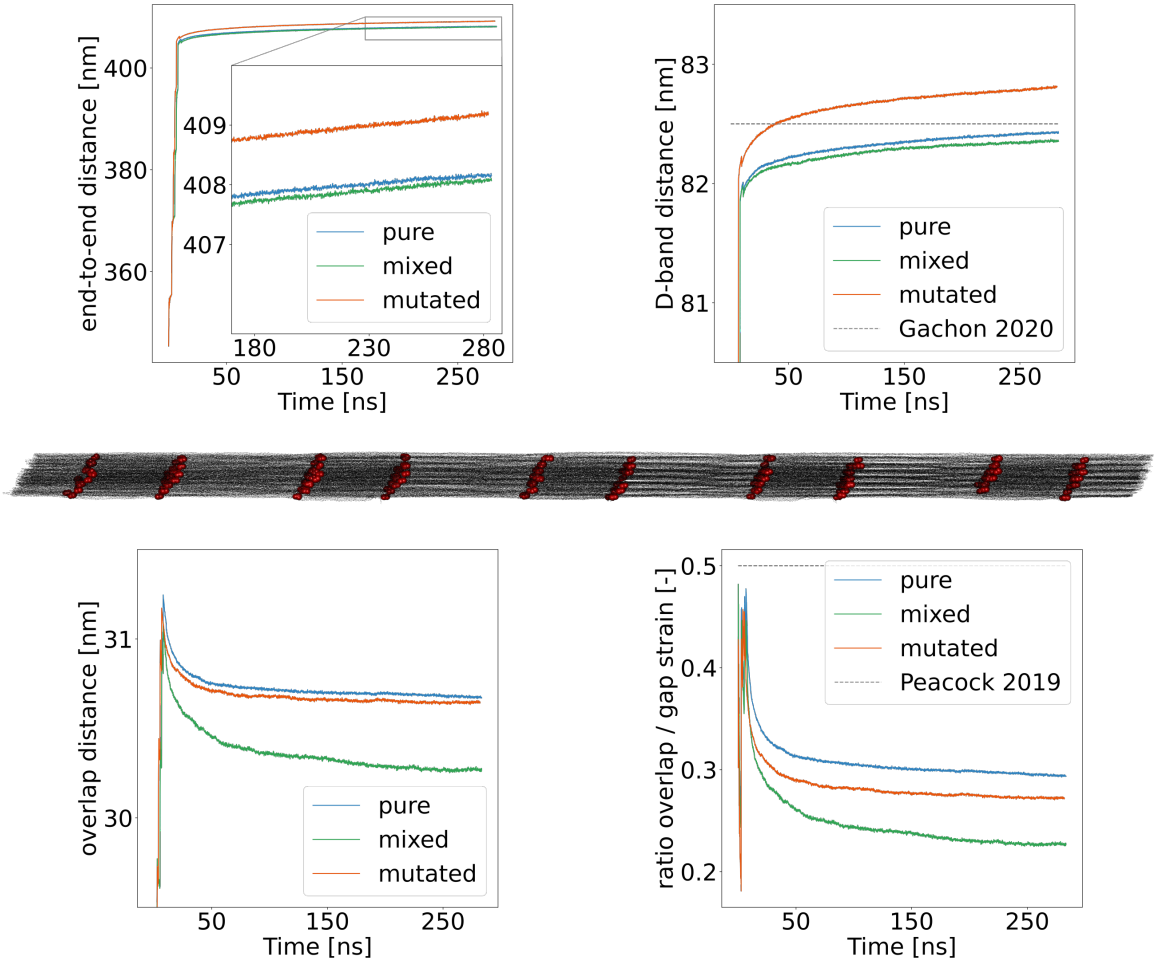

Figure 4.6. Collagen microfibrils under force. Mesoscopic observables from AA simulations of a pure trivalent (*blue*), mixed divalent-trivalent (*green*) and partly trivalent (*red*) crosslinked microfibril under force are the end-to-end (*top, left*), D-band lengthening (*top, right*), overlap distance (*bottom, left*) and the ratio between the gap and overlap elongation (*bottom, right*). Estimates from AA simulations are shown together with reference values for the D-band lengthening and the gap-overlap strain ratio from experiments [206–209].

We monitored the end-to-end distance (*top, left*), the D-band extension (*top, right*), the elongation of the overlap region (*bottom, left*) and the ratio between the overlap and gap extension (*bottom, right*) for each of the three differently crosslinked microfibrils (Fig. 4.6). Each of these distance measures increases sharply when the force starts acting right at the beginning of the simulation, before approaching a constant value asymptotically. During the course of the AA simulation under force, the 335 nm-long microfibril extends by approximately 23 %, with the partly trivalent crosslinked one stretching the most with more than 409 nm (*insight, red*) followed by the pure trivalent (*blue*) or mixed crosslinked microfibril (*green*) with around 408 nm. A similar trend is also observed for the D-band stretching, resulting from the sum of gap and overlap region, with an elongation of 82.4 nm and 82.8 nm for the pure trivalent and mixed divalent-trivalent as well as the partly trivalent crosslinked microfibril, respectively. Of note, the 22 % applied strain D-band elongation reproduces AA simulations of smaller systems under force (82.5 ± 1.0 nm) and atomic force microscopy (AFM) nanoindentation experiments from the literature (80 nm to 82.5 nm) [206–209].

In contrast, for the stretching of the overlap region, we observed a sudden increase in extension right at the beginning from 27.9 nm to 31.3 nm, followed by a gradual decrease approaching a constant value of 30.3 nm and 30.7 nm for the mixed divalent-trivalent crosslinked and the pure and partly trivalent crosslinked microfibril, respectively. Accordingly, the gap region for the partly trivalent and mixed crosslinked microfibril expands from 39.7 nm to 52.1 nm followed by the extension of the pure trivalent one with 51.7 nm. Hence, on average, we obtained an overlap-gap strain ratio of 25 % to 30 % (*bottom, right*), implying that the overlap region is roughly three-to-four times stiffer than the gap region. This fact agrees well with AFM nanoindentation experiments from the literature reporting a 25 % to 100 % increased stiffness in the overlap region compared to the gap region [208, 209].

Obviously, these mechanical properties proposed for the structural validation of the collagen microfibril strongly depend on the type of collagen molecule, like achilles or rat tail tendon, the experimental conditions, e.g., fibril humidity and surrounding temperature, and the force or strain applied to the collagen microfibril. Notwithstanding these environmental uncertainties, our collagen microfibrillar structure largely reproduces key structural and mechanical observables obtained from experiments reported in the literature, in particular, the D-band lengthening and the overlap-gap strain ratio.

4.4 Conclusions

In this chapter, we developed a computational workflow for building collagen microfibrils in the range of the entire molecule by combining the coordinate file of a single triple helix and its crystal symmetry information with an optimization algorithm on a Bravais lattice. We demonstrated the capabilities of ColBuilder2 by generating a 335 nm-long collagen microfibril with different crosslink configurations, such as a pure trivalent, a mixed divalent-trivalent and a partly mutated trivalent crosslinked microfibril. We further validated their structural stability by performing large-scale AA-MD simulations under force using the GENESIS MD engine on the Fugaku supercomputer. As a result, we not only found no unwinding and pulling out of collagen molecules, but also that our microfibrillar structure largely reproduces key structural and mechanical observables from experiments, like the D-band lengthening and the strain ratio between overlap and gap regions.

We propose to use ColBuilder2 for the generation of large collagen microfibrils for different species with individual shape, e.g., fibril length and diameter, and various crosslink specifications, to analyze the triple helical packing within the microfibril together with an experimental validation. Furthermore, combining ColBuilder2 with our Martini 3 force field parameterisation from the next chapter 5 enables large-scale CG-MD simulations of differently crosslinked collagen microfibrils under equilibrium conditions and under force.

Beyond ColBuilder2, we suggest to analyze pairwise forces within the collagen microfibril through force distribution analysis to identify locations of high-force concentration and determine the influence of crosslinking on the mechanical response [84, 211]. Finally, combining hybrid reactive MD schemes, e.g., Kinetic Monte Carlo/Molecular Dynamics (KIMMDY), with the collagen microfibril could benefit our understanding of covalent bond ruptures and the subsequent radical migration within the fibrillar network [200].

Overall, we demonstrated that ColBuilder2 is suitable to generate atomistic coordinate and topology files of collagen microfibrils of arbitrary shape from different species with various crosslink densities. ColBuilder2 can be extended to coarse-grained (CG) MD-simulations by incorporating the Martini 3 collagen model from chapter 5 into the topology generator. This enables CG-MD simulations of differently crosslinked collagen microfibrils, while still keeping an eye on the atomistic details to ensure, e.g., the differentiation between protein backbone and crosslink beads. By screening the response of various crosslinked collagen microfibrils to mechanical stress on longer timescales, insights into the maturation of divalent crosslinks across different species, i.e., important for aging and various diseases, like the bruck syndrome or osteogenesis imperfecta, might be gained [212].

5 Coarse-grained Modeling of Collagen for Molecular Dynamics under Force

In this chapter, we focused on developing a coarse-grained model for collagen using the Martini 3 force field. Our coarse-grained model reproduces the helical shape and mechanical bending stiffness of the collagen triple helix, as well as solvation properties of one divalent (pre-mature) and trivalent (mature) crosslink. We thus present Martini 3 force field parameters for the collagen microfibril that are ideally suited, after some refinement of the bond length potential, to be used in large-scale molecular dynamics simulations for predicting the influence of crosslinking on the collagen fibrillar network under mechanical load.

5.1 Introduction

Building on the previous chapter 4, collagen-based tissue features a hierarchical higher-order structure consisting of collagen fibers, which comprise many collagen microfibrils. Each microfibrillar structure is further composed of collagen triple helices, which are linked through divalent and trivalent crosslinks, e.g., hydroxylysine-keto-norleucine (HLKNL) and hydroxylysyl-pyridinoline (PYD), respectively [213, 214]. Due to the stiff nature of the triple helix, collagen is classified as a semi-flexible biopolymer and characterized by its persistence length (see section 1.2) [215]. Experimental studies reported persistence lengths ranging between 11 nm to 15 nm and 65 nm to 180 nm, depending on the experimental setup, such as the measurement device utilized and salt concentration [74, 75, 215]. Specifically, persistence lengths derived from electron microscopy, dynamic light scattering and rheological experiments were large compared to those obtained from atomic force microscopy imaging and optical tweezers stretching. Notwithstanding these variations, collagen is an outstanding protein that combines an increased backbone rigidity with a remarkable higher-order structure to regulate mechanical signals on the micro- and macroscale (see section 1.1).

For this reason, computational simulation methods ranging from the atomistic to the mesoscopic scale can be employed to obtain a more in-depth understanding of the processes involved in collagen's response to mechanical load. Starting from the bottom, all-atom molecular dynamics simulations of collagen were performed for a single molecule and a 67 nm-long fibrillar structure to study mechanical processes, like triple helical folding and stress concentration within crosslinked fibrils, respectively [216–219]. Buehler et al. performed all-atom (AA) molecular dynamics (MD) simulations of a 84 nm-long collagen molecule to derive a mesoscopic model of the collagen fibril in order to monitor the influence of crosslinking onto the stress-strain response [76]. Vaughan et al. further extended this model to characterize the influence of mineralization on collagen's response to mechanical load [220]. Although this mesoscopic model qualitatively reproduces the strength of collagen, it lacks a clear chemical and thermodynamic distinction between amino acid and crosslink type, and moreover it lacks the triple helical shape giving rise to the stiff nature of the collagen molecule. Gautieri et al. developed a coarse-grained model for the collagen triple helix, using the Martini 2 force field, by matching force-extension curves of small peptides, such as glycine-proline, glycine-proline-hydroxyproline and glycine-proline-hydroxyproline-glycine, to AA-MD simulations for parameterising the bond length, bond angle and torsion angle potential, respectively [77]. Hence, there exists no coarse-grained collagen model that considers the underlying atomistic structure of the triple helix and enables the simulation of large crosslinked microfibrils close to AA resolution.

Here, we present a preliminary coarse-grained (CG) model for collagen based on the Martini 3 force field to perform large-scale CG-MD simulations of collagen microfibrils under equilibrium conditions and under force [59]. By combining a non-equidistant bond length potential with a structure-based Gō-model, we maintained the collagen triple helix and fine-tuned the mechanical bending stiffness by matching the persistence length to AA simulations, using worm-like chain theory. We further performed non-equilibrium free energy calculations of the divalent HLKNL and trivalent PYD crosslink to select suitable bead types. We matched the force-stretching of a 67 nm-long collagen molecule and divalent crosslinked fibril to AA simulations by iterating over the bonded terms force constants to capture the response of the 335 nm-long microfibril to mechanical load.

Overall, our Martini 3 collagen model is expected to largely capture the force-extension from AA simulations and to perform CG-MD simulations of large collagen microfibrils under force. This, combined with ColBuilder2 from chapter 4, paves the way for a more comprehensive exploration of the interplay between crosslink configuration, microfibrillar structure and mechanical load from the microscopic to the mesoscopic scale.

5.2 Methods

AA- and CG-MD simulations were performed using GROMACS (version 2023) [148, 221]. Moreover, the Martini 3 force field was combined with Gō-like potentials to obtain a CG model for collagen, that is suitable for large-scale CG simulations under force [50, 59]. AA simulations were based on two different kind of atomistic force fields. The Amber99sb*-ildnp force field (Amber99) [222, 223] was selected to run simulations of the collagen triple helix and 67 nm-long divalent crosslinked fibril. The Chemistry at Harvard Macromolecular Mechanics (Charmm36) force field [224–227] was used to parameterise the divalent and trivalent crosslink through non-equilibrium MD-based free energy calculations.

5.2.1 All-Atom Simulation Protocol

While AA simulations of the collagen triple helix were based on Amber99, the divalent HLKNL and trivalent PYD crosslink were parameterised with Charmm36. This mixed force field parameterisation setup allowed us to obtain more accurate reference simulations for both crosslinks, since Amber99 requires capped protein termini for the topology generation, which falsifies the free energy estimates, and Charmm36 does not. Besides, we applied the Verlet scheme to cut off short-range electrostatic and Van-der-Waals interactions at 1 nm and the particle-mesh Ewald method to treat long-range electrostatics [95, 114, 154].

Collagen molecule and fibrils in water

We chose a single collagen triple helix with a contour length in the range of the D-band, that is 67 nm, to determine the Martini 3 force field parameters for the bonded interactions [73]. We solvated the collagen molecule in a rectangular simulation box with the TIP3P water and neutralized the solvated system with counter ions [224]. We applied the steepest descent algorithm to energy minimize the neutralized system, while position restraining the heavy atoms along the protein backbone with $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$. The velocity rescaling thermostat kept the temperature constant for 1 ns at 310 K with $\tau_T = 0.1\,\mathrm{ps}$ [101]. The pressure was stabilized at 1 bar operating the Parrinello-Rahman barostat for 2 ns with $\tau_P = 2\,\mathrm{ps}$ and a compressibility of $4.5 \cdot 10^{-4}\,\mathrm{bar^{-1}}$ [103, 104]. During these equilibration steps, a harmonic potential position restrained the heavy atoms of the collagen backbone with $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$. Next, we cut the 300 nm-long collagen triple helix into three 67 nm-long pieces, simulated each piece separately for 100 ns and merged the individual trajectories to ensure sufficient diversity of amino acid sequences along the collagen backbone.

For the fibrillar structures, we relied on AA simulations performed by Rennekamp and Jung. Rennekamp solvated the 67 nm-long collagen fibril with TIP3P water and neutralized the system with counter ions, before performing a 10 ns NVT equilibration at 300 K operating the velocity rescaling thermostat followed by another 10 ns NpT equilibration at 1 bar with the Parinello-Rahman barostat, while keeping the backbone atoms position restrained using the aforementioned force constant [206]. For the large 335 nm-long collagen microfibrils, AA simulations were performed by Jung at the Riken institute (see section 4.2.3).

5.2.2 Coarse-Grained Simulation Protocol

For generating a CG model for collagen with the Martini 3 force field, we relied on the *martinize2* transformation tool to obtain both coordinate and topology file of the Martini 3 force field, using an atomistic coordinate file as input [59, 228]. The common Martini 3 parameter setup was selected for the CG simulations regarding the neighbor list, electrostatics and Van-der-Waals interactions [50, 52]. In detail, the Verlet scheme updated the neighbor list every 20 to 40 steps, using a buffer tolerance of $5\,\mathrm{J\,mol^{-1}\,ps^{-1}}$, and the reaction-field algorithm cut off the electrostatics at 1.1 nm with a relative permittivity of $\epsilon_r = 15$ [58]. The Van-der-Waals interactions were expressed through a Lennard-Jones potential with a shift at 0.9 nm and cut-off at 1.1 nm [51]. Overall, we selected the Martini 3 simulation parameters suggested on their webportal to perform MD simulations (https://www.cgmartini.nl) .

Collagen molecule and fibrils in water

We simulated a single collagen molecule in water to derive the bonded parameters for the Martini 3 force field. We solvated each of the three collagen triple helices in a rectangular simulation box using Martini 3 water beads with a Van-der-Waals radius of 0.21 nm. Next, we neutralized the system with counter ions and energy minimized the system with the steepest descent algorithm, prior to stabilizing the temperature for 5 ns at 310 K operating the velocity rescaling thermostat with $\tau_T = 1$ ps. Pressure equilibration was performed with the Parrinello-Rahman barostat by keeping the pressure constant for 10 ns at 1 bar using a compressibility of $3 \cdot 10^{-5}\,\mathrm{bar^{-1}}$ and $\tau_P = 12$ ps. Starting from the equilibrated system, we performed production runs not only under equilibrium conditions for 100 ns, but also out-of-equilibrium applying constant force pulling to determine the force-stretching of our collagen triple helix. For the higher-order crystal structures, we followed the above outlined steps to obtain an energy minimized system and performed a subsequent two-step NVT and NpT equilibration procedure. In detail, we equilibrated the system under NVT conditions

with a reduced 2 fs timestep for 0.5 ns and continued for another 4.5 ns with a 10 fs time increment. The NpT equilibration was performed with a timestep of 1 fs for 2 ns, 2 fs for 3 ns and 10 fs for 8 ns, while position restraining the backbone beads with 1000 kJ mol^{-1} nm^{-2}. Finally, we obtained an equilibrated collagen molecule and fibril under isothermal-isobaric conditions, as a starting point for constant force pulling simulations.

5.2.3 Topology Generation for the Collagen Fibrils

Beginning from the microfibrillar structure from ColBuilder2, we obtained 231 coordinate files, one for the collagen microfibril and 230 for each molecule, together with a *connect* file, containing information about all pairs and triplets of crosslinked triple helices within the microfibril. In general, to obtain the topology of the collagen microfibril with Gō-Martini 3, we used a bottom-up approach by building the topology of each collagen molecule individually, before merging them according to their crosslinking information.

First, we started with pre-processing the coordinate file of a single collagen molecule and called the *martinize2* command to derive the Martini 3 force field topology and coordinate file [228]. We used the *contact map* analysis script to determine close contacts of the atomistic structure by combining a structural- and chemical-based approach [229]. We further provided the coordinate file from Martini 3 and the atomistic contact map as input for the *create_goVirt.py* script to obtain the Gō-Martini 3 topology and coordinate file of the collagen molecule by merging the physics- and structure-based modeling approaches [110].

Second, based on the crosslinked collagen triple helices within the fibrillar structure, as listed in the *connect* file from ColBuilder2, we merged topology files of pairs and triplets of crosslinked collagen triple helices together into single files. We next added bond length and bond angle potentials between the crosslink beads to the topology of the crosslinked collagen molecules. Third, and in contrast to the standard Gō-Martini 3 approach, we defined Gō-like Lennard-Jones interactions between virtual sites as *pairs*, instead of introducing new bead types, usually one for each virtual site, to enable the generation of a multi-million particle systems using *grompp* from GROMACS (version 2023) [221]. Of note, combining the original definition of virtual sites in Gō-Martini 3 with *grompp* is not recommended, due to large amount of memory and time required for generating run input files.

Finally, the topology file of a crosslinked collagen molecule is composed of pairs and triplets of collagen triple helices with their Gō-Martini 3 topology files and the bonded terms of the divalent and trivalent crosslink. Taken all topology files of crosslinked collagen molecules together, that is 101 separate topology files, we obtained our main topology for one of the three collagen microfibrils using the Gō-Martini 3 force field approach.

5.2.4 Non-equilibrium Free Energy Calculations

The selection of Martini 3 bead types for both crosslinks, namely the divalent HLKNL and trivalent PYD crosslink, was based on comparing partition coefficients from Martini 3 with experiments from the literature and AA simulations [59]. As outlined in section 2.3, the partition coefficient quantifies the distribution of a compound between two immiscible phases, such as octanol and water, and depends on the partitioning free energy of the solute (eq. 2.23). By performing non-equilibrium MD-based free energy calculations, we estimated the free energy associated with transferring a compound from water $S1$ to a hydrophobic solvent $S2$ using a thermodynamic cycle, $\Delta\Delta G_{S1\to S2} = \Delta G_{\emptyset\to S2} - \Delta G_{\emptyset\to S1}$ (Fig. 2.3). The solvation free energy of each compound in water $\Delta G_{\emptyset\to S1}$ and a hydrophobic solvent $\Delta G_{\emptyset\to S2}$ was estimated by thermodynamic integration (see section 2.3.2). Uncertainty quantification for the estimated free energy differences due to solvation ΔG was obtained by bootstrapping and error propagation.

