Computational Investigation of MUC2 Mucin Rheology and Pore Characteristics:

Coarse-Grained Molecular Dynamics Simulations of a Non-Gelling Polymer Model

Josephine Lavoie

A Thesis
in
The Department
of
Physics

Presented in Partial Fulfillment of the Requirements
for the Degree of
Master of Science (Physics) at
Concordia University
Montréal, Québec, Canada

April 2026

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Josephine Lavoie

Entitled: Computational Investigation of MUC2 MucinRheology and Pore Characteristics:Coarse-Grained Molecular DynamicsSimulations ofa Non-Gelling Polymer Model

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Physics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Examiner
Deniz Erol

Examiner
Valter Zazubovits

Thesis Supervisor(s)
Rachael (Reiy) Mansbach

Thesis Supervisor(s)

Approved by
Pablo Bianucci Chair of Department or Graduate Program Director

Pascale Sicotte Dean of Arts & Sciences

Abstract

Computational Investigation of MUC2 Mucin Rheology and Pore Characteristics:

Coarse-Grained Molecular Dynamics Simulations of a Non-Gelling Polymer Model

Josephine Lavoie

Mucus forms a viscoelastic barrier in the gastrointestinal tract; the gel-forming mucin MUC2 is its primary structural component. Because MUC2 is too large ($\sim$5000 amino acids) for atomistic simulation at gelation-relevant scales, this thesis develops a coarse-grained bead-spring polymer model and computational pipeline for MUC2, building on Ford et al., to determine whether physical interactions alone can reproduce gelation or whether explicit disulfide crosslinks are required.

A large-scale parameter screening (Phase 1: 4,410 simulations across five parameter dimensions and seven concentrations) with Lennard-Jones interactions revealed that the model behaves as a viscous liquid (median $\tan\delta = 3.1$, 87.9% sol-like). An optimized production study (Phase 2: 131 simulations) confirmed liquid-like behavior ($\tan\delta \gg 1$) with storage moduli $G' \approx 37$–174 Pa. Pore characterization yielded mean diameters of 12.5–14.1 nm; validation runs with full-length dimers produced pore sizes (mean $\sim$274 nm) within the experimental range. An exploratory chain length study (6 to 104 beads per dimer, timescales up to $\sim$420 μs) confirmed all conditions remain liquid-like. The model does not gel.

This result is consistent with the Parsons MUC5AC model, which achieved gelation only after incorporating stochastic disulfide bonds. The work establishes a validated computational pipeline including Green-Kubo rheology and Bhattacharya-Gubbins pore characterization, and documents two methodological pitfalls (a WCA interaction potential bug and NPT equilibration volume shifts) relevant to coarse-grained LAMMPS studies. Adding explicit disulfide crosslinks is the most important next step toward a complete MUC2 model.

Acknowledgments

I would like to thank my supervisor, Dr. Re Mansbach, for their guidance, patience, and support throughout this project. Thank you for believing in me – it has been such a pleasure working with your lab. I have learned a great deal from this welcoming group of highly capable individuals.

Thank you to my committee members, Drs Zazubovits and Erol, for your interest in my work and your involvement throughout its evolution.

A huge thank you to my community here in Montreal. I have gained an impressive slew of practice in science communication from being faced with such a wide array of curiosities about my project. I also deeply appreciate the ways in which you all have helped me stay grounded through such a daunting process. Life throughout this degree has been a much richer experience with all of you in it.

I am grateful for the support of my family, both far away and nearby, as there has never been a doubt in their minds that I would succeed in this endeavor. Your faith emboldened me.

Computational resources were provided by the Digital Research Alliance of Canada (formerly Compute Canada) through access to the Narval cluster. This work would not have been possible without these resources.

I also thank the Concordia University Department of Physics for supporting my graduate studies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Mucus lines most of the body's internal surfaces, including the gut, the lungs, and the reproductive tract, and acts as a barrier between cells and the external environment [1]. The molecular mechanisms underlying its gel-forming function are not fully understood. This thesis develops and validates a coarse-grained computational model and analysis pipeline for studying MUC2 mucin, the main structural component of intestinal mucus, with the goal of determining whether such a model can reproduce the experimentally observed sol-gel transition, or whether additional physics such as disulfide crosslinks is needed.

1.1 Biological Context

Mucus performs diverse functions depending on its location in the body. In the gut, it keeps bacteria away from the epithelium while still letting nutrients through [2, 3]. It plays analogous roles in the lungs and reproductive tract (trapping particles, regulating transport) [4, 5], but the intestinal barrier is the focus of this thesis [6]. The mucus layer thickness varies along the gastrointestinal tract, from ∼120 μm in the jejunum to ∼830 μm in the colon [7].

What makes mucus work in all these contexts is that it is viscoelastic: it behaves like a liquid on long timescales (it flows) but like a solid on short ones (it springs back). This dual character comes from the mucin glycoproteins that form its structural network [8].

1.2 MUC2 Mucin

This thesis focuses on MUC2, the predominant gel-forming mucin in the intestinal tract [9]. MUC2 is a large molecule (approximately 640 kDa per monomer, or roughly 5000 amino acids) with a distinctive bottle-brush structure [6]. The central region is heavily decorated with O-linked glycan chains that extend from the protein backbone. The N- and C-terminal domains are not glycosylated; these cysteine-rich protein regions mediate intermolecular disulfide crosslinking that creates the gel network [10, 11, 12].

MUC2 is also implicated in several diseases. In inflammatory bowel disease (IBD), the mucus layer is often thinner and more permeable to bacteria [9, 13]. In colorectal cancer, MUC2 expression is frequently altered [14]. In cystic fibrosis, MUC5B and MUC5AC in the lungs are more directly affected (cystic fibrosis is primarily a pulmonary disease), but changes in overall mucus properties also impact intestinal function [15].

The physical chemistry of mucin gelation involves hydrophobic interactions between non-glycosylated domains and pH-dependent conformational changes [16]. Bansil and Turner [10] reviewed how low pH promotes hydrophobic interactions between mucin domains, driving conformational changes that contribute to gelation of gastric mucins. Celli et al. [17] showed that *Helicobacter pylori* reduces mucus viscoelasticity by raising local pH through urease activity, providing direct experimental evidence for pH-dependent gel weakening. These findings motivate the use of hydrophobic interaction strength as a key parameter in the coarse-grained model developed in this thesis.

1.3 Experimental Rheology of Mucus

The simulation results in this thesis are ultimately compared to experimental rheology measurements, which come from two main techniques.

Bulk rheology. A mechanical rheometer measures the storage and loss moduli (G' and G'') as functions of frequency and strain amplitude. Native mucus typically gives $G' \sim 1$–1000 Pa depending on concentration, source, and measurement conditions [18, 19]. Sample preparation is a persistent challenge: native mucus is heterogeneous, and purification procedures may alter its properties [18].

Microrheology. The complementary approach uses embedded tracer particles to probe local mechanical properties through their thermal fluctuations (passive) or their response to applied forces from optical or magnetic tweezers (active). Microrheology can detect spatial heterogeneity and requires smaller sample volumes (see Squires and Mason [20] for a review). The method was introduced by Mason and Weitz [21], who showed that frequency-dependent viscoelastic moduli can be extracted from the thermal motion of embedded probe particles via the Generalized Stokes-Einstein Relation (GSER).

Meldrum et al. [22] used particle tracking microrheology to characterize porcine intestinal MUC2, demonstrating an abrupt sol-gel transition between 10 and 15 mg/mL where tracer particle motion changed from diffusive to sub-diffusive. Georgiades et al. [23] combined small-angle X-ray and neutron scattering with microrheology to characterize MUC2 and MUC5AC structure and dynamics, identifying the overlap concentration ($c^* < 1$ mg/mL) and entanglement concentration ($c_e \approx 25$–30 mg/mL) for these mucins.

1.4 Why Simulations?

Experimental characterization of mucus is difficult. The material is heterogeneous, exhibits inter-individual variability, and degrades rapidly following extraction [18]. Rheological measurements provide bulk mechanical properties but do not directly reveal molecular mechanisms. Microscopy can resolve structure but typically cannot capture dynamics at the molecular scale.

Simulations address these limitations directly. A model of the mucin molecules is constructed, placed in a simulation box, and evolved under controlled conditions. Quantities like stress fluctuations and pore sizes, which are difficult to measure experimentally, are computed directly from the trajectory. Parameters such as concentration, interaction strength, and chain length can be varied systematically in ways that would be impractical in the lab.

The tradeoff is that simulations require approximations. Atomistic simulation of full-length MUC2 (with ∼5000 amino acids per monomer) is computationally prohibitive for the length and time scales relevant to gelation. Coarse-grained models [24, 25], which represent groups of atoms as single interaction sites, make these scales accessible at the cost of molecular detail. Previous work has demonstrated that coarse-grained bead-spring models can capture key features of mucin behavior, including network structure [26] and sol-gel transitions [27]. The question addressed in this thesis is whether such a model, relying on physical (non-covalent) interactions alone, is sufficient to reproduce MUC2 gelation, or whether explicit chemical crosslinks are required.

1.5 Previous Computational Studies of Mucins

1.5.1 Atomistic Simulations

Atomistic simulations of mucins can resolve local structure in detail, but the system sizes and timescales relevant to gelation are out of reach. A single MUC2 monomer contains approximately 5000 amino acids; a simulation box containing enough monomers to form a gel at physiological concentrations (10–50 mg/mL) would require millions of atoms plus explicit water molecules. At current computational capabilities, such simulations remain computationally prohibitive at the microsecond-to-millisecond timescales relevant to rheology. This motivates the coarse-grained approaches described below, which sacrifice atomic detail to access the relevant length and time scales.

1.5.2 The Ford et al. Coarse-Grained Model

Earlier coarse-grained work by Gniewek and Kolinski [28, 29] used an on-lattice Monte Carlo model to study mucin structure, thermodynamics, and barrier properties but did not reach the timescales or system sizes needed for rheological characterization. Ford et al. [26] developed an off-lattice coarse-grained molecular dynamics model capable of simulating mucin gel networks at physiologically relevant concentrations and timescales. Their model, which forms the foundation of the present work, represents MUC5B (the primary airway mucin) as a bead-spring polymer using the Kremer-Grest framework [24].

Ford's model uses three bead types to represent the domain architecture: hydrophobic terminals (T), glycosylated regions (G, called H in this thesis), and linkers (L). Each bead is about 9 nm in diameter, matching the persistence length of mucin, and represents roughly 100 amino acids. Chains are connected by FENE bonds [24], and non-bonded interactions use Lennard-Jones potentials [24] with bead-type-dependent interaction strengths ($\varepsilon_{\mathrm{TT}} > \varepsilon_{\mathrm{TH}} > \varepsilon_{\mathrm{HH}}$). The specific ε values and their physical motivation are discussed in Section 3.2. Box sizes were calibrated to physiological concentrations (1.5–5% by weight).

Ford et al. demonstrated that this model reproduces qualitative features of healthy versus cystic fibrosis mucus: at elevated concentrations mimicking CF conditions, the simulated networks exhibited a phase transition to denser, more entangled structures [26]. Ford also computed equilibrium viscoelastic moduli from statistical mechanics formulas, though the primary focus was structural characterization of healthy vs. cystic fibrosis mucus networks.

1.5.3 The Parsons MUC5AC Model

Parsons [27] developed a coarse-grained LAMMPS model of MUC5AC mucin, building on the lattice-based Monte Carlo work of Gniewek and Kolinski [28]. Like the Ford model, Parsons used the Kremer-Grest bead-spring framework with Lennard-Jones non-bonded interactions and FENE bonds along the chain backbone. However, the two models differ in several important respects.

Parsons used four bead types (hydrophobic [**H**], disulfide [**S**], polar [**P**], and neutral [**N**]) to represent the domain architecture of MUC5AC monomers. Each monomer was modeled as 20 beads with a diameter of $\sigma = 25$ nm, corresponding to the Kuhn length of mucin. The Kuhn length $b_K = 2l_p$ is twice the persistence length; the Ford model and the present work use $\sigma = 9$ nm, which matches the persistence length $l_p \approx$ 8–9 nm measured for MUC2 and MUC5AC [23]. The choice of σ affects the number of beads per chain and thus the accessible system sizes and timescales. Two monomers were joined permanently to form 40-bead dimers. A Monte Carlo subroutine allowed disulfide bonds between **S** beads to form and break stochastically, enabling dynamic polymer chain growth and network rearrangement. The system comprised 2,000 dimers (80,000 beads) in a periodic box of side 117σ at a mass concentration of approximately 16 mg/mL (1.6 g/100 mL).

The model was validated against several experimental observables. The average polymer size of 4.89 ± 2.23 dimers fell within the experimentally observed range of 3–8 dimers per mucin polymer. The mesh size of 337.5 nm at a temperature of $T = 0.60$ matched the accepted range for mucus pore sizes (~300 nm). A sol-gel transition was observed between $T = 0.60$ and $T = 0.65$: below this critical temperature, the storage modulus exceeded the loss modulus across a wide frequency range, while above it the system behaved as a viscous liquid. Rheological properties were computed using the Mason passive microrheology method, which extracts $G'(\omega)$ and $G''(\omega)$ from the mean square displacement of simulated beads.

The Parsons model is the closest precedent to the present work and shows that the Kremer-Grest framework can produce a sol-gel transition with realistic rheology. The main differences from this thesis are the mucin type (MUC5AC vs MUC2), the rheological method (Mason microrheology vs Green-Kubo), and the scope (Parsons studied one parameter set at one concentration, while this thesis screens thousands of combinations).

1.5.4 Other Coarse-Grained Approaches

Other coarse-graining strategies exist but were not used here (see Noid [25] for a comprehensive review). MARTINI [30] operates at roughly 4:1 heavy-atom resolution and has been applied to glycosylated proteins, but its finer resolution limits the system sizes and timescales that are accessible. DPD [31] captures hydrodynamic interactions and has been applied to polymer gels, though not to mucins specifically. The Kremer-Grest framework [24] used by Ford and in this thesis is the coarsest approach that still preserves chain connectivity and domain-specific interactions. It sacrifices chemical specificity, but this is the tradeoff required to simulate hundreds of chains at gelation-relevant concentrations.

1.6 Overview of the Thesis

1.6.1 Gap Addressed by This Thesis

The computational models described above have shown that coarse-grained mucin simulations can reproduce some qualitative features of mucus, such as network structure, concentration dependence, and sol-gel transitions. But quantitative rheology from these models is still limited. Ford et al. [26] computed equilibrium moduli at two concentrations using statistical mechanics formulas; Parsons [27] computed frequency-dependent moduli for MUC5AC but at a single condition. Neither study systematically explored how the force field parameters affect the resulting rheological and structural properties, and direct comparison to experimental measurements was limited.

Full-length mucin dimers are expensive in computational modelling, and it is not clear how short a chain can be while still capturing the essential gelation physics. This thesis addresses these gaps: systematic parameter exploration across thousands of combinations, Green-Kubo rheology that can be compared to experiment, a chain length study that tests whether shorter models are viable proxies, and a code pipeline that is publicly available and documented for extension to other mucins or disease states.

1.6.2 Thesis Objectives

The primary goal of this thesis is to develop and validate a computational model and code pipeline for studying MUC2 gel rheological and structural properties. The specific objectives are:

1. **Validate the implicit solvent** using Flory scaling of test chains.
2. **Screen force field parameters** across a large parameter space to identify the parameter regime most favorable for gelation.
3. **Compute rheological properties** using the Green-Kubo formalism. This is the main analysis tool: storage and loss moduli from equilibrium stress fluctuations, compared against experimental values from the mucus literature.
4. **Characterize pore structure** using the Bhattacharya-Gubbins method and compare to experimental pore sizes from AFM and nanoparticle transport studies.
5. **Determine chain length scaling.** Full-length MUC2 ($\sim$5000 amino acids) is too expensive for atomistic simulation, so knowing how short a coarse-grained chain can be

while still capturing the essential physics would be useful for future higher-resolution studies. The chain length study tests four model sizes (6 to 104 beads per dimer) at two concentrations across three timescales.

6. **Document a reproducible workflow and code pipeline.** All simulation input files, analysis scripts, and figure generation code are publicly available (Appendix D) so that the methodology can be extended to other mucins or disease states.

1.6.3 Thesis Organization

Chapters 2 and 3 cover theory and methods. Chapter 4 validates the implicit solvent through Flory scaling. The main results are in Chapter 5, which is organized into Phase 1 parameter screening (4,410 simulations) and Phase 2 production with stronger hydrophobic parameters; a critical interaction potential bug discovered during the work is also documented there. An exploratory chain length study is summarized in Chapter 5 with full results in Appendix A. Chapter 6 interprets the results, and Chapter 7 gives conclusions and future directions. Appendices B and C document earlier results retained for completeness.

Chapter 2

Theory

The simulations and analysis in this thesis rely on polymer physics, linear viscoelasticity theory, and the Green-Kubo formalism.

2.1 Polymer Physics Background

2.1.1 Flory Scaling

A polymer chain in solution adopts a random coil conformation. The size of this coil depends on the chain length N (number of monomers) according to a power law [32, 33]:

$$R_g \sim N^{\nu} \tag{2.1}$$

where R_g is the radius of gyration and ν is the Flory exponent. The value of ν indicates the solvent conditions:

- $\nu = 1/3$: Poor solvent (chain collapses into a globule)
- $\nu = 1/2$: Theta solvent (ideal chain, like a random walk)
- $\nu \approx 0.588$: Good solvent (chain swells due to excluded volume) [34, 33]
- $\nu = 1$: Rigid rod (fully extended)

Measuring ν from simulations provides a useful validation test. A value of $\nu \approx 1$ indicates a problem, as chains are not equilibrating and remain rod-like. The expected value of ν depends on solvent conditions: $\nu \approx 0.5$ for theta solvent or $\nu \approx 0.588$ for good solvent. Both indicate flexible random coil behavior and confirm that the polymer is properly equilibrated.

Any value of ν other than 1 is consistent with a flexible chain; the specific value encodes the solvent quality.

2.1.2 Why This Matters for Mucin

Mucin molecules are long and flexible. In solution, they should adopt random coil conformations. If the simulation produces rod-like chains ($\nu \approx 1$), the rheological properties will be completely wrong because the conformations are unphysical. Validating Flory scaling using simple test chains is therefore a prerequisite for trusting that the implicit solvent environment (Langevin dynamics) produces correct polymer physics before applying it to mucin systems.

2.1.3 Chain Length Scaling

Different models of polymer physics predict how key dynamical quantities scale with chain length N in different regimes. Two quantities are central: the relaxation time τ (the characteristic time for a polymer chain to lose memory of its initial conformation) and the zero-shear viscosity η_0 (the viscosity of the polymer solution in the limit of zero shear rate).

The Rouse model [35, 36] describes the dynamics of unentangled polymer chains as a series of beads connected by springs, where each bead experiences friction from the solvent. The key predictions are:

- **Relaxation time:** $\tau_R \sim N^2$, where τ_R is the Rouse relaxation time and N is the chain length (number of monomers). Longer chains take quadratically longer to relax their conformations.

- **Zero-shear viscosity:** $\eta_0 \sim N$, increasing linearly with chain length.

For entangled polymers (chains long enough to form topological constraints), the reptation model [37] predicts steeper scaling: the reptation time $\tau_{\text{rep}} \sim N^3$ and $\eta_0 \sim N^{3.4}$. The 3.4 exponent is an empirical value measured from melt viscosity experiments on linear polymers of varying molecular weight [36, 38]; the theoretical prediction is N^3, with the additional 0.4 arising from constraint release and contour length fluctuation effects. The crossover from Rouse to reptation behavior occurs at the entanglement length N_e.

These scaling laws are important for interpreting chain length studies (Chapter 5), particularly the N^2 Rouse scaling used to set equilibration times for different chain lengths (Section 3.6). However, the scaling predictions assume fixed concentration; when concentration varies simultaneously with chain length (as occurs with NPT equilibration), the observed exponents reflect both effects and cannot be directly compared to Rouse or reptation predictions.

2.2 Rheological Properties of Polymer Models

This section introduces the rheological quantities used to characterize the mechanical response of polymer systems, including mucus. In oscillatory rheology, a small-amplitude sinusoidal strain $\gamma(t) = \gamma_0 \sin(\omega t)$ is applied and the resulting stress $\sigma(t)$ is decomposed into components in-phase and out-of-phase with the strain. The in-phase component measures elastic (energy-storing) response, while the out-of-phase component measures viscous (energy-dissipating) response. This decomposition defines the frequency-dependent storage and loss moduli, which are the primary outputs of the Green-Kubo analysis used in this thesis.

2.2.1 Storage and Loss Moduli

Viscoelastic materials have both elastic (solid-like) and viscous (liquid-like) responses. These are characterized by the frequency-dependent storage modulus $G'(\omega)$ and loss modulus $G''(\omega)$:

- $G'(\omega)$: The elastic (storage) component. Energy stored during deformation and recovered when the stress is released.
- $G''(\omega)$: The viscous (loss) component. Energy dissipated as heat during deformation.

For a purely elastic solid, $G'' = 0$. For a purely viscous liquid, $G' = 0$. Real materials like mucus have both, and the ratio indicates which behavior dominates. The mathematical definitions of $G'(\omega)$ and $G''(\omega)$ as Fourier transforms of the stress autocorrelation function are given in Section 2.3 (Equations 2.4–2.5).

2.2.2 Loss Tangent and Sol-Gel Transition

The loss tangent is defined as:

$$\tan\delta = \frac{G''(\omega)}{G'(\omega)} \tag{2.2}$$

The name “loss tangent” comes from the phase angle δ between the applied strain and the resulting stress in oscillatory rheology: a purely elastic material has $\delta = 0$ (stress perfectly in-phase with strain), a purely viscous material has $\delta = \pi/2$ (stress 90° out-of-phase), and viscoelastic materials fall between. The ratio $G''/G' = \tan\delta$ thus parametrizes the balance between viscous and elastic response:

- $\tan\delta > 1$ ($\delta > 45°$): Viscous/liquid-like (sol)
- $\tan\delta < 1$ ($\delta < 45°$): Elastic/solid-like (gel)
- $\tan\delta \approx 1$ ($\delta = 45°$): Critical gel point [39]

For mucus, this matters because the gel point determines whether the material flows or holds its shape. Healthy mucus needs to be gel-like enough to form a barrier but not so stiff that it cannot be cleared. Experimental studies on porcine intestinal MUC2 show that the sol-gel transition occurs between 10 and 15 mg/mL, where tracer particle motion changes abruptly from diffusive to sub-diffusive [22]. This physiological gelation concentration is a key target for model validation: to accurately capture the rheological properties of MUC2, the model should exhibit $\tan\delta \approx 1$ at concentrations near 10–15 mg/mL.

2.3 Green-Kubo Formalism

2.3.1 Stress Fluctuations and Rheology

The Green-Kubo relations connect equilibrium fluctuations to transport properties [40, 41]. For rheology, the key quantity is the stress autocorrelation function:

$$C(t) = \langle \sigma_{xy}(0)\sigma_{xy}(t) \rangle \tag{2.3}$$

where σ_{xy} is an off-diagonal component of the stress tensor. In practice, averaging over all three off-diagonal components (xy, xz, yz) improves statistics.

2.3.2 Computing Moduli from Stress Fluctuations

The frequency-dependent moduli are obtained by Fourier transforming $C(t)$:

$$G'(\omega) = \frac{V}{k_B T}\omega \int_0^\infty C(t)\sin(\omega t)\,dt \tag{2.4}$$

$$G''(\omega) = \frac{V}{k_B T}\omega \int_0^\infty C(t)\cos(\omega t)\,dt \tag{2.5}$$

where V is the system volume and T is the temperature.

A related quantity is the plateau modulus (or instantaneous modulus) obtained directly from the stress variance:

$$G'_{\text{plateau}} = \frac{V}{k_B T}\langle \sigma_{xy}^2 \rangle \tag{2.6}$$

This represents the high-frequency limit of $G'(\omega)$ and is equivalent to $C(t = 0)$ scaled by V/k_BT. Because it requires only the variance of the stress tensor (no Fourier transform), G'_{plateau} is numerically robust and serves as a useful diagnostic, but it is frequency-independent and cannot distinguish elastic from viscous response. Throughout this thesis, the loss tangent $\tan\delta = G''/G'$ is always computed from the frequency-dependent Green-Kubo moduli (Eqs. 2.4–2.5) at the same reference frequency, not from G'_{plateau}.

The physical interpretation: stress fluctuations at equilibrium encode information about how the material would respond to an applied deformation. Large, slowly-decaying fluctuations indicate a stiff, elastic material. Small, quickly-decaying fluctuations indicate a soft, viscous material. The sine transform in $G'(\omega)$ (Equation 2.4) and the cosine transform in $G''(\omega)$ (Equation 2.5) partition this information: the sine function, being an odd function that vanishes at $t = 0$ and grows with t, preferentially weights the long-time tail of $C(t)$, capturing the slowly-relaxing elastic component. The cosine function, being an even function that peaks at $t = 0$ and decays with t, preferentially weights the short-time behavior, capturing the fast viscous dissipation.

2.3.3 Why Green-Kubo Over the Mason Method

An alternative approach to computing the frequency-dependent moduli (Equations 2.4–2.5) is the Mason passive microrheology method [21], which Parsons [27] used for MUC5AC. Tracer particles are embedded in the simulated gel, their mean squared displacement is recorded, and frequency-dependent moduli are extracted via the Generalized Stokes-Einstein Relation (GSER) [42]. Green-Kubo was chosen here mainly because it does not require tracer particles: moduli come directly from the stress tensor, which LAMMPS computes natively. This avoids the additional parameters (probe size, probe-polymer interactions) and potential artifacts that tracers introduce, and in a coarse-grained model where the bead diameter ($\sigma = 9$ nm) is comparable to typical probe radii, the continuum assumption underlying the GSER is questionable anyway. Green-Kubo also gives the true bulk response of the simulation box rather than local properties near each probe, and it scales easily to large parameter sweeps since it requires only the pressure tensor output already being computed.

The trade-off is that Green-Kubo needs long trajectories for the stress autocorrelation to converge (Section 4.3), and the Fourier transform integrals are numerically sensitive.

2.3.4 Practical Considerations

The main practical challenges for the Green-Kubo approach are equilibration, trajectory length, and numerical sensitivity. The stress autocorrelation function needs to either decay to zero (for liquids) or reach a well-defined plateau (for gels); if it does neither, the system has not been equilibrated long enough for the stress correlations to fully relax, and the resulting moduli will be unreliable. The production run should be at least 10–100× longer than the longest relaxation time [36], and the Fourier transform integrals are sensitive to noise and truncation effects. These considerations motivated the validation studies in Chapter 4.

2.4 Pore Size Characterization

2.4.1 What Pore Size Means

In this thesis, "pore size" refers to the diameter of the largest sphere that can be inscribed in the void space between polymer beads at a given location in the simulation box. The distribution of these diameters across many random locations characterizes the network's permeability: the pore size distribution determines what can diffuse through a mucus gel. Small molecules pass easily, while bacteria (typically 1–2 μm) should be excluded [43]. For drug delivery nanoparticles, the pores need to be in the right size range: big enough for the particles to enter, but small enough that the gel still provides a functional barrier [18, 44]. Drug-mucin binding interactions further modulate transport through the mucus layer [45].

