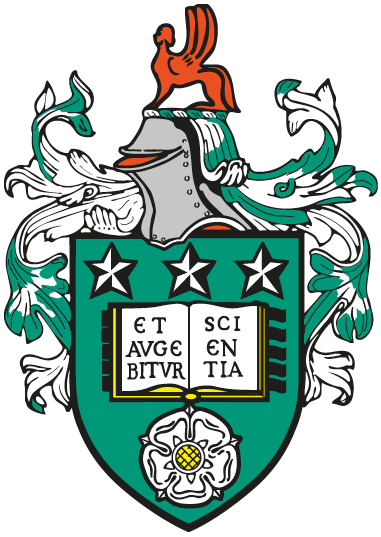
Exploring interactions of rod-like biological macromolecules through experimental and computational methods



Ryan Charles Cocking

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

Astbury Centre for Structural Molecular Biology

April 2024

Intellectual Property Statement

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Section 5.2.2 and Appendix B.3 contain work from a jointly authored review paper:

Molly S. C. Gravett*, Ryan C. Cocking, Alistair P. Curd, Oliver Harlen, Joanna Leng, Stephen P. Muench, Michelle Peckham, Daniel J. Read, Jarvellis F. Rogers, Robert C. Welch, Sarah A. Harris. "Moving in the mesoscale: Understanding the mechanics of cytoskeletal molecular motors by combining mesoscale simulations with imaging". *WIREs Comput Mol Sci.* 2022; 12:e1570. 10.1002/wcms.1570.

Molly S. C. Gravett: writing - original draft (lead). Sarah A. Harris: conceptualisation (lead); writing - original draft (equal); writing - review and editing (equal). All other authors: writing - original draft (equal); writing - review and editing (equal).

Figure 1.5 adapted with permission from *The Journal of Physical Chemistry B*. Figure 1.6A and Figure 1.6D adapted with permission from *Blood*. Figure 1.6B reproduced with permission from *Nanoscale*. Figure 1.6C adapted with permission from *Structure*. Figure 1.7 adapted with permission from Edwina F. Yeo at University College London. Figure 2.6A adapted with permission from *Cell*. Figure 2.6B adapted with permission from *Nature Structural & Molecular Biology*. Figure 2.7B donated by Kavitha Thirumurugan at the Vellore Institute of Technology. Figure 5.2 reproduced with permission from *Lab on a Chip*. Figure 5.3 adapted with permission from *Biophysical Journal*.

... at the end of the day, you know at the end of the day, anything - a benign object of any sort - you could take anything, you could take a *cheeseburger* and deconstruct it to its source!

— Devin Townsend, *Deconstruction*

Acknowledgements

I wish to thank the following people for being genuinely excellent: my supervisors - Sarah Harris, Oliver Harlen and David Brockwell - for their incredible support, professionalism and being a pleasure to work with; the other PIs in the FFEA team - Daniel Read, Jo Leng and Jon Pickering - for insightful discussions of all things mathematical and computational; Leon Willis and Alex Page in the squisher team, for being my guides in the weird world of wet lab experiments; Ben Hanson and Rob Welch for their help in getting to grips with FFEA and KOBRA, and for answering all my stupid questions about compilers; Molly Gravett, for being a wellspring of laughter and wisdom in the office; my fellow Wellcome trusters - Jon Machin, Upasana Sykora, Emma Ryder, Alex Flynn, Linda Makhlouf, and everyone in the years above us - for the warm welcome to Leeds and for getting through this together; my family - Mum, Dad, Eden, Estelle, Sasha the dog, and everyone else - for their eternal love and support, constantly telling me that I could do this, and making sure I had a roof over my head while I wrote this bloody thing; my friends from York and further afield - Aimee, James, Mike, Sean, Lowman, Matt, Rob, Ted, Ewan, Jacan, and everyone else - for being some of the best human beings on the planet; Thomas Heath of Walkeringham, for all the gossip and board games; Miso and Keeko the cats, who might as well be co-authors of the experiments chapter; and the musicians that brought me great joy by forming the audio backdrop to my time at Leeds:

Agent Fresco
Alcest
Alice in Chains
Allan Holdsworth
American Football
Andrew Hulshult
Andrew Prahlow
Aphex Twin
Artur Kordas
Autechre
Béla Bartók
Bert Jansch
Black Country, New Road
Blind Guardian
Boards of Canada
Brian Eno
Captain Beefheart
Cardiacs
Carpenter Brut
Com Truise
The Contortionist
Death
Devin Townsend
Dream Theater
Emperor
Frank Zappa
Frost*
Gentle Giant
Gojira
Gryphon
György Ligeti
Igor Stravinsky
Jaco Pastorius
Jean-Michel Jarre
John Adams
JPEGMAFIA
Käärijä
King Crimson
KGLW
Lili Boulanger
Lucas Pope
Machine Girl
Mahavishnu Orchestra
Massive Attack
Meshuggah
Michael Manring
Mike Keneally
Moondog
Mr Bungle
My Bloody Valentine
Nanoray
Nick Drake
Obscura
Opeth
Pantera
Pat Metheny
Pentangle
Philip Glass
pilotredsun
Plini
Porcupine Tree
Primus
Rush
Sea Power
SikTh
Sleepytime Gorilla Museum
Squarepusher
Steeleye Span
Steve Reich
Strapping Young Lad
Sunn O)))
Swans
System of a Down
Tears for Fears
They Might Be Giants
Tool
The Velvet Underground
VOLA
Yellow Magic Orchestra
Yoko Kanno

Abstract

The ability to solve the dynamics of biological macromolecules using computer simulation has been established as an essential tool of the biological sciences for over four decades. The most common technique is that of atomistic molecular dynamics, whereby every atom in a molecule is explicitly represented as a sphere of charge, and their trajectories are computed from interatomic potentials. Simulating the dynamics of particularly large and complex macromolecules containing millions of atoms or more remains computationally demanding, despite the critical importance of understanding such systems for addressing many research questions in molecular biology.

Coarse-grained molecular simulations address this by abstracting away atomic detail, in order to reduce the computational power required to capture larger length and time scales. Elastic rod models are a particular type of coarse-grained representation optimised to calculate the deformation of long and thin objects, which are a common occurrence throughout biology. The KirchhOff Biological Rod Algorithm (KOBRA) is one such model that was developed recently, however, it does not capture any kind of intermolecular interactions, thus limiting its application to systems where macromolecules are unlikely to come into contact with each other.

The work presented in this thesis concerns the continued development of the KOBRA simulation software, with a particular focus on the algorithms for repulsive steric interactions and attractive van der Waals forces. Fibrinogen, a large and fibrous blood protein, is used as an example macromolecule to place the results of the software development in a biological context. The computational performance of KOBRA is tested for systems in which hundreds of rods represent millions of atoms, the largest simulations to date with the software. Additional simulations of fibrin protofibrils, assembled from interacting fibrinogen rods, are also presented. Wet lab experiments are also presented that analyse the *in-vitro* mechanical response of fibrinogen to flow using a fluorescent labelling assay.

Contents

List of Figures

List of Tables

Abbreviations

AFINES	Active Filament Network Simulation
AFM	Atomic force microscopy
AMBER	Assisted Model Building with Energy Refinement
ATP	Adenosine triphosphate
AU	Absorbance units
BSA	Bovine serum albumin
CABS	C-Alpha, C-Beta, and Side chains
CFD	Computational fluid dynamics
CG	Coarse-grained
CHARMM	Chemistry at HARvard Molecular Mechanics
CPU	Central processing unit
CryoEM	Cryoelectron microscopy
Cys	Cysteine
CVD	Cardiovascular disease
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EFD	Extensional flow device
F-actin	Filamentous actin
FFEA	Fluctuating Finite Element Analysis
Fg	Fibrinogen
FpA/B	Fibrinopeptide A/B
GdnHCl	Guanidine hydrochloride
GPU	Graphics processing unit
GROMOS	GROningen MOlecular Simulation
H-bonds	Hydrogen bonds
HI	Hydrodynamic interactions
HPLC	High performance liquid chromatography
IAEDANS	5-[2-(2-iodoacetamido)ethyl]aminonaphthalene-1-sulfonic acid
IPA	2-iodo-N-phenylacetamide
KC	Kinetochore
KOBRA	KirchhOff Biological Rod Algorithm

LEBCs	Lees-Edwards boundary conditions
LJ	Lennard-Jones
MD	Molecular dynamics
MEDYAN	Mechanochemical Dynamics of Actin Networks
mRNA	Messenger ribonucleic acid
MSD	Mean squared displacement
MT	Microtubule
MW	Molecular weight
MWCO	Molecular weight cutoff
Myo5a	Myosin 5a
Myo5a-S1-6IQ	Myosin 5a subfragment-1 with all six isoleucine-glutamine domains
NAMD	Nanoscale Molecular Dynamics
Ndc80C	Nuclear division cycle 80 complex
NMR	Nuclear magnetic resonance
OPLS	Optimized Potentials for Liquid Simulations
PBCs	Periodic boundary conditions
PDB	Protein Data Bank
SASA	Solvent accessible surface area
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEM	Scanning electron microscopy
SURPASS	Single United Residue per Pre-averaged Secondary Structure fragment
TBS	Tris-buffered saline
TCEP	Tris(2-carboxyethyl)phosphine
TEM	Transmission electron microscopy
UNRES	UNited RESidue
UV	Ultraviolet
VDW	Van der Waals
WLC	Worm-like chain

Chapter 1

Introduction

1.1 Biomolecular simulations

The advent of programmable digital computers in the 1930s [1] and the dominance of Moore's law throughout the 20th century [2] propelled the advancement of molecular simulation. Some of the earliest simulations were simple statistical models of neutron diffusion during nuclear fission, conducted at Los Alamos National Laboratory in 1948 [3, 4]. The first molecular dynamics (MD) simulation of a protein, bovine pancreatic trypsin inhibitor, was published in 1977 [5] and contained around 500 atoms [6]. In the present day, the dynamics of billion-atom systems are run on supercomputers containing hundreds of thousands of CPU cores [7–9].

As the molecules of interest to researchers grow in complexity, it becomes necessary to deploy investigative techniques that can go where experiments cannot, where technological and practical barriers are reached [10]. The research challenges at the forefront of biology - the mechanisms of Alzheimer's disease [11] and cancers [12], rapid responses to future pandemics [13] and the design of next-generation biosynthetic materials [14], to name just a few - are incredibly complex problems that demand a multidisciplinary approach combining both experiment and computation [15]. Biomolecular simulations are therefore an essential component of the biological sciences toolkit.

1.1.1 Atomistic molecular dynamics

For the simulation of entire molecules, one of the most well known computational methods is atomistic MD, where every atom within a molecule is represented as a hard sphere. Interatomic interactions are represented by a sum of potential energy terms, collectively referred to as a force field, a generic form of which may be written as

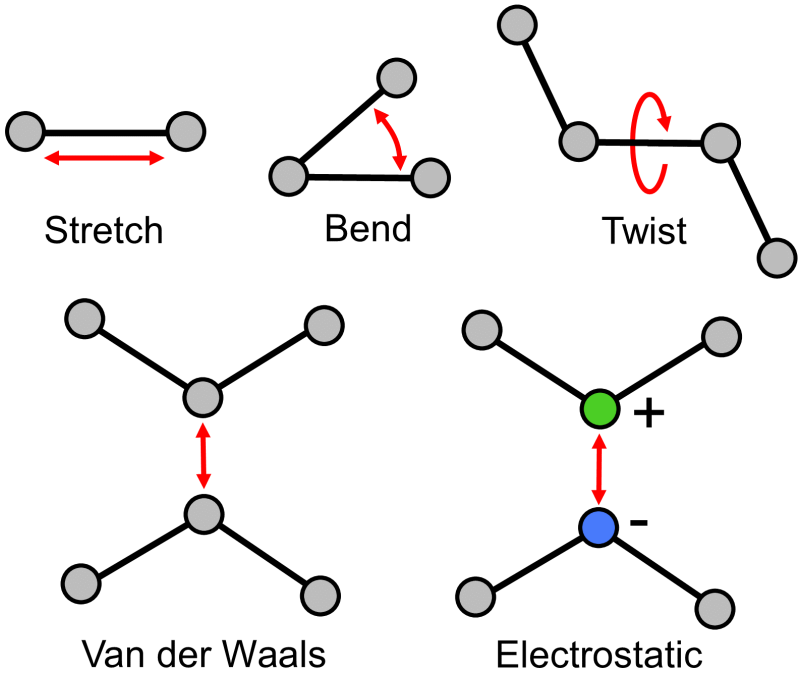


Figure 1.1: Typical contributions to a molecular dynamics force field from covalent bonds (top) and noncovalent interactions (bottom).

$$
\begin{aligned}
\mathcal{U}(\boldsymbol{r}^N) = & \sum_{\text{bonds}} k_r(r - r_0)^2 \\
& + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 \\
& + \sum_{\text{dihedrals}} V_n(1 + \cos[n\phi - \gamma]) \\
& + \sum_{i}^{N} \sum_{j>i}^{N} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) ,
\end{aligned}
\tag{1.1}
$$

where $\mathcal{U}(\boldsymbol{r}^N)$ is the total potential energy for N atoms as a function of their positions, $\boldsymbol{r}^N$ (Figure 1.1).

The stretching and bending between covalently bonded atoms relative to their equilibrium states are captured by the first and second terms of Equation 1.1, respectively, both of which are based on Hooke's law. The third term is a periodic function that captures torsional deformations or twists about covalent bonds. The fourth term models noncovalent interactions between pairs of atoms, i and j; the first part is van der Waals (VDW) interactions, given by the Lennard-Jones (LJ) potential, and the second is electrostatic interactions, given by the columb potential [16].

There are a number of well-established molecular simulation programs, but the force

fields that they employ may be packaged separately [10, 17]. Some commonly used programs include AMBER [18], GROMACS [19] NAMD [20] and OpenMM [21], which utilise force fields such as AMBER, GROMOS [22], CHARMM [23] and OPLS [24]. Force fields usually exist in several forms, each optimised for a particular type of molecule or environment. For example, AMBER has force field variants that represent proteins, nucleic acids, lipids, carbohydrates, ions or ligands [25].

The effects of the solvent on solute molecules must also be represented. In biomolecular simulations this solvent is usually liquid water at a temperature of around 300 K, or room temperature. Many solvent models exist, but they can be broadly classified into two groups: explicit, where the trajectories of water molecules and ions are directly calculated as part of the force field in Equation 1.1, and implicit, where the solvent is represented more abstractly, such as a mean force that acts on the solute [26].

The reduction of atomic detail means that implicit solvents typically require much less computing power than an explicit solvent over the course of a simulation. However, each calculation per element per time step may be more expensive in the implicit solvent, e.g. solving the electrostatic properties at a single grid point using the Poisson-Boltzmann equation, versus the potential energy of a single water molecule.

The choice of solvent model for a biomolecular simulation will of course depend on the system, although the solute is usually the main component of interest. Systems in which the solvent can significantly influence the behaviour of the solute, such as lipid bilayers, may be represented more accurately with an explicit model [16].

1.1.2 Coarse-grained simulations

An implicit solvent is an example of coarse-graining, where the fidelity of a simulation is intentionally reduced in order to free up computational resources so that larger systems can be simulated for longer times. The choice of which components to represent in a coarse-grained (CG) model will depend on the system in question, but usually the dominant sources of force are retained, as these will contribute the most to the overall dynamics of the system. In a sense, every molecular model is coarse-grained to some degree, because it is computationally intractable to exactly calculate the trajectories of every nucleus and electron in any molecule from first-principles quantum mechanics [16].

Particle-based CG models may group multiple atoms together as single spheres or pseudoatoms (Figure 1.2B, C) that respond to forces, thus reducing the degrees of freedom in the system and the size of the resulting trajectories [16]. Models such as MARTINI [27], CABS [28], UNRES [29] and SURPASS [30] may be characterised as N-bead, where N refers to the number of pseudoatoms used to represent a side chain and the portion of the backbone it bonds to. The corresponding force field may be physics-based and have a similar form to Equation 1.1, or utilise predetermined knowledge of the molecular structure

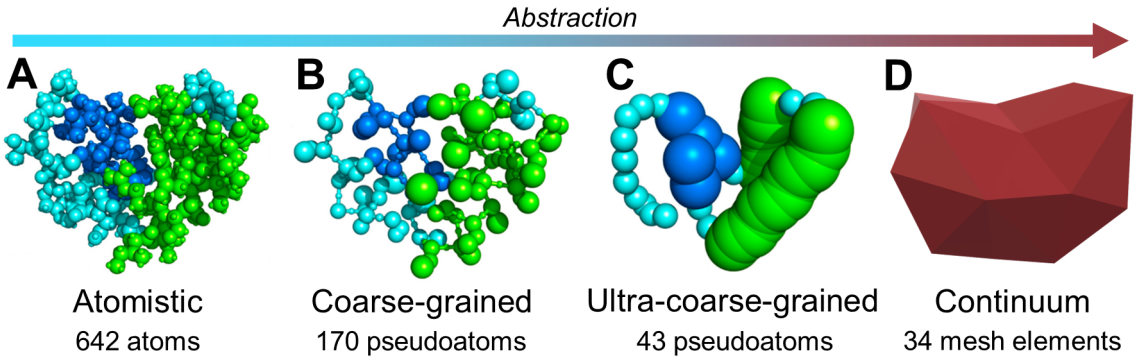


Figure 1.2: Increasingly coarse representations of crambin, a small plant protein (PDB: 1CRN [32]). **A**) Atomistic, atoms are represented as spheres. **B**) Coarse-grained, up to four pseudoatoms per side chain. **C**) Ultra-coarse-grained, one pseudoatom per side chain. **D**) Continuum, a tetrahedral mesh at 8 Å resolution. **A**, **B** and **C** adapted from [33].

and behaviour to further coarsen the system [31].