We used the Charmm-Gui input generator to obtain the coordinate and topology files from the Charmm36 force field for both crosslinks [227, 230, 231]. Prior to thermodynamic integration, we solvated each compound in water and octanol, and neutralized the system, if needed [120]. For PYD, for instance, we inserted one chlorid ion to compensate the charged nitrogen N^+, located at the pyridine ring, to obtain a system with net charge zero. Subsequently, we energy minimized the system using the steepest descent algorithm and performed a NVT equilibration for 50 ns operating the Berendsen thermostat at 310 K in each end state, $\lambda = 0$ and $\lambda = 1$ [99]. From both end state ensembles, we discarded the first 10 ns and extracted 100 starting configurations. Alchemical transitions were completed in both directions, $0 \to 1$ and $1 \to 0$, using an equally spaced coupling parameter λ and the GROMACS implementation for free energy calculations [221]. During the 2 ns to 10 ns thermodynamic integration simulations, temperature was implicitly kept constant at 310 K using a stochastic dynamics integrator, while pressure was maintained at 1 bar using the stochastic rescaling barostat with $\tau_P = 5$ ns and a compressibility of $4.5{\cdot}10^{-5}$ bar^{-1} [102, 155]. A Gapsys soft-core potential was selected ($\alpha_{LJ} = 0.85$, $\sigma_{LJ} = 0.3$ and $\alpha_Q = 0.3$) to prevent singularities when changing the Lennard-Jones and Coulomb interactions together [232]. Finally, we relied on the *pmx* package by Gapsys et al. to quantify the free energy of transferring each compound between solvents, and iterated over the selected bead types and mapping schemes to match the free energy estimates from Martini 3 to AA simulations [126, 232].

5.2.5 Collagen Molecule and Fibril under Force

MD simulations under force were completed to compare the elongation of a single collagen molecule between AA and CG resolution in the entropic and enthalpic force-extension regime, i.e., where entropic and enthalpic effects govern the response to mechanical stress, respectively. As explained in section 2.4, the physics-based model from Kratky and Porod is suitable to quantify the mechanical bending stiffness of a semi-flexible polymer in the entropic force-extension regime by its persistence length [30]. Marko and Siggia extended the worm-like chain theory towards conditions under force by introducing a linear force-extension relation to the Hamiltonian of a single chain to derive an interpolation equation for the force-stretching of a semi-flexible biopolymer (eq. 2.30) [34–36]. With their interpolation equation at hand, we were able to match the persistence length of the collagen triple helix from CG to AA simulations under force and to experiments from the literature.

Starting configurations were generated by aligning the collagen triple helix in the y-dimension and performing a short NpT equilibration with the aforementioned protocol, while keeping the C_α atoms position restrained with $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$. MD simulations under constant force were performed with force constants ranging between 30 to $1500\,\mathrm{pN}$ per helical strand, applied at both ends of the collagen backbone, pointing orthogonal to the helix alignment in the z-direction. Due to the orthogonal pulling, the collagen triple helix rotated and straightened until a steady-state under force was reached. From the constant force MD simulations in the entropic regime ($F_z \leq 100\,\mathrm{pN}$), we estimated the persistence length of a single molecule using the stretched worm-like chain theory (eq. 2.30).

In the enthalpic force-extension regime, we performed MD simulations of the $67\,\mathrm{nm}$-long collagen molecule and divalent or trivalent crosslinked fibril under force without and with extra rotational restraints to monitor the effect on the unwinding of the triple helices, as proposed by Zapp et al. [218]. Specifically, we applied torque restraints with a force constant of $2,000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ and $1,000,000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ for AA- and CG-MD simulations, respectively, and moreover performed constant force pulling simulations using force constants ranging from $300\,\mathrm{pN}$ to $1500\,\mathrm{pN}$. Note that the AA-MD simulations of the $67\,\mathrm{nm}$-long divalent crosslinked fibril under force were performed by Rennekamp [206, 218].

5.3 Results

As introduced in section 1.2, the Martini 3 force field relies on a building block principle for mapping a given AA structure to CG resolution, and on a mixed top-down and bottom-up parameterisation strategy for the non-bonded and bonded interactions, respectively [59, 161]. With the aim of performing CG simulations of the collagen microfibril with the Martini 3 force field, we split the complex fibrillar system into smaller blocks, for each of which a CG model was developed. Accordingly, the collagen microfibril is composed of many triple helices, which are linked through divalent HLKNL and trivalent PYD crosslinks. We hence followed the Martini 3 philosophy to parameterise each crosslink and the collagen triple helix individually, before assembling them into microfibrillar collagen structures.

Mapping scheme and bead type selection for crosslinks were optimized by comparing partition coefficients, obtained from alchemical transitions, with reference values and AA simulations. Furthermore, probability densities from Martini 3 simulations were matched to those from mapped AA ones using direct Boltzmann-inversion to determine the parameters for the bonded terms between protein backbone beads. Validation targets were defined through structural and mechanical properties, such as the triple helical shape, given by the rise per residue and residues per turn, the mechanical bending stiffness, characterized by the persistence length, and the force-extension in the enthalpic force-stretching regime. For now, we provide an overview about the major steps involved in setting up a CG description for collagen based on the Martini 3 force field in the flow-sheet diagram 5.2.

5.3.1 Mapping and Bead Type Selection for Crosslinks

Mapping complex molecules, such as HLKNL and PYD, from AA to CG resolution with Martini 3, determines all subsequent steps in the parameterisation process [55]. In Martini 3, on average, two-to-four heavy atoms plus the associated hydrogens are mapped to a single CG interaction site, located at the center-of-geometry of the underlying AA structure. Functional groups, e.g., carboxyl groups, are kept together, and the symmetry and shape of the atomistic structure is retained as much as possible [59, 111]. Following these guidelines, we defined a symmetrical plane in the center of the HLKNL crosslink, i.e., at the secondary amine group, which has an alcohol and a ketone group attached on each side (Fig. 5.1A, *right*). Bearing the symmetrical property of the crosslink in mind, we selected a small *SP1d* bead to represent the hydrogen donor characteristics of the secondary amine group, and used a regular four-to-one mapping for 1-propanol and 1-propanone. To differ-

entiate between both sides of the amine group, we chose a polar *P1* bead for the alcohol and an intermediate hydrophilic bead, i.e., the *N6a* bead, for the ketone group on the other side.

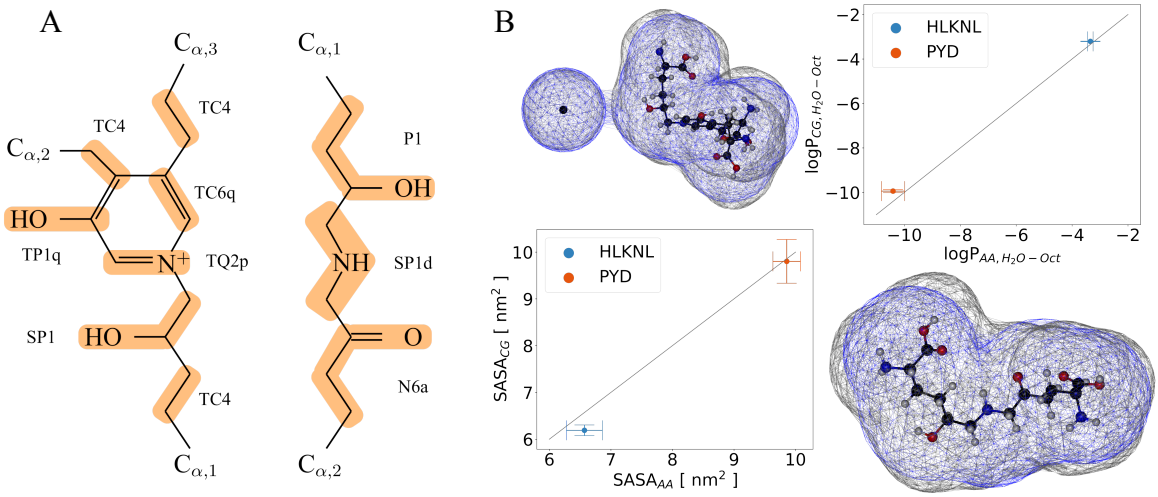

Figure 5.1. Mapping and bead type selection for crosslinks with Martini 3. A Mapping and bead types of the trivalent PYD (*left*) and divalent HLKNL (*right*) crosslink in Martini 3. **B** Non-equilibrium free energy calculations to compare the octanol-/water partition coefficient of each compound from CG to AA simulations (*top,right*). Bonded interactions were tuned by matching the solvent accessible surface area from Martini 3 to AA simulations (*bottom, left*). The Conolly surface for both divalent (*top, left*) and trivalent (*bottom, right*) crosslink agree well between fine-(*gray*) and coarse-grained (*blue*) resolution.

For PYD, only small and tiny beads were used to map the atomistic structure to CG resolution in order to emphasise the aromatic nature of the crosslink (Fig.5.1A, *left*). The central pyridine ring was represented by three tiny beads, namely one *TQ2p* bead for the nitrogen, one *TP1q* for the phenol and one *TC6q* for the benzene building block, forming a triangular structure reflecting its planar shape and rigidity. Due to the positive charged nitrogen, *q*-labels were introduced to distribute the charge over all tiny beads involved in the mapping of pyridine. Hence, according to the charge distribution in Amber99, charges of $q = 0.7\,\mathrm{e}$, $q = 0.2\,\mathrm{e}$ and $q = 0.1\,\mathrm{e}$ were assigned to the *TQ2p*, *TP1q* and *TC6q* bead, respectively. In addition, the trivalent crosslink features three chemical links connecting two triple helices (Fig.5.1A, *left*). The backbone beads from two strands of one triple helix, namely $C_{\alpha,2}$ and $C_{\alpha,3}$, were linked to the pyridine ring using a moderate polar *TC4* bead. The last link between the charged nitrogen and $C_{\alpha,1}$ was accomplished with one *SP1* bead representing the polar character of ethanol, and one more moderate *TC4* bead for ethane.

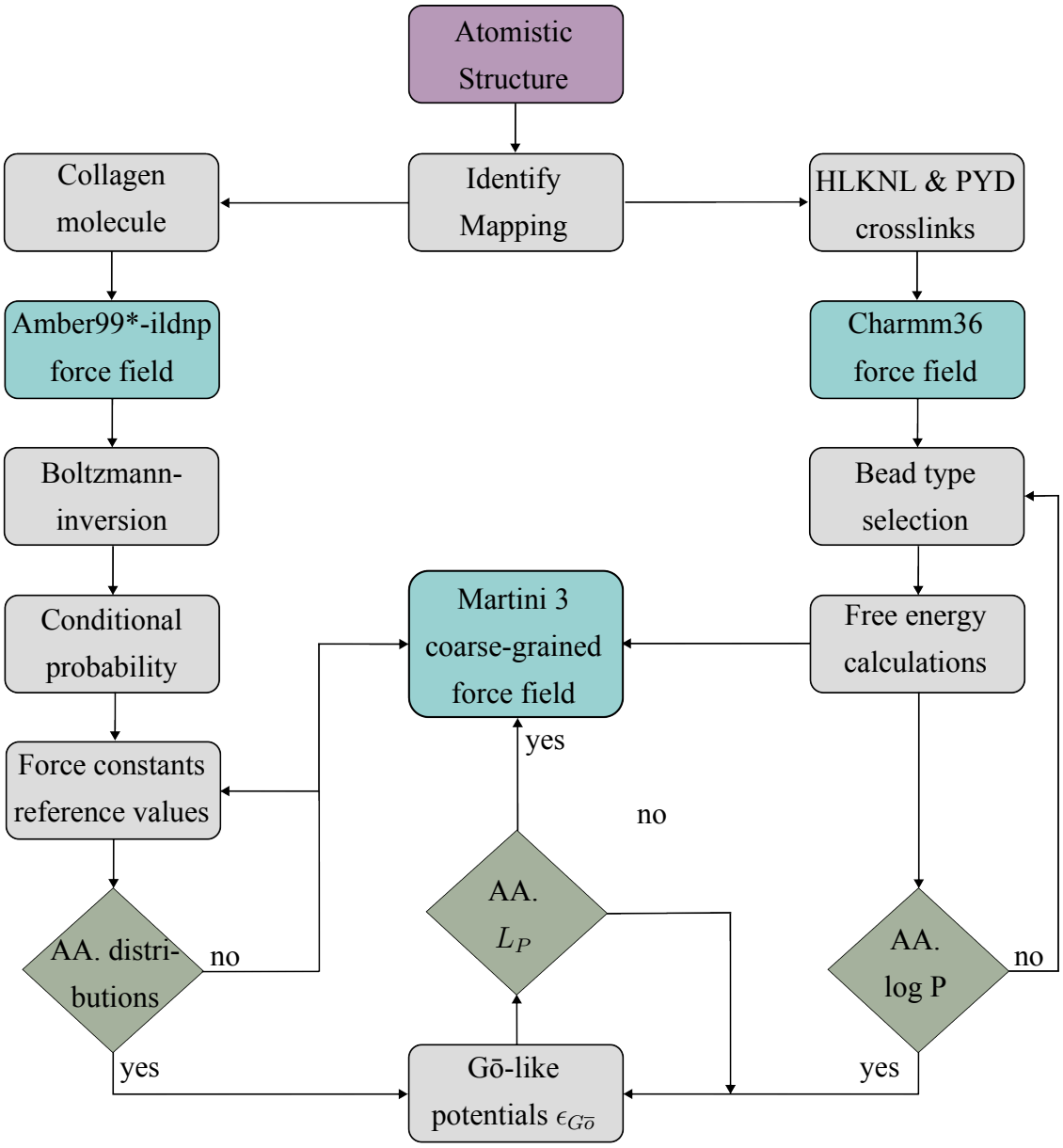

Figure 5.2. Overview for modeling collagen with Martini 3. Purple boxes represent the atomistic structures obtained from the ColBuilder webportal, e.g., the collagen triple helix, the fibril and the divalent and trivalent crosslink [203]. Turquoise boxes are the AA and Martini 3 force fields. Gray boxes are steps performed for parameterising the Martini 3 force field. Green rhombuses are decision boxes for validation and tuning purposes, such as matching bonded terms distributions (*left*) and partition coefficients to AA data (*right*). As soon as both conditions agree, Gō-like potentials were used to tailor the chain stiffness to AA simulations by matching the persistence length.

5.3.2 Partitioning Free Energy for Crosslinks

The bead type selection was based on comparing partition coefficients for both crosslinks from Martini 3 with AA simulations [58]. The partition coefficient of each compound was determined in an octanol-/water system using non-equilibrium MD-based free energy calculations. To estimate the free energy of transferring each crosslink from octanol to water, we combined the *pmx* package with a thermodynamic cycle, i.e., $\Delta\Delta G_{S1\rightarrow S2} = \Delta G_{\emptyset\rightarrow S2} - \Delta G_{\emptyset\rightarrow S1}$ (Fig. 2.3) [126, 232]. In detail, we performed four separate replicates with 100 backward and forward transitions each, yielding a cumulative simulation time of $4\,\mu$s, to ensure sufficient sampling of both crosslink compounds.

Building on this exhaustive sampling of the conformational space, we obtained an estimate for the free energy of partitioning for the divalent and trivalent crosslink, that is $\Delta\Delta G_{W\rightarrow O} = -19.06 \pm 0.07\,\mathrm{kJ\,mol^{-1}}$ as well as $\Delta\Delta G_{W\rightarrow O} = -58.98 \pm 0.43\,\mathrm{kJ\,mol^{-1}}$, respectively. In addition, we facilitated the experimental- and AA-based validation by calculating the partition coefficients with eq. 2.23 (Fig.5.1B). We observed sufficient agreement for the partition coefficients of the divalent and trivalent crosslink between Martini 3 (-3.05 ± 0.01 and -9.96 ± 0.08) and AA simulations (-3.11 ± 0.05 and -10.26 ± 0.44).

However, by comparing our MD-based partition coefficients of the divalent and trivalent crosslink with machine-learned predictions from *rdkit* [233] (-2.11 and -5.46) and the atom additive method *xlogp3* from PubChem [234] (-7.1 and -8.6) only a qualitative agreement was found. This is attributed to the chemical nature of each crosslink, which bears features, like a charged nitrogen in the pyridine ring, that are not well captured by the predictions.

5.3.3 Parameterising Bonded Terms for Crosslinks

In Martini 3, the parameterisation of bonded interactions between CG beads is based on a bottom-up approach by matching probability densities of the bonded terms in an iterative process to mapped AA simulations [111]. We followed the previously mentioned simulation protocol to obtain an equilibrated isothermal-isobaric ensemble for each crosslink in water and used a center-of-geometry based scheme to map the atomistic trajectory to CG resolution. From this mapped trajectory, we calculated bond distances, bond angles and torsion angles between crosslink beads and their respective probability distributions. By iterating over the force constants and reference values of the Martini 3 force field, we matched the CG probability densities to the mapped AA ones, following the standard procedure for parameterising new molecules with Martini 3 (Fig. A7-A8) [59, 235].

Accordingly, bond lengths between Martini 3 beads were rescaled by matching the solvent accessible surface area (SASA) to AA simulations to reproduce the shape of the atomistic structure. After rescaling, as expected, the SASA retrieved from Martini 3 simulations for the divalent ($6.2 \pm 1.1\,\mathrm{nm}^2$) and trivalent crosslink ($9.7 \pm 3.5\,\mathrm{nm}^2$) agree well with estimates from AA simulations ($6.5 \pm 2.9\,\mathrm{nm}^2$ and $9.8 \pm 2.2\,\mathrm{nm}^2$) (Fig. 5.1B, *left*). Tab. 5.1 lists the Martini 3 force field parameters for the trivalent (*above*) and divalent (*below*) crosslink.

Table 5.1. Martini 3 force field parameters for crosslinks. Martini 3 parameter set for 3 bonded terms of the trivalent PYD (*above*) and divalent HLKNL crosslink (*below*).

Bonded term	Reference value	Force constant
BB_1-TC4	0.270 nm	11000 kJ mol^{-1} nm^{-2}
TC4-SP1	0.250 nm	12000 kJ mol^{-1} nm^{-2}
SP1-TQ2p	0.260 nm	7000 kJ mol^{-1} nm^{-2}
TQ2p-TP1q	0.250 nm	70000 kJ mol^{-1} nm^{-2}
TQ2p-TC6q	0.270 nm	70000 kJ mol^{-1} nm^{-2}
TP1q-TC6q	0.395 nm	70000 kJ mol^{-1} nm^{-2}
TP1q-TC4	0.230 nm	12000 kJ mol^{-1} nm^{-2}
TC6q-TC4	0.290 nm	9000 kJ mol^{-1} nm^{-2}
BB_2-TC4	0.250 nm	9000 kJ mol^{-1} nm^{-2}
BB_3-TC4	0.250 nm	9000 kJ mol^{-1} nm^{-2}
BB_1-TC4-SP1	180°	150 kJ mol^{-1}
TC4-SP1-TQ2p	180°	150 kJ mol^{-1}
SP1-TQ2p-TP1q	180°	150 kJ mol^{-1}
SP1-TQ2p-TC6q	100°	150 kJ mol^{-1}
TP1q-TC4-BB_2	140°	150 kJ mol^{-1}
TC6q-TC4-BB_3	140°	150 kJ mol^{-1}
BB_1-P1	0.310 nm	10000 kJ mol^{-1} nm^{-2}
P1-SP2	0.415 nm	7000 kJ mol^{-1} nm^{-2}
SP2-N6a	0.365 nm	5000 kJ mol^{-1} nm^{-2}
N6a-BB_2	0.360 nm	9000 kJ mol^{-1} nm^{-2}
BB_2-N6a-SP2	140°	150 kJ mol^{-1}
N6a-SP2-P1	140°	50 kJ mol^{-1}
SP2-P1-BB_1	140°	150 kJ mol^{-1}

On a microscopic scale, we compared AA- (*gray*) and CG-based Conolly surfaces (*blue*) of each crosslink to identify the subtle differences in the molecular structure (Fig 5.1B). We found that the CG model for the PYD crosslink largely reproduces the Van-der-Waals surface area from the AA structure with slight deviations near the protein backbone beads. A similar trend is observed for the Conolly surface of the divalent crosslink. These slight deviations can be attributed to the loss of fine-grained information upon coarse-graining.

5.3.4 Parameterising Bonded Terms for the Collagen Molecule

The collagen molecule exhibits a triple helical structure that is the reason for its unique mechanical properties arising from the intra- and interhelical interactions. Therefore, we examined the parameterisation of bonded terms thoroughly. Beyond the standardized Martini 3 bonded potentials parameterisation scheme, we combined a conditional probability distribution with direct Boltzmann-inversion to reproduce the symmetry and shape of the collagen triple helix. For this purpose, we performed AA equilibrium simulations of three different 67 nm-long collagen molecules in water for 100 ns, before applying a center-of-geometry based mapping scheme to the protein backbone. From the mapped trajectories, we measured bond lengths, bond angles and torsion angles between backbone beads and used their probability distributions as our major parameterisation targets (Fig. 5.3A, *blue*).

The probability distributions for the bond angle and dihedral angle are unimodal with a peak at $138\,^\circ$ and $76\,^\circ$ and a width of approximately $17\,^\circ$ and $27\,^\circ$. For the bond length distribution, however, we observed a pronounced tail implying an underlying bimodal character (Fig. 5.3A, *left*). As a consequence of this, bond lengths between mapped backbone beads are non-equidistant, and thus amino acid sequence specific. We therefore performed direct Boltzmann-inversion by fitting a Gaussian $P(q)$ to the measured probability distributions to derive the Martini 3 force field parameters for each bonded term, separately [236]:

$$P(q) = C_q \cdot \exp\left(-\frac{1}{2}\frac{(q-\mu_q)^2}{\sigma_q^2}\right), \tag{5.1}$$

where q denotes a single degree of freedom, e.g., the bond distance r_{ij}, bond angle θ_{ijk} and torsion angle ϕ_{ijkl}, and μ_q, σ_q^2 and C_q are the mean, variance and amplitude of the fitted Gaussian, respectively. For Boltzmann-inversion, we assume each degree of freedom to be independent in the canonical ensemble, and thus obeying a Boltzmann distribution [236]:

$$P(q) = Z^{-1} \cdot \exp\left(-\beta \cdot V^{CG}(q)\right), \tag{5.2}$$

where Z is the partition function taken over the ensemble and $V^{CG}(q)$ the Martini 3 potential.