2.4.2 Bhattacharya-Gubbins Method

The Bhattacharya-Gubbins (B-G) random point insertion method [46] measures the distribution of pore sizes by:

1. Picking random points in the simulation box
2. For each point, finding the largest sphere that fits without overlapping any polymer beads

The output is a distribution of pore diameters, from which statistics (mean, median, percentiles) are computed.

Mathematically, for a test point $\mathbf{P}$, the optimization finds the center $\mathbf{C}$ of the largest inscribed sphere:

$$\max_{\mathbf{C}} R(\mathbf{C}) = \min_{i} |\mathbf{C} - \mathbf{r}_i| - r_{\text{bead}} \tag{2.7}$$

subject to $|\mathbf{C}-\mathbf{P}| \leq R(\mathbf{C})$, where $\mathbf{C}$ is the center of the inscribed sphere, $R(\mathbf{C})$ is its radius, $\mathbf{r}_i$ are the positions of the polymer beads (indexed by i), and r_{bead} is the bead radius.

The constraint ensures the sphere contains the test point. The optimization finds the largest such sphere. This gives the local "pore size" at that point.

2.5 Molecular Dynamics Simulations

Molecular dynamics (MD) is a computational method that simulates the motion of a collection of interacting particles by numerically integrating their equations of motion [47, 48]. Starting from an initial configuration of particle positions and velocities, MD computes the forces on each particle from a set of interaction potentials (collectively called the force field), advances the positions and velocities by a small timestep, and repeats. The resulting trajectory, a time series of particle positions, provides the raw data for computing thermodynamic, structural, and rheological properties.

This section describes the equations of motion, the two main interaction potentials used in this work (Lennard-Jones for non-bonded interactions and FENE for bonded interactions), and the Langevin dynamics extension that models implicit solvent effects. The specific parameter choices for these potentials are discussed in Chapter 3.

2.5.1 Equations of Motion

The starting point for MD is Newton's second law, applied to each particle i in the system. MD numerically integrates these equations using finite difference schemes such as the velocity Verlet algorithm [49, 50]:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i = -\nabla_i U \tag{2.8}$$

where m_i is the mass of particle i, $\mathbf{r}_i$ is its position vector, $\mathbf{F}_i$ is the total force on particle i, and U is the total potential energy of the system. The forces are derived from a force field: a set of mathematical functions that describe how particles interact. In most MD simulations (coarse-grained or atomistic), the force field includes both bonded interactions (between particles connected along a polymer chain) and non-bonded interactions (between all pairs of particles within a cutoff distance). Two non-bonded and bonded potentials are central to this work:

Lennard-Jones (LJ) potential. The LJ potential models non-bonded interactions between particles:

$$U_{\text{LJ}}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right], \quad r < r_c \tag{2.9}$$

where r is the distance between two particles, ε sets the interaction strength (depth of the potential well), σ sets the particle diameter (the distance at which $U = 0$), and r_c is a cutoff distance beyond which the potential is truncated to zero. The r^{-12} term provides short-range repulsion (preventing particle overlap), while the r^{-6} term provides longer-range attraction (modeling van der Waals forces). The potential has a minimum at $r = 2^{1/6}\sigma \approx 1.12\sigma$. If the cutoff is set to $r_c = 2^{1/6}\sigma$, only the repulsive portion is retained, yielding the Weeks-Chandler-Andersen (WCA) potential [51]. A larger cutoff (e.g., $r_c = 2.5\sigma$) includes attractive interactions, which are essential for modeling hydrophobic association.

In practice, truncating the potential at r_c introduces a discontinuity: $U(r_c) \neq 0$ and the force $F = -dU/dr$ has a step at r_c. LAMMPS handles this by shifting the potential so that $U(r_c) = 0$, ensuring continuous energy (the `lj/cut` pair style applies this shift automatically). The force still has a small discontinuity at r_c, but for cutoffs of 2.0σ or larger this discontinuity is negligible.

FENE (Finitely Extensible Nonlinear Elastic) potential. The FENE potential [52] models bonded interactions between consecutive beads in a polymer chain:

$$U_{\mathrm{FENE}}(r) = -\frac{1}{2}KR_0^2 \ln\left[1 - \left(\frac{r}{R_0}\right)^2\right] \tag{2.10}$$

where K is the spring constant and R_0 is the maximum bond extension. The potential diverges as $r \to R_0$, preventing bonds from stretching beyond a finite length. In practice, the FENE bond is combined with the repulsive part of the LJ potential to prevent bead overlap. The resulting FENE+LJ bond has a sharp potential well that maintains chain connectivity while allowing thermal fluctuations. The specific parameter choices (K, R_0, ε, σ) used in this work are discussed in Section 3.2.

2.5.2 Langevin Dynamics and Implicit Solvent

Langevin dynamics [53] is an extension of the basic MD equations of motion that implicitly models the effects of a solvent on a solute (for example, a polymer in solution) without simulating individual solvent molecules. The equation of motion for each particle i becomes:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i - \gamma\frac{d\mathbf{r}_i}{dt} + \sqrt{2\gamma k_B T}\,\boldsymbol{\eta}(t) \tag{2.11}$$

where $\mathbf{F}_i$ is the force from the interaction potentials (same as in Equation 2.8), γ is the friction coefficient (units of mass/time, representing the drag that the solvent would exert on the particle), k_B is Boltzmann's constant, T is the temperature, and $\boldsymbol{\eta}(t)$ is a Gaussian white noise vector with zero mean and unit variance. The equation adds two terms to Newton's second law: a friction term $-\gamma\dot{\mathbf{r}}_i$ that opposes motion (modeling the viscous drag of the solvent) and a random force term $\sqrt{2\gamma k_B T}\,\boldsymbol{\eta}(t)$ that models the thermal fluctuations arising from random collisions with solvent molecules. The amplitude of the random force is set by the fluctuation-dissipation theorem to ensure that the system reaches the correct thermal equilibrium at temperature T: stronger friction requires stronger random forcing to maintain the target temperature.

In the overdamped limit, where friction dominates inertia ($\gamma \gg m/\tau$), the acceleration term $m\ddot{\mathbf{r}}$ becomes negligible, and the equation of motion simplifies to Brownian dynamics (BD):

$$\gamma \frac{d\mathbf{r}_i}{dt} = \mathbf{F}_i + \sqrt{2\gamma k_B T}\,\boldsymbol{\eta}(t) \tag{2.12}$$

By eliminating the inertial term, BD removes the fastest vibrational timescales from the system, permitting timesteps 10–100× larger than Langevin dynamics while preserving the correct diffusive dynamics and equilibrium sampling [54]. Ford et al. [26] used overdamped Brownian dynamics for their MUC5B model, achieving comparable spatial resolution (9 nm beads) without the timestep limitations of Langevin dynamics. This thesis uses Langevin dynamics (Equation 2.11) because it is natively supported in LAMMPS via the `fix langevin` command and provides straightforward temperature control. The trade-off is a more restrictive timestep ($\Delta t \leq 0.012\tau$; see Section 4.1), which limits accessible simulation timescales to $\sim$420 μs. A future transition to Brownian dynamics could extend this toward the millisecond regime relevant to experimental rheometry.

Chapter 3

Methods

This chapter describes the polymer model of MUC2, the interaction potentials (force field), simulation protocols, and analysis pipelines.

3.1 Polymer Model of MUC2

3.1.1 Bead Types and Chain Architecture

Following Ford et al. [26], MUC2 is modeled as a linear bead-spring polymer where each bead represents approximately 100 amino acids plus their associated glycan decoration. The model maps the biological domain structure of MUC2 onto a sequence of beads with domain-specific interaction strengths, connected by spring-like bonds. Ford et al. use three bead types (T, G, and L), where L-beads represent linker regions adjacent to cysteine-rich domains that form inter-chain disulfide bonds. This thesis omits the L-bead because modeling cysteine knot formation would require a Monte Carlo bond-formation subroutine (as implemented by Parsons [27] for disulfide bonds in MUC5AC); the present work focuses on validating the baseline interaction parameters before introducing this additional complexity. The two bead types used are:

- **Type T (Terminal):** Represents the hydrophobic N- and C-terminal domains
- **Type H (Hydrophilic):** Represents the glycosylated central domain

The full 52-bead chain representing a MUC2 monomer has the pattern:

$$[\mathrm{T}]_{11} + [\mathrm{H}]_{35} + [\mathrm{T}]_{6} \tag{3.1}$$

meaning 11 T-beads at the N-terminus, 35 H-beads in the glycosylated region, and 6 T-beads at the C-terminus. A dimer consists of two identical monomers joined at their C-termini by the same FENE bond used between all adjacent beads, matching the biological C-terminal disulfide linkage of MUC2 [12]. This gives a 104-bead chain with the symmetric pattern $[T]_{11}$–$[H]_{35}$–$[T]_6$–$[T]_6$–$[H]_{35}$–$[T]_{11}$, with a T_{12} block at the junction and T_{11} blocks at each free end. The shorter chain lengths used in the chain length study ($N = 6$ and $N = 18$) are truncated versions that preserve the T–H–T domain order (Section 3.6).

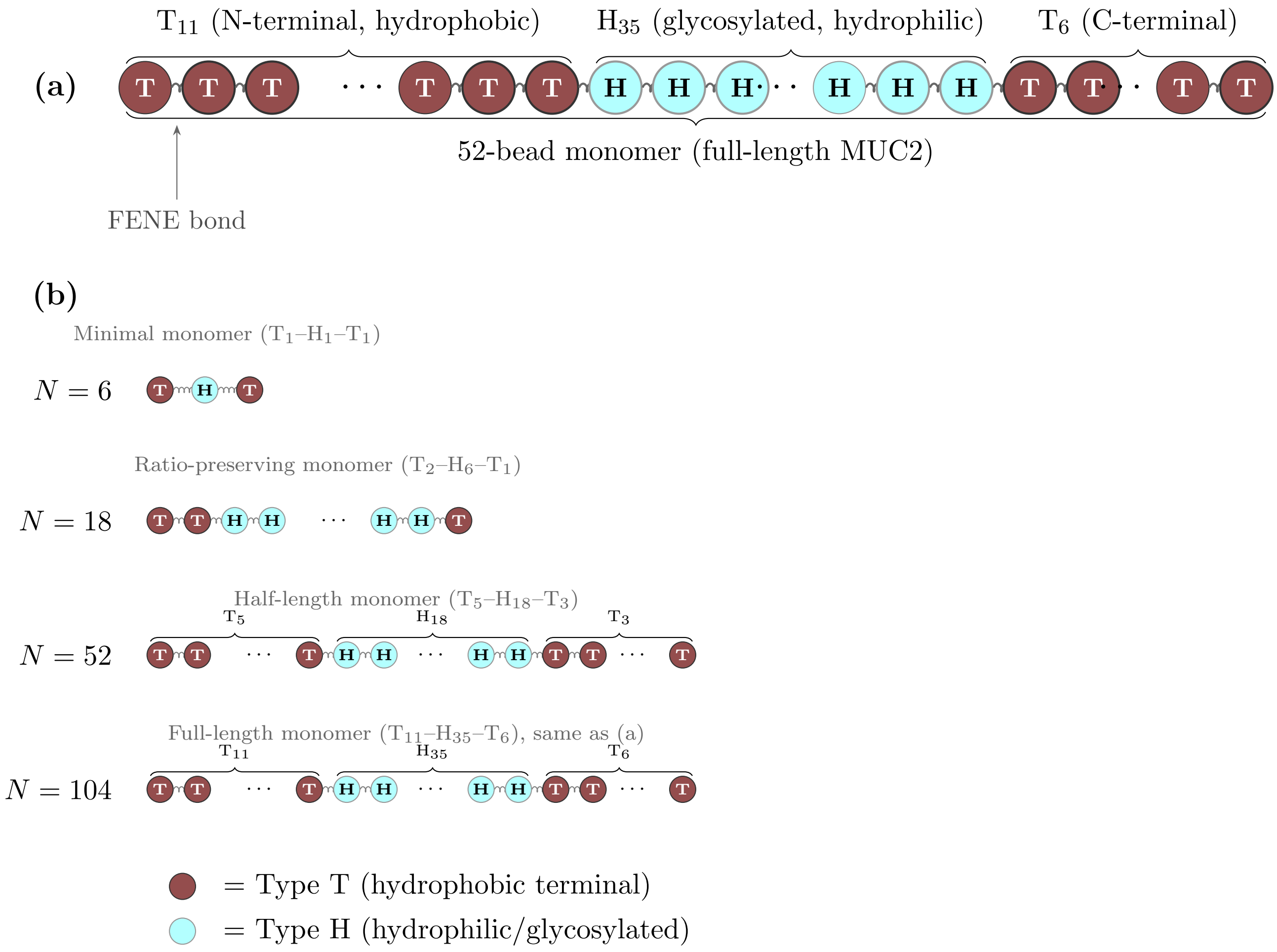


Figure 3.1: Coarse-grained bead-spring model of MUC2. **(a)** The 52-bead monomer showing the T_{11}–H_{35}–T_6 domain structure: hydrophobic terminal beads (T, brown) and hydrophilic glycosylated beads (H, light blue), connected by FENE bonds (springs). **(b)** Monomer composition for each of the four chain lengths used in the chain length study. Each simulation chain is a dimer of two such monomers joined at their C-termini, giving the pattern T_{11}–H_{35}–T_6–T_6–H_{35}–T_{11} for the full-length model. The T–H–T domain ratio is preserved at all lengths. Bead diameter $\sigma = 9$ nm.

3.1.2 Bead Size Justification

The choice of 9 nm bead diameter (representing approximately 100 amino acids) is grounded in two complementary physical arguments.

Persistence length matching. The persistence length l_p is the characteristic length scale over which a polymer maintains directional correlation of its backbone. Below l_p, the chain behaves essentially as a rigid rod; above l_p, thermal fluctuations dominate and the chain behaves as a flexible random walk. Coarse-graining to beads of size approximately equal to l_p is therefore a natural choice, as it captures this fundamental transition without losing essential physical behavior.

Experimental measurements of persistence length for gel-forming mucins are summarized in Table 3.1.

Table 3.1: Experimental persistence length measurements for gel-forming mucins.

Source	Mucin Type	l_p (nm)	Method
Georgiades et al. [23]	MUC5AC, MUC2	8 ± 2	SAXS/SANS
Hughes et al. [55]	MUC5B monomers	7 ± 2	TEM
Hughes et al. [55]	MUC5B polymers	15 ± 1	TEM
Hughes et al. [55]	MUC5B glycopeptides	18 ± 4	TEM
Round et al. [56]	Ocular mucins	36	AFM

The 9 nm bead diameter falls within the range of experimentally measured persistence lengths for gel-forming mucins (7–18 nm for the glycosylated domains). This choice captures the correlation length of the mucin backbone and represents the scale at which the chain transitions from rigid to flexible behavior.

In polymer physics, the Kuhn length b_K is related to persistence length by $b_K = 2l_p$ for worm-like chains (WLC). The worm-like chain model [57] describes a polymer as a continuously flexible rod characterized by l_p; it applies to semi-flexible polymers like mucins, where the glycosylated backbone is stiff at short scales ($< l_p$) and flexible at long scales ($> l_p$). Coarse-graining polymers to the Kuhn/persistence length scale is a well-established approach in bead-spring models [24, 57] because universal behavior emerges at scales larger than l_p, so computational efficiency is greatly improved without sacrificing essential physics.

Glycan brush diameter. The 9 nm bead size also approximates the effective cross-sectional diameter of the glycosylated mucin chain. The mucin domains are heavily O-glycosylated, with dense polysaccharide side chains comprising >50% of the molecular weight. The peptide backbone itself has a diameter of approximately 0.4 nm, but the glycan brush extends roughly 4–5 nm on each side, producing an effective diameter of approximately 8–10 nm for the glycosylated regions [26]. This “hydrodynamic diameter” is the effective size of the glycosylated chain as it would be measured by solution techniques such as dynamic light scattering, which probe the overall size of the particle including its hydration shell and flexible appendages.

3.1.3 Chain Length Models

To study how chain length affects properties, simulations were conducted with chains of varying lengths. Four models were used, spanning about a 20× range from 6 beads (minimal T-H-T dimer) to 104 beads (full MUC2 dimer). The complete specifications are provided in Table 3.5 (Section 3.6).

3.2 Force Field

3.2.1 Bonded Interactions

Consecutive beads are connected by FENE bonds (Equation 2.10) combined with the repulsive LJ component (Equation 2.9), using the standard Kremer-Grest parameterization [24]: spring constant $K = 30.0\,\varepsilon/\sigma^2$ and maximum extension $R_0 = 1.5\sigma$. The spring constant $K = 30$ is chosen to be large enough that bond lengths fluctuate by less than 10% around the equilibrium distance, preventing chain crossing while still permitting thermal motion [24]. The maximum extension $R_0 = 1.5\sigma$ ensures that beads cannot pass through each other even at high temperatures. The repulsive LJ component of the FENE bond uses a cutoff of $r_c = 2^{1/6}\sigma \approx 1.12\sigma$ (the WCA cutoff), so only the repulsive core contributes to the bond potential.

3.2.2 Non-bonded Interactions

Non-bonded beads interact via the truncated LJ potential (Equation 2.9) with a cutoff distance of $r_c = 2.5\sigma$ (Phase 1) or $r_c = 2.0\sigma$ (Phase 2 and chain length study). Both cutoffs include the attractive portion of the potential; the choice of 2.0σ vs. 2.5σ has negligible effect on the results (validated in Appendix C). The interaction strengths depend on bead types:

Table 3.2: Lennard-Jones interaction parameters. Values varied during parameter sweep.

Pair	ε (range)	Physical basis
T-T	0.3–0.7 (Phase 1); 0.7–1.0 (Phase 2)	Hydrophobic attraction (drives gelation)
T-H	0.3–0.6	Cross-interaction
H-H	0.2–0.5	Weak (steric repulsion of glycans)

The physical constraint $\varepsilon_{\mathrm{TT}} \geq \varepsilon_{\mathrm{TH}} \geq \varepsilon_{\mathrm{HH}}$ ensures that terminal-terminal interactions are strongest, which is necessary for network formation.

3.3 Unit System

Simulations use Lennard-Jones reduced units, a standard convention in coarse-grained MD where all quantities are expressed in terms of the fundamental scales set by the LJ potential: the particle diameter σ (length), the well depth ε (energy), and the particle mass m (mass). All other units are derived from these three: time is measured in $\tau = \sigma\sqrt{m/\varepsilon}$, temperature in ε/k_B, and stress in ε/σ^3. This makes the simulation results independent of the specific physical system until the mapping to SI units is chosen. The conversions to SI units are:

Table 3.3: Unit conversions from LJ to SI.

Quantity	LJ unit	SI value	Basis
Length	σ	9.0 nm	Persistence length / glycan brush diameter (Section 3.1.2)
Energy	ε	4.11×10^{-21} J	$k_B T$ at 298 K
Time	τ	1404 ps	$\sqrt{m\sigma^2/\varepsilon}$, $m = 1$ (LJ units)[1]
Stress	ε/σ^3	5638 Pa	Energy density

3.4 Simulation Protocol

3.4.1 System Generation

Initial configurations are generated using self-avoiding random walk (SARW) placement:

1. Place one complete chain at a time: a random starting position and orientation is chosen, then each subsequent bead is placed at a random bond angle from the previous bead, building a random coil conformation
2. Check for overlaps between the new chain and all existing beads (minimum distance 0.9σ)
3. If any overlap is detected, reject the entire chain and retry with a new random position
4. Typical acceptance rate: 85–95% depending on concentration

This avoids the "straight chain problem" where linear (rod-like) initial configurations take much longer to equilibrate, as confirmed by the Flory scaling validation (Section 4.2).

3.4.2 Equilibration

The Phase 1 screening protocol uses:

1. **Energy minimization:** 1000 steps of conjugate gradient minimization [58, 59], an iterative optimization algorithm that adjusts particle positions to reduce the total potential energy. Unlike steepest descent (which follows the local gradient), conjugate gradient uses information from previous steps to avoid oscillating back and forth, converging more efficiently to a low-energy configuration.

2. **Gentle heating:** 50,000 steps with high damping ($\gamma = 2.0\tau^{-1}$), temperature ramped from $T = 0.1$ to the target temperature. The high damping suppresses large particle velocities during the initial relaxation when the system is far from equilibrium.

3. **NVT production:** 600,000 steps with normal damping ($\gamma = 1.0\tau^{-1}$). The damping parameter $\gamma = 1.0\tau^{-1}$ follows the standard Kremer-Grest convention [24] and produces realistic diffusive dynamics on timescales longer than γ^{-1}.

Phase 2 production simulations use the same gentle-heating protocol followed by 200,000 steps of NPT relaxation at $P = 0$ and extended NVT production runs (2,000,000 steps) to improve stress sampling statistics. The chain length study uses a more extensive equilibration (400,000 total steps: 200,000 gentle heating plus 200,000 NPT relaxation at $P = 0$) to account for the longer relaxation times of full-length chains. Production runs use NVT (constant volume) at the post-NPT box size.

The NPT step uses the LAMMPS `fix nph iso 0.0 0.0 10.0` command combined with `fix langevin` for temperature control. This implements an isotropic Nosé–Hoover barostat [60, 61] that adjusts the box volume to maintain zero external pressure, with a pressure damping parameter of 10.0τ. The Langevin thermostat controls the temperature at the target value (e.g., $T = 0.9$ in reduced units). The zero-pressure target was chosen because there is no external confining pressure on the simulated polymer solution, and the system should find its equilibrium density at the specified temperature. However, as documented in Section 3.4.2, this zero-pressure NPT equilibration produces chain-length-dependent volume changes that shift effective concentrations from nominal values. In retrospect, using NVT throughout (as in Phase 1) would have been preferable, since the box sizes were already calibrated to target concentrations.

NPT Volume Equilibration and Effective Concentrations

The NPT equilibration step in the chain length study and Phase 2 allows the simulation box to expand or contract to relieve internal pressure. Because initial box sizes are calculated from a reference scaling relation (calibrated for $N = 52$ dimers at 5 mg/mL), the equilibrium box volume after NPT can differ significantly from the initial volume, particularly for chain lengths far from the reference. This means the *effective* polymer concentration during NVT production differs from the nominal (target) concentration.

Table 3.4 summarizes the effective concentrations after NPT equilibration. Phase 1 simulations, which use NVT throughout (no NPT step), are unaffected: nominal and effective concentrations are identical.

Table 3.4: Effective concentrations after NPT volume equilibration. Phase 1 uses no NPT (concentrations exact). Phase 2 ($N = 52$) shows concentration-dependent volume changes: low concentrations contract (box too large for the cohesive polymer network) while high concentrations expand slightly. The crossover occurs near 30 mg/mL. The chain length study shows dramatic chain-length-dependent volume changes: short chains ($N = 6, 18$) expand (gas-like dispersal), while long chains ($N = 52, 104$) contract (cohesive aggregation). Phase 2 values averaged across 3 replicates per set (12 total per concentration); chain length values across 5 replicates.

Study	N	c_{nom} (mg/mL)	ΔV (%)	c_{eff} (mg/mL)	$c_{\mathrm{eff}}/c_{\mathrm{nom}}$
Phase 1 (no NPT, concentrations exact)					
Phase 1	52	10–40	0	10–40	1.00
Phase 2 ($N = 52$, all 131 simulations parsed)					
Phase 2	52	10	−40	16.7	1.67
Phase 2	52	15	−40	25.0	1.67
Phase 2	52	20	−27	27.6	1.38
Phase 2	52	25	−13	29.0	1.16
Phase 2	52	30	−1	30.5	1.02
Phase 2	52	35	+6	33.3	0.95
Phase 2	52	40	+9	36.7	0.92
Phase 2	52	45	+10	41.1	0.91
Phase 2	52	50	+9	46.0	0.92
Phase 2	52	55	+6	51.9	0.94
Phase 2	52	60	+3	58.2	0.97
Chain length study (NPT, chain-length dependent)					
Chain length	6	10	+786	1.1	0.11
Chain length	6	30	+2.5	29.3	0.98
Chain length	18	10	+346	2.2	0.22
Chain length	18	30	+105	14.6	0.49
Chain length	52	10	−43	17.6	1.76
Chain length	52	30	−5	31.5	1.05
Chain length	104	10	−65	29.5	2.95
Chain length	104	30	−21	38.1	1.27

The pattern reflects the physics of the initial configuration: the box size is calibrated for $N = 52$ at 5 mg/mL, so chains near this length equilibrate close to their initial volume. Short chains ($N = 6, 18$) with fewer hydrophobic T-beads per chain cannot form a cohesive network, so NPT relaxation expands the box toward a dilute gas state. Conversely, long chains ($N = 104$) with more hydrophobic contacts per chain contract the box as the polymer network densifies. This systematic bias means that nominal concentrations in the chain length study should be interpreted as target values, with effective concentrations reported in Table 3.4. Implications for the chain length scaling analysis are discussed in Section 6.6.

3.4.3 Data Collection

During the NVT production phase, LAMMPS outputs three types of data:

- **Stress tensor:** The six independent components of the pressure tensor (P_{xx}, P_{yy}, P_{zz}, P_{xy}, P_{xz}, P_{yz}) are written every 50 timesteps. LAMMPS computes the stress tensor from the kinetic energy of the particles plus the virial contribution from pairwise and bonded forces [62]. The three off-diagonal components (P_{xy}, P_{xz}, P_{yz}) are the inputs for the Green-Kubo rheology analysis. At $\Delta t = 0.005\tau$, 50 steps corresponds to $0.25\tau \approx 0.35$ ns temporal resolution.

- **Trajectories:** Particle positions and types are written every 1000 timesteps for pore characterization and structural analysis.