Coarser still are continuum mechanics models, which treat objects as smooth, continuous materials and use a parameterised constitutive model to represent internal forces [34]. Unlike particle-based MD, continuum mechanics approximates the simulation domain as a continuous finite element mesh (Figure 1.2D) that deforms in response to forces [35, 36]. This abstraction of detail and the ability of meshes to adopt amorphous shapes means that continuum mechanics can reach much larger length and time scales than MD [37]. At sub-nanometre length scales the continuum approximation breaks down, in the physical sense due to atoms becoming sufficiently resolved such that they cannot be reasonably coarse-grained into continua [38], and in the technical sense due to large fluctuations in energies leading to large mesh deformations over short periods of time. Large, sudden deformations usually lead to numerical instabilities in the simulation [39].

1.1.3 The biological mesoscale

X-ray crystallography is the gold standard experimental technique for determining protein structures in atomic detail [40], and cryoelectron microscopy (cryoEM) can capture both near-atomic and lower resolution structures. The popularity of cryoEM has grown rapidly over the past two decades, and as of 2022 over 21,000 density maps have been published to the Electron Microscopy Data Bank [41].

However, compared to the atomistic regime, there are fewer established simulation techniques that can capture the biologically relevant dynamics of these larger structures. This is further complicated by the fact that many macromolecules are involved in multi-scale processes that are difficult to fully capture within a single simulation framework [42]. For example, in human cytoplasmic dynein, the hydrolysis of ATP (several angstroms) produces the energy required for a series of conformational changes (tens of nanometres) that are necessary for it to traverse the surface of microtubules (hundreds of nanometres

to tens of micrometres) through many stepping events (tens of milliseconds) in order to transport its molecular cargo [43, 44].

Although atomistic MD is capable of modelling a complete dynein monomer of around one million atoms including water [45], it is currently impossible to simulate the dynamics of such a highly detailed structure at the timescales for which it performs its biological function [44]. Likewise, continuum mechanics and ultra-CG simulations can reach the required length and time scales, but run the risk of neglecting the finer details that may significantly impact larger scale dynamics. There is a length scale gap between the atomistic and continuum regimes of biomolecular simulation [36, 38].

Models have been devised that specifically target the *biological mesoscale* in order to address some of the problems around inherently multi-scale biological systems. The mesoscale is defined as encompassing systems with length scales of 10 to 500 nm, and time scales of 1 μs to 1 s, although the precise bounds have varied historically [36, 37, 46]. This definition may be written more generally as exhibiting three characteristics:

1. The smallest lengths in the system are larger than atoms; the continuum approximation is obeyed.
2. Thermal fluctuations are important to macromolecular dynamics; the Péclet number is small to moderate.
3. Dynamics are overdamped and fluid inertia is neglected; the Reynolds number is small.

The Reynolds and Péclet numbers, Re and Pe, are dimensionless numbers that qualitatively describe the transport of matter in a system [47]. In an implicit solvent the diffusion coefficient of a spherical solute particle, subject to a viscous drag force, is given by the Stokes-Einstein equation,

$$D = \frac{k_B T}{6\pi\eta R} \ , \tag{1.2}$$

where D is the diffusion coefficient (units $m^2\ s^{-1}$), k_B is Boltzmann's constant, T is the temperature (units K), η is the dynamic viscosity of the fluid (units Pa s) and R is the hydrodynamic radius of the sphere (units m). This is valid in the low Reynolds number limit, which is given by

$$\mathrm{Re} = \frac{uL\rho}{\eta} \ , \tag{1.3}$$

where u is the flow speed (units $m\ s^{-1}$), L is the length scale of interest (units m) and ρ is the fluid density (units $kg\ m^{-3}$). In molecular modelling L is usually the largest linear dimension of the molecule being studied, e.g. length or radius, but at the macroscale it

may represent the geometry of the system, e.g. pipe diameter. For $\mathrm{Re} \ll 1$ viscous forces dominate over inertial forces within the fluid [48]. The Péclet number, Pe, is the ratio of the advection rate to the diffusion rate of a system,

$$\mathrm{Pe} = \frac{uL}{D} = \frac{\dot{\gamma}L^2}{D} \ , \tag{1.4}$$

where $\dot{\gamma}$ is the shear rate of the fluid. Transport is diffusion-dominated for $\mathrm{Pe} < 1$ and advection-dominated for $\mathrm{Pe} > 1$ [49]. As Pe increases, sufficiently flexible or anisotropic macromolecules may exhibit ordered arrangements, such as aligning in the flow direction [50, 51].

Fluctuating Finite Element Analysis (FFEA) is a mesoscale simulation software developed at the University of Leeds which uses a modified version of continuum mechanics that is able to capture thermal noise [37, 52]. Globular macromolecules are represented as 3D viscoelastic meshes composed of tetrahedral elements that thermally fluctuate. The equation of motion is given by the overdamped Cauchy momentum equation,

$$\rho \frac{D\boldsymbol{u}}{Dt} = \nabla \cdot (\boldsymbol{\sigma}^v + \boldsymbol{\sigma}^e + \boldsymbol{\pi}) = 0 \ , \tag{1.5}$$

where ρ is the density, $D\boldsymbol{u}/Dt$ is the material derivative of the velocity with respect to time, $\boldsymbol{\sigma}^v$ is the viscous stress arising from both internal (macromolecule) and external (solvent) viscosities, $\boldsymbol{\sigma}^e$ is the elastic stress and $\boldsymbol{\pi}$ is the stochastic thermal stress.

The form of the stresses in Equation 1.5 is known as the Kelvin-Voigt constitutive model, which is one of the simplest representations of a viscoelastic material. The thermal stress is chosen such that its statistical properties satisfy the fluctuation-dissipation theorem, which states that any form of viscous dissipation must have an associated stochastic thermal noise [53, 54].

Attempting to represent fibrous proteins with FFEA tetrahedral meshes leads to numerical instabilities, as the smallest elements are resolved at the length scale of the fibre diameter, which may only be a few atoms wide. It is therefore desirable to have mesoscale models specifically designed for modelling rod-like objects.

1.2 Biological rod models

Fibrous macromolecules are ubiquitous in biology and branch multiple length scales. A classic example is the DNA double helix, which, in addition to its function as the basis of protein synthesis, is also popular as a building block for biosynthetic structures due to its binding specificity and well-understood structural properties [55]. The eukaryotic cytoskeleton contains microtubules (MTs) and filamentous actin (F-actin) that act as tracks

along which motor proteins can transport cellular cargo; dynein and kinesin use MTs and myosin uses F-actin [43, 56]. Von Willebrand factor is a blood glycoprotein that mediates platelet adhesion to the vessel walls during clot formation [57], and may adopt compact or elongated structures depending on the fluid environment [58]. Polymeric substances such as keratin [59], collagen [60], fibrin [61], cellulose [62] and chitin [63], are found in a diverse range of organisms, yet often produce fibrous intermediate structures during their formation, which can span from the nanoscale to the macroscale [64].

1.2.1 Theory of elastic rods

Particle-based simulations are capable of modelling fibrous proteins, but a 'rod model' is used here here to mean some coarse-grained methodology in which the anisotropy of the macromolecule constitutes an integral part, e.g. the slender body approximation, which defines rods as having an aspect ratio $L/2R \gg 1$, where L is the length and R is the cross-sectional radius. Brief summaries of some popular rod models are presented here, with a particular focus on *elastic* rod models, which are able to represent the internal material properties of the rod and how these modulate its response to external forces.

Proteins are biological polymers and there is a well-established body of literature concerning their mathematical description [54]. An ideal or freely-jointed chain is the most basic polymer model, consisting of a discrete series of points with a fixed segment length between them. The points undertake a random walk due to thermal noise, and the orientation of the segments are independent of each other; the chain is highly flexible and does not capture material properties. Despite chains being anisotropic in shape when fully extended, the unconstrained random walk tends to produce compact, globular shapes, although these can extend and align in response to flow [51, 65]. Ideal chains are not generally classified as rod models.

For fibrous proteins exhibiting a higher degree of stiffness, i.e. semi-flexible, the worm-like-chain (WLC) or Kratky-Porod model is more suitable. A protein is represented as a linear chain of segments subject to thermal fluctuations at finite temperature. The orientations of two given segments exhibit an exponential decorrelation as the distance between them increases, which may be written as

$$\left\langle \boldsymbol{t}(s) \cdot \boldsymbol{t}(0) \right\rangle = e^{-s/L_p} \ , \tag{1.6}$$

where $\boldsymbol{t}$ is the tangent vector at some distance, s, along the chain, and L_p is the persistence length parameter over which the bending stiffness decays. The chain behaves as a rigid material for $s < L_p$ or a flexible material for $s > L_p$ [54, 66].

Cosserat rod models are more complex than WLCs, possessing more elastic degrees of freedom at a cost of being more computationally expensive. Considering the formulation

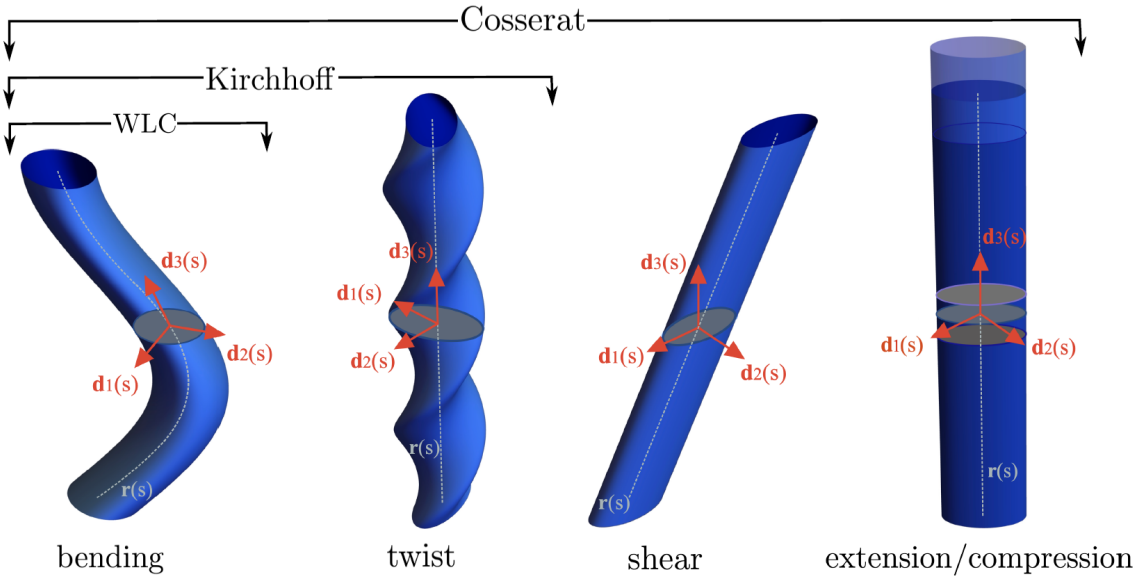


Figure 1.3: Four types of elastic deformation represented across the worm-like chain, Kirchhoff and Cosserat rod models. The material orientation frame, $\boldsymbol{d}_i$, is attached to the centreline, $\boldsymbol{r}(s)$, at a distance, s, along the rod. Adapted from [67].

by Corazza and Singh [67], a rod is defined as a continuous curve with a centreline vector, $\boldsymbol{r}(s)$, and material orientation frame, $\boldsymbol{d}_i(s)$, where $i = \{1, 2, 3\}$, that vary depending on the distance, s, along the curve. Two strain vectors, $\boldsymbol{u}(s)$ and $\boldsymbol{v}(s)$, characterise deformations in the rod and are given by

$$\begin{aligned} \frac{d\boldsymbol{d}_i}{ds} &= \boldsymbol{u} \times \boldsymbol{d}_i \\ \frac{d\boldsymbol{r}}{ds} &= \boldsymbol{v} \ , \end{aligned} \tag{1.7}$$

where $\boldsymbol{u}(s)$ is associated with bending and twisting, and $\boldsymbol{v}(s)$ with shearing and extension/compression (Figure 1.3). Kirchhoff rods are a subset of Cosserat rods that only represent bends and twists [68]. These frameworks do not prescribe that rods be discrete, but they are usually implemented as such in molecular simulations for practical reasons. Furthermore, the external forces that cause deformations are arbitrary, so thermal fluctuations are not included by default as with WLCs.

1.2.2 Applications to microscale systems and beyond

Microscale systems of the cytoskeletal reorganisation have been simulated by specialised programs that extend elastic rod models to capture more complex biophysical and biochemical effects. The 2D AFINES software can explore systems tens of micrometres in size, in times on the order of seconds. Discrete chains of F-actin are represented as po-

lar, extensible WLCs that sterically interact via an excluded volume potential [69]. By varying filament length and the concentration of cross-linking chemical species, distinct structural phases of actin networks were revealed that were bundled, polarity-sorted, and contracted [70].

At an equivalent spatiotemporal scale to AFINES, the 3D MEDYAN [71] and Cytosim [72] software packages represent cytoskeletal filaments as semi-flexible chains of interconnected segments that can sterically interact; MEDYAN captures stretching and bending deformations between segments, but Cytosim only allows bending. The dynamics of filaments and other diffusing structures such as molecular motors are solved in Cytosim and AFINES by overdamped Langevin dynamics, and in MEDYAN by a stochastic reaction–diffusion scheme coupled to mechanical energy minimization. Across all three models, the inter-filament mechanical bonds formed by motors and cross-linkers are represented as harmonic potentials. The MEDYAN software was recently extended to represent cellular membranes as 2D elastic sheets that can stretch, bend, and sterically repel filaments [73].

Virtual AFM simulations with MEDYAN revealed F-actin bundling and alignment in the direction of applied force was mediated by both mechanical and chemical cytoskeletal activity at fast and slow timescales, respectively [74]. Cytosim was used to investigate how the coarse graining of different ensembles of a motor protein, non-muscle myosin 2, affected their contractile dynamics [75].

The drag force experienced by macromolecules immersed in a viscous fluid decays with the dominant length scale in the system [76]. The anisotropy of rods therefore makes them particularly susceptible to hydrodynamic forces, more so than globular macromolecules with similar masses. Long-range hydrodynamic interactions additionally cause the dynamics of nearby rods to become coupled [77]. There are many mathematical frameworks that attempt to combine elasticity and hydrodynamics, which add a considerable degree of complexity to the modelling [78, 79].

The model of aquatic locomotion by Fauci and Peskin [80] represented a slender organism as an elastic rod that could stretch and bend, which was immersed in a 2D fluid field and driven by a sinusoidal force. Huang and Jawed [81] simulated the rotation of two tethered, helical Kirchhoff rods at low Reynolds number, and were able to reproduce the flagellar bundling that is typically observed in bacterial swimming. The macroscopic Cosserat model by Gazzola et al. [82] captured hydrodynamics, surface friction and motility, and was applied to a diverse set of systems including bacterial swimming, slithering locomotion, and the formation of twisted DNA structures under hundreds of newtons of tensile force.

Macroscopic elastic rod models are not limited to biological objects. Du et al. [83] applied a non-shearing Cosserat rod model to real-time simulation of electronic cables in the con-

text of computer-aided design for aerospace devices, where deformations are primarily driven by gravitational and frictional forces. Inextensible surgical thread was modelled as Kirchhoff rods by Wang et al. [84], devised to be sufficiently fast to allow for real-time haptic feedback in a surgical simulation program, and to handle large bending deformations required for accurate modelling of knots.

1.2.3 The KOBRA simulation software

The KirchhOff Biological Rod Algorithm (KOBRA) [85] is a mesoscale elastic rod model that was recently developed alongside the FFEA software, which by itself is incapable of modelling fibrous proteins in a numerically stable manner. Discrete rods can stretch, bend and twist, so KOBRA is technically an extended form of the Kirchhoff model (Figure 1.3). Dynamics are primarily driven by thermal noise, and interfaces can be defined between rods and tetrahedral meshes to transmit forces between large complexes. A summary of the physics is presented in Chapter 2.

Rod-rod interactions were not included in the first version of KOBRA. Consequently, rods are able to pass through each other in a nonphysical manner and attractive protein-protein interactions cannot be represented. Such features are an essential component of molecular simulation programs [16], so their development was the primary focus of this project.

Although development could proceed with generic rod structures, selecting an example protein would provide an opportunity to validate the software in a biologically relevant context. Fibrinogen, a fibrous blood protein [61], was ultimately chosen based on the following criteria:

1. Its structure has a slender aspect ratio, $L/2R \gg 1$
2. Its structure and function are relevant at the biological mesoscale: lengths of 10 to 500 nm, times of 1 μs to 1 s.
3. Its function critically depends on intermolecular interactions.
4. It is has an extensive body of literature, from which a rod model may be parameterised.

1.3 Fibrinogen

1.3.1 Physiological background

Haemostasis is the process by which platelets and other blood proteins form blood clots that staunch the flow of blood at an area of damaged tissue in vertebrates [57]. One such protein, fibrinogen (clotting factor I), undergoes thrombin-catalysed polymerisation during the final stages of the coagulation cascade into the filamentous fibrin network, which

provides critical strength and stability to the clot microstructure (Figure 1.4). Once bleeding has ceased and the clot is no longer needed, fibrin is broken down by plasmin into fragments during fibrinolysis [61, 86, 87].