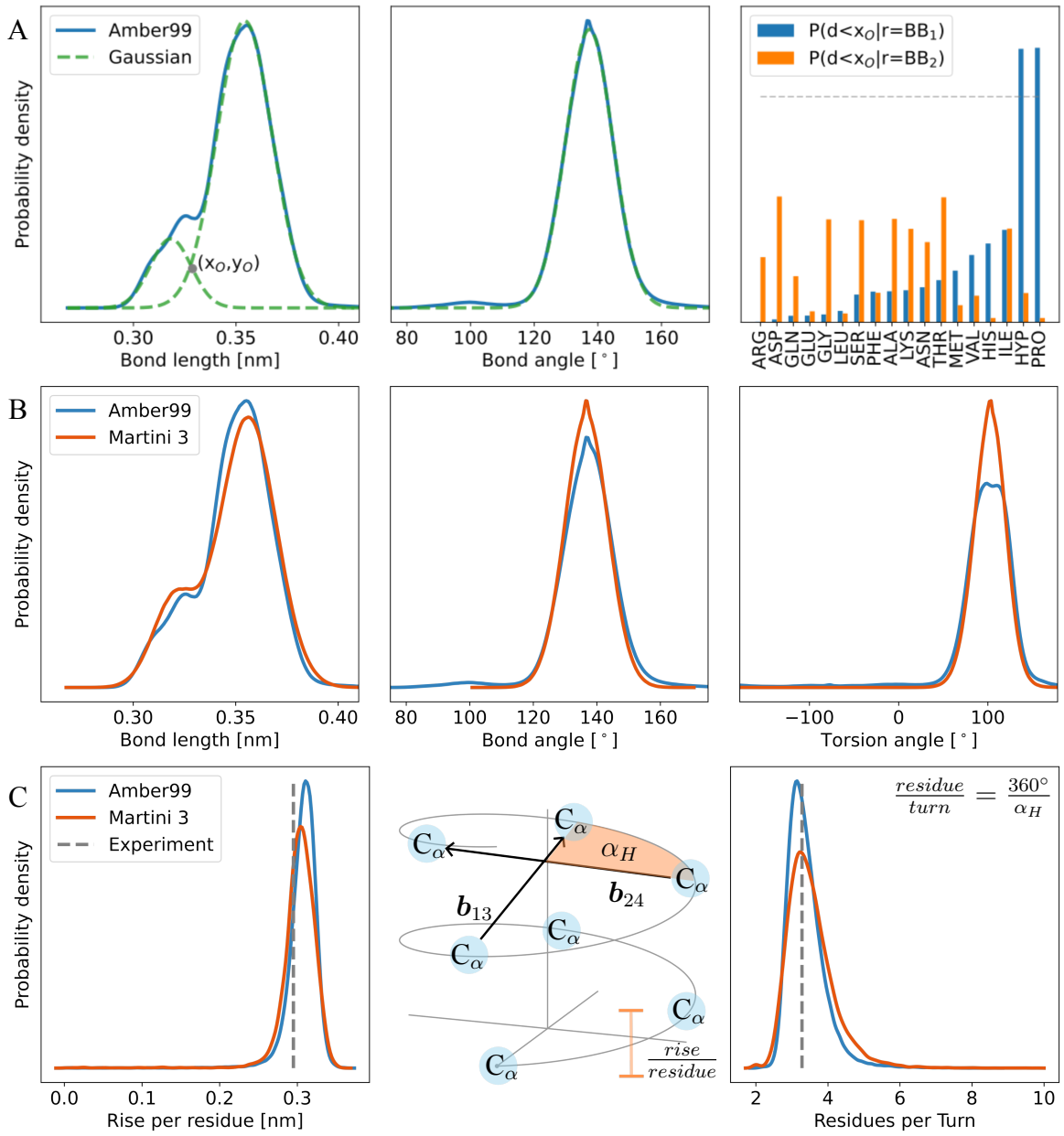

Figure 5.3. Bonded terms for the collagen triple helix A Direct Boltzmann-inversion (*green*) was performed to derive Martini 3 parameters from mapped AA trajectories (*blue*). We conditioned the bond length distribution on short distances $d < x_0$ to identify amino acid types involved in the tail of the distribution (*blue, right*). **B** Bond length (*left*), bond angle (*center*) and torsion angle (*right*) distributions from CG simulations (*orange*) are matched to AA ones (*blue*) by iterating over the Martini 3 force field parameters. **C** We performed a structural validation by comparing the rise per residue (*left*) and residues per turn (*right*) from Martini 3 with AA simulations and experiments [237–241].

By comparing the mean μ_q and variance σ_q^2 of the Gaussian function (eq. 5.1) to the Boltzmann-distribution (eq. 5.2), the reference values and force constants of each bonded term of the Martini 3 force field can be obtained in a straightforward manner. As an example, the reference distance r_{ij}^0 and force constant k_{ij}^b of the harmonic bond length potential $V_b^{CG}(r_{ij}) = k_{ij}^b \cdot (r_{ij} - r_{ij}^0)^2$ are derived according to $r_{ij}^0 = \mu_b$ and $1/k_{ij}^b = \beta\sigma_b^2$.

Due to the pronounced tail of the bond length probability distribution, we used a bimodal Gaussian, namely the sum of two Gaussians, to estimate the mean and variance of each distribution separately, and thereby take the non-equidistance of the mapped collagen backbone into account. We identified the first Gaussian, representing the tail of the distribution, at $\mu_{b,1} = 0.319\,\text{nm}$ with $\sigma_{b,1} = 0.011\,\text{nm}$, and the second one at $\mu_{b,2} = 0.355\,\text{nm}$ with $\sigma_{b,2} = 0.014\,\text{nm}$. Note that the latter represents the main part of the probability distribution. We further assigned bond distances to each amino acid type to identify those types of amino acids involved in short bond lengths. Therefore, we first determined the point of intersecting Gaussians at $x_O = 0.329\,\text{nm}$, second calculated Euclidean distances between adjacent backbone beads, and third filtered them under the condition $d < x_O$. Next, we assigned a probability to each residue type r depending on whether its backbone bead was the start BB_1 or end bead BB_2 of the bond. Hence, the conditional probability $P(d < x_O | r = \text{BB}_i)$ with $i = (1, 2)$ was constructed to quantify the likelihood of each residue type being involved in either short or long bond distances (Fig. 5.3A, *blue, right*). We found that mapped proline-like amino acids, e.g., hydroxyproline (HYP) and proline (PRO), are largely responsible for the tail of the bond length probability distribution.

Consequently, we chose two individual bond length potentials, one for arbitrary amino acid sequences and one for HYP/PRO-X combinations, where X is a random amino acid, to reproduce the bond length distribution from mapped AA simulations with the Martini 3 force field (Fig. 5.3B). By matching the bonded terms distributions from Martini 3 to mapped AA simulations, we derived a fine-tuned parameter set for our collagen model (Tab. 5.2). It is worth noting that the reference values for the bond length distributions are slightly above those derived from Boltzmann-inversion, however, still not completely matching the AA-based distributions, and thus suggesting too soft bond length potentials (see section 5.3.6).

Nonetheless, we finally validated the triple helical structure of our CG collagen model by comparing the helical pitch to AA simulations and experiments from the literature [237–242]. The helical pitch is a structural property that jointly quantifies the axial and radial degree of freedom and is given by two helical observables. The rise per residue (*left*) characterizes the axial degree of freedom of the collagen triple helix, while the residues per turn (*right*) quantifies its radial degree of freedom (Fig. 5.3C).

Table 5.2. Martini 3 force field parameters for the collagen model. Martini 3 parameters obtained from matching bonded terms distributions from CG to mapped AA simulations.

Bonded term	Reference value	Force constant
BB_1-BB_2 [a]	0.360 nm	14000 kJ mol^{-1} nm^{-2}
BB_1-BB_2 [b]	0.321 nm	24000 kJ mol^{-1} nm^{-2}
BB_1-BB_2-BB_3	138 °	152 kJ mol^{-1}
BB_1-BB_2-BB_3-BB_4	76 °	17 kJ mol^{-1}

[a] Bond length potential for all amino acid sequences, except those with HYP/PRO-X.

[b] Bond length potential for amino acid sequences with HYP/PRO-X.

First, we characterized the radial degree of freedom by counting the residues per turn (*left*) for each strand of the collagen triple helix using the protein helices analysis package from MDAnalysis [165, 243]. We more precisely selected four adjacent C_α atoms and linked them in sequence by Euclidean vectors to determine the respective bisectors $\boldsymbol{b}_{13}$ and $\boldsymbol{b}_{24}$ (*black*). Next, we calculated the dot product of the bisectors to estimate the enclosed angle, i.e., the helical twist α_H, and the residues per turn. To better compare the AA simulations with Martini 3, we determined the respective probability distribution to demonstrate that both AA and CG simulations exhibit a similar behaviour with peaks at 2.96 and 3.15 residues per turn, respectively. In comparison with experiments reported in the literature, both Martini 3 and AA simulations marginally underestimate this helical property by approximately $5\,\%$ and 10%, respectively [239, 240, 242].

Second, we again used the protein helix analysis package to describe the axial degree of freedom based on the rise per residue. We defined the local helix axis as the normal vector of the plane spanned by the two bisectors $\boldsymbol{b}_{13}$ and $\boldsymbol{b}_{24}$. We next projected the position vector between the inner two C_α atoms $\boldsymbol{r}_{23}$ onto the local helix axis, to estimate the rise per residue within groups of four C_α atoms. Once again, for validation purposes, we computed the corresponding probability distributions from AA and Martini 3 simulations. More precisely, the AA-based distribution exhibits a define peak at 0.307 nm and the Martini 3 one at 0.304 nm, which is slightly above experimental values obtained from X-ray fiber diffraction (0.290 nm) [237, 238]. Notwithstanding these deviations, we concluded that our CG collagen model captures both the axial and radial degree of freedom of each helical strand well enough, and largely reproduces the shape of the collagen triple helix.

5.3.5 Collagen Molecule under Force

We next improved our CG collagen model by fine-tuning both the mechanical bending and stretching stiffness in the entropic and enthalpic force-extension regime, such that persistence lengths from experiments and AA simulations are captured. Note that the entropic and enthalpic regime is given by the force range in which entropic and enthalpic effects dominate the biopolymer's response to mechanical load (see section 5.2.5). Specifically, we applied constant force pulling simulations of a single collagen triple helix using force constants between 30 pN to 100 pN per strand for the entropic force-extension regime, and 300 pN to 1500 pN per strand for the enthalpic force-stretching regime.

In the entropic force-extension regime, we applied the worm-like chain model (eq. 2.31) to match the persistence length of a single collagen triple helix from Martini 3 to AA simulations and experiments from the literature by adjusting the strength of the Gō-like potential $\epsilon_{G\bar{o}}$. In detail, we combined Gō-like potentials with the Martini 3 force field to maintain the triple helical structure of the collagen molecule. For this reason, we relied on the workflow and implementation proposed by Moreira and Poma, that is, first determine a Gō-like contact map from the atomic structure, second identify close contacts within the molecule, third assign virtual sites to each backbone bead, and finally introduce non-bonded Lennard-Jones interactions between virtual sites [229]. According to this, Gō-like potentials between virtual sites (*orange*) maintained the collagen triple helix during production runs by updating the positions of the respective CG backbone beads (Fig. 5.4, *black*).

By further adjusting the depth of the potential well $\epsilon_{G\bar{o}}$, the strength of the interhelical interactions can be tuned to match the persistence length from AA simulations. Hence, we selected a potential well of $\epsilon_{G\bar{o}} = 15\,\mathrm{kJ\,mol^{-1}}$ to reproduce the persistence length in the entropic force-extension regime. According to worm-like chain fits to the simulation data, the persistence length from AA (4.77 nm) and CG simulations (4.60 nm) with a respective contour length of 72.48 nm and 73.40 nm are in very close proximity (Fig. 5.4, *left*).

Experimental studies suggest persistence lengths for a 300 nm-long collagen molecule between 11 nm to 15 nm and 65 nm to 180 nm, dependent on the experimental setup (see section 5.1) [74, 75, 215]. While large persistence lengths were obtained from electron microscopy, dynamic light scattering and rheological experiments, the lower ones were measured by a combination of atomic force microscopy imaging and optical tweezers stretching with worm-like chain theory [74]. Since the latter more resemble our MD simulations, we chose this range of persistence lengths between 11 nm to 15 nm as a reference to confirm the mechanical bending stiffness of our Gō-Martini 3 collagen model.

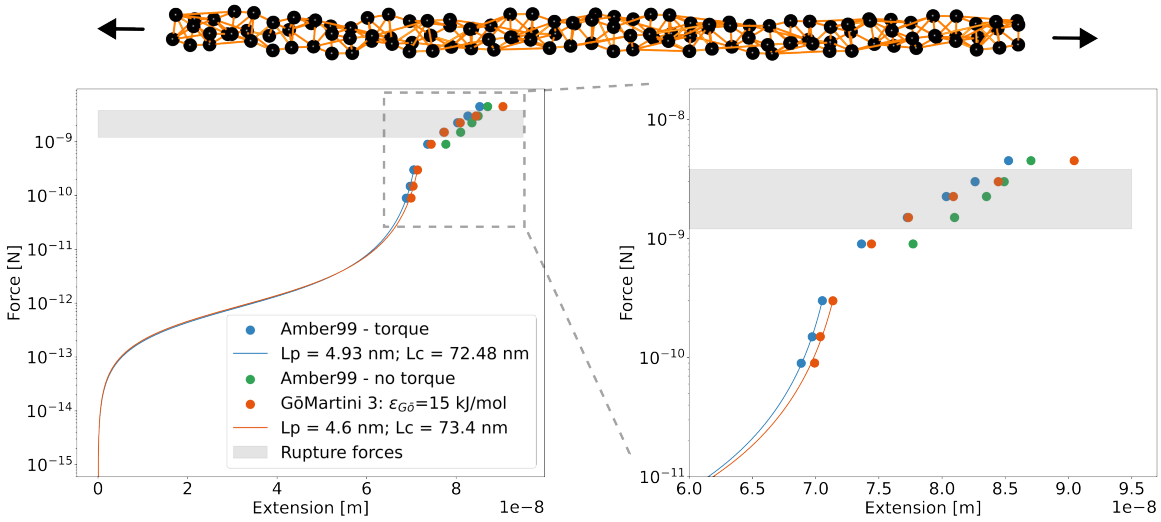

Figure 5.4. Force-extension of the collagen triple helix. Virtual sites for each collagen backbone bead (*black*) are introduced and linked by Gō-like potentials (*orange, top*). Combining such Gō-like potentials with Martini 3 (*bottom, orange*) to match the persistence length (eq. 2.30) of a single collagen molecule to AA simulations under force (*bottom, blue*). Note that the rotational restraints used by Zapp et al. to prevent the triple helix unwinding, increase the overall chain stiffness (*bottom, green* vs. *blue*) [218].

In the enthalpic force-stretching regime, we focused on pulling forces ranging from 300 pN to 1500 pN per strand, that is, where covalent bonds are likely to rupture (*gray*) and triple helices tend to unwind (Fig. 5.4, *right*). By comparing the collagen molecule force-stretching between AA and Gō-Martini 3 simulations, we observe an acceptable agreement, apart from the highest rupture force at 1500 pN. Specifically, for forces of 300 pN, 500 pN, 750 pN and 1000 pN, we calculated end-to-end distances of 73.63 nm, 77.25 nm, 80.34 nm and 82.61 nm as well as 74.41 nm, 77.33 nm, 80.09 nm and 84.44 nm from AA and CG simulations. Both models thus deviate marginally by less than 4% of the contour length, i.e., 67 nm. However, at the highest rupture force of 1500 pN applied here, the CG-based collagen triple helix stretches up to 90.45 nm, that is 5.19 nm more than the AA one.

Despite this absolute agreement, our CG model is not able to reproduce the slope of the AA-based force-stretching curve. In detail, we observe too much stretching at forces above 750 pN, suggesting that our CG model is not able to capture the bond stretching appropriately. This assumption is confirmed by looking at the bond length probability distribution (Fig. 5.3B, *left*), namely that the CG distribution is slightly wider than the AA one, suggesting that too small force constants for the bond length potentials were chosen.

Finally, it is important to comment on the triple helix unwinding, that was observed for both AA and CG simulations under force. On the atomistic scale, rotational restraints were introduced by Zapp et al. to prevent the triple helix unwinding, which, however, increases the overall stiffness of the collagen molecule (*blue*) compared to a situation without rotational restraints (Fig. 5.4, *green*) [218]. On the coarser level, we could prevent the triple helix unwinding by either choosing a deeper Gō-potential well $\epsilon_{G\bar{o}}$, which reduces the overall flexibility, or by extra rotational restraints, which restricts the time increment to 5 fs.

5.3.6 Collagen Fibril under Force

We now move from one single collagen molecule to an ensemble of more than 40 parallel aligned crosslinked triple helices forming a 67 nm-long fibrillar structure, as provided on the ColBuilder webserver [73, 203]. We chose this intermediate system size due to the D-band pattern of the collagen microfibril (see section 4.1), and its feasibility to be still simulated with AA-MD on standard high-performance computing clusters. Once again, we selected the force-stretching behaviour as our main observable for validating the mechanical stiffness of collagen. CG simulations of divalent and trivalent crosslinked collagen fibrils with and without rotational restraints were performed under constant force applying force constants ranging from 300 pN to 1500 pN and the respective end-to-end distances were measured.

Fig. 5.5A shows the force-stretching behaviour of the 67 nm-long divalent crosslinked collagen fibril from AA (*blue*) and Gō-Martini 3 simulations with (*red*) and without (*green*) additional torque restraints under force. We performed constant force pulling with force constants in the bond rupture regime, that is between 300 pN to 1500 pN per strand. More precisely, for pulling forces of 300 pN, 500 pN, 750 pN and 1000 pN per strand, we calculated end-to-end distances of 74.43 nm, 77.83 nm, 80.93 nm and 83.20 nm as well as 75.42 nm, 78.74 nm, 82.30 nm and 84.90 nm from AA- and CG-MD simulations under force, respectively, employing additional rotational restraints force constant of $2,000$ kJ mol^{-1} nm^{-2} and $1,000,000$ kJ mol^{-1} nm^{-2}, respectively. At the highest force (1500 pN per strand), our collagen CG model fails to reproduce AA simulations by more than 5 nm, which is to be expected, as protein backbone bonds are softer and likely to lengthen at high forces.

Without rotational restraints, slightly larger elongations of 75.42 nm, 82.30 nm, 84.90 nm and 91.93 nm were obtained. Similar to the previous single collagen molecule under force case, we observe an overstretching of our current Gō-Martini 3 model, in particular at high forces above 750 pN, which is still present even with such a high torque restraint force constant, that, however, restricts the time increment of CG simulations to 5 fs.

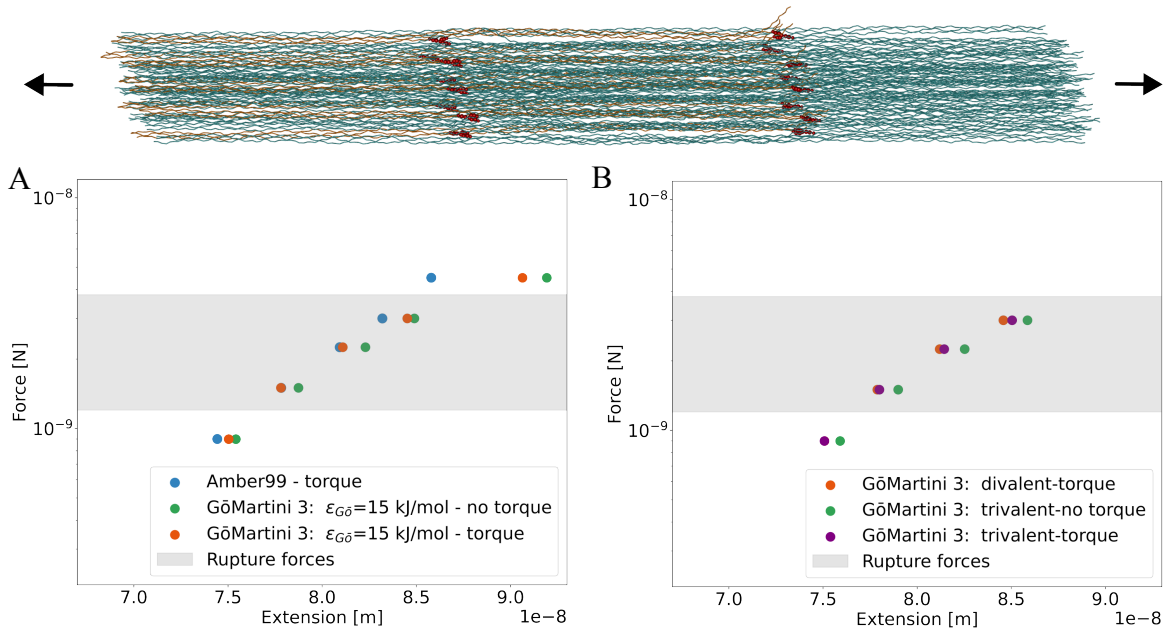

Figure 5.5. Force-stretching of the collagen fibrils. A Comparing the force-stretching of the 67 nm-long divalent crosslinked fibril from AA (*blue*) and Gō-Martini 3 simulations with (*green*) and without (*red*) torque restraints. **B** Force-stretching of the 67 nm-long divalent or trivalent crosslinked collagen fibril from Gō-Martini 3 simulations under force.

Next, we compared the force-stretching of the divalent and trivalent crosslinked collagen fibril for the exact same pulling forces using the Gō-Martini 3 model with and without rotational restraints (Fig. 5.5B). In detail, we obtained end-to-end distances of 75.07 nm, 78.00 nm, 81.44 nm and 85.05 nm for the trivalent crosslinked fibril with as well as 75.91 nm, 79.00 nm, 82.53 nm and 85.86 nm without extra torque restraints. Once more, such torque restraints increased the overall stiffness of the collagen fibril, both on the AA and CG scale, however, are not sufficient to explain the deviation between our collagen CG model and the AA simulations in the high-force regime. Of note, torque restraints were only required for rather small collagen fibrils, which lack sufficient crosslinking across triple helices to prevent unwinding, and therefore can be omitted when moving towards larger systems.