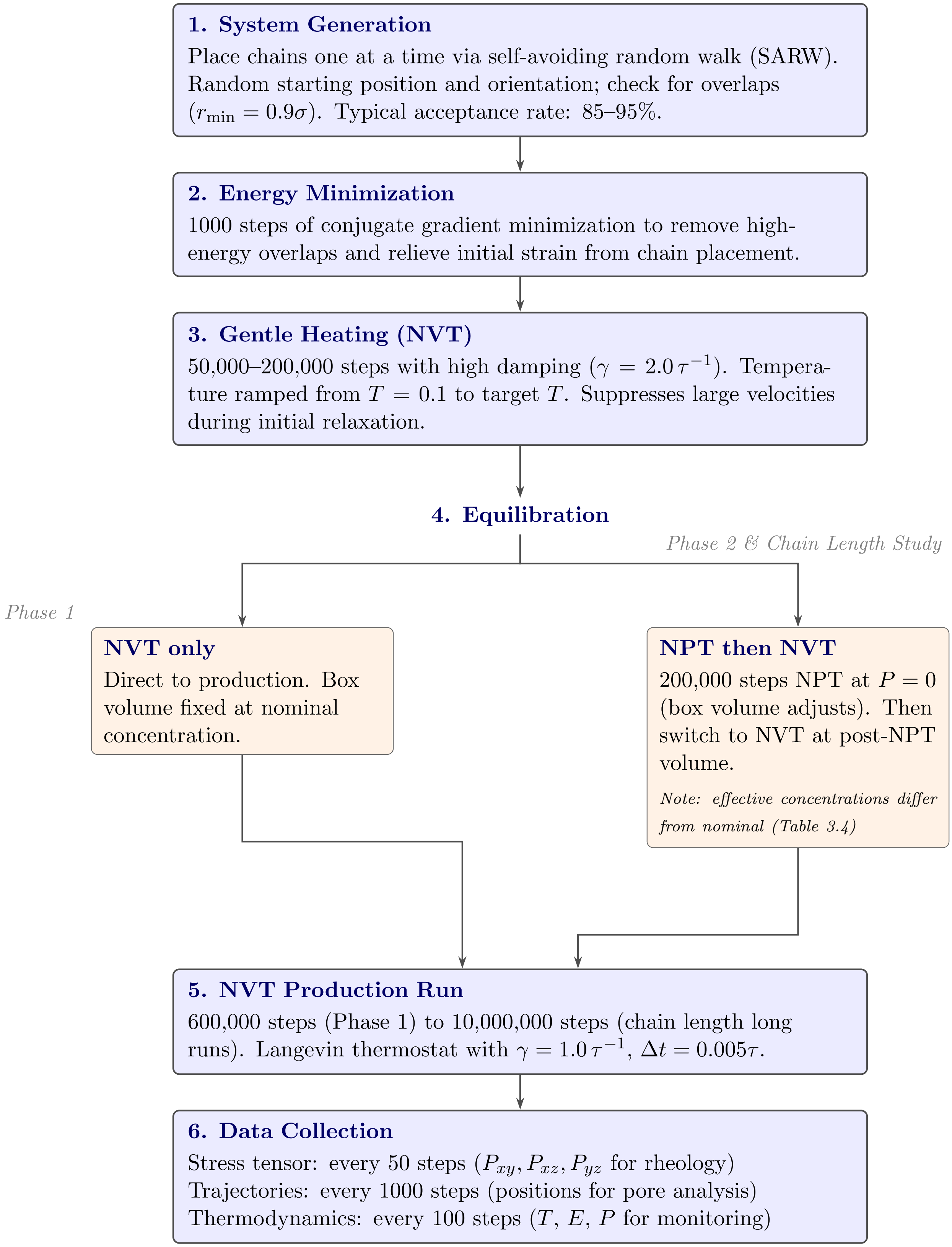
- **Thermodynamic properties:** Temperature, total energy, kinetic energy, potential energy, and pressure are written every 100 timesteps for equilibration monitoring.

3.5 Analysis Pipelines

3.5.1 Rheology Pipeline

The Green-Kubo rheology analysis proceeds as follows. All code is available in the `rheology_common.py` module (Appendix D).

1. **Load stress tensor time series.** The three off-diagonal stress components (P_{xy}, P_{xz}, P_{yz}) are read from the LAMMPS pressure output file. These are converted from LAMMPS pressure units to LJ stress units using the simulation volume.



1. System Generation

Place chains one at a time via self-avoiding random walk (SARW). Random starting position and orientation; check for overlaps ($r_{\min} = 0.9\sigma$). Typical acceptance rate: 85–95%.

2. Energy Minimization

1000 steps of conjugate gradient minimization to remove high-energy overlaps and relieve initial strain from chain placement.

3. Gentle Heating (NVT)

50,000–200,000 steps with high damping ($\gamma = 2.0\,\tau^{-1}$). Temperature ramped from $T = 0.1$ to target T. Suppresses large velocities during initial relaxation.

4. Equilibration

NVT only

Direct to production. Box volume fixed at nominal concentration.

NPT then NVT

200,000 steps NPT at $P = 0$ (box volume adjusts). Then switch to NVT at post-NPT volume.

Note: effective concentrations differ from nominal (Table 3.4)

5. NVT Production Run

600,000 steps (Phase 1) to 10,000,000 steps (chain length long runs). Langevin thermostat with $\gamma = 1.0\,\tau^{-1}$, $\Delta t = 0.005\tau$.

6. Data Collection

Stress tensor: every 50 steps (P_{xy}, P_{xz}, P_{yz} for rheology)
Trajectories: every 1000 steps (positions for pore analysis)
Thermodynamics: every 100 steps (T, E, P for monitoring)

Figure 3.2: Simulation pipeline for MUC2 coarse-grained molecular dynamics simulations. System generation uses self-avoiding random walk placement. After energy minimization and gentle heating, the equilibration protocol differs by study phase: Phase 1 uses NVT throughout (fixed box volume), while Phase 2 and the chain length study include an NPT step that allows the box to equilibrate at zero pressure before switching to NVT production. Data are collected at three frequencies during the production run.

2. **Compute the stress autocorrelation function (SACF).** For each off-diagonal component, the autocorrelation $C(t) = \langle P_{\alpha\beta}(0)P_{\alpha\beta}(t)\rangle$ is computed using the FFT-based Wiener-Khinchin method [63], which is $O(N \log N)$ rather than the $O(N^2)$ direct method.

3. **Apply frequency filtering.** The Fourier sine and cosine transforms (Equations 2.4–2.5) are only computed at frequencies for which the autocorrelation data contains $\geq$10 data points per oscillation period. This prevents aliasing and ensures adequate sampling of each frequency component.

4. **Integrate using Simpson's rule.** The Fourier integrals are evaluated numerically using Simpson's rule, which provides $O(h^4)$ accuracy for smooth functions. This replaced the initial trapezoidal integration ($O(h^2)$), which was the primary source of the early 400% integration errors.

5. **Apply 5-point moving average smoothing** to the SACF before Fourier transformation. This suppresses high-frequency noise (from thermal fluctuations in the stress tensor) without distorting the physically meaningful low-frequency content. The smoothing window (5 data points = 1.25τ) is much shorter than the relaxation times of interest.

6. **Average over 3 shear directions.** The moduli from P_{xy}, P_{xz}, and P_{yz} are averaged to improve statistics by a factor of $\sqrt{3}$.

7. **Ensemble average.** For conditions with multiple replicates ($n = 5$ for the chain length study, $n = 3$ for Phase 2), moduli are averaged across replicates and standard errors computed.

The accuracy of the pipeline was verified against analytical test cases: exponential decay functions $C(t) = Ae^{-t/\tau_c}$ with known Fourier transforms. After the improvements described above, integration errors were reduced to <1% on these test cases (Section 4.3).

3.5.2 Pore Characterization Pipeline

The Bhattacharya-Gubbins pore characterization analysis proceeds as follows:

1. **Load trajectory frames.** Particle positions are read from the LAMMPS dump file. Five frames are sampled from the production trajectory (evenly spaced) to capture temporal variation in pore structure.

2. **Generate random test points.** 5,000 random points (10,000 for the chain length study, to improve resolution for wider distributions) are generated uniformly within the simulation box for each frame.

3. **Filter points inside polymer beads.** Any test point closer than $r_{\text{bead}} = 0.5\sigma = 4.5$ nm to any bead center is discarded, as it lies inside the excluded volume of a polymer bead.

4. **Find the largest inscribed sphere.** For each valid test point, a KDTree spatial index (SciPy [64] `cKDTree`) identifies the nearest bead in $O(\log n)$ time. The pore diameter at that point is twice the distance from the test point to the nearest bead surface: $d_{\text{pore}} = 2(|\mathbf{P} - \mathbf{r}_{\text{nearest}}| - r_{\text{bead}})$. This is a simplified version of the full Bhattacharya-Gubbins optimization (Equation 2.7) that uses the test point directly as the sphere center rather than optimizing over $\mathbf{C}$.

5. **Accumulate pore diameter distribution** across all test points and frames, pooling across replicates where applicable.

6. **Compute statistics:** mean, median, interquartile range (IQR), and the full distribution for visualization.

Key parameters: bead radius $r_{\text{bead}} = 0.5\sigma = 4.5$ nm, probe radius $r_{\text{probe}} = 0.17$ nm. The probe radius sets a minimum detectable pore size ($2r_{\text{probe}} = 0.34$ nm): any void smaller than this is discarded, since no probe molecule could fit. The value 0.17 nm corresponds to the van der Waals radius of argon, following the original Bhattacharya-Gubbins convention [46]; in practice, all pores in these simulations are much larger than this threshold, so the choice of probe radius has negligible effect on the results. The KDTree data structure enables $O(\log n)$ nearest-neighbor queries, making the analysis tractable for systems with $\sim$100,000 beads. Periodic boundary conditions are applied when computing nearest-neighbor distances.

3.6 Chain Length Study Protocol

To test how chain length affects the simulation results, a systematic study was conducted using four chain lengths at two concentrations. The four chain lengths, ranging from a minimal 6-bead representation to the full 104-bead MUC2 dimer, are described in Table 3.5. Each preserves the T–H–T domain order of the MUC2 monomer at a different level of truncation. If shorter chains produce similar emergent behavior (comparable modulus scaling, pore structure, and chain conformations), they may serve as valid minimal models for future high-resolution studies where full-length MUC2 is too expensive.

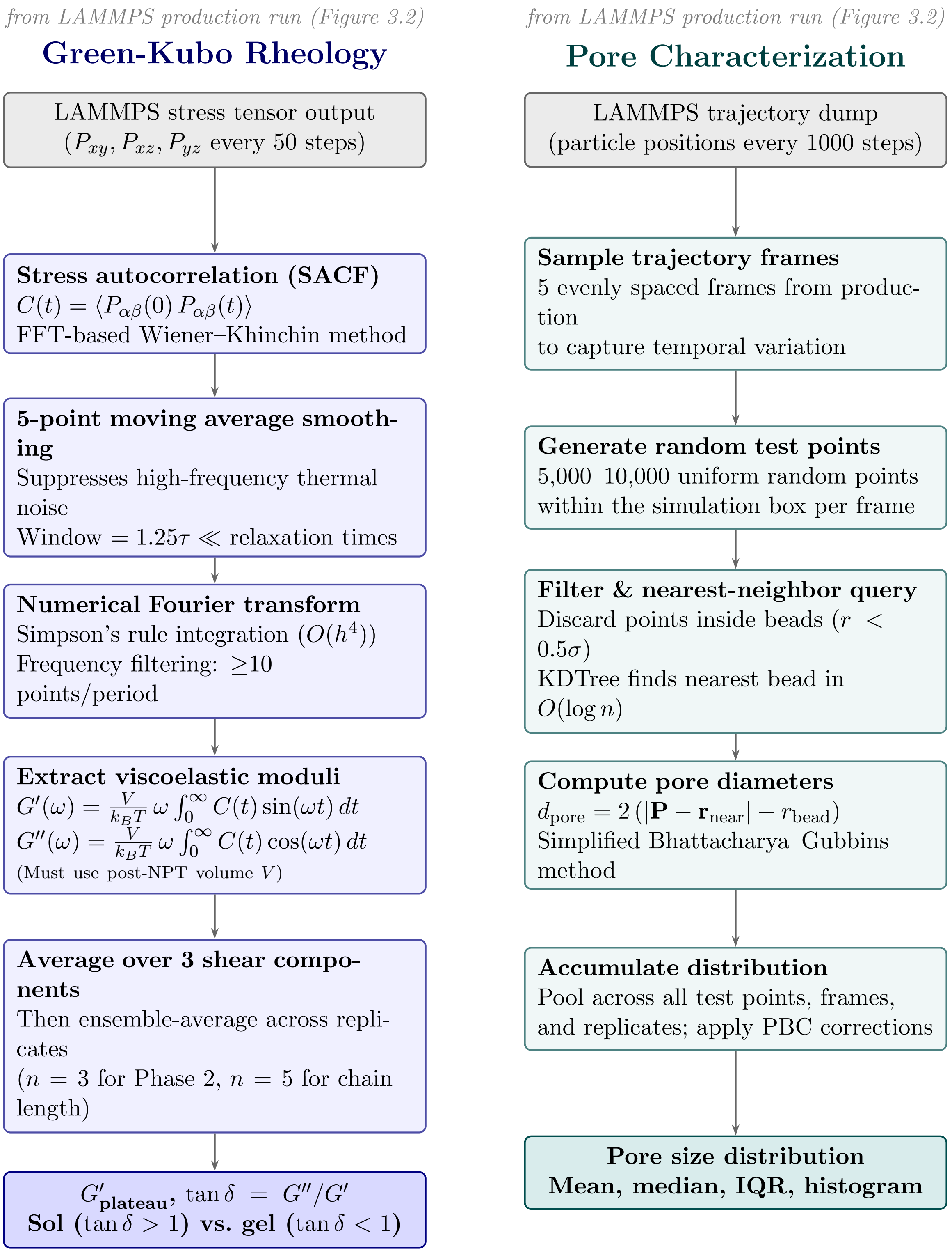


Figure 3.3: Analysis pipelines for the two primary observables. **Left:** Green-Kubo rheology pipeline, which transforms the stress tensor autocorrelation function into frequency-dependent storage (G') and loss (G'') moduli via numerical Fourier transform. The loss tangent $\tan\delta = G''/G'$ determines whether the material is sol-like ($\tan\delta > 1$) or gel-like ($\tan\delta < 1$). **Right:** Bhattacharya–Gubbins pore characterization pipeline, which estimates the pore size distribution by measuring the distance from random test points to the nearest polymer bead surface. Both pipelines read from the LAMMPS production output and are implemented in the shared `rheology_common.py` module.

3.6.1 Chain Length Models

The Ford et al. coarse-grained framework allows the MUC2 chain to be represented at different lengths while preserving the basic domain structure. Table 3.5 lists the four chain lengths tested:

Table 3.5: Chain length models used in the chain length dependency study.

Model	Pattern	N (beads/dimer)	Description
Pattern-preserving	T-H-T	6	Minimal representation
Ratio-preserving	T_2-H_6-T_1	18	Preserves T:H ratio
Half	T_5-H_{18}-T_3	52	Half-length MUC2
Full	T_{11}-H_{35}-T_6	104	Full MUC2 dimer

This provides chain lengths spanning nearly a 20× range, enabling identification of scaling relationships.

3.6.2 Concentration Selection

Two nominal concentrations were originally chosen to probe different physical regimes:

- **10 mg/mL (nominal):** At or below the overlap concentration for mucin [23], matching Phase 2's lowest concentration. At this concentration chains interact weakly, allowing intrinsic chain length effects to dominate.
- **30 mg/mL (nominal):** Above the overlap concentration, where chain entanglement becomes significant. Tests whether scaling relationships persist at higher concentrations.

Running at two concentrations enables verification that observed trends are physical rather than artifacts of a single condition. However, as discussed in Section 3.4.2, NPT equilibration at zero pressure causes chain-length-dependent volume changes that shift the effective concentrations considerably from the nominal values. For example, the shortest chains ($N = 6$) at nominal 10 mg/mL expand to an effective concentration of just 1.1 mg/mL, while the longest chains ($N = 104$) contract to 29.5 mg/mL. This means that chain length and effective concentration co-vary across the study, complicating the interpretation of apparent chain length scaling (see Section 6.6 for further discussion). Effective concentrations for all conditions are reported alongside nominal values in the results tables (Chapter 5).

3.6.3 Fixed Force Field Parameters

To isolate chain length and concentration effects, force field parameters were held constant. The primary chain length study uses the winning Phase 2 parameters (referred to as "Set D" in Chapter 5, where the full parameter set comparison is presented): $\varepsilon_{\mathrm{TT}} = 1.0$, $\varepsilon_{\mathrm{TH}} = 0.4$, $\varepsilon_{\mathrm{HH}} = 0.4$, $T = 0.9$, $r_c = 2.0\sigma$. These parameters produced the strongest elastic response in the Phase 2 optimization (Chapter 5). Simulations were conducted at three timescales. An earlier study using a different parameter set ("pre-correction": $\varepsilon_{\mathrm{TT}} = 0.4$, $\varepsilon_{\mathrm{TH}} = 0.4$, $\varepsilon_{\mathrm{HH}} = 0.3$, $T = 1.0$, $r_c = 2.5\sigma$), which had been identified as the best-performing set prior to the WCA bug correction (see Section 5.1.1 and Appendix C), at the short timescale is documented in Appendix B.

3.6.4 Simulation Protocol

The validated Phase 2 protocol was used with equilibration time scaled for chain length to account for the N^2 scaling of Rouse relaxation time:

Table 3.6: Equilibration scaling for chain length study.

Model	N	Gentle Equilibration (steps)
Pattern-preserving	6	200,000
Ratio-preserving	18	200,000
Half	52	200,000
Full	104	800,000

Rouse scaling ($\tau_R \sim N^2$) predicts that equilibration time should increase quadratically with chain length. In principle, all chain lengths should receive scaled equilibration times. In practice, only the full model ($N = 104$) was extended by a factor of 4, because the shorter chains ($N = 6$, 18, 52) equilibrate rapidly at 200,000 steps and showed no signs of incomplete relaxation (stable energy and temperature). The $N = 104$ chains, being 4× longer than the $N = 52$ baseline, required ∼4× the equilibration to avoid retaining memory of the initial conformation.

3.6.5 Replicates and Analysis

Five independent replicates were run for each chain length at each concentration, yielding 40 total simulations (4 chain lengths × 2 concentrations × 5 replicates). Each replicate uses a different random seed for both the initial chain placement (different SARW configurations) and the Langevin thermostat (different random velocity initialization), ensuring that replicates are statistically independent. Each simulation outputs:

- Stress tensor (every 50 steps) for Green-Kubo rheology
- Trajectory (every 1000 steps) for pore characterization

Analysis used the same validated pipelines as the Phase 2 production runs. Scaling exponents were extracted by fitting power laws to the chain length dependence at each concentration. As discussed in Section 3.4.2, the NPT equilibration produces chain-length-dependent volume changes that shift effective concentrations, so these scaling exponents should not be interpreted as pure chain length dependencies; they reflect a confound of chain length and concentration effects.

3.7 Computational Resources

All simulations were run on the Narval cluster (Digital Research Alliance of Canada) using LAMMPS version 20230802 [58, 59]. Analysis code was written in Python using NumPy [65], SciPy [64], and Matplotlib [66]. Each job used 4 CPUs and 8 GB of RAM.

The Phase 1 screening (4,410 simulations) required approximately 2–3 hours per simulation, totaling $\sim$44,000 CPU-hours. Phase 2 production (131 simulations) used 10–15 hours per simulation ($\sim$700 CPU-hours total). The chain length study comprised three campaigns at increasing timescales. The short-timescale campaign (40 simulations, 14 μs target) required 10–15 hours per simulation. The medium-timescale campaign (40 simulations, 420 μs target) was allocated 96 hours of walltime per job; actual production ranged from $\sim$120 μs (for the largest systems with $N = 104$, which complete fewer timesteps per hour due to the larger number of pairwise interactions) to $\sim$210 μs (for $N = 6$). The long-timescale campaign (40 simulations, 840 μs target) was allocated 168 hours; actual production ranged from $\sim$250 to $\sim$420 μs. The total computational cost of the chain length study was approximately 30,000 CPU-hours across 120 simulations. Longer simulations could be achieved through LAMMPS checkpoint/restart protocols spanning multiple SLURM jobs, but this was not implemented for this thesis.

Chapter 4

Validation Studies

Before applying the model to mucin systems, the implicit solvent environment was tested to confirm that LAMMPS with Langevin dynamics produces correct polymer physics.

4.1 Timestep and Timescale Validation

4.1.1 Timestep Selection

The integration timestep must be small enough to resolve the fastest motions in the system (bond vibrations) while being large enough to access the long timescales relevant to rheology. For FENE bonds with $K = 30.0\,\varepsilon/\sigma^2$ and $R_0 = 1.5\sigma$, the characteristic bond oscillation period is approximately $\tau_{\text{bond}} \sim 2\pi\sqrt{m/K} \approx 1.1\tau$. Standard practice requires the timestep to be at most $\sim$1/50 of this period.

The initial simulations used $\Delta t = 0.005\tau$ (7.0 ps), well within this stability criterion. The FENE-optimized timestep of $\Delta t = 0.012\tau$ (16.8 ps), used for the medium and long timescale chain length studies, provides a 2.4× speedup while remaining stable for all chain lengths tested (confirmed by monitoring bond length distributions and energy conservation). This optimized timestep was validated by comparing short test runs at $\Delta t = 0.005\tau$ and $\Delta t = 0.012\tau$: bond length distributions, radii of gyration, and stress tensor statistics were indistinguishable within statistical error. Larger timesteps were tested but proved unstable: at $\Delta t = 0.015\tau$, FENE bond length violations occurred sporadically (bonds exceeding the maximum extensibility $R_0 = 1.5\sigma$, causing LAMMPS to terminate with “Bad FENE bond” errors), and at $\Delta t = 0.02\tau$, simulations consistently crashed within the first few thousand steps. The $\Delta t = 0.012\tau$ value thus represents the practical upper limit for stable integration with this force field.

4.1.2 Accessible Timescales and Comparison to Prior Work

With the LJ time unit $\tau \approx 1.4$ ns (Table 3.3), the equivalent physical simulation timescales for a single SLURM job are:

- Phase 1 screening: 600,000 steps at $\Delta t = 0.005\tau = 3{,}000\tau \approx 4.2$ μs
- Phase 2 production: 2,000,000 steps at $\Delta t = 0.005\tau = 10{,}000\tau \approx 14$ μs
- Chain length medium: 25,000,000 steps at $\Delta t = 0.012\tau \approx 420$ μs (target)
- Chain length long: 50,000,000 steps at $\Delta t = 0.012\tau \approx 840$ μs (target)

The medium and long chain length simulations were run as single SLURM jobs (96 hours for medium, 168 hours for long). Longer simulations could be achieved through checkpoint/restart protocols spanning multiple jobs, but this was not implemented for this thesis. Larger systems ($N = 52$, 104) complete fewer timesteps per hour due to the $O(N)$ cost scaling of pairwise force calculations, so actual production times range from ∼120–210 μs (medium) and ∼250–420 μs (long) depending on chain length. The partial production data is fully usable for Green-Kubo analysis, as the stress autocorrelation function converges well before the production endpoint.

Ford et al. [26] used overdamped Brownian dynamics (no inertial term), which permits larger effective timesteps. Their 9 nm spatial resolution and ∼1 ns temporal resolution are comparable to this work, though their total simulation times were not explicitly reported. The present work extends the Ford framework by systematically exploring timescales from 4.2 μs to ∼420 μs. The chain length study (Chapter 5) reveals that moduli change with simulation timescale: this does not mean the stress autocorrelation fails to converge within each simulation, but rather that the system's underlying structure continues to evolve between simulation campaigns. Longer chains ($N = 52$, 104, with ∼5,000–10,000 beads per simulation box) develop stronger entanglement networks over hundreds of microseconds, producing different equilibrium moduli at each timescale because the system has not yet reached its final equilibrium state.

4.1.3 Gap to Experimental Timescales

Experimental bulk rheology operates on timescales of seconds to hours, while microrheology (particle tracking) accesses milliseconds to seconds. The longest simulations in this work (∼420 μs) are approximately 1,000× shorter than the millisecond timescales at which experimental rheometers operate. This timescale gap means that the simulated moduli represent the system's response at early times, before the polymer network has fully relaxed or matured. The chain length study (Chapter 5) demonstrates that moduli evolve notably between 14 μs and 420 μs, indicating that the system has not yet reached steady state. Bridging this gap to experimental timescales will require either checkpoint/restart protocols spanning multiple compute jobs or a transition to Brownian dynamics with larger timesteps.

As discussed in Section 2.5.2, Brownian dynamics eliminates the inertial term from the equation of motion, permitting timesteps 10–100× larger than the Langevin dynamics used here. A transition to Brownian dynamics, as used by Ford et al. [26], could extend the accessible simulation timescale from $\sim$420 μs toward the millisecond regime needed for direct comparison with experimental rheometry.

4.2 Flory Scaling Validation

4.2.1 Motivation

The simulations use Langevin dynamics to model implicit solvent effects, where friction and random forces replace explicit water molecules. Before applying this approach to mucin systems, it was essential to verify that LAMMPS with the Langevin thermostat produces correct polymer physics in this implicit solvent environment.

The Flory exponent ν (from $R_g \sim N^\nu$) provides a fundamental test: polymers in a proper solvent environment should adopt random coil conformations with $\nu \approx 0.5$–0.6 depending on solvent quality. This validation was performed using simple homogeneous test chains (not the mucin model) to isolate the solvent behavior from any mucin-specific effects.

4.2.2 Initial Observation: Straight Chain Initialization

Initial simulations used straight-line configurations (all beads collinear) as starting points, with the expectation that the implicit solvent dynamics would relax chains into random coil conformations. However, measurement of the Flory exponent yielded:

$$\nu = 1.054 \pm 0.004 \tag{4.1}$$

This rod-like scaling indicates that chains were not achieving equilibrium conformations within the allocated simulation time, despite appearing equilibrated by thermodynamic criteria such as temperature and total energy.

4.2.3 Solution: Self-Avoiding Random Walk Initialization

Implementation of self-avoiding random walk (SARW) initialization, where chains begin as random coils rather than straight rods, resolved this issue. With proper initialization and length-dependent equilibration times (longer chains require more time due to $\tau_R \sim N^2$ Rouse scaling), the measured exponent became:

$$\nu = 0.475 \pm 0.001 \tag{4.2}$$

This value (Figure 4.1) is close to the ideal chain prediction ($\nu = 0.5$), indicating near-theta solvent conditions in the implicit solvent environment. It falls slightly below the good-solvent prediction ($\nu \approx 0.588$), which is reasonable for a Langevin thermostat that does not fully capture explicit solvent swelling effects [36].