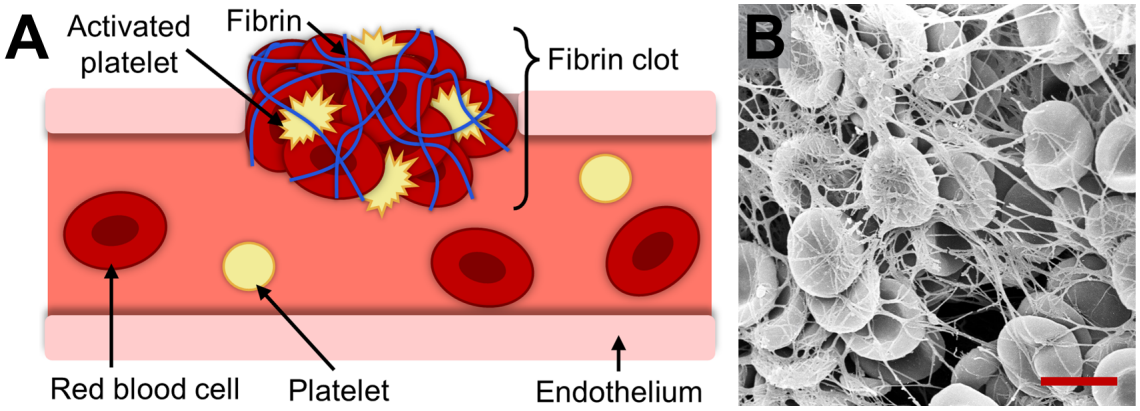


Figure 1.4: **A**) Cartoon of a fibrin-stabilised blood clot plugging a ruptured vessel. **B**) SEM image of a blood clot, with red blood cells and fibrin strands visible, scale bar 5 μm. **B** adapted from [88].

Cardiovascular diseases (CVDs) are the leading cause of death globally and are expected to account for over 22 million deaths per annum by 2030 [89], so investigations into its pathologies are of vital importance to public health. Myocardial infarctions, ischaemic strokes and thromboembolisms are CVDs in which blood clot structure is important in determining the structure and progression of the thrombus [87, 90, 91]. Fibrinogen polymerisation is an essential step of blood clot formation, so research into its function can therefore improve understanding of the molecular basis of these diseases and aid in the discovery of new therapeutics. In addition to its role in common CVDs, abnormal quantities or functionalities of fibrinogen characterise a set of rare hereditary diseases known as fibrinogenemias [92]; fibrinogen may also erroneously misfold to form amyloid fibrils, leading to kidney failure [93].

1.3.2 Molecular structure of fibrinogen

Fibrinogen is a 340 kDa hexameric glycoprotein complex with an end-to-end length of 45 nm and cross-sectional diameter of 5 nm at its widest point (Figure 1.5A, B). Its quaternary structure is a 'dimer of trimers': each half of a fibrinogen molecule consists of three protein chains, Aα (66.1 kDa), Bβ (54.4 kDa) and γ (48.5 kDa) [94], joined together by disulfide bonds [86].

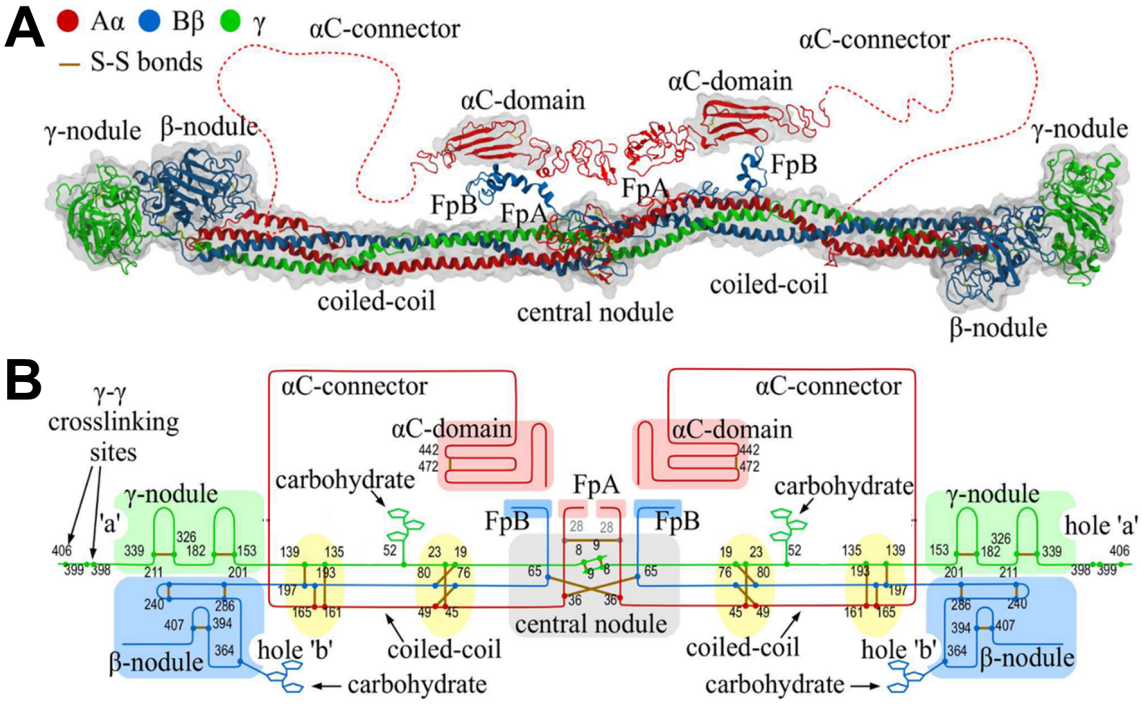


Figure 1.5: Native structure of human fibrinogen. **A)** X-ray crystal structure showing the central nodule, β-nodules, γ-nodules and coiled-coils (PDB: 3GHG [95]), combined *in-silico* with the NMR structure of recombinant bovine αC-domain fragment (PDB: 2JOR [96]), and the residues of fibrinopeptides A and B (FpA and FpB). No structure is available for the αC-connector regions, but they are shown for completeness. **B)** Schematic representation. Disulfide bonds, cross-linking sites and carbohydrates are labelled with residue numbers. Adapted from [97].

Both halves together give the molecule a characteristic shape with three globular nodules, two distal and one central, connected by triple coiled-coils with a diameter of 2 nm. The central 'E' nodule contains the N-termini of the chains, including the fibrinopeptides A and B (FpA and FpB), and the distal 'D' nodules contain the globular C-termini of Bβ and γ, with a diameter of 5 nm [95, 97]. The C-terminus of the Aα chain, referred to as the αC region, consists of an intrinsically disordered αC-connector (Aα221-391) that ends in the globular αC-domain (Aα392-610) [98], which is most commonly attached to the E nodule [99], but can extend out from the molecule up to lengths of ~ 21 nm [100]. Ongoing research is exploring the potential existence of small, folded structures within the unresolved sections of αC [101].

The high flexibility of the αC region means that it has not been successfully crystallised; the structure of human αC is unknown, but bovine αC-domain fragments have been determined by NMR [96] and fibrinogen molecules have been visualised with their αC regions by high resolution AFM on a graphite substrate [100, 102]. Around 70% of clottable fibrinogen exists *in-vivo* as a 'high molecular weight' variant with MW $= 340$ kDa and full length αC regions, and 25% as a 'low molecular weight' variant with MW $= 305$ kDa due to truncation of αC [103, 104].

1.3.3 Formation of the fibrin network

The polymerisation of fibrinogen into the fibrin network (Figure 1.6A) begins when it is converted into fibrin monomers in the presence of thrombin, which proteolytically cleaves FpA and FpB from the N-termini of Aα and Bβ chains, respectively, in the E nodule. This exposes A and B 'knobs' that go on to form non-covalent interactions between a and b 'holes' located in the C-termini of γ and Bβ chains, respectively, on other monomers [61, 86]. Monomers (Figure 1.6B) link together longitudinally via *knob-hole interactions* and covalent cross-links [90] to form half-staggered, double-stranded oligomers [105] that typically range from 2 to 25 monomers long (Figure 1.6C). Oligomers that reach a critical length of 20-25 monomers or 0.5-0.6 μm are known as protofibrils [106], which laterally aggregate to form branched, twisted fibres hundreds of nanometres in diameter (Figure 1.6D) that constitute the fibrin network (Figure 1.6E). *In-vitro* clotting experiments have shown the gel point of fibrinogen, characterised by a sudden increase in solution viscosity, to be on the order of minutes, with further branching and elongation of fibres continuing thereafter [107, 108].

A:a knob-hole interactions are essential for polymerisation, occurring earlier and exhibiting greater strength than B:b interactions [61]. Single-molecule pulling experiments by Litvinov et al. [110] found that A:a interactions have a peak rupture force of around 35 pN and exhibit catch-slip behaviour, where the application of force initially strengthens the bond up to a maximum value, after which the bond weakens.

B:b interactions are thought to play a role in lateral aggregation of protofibrils, since clots formed solely from A:a interactions were shown to have thinner clot fibres [87]; polymerisation with mutants lacking αC regions also yielded thinner fibres [102]. A simulation time of 10 ms is insufficient to sample the formation of B:b interactions via CG MD [106]. When fibrinogen is converted to fibrin, αC detaches from the E region and its length increases substantially [102], increasing the effective radius of oligomers [106] and reducing the critical length required for protofibril aggregation to occur.

1.3.4 Impact of flow on polymerisation

In a Newtonian fluid the viscous stress, τ (units Pa), experienced by immersed molecules due to flow is proportional to the flow velocity gradient,

$$\tau = \eta \frac{\partial u}{\partial y} \ , \tag{1.8}$$

where y is the axis perpendicular to the flow direction [111]. In reality, the highly crowded molecular environment of blood plasma causes it to exhibit non-Newtonian behaviour [112, 113], but this approximation is nonetheless useful for explaining some aspects of the viscous behaviour of fibrinogen [114].

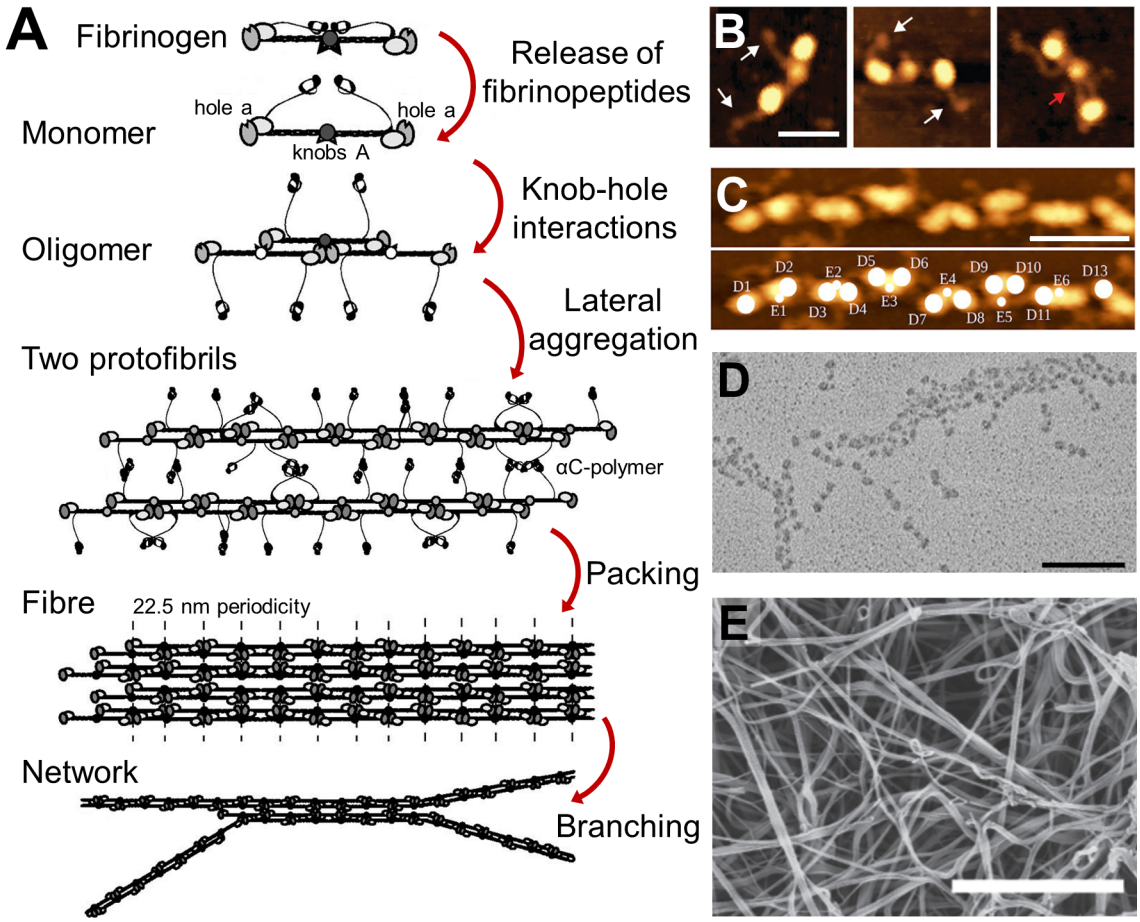


Figure 1.6: Fibrin polymerisation and products. **A**) Schematic representation [87]. **B**) High resolution AFM images of fibrinogen with visible αC regions. White arrows denote globular αC-domains. Red arrow denotes an αC-domain interacting with central E nodule. Scale bar 30 nm. Reproduced from [102]. **C**) High resolution AFM image of double-stranded fibrin oligomer. Geometric centres of D and E nodules are denoted with white dots. Scale bar 50 nm. Adapted from [106]. **D**) TEM image of fibrin fibres. Scale bar 100 nm. Adapted from [107]. **E**) SEM image of a fibrin gel formed in the absence of other blood plasma components. Scale bar 5 μm. Adapted from [109].

At the length scales of a fibrinogen molecule a blood vessel may be approximated as a straight pipe (Figure 1.7) [115]. At a sufficiently low Reynolds number, Re < 1000, the flow within the pipe is laminar, the velocity profile is parabolic with a maximum at the pipe centre, and the molecules experience shear flow [116].

The flow obeys a no-slip condition at the walls, where the velocity is zero but the shear stress is at a maximum; the wall shear stress is often used to characterise such systems. Fibrinogen must function whilst being continuously exposed to pulsatile shear flow with typical shear rates of 10 to 2000 s^{-1} in veins and arteries, respectively [117–119].

Under pathological vessel conditions, such as a partial blockage, the rapid reduction in vessel diameter (stenosis) results in localised regions of flow that act purely in the longitudinal axis and apply a stretching force across molecules, known as extensional or elongational flow [58, 120]. Severely stenosed arteries have been reported to exhibit wall shear stresses as high as 40,000 s^{-1} [118, 121].

Experiments have shown that shear flow significantly affects the microstructure and material properties of a blood clot, although the molecular mechanisms that cause this are not fully understood, particularly with respect to specific structural motifs such as αC regions or knob-hole interactions. This has clear consequences for patient health: clots made from thinner fibres are associated with an increased risk of thrombosis due to an enhanced resistance to fibrinolysis, whereas thicker fibres are linked to an increased bleeding risk due to a reduced resistance [91].

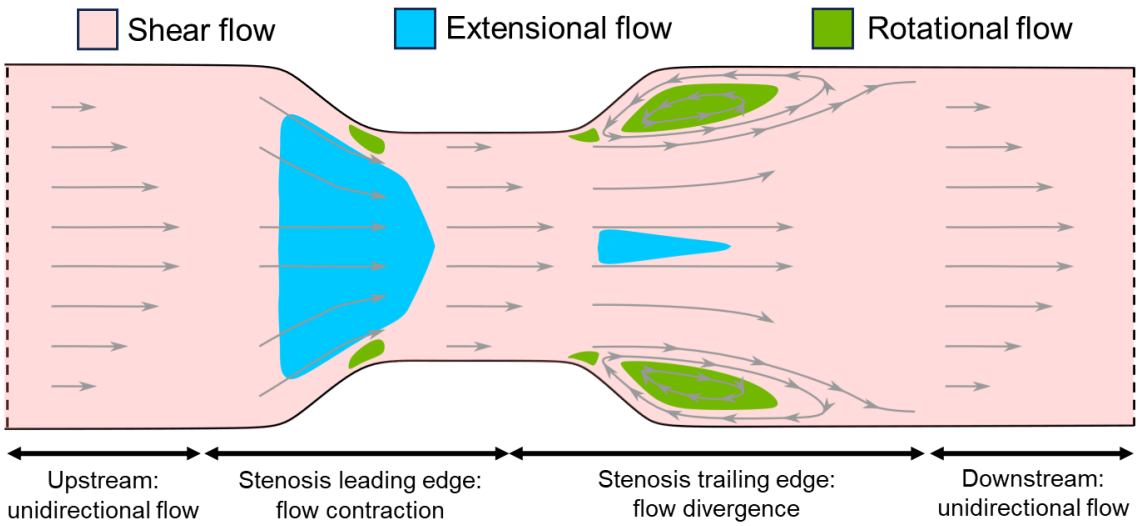


Figure 1.7: Schematic of the flow structures within a 2D slice of a stenosed blood vessel. Flow velocity is represented with grey arrows. Shear flow (pink) dominates upstream and downstream of the stenosis, and in most of the stenosis interior. Extensional flow (blue) occurs at the inlet of the stenosis, and to a lesser extent at the outlet. Rotational flow (green) mainly occurs in recirculation zones adjacent to the outlet. Adapted from [115].