5.3.7 Collagen Microfibril under Force

Notwithstanding the weakness, that the devised Gō-Martini 3 parameterisation is still preliminary, we performed CG-MD simulations under force, for each of the three 335 nm-long collagen microfibrils presented in chapter 4, to demonstrate the feasibility of using the Martini 3 force field for such covalently crosslinked networks of semi-flexible biopolymers.

Each microfibril is composed of 230 triple helices with 177 to 248 crosslinks, forming 101 crosslinked molecules with around 43 million atoms in total. Upon coarse-graining, this reduces to less than 4.5 million beads and comes in reach for high-performance computing clusters like Helix[1]. Due to the structural complexity of the microfibril, we briefly refer to the topology generation process of the Gō-Martini 3 model in section 5.2.3.

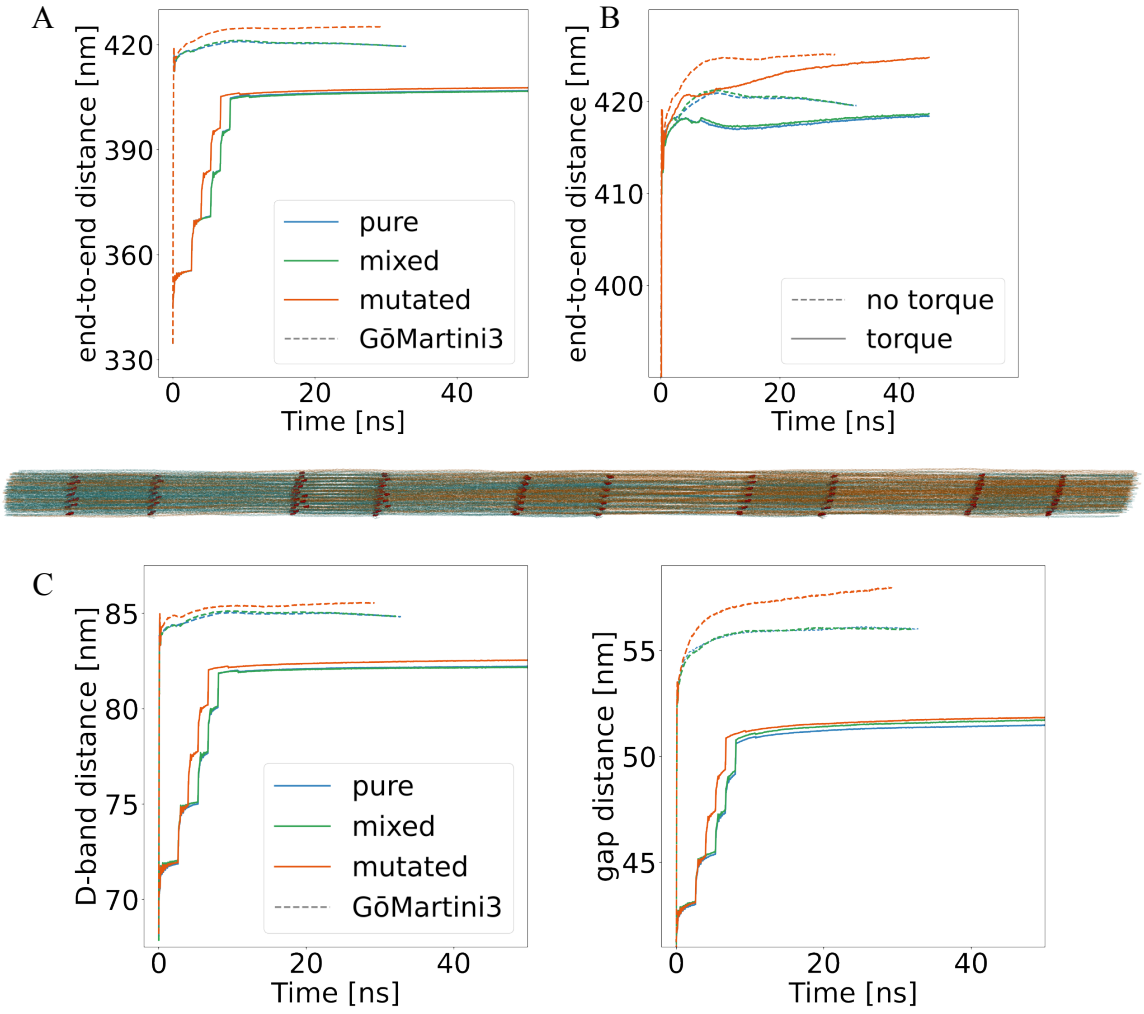

Figure 5.6. Force-stretching of the collagen microfibrils. **A** The overall end-to-end distance of the three collagen microfibrils under force (*center*) from AA simulations with (*solid*) and CG simulations without torque restraints (*dashed*). **B** The influence of torque restraints on the force-stretching of the collagen microfibrils from CG simulations. **C** Comparing the elongation of the D-band and gap region from AA to CG simulations under force.

[1]The bwFor cluster Helix is the high-performance computing cluster of the state of Baden-Württemberg.

Fig. 5.6A presents preliminary results, like the end-to-end distance (*top*), D-band and gap stretching (*bottom*) for each of the three collagen microfibrils, i.e., the pure trivalent (*blue*), the mixed divalent-trivalent (*green*) and the partly mutated trivalent crosslinked one (*red*), from AA- and CG-MD simulations with (*solid*) and without (*dashed*) rotational restraints under force. Due to the limited CG simulation time so far, we restrict ourselves to a more qualitative analysis, since we can not confirm convergence of our simulations.

Overall, we observed an excessive stretching of our collagen CG model compared to the AA simulations, which occurs regardless of the type of crosslink and distance measure, e.g., the D-band stretching and the end-to-end distance. More precisely, while our collagen CG model is able to capture slight differences on the microscopic scale, namely the influence of the crosslink density on the extension of the microfibril, it overstretches by roughly 15%, which might be due to a too soft bond length potential force constant (see section 5.3.5). Up to now, employing extra rotational restraints with the same force constants as before, seems to influence the overall stiffness of the collagen microfibril only implicitly, by post-poning the time at which maximum stretching is achieved (Fig. 5.6B).

Specifically, the pure trivalent and mixed divalent-trivalent crosslinked collagen microfibril demonstrate an overall lower elongation compared to the partly mutated trivalent crosslinked microfibril, that is evident on both AA- and CG-MD simulations under force (see section 4.3.4). As mentioned in chapter 4, this might be attributed to the reduced number of trivalent crosslinks, forming a thinner network of crosslinked collagen triple helices with fewer links, and therefore leading to a larger extension. Moreover, when comparing the gap stretching for the differently crosslinked microfibrils, it is obvious that the partly trivalent crosslinked one still elongates, suggesting that the mechanical forces are not yet evenly distributed within the fibrillar network. This could prompt the assumption that the distribution of forces is facilitated in highly crosslinked microfibrillar networks at such high mechanical loads. Obviously, this needs to be verified by monitoring longer simulation times and lower pulling forces, since our Gō-Martini 3 simulations of the differently crosslinked collagen microfibrils are not converged yet.

In other words, with decreasing number of crosslinks, forces are less equally distributed within the collagen microfibril, such that, in the event of a sudden mechanical perturbation, more locations of high-force concentration arise, and thus the likelihood of bond rupture events increases. Apparently, this is a conclusion that can only be drawn from a microscopic analysis of the bond rupture forces along the protein backbone, rather than by comparing the force-stretching of differently crosslinked microfibrils on the mesoscale.

5.4 Conclusions and Discussion

In this chapter, we derived a preliminary CG model for the semi-flexible biopolymer collagen using the Martini 3 force field, and validated our parameterisation against experimental and AA simulation data for structural and mechanical properties of a single molecule, such as the helical pitch and the entropic force-extension behaviour, besides thermodynamic and volumetric properties for the divalent (pre-mature) and trivalent (mature) crosslink.

Our Martini 3 model largely reproduces the octanol-/water partitioning and surface area of the divalent and trivalent crosslink (Fig. 5.1), and the overall shape of the collagen triple helix retrieved from AA simulations (Fig. 5.3). However, when comparing the force-stretching of the 67 nm-long collagen triple helix and divalent crosslinked fibril from Martini 3 with AA simulations, we observed strong deviations, in particular, for pulling forces above 750 pN per strand (Fig. 5.4-5.5). Nonetheless, we demonstrated that our CG model is suitable to perform MD simulations of the pure trivalent, mixed divalent-trivalent and partly mutated trivalent crosslinked collagen microfibril under force, since it captures slight differences in the force-extension observable due to varying crosslink configurations (Fig. 5.6).

We hence recommend to reparameterise the current Martini 3 force field for collagen to adequately capture the force-extension in the entropic and enthalpic regime by increasing the force constants of the bond length potentials. Most importantly, it should be critically assessed whether applying rotational restraints for preventing unwinding of the triple helices in microfibrillar structures under force are required, since they restrict the time increment of the simulations to 5 fs, and thereby cause a great loss of computational performance. In this regard, due to the limited size of the 67 nm-long collagen systems, e.g., single molecule and fibril, additional torque restraints were required, however, can be omitted when moving to larger systems (Fig. 5.6). This fact shows the capability of Martini 3 to perform simulations of large-scale collagen microfibrils under force without relying on additional constraints, that were introduced to mimic biological conditions [218]. According to this, we expect that our Martini 3 collagen model combined with the ColBuilder2 framework enables computational studies closer to its actual biological counterpart to explore the complex interplay between crosslinking and mechanical properties on long timescales (see chapter 4).

Moreover, we anticipate that this combination is suited to address questions about the influence of various crosslink configurations on the mechanical response of collagen microfibrils. Beside the crosslinks explored so far, the Martini 3 building block facilitates the introduction of yet unparameterised crosslink topologies, e.g., the trivalent pyrrole and synthetic genipin crosslink [244]. Incorporating new crosslink topologies into the collagen mi-

crofibril is expected to complement experimental studies in tissue engineering, for instance, by screening the response of various crosslinked microfibrils to mechanical load. We thus foresee that analyzing the force distribution within collagen microfibrils of various crosslink configurations might guide our understanding in engineering collagen-based scaffolds with tuneable mechanical properties to meet the divers requirements in tissue engineering for, among others, biomedical applications.

As already mentioned in the discussion about our Martini 3 PPE model in section 3.4, extending reactive MD simulations to the Martini 3 force field is the way to an in-depth understanding about the force distribution within large-scale semi-flexible (bio-)polymer networks. We predict that reactive CG-MD simulations of collagen microfibrils comprising various crosslink configurations could elucidate molecular processes that prevent macroscopic failure modes, such as the recently shown radical scavenging on dihydroxy-phenylalanine (DOPA) and the sacrificial bond ruptures in trivalent crosslinks [206, 218].

Expanding on this, by performing multiple bond ruptures within large-scale collagen microfibrils, while continuously monitoring the distribution of pairwise forces within the fibrillar network, could provide access to molecular processes governing macroscopic failure prevention, which can not be captured by AA simulations due to the long timescales involved. This is therefore expected to provide insights into potential repair mechanisms on the microscale, for instance, recombination of DOPA residues with nearby Lysine-like residues, i.e., rupture fragments from former crosslinks, to maintain the overall structural integrity and mechanical stability of collagen's fibrillar network[2].

[2]This idea was picked up from a joint discussion with Prof. Dr. Frauke Gräter in Kobe. (March 2023)

6 Summary and Conclusions

Mechanical degradation of polymer-based material under mechanical load can lead to macroscopic failure modes, such as mechanical fatigue, that is governed by microscopic processes within large semi-flexible (bio-)polymer assemblies. The aim of this thesis was to shed light on the structural alignment and response of large-scale semi-flexible (bio-)polymer networks, namely poly (*para*-phenylene ethynylene)s bulk systems and collagen microfibrils, to an external mechanical perturbation. We developed a coarse-grained model for each (bio-)polymer using the Martini 3 force field, identified locations of high-force concentration within semi-flexible PPE networks, derived an algorithm for the structural generation of collagen microfibrils with different crosslink configurations, and performed preliminary coarse-grained (CG) simulations of these collagen microfibrils under force.

Coarse-grained poly(*para*-phenylene ethynylene)s networks under shear

We developed a CG model for PPEs with the Martini 3 force field to enable computational studies of PPEs in large-scale assemblies under equilibrium and non-equilibrium conditions. To be more precise, we used an optimization geometrical approach to take the shape of the π-conjugated backbone into account, and applied an additional angular restriction to tune its mechanical bending stiffness by matching the persistence length to experiments from the literature. Our Martini 3 model reproduces key structural and thermodynamic observables of single PPE chains and mixtures, such as bulk density, interchain packing and π-stacking within bundles of semi-flexible polymers. We also show that chain entanglement increases at the expense of nematic order with growing polymer chain length (Fig. 3.8).

In addition, non-equilibrium shearing simulations revealed the formation of shear bands within bulk systems composed of intermediate- (60 monomers) and long-chain PPEs (120 monomers) leading to extreme shear rates and rupture forces in the high shear rate band (Fig. 3.9). Since this is not existent for networks of short chains (20 monomers), we concluded that intermediate- and long-chain PPEs in sheared bulk assemblies preferably break at the center until a lower chain length in the range of the persistence length is reached, upon which rupture forces are equally distributed and bond ruptures are less prominent (Fig. 3.10).

Even though shear banding was already observed by rheological experiments, the microscopic origin of this non-linearity still remains elusive [188, 191, 193, 197]. Recent mesoscopic simulations proposed flow-induced chain disentanglement and localized stochastic chain dynamics as main contributors to shear band formation in entangled polymer solutions [198, 199, 245]. This, however, neglects the influence of enthalpic effects on the response to large deformations, which play an important role for shear banding in entangled networks comprising semi-flexible PPEs (see section 3.4). Due to the chemical specification and thermodynamic-based parameterisation, we expect our Martini 3 model to provide meaningful insights into the non-linear response of semi-flexible PPE networks to mechanical load, for example, by elucidating the influence of the inter-PPE force balance on shear band formation, as proposed by Wang et al. from theoretical considerations [246].

Structure generation of collagen microfibrils

Up to now, our atomistic models for the collagen type I fibril were limited to one D-band (67 nm), such that many of the 300 nm-long triple helices were only crosslinked at one end, but not connected to other triple helices. We implemented a computational workflow, termed ColBuilder2, to build collagen microfibrils with tunable crosslink density by combining the crystal symmetry information with a lattice-based structure optimization procedure (Fig. 4.3). We demonstrated the versatility of our computational optimization scheme by generating three differently crosslinked 335 nm-long collagen microfibrils, that is one pure trivalent, one mixed divalent-trivalent and one partly mutated trivalent crosslinked collagen microfibril. We further collaborated with Jung to perform all-atom molecular dynamics (MD) simulations under force using the supercomputer at Riken to finally verify the structural properties of our atomistic models with experiments from the literature, such as end-to-end distance, D-band stretching and overlap-gap strain ratio (Fig. 4.6).

Coarse-grained collagen microfibril under force

We derived Martini 3 force field parameters for collagen microfibrils by first identifying relevant compounds of the fibrillar network, e.g., the divalent and trivalent crosslink as well as the collagen triple helix, and second by parameterising each compound individually. In detail, we combined the Martini 3 building block principle with non-equilibrium MD-based free energy calculations to select an appropriate mapping scheme and bead types for each crosslink by comparing their oil-/water partitioning and surface area with all-atom simulations (Fig. 5.1). Also, we relied on direct Boltzmann-inversion and non-equidistant bond length potentials to maintain the triple helical shape of collagen, and moreover superimposed

Gō-like potentials to refine the mechanical bending stiffness. Our Gō-Martini 3 model for collagen reproduces key structural and mechanical properties from AA simulations and experiments, such as the helical pitch and the stiff nature of the collagen triple helix, that is given by the persistence length in the entropic force-extension regime (Fig. 5.4A).

However, our parameterisation lacks the force-stretching of a single collagen molecule, fibril and microfibril in the enthalpic force-extension regime, in particular, for pulling forces above $750\,\mathrm{pN}$ (Fig. 5.4B-5.5A). At such high pulling forces, our CG model exhibits too much elongation, which can be attributed to soft bond length potentials arising from too low force constants (Fig. 5.3B, *left*). Notwithstanding this preliminary state, our Gō-Martini 3 model for collagen captures the force-stretching of the three differently crosslinked collagen microfibrils from AA simulations, and is therefore suited to determine the distribution of pairwise forces within microfibrillar networks, specifically to detect bonds between segments along the protein backbone and along crosslinks that are prone to rupture.

Furthermore, the building block principle of the Martini 3 force field facilitates the incorporation of yet unparameterised crosslink topologies, such as the trivalent pyrrole, the synthetic glutaraldehyde or genipin crosslink, to analyze their influence on the force distribution within the collagen microfibril under mechanical load [244, 247]. In this regard, we predict that our Martini 3 model could guide experimental studies, concerned with rendering collagen-based scaffolds towards certain mechanical properties, by efficiently screening the response of collagen microfibrils with various crosslink configurations to mechanical stress.

6.1 Perspective on semi-flexible Polymers under Force

The Martini 3 force field

Modeling semi-flexible polymers, e.g., collagen or PPE, with the Martini 3 force field largely focused on reproducing mechanical, structural and solvation properties from the atomistic models and experiments from the literature. In particular, modeling complex molecular structures with the Martini 3 building block principle is not straightforward. For example, while the π-conjugated backbone of PPEs was captured with a geometrical model, maintaining the collagen triple helix under force posed quiet a challenge.

In practice, solely applying the Martini 3 force field to maintain the helical shape of a particular molecule is not sufficient, such that structure-based models, e.g., elastic-network and Gō-like potentials are commonly used (see section 5.3.5) [115, 248]. Specifically, preserving the triple helix of the $67\,\mathrm{nm}$-long collagen molecule and divalent crosslinked fibril under force required additional rotational restraints at each protein termini (both AA and

CG) [218]. Obviously, such restraints introduce an unwanted bias to MD simulations, however, are required for small fibrillar systems to reduce boundary effects, e.g., triple helix unwinding, and can be omitted for larger microfibrillar systems (Fig. 5.5-5.6). This demonstrates the success of the Martini 3 force field by enabling large CG-MD simulations of various crosslinked collagen microfibrils under force on long timescales in order to get closer to their real biological counterparts on the micrometer scale.

Yet, as already known from the Martini 2 force field, modeling biopolymers with complex helical shapes (e.g. DNA) required new bead types to capture fine structural details, namely the tiny bead (see section 1.2). Uusitalo et al., for instance, combined such tiny beads with an elastic-network model to reproduce the double helical shape and persistence length of RNA from AA simulations and experiments, however, was limited to a 10 fs time increment [248]. Although, the reduced time step issue was resolved in the Martini 3 force field by reparameterised cross-interactions between beads of different size, a well-defined parameter set, and more preferably detailed instructions on modeling complex helical shapes are still missing. Promising progress in this matter is pursued by Grünewald, who aims to improve the way helices are modeled in general[1]. According to Grünewald, the future helical model in the Martini 3 force field might rely on a single CG bead, that is located at the C_α atom of the atomistic structure and linked through constraints to four virtual sites, each of which represents the N-, C-, O- and H-atom in direct proximity. Besides, this atomistic description of the protein backbone is expected to become handy when incorporating hybrid MD-schemes, like Kinetic Monte Carlo/Molecular Dynamics (KIMMDY) [200] (see section 3.4), into the Martini 3 force field to enable reactive CG-MD simulations.

Large poly(*para*-phenylene ethynylene)s bulks and collagen microfibrils under force

Expanding on this, we predict that incorporating KIMMDY into the Martini 3 force field might proof fruitful to determine rupture events in semi-flexible (bio-)polymer networks, and thereby gain insights into the molecular processes leading to macroscopic failure modes. For example, performing multiple rupture events in collagen microfibrils is assumed to affect the force distribution within the fibrillar network by, among others, altering force transmission pathways and the capability to provide as a structural scaffold. Furthermore, multiple rupture events might even reveal evolutionary-based prevention mechanisms for catastrophic failure, such as radical scavenging on DOPA subsequent to homolytic bond cleavage and possible crosslink formation with close by Lysine-like residues[2](see section 5.4) [218,

[1] Based on a personal conversation with Dr. Fabian Grünewald, core-developer of Martini 3. (17.07.2023)

[2] This idea was picked up from a joint discussion with Prof. Dr. Frauke Gräter in Kobe. (March 2023)

249]. In a way, knowledge about these evolutionary-based protection mechanisms on the molecular scale is anticipated to guide post-functionalization in tissue engineering aiming to slow down biopolymer-based matter degradation and aging processes [206].

Beside collagen, combining reactive CG-MD simulations with our Martini 3 PPE model could help elucidate the time-evolution of shear banding within networks comprising semi-flexible polymers under shear-flow on the molecular scale. In detail, we expect that through continuous bond scissions of long-chain PPEs a complex many-body network of chains with different lengths arises, that more closely resembles the cryo-milling experiments (see section 3.4). This might shed light into molecular processes governing the formation and depletion of shear banding in entangled semi-flexible polymer networks, for example, bond rupture induced chain-disentanglement leading to bulk systems containing short-chain PPEs.

Finally inspired by nature, Fitzpatrick et al. combined brownian dynamics simulations with microrheological experiments to reveal a non-linear visoelastic response of composites comprising stiff actin and soft DNA biopolymers to mechanical load [174, 250]. Accordingly, we propose the Martini 3 force field to probe blends of stiff and soft polymers, e.g., PPE and polymethyl methacrylate (see section 3.1), to observe a similar non-linearity under force. Probing the response of such composites to shear-flow on the microscale could provide meaningful insights into their processability for, e.g., additive manufacturing processes. More visionary, incorporating PPEs into composites could leverage their potential as radical sponges, e.g., similar to DOPA in collagen-based tissue [249]. This is expected to slow down mechanical degradation and could increase the lifespan of polymer-based matter under cyclic mechanical load by reducing the likelihood of mechanical failure modes, e.g., polymer fatigue due to homolytic bond scission.