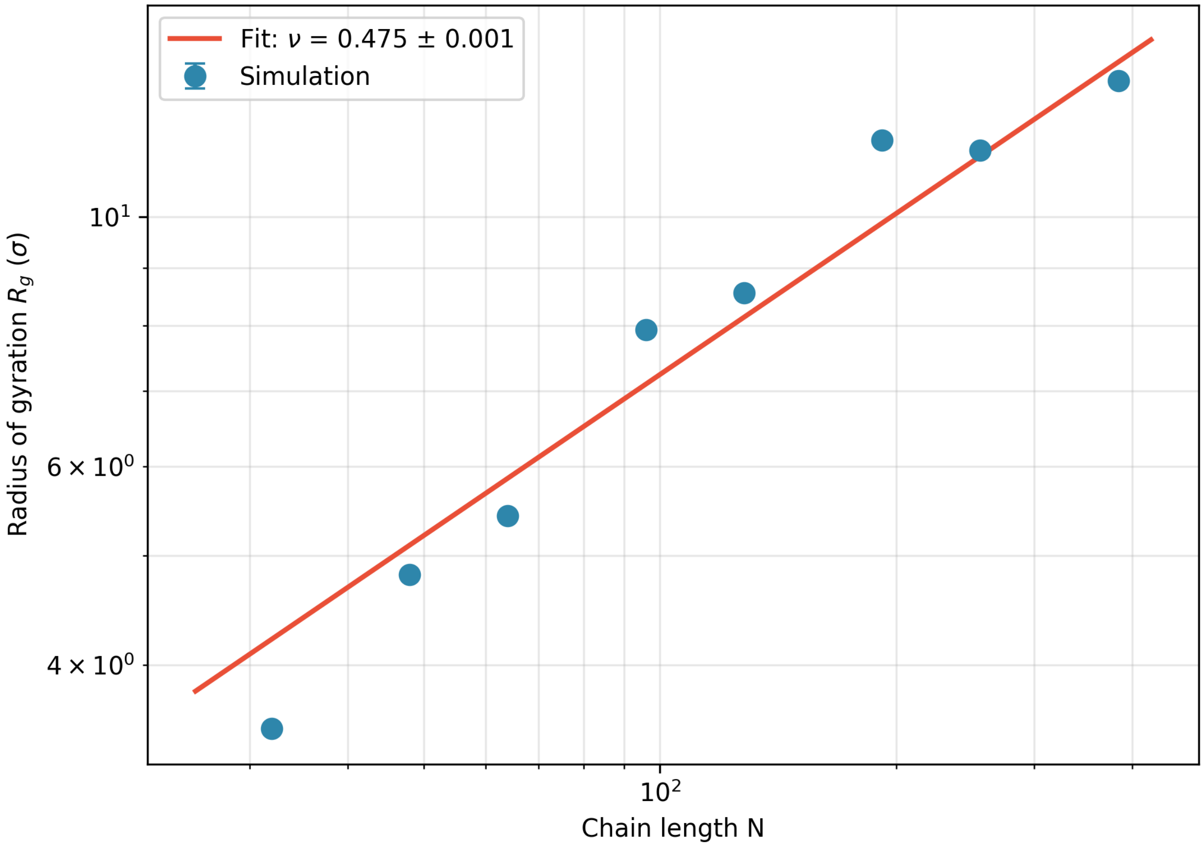


Figure 4.1: Flory scaling validation of the implicit solvent environment. Homogeneous test chains (N = 32–384 beads) in LAMMPS with Langevin dynamics show $R_g \sim N^{0.475\pm0.001}$, confirming near-ideal chain conformations ($\nu = 0.5$) expected for implicit solvent. Error bars represent standard errors from time-averaged production data (3,000–10,000 data points per chain length); bars are present but smaller than the markers for all points.

4.2.4 Validation Results

The Flory scaling analysis confirmed that LAMMPS with Langevin dynamics produces valid implicit solvent behavior: the measured exponent of $\nu \approx 0.5$ confirms proper random coil statistics. It also showed that initial conformations matter: polymer simulations can appear equilibrated by thermodynamic criteria (temperature, energy) while retaining unphysical rod-like conformations from straight-chain initialization. The Flory exponent proved to be a sensitive diagnostic that readily detects equilibration problems that would otherwise go unnoticed. Since the Rouse time scales as $\tau_R \sim N^2$, longer chains require quadratically more equilibration time; this informed the chain-length-dependent equilibration protocol used in the chain length study (Section 3.6).

4.3 Integration Quality

The Green-Kubo integrals (Equations 2.4–2.5) require careful numerical treatment because the Fourier sine and cosine transforms of the stress autocorrelation function are sensitive to sampling density, integration scheme, and noise. Initial implementations exhibited errors up to 400%, primarily from three sources: undersampling at high frequencies (insufficient data points per oscillation period), error accumulation from trapezoidal integration of oscillatory integrands, and noise amplification in the Fourier transform at short correlation times.

These issues were resolved through a combination of techniques. Frequency filtering restricts modulus computation to frequencies with at least 10 data points per oscillation period, ensuring adequate sampling of each frequency component. Simpson's rule replaces trapezoidal integration, providing significantly better accuracy for the smooth stress autocorrelation function. A 5-point moving average is applied to the autocorrelation function before Fourier transformation to suppress high-frequency noise without distorting the physically meaningful low-frequency content. Finally, ensemble averaging across multiple replicates ($n = 5$ for most conditions) reduces statistical noise in the final modulus estimates. These modifications reduced integration errors to <1% on analytical test cases (known exponential decay functions with analytically computable Fourier transforms).

4.4 Pore Analysis Validation

To validate the Bhattacharya-Gubbins pore characterization pipeline [46], Figure 4.2 shows the pore size distribution from dedicated NVT validation runs ($N = 104$ full-length dimer, $c = 10$ mg/mL, 5 replicates, 29–41$\times 10^6$ production steps at $\Delta t = 0.012\tau$, corresponding to $\sim$490–690 μs of production time per replicate) using the optimized Set D parameters ($\varepsilon_{\mathrm{TT}} = 1.0$, $T = 0.9$). These simulations are separate from the chain length study: they use NVT equilibration throughout (no NPT step), so the nominal concentration equals the actual concentration. The simulated pore sizes are compared to experimental reference ranges from the literature.

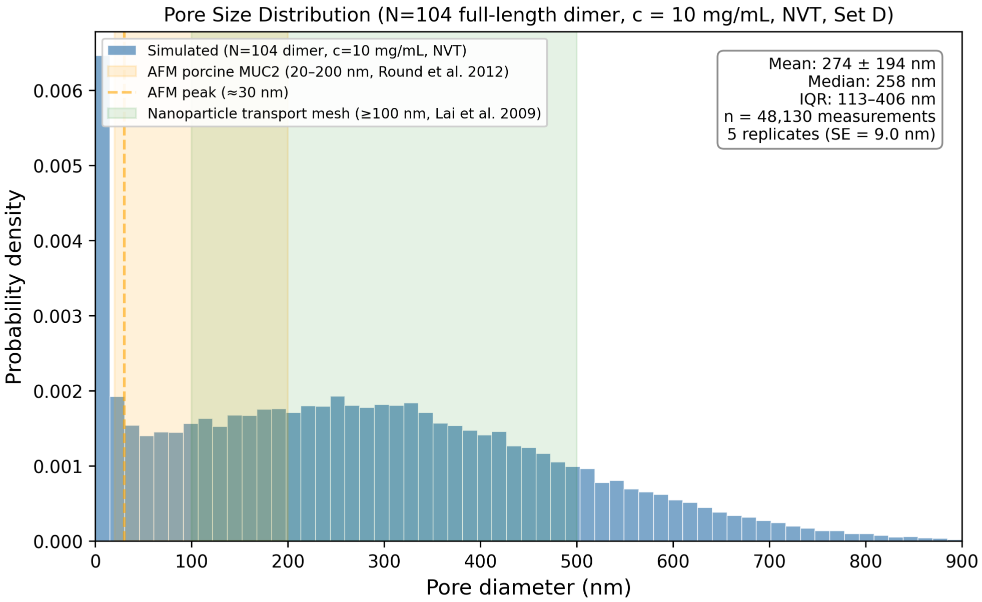


Figure 4.2: Pore size distribution from full-length ($N = 104$) MUC2 dimer simulations at $c = 10$ mg/mL (NVT, Set D parameters), validating the Bhattacharya-Gubbins analysis pipeline. Shaded regions indicate experimental pore size ranges: AFM imaging of porcine intestinal MUC2 (20–200 nm, peak at 30 nm) [67] and mesh sizes inferred from nanoparticle transport studies (∼100 nm and above, reviewed in [18]). The simulated distribution (mean 274 nm, SD 194 nm, median 258 nm, $n = 48{,}130$ across 5 replicates) overlaps with the nanoparticle transport range and exceeds most of the AFM range (peak at 30 nm), consistent with the lower concentration of these validation runs compared to native mucus.

The pore sizes from this simulation ($c = 10$ mg/mL) are comparable to experimental measurements, though direct quantitative comparison is complicated by differences in mucin type (MUC2 vs MUC5B), preparation method, and measurement technique. The key validation is that the Bhattacharya-Gubbins method produces physically reasonable pore sizes in the correct order of magnitude, supporting its use for the systematic pore characterization in Chapter 5.

Chapter 5

Results

This chapter presents results from three simulation campaigns: Phase 1 parameter screening with corrected full Lennard-Jones interactions, Phase 2 optimized production simulations with stronger hydrophobic parameters, and a systematic chain length study. The chain length study was intended to establish scaling relationships, but NPT volume equilibration produced chain-length-dependent box size changes that shifted effective concentrations far from nominal values, complicating the interpretation of apparent scaling exponents. The discovery and correction of a critical interaction potential bug (WCA) is also documented, as it fundamentally altered the interpretation of the model's capabilities.

5.1 Methodological Errors Discovered During Analysis

Two methodological errors were discovered during this work and are documented here before the corrected results.

5.1.1 WCA Interaction Potential Bug

The initial sweep of 4,410 simulations inadvertently used the Weeks-Chandler-Andersen (WCA) potential, the repulsive-only portion of the Lennard-Jones potential, instead of the intended full LJ potential. The bug originated in the LAMMPS input file:

```
pair_style lj/cut 2.5          # Global cutoff = 2.5 sigma
pair_coeff * * 1.0 1.0 1.1225  # BUG: overrides to 1.1225!
```

The per-pair cutoff of $1.1225\sigma = 2^{1/6}\sigma$ is exactly where the LJ potential crosses zero. At this cutoff, only the repulsive core is sampled and all attractive interactions are eliminated. As a result, the ε parameters become irrelevant and only hard-sphere repulsion matters. This explained why all 4,410 parameter combinations initially produced nearly identical rheology (variation <1%), which was the key observation that prompted investigation.

The fix was to explicitly set the cutoff in each `pair_coeff` line to match the `pair_style` cutoff (2.0, 2.5, or 3.0σ). All subsequent simulations (Phase 1, Phase 2, and the chain length study) use the corrected full LJ interactions. The WCA results are documented in Appendix C for methodological context.

5.1.2 NPT Volume Equilibration Effect

A second methodological issue, discovered during post-hoc analysis, affects the Phase 2 and chain length study results. These simulations use NPT equilibration with the pressure target mistakenly set to $P = 0$ (intended to let the system find its equilibrium density, but in practice allowing the box to expand or contract freely). Because the initial box sizes are calibrated from a reference scaling relation, the equilibrium box volume after NPT can differ appreciably from the initial volume, meaning the effective polymer concentration during production differs from the nominal (target) concentration.

The Green-Kubo formulas (Equations 2.4–2.5) require the actual production volume V, but the analysis scripts used the initial (pre-NPT) box volume from simulation metadata. This introduces a systematic error in the reported G' and G'' values by a factor of $V_{\text{actual}}/V_{\text{initial}}$. The loss tangent $\tan\delta = G''/G'$ is unaffected because both moduli scale by the same factor.

All modulus values reported in Sections 5.3 and 5.4 have been analytically corrected using NPT volume ratios: Phase 2 uses per-simulation volume ratios (131 individual corrections), while the chain length tables use per-concentration mean volume ratios. The effective concentrations after NPT equilibration are tabulated in Table 3.4 (Section 3.4.2). Phase 1 simulations use NVT throughout (no NPT step) and are unaffected.

5.2 Phase 1: Parameter Screening

With the corrected full Lennard-Jones potential, Phase 1 re-screened the same five-dimensional parameter space across seven concentrations (10–40 mg/mL).

5.2.1 Parameter Space

After applying physical constraints (requiring $\varepsilon_{\text{TT}} \geq \varepsilon_{\text{TH}} \geq \varepsilon_{\text{HH}}$), 4,410 parameter combinations were targeted. All 4,410 completed successfully.

Table 5.1: Parameter ranges explored in Phase 1 screening.

Parameter	Values Tested	Count
$\varepsilon_{\mathrm{TT}}$ (terminal-terminal)	0.3, 0.4, 0.5, 0.6, 0.7	5
$\varepsilon_{\mathrm{TH}}$ (cross-interaction)	0.3, 0.4, 0.5, 0.6	4
$\varepsilon_{\mathrm{HH}}$ (glycan-glycan)	0.2, 0.3, 0.4, 0.5	4
Temperature	0.8, 0.9, 1.0, 1.1, 1.2	5
LJ cutoff	2.0σ, 2.5σ, 3.0σ	3
Concentration	10, 15, 20, 25, 30, 35, 40 mg/mL	7

5.2.2 Rheological Results: Viscous Liquid Behavior

The results with correct full LJ interactions are *dramatically different* from the WCA screening. With attractive interactions enabled, the vast majority of parameter combinations produce viscous liquid (sol) behavior:

Table 5.2: Phase 1 rheological results by concentration. Values are mean ± SE ($n = 630$ per concentration). G' and G'' from Green-Kubo analysis; $\tan\delta$ is the median across simulations at each concentration.

c (mg/mL)	G' (LJ)	G'' (LJ)	Median $\tan\delta$	$\tan\delta < 1$
10.0	9.6×10^{-4}	6.4×10^{-3}	3.0	77/630
15.0	1.9×10^{-3}	1.5×10^{-2}	3.1	72/630
20.0	5.2×10^{-3}	2.8×10^{-2}	3.1	78/630
25.0	6.5×10^{-3}	4.4×10^{-2}	3.2	79/630
30.0	1.0×10^{-2}	6.5×10^{-2}	3.0	75/630
35.0	1.4×10^{-2}	8.8×10^{-2}	3.2	71/630
40.0	1.9×10^{-2}	1.1×10^{-1}	2.8	81/630

With a median $\tan\delta$ of 3.1 across all 4,410 simulations, the model behaves overwhelmingly as a viscous liquid (87.9% sol-like). Both G' and G'' increase monotonically with concentration, but the median $\tan\delta$ remains approximately constant ($\sim$3) at all concentrations, indicating that the sol-like character is robust across the tested range. The median rather than the mean is used because 37% of simulations produce negative G' values (and thus undefined or negative $\tan\delta$), reflecting Green-Kubo convergence limitations at the short 4.2 μs production timescale. The median is robust to these outliers and provides a reliable summary statistic for the skewed $\tan\delta$ distribution.

5.2.3 Concentration Scaling

The storage modulus scales as $G' \propto c^{2.2}$ (Figure 5.1), consistent with semi-dilute entangled polymer solution behavior. This quadratic scaling contrasts sharply with the linear scaling ($G' \propto c^{1.0}$) observed in the WCA screening (Appendix C), confirming that the attractive component of the LJ potential qualitatively changes the polymer physics.

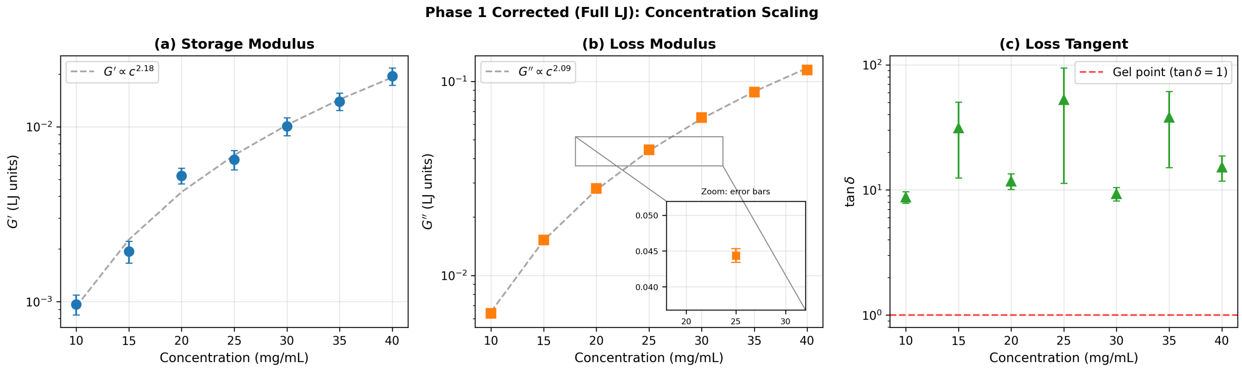


Figure 5.1: Phase 1 concentration scaling with corrected full LJ interactions. (a) Storage modulus $G' \propto c^{2.2}$ (semi-dilute entangled regime). (b) Loss modulus $G'' \propto c^{2.1}$, scaling similarly to G'. (c) Loss tangent $\tan\delta$ is approximately concentration-independent ($\sim$3), remaining $\gg 1$ (sol-like) at all concentrations tested. Error bars represent standard error of the mean across $n = 630$ simulations at each concentration; zoom inset in (b) shows error bars are present but small due to the large sample size.

5.2.4 Parameter Sensitivity

With corrected interactions, the variance in rheological properties is distributed across multiple parameters. The parameter sensitivity was quantified by computing the fraction of total variance in each rheological property (G', G'', $\tan\delta$) explained by each parameter, using one-way analysis of variance (ANOVA). For each parameter, simulations were grouped by the parameter's discrete values (e.g., $\varepsilon_{\mathrm{TT}} \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$), and the between-group sum of squares was divided by the total sum of squares to obtain the variance explained. Hydrophobic attraction $\varepsilon_{\mathrm{TT}}$ explains the largest share of G' variance (7.9%), while concentration dominates G'' variance (54.7%) and $\varepsilon_{\mathrm{TT}}$ and $\varepsilon_{\mathrm{HH}}$ together account for most $\tan\delta$ variance:

Table 5.3: Phase 1: Variance explained by each parameter.

Parameter	G' (%)	G'' (%)	$\tan\delta$ (%)
Concentration c	4.0	54.7	0.0
$\varepsilon_{\mathrm{TT}}$	7.9	2.8	29.3
$\varepsilon_{\mathrm{TH}}$	0.0	0.0	0.0
$\varepsilon_{\mathrm{HH}}$	0.0	0.0	24.9
Temperature T	0.0	1.5	0.0
LJ cutoff r_c	0.2	0.0	0.1

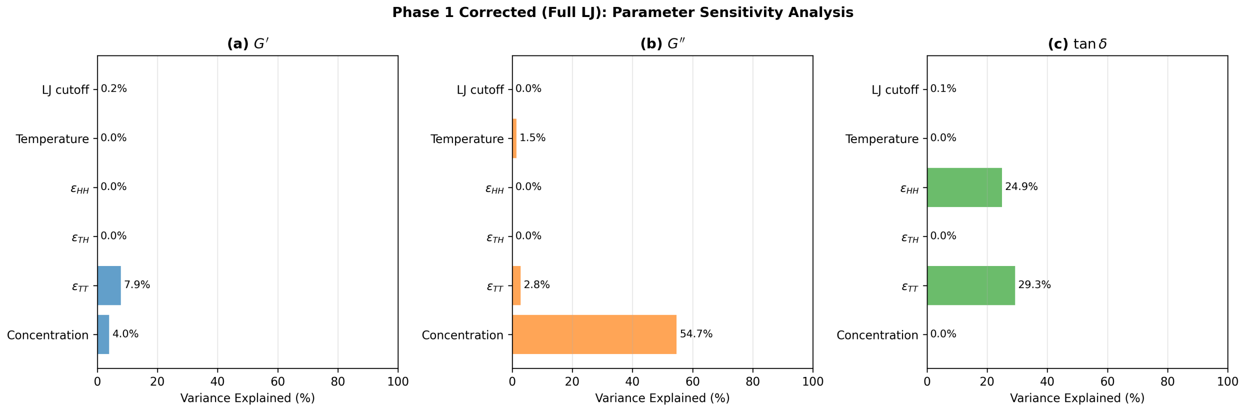


Figure 5.2: Phase 1 parameter sensitivity analysis. $\varepsilon_{\mathrm{TT}}$ explains the largest share of G' variance (7.9%), concentration dominates G'' variance (54.7%), and $\varepsilon_{\mathrm{TT}}$ and $\varepsilon_{\mathrm{HH}}$ together dominate $\tan\delta$ variance (54.2%).

5.2.5 ”Gelation” Boundary

While 87.9% of simulations are sol-like ($\tan\delta > 1$), the remaining 12.1% (533/4,410) show $\tan\delta < 1$. However, this “gel-like” fraction is largely attributable to Green-Kubo convergence limitations: 45% of these 533 simulations have *negative* G' (physically meaningless), and only 235 (5.3% of all simulations) have both positive $G' > 0.01$ (LJ units) and $\tan\delta < 1$. The short 4.2 μs production time is insufficient for reliable Green-Kubo convergence across much of the parameter space.

Among the 235 simulations with both positive $G' > 0.01$ and $\tan\delta < 1$, the pattern is clear: they cluster at high $\varepsilon_{\mathrm{TT}}$ (≥ 0.5) and $T = 0.9$. The conditions most favorable for elastic response, defined as the lowest median $\tan\delta$ (closest to gelation), are at the highest hydrophobic attraction ($\varepsilon_{\mathrm{TT}} = 0.7$, maximum tested) and $T = 0.9$, the lowest temperature that avoids unphysical clustering ($T = 0.8$ produced dense aggregates that did not equilibrate into extended networks). This region of parameter space is as close to the gelation boundary as Phase 1 got, and thus motivated Phase 2.

5.3 Phase 2: Optimized Production Simulations

Based on the Phase 1 gelation boundary, Phase 2 systematically explored stronger hydrophobic attraction to determine if gelation is achievable across a wide range of concentrations. Note that while Phase 1 reports moduli in reduced (LJ) units to facilitate comparison across ∼4,400 parameter combinations, Phase 2 and the chain length study report moduli in SI units (Pa) using the conversion factor $\varepsilon/\sigma^3 = 5638$ Pa (Table 3.3). Four parameter sets were tested with $\varepsilon_{\mathrm{TT}}$ values at and above the Phase 1 maximum:

Table 5.4: Phase 2 parameter sets. All sets use $\varepsilon_{\mathrm{TH}} = 0.4$, $\varepsilon_{\mathrm{HH}} = 0.4$, $T = 0.9$, $r_c = 2.0\sigma$ (from Phase 1 gelation boundary).

Set	$\varepsilon_{\mathrm{TT}}$	n	Description
A	0.7	33	Phase 1 winner
B	0.8	33	Moderate increase
C	0.9	32	Strong increase
D	1.0	33	Maximum tested

Simulations spanned 11 target concentrations (10–60 mg/mL) with 3 replicates per condition, yielding 131 completed simulations (one replicate missing from Set C at 25 mg/mL).

5.3.1 Rheological Results: Liquid-Like Behavior

All 131 Phase 2 simulations are liquid-like ($\tan\delta > 1$). Table 5.5 summarizes the per-set results.

Table 5.5: Phase 2 summary by parameter set (volume-corrected). Ranges span the 11 concentrations tested (10–60 mg/mL nominal). G'_{plateau} from stress variance; G' and G'' from Green-Kubo analysis. Negative G' values reflect Green-Kubo convergence failures at the 14 μs timescale. $\tan\delta$ is the median across all simulations in each set. All values corrected using per-simulation NPT volume ratios. Per-concentration breakdown is in Table 5.6. All conditions are liquid-like ($\tan\delta > 1$).

Set	ε_{TT}	G'_{plateau} range (kPa)	G' range (Pa)	G'' range (Pa)	$\tan\delta$ (median)	Sol conditions
A	0.7	23–119	−179 to 319	52–967	7.5	33/33 (100%)
B	0.8	38–138	−225 to 543	280–1,180	4.9	33/33 (100%)
C	0.9	42–141	−412 to 634	313–1,390	4.6	32/32 (100%)
D	1.0	46–147	−116 to 520	335–1,257	5.8	33/33 (100%)
Total sol conditions:						**131/131 (100%)**

Like Phase 1, the optimized Phase 2 parameters produce overwhelmingly liquid-like behavior. The volume-corrected G'_{plateau} spans 23–147 kPa across the concentration range, while the frequency-dependent Green-Kubo G' ranges from −412 to 634 Pa, with negative values reflecting convergence failures at the 14 μs timescale. The loss modulus G'' ranges from 52 to 1,390 Pa across all sets, consistently exceeding G' and confirming viscous-dominated response. The median $\tan\delta \approx 5$–8 across all sets confirms firmly liquid-like response. Per-concentration breakdown is in Table 5.6.

5.3.2 Hydrophobic Strength Controls Plateau Modulus

The volume-corrected plateau modulus G'_{plateau} (from stress variance) increases monotonically with ε_{TT} (Figure 5.3a): Set A ($\varepsilon_{\text{TT}} = 0.7$) spans 23–119 kPa while Set D ($\varepsilon_{\text{TT}} = 1.0$) spans 46–147 kPa, with both the minimum and maximum shifting upward. In all cases, $\tan\delta \gg 1$ (liquid-like; Figure 5.3d), indicating that the system remains a viscous liquid regardless of ε_{TT} at these simulation timescales.

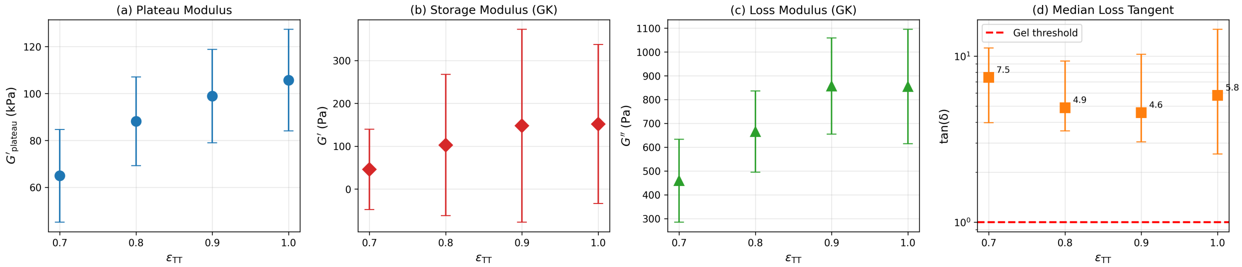


Figure 5.3: Phase 2: Parameter set comparison. (a) Mean plateau modulus G'_{plateau} vs ε_{TT}, increasing monotonically from 65 kPa to 106 kPa. (b) Mean Green-Kubo storage modulus G' vs ε_{TT}. (c) Mean Green-Kubo loss modulus G'' vs ε_{TT}. (d) Median loss tangent $\tan\delta$ vs ε_{TT}, remaining $\gg 1$ for all sets, indicating liquid-like behavior (Table 5.5). Error bars in (a)–(c): SD across all simulations within each set (spanning 11 concentrations); in (d): interquartile range.