Clots formed in stationary fluids have fibres with isotropic orientations, whereas clots

Fibrin structure	L (nm)	D (μm^2 s^{-1})	Pe		
			$\dot{\gamma}_{ve} = 100$ s^{-1}	$\dot{\gamma}_{ar} = 2000$ s^{-1}	$\dot{\gamma}_{st} = 10{,}000$ s^{-1}
Monomer	45	20	0.01	0.2	1
Protofibril	600	4	9	180	900

Table 1.1: Péclet numbers computed for a fibrin monomer and protofibril in venous, arterial [118] and stenotic [124] conditions. Lengths and diffusivities obtained from [106].

formed at low shear rates (5-100 s^{-1}) contain many fibres that are aligned with the flow direction [122], corresponding to a more heterogeneous microstructure [117, 123]. Although flow itself can increase the overall thickness of fibres and subsequently increase the clot stiffness [117], increased levels of thrombin have been shown to produce denser clots with thinner fibres that have more branch points [87, 91, 109]. The highly dynamic nature of intermediate fibrin structures makes it difficult to conduct experiments that assess the impact of flow on specific protein domains during polymerisation.

The physical dimensions of fibrinogen place it suitably within the biological mesoscale and hence for simulation in KOBRA. The thermal noise must, however, be compared with the physiological and pathological flows that fibrinogen experiences. If flow forces are significantly larger than diffusive forces, a coarser simulation framework that places a greater emphasis on hydrodynamics might be more suitable.

For Péclet numbers greater than 1, fibres tend to align with the flow direction and macromolecular dynamics begin to be dominated by flow over diffusion. Table 1.1 shows six values of Pe calculated from Equation 1.4 for a fibrin monomer and protofibril for a range of shear rates.

Flow tends to dominate for protofibrils with a range of $9 < \text{Pe} < 900$ due to their greater length, but with a lower limit of $\text{Pe} = 0.01$ for monomers, the polymerisation of monomers into protofibrils falls within the biological mesoscale as defined in §1.1.3. Theoretically, thermal noise and flow are both important at different stages of polymerisation, so a molecular simulation would ideally account for both effects.

1.4 Thesis outline

Chapter 2 is a more detailed summary of the KOBRA software as it existed in 2020, prior to this project. It describes the rod discretisation, internal elastic energy calculations, force transfer from rods to FFEA tetrahedra, parameterisation from atomistic simulations, and examples of studies using the software.

The software development of KOBRA during this project is discussed in Chapter 3, primarily focussing on repsulsive steric and attractive VDW rod-rod interactions, in addition to periodic boundary conditions. Validation of the updated software is presented in the form of unit and integration tests, and benchmark simulations measuring parallel scal-

ability and attainable simulation time. Additional simulations are presented in a more biologically relevant context, in which interacting rods are assembled into a larger structure of a fibrin protofibril, and the interaction strength and bending stiffness are varied in a parameter sweep.

An experimental project is presented in Chapter 4, wherein the structural response of fibrinogen to both physiological and pathological flows is quantified *in-vitro*, using a specific fluoroescent labelling assay. A pre-existing experimental workflow, initially designed for small, globular proteins in industrial contexts, was adapted to gauge its suitability for studying larger, fibrous proteins such as fibrinogen.

Chapter 5 provides a summary of the project, along with a discussion of the most desirable features for the next phase of KOBRA software development, and how these should integrate with the FFEA tetrahedral software.

Chapter 2

Overview of the KOBRA elastic rod model

This chapter provides an overview of the KirchhOff Biological Rod Algorithm (KOBRA) developed by Welch et al. [85]. The author of this thesis did not contribute to the initial version of KOBRA or its associated paper.

The KOBRA model was derived using the framework of the discrete elastic rod model of Bergou et al. [68], but extended to be applicable to biological macromolecules through the addition of thermal noise, viscous drag, extensional energy, hinge regions, and equilibrium twist. It was developed within the same codebase as the Fluctuating Finite Element Analysis (FFEA) simulation software, and both are run from the same program.

2.1 Definition and dynamics

The KOBRA elastic rod model represents long, thin proteins as chains of extensible rod elements connected end-to-end (Figure 2.1). Each rod element[1], $\boldsymbol{p}_i$, is defined as the line connecting two node positions, $\boldsymbol{r}_i$, such that $\boldsymbol{r}_{i+1} = \boldsymbol{r}_i + \boldsymbol{p}_i$, where for a rod of N nodes there are $N-1$ elements, and the node index $i = 0, 1, 2, ...N-1$, together with the angle, θ_i, that represents the rotation about the axis $\boldsymbol{p}_i$. The material axes, $\boldsymbol{m}_i$ and $\boldsymbol{n}_i$, are both perpendicular to the unit vector along the axis, $\boldsymbol{l}_i = \boldsymbol{p}_i/|\boldsymbol{p}_i|$, such that $\boldsymbol{l}_i \cdot \boldsymbol{m}_i = \boldsymbol{m}_i \cdot \boldsymbol{n}_i = \boldsymbol{n}_i \cdot \boldsymbol{l}_i = 0$, and $||\boldsymbol{l}_i|| = ||\boldsymbol{m}_i|| = ||\boldsymbol{n}_i|| = 1$.

Each rod has two distinct structures: the current state defined by $\boldsymbol{r}_i$ and $\boldsymbol{m}_i$, and the equilibrium state defined by $\widetilde{\boldsymbol{r}}_i$ and $\widetilde{\boldsymbol{m}}_i$ [2]. The current structure evolves with time in response to deformations, whereas the equilibrium structure remains constant. Elastic energies

[1] For the remainder of this thesis, unless otherwise stated, a 'rod' refers to the complete chain and an 'element' refers to a discrete segment of that chain.

[2] $\boldsymbol{r}_i$ and $\widetilde{\boldsymbol{r}}_i$ represent the same physical quantity, in this case the position of a rod node, but for the current and equilibrium structures of the same rod, respectively.

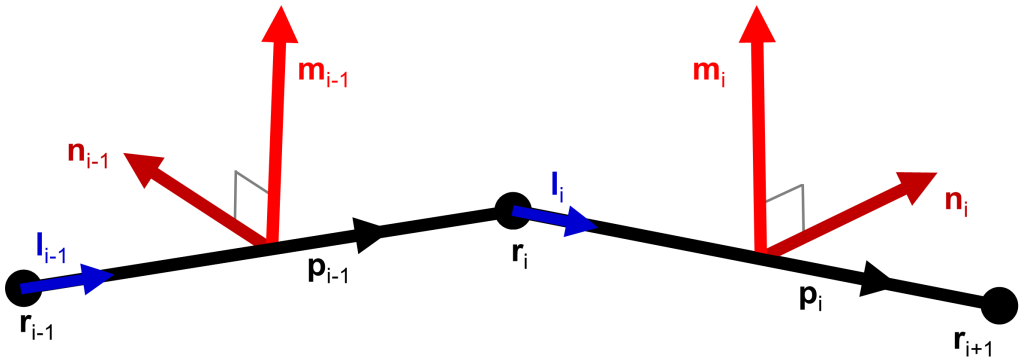


Figure 2.1: Discretised curve defining two adjacent elements within a KOBRA rod. The i-th element, $\boldsymbol{p}_i$ (black arrows), is defined as the line connecting two nodes, $\boldsymbol{r}_i$ and $\boldsymbol{r}_{i+1}$ (black circles). The unit vector, $\boldsymbol{l}_i = \boldsymbol{p}_i/|\boldsymbol{p}_i|$ (blue arrows), points along the rod axis. The material axes, $\boldsymbol{m}_i$ and $\boldsymbol{n}_i$ (red arrows), are both perpendicular to $\boldsymbol{l}_i$. Adapted from [85].

arise within the rod as the current structure is deformed away from equilibrium by thermal noise, which produces an elastic force and torque on the rod elements. The net force on the node at position $\boldsymbol{r}_i$ is given by

$$\boldsymbol{F}_i = \frac{\partial E}{\partial \boldsymbol{r}_i} \approx \frac{E(\boldsymbol{r}_i + \Delta\boldsymbol{r}_i) - E(\boldsymbol{r}_i - \Delta\boldsymbol{r}_i)}{2\Delta\boldsymbol{r}_i} \ , \tag{2.1}$$

where $\boldsymbol{F}_i$ is the force vector acting on node i, $\Delta\boldsymbol{r}_i$ is a small virtual displacement from $\boldsymbol{r}_i$, and E is the total elastic energy, obtained by summing the component energies associated with extension, torsion, and bending (Figure 2.2). The deformations in the rod due to the displacements $\Delta\boldsymbol{r}_i$ are assumed to be sufficiently small such that the elastic energies are within the linear regime. Similarly, the torque on the node is given by

$$\boldsymbol{\tau}_i = \frac{\partial E}{\partial \theta_i}\boldsymbol{l}_i \approx \frac{E(\theta_i + \Delta\theta_i) - E(\theta_i - \Delta\theta_i)}{2\theta_i}\boldsymbol{l}_i \ , \tag{2.2}$$

where $\Delta\theta_i$ is a small virtual rotation about the unit rod axis, $\boldsymbol{l}_i$.

The translational and rotational dynamics of the rods are given by stochastic equations for the positions of the nodes, $\boldsymbol{r}_i$, and the rotation angle, θ_i, given by

$$d\boldsymbol{r}_i = \frac{dt}{\zeta}(\boldsymbol{F}_i + \boldsymbol{f}_i) \ , \tag{2.3}$$

where $\boldsymbol{F}_i$ is the internal elastic force, $\boldsymbol{f}_i$ the random thermal force and ζ is the translational viscous drag coefficient, together with

$$d\theta_i = \frac{dt}{\zeta_\theta}(\tau_i + g_i) \ , \tag{2.4}$$

where τ_i and g_i are respectively the elastic and random thermal torques acting along the element axis, and ζ_θ is the rotational drag coefficient. The expression for translational drag assumes that elements are spheres, $\zeta = 6\pi\eta R$, but for rotational drag they are treated as cylinders, $\zeta_\theta = 8\pi\eta R^2|\boldsymbol{p}_i|$, where η is the dynamic viscosity of the fluid medium and R is the hydrodynamic radius [85]. The thermal noise terms, f_i and g_i, are derived from the Fluctuation-Dissipation theorem [53].

2.2 Elastic energy of deformation

2.2.1 Stretch

The extensional energy associated with a change in length along the rod element axis (Figure 2.2A) is given by Hooke's law for springs as

$$E_{stretch,i} = \frac{1}{2}k_i(|\boldsymbol{p}_i| - |\widetilde{\boldsymbol{p}}_i|)^2 \ , \tag{2.5}$$

where $|\boldsymbol{p}_i|$ is the length of the i-th element, $|\widetilde{\boldsymbol{p}}_i|$ is its equilibrium length, and $\boldsymbol{k}_i$ is the spring constant, given by

$$k_i = \frac{\kappa_{s,i}}{|\widetilde{\boldsymbol{p}}_{\boldsymbol{i}}|} \ , \tag{2.6}$$

where $\kappa_{s,i}$ is the stretching constant (units N), which is equal to the product of the Young's modulus and the cross-sectional area.

2.2.2 Twist

The torsional energy due to a rotation is defined between two adjacent elements about a single node (Figure 2.2C) and is given by

$$E_{twist,i} = \frac{\beta_i}{L_i}\left(\mathtt{mod}\left(\Delta\theta_i - \widetilde{\Delta\theta}_i + \pi, 2\pi\right) - \pi\right)^2 \ , \tag{2.7}$$

where β_i is the torsional constant (units N m^2), and $\Delta\theta_i$ and $\widetilde{\Delta\theta}_i$ are the angles between a pair of material axes in the current and equilibrium configurations, respectively. The length of the integration domain, L_i , is a prefactor used to convert an integrated quantity to a discrete one [68], and is written as

$$L_i = \frac{|\boldsymbol{p}_i| + |\boldsymbol{p}_{i-1}|}{2} \ . \tag{2.8}$$

The periodic function defined with the modulo operator in Equation 2.7 ensures the torsional energy remains continuous, provided that $-\pi < \Delta\theta < \pi$.

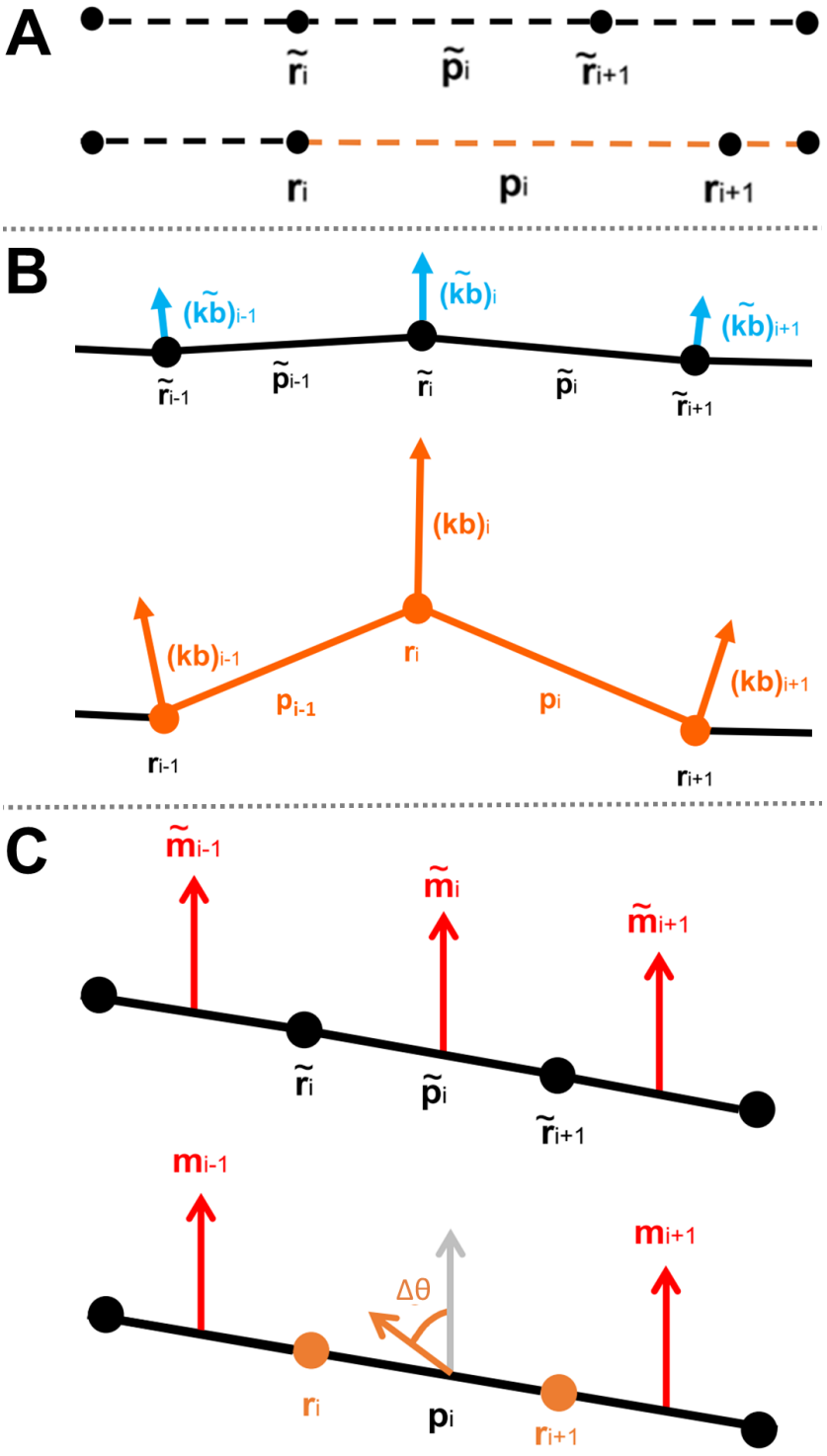


Figure 2.2: Schematic of the three types of elastic deformation that occur in KOBRA rods relative to their equilibrium structure. **A**) Stretching is defined by the increase in length of element $\boldsymbol{p}_i$ due to the movement of nodes $\boldsymbol{r}_i$ and $\boldsymbol{r}_{i+1}$. **B**) Bending at node $\boldsymbol{r}_i$ is defined by the change in length of the associated curvature binormal, $(k\boldsymbol{b})_i$, which represents the change in direction of $\boldsymbol{p}_i$ and $\boldsymbol{p}_{i-1}$. **C**) Twisting about node $\boldsymbol{r}_i$ is defined by the angle, $\Delta\theta$, between adjacent material axes $\boldsymbol{m}_i$ and $\boldsymbol{m}_{i+1}$. Adapted from [85].

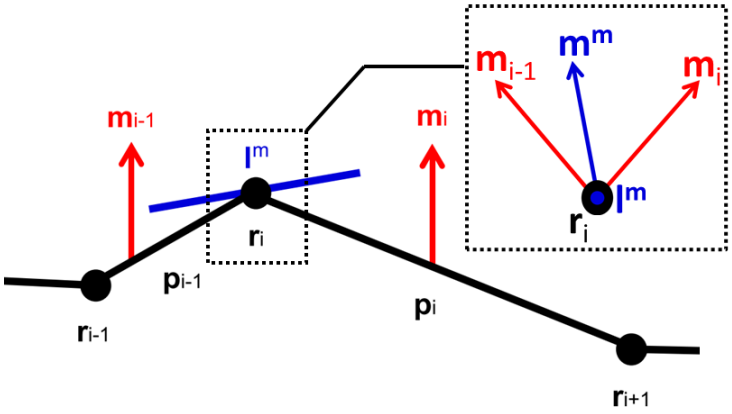


Figure 2.3: Schematic of mutual element used during the bend energy calculation, represented by the unit vector $\boldsymbol{l}^m$ and defined at node $\boldsymbol{r}_i$. **Inset**) The mutual material axis, $\boldsymbol{m}^m$, is constructed from a weighted average of the material axes adjacent to the bent node, $\boldsymbol{m}_i$ and $\boldsymbol{m}_{i-1}$ ($\boldsymbol{l}^m$ points into page). Adapted from [85].