Beyond the Martini 3 force field

Molecular simulations were recently disrupted by machine-learning (ML) approaches, that are expected to overcome the limitations of our physics-based simulation methods, i.e., limited computational resources to access the time and length scales required for certain biological processes [251]. Although, we now leave the well-defined boundaries of the Martini 3 force field to the promising field of ML-CG modeling, the basic steps still remain, that is identifying a mapping scheme, defining the potential energy surface and validating with higher-resolution and experimental data. While most ML-CG force fields focused on capturing the potential energy surface of the molecular structure (e.g. CGnets), only some attempted to tackle the far more complex issue of data-driven mapping schemes arising from the dimensionality reduction [252–254]. Recently, Majewski et al. introduced a single

neural-network architecture, that was able to capture the thermodynamics of twelve structurally diverse proteins, and therefore paves the way towards transferable CG force fields, that is the holy grail in the field of CG modeling [255].

Even though the current ML-CG force fields are not yet applicable to systems on the scale of the semi-flexible (bio-)polymer networks probed in this thesis, we expect such force fields to change the field of molecular simulations permanently. Furthermore, combining large language models to guide chemical intuition, e.g., an optimized mapping scheme, with quantum mechanical calculations to capture essential physics, e.g., a delocalized π-electron system, could proof fruitful in the future. Specifically, such visionary ML-based approaches might be suitable to automate CGing for high-throughput screening applications, e.g., drug discovery and design, and to overcome current limitations in MD simulations, that is balancing biomolecular processes, chemical accuracy and computational resources [256, 257].

A Appendix

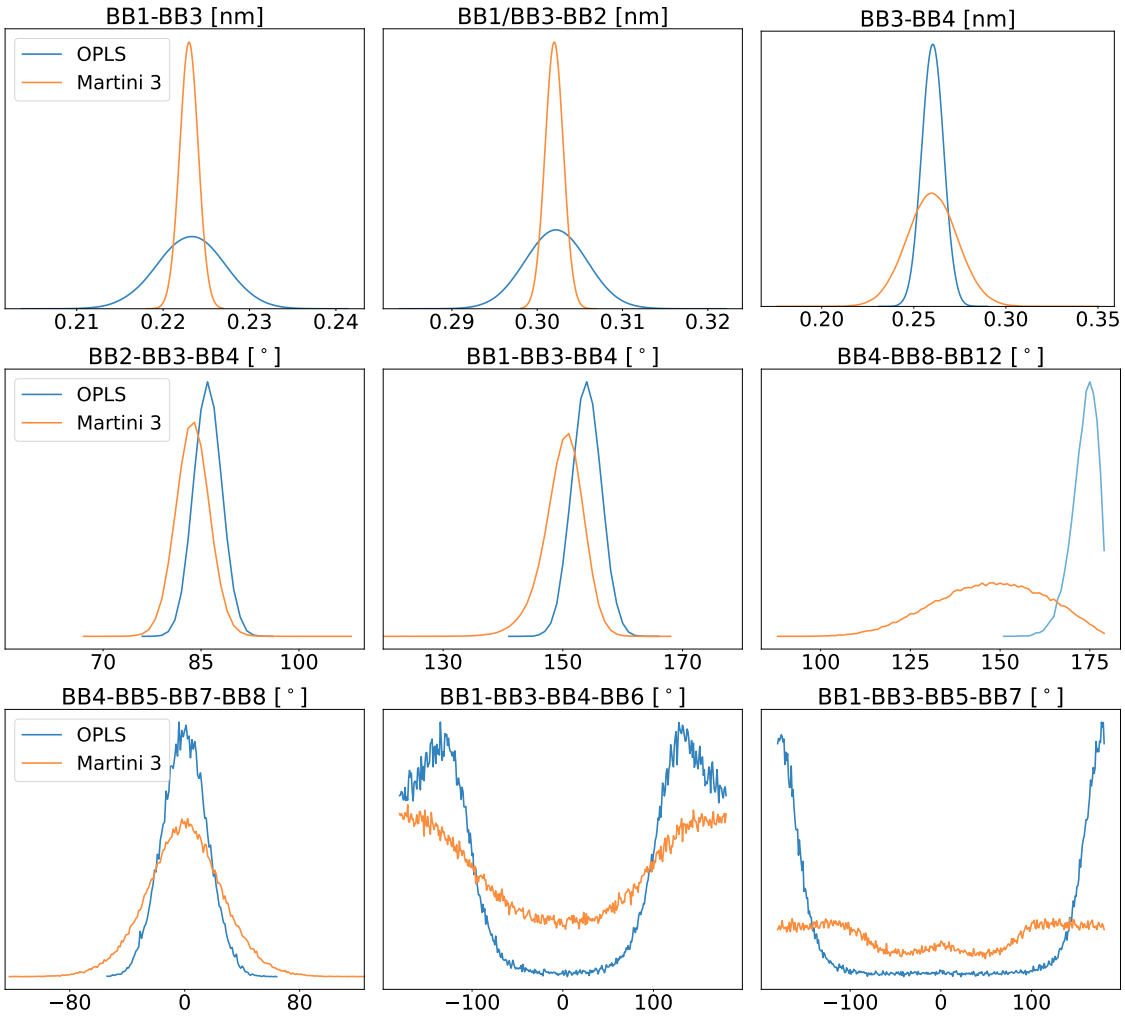

Figure A1. Matching probability densities based on a center-of-geometry mapping. Probability densities for bond length (*top*), bond angle (*middle*) and dihedral angle potentials (*bottom*) obtained from a center-of-geometry mapping. Good agreement for probability densities from Martini 3 and mapped AA simulations. However, probability densities for dihedral angles were wider than the AA ones.

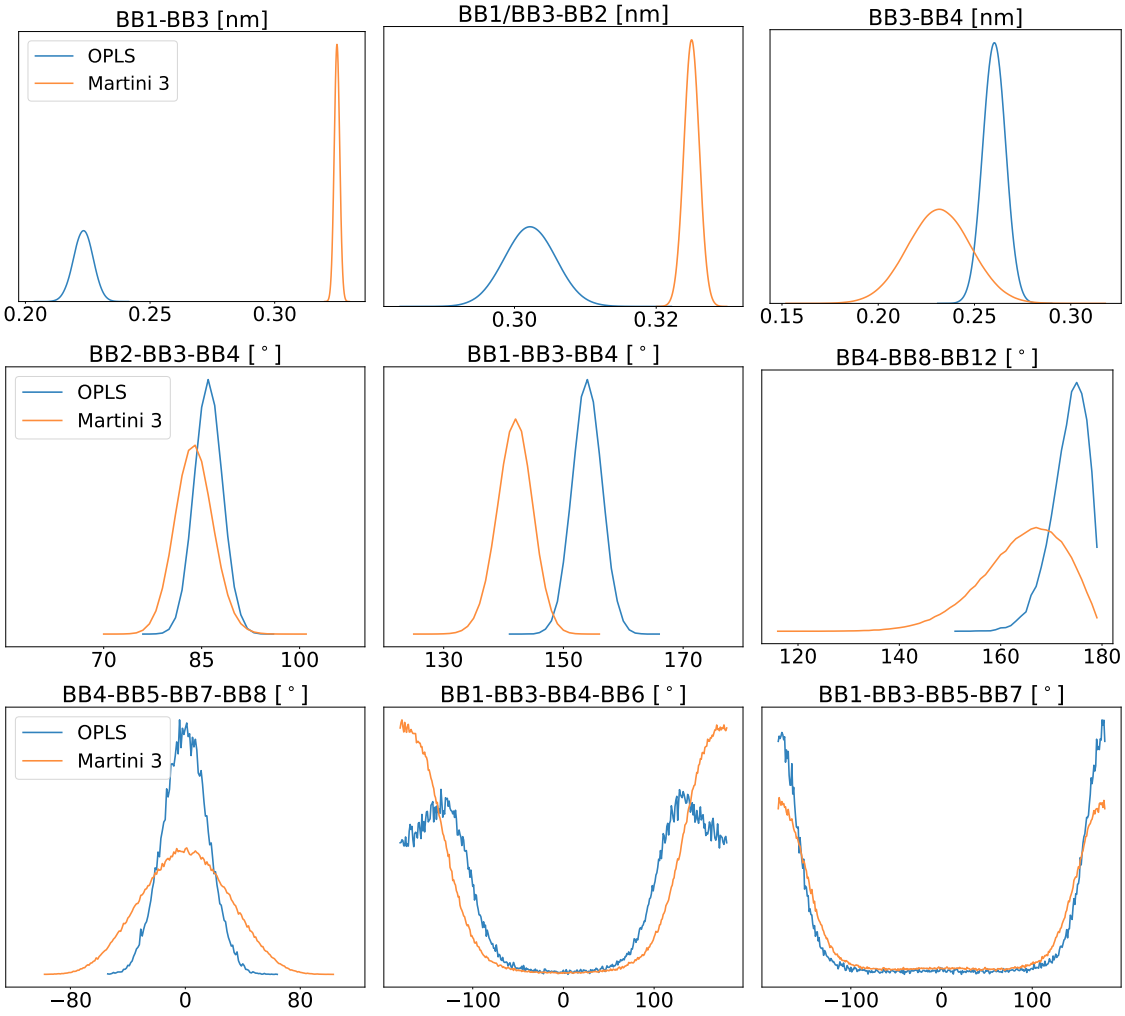

Figure A2. Comparing probability densities obtained from the geometrical model. Probability densities for bond length (*top*), bond angle (*middle*) and dihedral angle potentials (*bottom*) partly deviate from the mapped AA ones. The additional harmonic bond angle potential reduces the width of the dihedral angle probability densities.

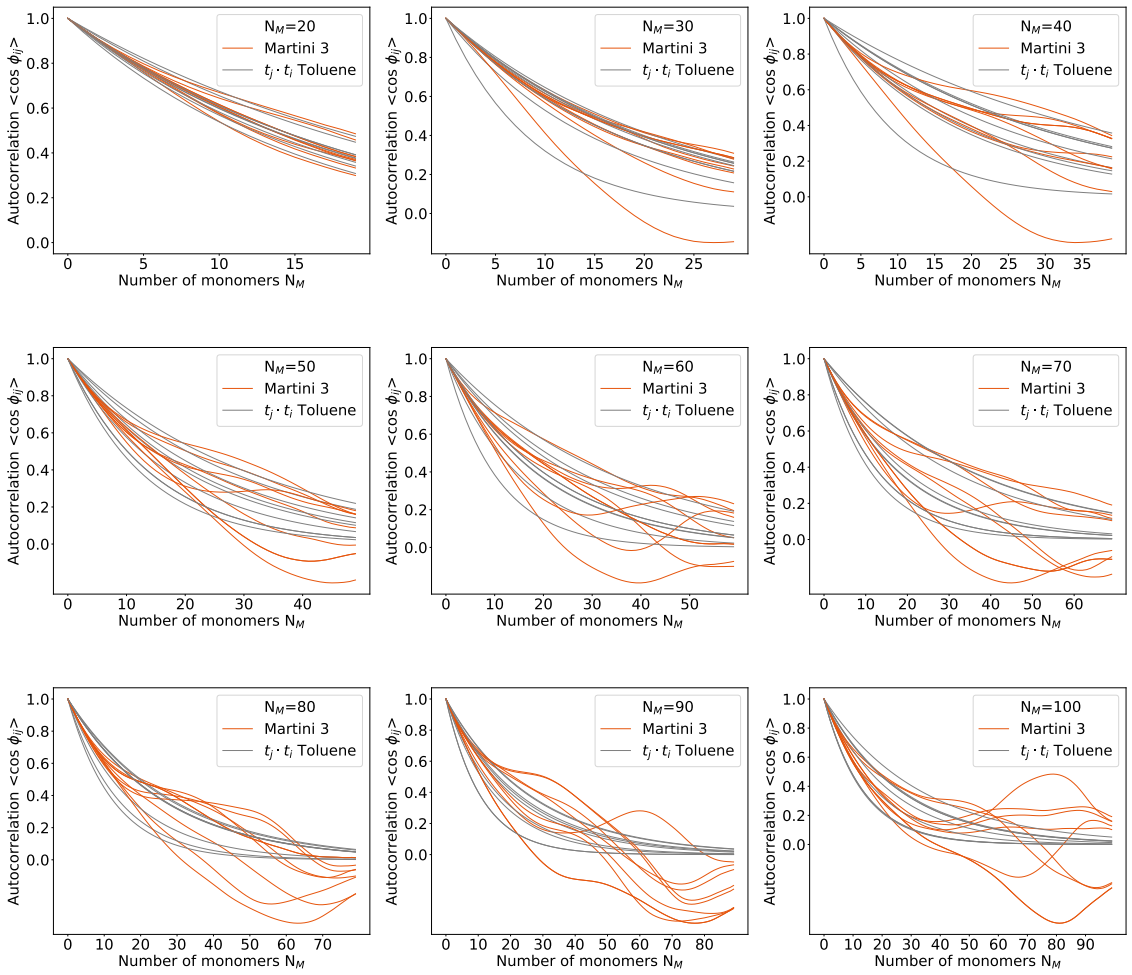

Figure A3. Persistence length for PPE in water. Fitting the unit tangent vector autocorrelation (eq. 2.29) to a Martini 3 trajectory for a single 20- to 100-monomer long PPE chain in water.

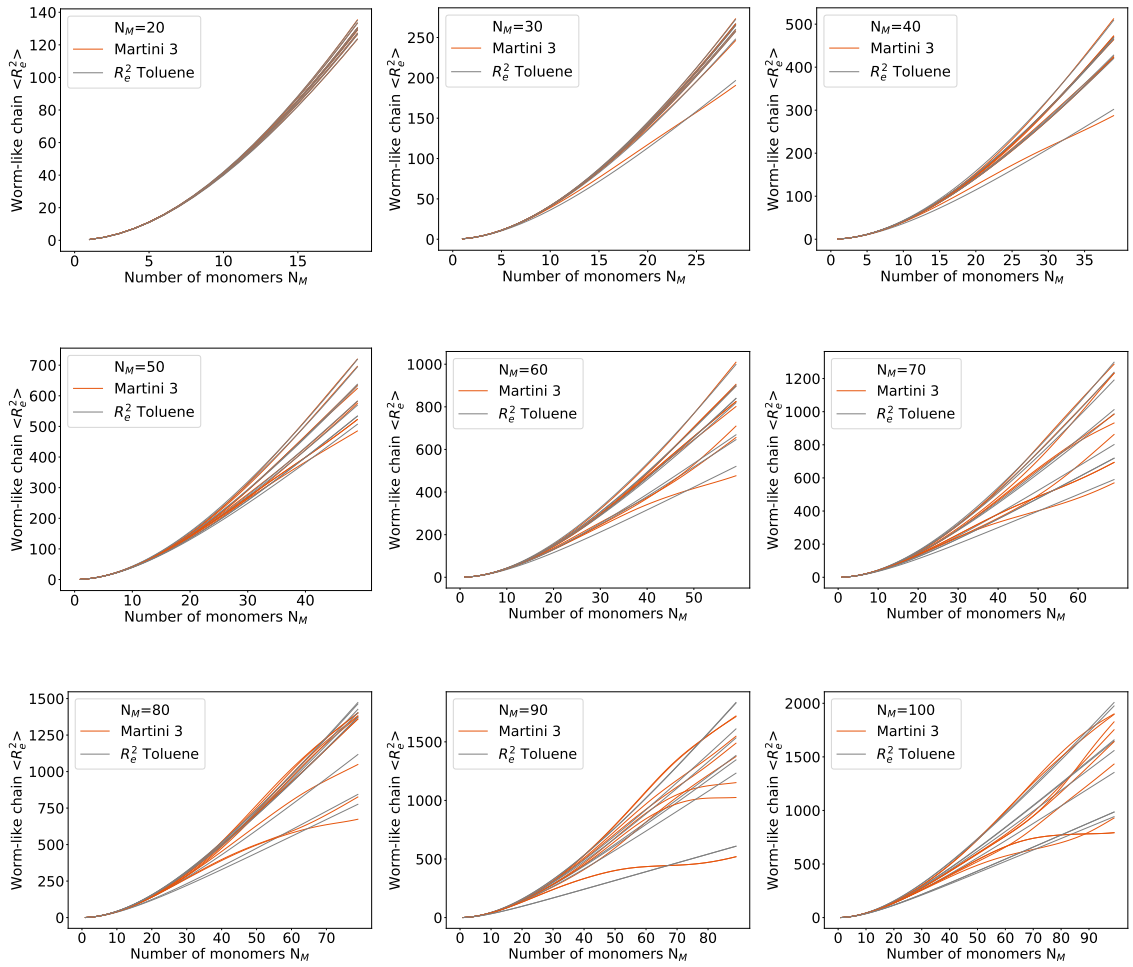

Figure A4. Persistence length for PPE in water. Fitting the squared end-to-end distances from the worm-like chain theory (eq. 2.30) to a Martini 3 trajectory for a single 20- to 100-monomer long PPE chain in water.

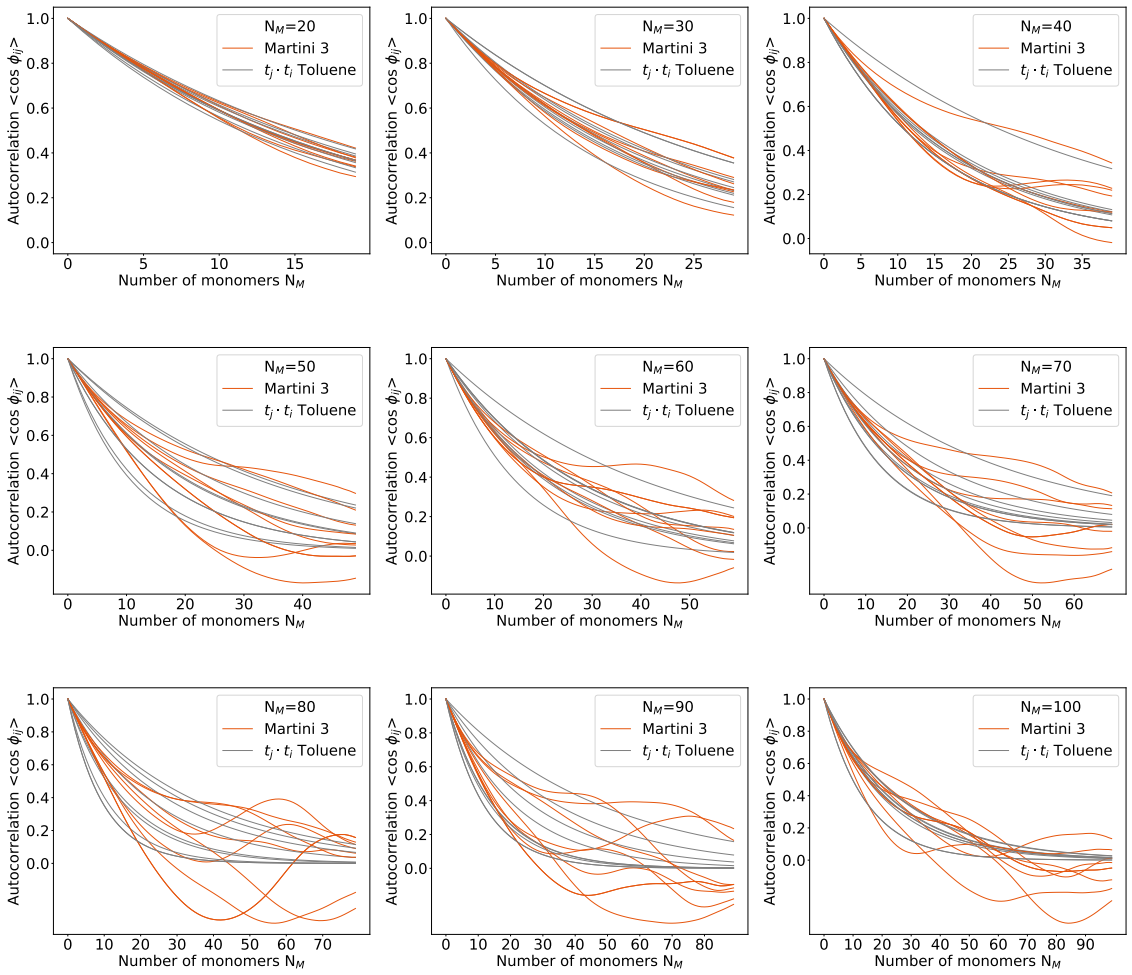

Figure A5. Persistence length for PPE in toluene. Fitting the unit tangent vector auto correlation (eq. 2.29) to a Martini 3 trajectory for a single 20- to 100-monomer long PPE chain in toluene.