5.3.3 Concentration Dependence

All concentrations tested (nominal 10–60 mg/mL) are liquid-like. Phase 2 simulations use NPT equilibration, which changes the box volume in a concentration-dependent manner (Table 3.4): low-concentration boxes contract ($\Delta V \approx -36\%$ at 10 mg/mL), while high-concentration boxes expand slightly ($\Delta V \approx +9\%$ at 40–50 mg/mL), with a crossover near 30 mg/mL. The nominal 10–60 mg/mL range corresponds to effective concentrations of approximately 17–58 mg/mL. All G', G'', and G'_{plateau} values in Table 5.6 have been volume-corrected using the per-concentration NPT volume ratios from all 131 simulations. The correction factor is $V_{\text{prod}}/V_{\text{init}} = 1/r$ where $r = c_{\text{eff}}/c_{\text{nom}}$. Note that $\tan\delta = G''/G'$ is unaffected by volume correction since the same factor applies to both moduli:

Table 5.6: Phase 2 results by concentration, volume-corrected using per-concentration NPT volume ratios (Table 3.4). G'_{plateau} from stress variance (mean ± SD across n simulations at each concentration, pooling all four parameter sets). G', G'', and $\tan\delta$ from Green-Kubo analysis; these are per-concentration means across all parameter sets. Green-Kubo G' and G'' are reported without uncertainty because the large inter-set spread ($\varepsilon_{\text{TT}} = 0.7$–$1.0$) makes the SD misleading as a measure of precision. All concentrations are liquid-like ($\tan\delta > 1$).

c_{nom}	c_{eff}	n	G'_{plateau} (kPa)	G' (Pa)	G'' (Pa)	$\tan\delta$
10.0	16.7	12	40.0 ± 14.7	58	299	8.3
15.0	25.0	12	58.8 ± 15.0	60	471	16.6
20.0	27.6	12	64.3 ± 12.9	148	573	5.9
25.0	29.0	11	66.8 ± 8.7	104	597	19.0
30.0	30.5	12	70.4 ± 6.2	37	557	13.0
35.0	33.3	12	76.6 ± 8.0	106	565	12.1
40.0	36.7	12	84.9 ± 9.5	95	697	7.2
45.0	41.1	12	95.2 ± 12.0	131	792	11.3
50.0	46.0	12	106.3 ± 11.8	151	911	7.0
55.0	51.9	12	120.2 ± 14.5	174	876	8.5
60.0	58.2	12	134.6 ± 17.6	62	917	8.5

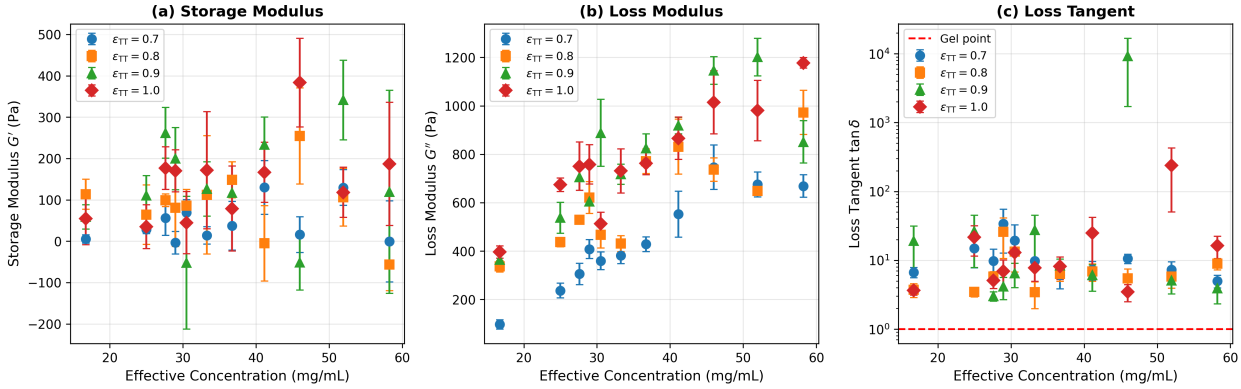


Figure 5.4: Phase 2: Volume-corrected Green-Kubo rheology vs effective concentration for all four parameter sets (Sets A–D, $\varepsilon_{\text{TT}} = 0.7$–$1.0$). (a) Storage modulus G', (b) loss modulus G'', and (c) loss tangent $\tan\delta$ vs c_{eff} (NPT-corrected). All conditions remain liquid-like ($\tan\delta > 1$). Error bars: SE ($n = 3$ replicates per condition).

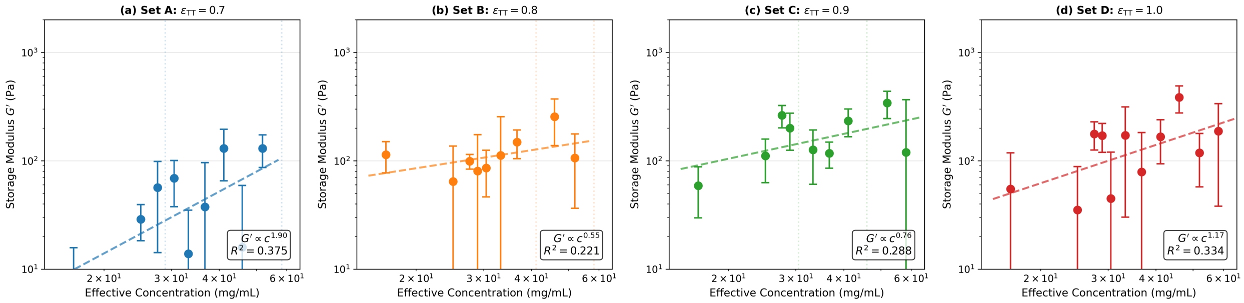


Figure 5.5: Concentration scaling of volume-corrected G' from Green-Kubo analysis for all Phase 2 parameter sets (A–D). Each panel shows G' vs effective concentration (NPT-corrected) on log-log axes with power-law fits. Concentrations where the mean G' is negative are excluded from fits, reflecting Green-Kubo convergence noise at the 14 μs timescale. Error bars: SE ($n = 3$ replicates per condition).

5.3.4 Pore Characterization

Pore size characterization was performed on all 131 Phase 2 simulations using the Bhattacharya-Gubbins method (5,000 test points per frame, probe radius 0.17 nm).

Table 5.7: Phase 2 pore size by parameter set.

Set	ε_{TT}	Mean Diam. (nm)	Median Diam. (nm)	Mean n_{pores}
A	0.7	12.5 ± 2.0	12.0 ± 2.0	697
B	0.8	13.2 ± 2.6	12.9 ± 2.7	758
C	0.9	13.6 ± 2.9	13.4 ± 3.1	827
D	1.0	14.1 ± 3.1	13.9 ± 3.4	880
Overall		13.3 ± 2.7	13.0 ± 2.9	790

Pore diameter increases with ε_{TT}, from 12.5 nm for Set A ($\varepsilon_{TT} = 0.7$, the weakest Phase 2 hydrophobic attraction) to 14.1 nm for Set D ($\varepsilon_{TT} = 1.0$, the strongest), consistent with stronger aggregation creating larger voids between polymer clusters (Table 5.7). The concentration dependence is non-monotonic (Figure 5.6a): pore diameter peaks at ∼17 nm near effective concentrations of ∼30 mg/mL, then decreases at higher concentrations as the box becomes densely packed with polymer.

At low effective concentrations ($c_{\text{eff}} \approx$ 17–25 mg/mL), the chains form isolated clusters with relatively large inter-cluster voids, producing the largest pores. As concentration increases, the clusters begin to overlap and fill space more uniformly, reducing the void sizes. The anti-correlation between pore size and plateau modulus at intermediate concentrations (Figure 5.6c) suggests that the structural transition from clustered to uniform packing is accompanied by a shift in mechanical response. These Phase 2 pore sizes (12–17 nm) are considerably smaller than experimental estimates for native mucus (20–400 nm from AFM and nanoparticle transport studies), likely reflecting the tight aggregation that occurs at the 14 μs simulation timescale before the network has fully matured.

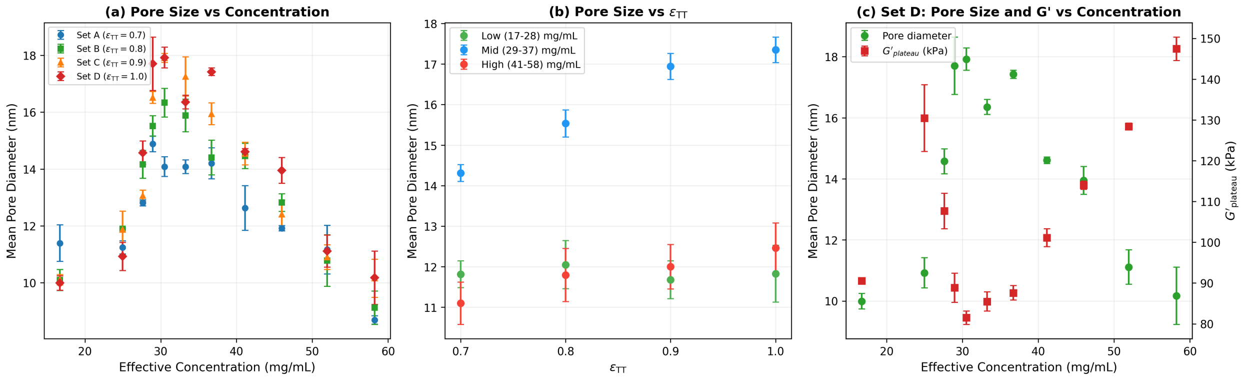


Figure 5.6: Phase 2 pore size analysis using NPT-corrected effective concentrations. (a) Mean pore diameter vs effective concentration by parameter set, showing non-monotonic (dome-shaped) pattern peaking near 30 mg/mL. (b) Pore size vs ε_{TT} grouped by effective concentration range: stronger hydrophobic attraction consistently produces larger pores. (c) Set D dual-axis comparison of pore diameter and G'_{plateau} vs effective concentration, showing that pore size and modulus are anti-correlated at mid-range concentrations. Error bars represent standard error of the mean across replicate simulations.

5.3.5 Phase 1 vs Phase 2 Comparison

Phase 2 results are qualitatively consistent with Phase 1, as both show liquid-like behavior:

Table 5.8: Phase 1 vs Phase 2 comparison. Both phases show liquid-like behavior.

Property	Phase 1	Phase 2
Parameter sets	Standard sweep	Optimized from winner
ε_{TT} range	0.3–0.7	0.7–1.0
Temperature	0.8–1.2	0.9 (fixed)
$\tan\delta$ (median)	∼3	∼6
Sol-like ($\tan\delta > 1$)	3,877/4,410 (87.9%)	131/131 (100%)
G' range (per-conc. mean)	∼5–110 Pa	∼37–174 Pa
G'' range	∼36–620 Pa	∼52–1,390 Pa
Physical interpretation	Viscous liquid	Viscous liquid

Phase 2 produces moderately higher per-concentration mean storage modulus G' (∼37–174 Pa vs. ∼5–110 Pa in Phase 1) and higher loss modulus G'' (∼52–1,390 Pa vs. ∼36–

620 Pa). Both phases maintain $\tan\delta \gg 1$, confirming liquid-like behavior at the 14 μs simulation timescale. The interpretation of these differences is discussed in Section 6.3.

5.4 Chain Length Study (Exploratory)

An exploratory chain length study was conducted using four coarse-grained models ($N = 6$, 18, 52, and 104 beads per dimer) at two nominal concentrations (10 and 30 mg/mL) with five replicates each, using Set D parameters ($\varepsilon_{\mathrm{TT}} = 1.0$, $T = 0.9$) at three timescales (14 μs to $\sim$420 μs). Due to the NPT volume equilibration issue (Section 3.4.2), chain length and effective concentration co-vary across the study (Table 3.4), making it difficult to isolate pure chain length effects. The study is therefore labeled exploratory, and the full results are presented in Appendix A.

The principal findings are:

- **All conditions remain liquid-like** ($\tan\delta \gg 1$) at all timescales, consistent with Phase 1 and Phase 2.
- Short chains ($N = 6$, 18) at low effective concentration relax and lose elastic structure at longer timescales, while longer chains ($N = 52$, 104) at higher effective concentration develop stronger elastic response over time.
- Pore sizes are strongly dependent on effective concentration, with short chains at low c_{eff} developing very large pores at extended timescales.
- The apparent chain length scaling exponents are confounded with NPT-induced concentration differences and should not be interpreted as pure chain length effects.

5.5 Summary of Results

1. **WCA bug discovered and corrected.** The initial parameter sweep used repulsive-only (WCA) interactions due to a cutoff specification error.
2. **Phase 1: Overwhelmingly liquid-like.** With correct full LJ interactions, 4,410 parameter combinations produce a median $\tan\delta$ of 3.1 (87.9% sol-like). Both G' and G'' increase with concentration, but the median $\tan\delta$ remains approximately constant ($\sim$3) across all concentrations. Green-Kubo convergence is limited at the 4.2 μs timescale (37% negative G'). The strongest elastic response occurs at $\varepsilon_{\mathrm{TT}} = 0.7$, $T = 0.9$. Concentration scaling follows $G' \propto c^{2.2}$.

3. **Phase 2: Liquid-like with stronger interactions.** Testing $\varepsilon_{\mathrm{TT}} = 0.7$–$1.0$ at $T = 0.9$ across 11 concentrations yields liquid-like behavior ($\tan\delta \gg 1$) in all 131 simulations. Volume-corrected plateau modulus G'_{plateau} spans 23–147 kPa across concentrations and parameter sets, reflecting the high-frequency stress variance. Stronger hydrophobic attraction increases both moduli but does not achieve gelation at the 14 μs timescale.

4. **Phase 2 pores:** Mean pore diameter of 13.3 nm, increasing with $\varepsilon_{\mathrm{TT}}$ (12.5–14.1 nm) and showing non-monotonic concentration dependence.

5. **Chain length study (exploratory, Set D):** All conditions are **liquid-like** ($\tan\delta \gg 1$) at all timescales (14 μs to $\sim$420 μs), consistent with Phase 1 and Phase 2. NPT volume equilibration shifts effective concentrations from nominal values (Table 3.4), confounding chain length and concentration effects. Short chains relax at longer timescales while long chains strengthen; pore sizes are strongly concentration-dependent. Full results are in Appendix A.

6. **Computational constraints:** The Narval cluster enforces a maximum 168-hour (7-day) walltime per job. With 100,000 beads and FENE-optimized $\Delta t = 0.012\tau$, achievable simulation times are $\sim$120–420 μs per job depending on system size, orders of magnitude short of the μs–ms timescales needed for direct comparison with experimental rheology. Medium (96-hour walltime) and long (168-hour walltime) runs both achieved only 30–50% of target production steps, with larger systems completing fewer steps due to denser neighbor lists.

Chapter 6

Discussion

6.1 Lessons from the WCA Bug

The discovery that the initial 4,410-simulation sweep used WCA (repulsive-only) instead of full LJ interactions is perhaps the most consequential methodological finding of this thesis. The bug was subtle: the LAMMPS input file specified a global cutoff of 2.5σ via `pair_style`, but the per-pair cutoff of 1.1225σ in `pair_coeff` silently overrode it.

The key diagnostic was that all 4,410 parameter combinations produced nearly identical rheology (variation <1%). With WCA interactions, the ε values become irrelevant because only the repulsive core is sampled. This physically implausible uniformity prompted investigation of the actual potential being computed.

Practical implications for the field. The LAMMPS cutoff hierarchy (`pair_coeff` overrides `pair_style`) is documented but easy to overlook, particularly when following templates from published work. This experience suggests that **explicit verification of interaction potentials**, for example by checking energy profiles or comparing to analytical LJ values, should be standard practice when setting up new simulations.

6.2 Phase 1: The Gelation Landscape

6.2.1 A Viscous Liquid

The Phase 1 screening reveals that the coarse-grained MUC2 model is overwhelmingly a viscous liquid under standard parameter conditions. With a median $\tan\delta$ of 3.1 across 4,410 simulations, the loss modulus dominates the storage modulus by roughly 3:1. This is consistent with a dilute-to-semidilute polymer solution where chains interact but do not form a persistent network. The Green-Kubo convergence is limited at the 4.2 μs production timescale (37% of simulations produce negative G' values), but the median $\tan\delta$ provides a robust summary: the system is sol-like across all concentrations tested.

The $G' \propto c^{2.2}$ scaling is consistent with semi-dilute entangled behavior [57, 33]. This is expected because at concentrations of 10–40 mg/mL, the 52-bead chains are above the overlap concentration ($c^* < 1$ mg/mL for mucins [23]), so chains interpenetrate and form transient entanglements that contribute to elasticity. The liquid-like behavior ($\tan\delta \gg 1$) indicates that these entanglements are not sufficient for gelation on the accessible timescale, consistent with the absence of permanent crosslinks in the model.

6.2.2 The Gelation Boundary

The strongest elastic response in Phase 1 occurs consistently at $\varepsilon_{\mathrm{TT}} = 0.7$ (maximum tested) and $T = 0.9$, the lowest temperature producing physically reasonable behavior ($T = 0.8$ was tested but caused phase separation artifacts). This defines a gelation boundary in parameter space and motivated Phase 2's exploration of even stronger T-T attractions ($\varepsilon_{\mathrm{TT}} = 0.7$–$1.0$).

6.3 Phase 2: Hydrophobic Attraction and Elastic Response

6.3.1 Phase 2: Consistency with Phase 1

Phase 2 is consistent with Phase 1: the model is still liquid-like ($\tan\delta \gg 1$) under all conditions at the 14 μs timescale. The optimized parameters ($\varepsilon_{\mathrm{TT}} = 0.7$–$1.0$, $T = 0.9$) give per-concentration mean G' ranging from 37 Pa (at $c_{\mathrm{nom}} = 30$ mg/mL) to 174 Pa (at $c_{\mathrm{nom}} = 55$ mg/mL), with values at individual concentrations reflecting Green-Kubo convergence noise (Table 5.6), somewhat higher than Phase 1 ($G' \approx 5$–110 Pa after SI conversion). So stronger hydrophobic attraction shifts the elastic (storage) modulus upward. The loss modulus G'' also increases with $\varepsilon_{\mathrm{TT}}$ (from ∼52–967 Pa in Set A to ∼335–1,257 Pa in Set D), indicating that the viscous dissipation also increases, likely because the stronger hydrophobic contacts create more inter-chain friction. Both moduli shift upward, but the increase is not enough to cross the gelation threshold at these timescales.

6.3.2 Hydrophobic Attraction Controls Plateau Modulus

Increasing $\varepsilon_{\mathrm{TT}}$ increases the mean plateau modulus across concentrations (64.9 kPa → 105.7 kPa, a 63% increase; Figure 5.3a) and the per-concentration mean Green-Kubo modulus

($G' \approx$ 37–174 Pa; Table 5.6). However, $\tan\delta$ remains $\gg 1$ in all cases, so the system has not crossed the gelation threshold at these timescales. Whether gelation would eventually occur at longer timescales (μs–ms) remains an open question.

6.3.3 Modulus Values and Timescale Considerations

The Phase 2 plateau moduli (23–147 kPa across concentrations and parameter sets) represent the high-frequency elastic response from stress variance and exceed the physiological range for mucus gels (10–100 Pa) by 2–3 orders of magnitude. The frequency-dependent per-concentration mean Green-Kubo moduli ($G' \approx$ 37–174 Pa; Table 5.6) are much closer to physiological values, suggesting that the model captures the correct order of magnitude for elastic response at the accessible frequencies. For comparison, Wagner et al. [68] measured $G' \approx$ 1–10 Pa for reconstituted MUC5AC gels at 1–15 mg/mL using bulk oscillatory rheology, while native mucus samples typically yield $G' \sim$ 10–100 Pa [18, 19].

The chain length study (Set D) provides direct evidence for timescale dependence of moduli. Comparing Set D across three timescales (holding parameters constant), short chains at nominal 10 mg/mL show dramatically reduced volume-corrected G': $N = 18$ ($c_{\text{eff}} = 2.2$ mg/mL) drops from 4,790 Pa (14 μs) to 364 Pa (∼120–210 μs) to 327 Pa (∼250–420 μs); $N = 6$ ($c_{\text{eff}} = 1.1$ mg/mL) drops from 2,200 Pa to 3.6 Pa (medium) and stabilizes near 4 Pa (long). Conversely, long chains show *increased* G': $N = 52$ ($c_{\text{eff}} = 17.6$ mg/mL) rises from 40 kPa to 890 kPa and stabilizes, and $N = 104$ ($c_{\text{eff}} = 29.5$ mg/mL) from 71 kPa to 614 kPa. The convergence of G' between medium and long timescales for both short and long chains suggests the system has reached a quasi-steady state, but this steady state remains liquid-like ($\tan\delta \gg 1$) at all concentrations and chain lengths, indicating that gelation does not occur on these timescales. This chain-length-dependent timescale response highlights the importance of matching simulation timescale to chain relaxation time: short chains equilibrate within 14 μs and lose elastic structure, while long chains continue developing entanglements that increase G' until stabilizing by ∼250–420 μs. Note that these chain lengths are compared at very different effective concentrations (1.1–29.5 mg/mL at nominal 10; Table 3.4), so the apparent chain length dependence is confounded with concentration effects.

6.3.4 Pore Structure in Phase 2

The Phase 2 pore diameters (12.5–14.1 nm) are notably smaller than both the pre-correction chain length study pores (70–200 nm; Appendix B) and experimental measurements (20–400 nm). Part of this difference reflects the stronger attractive interactions in Phase 2, which promote tight aggregation and leave smaller voids. However, the coarse-grained representation itself may also contribute: the glycosylated regions are modeled as solid spheres, whereas real mucin glycans form flexible bottle-brush structures with considerable void space between the polysaccharide side chains. This means the model beads fill more volume than the real glycans do, potentially compressing the pore space and producing systematically smaller pores than experiment. The increase in pore size with $\varepsilon_{\mathrm{TT}}$ (from 12.5 nm in Set A to 14.1 nm in Set D) is counterintuitive but consistent with stronger attraction producing more compact aggregates separated by larger inter-aggregate voids. The pore analysis is independent of the rheological classification.

The chain length study (Section 6.4) extends this picture with pore analysis across all four chain lengths at three timescales, revealing a timescale-dependent bifurcation between short and long chains at low concentration (Section 6.4).

6.4 Chain Length Scaling

6.4.1 Modulus Scaling: Timescale-Dependent Chain Length Exponents

A critical complication in interpreting the chain length scaling is the NPT volume equilibration effect (Section 3.4.2). At nominal 10 mg/mL, effective concentrations range from 1.1 mg/mL ($N = 6$, box expansion) to 29.5 mg/mL ($N = 104$, box contraction). Since G' depends on both chain length and concentration, the apparent scaling exponent α in $G' \propto N^{\alpha}$ reflects both effects and cannot be interpreted as a pure chain length dependence.

After applying volume corrections to the Green-Kubo analysis (rescaling G' by $V_{\mathrm{actual}}/V_{\mathrm{initial}} = c_{\mathrm{nom}}/c_{\mathrm{eff}}$; see Table A.1), the Set D short-timescale exponent at nominal 10 mg/mL decreases from $\alpha \approx 2.4$ to $\alpha \approx 1.3$. The uncorrected exponent was inflated because longer chains contract to higher effective concentrations (boosting their G') while shorter chains expand to lower effective concentrations (depressing their G'). The corrected exponent of $\alpha \approx 1.3$ is lower than the uncorrected value, though it remains confounded with concentration and should not be compared directly to theoretical predictions (e.g., Rouse or reptation scaling) that assume fixed concentration. An earlier study with weaker pre-correction parameters (Appendix B) yielded a similar uncorrected short-timescale exponent ($\alpha \approx 2.3$); that study also used NPT and would require similar correction.

At medium timescale ($\sim$120–210 μs), the chain-length-dependent relaxation remains a robust observation even after volume correction: short chains ($N = 6, 18$) relax and lose elastic structure, while long chains ($N = 52, 104$) develop stronger elastic response over time. However, the specific scaling exponents should not be interpreted quantitatively given the concentration confound.

6.4.2 The Anomalous $N = 104$ Concentration Inversion

The uncorrected analysis showed inverted concentration scaling ($G'(c_{\text{nom}} = 10) > G'(c_{\text{nom}} = 30)$) for $N \geq 52$ across all three timescales. Volume correction reveals that this inversion has two distinct components:

At the short timescale (14 μs), the inversion disappears after volume correction: for $N = 104$, $G'(c_{\text{eff}} = 29.5) = 70{,}900$ Pa $< G'(c_{\text{eff}} = 38.1) = 90{,}300$ Pa, restoring normal concentration scaling. The apparent inversion in uncorrected data was entirely an artifact of the NPT volume effect: the nominal 10 mg/mL system contracted to a higher effective concentration (29.5 mg/mL) than the nominal 30 mg/mL system (38.1 mg/mL), but both had similar G' values. The original analysis, which used pre-NPT volumes, overcounted G' for the contracted (nominal 10) system and undercounted for the less-contracted (nominal 30) system.

At medium and long timescales ($\sim$120–420 μs), however, the inversion persists even after volume correction. For $N = 52$ at the long timescale, $G'(c_{\text{eff}} = 17.6) = 890{,}500$ Pa $> G'(c_{\text{eff}} = 31.5) = 397{,}200$ Pa. This genuine inversion suggests kinetic trapping: at lower effective concentration, long chains form isolated aggregates that become trapped in high-stress conformations over extended simulation times, while at higher concentration, inter-aggregate interactions may facilitate relaxation. The volume correction thus resolves the short-timescale anomaly but reveals a genuine kinetic trapping phenomenon at longer timescales.

6.4.3 Pore Sizes

The Set D pore analysis across all four chain lengths at three timescales (Tables A.4–A.6) shows pore sizes strongly dependent on effective concentration, which varies dramatically due to NPT volume changes. At the short timescale and nominal 10 mg/mL, $N = 6$ ($c_{\text{eff}} = 1.1$ mg/mL) produces the largest pores (median 273 nm) while $N = 52$ ($c_{\text{eff}} = 17.6$ mg/mL) produces the smallest (median 91 nm). At medium and long timescales, a striking bifurcation emerges: short chains at low effective concentration develop very large pores ($N = 6$: median $\sim$2,800 nm; $N = 18$: median $\sim$650 nm), while longer chains ($N = 52$, 104) increase from their short-timescale values (91 and 143 nm) to $\sim$300 nm at medium timescale and stabilize there. At nominal 30 mg/mL, $N = 18$ and $N = 52$ show significant contraction from short to medium timescales (46→13 nm and 49→18 nm respectively), while $N = 6$ and $N = 104$ change modestly. Once past the medium timescale, all pore sizes at

30 mg/mL stabilize (median 13–43 nm), consistent with the rheological observation that G' also converges between medium and long timescales.

The pre-correction chain length study (Appendix B), with weaker parameters ($\varepsilon_{\mathrm{TT}} = 0.4$, $T = 1.0$), produced pore sizes (median 70–200 nm) in good agreement with experimental measurements: AFM studies of porcine intestinal MUC2 report 20–200 nm [67], and nanoparticle transport studies suggest mesh sizes of $\sim$100 nm and above (reviewed in [18]). Set D's generally larger pores at low concentrations reflect stronger aggregation creating larger inter-cluster voids rather than a uniform polymer mesh. At 30 mg/mL, both parameter sets converge on similar pore sizes (34–49 nm for Set D vs 72–126 nm for the pre-correction set), suggesting that at higher concentrations, packing effects dominate over interaction strength.

6.4.4 Implications for Future Atomistic Simulations

The volume-corrected short-timescale scaling ($G' \propto N^{1.3}$, Set D, nominal 10 mg/mL) provides a sound methodological approach for selecting chain lengths, though effectively the exponent is confounded with NPT-induced concentration differences (Section 3.4.2). However, the observation that longer chains produce higher G' is robust across timescales and both parameter sets, even if the specific scaling exponents should be interpreted with much caution given the concentration issues.

The $N = 18$ model at nominal 10 mg/mL ($c_{\mathrm{eff}} = 2.2$ mg/mL) produces volume-corrected $G' = 327$ Pa at the long timescale ($\sim$250–420 μs), above the physiological range (10–100 Pa) but at an effective concentration well below physiological values. Future chain length studies should use NVT-only equilibration (as in Phase 1) or carefully validate post-NPT concentrations to avoid this systematic confound.

6.5 Comparison to Prior Work

Ford et al. [26] established the coarse-grained modeling approach for gel-forming mucins, focusing on MUC5B assembly and aggregation kinetics at two concentrations (1.5 and 5 mg/mL) while exploring ranges of hydrophobic and electrostatic interaction strengths. Ford also computed equilibrium linear viscoelastic moduli from statistical physics summation formulas.