2.2.3 Bend

The definition of the bending energy must be able to account for flexible hinges and anisotropic bending stiffness, both of which are common in biological structures such as coiled coils. The bending energy at a rod node is given by

$$E_{bend,i} = \frac{1}{2\widetilde{L}_i}\left[(\boldsymbol{\omega}_i^m - \widetilde{\boldsymbol{\omega}}_i^m)^T \cdot \boldsymbol{B}_i \cdot (\boldsymbol{\omega_i}^m - \widetilde{\boldsymbol{\omega}}_i^m)\right] \ , \tag{2.9}$$

where $\boldsymbol{\omega_i}^m$ and $\widetilde{\boldsymbol{\omega}}_i^m$ are the current and equilibrium curvature vectors, respectively, and $\boldsymbol{B}_i$ is the positive-definite 2x2 bending stiffness matrix (units m^4 Pa). The curvature is a two-vector that represents the bending of the rod centreline at node i expressed in terms of the local material axes, and is given by

$$\boldsymbol{\omega}_i^m = \begin{pmatrix} (k\boldsymbol{b})_i \cdot \boldsymbol{n}^m \\ -(k\boldsymbol{b})_i \cdot \boldsymbol{m}^m \end{pmatrix} \ , \tag{2.10}$$

where $(k\boldsymbol{b})_i$ is the curvature binormal vector of node i (Figure 2.2B), and $\boldsymbol{n}^m = \boldsymbol{m}^m \times \boldsymbol{l}^m$, where $\boldsymbol{m}^m$ is the mutual material axis, and $\boldsymbol{l}^m$ is the unit vector of the mutual element defined at $\boldsymbol{r}_i$ (Figure 2.3).

The curvature binormal represents the change in direction between two adjacent rod elements as the result of a bend, and is given by

$$(k\boldsymbol{b})_i = \frac{2\boldsymbol{p}_{i-1} \times \boldsymbol{p}_i}{|\boldsymbol{p}_i||\boldsymbol{p}_{i-1}| + \boldsymbol{p}_{i-1} \cdot \boldsymbol{p}_i} \ , \tag{2.11}$$

where the maximum bending energy would require the two rod elements to be overlap-

ping each other, which is very unlikely to happen. Nonetheless, Equation 2.9 is able to account for arbitrarily large bending angles.

The constants used to calculate the elastic energies in Equations 2.5, 2.7, and 2.9 are the rod material parameters: κ_s, β and $\boldsymbol{B}$, respectively, which may be dependent on the position along the molecule. These constants modulate how rods respond to forces during a KOBRA simulation, being somewhat analogous to the internal viscosities and elastic moduli of tetrahedral meshes in FFEA.

2.3 Parameterisation from atomistic simulations

The equilibrium structure and material parameters of rods may be defined manually, using experimentally-derived values [125], or extracted from molecular dynamics trajectories [85]. Single-chain PDB files and coiled coils have been successfully parameterised for KOBRA rods [85, 126], although a generalised workflow for *any* fibrous protein has not yet been developed[3]. This section is written in the context of coiled coils.

Given an atomistic model of a protein and an accompanying molecular dynamics trajectory, the first step in parameterisation is to map the latter onto a coarse-grained rod trajectory (Figure 2.4A). The node positions, $\boldsymbol{r}_i$, at index i are calculated by averaging the atomic positions, $\boldsymbol{r}_k^{atom}$, at index k over a number of clusters,

$$\boldsymbol{r}_i = \frac{1}{k_{max} - k_{min}} \sum_{k=k_{min}}^{k_{max}} \boldsymbol{r}_k^{atom} \quad , \tag{2.12}$$

where k_{max} and k_{min} are atom indices at the edges of the i-th cluster. The optimal cluster size, $k_{max} - k_{min}$, will depend upon the system being discretised, but should be chosen such that structural features contributing to the overall dynamics (e.g. hinges or intrinsic twist) are preserved, whilst averaging out small length scale fluctuations. A cluster size that is too large will produce an overly coarse rod that is limited in its ability to sample conformational space.

Coiled coils and many other proteins have an intrinsic twist at equilibrium that must also be captured. The material axis of the first element, $\boldsymbol{m}_0$, is calculated by averaging displacements between atom pairs located in opposite chains of the coiled coil (Figure 2.4B). Subsequent material axes, $\boldsymbol{m}_i$, are then calculated such that there are no large angular differences between adjacent axes [68].

The rod material parameters, κ_s, β and $\boldsymbol{B}$, are selected such that the local mean square fluctuations of stretching, twisting and bending in the coarse rod trajectory match those

[3]The FFEATools suite, written in Python and packaged separately from the core C++ program, contains many of the functions used during parameterisation.

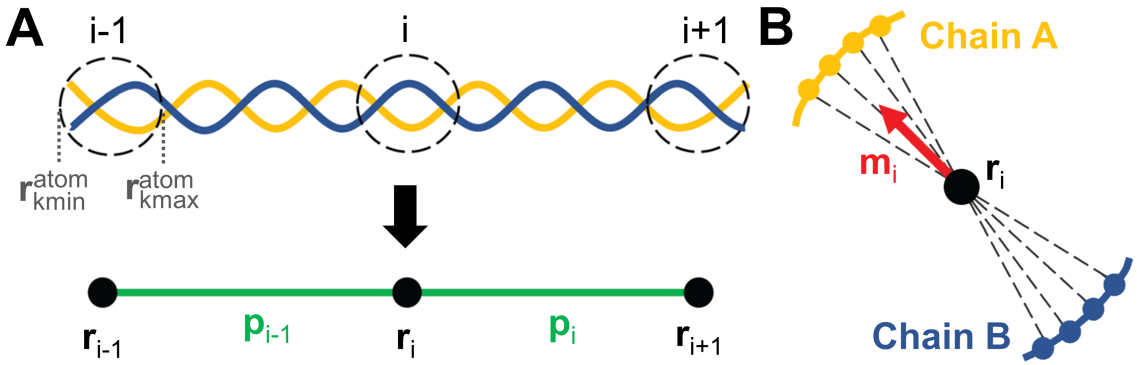


Figure 2.4: KOBRA discretisation of a coiled coil. **A**) Node positions are determined from atomic cluster averaging. **B**) The initial material axis is determined by averaging displacements between atom pairs (dashed lines). Perspective is down the helical axis. Adapted from [127].

of the atomistic trajectory [85]. The quality of the rod model can be determined using principal component analysis, a statistical technique that reduces the dimensionality of a dataset by considering only the most dominant features [128]. If the principle components of motion in both trajectories are highly correlated, then this indicates that the rod model is able to reproduce some of the large-scale dynamics observed in the atomistic simulations [127].

2.4 Rod-tetrahedra interfaces

For molecules with globular as well as filamentous regions it is necessary to be able to join together the KOBRA rod elements with the space-filling FFEA tetrahedral elements[4]. This requires a method to transfer forces between these two representations, described here with the 'rod-to-tetrahedra' attachment convention. At an interface, the end node of a rod, $\boldsymbol{r}_N$, becomes the attachment node, $\boldsymbol{r}_a$, that sits on the outer surface of the attachment tetrahedron. A fictional rod-attachment element, $\boldsymbol{p}_a$, extends from $\boldsymbol{r}_a$ and points towards the centroid of the tetrahedron (Figure 2.5), with an accompanying material axis, $\boldsymbol{m}_a$; these represent a connection only, and not an extension of the physical size of the rod.

The tetrahedral mesh is translated and rotated during initialisation such that there is no twist between $\boldsymbol{p}_a$ and $\boldsymbol{p}_{N-1}$. During the simulation $\boldsymbol{r}_a$ is determined by the position of the attachment tetrahedron, such that forces acting on the mesh are transmitted through to the rod.

Despite their usefulness in constructing mesoscale systems of greater complexity, rod-tetrahedra interfaces require further development in KOBRA. Even with the stipulation that interfaces should begin close to the equilibrium rod structure for optimal numerical stability [127], there is still a chance for undesirable behaviour to occur without consid-

[4]KOBRA and FFEA share the same codebase, despite being named differently. This is to reflect the fact that rods and tetrahedra are distinct objects within a simulation, and are subject to different physical models.

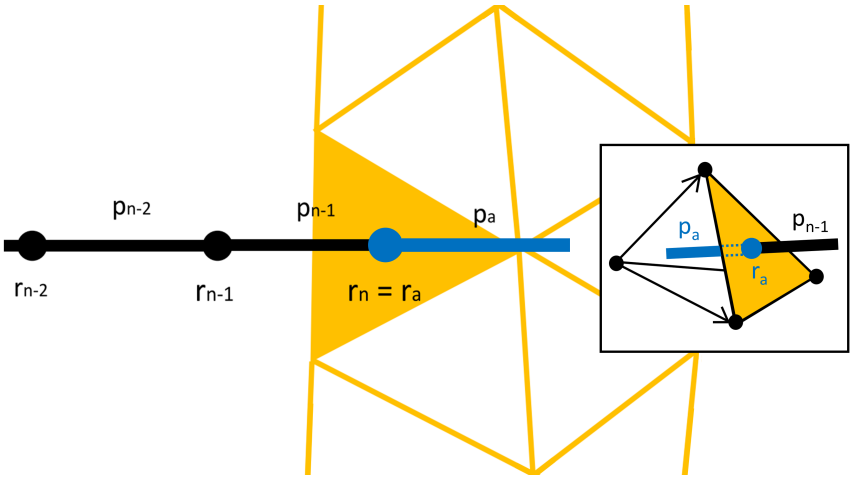


Figure 2.5: Schematic of a rod-to-tetrahedra interface. The attachment node at the end of the rod, $\boldsymbol{r}_n = \boldsymbol{r}_a$, sits on the outer face of the attachment tetrahedron. **Inset**) Tetrahedron structure. The attachment element, $\boldsymbol{p}_a$, connects to the outer face (yellow) and extends towards the centre. Adapted from [127].

erable tweaking of the simulation parameters. For example, for a given interface, if a large twisting force is applied to a rod that is very stiff in comparison to the stiffness of the tetrahedral mesh, the force through the interface may cause the attachment to invert the tetrahedron and crash the simulation. Inversion occurs when the nodes of unstable tetrahedral elements move too far in one direction and pass through a face, turning the element inside-out and resulting in an unphysical negative volume [126].

2.5 Previous studies using KOBRA

2.5.1 Kinetochore-microtubule attachment

The kinetochore (KC) is a protein complex that is critical for the correct segregation of chromosomes during mitotic division [129, 130]. The outer KC consists of at least three molecules [131] of nuclear division cycle 80 complex (Ndc80C), a 60 nm long protein comprised primarily of a coiled coil on either side of a flexible hinge [132], that attach to a spindle microtubule (MT) via a DAM1 ring protein, either perpendicular to the MT surface or at its end, parallel to the MT axis (Figure 2.6A, B). The inner KC connects the chromosome to Ndc80C, allowing the transmission of force through the MT to the chromosome necessary for separation.

Side-on kinetochore-microtubule attachment was modelled by Welch [127] as a test case for the rod-tetrahedra interface code. The outer KC was represented with three Ndc80C molecules parameterised from atomistic MD (Figure 2.6C) as rods with a central, flexible hinge region and a rod-tetrahedra interface at either end. The top meshes were attached to a spherical abstraction of the inner KC via elastic springs and the lower meshes sat hovering just above the MT surface (Figure 2.6D). The DAM1 ring was not represented

due to the absence of rod-rod interactions in KOBRA at the time. The binding behaviour and flexibility of Ndc80C was analysed in four system configurations, and it was found that the flexibility of Ndc80C was restricted by its interactions with the cargo and the MT, and the hinge allowed it to bind more easily to sites on the MT surface.

2.5.2 Myosin 5a lever structure and flexibility

Myosin 5a (Myo5a) is a motor protein driven by ATP hydrolysis that binds and 'walks' along cytoskeletal filamentous actin (F-actin) [135], in order to transport cellular cargo such as vesicles and organelles to the cell perimeter [136]. The full protein is a dimer that broadly consists of two motor domains that bind F-actin, each with a lever domain, where they dimerise at the tail domain that binds cargo (Figure 2.7A, B) [137]. There remains a lack of understanding of how the mechanics of individual components of Myo5a, particularly the lever domain, contribute to its overall transport [126].

The structure of Myo5a subfragment-1 with its full-length lever (Myo5a-S1-6IQ) was determined using cryoEM (Figure 2.7C) and its flexibility was characterised as being isotropic in both bending directions, but varying throughout the structure [125]. A rod-tetrahedra model of the dimer was built from two copies of the monomeric structure and found to be experimentally consistent (Figure 2.7D), with a novel meshing algorithm being developed to increase the number of accessible timesteps. Simulations of out-of-equilibrium strained states were run in an attempt to replicate walking of Myo5a [126].

2.6 Extending the rod model

In order to make the most efficient use of development time during this project, it is necessary to distinguish between software features that are *essential* versus *desirable*. Essential features significantly impact the fundamental modelling capabilities of KOBRA and are imperative for achieving physical realism; therefore, they should be given the highest priority. In contrast, desirable features may primarily aim to enhance the user experience or expand the software's scope, and a wishlist of such features could potentially be endless; only those deemed most pertinent to the study of fibrinogen are presented here.

2.6.1 Essential functionality

Coarse-grained simulations necessarily have to make compromises between accuracy and computational tractability [36, 46], however, KOBRA is severely limited due to its lack of interactions between rods, specifically short-range steric repulsion of overlapping atoms and mid-range attraction e.g. from VDW forces. Such physics is critical to modelling the biological mesoscale, especially considering the complex and crowded nature of the cellular environment [138], wherein most (if not all) molecular processes crucially depend on interactions between proteins. These effects are already captured by FFEA [37] and

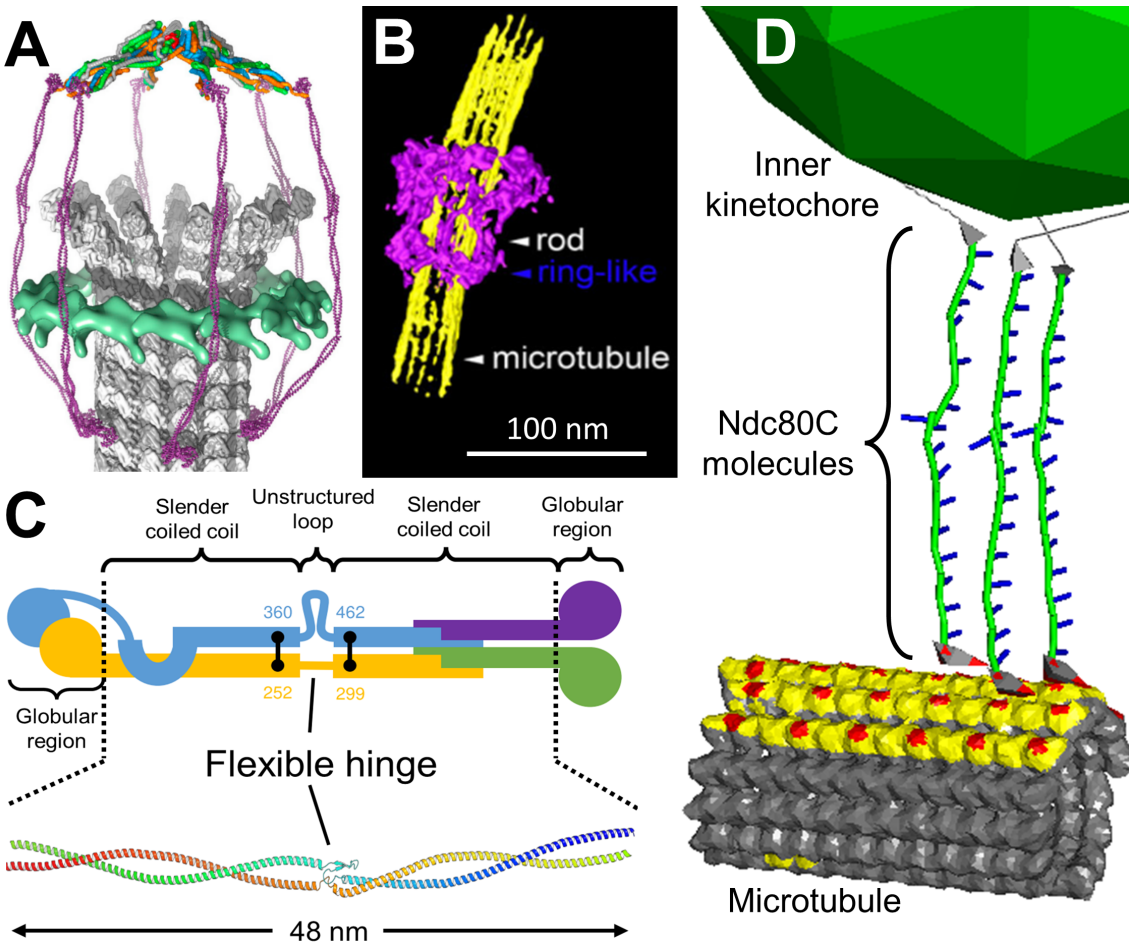


Figure 2.6: Previous study of kinetochore-microtubule attachment using KOBRA and FFEA. **A**) Proposed structure of the KC, including the DAM1 ring (green) and six Ndc80C proteins (purple), attached end-on to a MT (grey). The DAM1 structure was obtained from cryoEM and the Ndc80C globular domains from X-ray crystallography, but the remaining structures are hypothetical. Adapted from [133]. **B**) Projection of a 3D electron tomogram of a KC (pink) attached to a MT (yellow). A ring-like structure and one of several rods are labelled . Adapted from [134]. **C**) **Top:** Schematic of Ndc80C protein with cross-links (black) marked at their respective residues. The four subunits are differentiated by colour. **Bottom:** The equilibrium atomic model of the hinge and coiled-coil sections, constructed partly from the 'Bonsai' molecule (2VE7) [132]. Adapted from [127]. **D**) Simulation frame of side-on kinetochore-microtubule attachment, comprised of three Ndc80C complexes with their lower ends suspended above the MT surface, and the upper ends attached to a sphere (radius = 50 nm) via elastic springs. Each Ndc80C is represented as a rod (coiled-coils and hinge) connected at each end by a small tetrahedral mesh (globular domain). The sphere is an abstraction of the inner KC. Red and yellow tetrahedral elements denote attractive VDW regions of different energies. Grey elements, the sphere and rods are non-interacting. Adapted from [127].