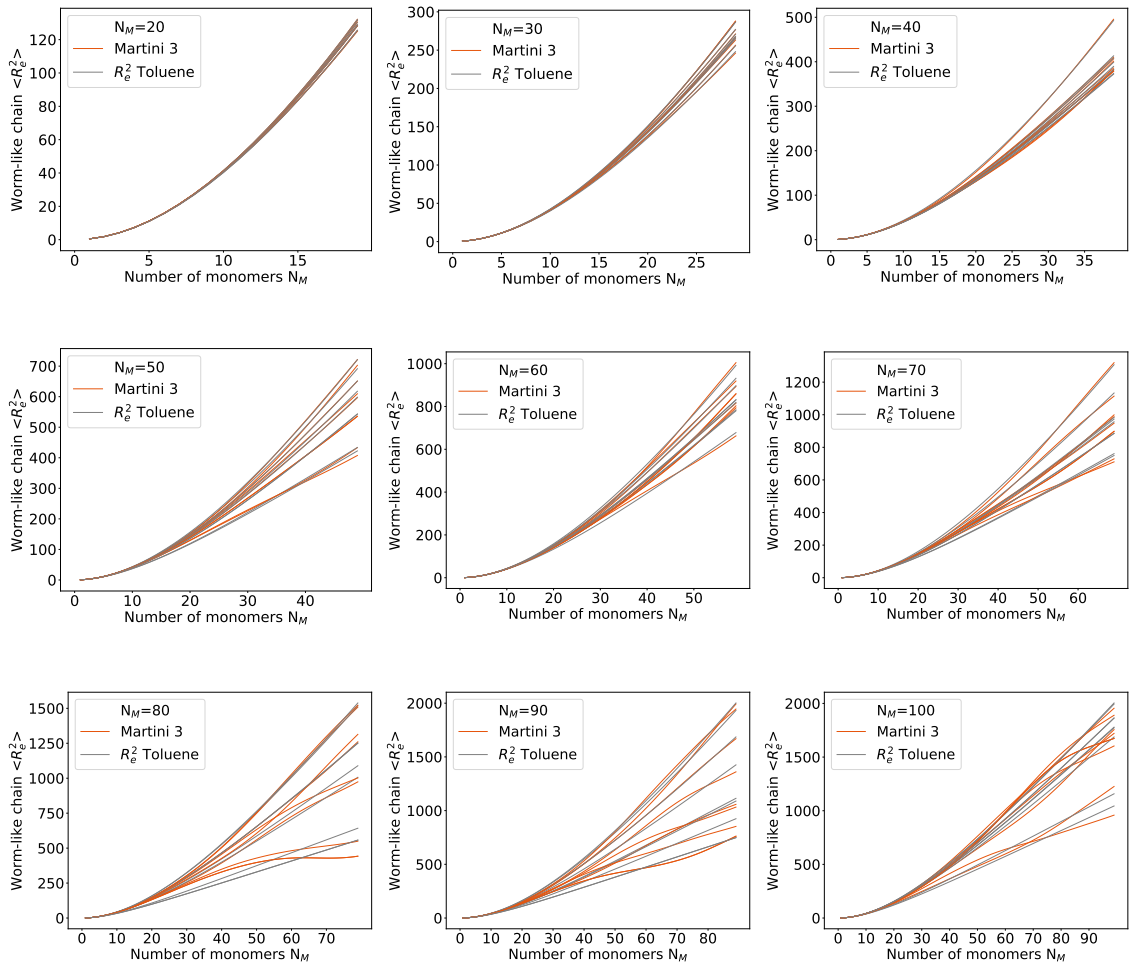

Figure A6. Persistence length for PPE in toluene. Fitting the squared end-to-end distances from the worm-like chain theory (eq. 2.30) to a Martini 3 trajectory for a single 20- to 100-monomer long PPE chain in toluene.

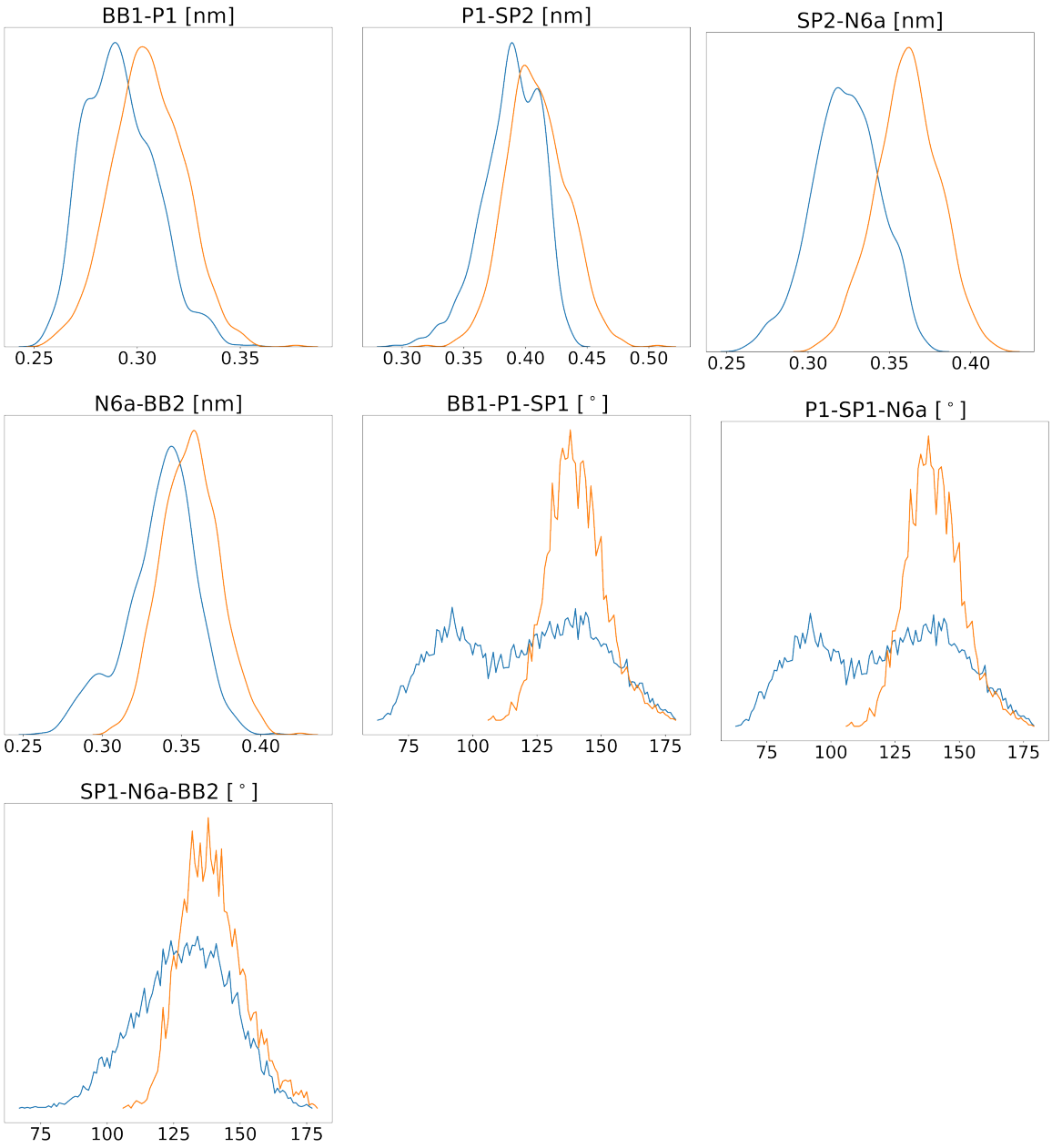

Figure A7. Bonded terms for divalent HLKNL crosslink. Probability densities for bonded terms of the divalent HLKNL crosslink from the mapped AA trajectory (*blue*) and Martini 3 simulations (*orange*).

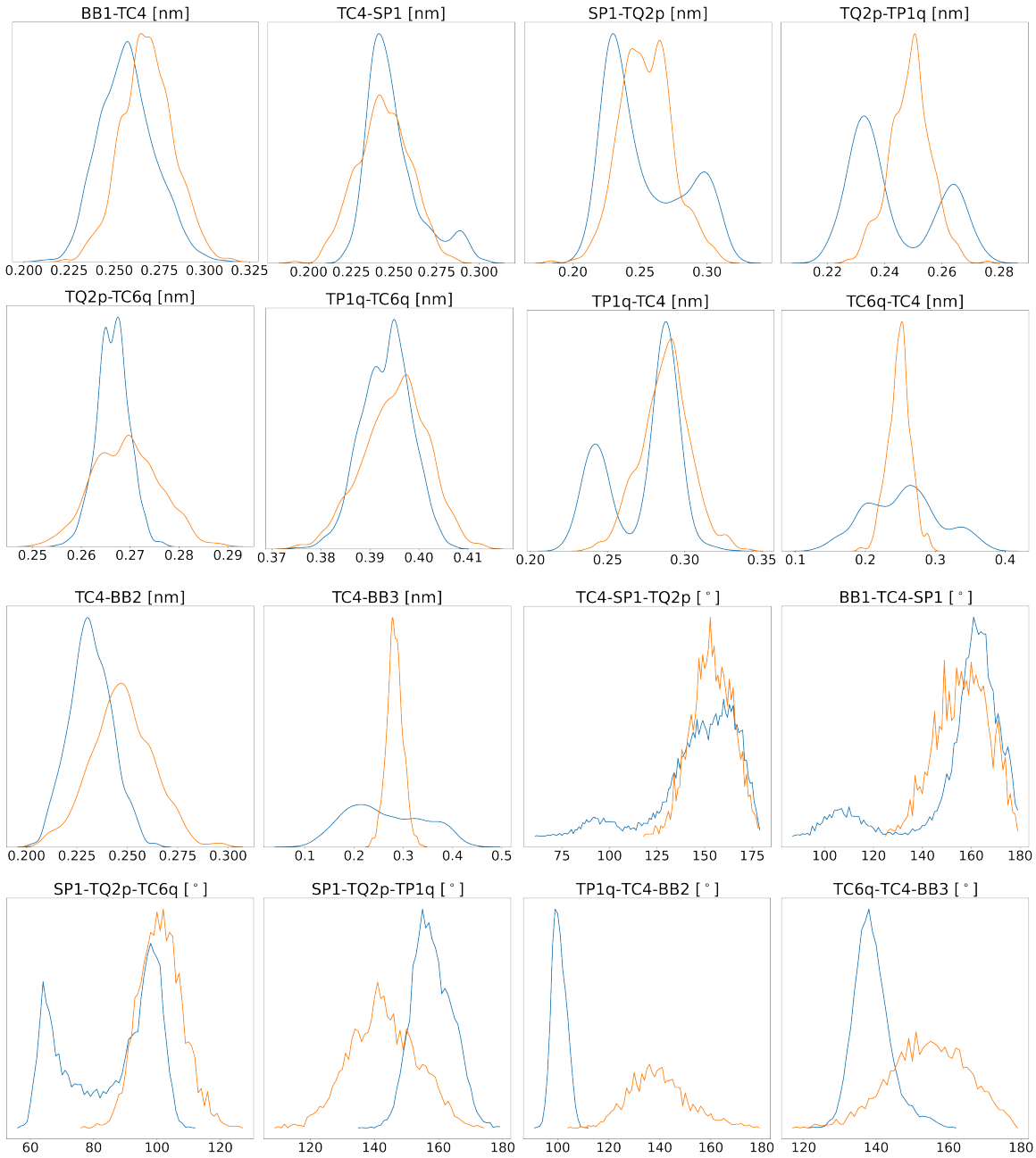

Figure A8. Bonded terms for trivalent PYD crosslink. Probability densities for bonded terms of the trivalent PYD crosslink from the mapped AA trajectory (*blue*) and Martini 3 simulations (*orange*).

Bibliography

[1] E. Frey et al. "Statistical Mechanics of Semiflexible Polymers: Theory and Experiment", *Dynamical Networks in Physics and Biology*. Springer Berlin Heidelberg, 1998. DOI: 10.1007/978-3-662-03524-5_9.

[2] Holly C. Betts et al. Integrated Genomic and Fossil Evidence Illuminates Life's Early Evolution and Eukaryote Origin, *Nat Ecol Evol*, **2**, 10, 2018. DOI: 10.1038/s41559-018-0644-x.

[3] Jan Kierfeld et al. "Semiflexible Polymers and Filaments: From Variational Problems to Fluctuations", *AIP Conference Proceedings*. Vol. 1002. AIP, 2008. DOI: 10.1063/1.2918091.

[4] Lotte Gerrits, Roel Hammink, and Paul H. J. Kouwer. Semiflexible Polymer Scaffolds: An Overview of Conjugation Strategies, *Polym. Chem.*, **12**, 10, 2021. DOI: 10.1039/D0PY01662D.

[5] Franck J. Vernerey. Mechanics of Transient Semi-Flexible Networks: Soft-elasticity, Stress Relaxation and Remodeling, *Journal of the Mechanics and Physics of Solids*, **160**, 2022. DOI: 10.1016/j.jmps.2022.104776.

[6] Enrique M. De La Cruz and Margaret L. Gardel. Actin Mechanics and Fragmentation, *Journal of Biological Chemistry*, **290**, 28, 2015. DOI: 10.1074/jbc.R115.636472.

[7] Ossama Moujaber and Ursula Stochaj. The Cytoskeleton as Regulator of Cell Signaling Pathways, *Trends in Biochemical Sciences*, **45**, 2, 2020. DOI: 10.1016/j.tibs.2019.11.003.

[8] Thomas D. Pollard and John A. Cooper. Actin, a Central Player in Cell Shape and Movement, *Science*, **326**, 5957, 2009. DOI: 10.1126/science.1175862.

[9] Anja Schmidt and Michael N. Hall. Signaling to the Actin Cytoskeleton, *Annu. Rev. Cell Dev. Biol.*, **14**, 1, 1998. DOI: 10.1146/annurev.cellbio.14.1.305.

[10] Peter Fratzl. Collagen: Structure and Mechanics. New York: Springer, 2010.

[11] Caroline Bonnans, Jonathan Chou, and Zena Werb. Remodelling the Extracellular Matrix in Development and Disease, *Nat Rev Mol Cell Biol*, **15**, 12, 2014. DOI: 10.1038/nrm3904.

[12] Paavo V. Komi. Relevance of in Vivo Force Measurements to Human Biomechanics, *Journal of Biomechanics*, **23**, 1990. DOI: 10.1016/0021-9290(90)90038-5.

[13] Ehsan Ban et al. Editorial: The Mechanobiology of Collagen Remodeling in Health and Disease, *Front. Mech. Eng.*, **9**, 2023. DOI: 10.3389/fmech.2023.1211250.

[14] P. Fratzl. "Collagen: Structure and Mechanics, an Introduction", *Collagen*. Springer US, 2008. DOI: 10.1007/978-0-387-73906-9_1.

[15] Francesco Copes et al. Collagen-Based Tissue Engineering Strategies for Vascular Medicine, *Front. Bioeng. Biotechnol.*, **7**, 2019. DOI: 10.3389/fbioe.2019.00166.

[16] Qasim A. Majid et al. Natural Biomaterials for Cardiac Tissue Engineering: A Highly Biocompatible Solution, *Front. Cardiovasc. Med.*, **7**, 2020. DOI: 10.3389/fcvm.2020.554597.

[17] Bhuvaneswari Gurumurthy and Amol V. Janorkar. Improvements in Mechanical Properties of Collagen-Based Scaffolds for Tissue Engineering, *Current Opinion in Biomedical Engineering*, **17**, 2021. DOI: 10.1016/j.cobme.2020.100253.

[18] Lauren A Hapach et al. Manipulation of *in Vitro* Collagen Matrix Architecture for Scaffolds of Improved Physiological Relevance, *Phys. Biol.*, **12**, 6, 2015. DOI: 10.1088/1478-3975/12/6/061002.

[19] C. P. Broedersz and F. C. MacKintosh. Modeling Semiflexible Polymer Networks, *Rev. Mod. Phys.*, **86**, 3, 2014. DOI: 10.1103/RevModPhys.86.995.

[20] Carsten Schuldt et al. Tuning Synthetic Semiflexible Networks by Bending Stiffness, *Phys. Rev. Lett.*, **117**, 19, 2016. DOI: 10.1103/PhysRevLett.117.197801.

[21] Anna Baun et al. Rigidification of Poly(*p* -Phenylene)s through *Ortho* -Phenyl Substitution, *Macromolecules*, **53**, 14, 2020. DOI: 10.1021/acs.macromol.0c00810.

[22] Zhiqiang Shen et al. Carbon Nanotube Length Governs the Viscoelasticity and Permeability of Buckypaper, *Polymers*, **9**, 12, 2017. DOI: 10.3390/polym9040115.

[23] Christof Ecker et al. Glassy State of Single Dendronized Polymer Chains, *Macromolecules*, **37**, 7, 2004. DOI: 10.1021/ma035079h.

[24] Katrin Wunderlich, Klaus Müllen, and George Fytas. "Shape Persistence in Polymers and Supramolecular Assemblies", Matti Knaapila. *Materials and Energy*. 2018. DOI: 10.1142/9789813225763_0001.

[25] Patricia M. Cotts, Timothy M. Swager, and Qin Zhou. Equilibrium Flexibility of a Rigid Linear Conjugated Polymer, *Macromolecules*, **29**, 23, 1996. DOI: 10.1021/ma9602583.

[26] Uwe H. F. Bunz. Poly(Aryleneethynylene)s, *Macromol. Rapid Commun.*, **30**, 9-10, 2009. DOI: 10.1002/marc.200800775.

[27] Uwe H. F. Bunz et al. Poly(Aryleneethynylene)s (PAE) as Paradigmatic Sensor Cores, *Chem. Soc. Rev.*, **44**, 13, 2015. DOI: 10.1039/C4CS00267A.

[28] Emanuel Smarsly et al. Printing Poly(*p* -Phenyleneethynylene) PLEDs, *ACS Appl. Mater. Interfaces*, **11**, 3, 2019. DOI: 10.1021/acsami.8b18827.

[29] Benhua Wang et al. Poly(*Para* -Phenyleneethynylene)-Sensor Arrays Discriminate 22 Different Teas, *ACS Sens.*, **3**, 2, 2018. DOI: 10.1021/acssensors.7b00943.

[30] O. Kratky and G. Porod. Röntgenuntersuchung Gelöster Fadenmoleküle, *Recl. Trav. Chim. Pays-Bas*, **68**, 12, 1949. DOI: 10.1002/recl.19490681203.

[31] Masao Doi and Samuel F. Edwards. The Theory of Polymer Dynamics. 73. Oxford: Clarendon Press, 2013.

[32] Michael Rubinstein and Ralph H. Colby. Polymer Physics. Oxford: Oxford Univ. Press, 2014. 440 pp.

[33] Adelheid Godt et al. How Flexible Are Poly(para-phenyleneethynylene)s?, *Angew. Chem.*, **118**, 45, 2006. DOI: 10.1002/ange.200602807.

[34] J. F. Marko and E. D. Siggia. Statistical Mechanics of Supercoiled DNA, *Phys. Rev. E*, **52**, 3, 1995. DOI: 10.1103/PhysRevE.52.2912.

[35] C. Bustamante et al. Entropic Elasticity of λ-Phage DNA, *Science*, **265**, 5178, 1994. DOI: 10.1126/science.8079175.

[36] John F. Marko and Eric D. Siggia. Stretching DNA, *Macromolecules*, **28**, 26, 1995. DOI: 10.1021/ma00130a008.

[37] Claus Heussinger, Mark Bathe, and Erwin Frey. Statistical Mechanics of Semiflexible Bundles of Wormlike Polymer Chains, *Phys. Rev. Lett.*, **99**, 4, 2007. DOI: 10.1103/PhysRevLett.99.048101.

[38] Mark Bathe et al. Cytoskeletal Bundle Mechanics, *Biophysical Journal*, **94**, 8, 2008. DOI: 10.1529/biophysj.107.119743.

[39] Klaus Kroy and Jens Glaser. The Glassy Wormlike Chain, *New J. Phys.*, **9**, 11, 2007. DOI: 10.1088/1367-2630/9/11/416.

[40] Lars Wolff, Pablo Fernandez, and Klaus Kroy. Inelastic Mechanics of Sticky Biopolymer Networks, *New J. Phys.*, **12**, 5, 2010. DOI: 10.1088/1367-2630/12/5/053024.

[41] David C. Morse. Tube Diameter in Tightly Entangled Solutions of Semiflexible Polymers, *Phys. Rev. E*, **63**, 3, 2001. DOI: 10.1103/PhysRevE.63.031502.

[42] Moumita Das, F. C. MacKintosh, and Alex J. Levine. Effective Medium Theory of Semiflexible Filamentous Networks, *Phys. Rev. Lett.*, **99**, 3, 2007. DOI: 10.1103/PhysRevLett.99.038101.

[43] Sihan Chen, Tomer Markovich, and Fred C. MacKintosh. Nonaffine Deformation of Semiflexible Polymer and Fiber Networks, *Phys. Rev. Lett.*, **130**, 8, 2023. DOI: 10.1103/PhysRevLett.130.088101.

[44] Arash Nikoubashman and Michael P. Howard. Equilibrium Dynamics and Shear Rheology of Semiflexible Polymers in Solution, *Macromolecules*, **50**, 20, 2017. DOI: 10.1021/acs.macromol.7b01876.

[45] Raghunath Chelakkot, Roland G. Winkler, and Gerhard Gompper. Migration of Semiflexible Polymers in Microcapillary Flow, *EPL*, **91**, 1, 2010. DOI: 10.1209/0295-5075/91/14001.

[46] Philipp S. Lang, Benedikt Obermayer, and Erwin Frey. Dynamics of a Semiflexible Polymer or Polymer Ring in Shear Flow, *Phys. Rev. E*, **89**, 2, 2014. DOI: 10.1103/PhysRevE.89.022606.

[47] Antonio Lamura and Roland G. Winkler. Tethered Semiflexible Polymer under Large Amplitude Oscillatory Shear, *Polymers*, **11**, 4, 2019. DOI: 10.3390/polym11040737.

[48] Antonio Lamura. Self-Attractive Semiflexible Polymers under an External Force Field, *Polymers*, **14**, 21, 2022. DOI: 10.3390/polym14214762.

[49] Dmitry G. Luchinsky et al. Welding Dynamics in an Atomistic Model of an Amorphous Polymer Blend with Polymer–Polymer Interface, *Journal of Polymer Science*, **58**, 15, 2020. DOI: 10.1002/pol.20190253.

[50] Siewert J. Marrink et al. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations, *J. Phys. Chem. B*, **111**, 27, 2007. DOI: 10.1021/jp071097f.

[51] Riccardo Alessandri et al. Pitfalls of the Martini Model, *J. Chem. Theory Comput.*, **15**, 10, 2019. DOI: 10.1021/acs.jctc.9b00473.

[52] Siewert J. Marrink, Alex H. de Vries, and Alan E. Mark. Coarse Grained Model for Semiquantitative Lipid Simulations, *J. Phys. Chem. B*, **108**, 2, 2004. DOI: 10.1021/jp036508g.

[53] Luca Monticelli et al. The MARTINI Coarse-Grained Force Field: Extension to Proteins, *J. Chem. Theory Comput.*, **4**, 5, 2008. DOI: 10.1021/ct700324x.

[54] Fabian Grünewald et al. Martini 3 Coarse-Grained Force Field for Carbohydrates, *J. Chem. Theory Comput.*, **18**, 12, 2022. DOI: 10.1021/acs.jctc.2c00757.

[55] Fabian Grünewald et al. Polyply; a Python Suite for Facilitating Simulations of Macromolecules and Nanomaterials, *Nat Commun*, **13**, 1, 2022. DOI: 10.1038/s41467-021-27627-4.

[56] Luis Itza Vazquez-Salazar et al. Martini Coarse-Grained Models of Imidazolium-Based Ionic Liquids: From Nanostructural Organization to Liquid–Liquid Extraction, *Green Chem.*, **22**, 21, 2020. DOI: 10.1039/D0GC01823F.