The most significant difference between this work and Ford's is probably the scale of parameter exploration: 4,410 combinations across five dimensions versus Ford's more focused approach. This made it possible to map out the gelation boundary systematically and identify which parameters actually matter ($\varepsilon_{\mathrm{TT}}$ and concentration explain most of the variance). The rheological method is also different: Green-Kubo stress autocorrelation gives frequency-dependent moduli directly from the pressure tensor [69], which is complementary to Ford's statistical mechanics approach and allows more direct comparison with experimental oscillatory rheology.

The chain length study provides scaling data across four model sizes ($N = 6$–104), which Ford did not attempt, though the NPT volume confound (Section 3.4.2) complicates quantitative interpretation of the exponents. The Bhattacharya-Gubbins pore characterization serves as a structural validation that was also absent from Ford's work, while the WCA bug documentation is a cautionary example for others working with LAMMPS cutoff specifications.

6.6 Limitations

6.6.1 Coarse-Graining Approximations

Each bead represents ∼100 amino acids. This loses all atomic-level detail: specific hydrogen bonds, glycan branching patterns, sequence-specific interactions, and ion binding. The interaction parameters are empirical rather than derived from atomistic simulations.

6.6.2 NPT Volume Equilibration Bias

The chain length study and Phase 2 simulations use NPT equilibration, which produces chain-length-dependent box volume changes (Table 3.4). This has two consequences:

1. **Effective concentrations differ from nominal values.** At nominal 10 mg/mL, effective concentrations range from 1.1 mg/mL ($N = 6$, +786% volume expansion) to 29.5 mg/mL ($N = 104$, −65% contraction). This means the chain lengths are not compared at equal concentrations, confounding the apparent chain length scaling exponents.

2. **The Green-Kubo volume prefactor was computed with pre-NPT box volumes.** The Green-Kubo formula $G = (V/k_BT)\langle P_{xy}(t)P_{xy}(0)\rangle$ requires the actual production volume, but the analysis scripts used the initial (pre-NPT) box volume from simulation metadata. All G' and G'' values in this thesis have been analytically corrected by the factor $V_{\text{actual}}/V_{\text{initial}} = c_{\text{nom}}/c_{\text{eff}}$ (Tables A.1–A.3). For the chain length tables, this correction is approximate because it uses mean volume changes averaged across replicates rather than per-replicate volumes; Phase 2 corrections use

per-simulation volume ratios. The loss tangent $\tan\delta = G''/G'$ is unaffected because both moduli scale by the same factor.

Phase 1 simulations use NVT throughout (no NPT step) and are unaffected by this issue. Phase 2 simulations ($N = 52$) show moderate concentration-dependent volume changes ($c_{\text{eff}}/c_{\text{nom}} = 0.91$–$1.67$; Table 3.4), with all 131 simulations volume-corrected using per-simulation NPT volume ratios.

In hindsight, using NVT throughout (as in Phase 1) would have avoided the concentration confound entirely. Future studies should either (a) use NVT-only equilibration to preserve target concentrations, or (b) record the post-NPT box volume and use it directly in the Green-Kubo analysis.

6.6.3 Missing Physics: No Crosslinks and Implicit Solvent

Two related simplifications probably explain why the model does not gel. First, real MUC2 networks include disulfide bonds that create semi-permanent chemical crosslinks between C-terminal cysteine-rich domains (semi-permanent because disulfide bonds can be broken and reformed under reducing conditions). The current model relies entirely on physical (non-covalent) interactions, which produce transient inter-chain contacts rather than the permanent connectivity needed for a percolating network [11, 12]. Without covalent crosslinks, chains can always disentangle on sufficiently long timescales, which is likely why $\tan\delta \gg 1$ persists even at the longest accessible simulation times. Parsons [27] achieved a sol-gel transition for MUC5AC precisely because the model included a Monte Carlo subroutine for stochastic disulfide bond formation and breaking, which suggests that chemical crosslinks may be necessary for gelation in bead-spring mucin models. Second, the Langevin thermostat does not capture specific solvation, ion-mediated interactions, or hydrodynamic effects, any of which could be important for the concentration-dependent gelation seen experimentally. Cao and Forest [70] demonstrated in a related polymer system that transient crosslinks can dramatically tune rheological properties, producing non-monotonic behavior in the stress relaxation modulus depending on binding and unbinding affinities. Adding explicit crosslinks is certainly a useful extension for future work. The importance of disulfide crosslinks may also help explain the well-documented experimental observation that processed or purified mucins often fail to gel properly: if the purification process disrupts disulfide bonds (through mechanical shearing or reducing conditions), the resulting mucin solution would lack the crosslink connectivity needed for gelation, much like the present model. A model with reactive disulfide bonds could test this hypothesis directly by comparing gel properties with and without crosslink formation.

6.6.4 Timescale Gap

The short-timescale simulations access ∼14 μs of dynamics; experimental rheology typically operates at oscillation frequencies of 0.01–100 rad/s, corresponding to timescales of 0.06–600 s [18, 8]. The FENE-optimized timestep ($\Delta t = 0.012\tau$, providing a 2.4× speedup) extends the accessible timescale to ∼250–420 μs for long-timescale runs, but this remains orders of magnitude short of the seconds-to-hours regime probed experimentally. The Narval cluster enforces a maximum 168-hour walltime per job, limiting achievable simulation times in a single job submission to ∼250–420 μs for 100,000-bead systems. Reaching millisecond timescales for direct experimental comparison would require either checkpoint/restart protocols spanning multiple SLURM jobs or a transition to Brownian dynamics, which permits ∼60× larger timesteps but requires specialized integration schemes not available in standard LAMMPS.

Chapter 7

Conclusions

7.1 Summary of Accomplishments

The central result is that the coarse-grained MUC2 model is a viscous liquid under all conditions tested. The corrected Phase 1 screening (4,410 simulations with full LJ interactions) gives a median $\tan\delta = 3.1$ (87.9% sol-like), with $G' \propto c^{2.2}$ scaling consistent with semi-dilute entangled behavior. Phase 2 (131 simulations with $\varepsilon_{\mathrm{TT}} = 0.7$–$1.0$ at $T = 0.9$) shifts G' upward ($\approx$ 37–174 Pa vs. $\sim$5–110 Pa in Phase 1) but does not produce gelation at the 14 μs timescale. The model does not gel.

The WCA bug was probably the most consequential discovery. The initial 4,410-simulation sweep was unknowingly using repulsive-only interactions due to a LAMMPS cutoff specification issue, which is why all parameter combinations produced identical rheology. Finding and correcting this error reshaped the entire project and serves as a cautionary example for others using similar LAMMPS configurations.

The chain length study (N = 6–104, Set D parameters, three timescales, five replicates) shows that longer chains consistently produce higher G', but NPT equilibration shifts box volumes enough that effective concentrations differ substantially from nominal ones (Table 3.4). The apparent scaling exponents mix chain length and concentration effects; after volume correction the short-timescale exponent drops to $G' \propto N^{1.3}$ (nominal 10 mg/mL). The $N = 18$ model at $c_{\mathrm{eff}} = 2.2$ mg/mL gives volume-corrected $G' = 327$ Pa at the longest timescale. Pore analysis shows a timescale-dependent bifurcation: short chains at low effective concentration develop very large pores (median $\sim$2,800 nm for $N = 6$, $\sim$650 nm for $N = 18$), while longer chains maintain stable pore structure.

On the validation side: Flory scaling of test chains ($\nu = 0.475 \pm 0.001$) confirmed the implicit solvent works, and the Bhattacharya-Gubbins pore characterization gives sizes in or near experimental ranges (mean 274 nm, median 258 nm for $N = 104$ at $c = 10$ mg/mL vs. AFM: 20–200 nm). The complete workflow is scripted and tested on Digital Research Alliance of Canada clusters.

7.2 Broader Implications

The most immediately useful result is probably the $N = 18$ model. At $c_{\text{eff}} = 2.2$ mg/mL it gives volume-corrected $G' = 327$ Pa at the longest accessible timescale, which is in a physically interesting range, and the chain is short enough that higher-resolution studies (atomistic or hybrid) are feasible. The qualitative trend that longer chains produce higher G' holds across all three timescales, even though the specific exponents are confounded with NPT-induced concentration differences (Section 6.6.2). Future work should use NVT-only equilibration to avoid this problem.

The chain length scaling gives a rough prediction for how mucin degradation (for instance by bacterial proteases in IBD) would affect both stiffness and pore size. The fact that hydrophobic attraction strength is what controls the elastic response points to the terminal hydrophobic domains as the key players in network formation, which is consistent with what is known experimentally about mucin self-assembly [10, 22]. For mucin-mimetic biomaterials, this suggests that strong hydrophobic association domains are probably necessary, though whether gelation requires even stronger interactions or just more time for network maturation is still an open question. The WCA bug serves as a cautionary example for the coarse-grained simulation community. The implication being that LAMMPS cutoff hierarchies can silently override intended physics, and similar errors may affect other studies using comparable configurations. The two-phase screening/production approach and the cross-validation between independent simulation campaigns might be useful patterns for other computational polymer studies.

7.3 Future Directions

The two most important extensions are longer timescales and explicit crosslinks. The chain length study shows that G' values converge by $\sim$250–420 μs but gelation does not occur ($\tan\delta \gg 1$ at all timescales). Reaching the millisecond regime needed for direct experimental comparison would require either checkpoint/restart protocols spanning multiple SLURM jobs or a transition to Brownian dynamics ($\sim 60\times$ larger timesteps). Adding disulfide bonds between chains would provide the permanent connectivity that is probably missing: combined with the physical interactions characterized here, chemical crosslinks could produce a dual-network gel closer to native MUC2.

Beyond these, the framework could be applied to other mucins (MUC5AC and MUC5B have different domain architectures), disease modeling (truncated chains for IBD [13, 9], altered glycosylation for CF [15]), or nanoparticle diffusion simulations for drug delivery applications [44].

7.4 Final Remarks

The goal of this thesis was to build a computational framework for studying MUC2 gel properties and to determine whether shorter chain models could serve as proxies for full-length mucin. The model does not gel. Under all conditions tested, at all accessible timescales (14–420 μs), the system remains a viscous liquid. Stronger hydrophobic attraction helps (per-concentration mean $G' \approx 37$–174 Pa with optimized parameters, vs. $\sim$5–110 Pa in Phase 1), but $\tan\delta$ stays well above 1.

The project required several methodological corrections before reliable results could be obtained. The initial 4,410-simulation sweep used WCA (repulsive-only) interactions due to a LAMMPS cutoff specification error; correcting this required re-running the full parameter sweep and fundamentally changed the results. The corrected screening, then the focused 131-simulation production campaign, and finally the chain length study each built on the previous phase. NPT equilibration was also found to shift effective concentrations by factors of 3–9× from nominal values, a systematic bias that is straightforward to check for but was not identified until post-hoc analysis.

The chain length study shows the expected qualitative trend (longer chains produce stiffer systems), but the specific scaling exponents are confounded with concentration differences from NPT (Section 6.6.2). The $N = 18$ model at $c_{\text{eff}} = 2.2$ mg/mL producing $G' = 327$ Pa at the longest timescale is probably the most promising result for future work, since it is short enough for higher-resolution simulations while still showing interesting rheology.

Achieving gelation in these simulations will likely require timescales well beyond the current $\sim$420 μs limit, and probably explicit crosslinks as well. Beyond the specific results, this thesis makes several methodological contributions: a validated Green-Kubo rheology pipeline that can be applied to any LAMMPS polymer simulation, a systematic screening approach that maps the gelation boundary in a five-dimensional parameter space, and a publicly available code pipeline (Appendix D) that can be extended to other mucins or disease states. The methodological lessons learned, particularly verifying interaction potentials and checking post-NPT box volumes, are relevant to the broader coarse-grained simulation community. The methodological lessons cost real time to learn, but the pipeline that came out of it is ready for future work on MUC2 and other mucins.

Bibliography

[1] Rama Bansil and Bradley S. Turner. The biology of mucus: Composition, synthesis and organization. *Advanced Drug Delivery Reviews*, 124:3–15, 2018.

[2] Malin E. V. Johansson, Mia Phillipson, Joel Petersson, Anna Velcich, Lena Holm, and Gunnar C. Hansson. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proceedings of the National Academy of Sciences*, 105(39):15064–15069, 2008.

[3] Michael A. McGuckin, Sara K. Lindén, Philip Sutton, and Timothy H. Florin. Mucin dynamics and enteric pathogens. *Nature Reviews Microbiology*, 9(4):265–278, 2011.

[4] Michael R. Knowles and Richard C. Boucher. Mucus clearance as a primary innate defense mechanism for mammalian airways. *The Journal of Clinical Investigation*, 109(5):571–577, 2002.

[5] John V. Fahy and Burton F. Dickey. Airway mucus function and dysfunction. *New England Journal of Medicine*, 363(23):2233–2247, 2010.

[6] David J. Thornton, Karine Rousseau, and Michael A. McGuckin. Structure and function of the polymeric mucins in airways mucus. *Annual Review of Physiology*, 70:459–486, 2008.

[7] Charles Atuma, Vicki Strugala, Adrian Allen, and Lena Holm. The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. *American Journal of Physiology – Gastrointestinal and Liver Physiology*, 280(5):G922–G929, 2001.

[8] Caroline E. Wagner, Kathleen M. Wheeler, and Katharina Ribbeck. Mucins and their role in shaping the functions of mucus barriers. *Annual Review of Cell and Developmental Biology*, 34:189–215, 2018.

[9] Malin E. V. Johansson, Henrik Sjövall, and Gunnar C. Hansson. The gastrointestinal mucus system in health and disease. *Nature Reviews Gastroenterology & Hepatology*, 10(6):352–361, 2013.

[10] Rama Bansil and Bradley S. Turner. Mucin structure, aggregation, physiological functions and biomedical applications. *Current Opinion in Colloid & Interface Science*, 11:164–170, 2006.

[11] Daniel Ambort, Malin E. V. Johansson, Jenny K. Gustafsson, Anna Ermund, and Gunnar C. Hansson. Perspectives on mucus properties and formation—lessons from the biochemical world. *Cold Spring Harbor Perspectives in Medicine*, 2(11):a014159, 2012.

[12] Gabriel Javitt, María Luisa Gómez Calvo, Lis Albert, Nava Reznik, Tal Ilani, Ron Diskin, and Deborah Fass. Intestinal gel-forming mucins polymerize by disulfide-mediated dimerization of D3 domains. *Journal of Molecular Biology*, 431(19):3740–3752, 2019.

[13] Maria Van der Sluis, Barbara A. E. De Koning, Adrianus C. J. M. De Bruijn, Anna Velcich, Jules P. P. Meijerink, Johannes B. Van Goudoever, Hans A. Büller, Jan Dekker, Isabelle Van Seuningen, Ingrid B. Renes, and Alexandra W. C. Einerhand. Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology*, 131(1):117–129, 2006.

[14] Johannes Betge, Nora I. Schneider, Lars Harbaum, Marion J. Pollheimer, Richard A. Lindtner, Peter Kornprat, Matthias P. Ebert, and Cord Langner. MUC1, MUC2, MUC5AC, and MUC6 in colorectal cancer: expression profiles and clinical significance. *Virchows Archiv*, 469(3):255–265, 2016.

[15] Cameron B. Morrison, Matthew R. Markovetz, and Camille Ehre. Mucus, mucins, and cystic fibrosis. *Pediatric Pulmonology*, 54(S3):S84–S96, 2019.

[16] Bogdan Barz, Bradley S. Turner, Rama Bansil, and Brigita Urbanc. Folding of pig gastric mucin non-glycosylated domains: a discrete molecular dynamics study. *Journal of Biological Physics*, 38(4):681–703, 2012.

[17] Jonathan P. Celli, Bradley S. Turner, Nezam H. Afdhal, Sarah Keates, Ionita Ghiran, Ciaran P. Kelly, Randy H. Ewoldt, Gareth H. McKinley, Peter So, Shyamsunder Erramilli, and Rama Bansil. Helicobacter pylori moves through mucus by reducing mucin viscoelasticity. *Proceedings of the National Academy of Sciences*, 106(34):14321–14326, 2009.

[18] Samuel K. Lai, Ying-Ying Wang, Denis Wirtz, and Justin Hanes. Micro- and macrorheology of mucus. *Advanced Drug Delivery Reviews*, 61(2):86–100, 2009.

[19] David B. Hill, Paula A. Vasquez, John Mellnik, Scott A. McKinley, Aaron Vose, Frank Mu, Ashley G. Henderson, Scott H. Donaldson, Neil E. Alexis, Richard C. Boucher, and M. Gregory Forest. A biophysical basis for mucus solids concentration as a candidate biomarker for airways disease. *PLoS ONE*, 9(2):e87681, 2014.

[20] Todd M. Squires and Thomas G. Mason. Fluid mechanics of microrheology. *Annual Review of Fluid Mechanics*, 42:413–438, 2010.

[21] T. G. Mason and D. A. Weitz. Optical measurements of frequency-dependent linear viscoelastic moduli of complex fluids. *Physical Review Letters*, 74(7):1250–1253, 1995.

[22] Oliver W. Meldrum, Gleb E. Yakubov, Mauricio R. Bonilla, Omkar Deshmukh, Michael A. McGuckin, and Michael J. Gidley. Mucin gel assembly is controlled by a collective action of non-mucin proteins, disulfide bridges, Ca^{2+}-mediated links, and hydrogen bonding. *Scientific Reports*, 8:5802, 2018.

[23] Pantelis Georgiades, Emanuela di Cola, Richard K. Heenan, Paul D. A. Pudney, David J. Thornton, and Thomas A. Waigh. A combined small-angle X-ray and neutron scattering study of the structure of purified soluble gastrointestinal mucins. *Biopolymers*, 101(12):1154–1164, 2014.

[24] Kurt Kremer and Gary S. Grest. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics*, 92(8):5057–5086, 1990.

[25] W. G. Noid. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics*, 139(9):090901, 2013.

[26] Andrew G. Ford, Xue-Zheng Cao, Micah J. Papanikolas, Takafumi Kato, Richard C. Boucher, Matthew R. Markovetz, David B. Hill, Ronit Freeman, and M. Gregory Forest. Molecular dynamics simulations to explore the structure and rheological properties of normal and hyperconcentrated airway mucus. *Studies in Applied Mathematics*, 147(4):1369–1387, 2021.

[27] Jon M. Parsons. A computational model of mucus. Master's thesis, San Diego State University, San Diego, CA, 2021.

[28] Paweł Gniewek and Andrzej Kolinski. Coarse-grained Monte Carlo simulations of mucus: structure, dynamics, and thermodynamics. *Biophysical Journal*, 99(11):3507–3516, 2010.

[29] Paweł Gniewek and Andrzej Kolinski. Coarse-grained modeling of mucus barrier properties. *Biophysical Journal*, 102(2):195–200, 2012.

[30] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*, 111(27):7812–7824, 2007.

[31] Robert D. Groot and Patrick B. Warren. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *The Journal of Chemical Physics*, 107(11):4423–4435, 1997.

[32] Paul J. Flory. *Principles of Polymer Chemistry.* Cornell University Press, Ithaca, NY, 1953.

[33] Pierre-Gilles de Gennes. *Scaling Concepts in Polymer Physics.* Cornell University Press, Ithaca, NY, 1979.

[34] J. C. Le Guillou and J. Zinn-Justin. Critical exponents for the n-vector model in three dimensions from field theory. *Physical Review Letters*, 39:95–98, 1977.

[35] Prince E. Rouse, Jr. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *The Journal of Chemical Physics*, 21(7):1272–1280, 1953.

[36] Masao Doi and Sam F. Edwards. *The Theory of Polymer Dynamics.* Oxford University Press, Oxford, 1988.

[37] Pierre-Gilles de Gennes. Reptation of a polymer chain in the presence of fixed obstacles. *The Journal of Chemical Physics*, 55:572–579, 1971.

[38] John D. Ferry. *Viscoelastic Properties of Polymers.* John Wiley & Sons, New York, 3rd edition, 1980.

[39] H. Henning Winter and François Chambon. Analysis of linear viscoelasticity of a crosslinking polymer at the gel point. *Journal of Rheology*, 30(2):367–382, 1986.

[40] Melville S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. II. Irreversible processes in fluids. *The Journal of Chemical Physics*, 22(3):398–413, 1954.

[41] Ryogo Kubo. Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957.

[42] Thomas G. Mason. Estimating the viscoelastic moduli of complex fluids using the generalized Stokes–Einstein equation. *Rheologica Acta*, 39:371–378, 2000.

[43] Oliver Lieleg, Ioana Vladescu, and Katharina Ribbeck. Characterization of particle translocation through mucin hydrogels. *Biophysical Journal*, 98(9):1782–1789, 2010.

[44] Marie Boegh and Hanne Mørck Nielsen. Mucus as a barrier to drug delivery – understanding and mimicking the barrier properties. *Basic & Clinical Pharmacology & Toxicology*, 116(3):179–186, 2015.

[45] Jacob Witten, Tahoura Samad, and Katharina Ribbeck. Molecular characterization of mucus binding. *Biomacromolecules*, 20(4):1505–1513, 2019.

[46] Supriyo Bhattacharya and Keith E. Gubbins. Fast method for computing pore size distributions of model materials. *Langmuir*, 22:7726–7731, 2006.

[47] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, Oxford, 2nd edition, 2017.

[48] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, San Diego, 2nd edition, 2002.

[49] Loup Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159(1):98–103, 1967.

[50] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. 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):637–649, 1982.

[51] John D. Weeks, David Chandler, and Hans C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of Chemical Physics*, 54(12):5237–5247, 1971.

[52] R. Byron Bird, Robert C. Armstrong, and Ole Hassager. *Dynamics of Polymeric Liquids*, volume 2: Kinetic Theory. John Wiley & Sons, New York, 2nd edition, 1987.

[53] T. Schneider and E. Stoll. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B*, 17(3):1302–1322, 1978.

[54] Donald L. Ermak and J. A. McCammon. Brownian dynamics with hydrodynamic interactions. *The Journal of Chemical Physics*, 69(4):1352–1360, 1978.

[55] Gareth W. Hughes, Caroline Ridley, Richard Collins, Alan Roseman, Robert Ford, and David J. Thornton. The MUC5B mucin polymer is dominated by repeating structural motifs and its topology is regulated by calcium and pH. *Scientific Reports*, 9:17350, 2019.

[56] Andrew N. Round, Mark Berry, Terrence J. McMaster, Stefan Stoll, David Gowers, Anthony P. Corfield, and Mervyn J. Miles. Heterogeneity and persistence length in human ocular mucins. *Biophysical Journal*, 83(3):1661–1670, 2002.

[57] Michael Rubinstein and Ralph H. Colby. *Polymer Physics*. Oxford University Press, Oxford, 2003.

[58] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117(1):1–19, 1995.

[59] Aidan P. Thompson, H. Metin Aktulga, Richard Berger, Dan S. Bolintineanu, W. Michael Brown, Paul S. Crozier, Pieter J. in 't Veld, Axel Kohlmeyer, Stan G. Moore, Trung Dac Nguyen, Ray Shan, Mark J. Stevens, Julien Tranchida, Christian Trott, and Steven J. Plimpton. LAMMPS – a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, 271:108171, 2022.

[60] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[61] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.

[62] Aidan P. Thompson, Steven J. Plimpton, and William Mattson. General formulation of pressure and stress tensor for arbitrary many-body interaction potentials under periodic boundary conditions. *The Journal of Chemical Physics*, 131(15):154107, 2009.

[63] Jorge Ramírez, Sathish K. Sukumaran, Bart Vorselaars, and Alexei E. Likhtman. Efficient on the fly calculation of time correlation functions in computer simulations. *The Journal of Chemical Physics*, 133(15):154103, 2010.

[64] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272, 2020.

[65] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585:357–362, 2020.

[66] John D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[67] A. N. Round, N. M. Rigby, A. Garcia de la Torre, Adam Macierzanka, E. N. C. Mills, and Alan R. Mackie. Lamellar structures of MUC2-rich mucin: a potential role in governing the barrier and lubricating functions of intestinal mucus. *Biomacromolecules*, 13(10):3253–3261, 2012.

[68] Caroline E. Wagner, Bradley S. Turner, Michael Rubinstein, Gareth H. McKinley, and Katharina Ribbeck. A rheological study of the association and dynamics of MUC5AC gels. *Biomacromolecules*, 18(11):3654–3664, 2017.

[69] Alexei E. Likhtman, Sathish K. Sukumaran, and Jorge Ramírez. Linear viscoelasticity from molecular dynamics simulation of entangled polymers. *Macromolecules*, 40(18):6748–6757, 2007.

[70] Xue-Zheng Cao and M. Gregory Forest. Rheological tuning of entangled polymer networks by transient cross-links. *The Journal of Physical Chemistry B*, 123(5):974–982, 2019.

Appendix A

Chain Length Study (Exploratory, Set D Parameters)

Note: This section presents exploratory results from the chain length study. Due to the NPT volume equilibration issue (Section 3.4.2), chain length and effective concentration co-vary across the study, making it difficult to isolate pure chain length effects. The qualitative trends are robust, but the specific scaling exponents should be interpreted with caution.

A central question motivating this work is: how short can a coarse-grained mucin model be while still capturing the essential polymer physics? This question has practical implications for future atomistic simulations, where the computational cost of simulating full-length MUC2 (∼5000 amino acids) is prohibitive. If shorter chain segments produce similar emergent behavior, they may serve as valid minimal models for higher-resolution studies.

The chain length study used four coarse-grained models (N = 6, 18, 52, and 104 beads per dimer) at two nominal concentrations (10 and 30 mg/mL) with five replicates each (40 simulations total). As discussed in Section 3.4.2, NPT equilibration causes chain-length-dependent volume changes that shift effective concentrations markedly from these nominal values (Table 3.4). Concentrations reported in this section are nominal unless otherwise noted. The primary results presented here use **Set D parameters** ($\varepsilon_{\mathrm{TT}} = 1.0$, $\varepsilon_{\mathrm{TH}} = 0.4$, $\varepsilon_{\mathrm{HH}} = 0.4$, $T = 0.9$, $r_c = 2.0\sigma$), the Phase 2 winner, at three timescales: **short** (14 μs, $\Delta t = 0.005\tau$), **medium** (∼120–210 μs actual of 420 μs target, $\Delta t = 0.012\tau$ with FENE-optimized timestep ramp), and **long** (∼250–420 μs actual of 840 μs target, $\Delta t = 0.012\tau$). All 40 short simulations completed. Of 40 medium simulations, 7 completed fully, 32 reached 60–99% of production before the 96-hour SLURM walltime limit, and 1 produced no stress output. Of 40 long simulations, 5 completed fully and 35 reached 30–50% of production before the 168-hour SLURM walltime limit. The partial production data is fully usable for Green-Kubo rheological analysis, as the stress autocorrelation converges well before the production endpoint.

An earlier chain length study using pre-correction parameters ($\varepsilon_{\mathrm{TT}} = 0.4$, $T = 1.0$) is documented in Appendix B. The pre-correction study produced artificially elastic-dominated response ($\tan\delta \approx 0.001$) due to weaker interactions and higher temperature; the Set D results presented here, which are liquid-like ($\tan\delta \gg 1$) consistent with Phase 1 and Phase 2, represent the corrected chain length analysis.