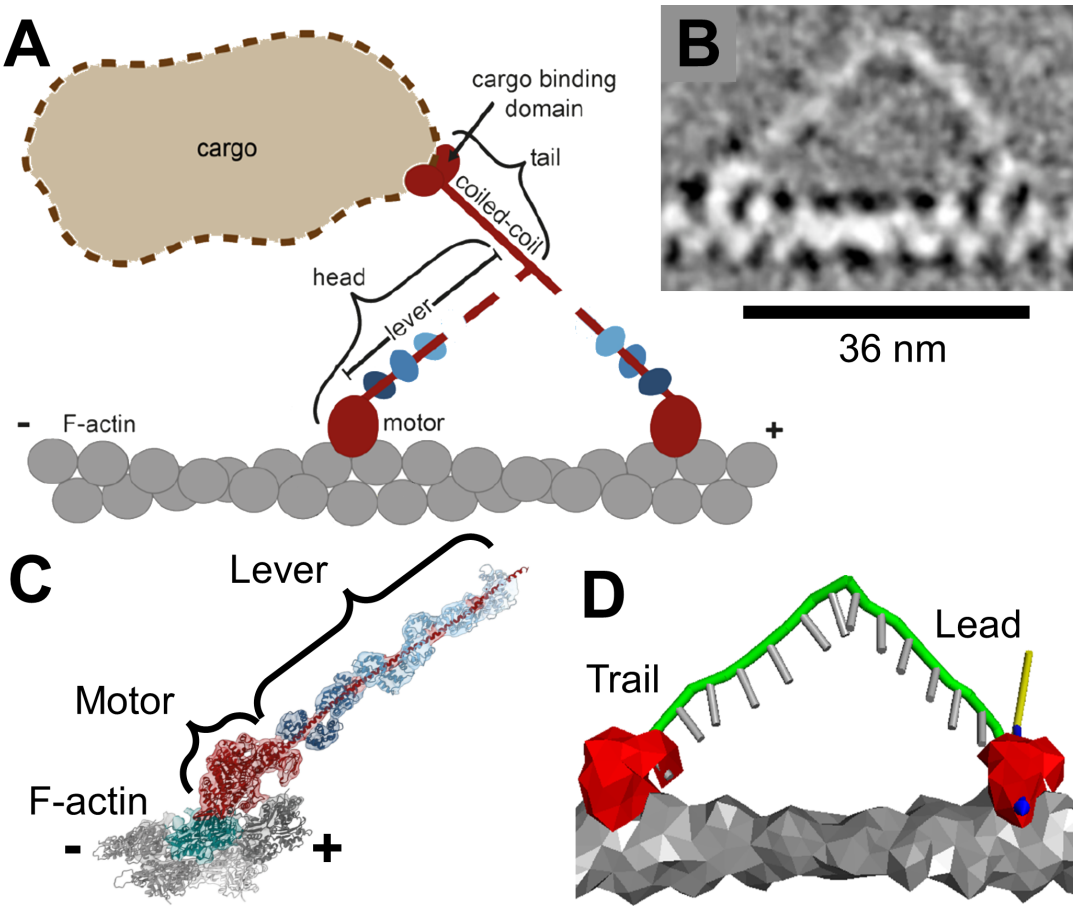


Figure 2.7: Previous study of Myo5a lever structure and walking mechanics using KOBRA and FFEA. **A**) Schematic of Myo5a walking on F-actin with domains highlighted. Heavy chains are coloured red and light chains blue. Adapted from [126]. **B**) Cryoelectron micrograph of dimeric Myo5a walking along F-actin. Adapted from [36]. **C**) CryoEM structure of Myo5a-S1-6IQ bound to F-actin, with fitted pseudoatomic model [125]. **D**) Final simulation frame of Myo5a in a non-equilibrium strained state. The model was constructed from the monomeric structure [125], with the motors represented as tetrahedra (red) and the levers as rods (green). During the simulation, translational (blue) and rotational (yellow) forces were applied to the lead motor to create a strained state. Adapted from [126].

other mesoscale models [36, 71, 139], so it is clearly a priority to update KOBRA to have an equivalent level of physical description.

The absence of steric interactions means that rods in KOBRA can pass through each other. Their application is hence limited to specific systems with protein volume fractions below a certain threshold [70], wherein rods would be unlikely to come into contact. The KC-MT system in Figure 2.6 was deemed sufficiently sparse, although this only simulated three Ndc80C molecules, the minimum number thought to be required for attachment [131].

2.6.2 Desirable features

A feature common to molecular simulation programs, yet absent from KOBRA, is periodic boundary conditions (PBCs). This computational method involves creating an artificial boundary that encloses the simulation box, such that particles that cross one edge of the box reappear at the opposite edge with no changes to momentum, effectively simulating an 'infinite' spatial domain [140]. Whilst not as critical as intermolecular interactions to simulating biological macromolecules, PBCs are necessary to describe bulk materials, such as proteins in an aqueous solution without the presence of a substrate or physical boundary.

Fluid-structure interactions of KOBRA rods are modelled as an implicit solvent of spherical (translational) and cylindrical (rotational) viscous drag against a static background fluid, with no long-range hydrodynamic correlations. Drag against a *moving* background would make it trivial to set up unidirectional flow conditions in a KOBRA simulation, by defining a single background field vector at input. Shearing flow requires a special case of PBCs known as Lees-Edwards boundary conditions (LEBCs) [141], which were previously implemented for tetrahedra and used to study packed colloid interactions [142], but were not factored into the main codebase.

Chapter 3

Development of rod-rod interactions in KOBRA

The previous chapter gave a summary of the KirchhOff Biological Rod Algorithm (KOBRA) software as it existed in 2020. A given protein with some length, L, cross-sectional radius, R, and a large aspect ratio, $L/2R \gg 1$, is discretised in KOBRA as a rod: a linear series of interconnected nodes with three elastic degrees of freedom - stretch, twist, and bend - that act to minimise deformations from thermal fluctuations by moving the rod towards its equilibrium structure. The line segments between nodes are the rod elements, which do not interact with neighbouring elements. Only thermal and elastic forces influence rod dynamics, regardless of proximity to other rods [85].

Fibrinogen is a fibrous protein, with $L/2R \sim 9$, but the original version of KOBRA is insufficient to model key aspects of its biological function. The polymerisation of fibrin(ogen) monomers into the fibrin network involves many molecules being in close proximity, and multiple types of intermolecular interaction, with noncovalent knob-hole 'bonds' being some of the earliest to form in the process [61, 106]. Furthermore, the *in-vivo* environment of fibrinogen is an aqueous suspension (blood plasma), so simulations must also capture the bulk fluid properties.

This chapter will focus on the continued development of the KOBRA software, aimed at extending its capabilities to model biomolecular interactions between rods, both repulsive and attractive, in addition to periodic boundary conditions. Sections detailing unit and integration tests, performance benchmarks of simulations containing large numbers of rods, and simulations of protofibrils assembled from smaller fibrinogen rods, are also presented. The software development that forms this chapter is the sole work of the author of this thesis.

3.1 Mesoscale biomolecular interactions

Steric interactions, in which atoms repel each other sharply when their electron shells overlap, are a key property of physical systems and arise from quantum mechanical effects [16]. The representation of steric effects in biomolecular simulations is essential to ensure that objects do not pass through each other when they are in close proximity. This is often achieved by applying a repulsive potential energy between the molecules being simulated. An exception to this is simulations of highly dilute suspensions, which are able to neglect steric interactions due to molecules being sufficiently far apart that the probability of interaction is minimal (§B.1) [54, 70].

Van der Waals (VDW) interactions are nonspecific and mid-ranged forces that arise between molecules due to an induced dipole moment, which is also quantum mechanical in origin. Fluctuations in the electron shells of an atom can give rise to an instantaneous electric dipole, which then induces further dipoles in neighbouring atoms, resulting in attractive or repulsive forces between them. They are fairly weak interactions that occur over distances of 5-10 Å, with the strength decreasing rapidly as the distance increases [16].

Hydrogen bonds (H-bonds) are short-ranged attractive forces that arise between 'donor' atoms with a partial positive charge and 'acceptor' atoms with a partial negative charge. With hydrogen as the donor, the most common acceptor atoms are nitrogen and oxygen. Although hydrogen bonds are weak individually, when grouped together they confer significant structural stability, such as in antiparallel β-sheets [143] and the DNA double helix [144]. The effective distance of most hydrogen bonds in proteins is 1-3 Å [145].

Covalent bonds, on the order of 1-2 Å in length [146], are sufficiently strong that in atomistic simulations they are represented as permanent harmonic potentials, whereas VDW and H-bonds are treated as non-bonded interactions (Equation 1.1). However, electrostatic effects are screened at distances larger than the Debye length of the solution, which in physiological environments is around 0.7 nm [147]. This screening means that VDW interactions, H-bonds and long-range electrostatics can all be treated as a similar kind of potential in mesoscale simulations [36]. Protein-protein interactions may be modelled as acting between entire domains, rather than specific atoms or residues [52, 71, 139, 148].

3.2 Hard-core steric repulsion is unsuitable for mesoscale modelling

The Lennard-Jones (LJ) potential is a well-known analytical expression that approximates both steric and VDW forces, and is given by

$$U_{LJ} = 4\epsilon\left[\left(\frac{\sigma}{s}\right)^{12} - \left(\frac{\sigma}{s}\right)^{6}\right] , \tag{3.1}$$

where s is the distance between the centres of two interacting particles, ϵ is the maximum potential energy of the interaction, and σ is the distance at which the potential is zero. The distance at which the potential well is deepest, $U_{LJ} = -\epsilon$, is given by $s_{min} = 2^{1/6}\sigma$. The repulsive and attractive force regimes come from the $1/s^{12}$ and $1/s^{6}$ terms, respectively.

The two adjustable parameters and functional form of the LJ potential lend it a versatility that is useful for modelling generic attractive interactions at the biological mesoscale, when the exact form may not be known [36].

Atoms in MD simulations typically exhibit 'hard-core' steric repulsion, where their charge radii are not permitted to overlap. The s^{12} term in Equation 3.1 encapsulates the phenomenon by causing the potential energy to tend to infinity as the inter-atomic distance goes to zero.

While such a function does in theory give the required hard-core repulsion, in practice the steepness of this potential makes it unsuitable for CG simulations. Consider at some simulation time, t_0, that two rod elements are close, but not intersecting. In the following time step, t_1, thermal noise may cause the rods to have a small intersection. With the potential of Equation 3.1, even a very small overlap causes a huge rise in potential energy, and the large resulting force may be sufficient to destabilise the simulation.

Preventing this spike requires either: **1**) reducing the time step to resolve the small-scale motion against this stiff spring response, severely limiting the time scales reachable by KOBRA), or **2**) modifying the LJ potential to give a softer repulsion, wherein the repulsive force allows some inter-penetration but still prevents rods from passing through one another. In the case of the tetrahedral FFEA simulation a repulsive steric force proportional to the volume of overlap between intersecting tetrahedral elements was implemented, for which a fast pre-existing algorithm was used [37, 149].

Figure 3.1 shows a modified LJ potential alongside its original form, where s now represents the surface-surface distance, such that $s < 0$ corresponds to an overlapped configuration. At $s < s_{min}$, the sharp increase in U_{LJ} has been replaced by a softer potential (I) that requires an interpolative potential (II) to bridge the attractive part of the LJ 6-12 potential (III) [39]. The three potential regimes may be written together as

$$U = \begin{cases} ks^2 , & s < 0 \\ \epsilon\left[2\left(\frac{s}{s_{min}}\right)^3 - 3\left(\frac{s}{s_{min}}\right)^2\right] , & 0 \leq s < s_{min} \\ 4\epsilon\left[\left(\frac{\sigma}{s}\right)^{12} - \left(\frac{\sigma}{s}\right)^{6}\right] , & s \geq s_{min} \end{cases} , \tag{3.2}$$

and are each discussed in further detail throughout this chapter.

The steric potential has maximum of 50 k_BT at $s = -2R = -5s_{min}$, which is sufficient to make rod crossing highly improbable. The attractive potential minimum of $\epsilon = 8.5$ k_BT was calculated from the force distribution of fibrinogen knob-hole interactions [110].

The modified LJ potential has been previously used in FFEA tetrahedra to approximate both specific and non-specific attractive interactions between the surfaces of proteins [37, 39, 52]. The rod implementation of this is described in §3.3.5 and §3.4.3.

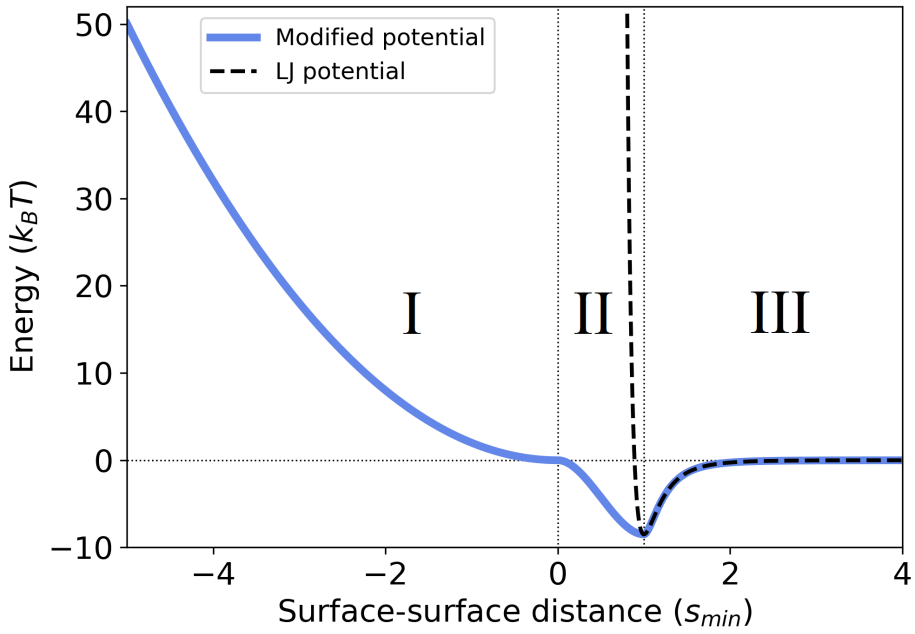


Figure 3.1: Modified Lennard-Jones potential (blue) with the soft steric (I), interpolative (II), and 6-12 (III) regimes. Attractive potential well parameterised from fibrinogen knob-hole interactions [110].

3.3 Repulsive steric interactions

3.3.1 Crude distance calculation and cutoff

A significant amount of computation time can be saved by neglecting interactions between elements that are sufficiently far apart for the interaction energy to be negligible. For a rod element, i, the steric cutoff radius is equivalent to the element length, $|\boldsymbol{p}_i|$, plus its cross-sectional radius, R_i, originating from the midpoint along its axis (centreline), $\boldsymbol{r}_i^{mid} = \boldsymbol{p}_i + \frac{1}{2}\boldsymbol{r}_i$. The rod axis vector, $\boldsymbol{p}_i$, is defined as the line connecting two node positions on the same rod, $\boldsymbol{r}_i$ and $\boldsymbol{r}_{i+1}$, such that $\boldsymbol{p}_i = \boldsymbol{r}_{i+1} - \boldsymbol{r}_i$. A 'crude' distance is calculated between the midpoints of a pair of elements, $|\boldsymbol{r}_{ij}^{mid}| = |\boldsymbol{r}_j^{mid} - \boldsymbol{r}_i^{mid}|$, where j is a different, non-adjacent rod element. If the Boolean condition,

$$|\boldsymbol{r}_{ij}^{mid}| < \texttt{max}\big(|\boldsymbol{p}_i|,\ |\boldsymbol{p}_j|\big) + R_i + R_j \ , \tag{3.3}$$

is `true`, then j is within the cutoff radius of i, indicating the pair might be involved in a steric collision; `false` results are discarded. The inter-element midpoint distance is crude in the sense that it does not account for the relative orientation of the elements, but is fast to calculate and so provides a rapid means to exclude non-interacting pairs of elements. Equation 3.3 is computed between i and every other element in the system (Figure 3.2). The next element along the rod is selected in place of i, and the procedure is repeated, until all elements on all rods have checked for potentially colliding elements.