[57] Cécile Hilpert et al. Facilitating CG Simulations with MAD: The MArtini Database Server, *J. Chem. Inf. Model.*, **63**, 3, 2023. DOI: 10.1021/acs.jcim.2c01375.

[58] Siewert J. Marrink and D. Peter Tieleman. Perspective on the Martini Model, *Chem. Soc. Rev.*, **42**, 16, 2013. DOI: 10.1039/c3cs60093a.

[59] Paulo C. T. Souza et al. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics, *Nat Methods*, **18**, 4, 2021. DOI: 10.1038/s41592-021-01098-3.

[60] H. Jelger Risselada. Martini 3: A Coarse-Grained Force Field with an Eye for Atomic Detail, *Nat Methods*, **18**, 4, 2021. DOI: 10.1038/s41592-021-01111-9.

[61] F. Emil Thomasen et al. Improving Martini 3 for Disordered and Multidomain Proteins, *J. Chem. Theory Comput.*, **18**, 4, 2022. DOI: 10.1021/acs.jctc.1c01042.

[62] Ainara Claveras Cabezudo et al. Scaling Protein–Water Interactions in the Martini 3 Coarse-Grained Force Field to Simulate Transmembrane Helix Dimers in Different Lipid Environments, *J. Chem. Theory Comput.*, **19**, 7, 2023. DOI: 10.1021/acs.jctc.2c00950.

[63] Gül H. Zerze. Optimizing the Martini 3 Force Field Reveals the Effects of the Intricate Balance between Protein–Water Interaction Strength and Salt Concentration on Biomolecular Condensate Formation, *J. Chem. Theory Comput.*, 2023. DOI: 10.1021/acs.jctc.2c01273.

[64] Jan A. Stevens et al. Molecular Dynamics Simulation of an Entire Cell, *Front. Chem.*, **11**, 2023. DOI: 10.3389/fchem.2023.1106495.

[65] Weria Pezeshkian et al. Molecular Architecture and Dynamics of SARS-CoV-2 Envelope by Integrative Modeling, *Structure*, **31**, 4, 2023. DOI: 10.1016/j.str.2023.02.006.

[66] André P. Carvalho et al. Sticky-MARTINI as a Reactive Coarse-Grained Model for Molecular Dynamics Simulations of Silica Polymerization, *npj Comput Mater*, **8**, 1, 2022. DOI: 10.1038/s41524-022-00722-w.

[67] Selim Sami and Siewert J. Marrink. Reactive Martini: Chemical Reactions in Coarse-Grained Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, **19**, 13, 2023. DOI: 10.1021/acs.jctc.2c01186.

[68] H. Schnablegger et al. Investigations on the Morphology of Poly-1,4-(Phenylethyl Xylylene) in Solution, *Acta Polym.*, **50**, 11-12, 1999. DOI: 10.1002/(SICI)1521-4044(19991201)50:11/12<391::AID-APOL391>3.0.CO;2-2.

[69] Uwe H. F. Bunz et al. Solid-State Structures of Phenyleneethynylenes: Comparison of Monomers and Polymers, *Chem. Mater.*, **11**, 6, 1999. DOI: 10.1021/cm990036u.

[70] Behnaz Bagheri, Björn Baumeier, and Mikko Karttunen. Getting Excited: Challenges in Quantum-Classical Studies of Excitons in Polymeric Systems, *Phys. Chem. Chem. Phys.*, **18**, 44, 2016. DOI: 10.1039/C6CP02944B.

[71] Manuel Hodecker et al. Twisting and Bending Photo-Excited Phenylethynylbenzenes – a Theoretical Analysis, *Phys. Chem. Chem. Phys.*, **22**, 18, 2020. DOI: 10.1039/D0CP01662D.

[72] Sidath Wijesinghe et al. Conformation of Ionizable Poly *Para* Phenylene Ethynylene in Dilute Solutions, *J. Polym. Sci. Part B: Polym. Phys.*, **54**, 5, 2016. DOI: 10.1002/polb.23949.

[73] Joseph P. R. O. Orgel et al. Microfibrillar Structure of Type I Collagen *in Situ*, *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 24, 2006. DOI: 10.1073/pnas.0502718103.

[74] Yu-Long Sun et al. Direct Quantification of the Flexibility of Type I Collagen Monomer, *Biochemical and Biophysical Research Communications*, **295**, 2, 2002. DOI: 10.1016/S0006-291X(02)00685-X.

[75] Nagmeh Rezaei, Aaron Lyons, and Nancy R. Forde. Environmentally Controlled Curvature of Single Collagen Proteins, *Biophysical Journal*, **115**, 8, 2018. DOI: 10.1016/j.bpj.2018.09.003.

[76] Baptiste Depalle et al. Influence of Cross-Link Structure, Density and Mechanical Properties in the Mesoscale Deformation Mechanisms of Collagen Fibrils, *Journal of the Mechanical Behavior of Biomedical Materials*, **52**, 2015. DOI: 10.1016/j.jmbbm.2014.07.008.

[77] Alfonso Gautieri et al. Coarse-Grained Model of Collagen Molecules Using an Extended MARTINI Force Field, *J. Chem. Theory Comput.*, **6**, 4, 2010. DOI: 10.1021/ct100015v.

[78] B. J. Alder and T. E. Wainwright. Phase Transition for a Hard Sphere System, *The Journal of Chemical Physics*, **27**, 5, 1957. DOI: 10.1063/1.1743957.

[79] B. J. Alder and T. E. Wainwright. Studies in Molecular Dynamics. I. General Method, *The Journal of Chemical Physics*, **31**, 2, 1959. DOI: 10.1063/1.1730376.

[80] A. Rahman. Correlations in the Motion of Atoms in Liquid Argon, *Phys. Rev.*, **136**, 2A 1964. DOI: 10.1103/PhysRev.136.A405.

[81] Jaewoon Jung et al. New Parallel Computing Algorithm of Molecular Dynamics for Extremely Huge Scale Biological Systems, *J Comput Chem*, **42**, 4, 2021. DOI: 10.1002/jcc.26450.

[82] Mark Abraham et al. *GROMACS 2023.2 Source Code*. Version 2023.2. Zenodo, 2023. DOI: 10.5281/ZENODO.8134397.

[83] Florian Sebastian Franz. Molecular Dynamics of the Membrane-Talin-Vinculin Axis under Force, 2021. DOI: 10.11588/HEIDOK.00030799.

[84] Ion Bogdan Costescu. Time-Resolved Force Distribution Analysis for Molecular Communication, 2013. DOI: 10.11588/HEIDOK.00015526.

[85] Mark E. Tuckerman and Glenn J. Martyna. Understanding Modern Molecular Dynamics: Techniques and Applications, *J. Phys. Chem. B*, **104**, 2, 2000. DOI: 10.1021/jp992433y.

[86] Ulrich Schwarz. Theoretical Biophysics. Heidelberg Universiy, Institute for Theoretical Physics, 2021.

[87] Jean-Paul Ryckaert and Andre Bellemans. Molecular Dynamics of Liquid Alkanes, *Faraday Discuss. Chem. Soc.*, **66**, 1978. DOI: 10.1039/dc9786600095.

[88] Loup Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules, *Phys. Rev.*, **159**, 1, 1967. DOI: 10.1103/PhysRev.159.98.

[89] William C. Swope et al. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters, *The Journal of Chemical Physics*, **76**, 1, 1982. DOI: 10.1063/1.442716.

[90] D. C. Rapaport. The Art of Molecular Dynamics Simulation. Cambridge University Press, 2004. 549 pp.

[91] Berk Hess et al. LINCS: A Linear Constraint Solver for Molecular Simulations, *J. Comput. Chem.*, **18**, 12, 1997. DOI: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[92] Shuichi Miyamoto and Peter A. Kollman. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models, *J. Comput. Chem.*, **13**, 8, 1992. DOI: 10.1002/jcc.540130805.

[93] Hans C Andersen. Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations, *Journal of Computational Physics*, **52**, 1, 1983. DOI: 10.1016/0021-9991(83)90014-1.

[94] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale, *Ann. Phys.*, **369**, 3, 1921. DOI: 10.1002/andp.19213690304.

[95] Tom Darden, Darrin York, and Lee Pedersen. Particle Mesh Ewald: An N Log(N) Method for Ewald Sums in Large Systems, *The Journal of Chemical Physics*, **98**, 12, 1993. DOI: 10.1063/1.464397.

[96] J.A. Barker and R.O. Watts. Monte Carlo Studies of the Dielectric Properties of Water-like Models, *Molecular Physics*, **26**, 3, 1973. DOI: 10.1080/00268977300102101.

[97] R.O. Watts. Monte Carlo Studies of Liquid Water, *Molecular Physics*, **28**, 4, 1974. DOI: 10.1080/00268977400102381.

[98] R. Clausius. XVI. *On a Mechanical Theorem Applicable to Heat*, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **40**, 265, 1870. DOI: 10.1080/14786447008640370.

[99] H. J. C. Berendsen et al. Molecular Dynamics with Coupling to an External Bath, *The Journal of Chemical Physics*, **81**, 8, 1984. DOI: 10.1063/1.448118.

[100] Norbert Wiener. Collected Works with Commentaries. Mathematicians of Our Time 10, 15, 20, 23. Cambridge, Mass: MIT Press, 1976. 4 pp.

[101] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical Sampling through Velocity Rescaling, *J.Chem. Phys.*, **126**, 1, 2007. DOI: 10.1063/1.2408420.

[102] Mattia Bernetti and Giovanni Bussi. Pressure Control Using Stochastic Cell Rescaling, **153**, 11, 2020. DOI: 10.1063/5.0020514.

[103] M. Parrinello and A. Rahman. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method, **52**, 12, 1981. DOI: 10.1063/1.328693.

[104] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant Pressure Molecular Dynamics Algorithms, *The Journal of Chemical Physics*, **101**, 5, 1994. DOI: 10.1063/1.467468.

[105] A W Lees and S F Edwards. The Computer Study of Transport Processes under Extreme Conditions, *J. Phys. C: Solid State Phys.*, **5**, 15, 1972. DOI: 10.1088/0022-3719/5/15/006.

[106] Denis J. Evans and G. P. Morriss. Nonlinear-Response Theory for Steady Planar Couette Flow, *Phys. Rev. A*, **30**, 3, 1984. DOI: 10.1103/PhysRevA.30.1528.

[107] Michael Levitt and Arieh Warshel. Computer Simulation of Protein Folding, *Nature*, **253**, 5494, 1975. DOI: 10.1038/253694a0.

[108] Michael Levitt. Birth and Future of Multiscale Modeling for Macromolecular Systems (Nobel Lecture), *Angew. Chem. Int. Ed.*, **53**, 38, 2014. DOI: 10.1002/anie.201403691.

[109] Karol Wołek, Àngel Gómez-Sicilia, and Marek Cieplak. Determination of Contact Maps in Proteins: A Combination of Structural and Chemical Approaches, *The Journal of Chemical Physics*, **143**, 24, 2015. DOI: 10.1063/1.4929599.

[110] Adolfo B. Poma, Marek Cieplak, and Panagiotis E. Theodorakis. Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins, *J. Chem. Theory Comput.*, **13**, 3, 2017. DOI: 10.1021/acs.jctc.6b00986.

[111] Riccardo Alessandri, Fabian Grünewald, and Siewert J. Marrink. The Martini Model in Materials Science, *Adv. Mater.*, **33**, 24, 2021. DOI: 10.1002/adma.202008635.

[112] Fabian Grunewald et al. Transferable MARTINI Model of Poly(Ethylene Oxide), *J. Phys. Chem. B*, **122**, 29, 2018. DOI: 10.1021/acs.jpcb.8b04760.

[113] Jaakko J. Uusitalo et al. Martini Coarse-Grained Force Field: Extension to DNA, *J. Chem. Theory Comput.*, **11**, 8, 2015. DOI: 10.1021/acs.jctc.5b00286.

[114] Frank Eisenhaber et al. The Double Cubic Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies, *J. Comput. Chem.*, **16**, 3, 1995. DOI: 10.1002/jcc.540160303.

[115] Xavier Periole et al. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition, *J. Chem. Theory Comput.*, **5**, 9, 2009. DOI: 10.1021/ct9002114.

[116] Shoji Takada. Gō Model Revisited, *BIOPHYSICS*, **16**, 0, 2019. DOI: 10.2142/biophysico.16.0_248.

[117] Hiroshi Taketomi, Yuzo Ueda, and Nobuhiro Gō. Studies of Protein Folding, Unfolding and Fluctuations by Computer Simulation: The Effect of Specific Amino Acid Sequence Represented by Specific Inter-Unit Interactions, *International Journal of Peptide and Protein Research*, **7**, 6, 2009. DOI: 10.1111/j.1399-3011.1975.tb02465.x.

[118] Vladimir Sobolev et al. Molecular Docking Using Surface Complementarity: Docking Using Surface Complementary, *Proteins*, **25**, 1, 1996. DOI: 10.1002/(SICI)1097-0134(199605)25:1<120::AID-PROT10>3.0.CO;2-M.

[119] Ch Chipot and A. Pohorille, eds. Free Energy Calculations: Theory and Applications in Chemistry and Biology. Springer Series in Chemical Physics 86. Berlin ; New York: Springer, 2007. 517 pp.

[120] John G. Kirkwood. Statistical Mechanics of Fluid Mixtures, *The Journal of Chemical Physics*, **3**, 5, 1935. DOI: 10.1063/1.1749657.

[121] Johannes Kästner. Umbrella Sampling: Umbrella Sampling, *WIREs Comput Mol Sci*, **1**, 6, 2011. DOI: 10.1002/wcms.66.

[122] Shankar Kumar et al. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method, *J. Comput. Chem.*, **13**, 8, 1992. DOI: 10.1002/jcc.540130812.

[123] Benoît Roux. The Calculation of the Potential of Mean Force Using Computer Simulations, *Computer Physics Communications*, **91**, 1-3, 1995. DOI: 10.1016/0010-4655(95)00053-I.

[124] Jochen S. Hub, Bert L. de Groot, and David van der Spoel. G-Wham A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates, *J. Chem. Theory Comput.*, **6**, 12, 2010. DOI: 10.1021/ct100494z.

[125] Lucas Diedrich et al. *Energetics and Kinetics of Membrane Permeation of Photoresists for Bioprinting*. preprint. Biophysics, 2023. DOI: 10.1101/2023.03.27.534360.

[126] Daniel Seeliger and Bert L. de Groot. Protein Thermostability Calculations Using Alchemical Free Energy Simulations, *Biophysical Journal*, **98**, 10, 2010. DOI: 10.1016/j.bpj.2010.01.051.

[127] Gavin E. Crooks. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems, *Journal of Statistical Physics*, **90**, 5/6, 1998. DOI: 10.1023/A:1023208217925.

[128] Charles H Bennett. Efficient Estimation of Free Energy Differences from Monte Carlo Data, *Journal of Computational Physics*, **22**, 2, 1976. DOI: 10.1016/0021-9991(76)90078-4.

[129] Michael R. Shirts et al. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods, *Phys. Rev. Lett.*, **91**, 14, 2003. DOI: 10.1103/PhysRevLett.91.140601.

[130] D. Beysens and G. Forgacs, eds. Dynamical Networks in Physics and Biology. Centre de Physique Des Houches 10. Springer, 1998.

[131] C. Bouchiat et al. Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements, *Biophysical Journal*, **76**, 1, 1999. DOI: 10.1016/S0006-3495(99)77207-3.

[132] Theo Odijk. Stiff Chains and Filaments under Tension, *Macromolecules*, **28**, 20, 1995. DOI: 10.1021/ma00124a044.

[133] Matthias Brosz et al. Martini 3 Coarse-Grained Force Field for Poly(*Para* -Phenylene Ethynylene)s, *Phys. Chem. Chem. Phys.*, **24**, 17, 2022. DOI: 10.1039/D1CP04237H.

[134] Uwe H. F. Bunz. Poly(*p*- Phenyleneethynylene)s by Alkyne Metathesis, *Acc. Chem. Res.*, **34**, 12, 2001. DOI: 10.1021/ar010092c.

[135] Robert Deans et al. A Poly(*p* -Phenyleneethynylene) with a Highly Emissive Aggregated Phase, *J. Am. Chem. Soc.*, **122**, 35, 2000. DOI: 10.1021/ja0007298.

[136] Fabian Markl et al. Printing PPEs: Fundamental Structure–Property Relationships, *ACS Macro Lett.*, **3**, 8, 2014. DOI: 10.1021/mz500340t.

[137] Jordan H. Wosnick, Charlene M. Mello, and Timothy M. Swager. Synthesis and Application of Poly(Phenylene Ethynylene)s for Bioconjugation: A Conjugated Polymer-Based Fluorogenic Probe for Proteases, *J. Am. Chem. Soc.*, **127**, 10, 2005. DOI: 10.1021/ja043134b.

[138] Joong Ho Moon and Timothy M. Swager. Poly(*p* -Phenylene Ethynylene) Brushes, *Macromolecules*, **35**, 16, 2002. DOI: 10.1021/ma025539r.

[139] Joong Ho Moon et al. Live-Cell-Permeable Poly(p-Phenylene Ethynylene), *Angew. Chem. Int. Ed.*, **46**, 43, 2007. DOI: 10.1002/anie.200701991.

[140] B. Bagheri, M. Karttunen, and B. Baumeier. Solvent Effects on Optical Excitations of Poly Para Phenylene Ethynylene Studied by QM/MM Simulations Based on Many-Body Green's Functions Theory, *Eur. Phys. J. Spec. Top.*, **225**, 8-9, 2016. DOI: 10.1140/epjst/e2016-60144-5.

[141] Ferry Kienberger et al. Static and Dynamical Properties of Single Poly(Ethylene Glycol) Molecules Investigated by Force Spectroscopy, *Single Mol.*, **1**, 2, 2000. DOI: 10.1002/1438-5171(200006)1:2<123::AID-SIMO123>3.0.CO;2-3.

[142] Ramnath Ramachandran et al. Persistence Length of Short-Chain Branched Polyethylene, *Macromolecules*, **41**, 24, 2008. DOI: 10.1021/ma801775n.

[143] C. A. Tweedie et al. Enhanced Stiffness of Amorphous Polymer Surfaces under Confinement of Localized Contact Loads, *Adv. Mater.*, **19**, 18, 2007. DOI: 10.1002/adma.200602846.

[144] D. Y. Yoon and P. J. Flory. Persistence Vectors for Polypropylene, Polystyrene, and Poly(Methyl Methacrylate) Chains, *J. Polym. Sci. Polym. Phys. Ed.*, **14**, 8, 1976. DOI: 10.1002/pol.1976.180140806.

[145] Sabina Maskey et al. Structure of Rigid Polymers Confined to Nanoparticles: Molecular Dynamics Simulations Insight, *Langmuir*, **32**, 8, 2016. DOI: 10.1021/acs.langmuir.5b04568.

[146] James Sangster. Octanol□Water Partition Coefficients of Simple Organic Compounds, *Journal of Physical and Chemical Reference Data*, **18**, 3, 1989. DOI: 10.1063/1.555833.

[147] Corwin Hansch, Albert Leo, and D. H. Hoekman, eds. Exploring QSAR. ACS Professional Reference Book. American Chemical Society, 1995.

[148] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation, *Computer Physics Communications*, **91**, 1-3, 1995. DOI: 10.1016/0010-4655(95)00042-E.

[149] Mark James Abraham et al. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers, *SoftwareX*, **1–2**, 2015. DOI: 10.1016/j.softx.2015.06.001.

[150] Katsuhiko Okuyama et al. Electronic Spectra of Tolan in a Supersonic Free Jet, *J. Phys. Chem.*, **88**, 9, 1984. DOI: 10.1021/j150653a010.

[151] Carrie E. Halkyard et al. Evidence of Aggregate Formation for 2,5-Dialkylpoly(*p* -Phenyleneethynylenes) in Solution and Thin Films, *Macromolecules*, **31**, 25, 1998. DOI: 10.1021/ma9812332.

[152] Shinji Toyota. Rotational Isomerism Involving Acetylene Carbon, *Chem. Rev.*, **110**, 9, 2010. DOI: 10.1021/cr1000628.

[153] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The Missing Term in Effective Pair Potentials, *J. Phys. Chem.*, **91**, 24, 1987. DOI: 10.1021/j100308a038.

[154] Szilárd Páll and Berk Hess. A Flexible Algorithm for Calculating Pair Interactions on SIMD Architectures, *Computer Physics Communications*, **184**, 12, 2013. DOI: 10.1016/j.cpc.2013.06.003.

[155] N. Goga et al. Efficient Algorithms for Langevin and DPD Dynamics, *J. Chem. Theory Comput.*, **8**, 10, 2012. DOI: 10.1021/ct3000876.

[156] Djurre H. de Jong et al. Martini Straight: Boosting Performance Using a Shorter Cutoff and GPUs, *Computer Physics Communications*, **199**, 2016. DOI: 10.1016/j.cpc.2015.09.014.

[157] *Martini Coarse-Grained Force Field for Biomolecules*.

[158] Fabian Grünewald et al. “Protocol for Simulations of PEGylated Proteins with Martini 3”, *Structural Genomics*. Vol. 2199. 2021. DOI: 10.1007/978-1-0716-0892-0_18.

[159] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing, *Science*, **220**, 4598, 1983. DOI: 10.1126/science.220.4598.671.

[160] Riccardo Alessandri et al. Bulk Heterojunction Morphologies with Atomistic Resolution from Coarse-Grain Solvent Evaporation Simulations, *J. Am. Chem. Soc.*, **139**, 10, 2017. DOI: 10.1021/jacs.6b11717.

[161] Riccardo Alessandri et al. Resolving Donor–Acceptor Interfaces and Charge Carrier Energy Levels of Organic Semiconductors with Polar Side Chains, *Adv. Funct. Mater.*, **30**, 46, 2020. DOI: 10.1002/adfm.202004799.