A.1 Rheology

A.1.1 Short Timescale (14 μs)

Table A.1: Chain length test short timescale (14 μs) rheological results. Set D parameters ($\varepsilon_{\mathrm{TT}} = 1.0$, $T = 0.9$), mean ± SD across $n = 5$ replicates. c_{eff} is the effective concentration after NPT volume equilibration (Table 3.4). G' values are volume-corrected: values here are rescaled by $V_{\mathrm{actual}}/V_{\mathrm{initial}} = c_{\mathrm{nom}}/c_{\mathrm{eff}}$. $\tan\delta = G''/G'$ is unaffected by the volume correction (both moduli scale identically).

N	c_{nom} (mg/mL)	c_{eff} (mg/mL)	G' (Pa)	$\tan\delta$
6	10	1.1	$2{,}200 \pm 36$	24 ± 24
6	30	29.3	$60{,}900 \pm 2{,}100$	181 ± 284
18	10	2.2	$4{,}790 \pm 18$	456 ± 895
18	30	14.6	$31{,}400 \pm 800$	22 ± 30
52	10	17.6	$40{,}300 \pm 1{,}100$	22 ± 37
52	30	31.5	$72{,}400 \pm 560$	6.4 ± 3.0
104	10	29.5	$70{,}900 \pm 7{,}200$	16 ± 14
104	30	38.1	$90{,}300 \pm 3{,}500$	7.4 ± 4.3

All conditions produce $\tan\delta \gg 1$ (liquid-like), consistent with Phase 1 results (median $\tan\delta = 3.1$, 87.9% sol-like). An earlier chain length study using weaker pre-correction parameters ($\varepsilon_{\mathrm{TT}} = 0.4$, $T = 1.0$; Appendix B) produced artificially elastic-dominated response ($\tan\delta \approx 0.001$) due to weaker attractions and higher temperature suppressing viscous dissipation, a parameter set effect rather than a timescale effect.

A.1.2 Medium Timescale (∼120–210 μs)

Table A.2: Chain length test medium timescale rheological results. Target production: 420 μs; actual: ∼120–210 μs due to Narval 96-hour walltime (larger systems complete fewer steps). Set D parameters ($\varepsilon_{\mathrm{TT}} = 1.0$, $T = 0.9$), mean ± SD across $n = 4$–5 replicates. c_{eff} from Table 3.4. G' volume-corrected (see Table A.1 caption). $\tan\delta$ unaffected.

N	c_{nom} (mg/mL)	c_{eff} (mg/mL)	G' (Pa)	$\tan\delta$
6	10	1.1	3.6 ± 0.1	28 ± 15
6	30	29.3	$224{,}700 \pm 5{,}500$	35 ± 52
18	10	2.2	364 ± 20	35 ± 53
18	30	14.6	$630{,}800 \pm 12{,}800$	34 ± 10
52	10	17.6	$898{,}400 \pm 12{,}000$	80 ± 260
52	30	31.5	$586{,}800 \pm 15{,}600$	26 ± 2.3
104	10	29.5	$613{,}900 \pm 3{,}000$	23 ± 2.3
104	30	38.1	$297{,}800 \pm 1{,}100$	26 ± 5.8

A.1.3 Long Timescale (∼250–420 μs)

Table A.3: Chain length test long timescale rheological results. Target production: 840 μs; actual: ∼250–420 μs due to Narval 168-hour walltime. Set D parameters ($\varepsilon_{\mathrm{TT}} = 1.0$, $T = 0.9$), mean ± SD across $n = 5$ replicates. c_{eff} from Table 3.4. G' volume-corrected (see Table A.1 caption). $\tan\delta$ unaffected.

N	c_{nom} (mg/mL)	c_{eff} (mg/mL)	G' (Pa)	$\tan\delta$
6	10	1.1	4 ± 0.1	15 ± 8
6	30	29.3	$220{,}400 \pm 5{,}100$	6.2 ± 3.1
18	10	2.2	327 ± 18	18 ± 23
18	30	14.6	$540{,}000 \pm 7{,}600$	9.3 ± 6.6
52	10	17.6	$890{,}500 \pm 6{,}000$	24 ± 19
52	30	31.5	$397{,}200 \pm 9{,}700$	7.6 ± 2.4
104	10	29.5	$613{,}900 \pm 3{,}600$	46 ± 77
104	30	38.1	$478{,}200 \pm 6{,}000$	66 ± 69

A.1.4 Rheology Summary

The most robust finding across all chain lengths, concentrations, and timescales is that the system remains liquid-like: $\tan\delta \gg 1$ in every condition tested, ranging from 6 to 456. Gelation ($\tan\delta < 1$) is not achieved even at the longest accessible timescale (∼250–420 μs). The large $\tan\delta$ variance for $N = 104$ (46 ± 77 at 10 mg/mL, 66 ± 69 at 30 mg/mL) reflects

Green-Kubo convergence challenges for the longest chains, but the conclusion of liquid-like behavior is unambiguous.

The timescale dependence of G' reveals a striking chain-length-dependent bifurcation (Figure A.2). Short chains relax and lose elastic structure at longer timescales: $N = 6$ ($c_{\text{eff}} = 1.1$ mg/mL) drops from 2,200 Pa at 14 μs to 3.6 Pa at ~120–210 μs, stabilizing at 4 Pa at the long timescale, effectively collapsing to near-zero. Under NPT equilibration at $P = 0$, these minimal chains exhibit ~786% box volume expansion, producing a dilute-gas state at $c_{\text{eff}} = 1.1$ mg/mL. At this extremely low effective concentration, no network formation is expected regardless of chain length; the low G' reflects testing in the wrong phase rather than an intrinsic inability of 6-bead chains to form networks. Similarly, $N = 18$ ($c_{\text{eff}} = 2.2$ mg/mL) decreases from 4,790 Pa (14 μs) through 364 Pa (~120–210 μs) to 327 Pa (~250–420 μs), confirming convergence. Although this value exceeds the physiological mucus range (10–100 Pa), the effective concentration (2.2 mg/mL) is well below physiological values (10–50 mg/mL), preventing direct comparison.

In contrast, long chains strengthen over time: $N = 52$ increases from 40,300 to 890,500 Pa and $N = 104$ from 70,900 to 613,900 Pa at nominal 10 mg/mL (Figure A.1). For both $N = 52$ ($c_{\text{eff}} = 17.6$ mg/mL) and $N = 104$ ($c_{\text{eff}} = 29.5$ mg/mL), G' values at the medium and long timescales are similar (~0.6–0.9 MPa), suggesting the modulus has converged. Note that effective concentrations differ considerably across chain lengths (Table 3.4), complicating direct comparison.

This concentration confound is evident in the apparent chain length scaling (Figure A.3). At nominal 10 mg/mL, effective concentrations range from 1.1 mg/mL ($N = 6$) to 29.5 mg/mL ($N = 104$), spanning a 27-fold range. Volume correction reduces the short-timescale scaling exponent from $G' \propto N^{2.4}$ to $G' \propto N^{1.3}$, but because chain length and effective concentration co-vary, this exponent cannot be interpreted as a pure chain length scaling (see Section 6.6).

After volume correction, the apparent concentration inversion for $N \geq 52$ at the short timescale disappears: $G'(c_{\text{eff}} = 17.6) < G'(c_{\text{eff}} = 31.5)$ for $N = 52$ and $G'(c_{\text{eff}} = 29.5) < G'(c_{\text{eff}} = 38.1)$ for $N = 104$, consistent with normal concentration scaling. However, at medium and long timescales the inversion persists even after volume correction, suggesting genuine kinetic trapping effects at longer timescales.

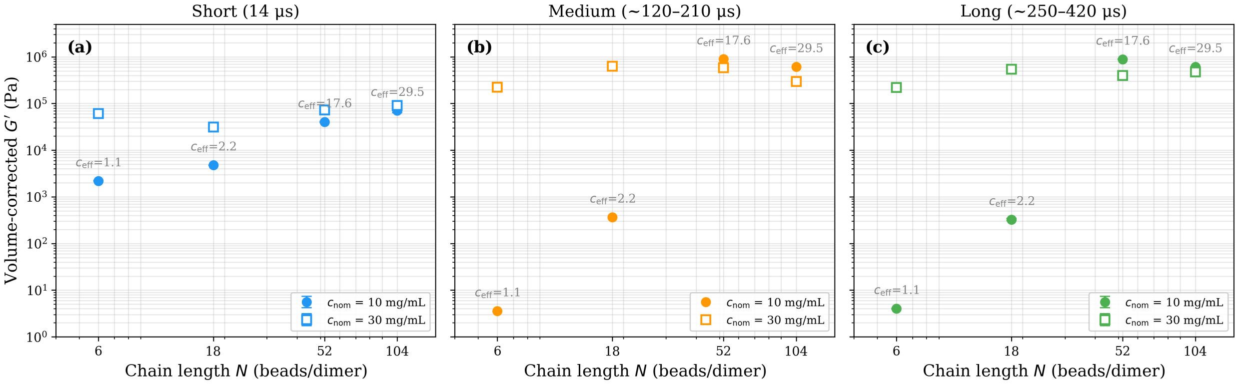


Figure A.1: Volume-corrected storage modulus G' vs chain length N at three timescales. (a) Short (14 μs), (b) medium ($\sim$120–210 μs), (c) long ($\sim$250–420 μs). Filled symbols: nominal 10 mg/mL; open symbols: nominal 30 mg/mL. Annotations show effective concentrations (c_{eff}) after NPT equilibration for the nominal 10 mg/mL series, illustrating how shorter chains expand to very dilute effective concentrations. At short timescale (a), G' increases monotonically with N at both concentrations. At medium and long timescales (b, c), the shortest chains ($N = 6$, 18) relax to low G', while longer chains ($N \geq 52$) develop stronger elastic networks. Error bars represent SD across $n = 5$ replicates.

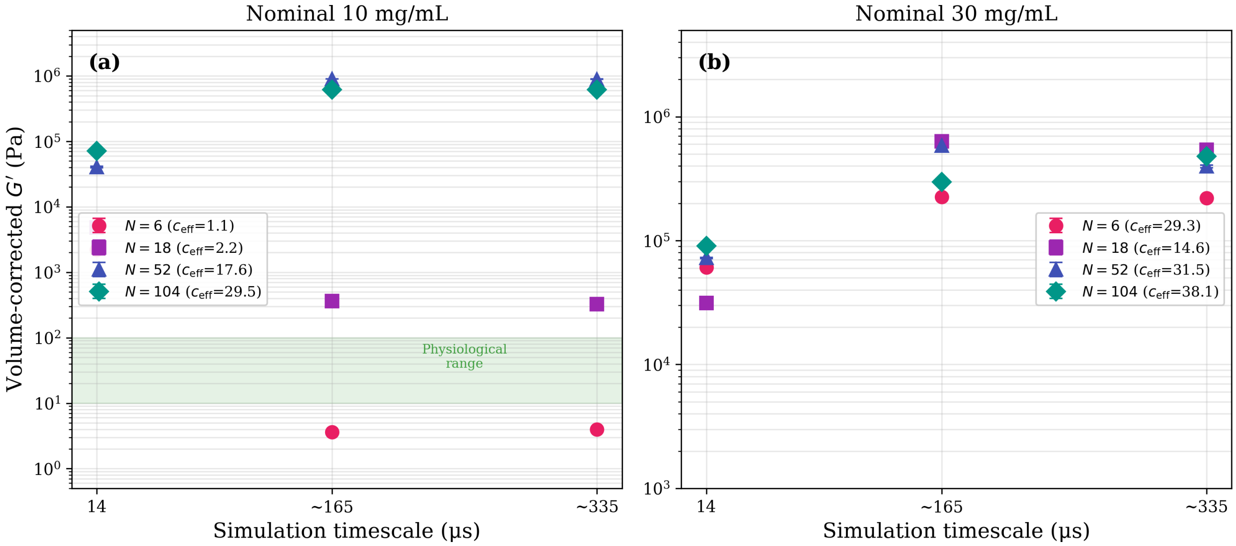


Figure A.2: Timescale evolution of volume-corrected G' for each chain length. (a) Nominal 10 mg/mL (effective concentrations in legend). Short chains relax: $N = 6$ ($c_{\text{eff}} = 1.1$ mg/mL) drops from 2,200 Pa to ∼4 Pa; $N = 18$ ($c_{\text{eff}} = 2.2$ mg/mL) drops from 4,790 Pa to 327 Pa. Long chains strengthen: $N = 52$ and 104 increase by $> 10\times$ from short to medium timescale, then stabilize. Green shaded region indicates physiological mucus range (10–100 Pa). (b) Nominal 30 mg/mL: all chain lengths show more modest timescale dependence; $N = 18$ ($c_{\text{eff}} = 14.6$ mg/mL) uniquely increases with timescale. Error bars: SD ($n = 5$); absent at medium timescale.

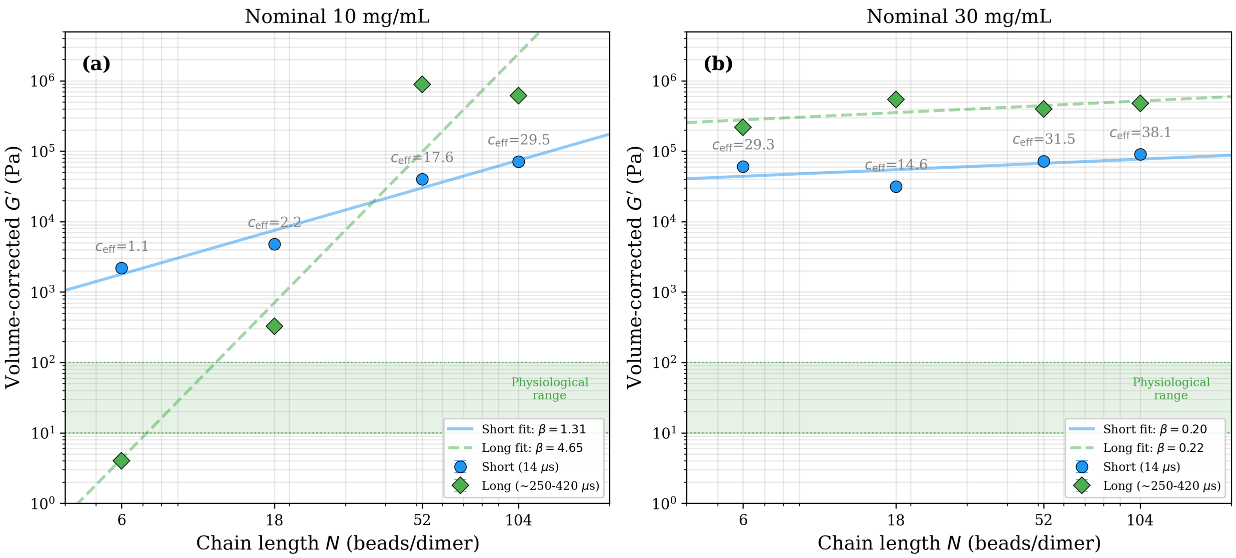


Figure A.3: Volume-corrected G' vs chain length with power-law fits, analogous to the pre-correction scaling analysis (Appendix B). (a) Nominal 10 mg/mL: short-timescale scaling exponent $\beta = 1.3$ (volume-corrected; $\beta = 2.4$ before correction). The effective concentration range (1.1–29.5 mg/mL) spans a 27-fold range, making the apparent exponent a confound of chain length and concentration effects. At longer timescales, short chains relax while long chains strengthen, producing a steeper apparent scaling. (b) Nominal 30 mg/mL: weaker scaling ($\beta \approx 0.2$) as effective concentrations are more similar across chain lengths (14.6–38.1 mg/mL). Green band: physiological mucus range (10–100 Pa). Error bars: SD ($n = 5$).

A.2 Pore Size Distribution

Bhattacharya-Gubbins pore analysis (10,000 test points × 5 frames per simulation, increased from the 5,000 used in Phase 2 to improve resolution for the wider pore distributions expected at low effective concentrations; KDTree-accelerated) was performed on all 40 Set D simulations at each of the three timescales (short, medium, and long), yielding pore size distributions for all four chain lengths at all timescales.

A.2.1 Short Timescale (14 μs)

Table A.4: Set D pore sizes by chain length at the short timescale (14 μs). Effective concentrations from NPT volume equilibration (Table 3.4). Pre-correction pore results with weaker parameters are in Appendix B.

N	c_{nom}	c_{eff}	Reps	Mean Pore (nm)	Median Pore (nm)	IQR (nm)
6	10	1.1	5	435 ± 3	273	80–779
6	30	29.3	5	43 ± 0.2	39	12–65
18	10	2.2	5	242 ± 1	125	66–411
18	30	14.6	5	67 ± 1	46	19–94
52	10	17.6	5	118 ± 2	91	27–182
52	30	31.5	5	75 ± 2	49	8–121
104	10	29.5	5	174 ± 5	143	31–287
104	30	38.1	5	65 ± 1	34	6–104

A.2.2 Medium Timescale (∼120–210 μs)

Table A.5: Set D pore sizes by chain length at the medium timescale (∼120–210 μs). Effective concentrations from NPT volume equilibration (Table 3.4).

N	c_{nom}	c_{eff}	Reps	Mean Pore (nm)	Median Pore (nm)	IQR (nm)
6	10	1.1	5	$2{,}725 \pm 21$	$2{,}808 \pm 23$	702–4,408
6	30	29.3	5	50 ± 1	31 ± 1	5–83
18	10	2.2	5	745 ± 14	619 ± 17	165–1,267
18	30	14.6	5	33 ± 1	13 ± 0	5–51
52	10	17.6	4	275 ± 1	299 ± 1	134–422
52	30	31.5	5	56 ± 1	18 ± 0	5–103
104	10	29.5	2	275 ± 0	299 ± 1	140–418
104	30	38.1	4	74 ± 0	43 ± 0	5–139

A.2.3 Long Timescale (∼250–420 μs)

Table A.6: Set D pore sizes by chain length at the long timescale (∼250–420 μs). Effective concentrations from NPT volume equilibration (Table 3.4).

N	c_{nom}	c_{eff}	Reps	Mean Pore (nm)	Median Pore (nm)	IQR (nm)
6	10	1.1	5	$2{,}732 \pm 20$	$2{,}819 \pm 23$	709–4,416
6	30	29.3	5	49 ± 0	33 ± 1	5–80
18	10	2.2	4	776 ± 10	650 ± 9	171–1,320
18	30	14.6	4	35 ± 0	13 ± 0	5–56
52	10	17.6	5	275 ± 0	298 ± 0	134–422
52	30	31.5	5	55 ± 0	20 ± 1	5–103
104	10	29.5	5	275 ± 0	299 ± 0	139–418
104	30	38.1	5	73 ± 1	42 ± 1	5–137

A.2.4 Pore Summary

Pore structure shows striking timescale dependence that varies with chain length. At the short timescale, pore size depends strongly on effective concentration: at nominal 10 mg/mL, effective concentrations range from 1.1 mg/mL ($N = 6$) to 29.5 mg/mL ($N = 104$), making direct chain length comparisons unreliable. At nominal 30 mg/mL, effective concentrations are more similar (14.6–38.1 mg/mL) and pore sizes converge to median values of 34–49 nm.

At medium and long timescales, a clear bifurcation emerges (Tables A.5–A.6). Short chains at low effective concentration develop very large pores: $N = 6$ at $c_{\text{eff}} = 1.1$ mg/mL produces median pore diameters of ∼2,800 nm at both medium and long timescales, up from 273 nm at the short timescale, reflecting the extreme NPT expansion of these minimal chains into a dilute-gas state. Similarly, $N = 18$ at $c_{\text{eff}} = 2.2$ mg/mL increases from median 125 nm (short) to 619 nm (medium) to 650 nm (long), stabilizing between medium and long timescales (∼5% change). In contrast, longer chains ($N = 104$) at higher effective concentrations (29.5, 38.1 mg/mL) show stable pore sizes across all three timescales (median ∼300 nm at 10 mg/mL, ∼20–43 nm at 30 mg/mL), indicating that their pore structure is already equilibrated at 14 μs. At nominal 30 mg/mL, all chain lengths produce small pores (median 13–49 nm), but the short-to-medium transition reveals notable reorganization: $N = 18$ contracts from 46 to 13 nm and $N = 52$ from 49 to 18 nm, while $N = 6$ (39 to 31 nm) and $N = 104$ (34 to 43 nm) change modestly. Once past the medium timescale, all pore sizes stabilize (<10% change from medium to long). For comparison, the pre-correction chain length study with weaker parameters ($\varepsilon_{\text{TT}} = 0.4$, $T = 1.0$; Appendix B) produced median pores of 70–200 nm, within the experimental range measured by AFM (20–200 nm).

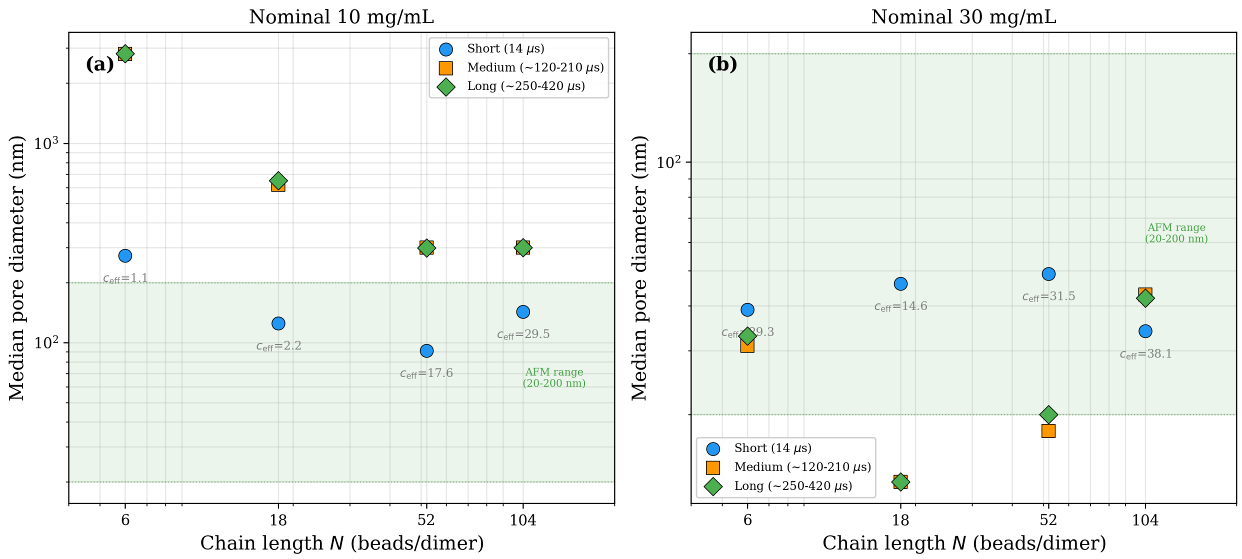


Figure A.4: Set D median pore diameter vs. chain length across all three timescales. (a) At nominal 10 mg/mL, a pronounced bifurcation emerges: short chains ($N = 6$, 18) at low c_{eff} develop very large pores at extended timescales (up to ∼2,800 nm for $N = 6$), while longer chains ($N = 52$, 104) at higher c_{eff} show stable pore structure at medium and long timescales (∼300 nm), though short-timescale values differ (91 and 143 nm respectively). (b) At nominal 30 mg/mL, all chain lengths produce small pores (13–49 nm), but $N = 18$ and $N = 52$ show significant contraction from short to medium timescales (46→13 nm and 49→18 nm), stabilizing thereafter, as effective concentrations are more similar (14.6–38.1 mg/mL). The green shaded region indicates the AFM experimental range (20–200 nm).

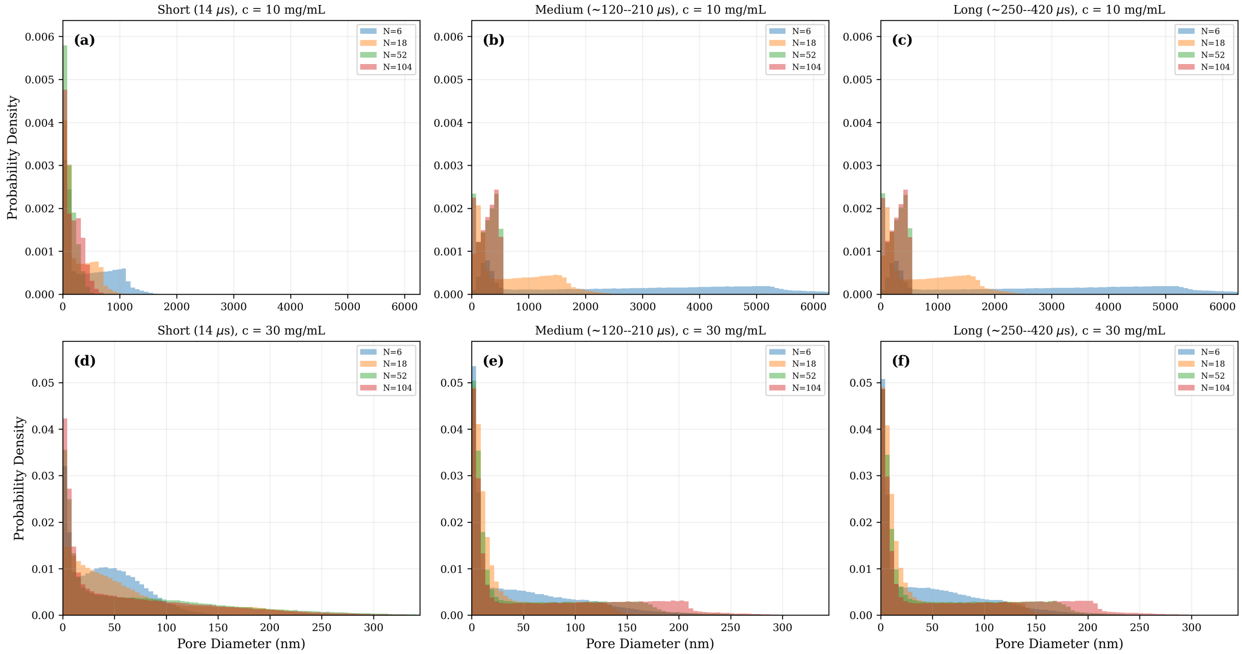


Figure A.5: Set D pore size probability density distributions across all three timescales. Columns: short (14 μs), medium (∼120–210 μs), and long (∼250–420 μs). Top row: nominal 10 mg/mL; bottom row: nominal 30 mg/mL. All panels show histograms from raw Bhattacharya–Gubbins diameter data pooled across replicates. At 10 mg/mL, the short-chain distributions (N = 6, 18) shift dramatically rightward at extended timescales, while $N = 52$ and $N = 104$ distributions remain stable. At 30 mg/mL, all distributions are narrow and concentrated below 100 nm at all timescales.