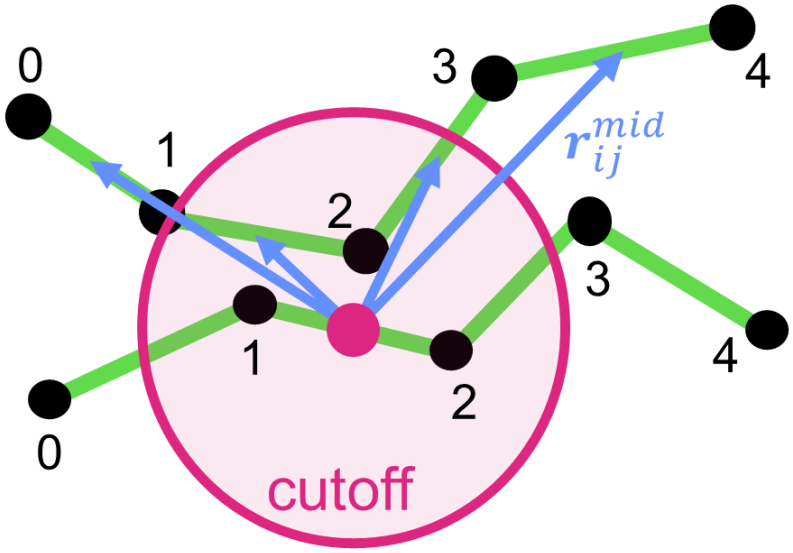


Figure 3.2: Schematic of the crude steric cutoff calculation. The midpoint distance is measured from element 1 on the bottom rod, to all other elements on the top rod (blue arrows). The midpoints of elements 1 and 2 on the top are within the cutoff radius of element 1 on the bottom, so they may potentially interact.

Self-interactions have not yet been implemented into KOBRA, so i and j in Equation 3.3 are assumed to exist on separate rods, as illustrated in Figure 3.2. They are necessary for modelling highly flexible and/or coiled polymers, however, it is assumed that the molecules under consideration are sufficiently stiff that their total length is significantly shorter than their persistence length, so that self-crossing events can be neglected. Implicitly, adjacent elements are also not included; their energetic contributions are accounted for by calculating the internal elastic energy calculation.

The above method is simple to implement, but requires a distance calculation between all possible element pairs. For very large systems this is potentially more computationally expensive than grid or voxel-based 'bucketing' methods that only consider interactions between elements within a similar region of space [142]. In sparse systems where rods are far apart, the steric cutoff radius is likely to be small relative to the system size, resulting

in a large number of redundant crude distance calculations.

3.3.2 Minimum displacement between two infinite lines

In order to determine the actual closest distance between two rod elements, i and j, their relative orientations must be taken into account. To do this the minimum displacement between two infinite straight lines, formed by extending the line segments corresponding to i and j, is considered. The pair of points, $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$, that form the minimum displacement between the centrelines, $\boldsymbol{c}_{ij} = \boldsymbol{c}_j - \boldsymbol{c}_i$, can be calculated as

$$\begin{aligned} \boldsymbol{c}_i &= \boldsymbol{r}_i + \left[\frac{(\boldsymbol{r}_j - \boldsymbol{r}_i)\cdot\left[\boldsymbol{l}_i \times (\boldsymbol{l}_i \times \boldsymbol{l}_j)\right]}{\boldsymbol{l}_i \cdot \left[\boldsymbol{l}_i \times (\boldsymbol{l}_i \times \boldsymbol{l}_j)\right]}\right]\boldsymbol{l}_i \\ \boldsymbol{c}_j &= \boldsymbol{r}_j + \left[\frac{(\boldsymbol{r}_i - \boldsymbol{r}_j)\cdot\left[\boldsymbol{l}_j \times (\boldsymbol{l}_i \times \boldsymbol{l}_j)\right]}{\boldsymbol{l}_j \cdot \left[\boldsymbol{l}_j \times (\boldsymbol{l}_i \times \boldsymbol{l}_j)\right]}\right]\boldsymbol{l}_j \quad , \end{aligned} \tag{3.4}$$

where i and j are indices denoting separate rod elements, $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ are the first node positions on each element, and $\boldsymbol{l}_i$ and $\boldsymbol{l}_j$ are the unit vectors pointing along the element axes (previously defined in Figure 2.1). The displacement calculation is illustrated schematically in Figure 3.3.

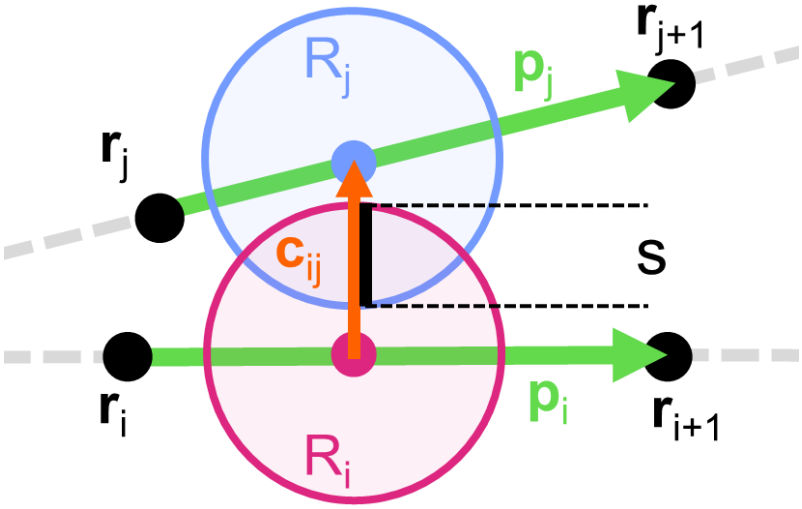


Figure 3.3: Schematic of the minimum centreline displacement, $\boldsymbol{c}_{ij}$, between two elements, i and j, during a steric collision. Elements are initially assumed to lie along infinite lines (grey dashes). Element radii, R_i and R_j, define the effective steric volume of the rods. The surface-surface distance, s, is negative during a steric collision.

Equation 3.4 is undefined if $\boldsymbol{l}_i$ and $\boldsymbol{l}_j$ are exactly parallel, due to a division by zero. Despite translational and rotational thermal noise causing elements to constantly move, and three elastic degrees of freedom, the probability that two nearby elements will become parallel, although very low, increases with simulation time. A simple workaround is to define $\boldsymbol{c}_i$

and $\boldsymbol{c}_j$ as the midpoints of their respective elements if they are parallel. The inaccuracy of this approach is offset by the low probability.

3.3.3 Discrete element correction

Because Equation 3.4 assumes that both rod elements lie along infinite lines, $\boldsymbol{c}_i$ and/or $\boldsymbol{c}_j$ occasionally get placed outside the discrete bounds of their elements.

To solve this, a constraint is applied to the placement of $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ by re-assigning them to the nearest node if they lie outside their respective element. For element i, the point $\boldsymbol{c}_i$ is altered according to the mutually exclusive conditions,

$$\boldsymbol{c}_i = \begin{cases} \boldsymbol{r}_i, & \boldsymbol{p}_i \cdot \vec{\boldsymbol{cr}} \leq 0 \\ \boldsymbol{r}_i + \boldsymbol{p}_i, & \boldsymbol{p_i} \cdot \vec{\boldsymbol{cr}} \geq |\boldsymbol{p}_i|^2 \end{cases}, \tag{3.5}$$

where $\vec{\boldsymbol{cr}} = \boldsymbol{c}_i - \boldsymbol{r}_i$. Equation 3.5 is also applied to $\boldsymbol{c}_j$ [127, 150].

This constraint is sufficient for most configurations of element pairs during a simulation, however, a correction is required for cases where elements are 'nearly parallel' and cause erroneous calculations of $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ (Figure 3.4). For ease of description, it is assumed here that $\boldsymbol{p}_i$ points along the Cartesian x axis. Four traits were observed that, when combined, reproduced the bug:

1. Elements are separated in x by over half an element length.
2. Element centrelines do not intersect.
3. The angle, θ, between $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ in the x-y plane, is sufficiently small
4. The angle, ϕ, between $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ in the x-z plane, is sufficiently small

Typical angles observed were $\theta, \phi < 10^\circ$ for elements with aspect ratio, $|\boldsymbol{p}|/2R \geq 2$, although this was not rigorously quantified.

In the nearly parallel case, the points initially calculated by Equation 3.4 are outside one or both finite elements i and j. Equation 3.5 reassigns these erroneous points to the corresponding nearest nodes, however, the offset in x between i and j means that the new displacement is not a minimum.

To overcome this, distances are measured from $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ to the nodes on the opposite element. If any are smaller than the displacement obtained from the constraint of Equation 3.5, a 'better' minimum displacement is chosen by reassigning $\boldsymbol{c}_i$ and/or $\boldsymbol{c}_j$ to the appropriate nodes (Figure 3.4).

The crude cutoff and inter-element displacement calculations occur sequentially: if Equation 3.3 returns `true` for elements i and j, then Equations 3.4 and 3.5 are immediately

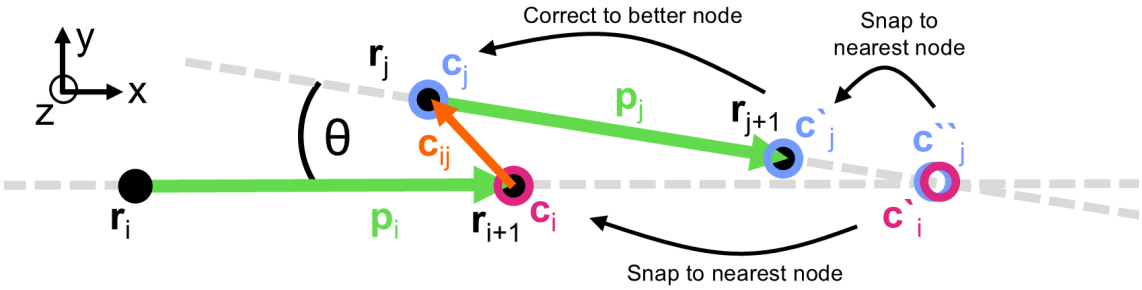


Figure 3.4: Schematic of a correction to the centreline displacement, due to an edge case in which the rod elements are nearly parallel. Both points are initially placed outside the bounds of their discrete elements, $\boldsymbol{c}_i'$ and $\boldsymbol{c}_j''$, at the mutual point where the infinite lines (grey dashes) cross. Each point is snapped to its nearest node, $\boldsymbol{c}_i'$ to $\boldsymbol{r}_{i+1}$ and $\boldsymbol{c}_j''$ to $\boldsymbol{r}_{j+1}$, however, node $\boldsymbol{r}_j$ is clearly the better choice to minimise the displacement. The correction places $\boldsymbol{c}_j'$ onto $\boldsymbol{r}_j$, to give the final pair of points forming the displacement, $\boldsymbol{c}_{ij}$.

applied.

3.3.4 Nearest neighbour assignment

In the steric calculation each element is defined to have an effective volume comprised of a 3D capsule: a cylinder of radius R from the axis vector, $\boldsymbol{p}$, capped at both ends by hemispheres of radius R from both node positions, $\boldsymbol{r}$ and $\boldsymbol{r} + \boldsymbol{p}$. Thus if the Boolean condition,

$$|\boldsymbol{c}_{ij}| > 0 \;\texttt{and}\; |\boldsymbol{c}_{ij}| < R_i + R_j \;\;, \tag{3.6}$$

returns `true`, this indicates that the volumes of i and j are partially overlapping and hence sterically interacting. If $|\boldsymbol{c}_{ij}| = 0$ then the element axis vectors have intersected and the simulation intentionally crashes, but this scenario can be avoided by selecting an appropriate time step (§B.2).

Interactions are recorded in a neighbour list, implemented as a dynamic 2D array. Each array element stores data on a single interaction event, such as rod identifiers, node positions, radii, etc.

During nearest neighbour assignment, every rod in the system is looped over pairwise, but the elements of each rod pair are not, so for N rods with M elements each, the time complexity scales as order $\mathcal{O}(\frac{1}{2}N[N-1]M^2)$. The end result is that there is some duplicated data across multiple rods, e.g. element i on rod 1 stores data on the interaction with element j on rod 2, but element j *also* stores a copy of that data. Whilst it would be computationally more efficient to avoid this duplication, nearest neighbour assignment only occurs once per time step, before the internal elastic energies of the rods are calculated.

3.3.5 Soft repulsive potential

The final step in the steric collision algorithm is to determine the energy, and hence force, of repulsion for each interaction. As discussed above the quadratic potential shown in Figure 3.1 is used,

$$U_{steric} = ks^2 = k\big[|\boldsymbol{c}_{ij}| - (R_i + R_j)\big]^2 \ , \tag{3.7}$$

where k is a constant defining the strength of the repulsive force, s is the surface-surface distance between the elements, and for steric interactions $s < 0$. Here k is chosen such that the steric potential has a maximum of 50 k_BT when the elements have fully overlapped, i.e. $s = -2R$, which for $R = 2.5$ nm gives $k = U/s^2 = 0.0082$ N m^{-1}. This provides an energy barrier much greater than k_BT such that it is highly unlikely that thermal fluctuations can cause elements to pass through another, but not so large as to make the system overly stiff.

Scaling the repulsive force with the element radius means that thinner rods will collide with greater stiffness, however, even with $R = 1$ nm, below which the CG approach of KOBRA becomes redundant, the potential remains softer than Lennard-Jones. For comparison, the diameter of a DNA double helix is ~ 2 nm [151].

The magnitude of the repulsive force that acts on both elements, $\boldsymbol{F}_{elem}$, is calculated from the central difference gradient of the energy,

$$|\boldsymbol{F}_{elem}| = -\frac{dU}{ds} \simeq \frac{\Delta U}{\Delta s} = \frac{U(s + \Delta s) - U(s - \Delta s)}{2\Delta s} \ , \tag{3.8}$$

where Δs is a small change to the surface-surface distance.

Equal and opposite forces are projected onto elements i and j along the line of the inter-element displacement. The force must be interpolated onto the nodes that form each element, $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, to avoid introducing torque into the rod, and to easily include the steric force in the rod node position update at the end of the time step. The magnitude of the node forces are weighted by their point of action along the element centreline,

$$\begin{aligned} \boldsymbol{F}_1 &= w_1 \boldsymbol{F}_{elem}, \qquad w_1 = \left[\frac{|\boldsymbol{p}| - l_1}{|\boldsymbol{p}|}\right] \\ \boldsymbol{F}_2 &= w_2 \boldsymbol{F}_{elem}, \qquad w_2 = \left[\frac{|\boldsymbol{p}| - l_2}{|\boldsymbol{p}|}\right] \ , \end{aligned} \tag{3.9}$$

where w_1 and w_2 are weights, $0 \leq w \leq 1$, and l_1 and l_2 are intra-element distances,

$$\begin{aligned} l_1 &= |\boldsymbol{c} - \boldsymbol{r}_1| \\ l_2 &= |\boldsymbol{p}| - l_1 \ , \end{aligned} \tag{3.10}$$

such that a force acting entirely at the first node of an element, $\boldsymbol{c} = \boldsymbol{r}_1$, will have $l_1 = 0$, hence $\boldsymbol{F}_1 = \boldsymbol{F}_{elem}$ and $|\boldsymbol{F}_2| = 0$.

Equations 3.7 to 3.10 represent the force calculation from a single steric interaction event. The interpolated node forces are summed up over all interactions between an element and its neighbours. The net steric force on a node, required for the rod dynamics update, is determined once both adjoining elements (excluding the ends of the rod) have completed the sum over their neighbours.

3.4 Attractive van der Waals interactions

In addition to steric repulsive forces, there are also attractive forces between molecules in the form of VDW interactions. The detailed electrostatic interactions that give rise to these forces are not modelled within KOBRA. Here, pairwise attractive forces are introduced that act between user-specified positions on molecules. Whilst VDW interactions are generally weak and nonspecific, occurring throughout most parts of a molecule, the implementation described in this section may also represent stronger specific interactions, such as knob-hole interactions between fibrin monomers during the early stages of polymerisation [61].

3.4.1 Definition of van der Waals sites

A VDW site is defined in KOBRA as a point lying on the rod centreline that exhibits an attractive, isotropic potential field, with the range and strength defined by the user. Attractive interactions occur between VDW sites, rather than between rod elements as with repulsive steric interactions.

When a rod is created using FFEATools, the user can add any number of VDW sites to it, which are written to a VDW parameter file, `.rodvdw`. Each row consists of an integer, $n_{site} = \{0, 1, ...6\}$, corresponding to a VDW interaction type, and a float, $0 \leq L_{site} \leq 1$, the fraction of the rod centreline length[1] at which the site is located (Figure 3.5). Specifying a single float and calculating the 3D spatial position internally makes the setup of a simulation easier for the user. Sites are not bound to the rod discretisation. One `.rodvdw` file is required for each unique configuration of VDW sites, and multiple rods may use the same file.

[1] This length does not include the steric radius at both ends of the rod.

A

File: curvy.rodvdw

```
1  ffea rod vdw file
2  num_sites 5
3  vdw params:
4  0, 0.000
5  0, 0.250
6  0, 0.500
7  0, 0.750
8  0, 1.000
```

B

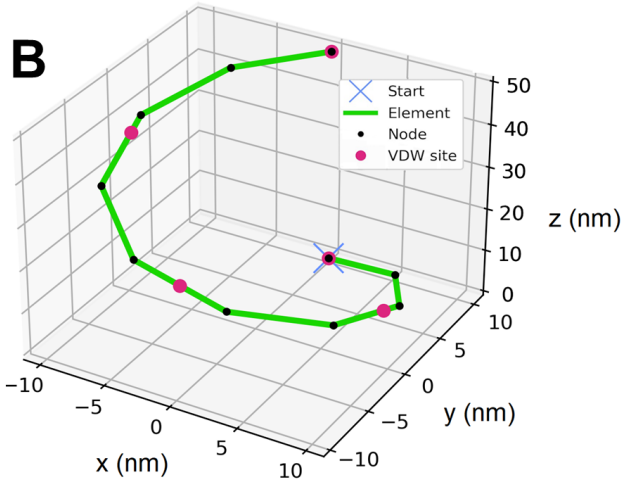


Figure 3.5: **A**) Sample `.rodvdw` file with five VDW sites, generated with FFEATools. Interaction types, n_{site}, are on the left, and centreline length fractions, L_{site} are on the right. **B**) 3D plot of a curved rod with VDW sites positioned according to **A**.