[162] Chen-Hui Shu et al. On-Surface Synthesis of Poly(p-Phenylene Ethynylene) Molecular Wires via in Situ Formation of Carbon-Carbon Triple Bond, *Nat Commun*, **9**, 1, 2018. DOI: 10.1038/s41467-018-04681-z.

[163] Caitlin C. Bannan et al. Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water, *J. Chem. Theory Comput.*, **12**, 8, 2016. DOI: 10.1021/acs.jctc.6b00449.

[164] Naomi Nitschke, Kalina Atkovska, and Jochen S. Hub. Accelerating Potential of Mean Force Calculations for Lipid Membrane Permeation: System Size, Reaction Coordinate, Solute-Solute Distance, and Cutoffs, *The Journal of Chemical Physics*, **145**, 12, 2016. DOI: 10.1063/1.4963192.

[165] Richard Gowers et al. “MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations”, Python in Science Conference. Austin, Texas, 2016. DOI: 10.25080/Majora-629e541a-00e.

[166] Charles R. Harris et al. Array Programming with NumPy, *Nature*, **585**, 7825, 2020. DOI: 10.1038/s41586-020-2649-2.

[167] Hsiao-Ping Hsu, Wolfgang Paul, and Kurt Binder. Standard Definitions of Persistence Length Do Not Describe the Local “Intrinsic” Stiffness of Real Polymer Chains, *Macromolecules*, **43**, 6, 2010. DOI: 10.1021/ma902715e.

[168] P Gutjahr, R Lipowsky, and J Kierfeld. Persistence Length of Semiflexible Polymers and Bending Rigidity Renormalization, *Europhys. Lett.*, **76**, 6, 2006. DOI: 10.1209/epl/i2006-10390-3.

[169] H. Benoit and P. Doty. Light Scattering from Non-Gaussian Chains, *J. Phys. Chem.*, **57**, 9, 1953. DOI: 10.1021/j150510a025.

[170] Pauli Virtanen, Ralf Gommers, and Travis E. Oliphant. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nat Methods*, **17**, 3, 2020. DOI: 10.1038/s41592-019-0686-2.

[171] James B. Carrell. Groups, Matrices, and Vector Spaces: A Group Theoretic Approach to Linear Algebra. New York, NY: Springer, 2017. 410 pp.

[172] Michael Waskom. Seaborn: Statistical Data Visualization, *JOSS*, **6**, 60, 2021. DOI: 10.21105/joss.03021.

[173] Mark E. Tuckerman. Statistical Mechanics: Theory and Molecular Simulation. Oxford ; New York: Oxford University Press, 2010. 696 pp.

[174] Robert Fitzpatrick et al. Synergistic Interactions Between DNA and Actin Trigger Emergent Viscoelastic Behavior, *Phys. Rev. Lett.*, **121**, 25, 2018. DOI: 10.1103/PhysRevLett.121.257801.

[175] Tristan Bereau and Kurt Kremer. Automated Parametrization of the Coarse-Grained Martini Force Field for Small Organic Molecules, *J. Chem. Theory Comput.*, **11**, 6, 2015. DOI: 10.1021/acs.jctc.5b00056.

[176] James A. Graham, Jonathan W. Essex, and Syma Khalid. PyCGTOOL: Automated Generation of Coarse-Grained Molecular Dynamics Models from Atomistic Trajectories, *J. Chem. Inf. Model.*, **57**, 4, 2017. DOI: 10.1021/acs.jcim.7b00096.

[177] Charly Empereur-Mot et al. *Swarm-CG* : Automatic Parametrization of Bonded Terms in MARTINI-Based Coarse-Grained Models of Simple to Complex Molecules *via* Fuzzy Self-Tuning Particle Swarm Optimization, *ACS Omega*, **5**, 50, 2020. DOI: 10.1021/acsomega.0c05469.

[178] G.D. Wignall, D.G.H. Ballard, and J. Schelten. Measurements of Persistence Length and Temperature Dependence of the Radius of Gyration in Bulk Atactic Polystyrene, *European Polymer Journal*, **10**, 9, 1974. DOI: 10.1016/0014-3057(74)90142-6.

[179] Vikhyaat Ahlawat, Surya Pratap S. Deopa, and Shivprasad Patil. Quantitative Elasticity of Flexible Polymer Chains Using Interferometer-Based AFM, *Nanomaterials*, **12**, 3, 2022. DOI: 10.3390/nano12030526.

[180] Lioba Kloppenburg et al. Poly(*p* -Phenyleneethynylene)s *Are* Thermotropic Liquid Crystalline, *Macromolecules*, **32**, 13, 1999. DOI: 10.1021/ma990020p.

[181] David Ofer, Timothy M. Swager, and Mark S. Wrighton. Solid-State Ordering and Potential Dependence of Conductivity in Poly(2,5-Dialkoxy-p-Phenyleneethynylene), *Chem. Mater.*, **7**, 2, 1995. DOI: 10.1021/cm00050a029.

[182] Mutasem Omar Sinnokrot and C. David Sherrill. Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations, *J. Phys. Chem. A*, **108**, 46, 2004. DOI: 10.1021/jp0469517.

[183] Thiago Messias Cardozo et al. Dynamics of Benzene Excimer Formation from the Parallel-Displaced Dimer, *Phys. Chem. Chem. Phys.*, **21**, 26, 2019. DOI: 10.1039/C8CP06354K.

[184] Jan Hermann, Dario Alfè, and Alexandre Tkatchenko. Nanoscale π–π Stacked Mole-Cules Are Bound by Collective Charge Fluctuations, *Nat Commun*, **8**, 1, 2017. DOI: 10.1038/ncomms14052.

[185] Hong Li et al. Poly((2,5-Dialkoxy-*p*-Phenylene)Ethynylene-*p*-Phenyleneethynylene)s and Their Model Compounds, *Macromolecules*, **31**, 1, 1998. DOI: 10.1021/ma970899d.

[186] Nicholas Michelarakis et al. Longitudinal Strand Ordering Leads to Shear Thinning in Nafion, *Phys. Chem. Chem. Phys.*, **23**, 45, 2021. DOI: 10.1039/D1CP02024B.

[187] S. Stalter et al. Molecular Dynamics Simulations in Hybrid Particle-Continuum Sch-Emes: Pitfalls and Caveats, *Computer Physics Communications*, **224**, 2018. DOI: 10.1016/j.cpc.2017.10.016.

[188] Shi Qing Wang. From Wall Slip to Bulk Shear Banding in Entangled Polymer Solutions, *Macromol. Chem. Phys.*, **220**, 1, 2019. DOI: 10.1002/macp.201800327.

[189] Yongjin Ruan et al. Shear Banding in Entangled Polymers: Stress Plateau, Banding Location, and Lever Rule, *ACS Macro Lett.*, **10**, 12, 2021. DOI: 10.1021/acsmacrolett.1c00518.

[190] Kai Pahnke et al. Entropy-Driven Selectivity for Chain Scission: Where Macromolecules Cleave, *Angew. Chem. Int. Ed.*, **55**, 4, 2016. DOI: 10.1002/anie.201508531.

[191] Sandra Lerouge and Peter D. Olmsted. Non-Local Effects in Shear Banding of Polymeric Flows, *Front. Phys.*, **7**, 2020. DOI: 10.3389/fphy.2019.00246.

[192] Prashant Tapadia, Sham Ravindranath, and Shi-Qing Wang. Banding in Entangled Polymer Fluids under Oscillatory Shearing, *Phys. Rev. Lett.*, **96**, 19, 2006. DOI: 10.1103/PhysRevLett.96.196001.

[193] Shi-Qing Wang, S. Ravindranath, and P. E. Boukany. Homogeneous Shear, Wall Slip, and Shear Banding of Entangled Polymeric Liquids in Simple-Shear Rheometry: A Roadmap of Nonlinear Rheology, *Macromolecules*, **44**, 2, 2011. DOI: 10.1021/ma101223q.

[194] Michael C. Burroughs et al. Coupled Nonhomogeneous Flows and Flow-Enhanced Concentration Fluctuations during Startup Shear of Entangled Polymer Solutions, *Phys. Rev. Fluids*, **5**, 4, 2020. DOI: 10.1103/PhysRevFluids.5.043301.

[195] Michael C. Burroughs et al. Flow-Concentration Coupling Determines Features of Nonhomogeneous Flow and Shear Banding in Entangled Polymer Solutions, *Journal of Rheology*, **67**, 1, 2023. DOI: 10.1122/8.0000469.

[196] Pouyan E. Boukany and Shi-Qing Wang. Shear Banding or Not in Entangled DNA Solutions, *Macromolecules*, **43**, 17, 2010. DOI: 10.1021/ma101267b.

[197] Pouyan E. Boukany and Shi-Qing Wang. Exploring the Transition from Wall Slip to Bulk Shearing Banding in Well-Entangled DNA Solutions, *Soft Matter*, **5**, 4, 2009. DOI: 10.1039/B804791J.

[198] Mahdi Boudaghi, Brian J. Edwards, and Bamin Khomami. Molecular Processes Leading to Shear Banding in Entangled Polymeric Solutions, *Polymers*, **15**, 15, 2023. DOI: 10.3390/polym15153264.

[199] Mouge Mohagheghi and Bamin Khomami. Molecular Processes Leading to Shear Banding in Well Entangled Polymeric Melts, *ACS Macro Lett.*, **4**, 7, 2015. DOI: 10.1021/acsmacrolett.5b00238.

[200] Benedikt Rennekamp et al. Hybrid Kinetic Monte Carlo/Molecular Dynamics Simulations of Bond Scissions in Proteins, *J. Chem. Theory Comput.*, **16**, 1, 2020. DOI: 10.1021/acs.jctc.9b00786.

[201] A.J. Hodge and J.A. Petruska. Recent Studies with the Electron Microscope on Ordered Aggregates of the Tropocollagen Molecule, *Aspects of Protein Structure*, **1**, 1963.

[202] Anthony L. Mescher and Luiz Carlos Uchôa Junqueira. Junqueira's Basic Histology: Text and Atlas. Fourteenth edition. New York Chicago San Francisco Athens London Madrid Mexico City Milan New Delhi Singapore Sydney Toronto: McGraw-Hill Education, 2016. 560 pp.

[203] Agnieszka Obarska-Kosinska et al. ColBuilder: A Server to Build Collagen Fibril Models, *Biophysical Journal*, **120**, 17, 2021. DOI: 10.1016/j.bpj.2021.07.009.

[204] Eric F. Pettersen et al. UCSF Chimera?A Visualization System for Exploratory Research and Analysis, *J. Comput. Chem.*, **25**, 13, 2004. DOI: 10.1002/jcc.20084.

[205] Dennis W. Bennett. Understanding Single-Crystal x-Ray Crystallography. Weinheim: Wiley-VCH, 2010.

[206] Benedikt Rennekamp et al. Collagen Breaks at Weak Sacrificial Bonds Taming Its Mechanoradicals, *Nat Commun*, **14**, 1, 2023. DOI: 10.1038/s41467-023-37726-z.

[207] Emilie Gachon and Patrick Mesquida. Stretching Single Collagen Fibrils Reveals Nonlinear Mechanical Behavior, *Biophysical Journal*, **118**, 6, 2020. DOI: 10.1016/j.bpj.2020.01.038.

[208] Chris J. Peacock and Laurent Kreplak. Nanomechanical Mapping of Single Collagen Fibrils under Tension, *Nanoscale*, **11**, 30, 2019. DOI: 10.1039/C9NR02644D.

[209] Majid Minary-Jolandan and Min-Feng Yu. Nanomechanical Heterogeneity in the Gap and Overlap Regions of Type I Collagen Fibrils with Implications for Bone Heterogeneity, *Biomacromolecules*, **10**, 9, 2009. DOI: 10.1021/bm900519v.

[210] Bogdan I Costescu and Frauke Gräter. Time-Resolved Force Distribution Analysis, *BMC Biophys*, **6**, 1, 2013. DOI: 10.1186/2046-1682-6-5.

[211] Wolfram Stacklies, Christian Seifert, and Frauke Graeter. Implementation of Force Distribution Analysis for Molecular Dynamics Simulations, *BMC Bioinformatics*, **12**, 1, 2011. DOI: 10.1186/1471-2105-12-101.

[212] Charlotte Gistelinck et al. Abnormal Bone Collagen Cross;Inking in Osteogenesis Imperfecta/Bruck Syndrome Caused by Compound Heterozygous *PLOD2* Mutations, *JBMR Plus*, **5**, 3, 2021. doi: 10.1002/jbm4.10454.

[213] N.C. Avery and A.J. Bailey. "Restraining Cross-Links Responsible for the Mechanical Properties of Collagen Fibers: Natural and Artificial", *Collagen*. Ed. by Peter Fratzl. Boston, MA: Springer US, 2008. doi: 10.1007/978-0-387-73906-9_4.

[214] T.J. Wess. "Collagen Fibrillar Structure and Hierarchies", *Collagen*. Ed. by Peter Fratzl. Boston, MA: Springer US, 2008. doi: 10.1007/978-0-387-73906-9_3.

[215] Esmaeal Ghavanloo. Persistence Length of Collagen Molecules Based on Nonlocal Viscoelastic Model, *J Biol Phys*, **43**, 4, 2017. doi: 10.1007/s10867-017-9467-2.

[216] Ken-Ichi Saitoh et al. Molecular Dynamics Study of Collagen Fibrils: Relation between Mechanical Properties and Molecular Chirality, *JBNB*, **11**, 04, 2020. doi: 10.4236/jbnb.2020.114017.

[217] Julian Hartmann and Martin Zacharias. Mechanism of Collagen Folding Propagation Studied by Molecular Dynamics Simulations, *PLoS Comput Biol*, **17**, 6, 2021. Ed. by Alexander MacKerell. doi: 10.1371/journal.pcbi.1009079.

[218] Christopher Zapp et al. Mechanoradicals in Tensed Tendon Collagen as a Source of Oxidative Stress, *Nat Comm.*, **11**, 1, 2020. doi: 10.1038/s41467-020-15567-4.

[219] James Rowe and Konstantin Röder. Chemical Bonds in Collagen Rupture Selectively under Tensile Stress, *Phys. Chem. Chem. Phys.*, **25**, 3, 2023. doi: 10.1039/D2CP05051J.

[220] Mahdi Tavakol and Ted J. Vaughan. A Coarse-Grained Molecular Dynamics Investigation of the Role of Mineral Arrangement on the Mechanical Properties of Mineralized Collagen Fibrils, *J. R. Soc. Interface.*, **20**, 198, 2023. doi: 10.1098/rsif.2022.0803.

[221] Mark Abraham et al. GROMACS 2023.2 Manual. Version 2023.2, 2023. doi: 10.5281/ZENODO.7588710.

[222] Robert B. Best and Gerhard Hummer. Optimized Molecular Dynamics Force Fields Applied to the Helix−Coil Transition of Polypeptides, *J. Phys. Chem. B*, **113**, 26, 2009. doi: 10.1021/jp901540t.

[223] Kresten Lindorff-Larsen et al. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field: Improved Protein Side-Chain Potentials, *Proteins*, **78**, 8, 2010. doi: 10.1002/prot.22711.

[224] William L. Jorgensen et al. Comparison of Simple Potential Functions for Simulating Liquid Water, *The Journal of Chemical Physics*, **79**, 2, 1983. doi: 10.1063/1.445869.

[225] Stewart R. Durell, Bernard R. Brooks, and Arieh Ben-Naim. Solvent-Induced Forces between Two Hydrophilic Groups, *J. Phys. Chem.*, **98**, 8, 1994. doi: 10.1021/j100059a038.

[226] Dmitrii Beglov and Benoît Roux. Finite Representation of an Infinite Bulk System: Solvent Boundary Potential for Computer Simulations, *J. Chem. Phys.*, **100**, 12, 1994. doi: 10.1063/1.466711.

[227] Jing Huang and Alexander D. MacKerell. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data, *J. Comput. Chem.*, **34**, 25, 2013. doi: 10.1002/jcc.23354.

[228] Peter C. Kroon et al. Martinize2 and Vermouth: Unified Framework for Topology Generation. Version 2, 2022. doi: 10.48550/ARXIV.2212.01191.

[229] Rodrigo A. Moreira et al. *All Atom Simulations Snapshots and Contact Maps Analysis Scripts for SARS-CoV-2002 and SARS-CoV-2 Spike Proteins with and without ACE2 Enzyme*. Version 0.1. 2020. doi: 10.5281/ZENODO.3817447.

[230] B. R. Brooks et al. CHARMM: The Biomolecular Simulation Program, *J. Comput. Chem.*, **30**, 10, 2009. doi: 10.1002/jcc.21287.

[231] Jumin Lee et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field, *J. Chem. Theory Comput.*, **12**, 1, 2016. doi: 10.1021/acs.jctc.5b00935.

[232] Vytautas Gapsys et al. Pmx: Automated Protein Structure and Topology Generation for Alchemical Perturbations, *J. Comput. Chem.*, **36**, 5, 2015. doi: 10.1002/jcc.23804.

[233] Greg Landrum et al. *Rdkit: (Q1 2023)*. Version Release_2023_03_2. Zenodo, 2023. doi: 10.5281/ZENODO.591637.

[234] Tiejun Cheng et al. Computation of Octanol−Water Partition Coefficients by Guiding an Additive Model with Knowledge, *J. Chem. Inf. Model.*, **47**, 6, 2007. doi: 10.1021/ci700257y.

[235] Riccardo Alessandri et al. Martini 3 Coarse\□Grained Force Field: Small Molecules, *Advcd Theory and Sims*, **5**, 1, 2022. doi: 10.1002/adts.202100391.

[236] Victor Rühle et al. Versatile Object-Oriented Toolkit for Coarse-Graining Applications, *J. Chem. Theory Comput.*, **5**, 12, 2009. doi: 10.1021/ct900369w.

[237] G. N. Ramachandran and Gopinath Kartha. Structure of Collagen, *Nature*, **176**, 4482, 1955. doi: 10.1038/176593a0.

[238] G. N. Ramachandran and V. Sasisekharan. Structure of Collagen, *Nature*, **190**, 4780, 1961. doi: 10.1038/1901004a0.

[239] Alexander Rich and F. H. C. Crick. The Structure of Collagen, *Nature*, **176**, 4489, 1955. doi: 10.1038/176915a0.

[240] Alexander Rich and F.H.C. Crick. The Molecular Structure of Collagen, *Journal of Molecular Biology*, **3**, 5, 1961. doi: 10.1016/S0022-2836(61)80016-8.

[241] Matthew D. Shoulders and Ronald T. Raines. Collagen Structure and Stability, *Annu. Rev. Biochem.*, **78**, 1, 2009. doi: 10.1146/annurev.biochem.77.032207.120833.

[242] Pauline M. Cowan, Stewart McGavin, and A. C. T. North. The Polypeptide Chain Configuration of Collagen, *Nature*, **176**, 4492, 1955. doi: 10.1038/1761062a0.

[243] Naveen Michaud-Agrawal et al. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations, *J. Comput. Chem.*, **32**, 10, 2011. doi: 10.1002/jcc.21787.

[244] Malavika Nair, Serena M. Best, and Ruth E. Cameron. Crosslinking Collagen Constructs: Achieving Cellular Selectivity Through Modifications of Physical and Chemical Properties, *Applied Sciences*, **10**, 19, 2020. doi: 10.3390/app10196911.

[245] Mahdi Boudaghi, Brian J. Edwards, and Bamin Khomami. Microstructural Evolution and Reverse Flow in Shear-Banding of Entangled Polymer Melts, *Soft Matter*, **19**, 3, 2023. doi: 10.1039/D2SM01337A.

[246] Shi-Qing Wang et al. New Theoretical Considerations in Polymer Rheology: Elastic Breakdown of Chain Entanglement Network, *The Journal of Chemical Physics*, **127**, 6, 2007. doi: 10.1063/1.2753156.

[247] Katarzyna Adamiak and Alina Sionkowska. Current Methods of Collagen Cross-Linking: Review, *International Journal of Biological Macromolecules*, **161**, 2020. doi: 10.1016/j.ijbiomac.2020.06.075.

[248] Jaakko J. Uusitalo et al. Martini Coarse-Grained Force Field: Extension to RNA, *Biophysical Journal*, **113**, 2, 2017. doi: 10.1016/j.bpj.2017.05.043.

[249] Markus Kurth et al. DOPA Residues Endow Collagen with Radical Scavenging Capacity, *Angew Chem Int Ed*, **62**, 24, 2023. doi: 10.1002/anie.202216610.

[250] Davide Michieletto, Robert Fitzpatrick, and Rae M. Robertson-Anderson. Maximally Stiffening Composites Require Maximally Coupled Rather than Maximally Entangled Polymer Species, *Soft Matter*, **15**, 33, 2019. doi: 10.1039/C9SM01461F.

[251] Frank Noé et al. Machine Learning for Molecular Simulation, *Annu. Rev. Phys. Chem.*, **71**, 1, 2020. doi: 10.1146/annurev-physchem-042018-052331.

[252] Jiang Wang et al. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields, *ACS Cent. Sci.*, **5**, 5, 2019. doi: 10.1021/acscentsci.8b00913.

[253] Lorenzo Boninsegna, Ralf Banisch, and Cecilia Clementi. A Data-Driven Perspective on the Hierarchical Assembly of Molecular Structures, *J. Chem. Theory Comput.*, **14**, 1, 2018. doi: 10.1021/acs.jctc.7b00990.

[254] Thomas T. Foley, M. Scott Shell, and W. G. Noid. The Impact of Resolution upon Entropy and Information in Coarse-Grained Models, *The Journal of Chemical Physics*, **143**, 24, 2015. doi: 10.1063/1.4929836.

[255] Maciej Majewski et al. Machine Learning Coarse-Grained Potentials of Protein Thermodynamics, *Nat Commun*, **14**, 1, 2023. doi: 10.1038/s41467-023-41343-1.

[256] Andrew D. White. The Future of Chemistry Is Language, *Nat Rev Chem*, **7**, 7, 2023. doi: 10.1038/s41570-023-00502-0.

[257] Paulo C. T. Souza et al. Perspectives on High-Throughput Ligand/Protein Docking With Martini MD Simulations, *Front. Mol. Biosci.*, **8**, 2021. doi: 10.3389/fmolb.2021.657222.