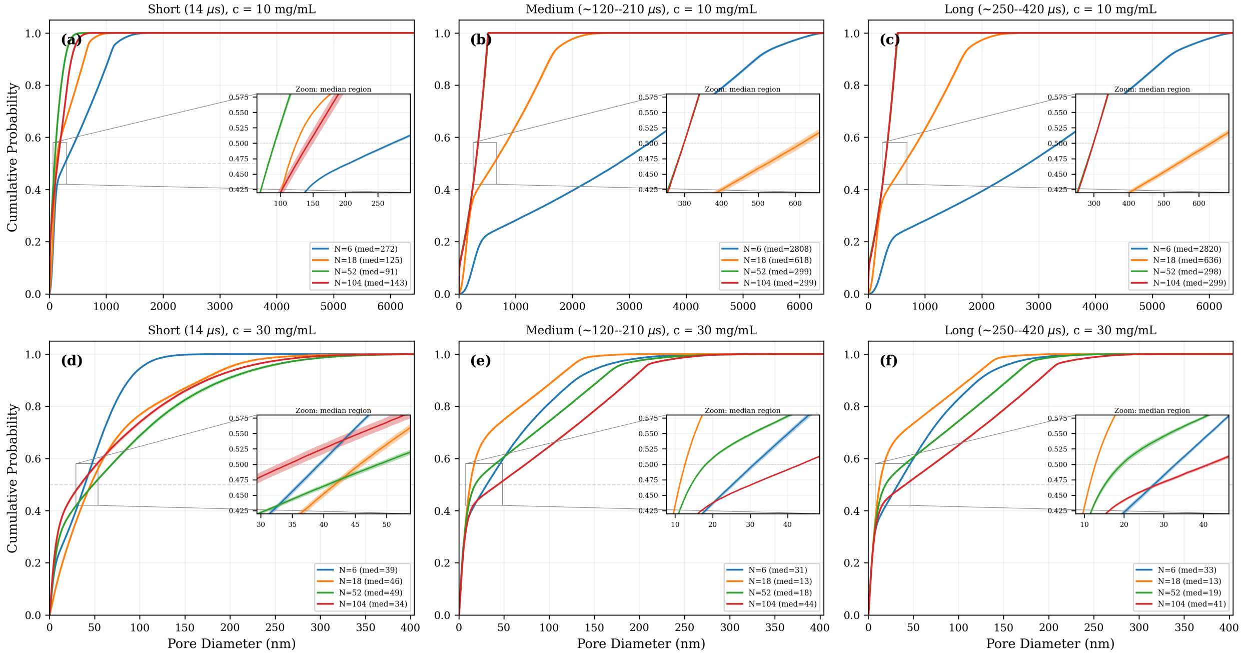


Figure A.6: Set D pore size cumulative distribution functions across all three timescales. Columns: short (14 μs), medium (~120–210 μs), and long (~250–420 μs). Top row: nominal 10 mg/mL; bottom row: nominal 30 mg/mL. All CDFs are computed from raw Bhattacharya–Gubbins diameter data. Transparent bands show ±1 SE across replicates; insets zoom into the median region to resolve the bands. At 10 mg/mL, the dramatic rightward shift of $N = 6$ and $N = 18$ CDFs from short to medium/long timescales contrasts with the stable distributions of $N = 52$ and $N = 104$, reflecting the bifurcation also visible in Figure A.4. At 30 mg/mL, all chain lengths produce similar, stable distributions across timescales.

Appendix B

Chain Length Study with Pre-Correction Parameters

This appendix documents the chain length study conducted with pre-correction parameters ($\varepsilon_{\mathrm{TT}} = 0.4$, $\varepsilon_{\mathrm{TH}} = 0.4$, $\varepsilon_{\mathrm{HH}} = 0.3$, $T = 1.0$, $r_c = 2.5\sigma$) at the short timescale (14 μs). These parameters were identified as the best-performing set during early screening, before the WCA interaction potential bug was discovered and corrected (Section 5.1.1). These results are retained for completeness but should be interpreted in context of the Set D results (Appendix A), which use the Phase 2 winner parameters and span three timescales. The weaker interactions and higher temperature produce artificially elastic-dominated response ($\tan\delta \approx 0.001$), in contrast to the liquid-like behavior ($\tan\delta \gg 1$) observed with Set D parameters at the same timescale.

B.1 Rheological Response

Simulations were conducted using all four coarse-grained models (N = 6, 18, 52, and 104 beads per dimer) at two concentrations (10 and 30 mg/mL) with five replicates each (40 simulations total).

Table B.1: Pre-correction chain length test: complete rheological results. Values are mean ± SE ($n = 5$ replicates). Pre-correction parameters ($\varepsilon_{\mathrm{TT}} = 0.4$, $T = 1.0$).

N	c (mg/mL)	G' (Pa)	G'' (Pa)	$\tan\delta$
6	10	54.4 ± 0.1	0.060 ± 0.004	0.0011 ± 0.0001
6	30	193.1 ± 0.8	0.18 ± 0.01	0.0009 ± 0.0001
18	10	628.7 ± 1.7	0.69 ± 0.02	0.0011 ± 0.0001
18	30	2400 ± 4	3.2 ± 0.3	0.0013 ± 0.0001
52	10	3920 ± 51	6.2 ± 0.3	0.0016 ± 0.0001
52	30	21577 ± 680	34.0 ± 1.7	0.0016 ± 0.0001
104	10	101064 ± 5145	184 ± 19	0.0018 ± 0.0002
104	30	62415 ± 957	120 ± 5	0.0019 ± 0.0001

The modulus scales as a power law with chain length ($G' \propto N^{\alpha}$):

- At 10 mg/mL: $\alpha = 2.49$
- At 30 mg/mL: $\alpha = 2.04$
- Average: $\alpha \approx 2.3$

This exponent is confounded with NPT-induced concentration differences (the same chain lengths at nominal 10 mg/mL have effective concentrations ranging from 1.1 to 29.5 mg/mL; Table 3.4), so it cannot be interpreted as a pure chain length dependence. The Set D analysis with volume correction yields $\alpha \approx 1.3$ at nominal 10 mg/mL (Section 6.4).

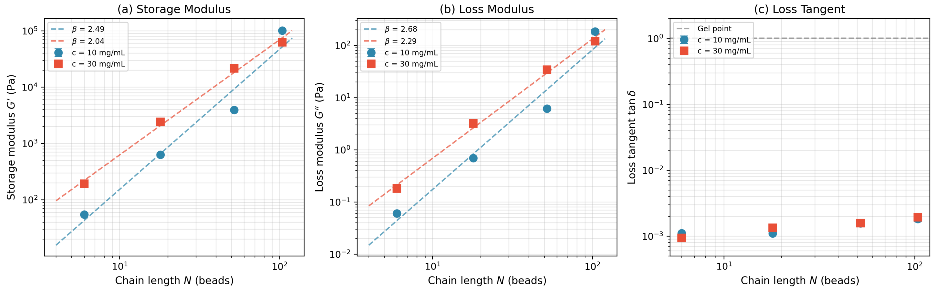


Figure B.1: Pre-correction chain length dependence of rheological properties. (a) Storage modulus $G' \propto N^{2.49}$ (10 mg/mL) and $G' \propto N^{2.04}$ (30 mg/mL). (b) Loss modulus shows similar scaling. (c) Loss tangent $\tan\delta \approx 0.001$–0.002 is nearly constant across chain lengths. The low $\tan\delta$ values reflect the weaker pre-correction parameters, not true gelation, as Set D at the same timescale produces $\tan\delta \gg 1$ (Table A.1). Error bars: SE ($n = 5$ replicates).

The $\tan\delta$ remains $\sim$0.001–0.002 across all chain lengths. These low values reflect the pre-correction parameter set effect (weaker attractions, higher temperature), not true gelation, as Set D at the same 14 μs timescale produces $\tan\delta \gg 1$ (Table A.1). Only $N = 6$ at 10 mg/mL ($G' = 54$ Pa) falls within the experimental mucus range (10–100 Pa); all other conditions exceed this range. The full-length dimer ($N = 104$) shows anomalous behavior: $G'(10 \text{ mg/mL}) > G'(30 \text{ mg/mL})$ (101 kPa vs 62 kPa). This inversion also appears in the Set D results, though at the short timescale it disappears after volume correction (Appendix A).

B.2 Pore Size Distribution

Table B.2: Pre-correction chain length test: pore size statistics. Median pore diameter ± SE ($n = 5$ replicates, ~21,000 pore measurements per condition).

N	c (mg/mL)	Median Pore (nm)	IQR (nm)
6	10	198 ± 1	160–240
6	30	126 ± 1	95–165
18	10	178 ± 2	120–250
18	30	109 ± 2	55–175
52	10	204 ± 3	110–310
52	30	117 ± 3	45–200
104	10	101 ± 3	35–165
104	30	72 ± 4	20–130

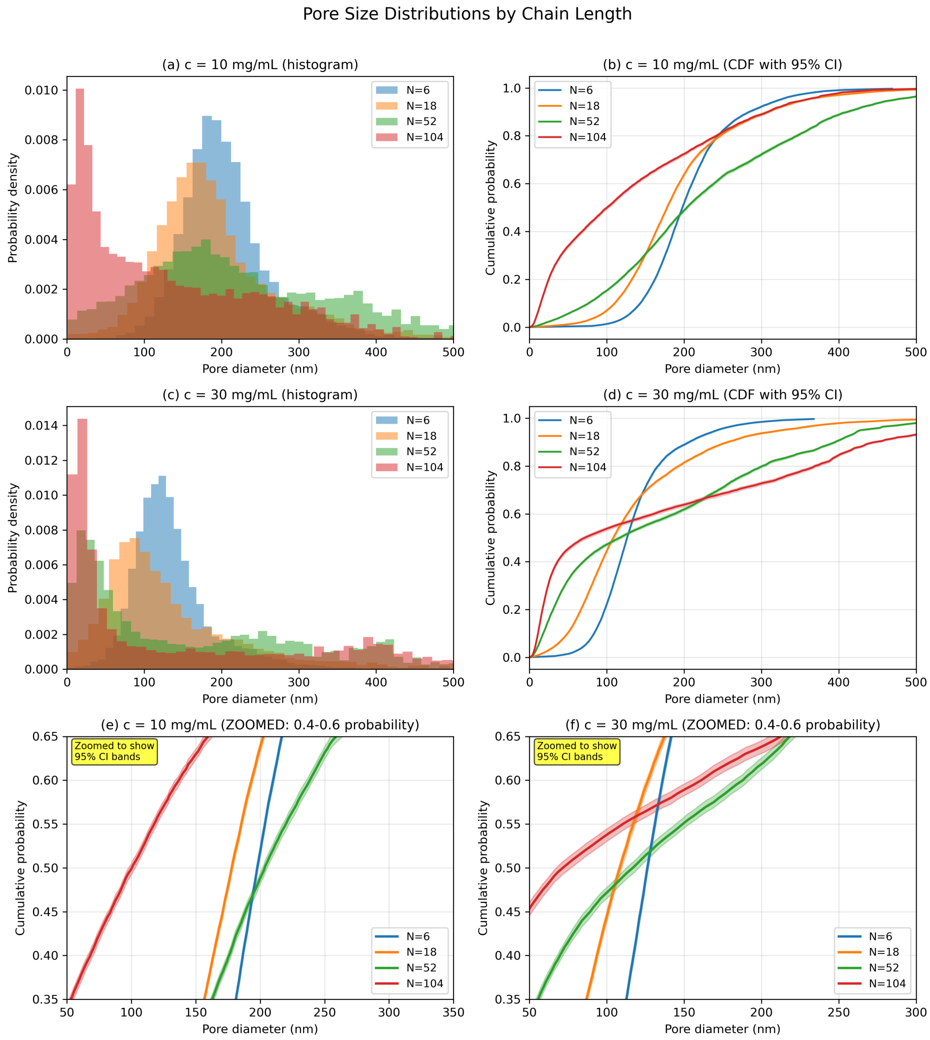


Figure B.2: Pre-correction pore size distributions by chain length. (a, c) Probability density histograms at 10 and 30 mg/mL. (b, d) Cumulative distribution functions with 95% bootstrap confidence intervals. (e, f) Zoomed CDFs showing error bands around the median. Distributions are right-skewed (skewness 0.5–1.6).

Pore size decreases ∼40% from 10 to 30 mg/mL. The longest chains fill space most efficiently, reducing both mode and median pore diameters. Median pores (70–200 nm) fall within experimental ranges (AFM: 20–200 nm, SEM: 100–400 nm) and would exclude bacteria (1–2 μm) while permitting nanoparticle diffusion.Set D with stronger parameters produces median pores of 619–650 nm at $c_{\text{eff}} = 2.2$ mg/mL for $N = 18$ at medium/long timescales (Tables A.5–A.6), reflecting stronger aggregation creating larger inter-cluster voids.

Appendix C

WCA Parameter Sweep (Historical)

This appendix documents results from the initial parameter sweep (4,410 simulations), which was affected by the WCA bug described in Section 5.1.1. These results are retained for methodological context: they illustrate what purely repulsive (WCA) polymer solutions look like and demonstrate why verifying interaction potentials is essential.

C.1 The WCA Bug

In LAMMPS, the `pair_coeff` cutoff overrides the global `pair_style` cutoff. The initial simulations specified:

```
pair_style lj/cut 2.5          # Global cutoff 2.5 sigma
pair_coeff * * 1.0 1.0 1.1225  # Per-pair cutoff overrides to 1.1225
```

The cutoff $r_c = 1.1225\sigma = 2^{1/6}\sigma$ is where the LJ potential crosses zero. At this cutoff, only the repulsive core is sampled (the WCA potential), and all attractive interactions are eliminated. The ε parameters become irrelevant because only hard-sphere-like repulsion remains.

C.2 WCA Sweep Results

With repulsive-only interactions, all 4,410 parameter combinations produced nearly identical rheology:

- $G' \propto c^{1.0}$ (dilute solution scaling, vs $c^{2.2}$ with full LJ)
- $\tan\delta \approx 0.014$ (apparently gel-like, but from entropic elasticity at simulation timescales)

- Variation across parameter sets <1% (confirming ε values are irrelevant with WCA)
- Absolute G' values ∼180× inflated compared to correctly equilibrated full LJ runs

The key diagnostic was the suspicious uniformity: with 4,410 different parameter combinations producing identical results, something was clearly wrong with the physics.

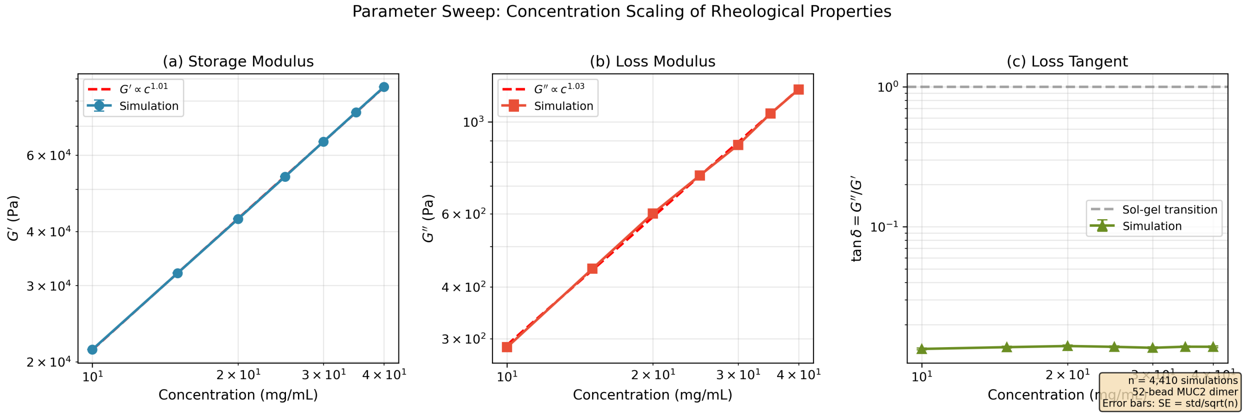


Figure C.1: Concentration scaling under WCA (repulsive-only) interactions. (a) G' and (b) G'' show linear scaling with concentration ($G' \propto c^{1.0}$), reflecting dilute solution behavior. (c) Loss tangent $\tan\delta \approx 0.014$ is concentration-independent. With corrected full LJ interactions, Phase 1 shows $G' \propto c^{2.2}$ (semi-dilute regime). Error bars: SE with $n = 630$ simulations per concentration.

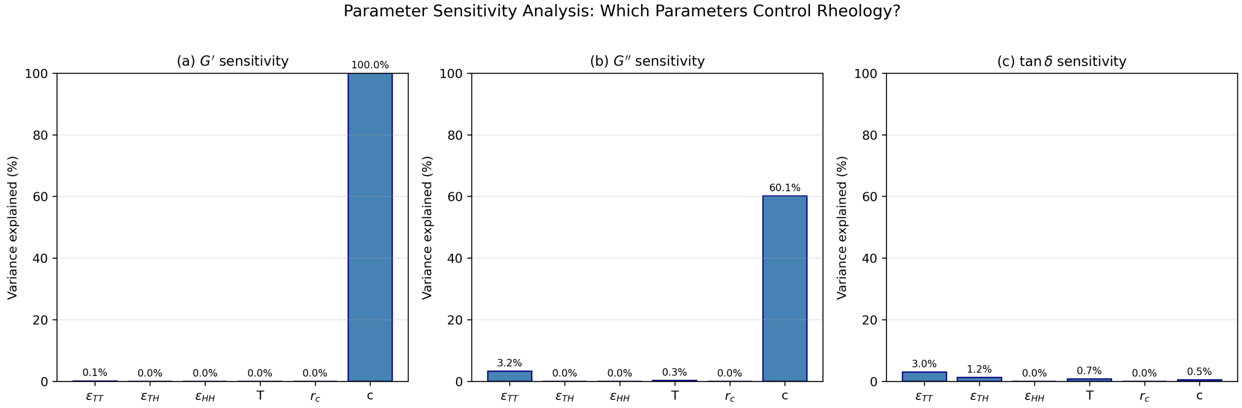


Figure C.2: WCA parameter sensitivity analysis. Variance explained by each parameter for (a) G', (b) G'', and (c) $\tan\delta$. Under WCA interactions, concentration explains nearly 100% of G' variance because ε values are irrelevant. With full LJ interactions, force field parameters produce 10–40× differences (see Phase 1, Chapter 5).

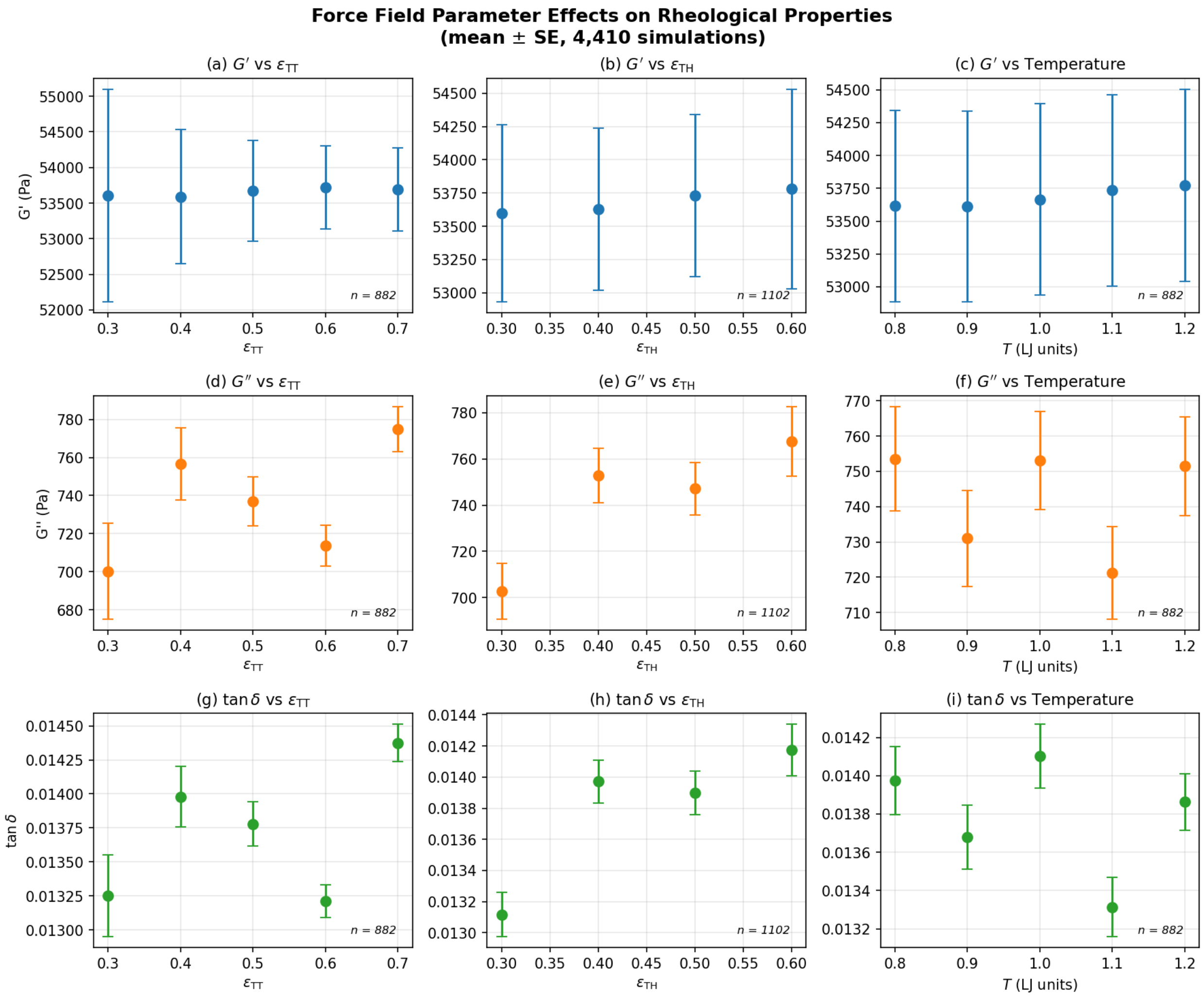


Figure C.3: WCA force field parameter effects on rheological properties. Top row: G' vs (a) $\varepsilon_{\mathrm{TT}}$, (b) $\varepsilon_{\mathrm{TH}}$, (c) temperature. Middle row: G'' vs the same parameters. Bottom row: $\tan\delta$. G' shows no dependence on any force field parameter under WCA interactions. G'' and $\tan\delta$ show weak (<10%) but statistically significant trends. Error bars: SE.

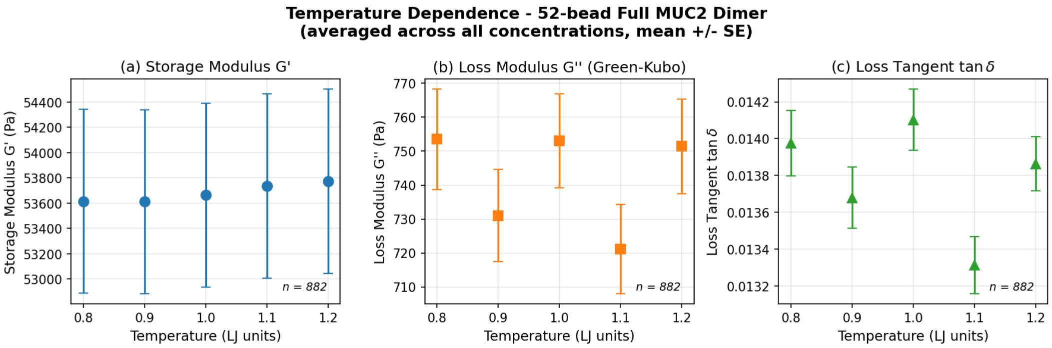


Figure C.4: WCA temperature dependence (averaged across all concentrations). (a) $G' \approx$ 53,600 Pa is constant with temperature within error. (b) $G'' \approx 740$ Pa shows no temperature dependence. (c) $\tan\delta \approx 0.014$ remains constant. Absolute values are $\sim 180\times$ inflated; $\tan\delta$ is reliable. Error bars: SE with $n = 882$ simulations per temperature.

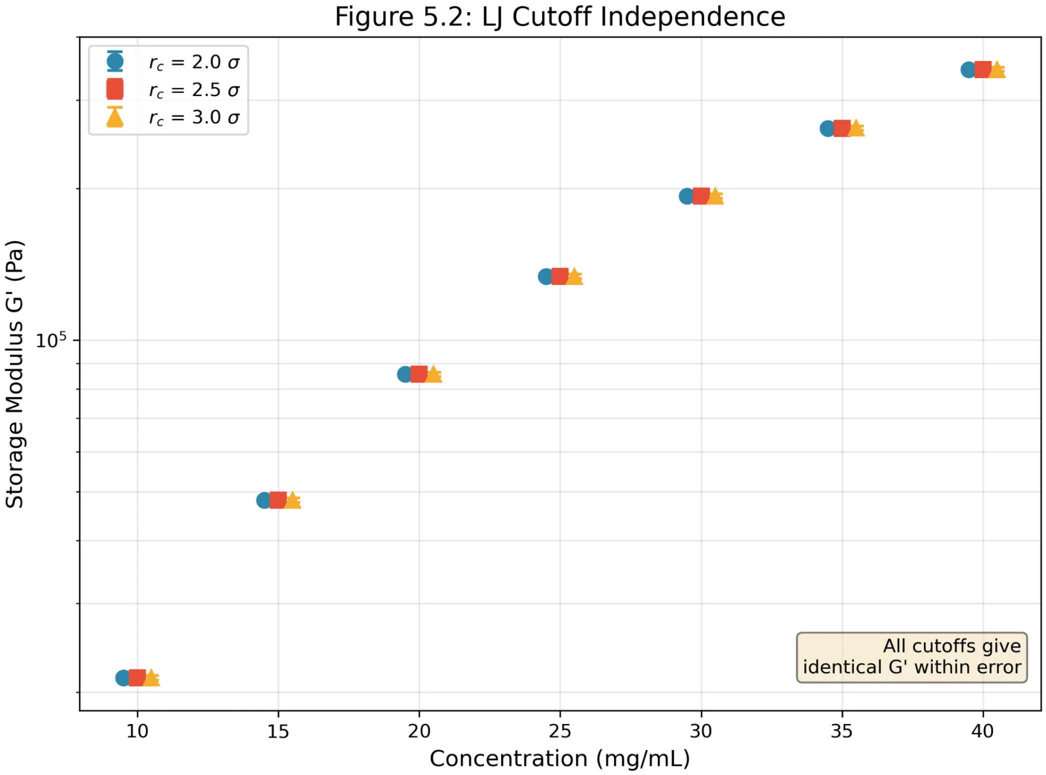


Figure C.5: WCA LJ cutoff independence. Storage modulus G' is identical within error for $r_c = 2.0$, 2.5, and 3.0σ at all concentrations, validating use of $r_c = 2.5\sigma$ as computational compromise. $n =$ 1,470 simulations per cutoff value.

Appendix D

Simulation Code

All simulation input files, analysis scripts, and figure generation code are available in a public GitHub repository:

`https://github.com/jojolavo/muc2_simulations`

The repository includes:

- LAMMPS input scripts for Phase 1 screening, Phase 2 production, and chain length study simulations
- Chain generation scripts for creating MUC2 dimer models at various chain lengths ($N = 6$–104 beads)
- Green-Kubo rheology analysis pipeline (`rheology_common.py`)
- Bhattacharya-Gubbins pore characterization scripts
- Job submission scripts for the Narval compute cluster