An additional file is required to define the Lennard-Jones (LJ) parameters, `.rodlj`, which stores a matrix of tuples, $P_{kl} = (\sigma_{kl}, \epsilon_{kl})$, where k and l are integers corresponding to a VDW interaction type, and $(\sigma_{kl}, \epsilon_{kl})$ have units of (length, energy). During the neighbour list update, k and l are used to reference $\boldsymbol{P}$ to obtain the LJ parameters for the interaction between each pair of VDW sites. With seven indices available, up to 49 unique LJ interactions may be defined. One `.rodlj` file is required per simulation.

3.4.2 Distance calculation

Much of the logic of inter-site distance calculations and assignment to the VDW neighbour list is similar to that previously described for the steric interactions, although there are some key differences that require elaboration. Interactions occur between pairs of VDW sites within a cutoff radius of each other, given by

$$|\boldsymbol{c_j} - \boldsymbol{c_i}| < 5s_{min} + R_i + R_j \ , \tag{3.11}$$

where $\boldsymbol{c_i}$ and $\boldsymbol{c_j}$ are the positions of the VDW sites on the centrelines of their respective rod elements, i and j, and s_{min} is the surface-surface distance at which the Lennard-Jones potential well is deepest, $U_{LJ} = -\epsilon$. A cutoff of $5s_{min}$ yields $U_{LJ} \simeq -10^{-4}\epsilon$, which is sufficiently small as to be negligible for influencing rod dynamics. The general surface-surface distance is given by

$$s = |\boldsymbol{c_j} - \boldsymbol{c_i}| - (R_i + R_j) \ . \tag{3.12}$$

The minimum displacement between potentially interacting sites is greatly simplified

compared to Equation 3.4 due to the sites being points rather than line segments. The points forming the displacement, $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$, are equivalent to the positions of the sites on each rod. Corrections to the displacement, $\boldsymbol{c}_{ij}$, are not required.

Each rod stores VDW interactions in a second neighbour list, decoupled from steric interactions. The list structures are identical, with the exception that VDW lists are additionally initialised with LJ parameters, σ_{kl} and ϵ_{kl}, read from the `.rodlj` file, where k and l denote interaction types.

3.4.3 Modified attractive potential

A major difference between the VDW and steric interactions is in the potential energies used to calculate the force. In §3.8 of Hanson [39] a soft-core Lennard-Jones potential is defined for the FFEA tetrahedral model, which has been adapted here for KOBRA rods.

This modified LJ potential, illustrated in Figure 3.1, shows interpolative and attractive potential energy regimes for $s > 0$, where s is given by Equation 3.12. An interpolative region is required for the smooth transition between soft steric repulsion and LJ attraction, in the range $0 \leq s < s_{min}$. This was previously calculated by Hanson [39] using the boundary conditions $U = 0$, $dU/ds = 0$ at $s = 0$, and $U = -\epsilon$, $dU/ds = 0$ at $s = s_{min}$, yielding a polynomial potential of the form,

$$U^{int} = \epsilon \left[2\left(\frac{s}{s_{min}}\right)^3 - 3\left(\frac{s}{s_{min}}\right)^2 \right] \ . \tag{3.13}$$

For the range $s_{min} \leq s < 4s_{min}$, beyond which VDW interactions are neglected, the standard form of the LJ potential, Equation 3.1, is used, which contains no spikes in potential energy.

Although VDW interactions occur between sites, the resulting forces must still be interpolated onto the nodes of their parent elements in order to affect the rod dynamics. Forces are calculated from the negative energy gradient, as in Equation 3.8, where F_{elem} is now the attractive force acting on both parent elements, i and j, as the result of an interaction between a pair of VDW sites on separate rods. Interpolation proceeds identically to Equations 3.9 and 3.10.

3.5 Periodic boundary conditions

It is computationally intractable to represent the entire volume of a particle suspension at the molecular level. Molecular simulations must therefore be limited to a finite space or unit cell, often a cuboidal box with dimensions on the order of 10 nm to 1 μm.

Imposing periodic boundary conditions (PBCs) on the boundaries of this computational

domain provides a means to approximate the bulk properties of a much larger domain by introducing periodic images of the simulation box (Figure 3.6) [16, 140]. Since each rod now has an infinite set of images a modification is applied to the rod-rod displacement calculation (Equation 3.4) to capture interactions between rods in neighbouring periodic cells to find the nearest available periodic images - the minimum image criterion [152].

Considering the midpoints of a pair of rod elements, $\boldsymbol{r}_i^{mid}$, and $\boldsymbol{r}_j^{mid}$, the integer coordinates of the periodic cell nearest to element j are given by

$$t_\alpha = \texttt{floor}\left(\frac{r_{j,\alpha}^{mid} - r_{i,\alpha}^{mid} + 0.5h}{h}\right) \ , \tag{3.14}$$

where $\alpha = \{0, 1, 2\}$ is the vector index, $t_\alpha = \{-1, 0, 1\}$ in a given Cartesian axis, `floor` is a function that returns the largest integer less than or equal to its argument, and h is the width of the simulation box. The coordinates of the nearest periodic image of $\boldsymbol{r}_j$ are given by subtracting the cell dimensions,

$$\boldsymbol{r}_j' = \boldsymbol{r}_j - h\boldsymbol{t} \ , \tag{3.15}$$

where $\boldsymbol{r}_i^{mid}$ and $\boldsymbol{r}_j^{mid}$ are subsequently re-calculated so that periodicity is factored into the the crude cutoff radius (§3.3.1).

The 'wrap-around' feature of PBCs was also implemented: when the centroid of a rod crosses a simulation box boundary, it will re-appear at the opposite boundary with its velocity unchanged. If the dimension of the box is too small, it is possible that a rod could interact with two different periodic images of the same rod. However, this is easily avoided by preventing the user from running simulations of rods with contour lengths greater than $h/2$.

3.6 Testing

Unit and integration testing are crucial to effective software development by verifying functionality and improving maintainability [153]. Unit testing focusses on individual components in isolation, ensuring that each behaves as expected. Integration testing assesses how these units interact and run together as a complete system [154, 155].

Unit tests in KOBRA (and FFEA) mainly operate within a restricted 'test mode' of the software that allows routines to be called without needing to run a simulation. Conversely, integration tests usually involve running complete simulations and analysing the trajectories for specific outcomes. Both unit and integration tests may involve additional setup or analysis with FFEATools.

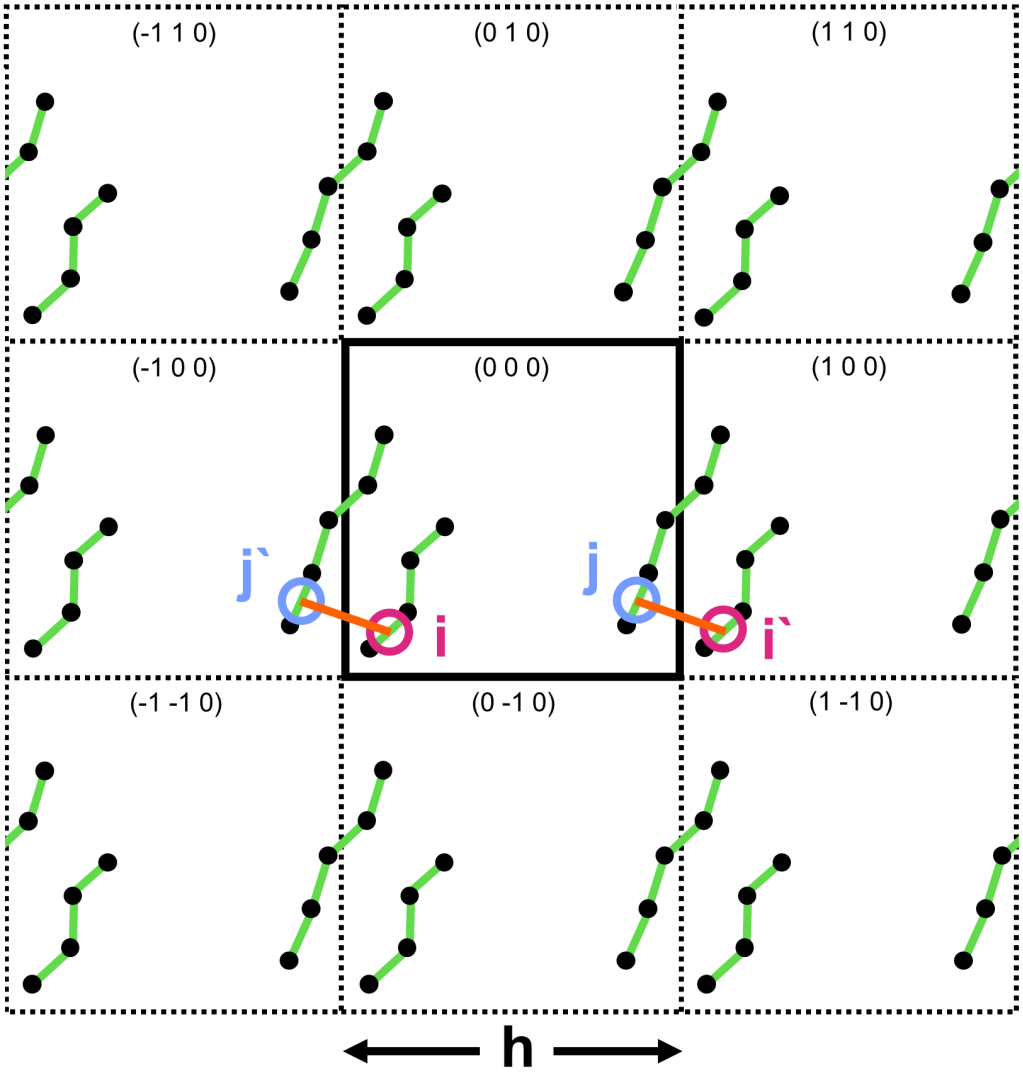


Figure 3.6: Schematic of rod periodic boundary conditions in 2D. The minimum image criterion ensures elements will only interact with the closest image. The distance between two elements in the central box, i and j, is in fact shorter when considering the distance over a periodic boundary: i interacts with j', and j interacts with i'. The periodic cell coordinates, $\boldsymbol{t}$, are shown at the top of each cell.

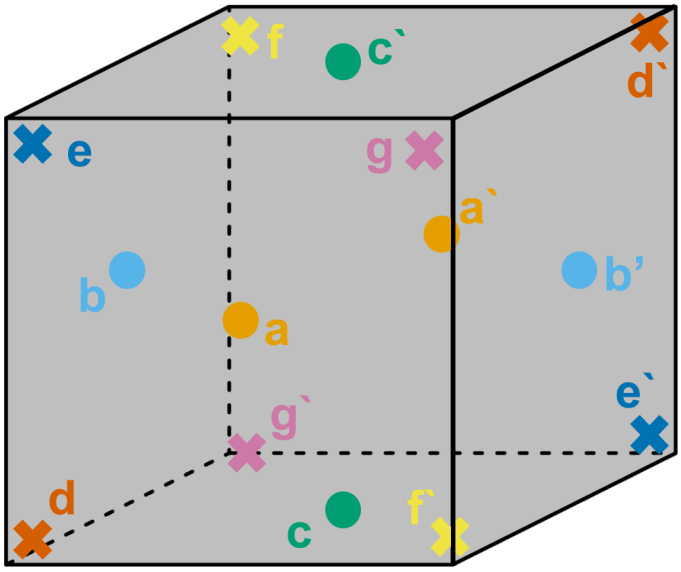


Figure 3.7: Pairs of points, $\boldsymbol{a}$ and $\boldsymbol{a}'$, $\boldsymbol{b}$ and $\boldsymbol{b}'$, etc. for vector distances calculated during the **nearest_image_pbc** test. Only the central simulation box is shown, but periodic boundary conditions are enabled.

The rest of this section comprises a brief description of each test that was written during the development of rod-rod interactions in KOBRA. Thermal noise is disabled.

3.6.1 Unit tests

nearest_image_pbc

In order to test that the periodic cell calculation of Equation 3.14 is correct, seven pairs of points, $\boldsymbol{a}$ and $\boldsymbol{a}'$, $\boldsymbol{b}$ and $\boldsymbol{b}'$, etc. are positioned around the simulation box (Figure 3.7) and the distance between them is calculated, each with a unique coordinate of the nearest periodic cell, $\boldsymbol{t}$. The test passes if every calculation returns the expected value for $\boldsymbol{t}$.

point_lies_within_rod_element

As its name suggests, only points in space that lie between the end-points of a rod element are considered to be part of that element. A single rod element is considered. Multiple arbitrary points on the centreline, $\boldsymbol{c}$, are generated from the parametric line equation, $\boldsymbol{c} = \boldsymbol{r} + t\boldsymbol{p}$, where t is a scalar value in the range $-1 \leq t \leq 2$. The test passes if $\boldsymbol{c}$ is adjusted to remain within the finite length of the rod centreline for all values of t.

rod_neighbour_list_construction

Four rod pair configurations are generated with FFEATools, with 4 nodes, radii $R = 1$ nm, length $L = 10$ nm, and both rods initially along the x-axis. Each second rod in a pair is translated in y by $\Delta y = \{0, 1, -1, 10\}$ nm and rotated in the x-y plane by $\Delta\theta =$

$\{0, 3, -3, 10\}^\circ$. The extent of overlap between both rods and the number of steric neighbours are known in advance for each configuration:

1. Parallel, full overlap. Two steric neighbours per element. $\Delta y = 0$ nm, $\Delta\theta = 0^\circ$.
2. Oblique, partial overlap. One steric neighbour per element. $\Delta y = 1$ nm, $\Delta\theta = 3^\circ$.
3. Oblique, partial overlap. One steric neighbour per element. $\Delta y = -1$ nm, $\Delta\theta = -3^\circ$.
4. Oblique, no overlap. Zero neighbours. $\Delta y = 10$ nm, $\Delta\theta = 10^\circ$.

The test passes if the expected total number of neighbours are calculated.

rod_steric_lj_potential

Two circles with equivalent radii, R, and centres at points $\boldsymbol{a}$ and $\boldsymbol{b}$, are moved apart in a straight line from a starting position of complete steric overlap, $s = -2R$, where s is the surface-surface distance. The increasing values of s are sufficient to pass through all three rod-rod interaction regimes: soft steric repulsion, interpolative, and attractive LJ 6-12. The test passes if the modified potential energy profile of Figure 3.1 is reproduced.

rod_vdw_site_placement

Create two rods of equal centreline length, straight and curved, in FFEATools, each with 5 uniformly-spaced VDW sites. The test passes if the KOBRA software can load the rod VDW sites and determine their positions in 3D space.

3.6.2 Integration tests

interactPBC_equalEnergy

The purpose of this test is to check the functioning of the steric repulson between overlapping rods, including the case where the overlap is between different periodic images. A KOBRA simulation of two pairs of rods is run for 5 steps with time step, $\Delta t = 10$ ps. Rods are straight and point along the z-axis, with 6 nodes, $R = 2.5$ nm, $L = 50$ nm, and box width $h = 200$ nm. Both rods in each pair are partially overlapping such that $s = -R$, on all elements. Pairs are positioned along the y-axis, with the first pair located at the box centre, and each rod of the second pair being located at opposite edges of the box, such that they interact over a periodic boundary. The test passes if the following crtieria are met at the end of the simulation:

- Central rods, a and b, repel towards the edges of the simulation box, away from the centre: $|\boldsymbol{c}_{ij}(t_{max})| - |\boldsymbol{c}_{ij}(0)| > 0$.
- Edge rods, c and d, repel towards the centre of the simulation box, away from the edges: $|\boldsymbol{c}_{ij}(t_{max})| - |\boldsymbol{c}_{ij}(0)| < 0$.

- Steric potential energies are equivalent across any given element, i, on all rods: $U_i^a = U_i^b = U_i^c = U_i^d$.
- All rods have some finite interaction energy.

pbc_wrap

If the centroid of a rod moves across a periodic boundary, it should reappear at the opposite side with its velocity unchanged. A KOBRA simulation of a straight rod with $M = 5$ elements, pointing in the z-axis, is run for 300 steps with $\Delta t = 1$ ps. A constant velocity, u_y, is applied in the positive y direction. Periodic boundary conditions are enabled. The velocity is chosen such that the time taken for the rod centroid to reach the edge of the simulation box, $y = h$, occurs at a known simulation time of $t = t_h$. The test passes if the y position of node 2 falls within the range $h/2 < y < h$ at $t = t_h - \Delta t$, and $0 < y < h/2$ at $t = t_h + \Delta t$.

stericInteract_fullySeparate

The most complex test presented in this chapter. In the absence of thermal noise, partially overlapped rod elements should repel each other until they are fully separated, at which point they stop moving. Input files are generated for many configurations of straight rod pairs (Figure 3.8), attempting to cover edge cases in the inter-element displacement. Most configurations begin in a partially overlapped state (group I), and some in a fully overlapped state with their centrelines directly intersecting, sometimes passing through the other side of a rod (group II).

A KOBRA simulation of every configuration is run for 250 steps with $\Delta t = 1$ ps. The test passes if group I rods are fully separated and no longer overlapping at the end of the simulation, and group II rods immediately raise an error at the first time step.

stericInteract_initForceSymmetric

For a pair of rod elements in a symmetric configuration, the corresponding steric repulsion force should also be symmetric. A pair of straight rods, a and b, with radii, R, and $M = 5$ elements, are generated in a perpendicular cross-shaped configuration (Figure 3.8B). The rods begin partially overlapped such that $s = -R$ at their central elements only, i.e. from $\boldsymbol{r}_2$ to $\boldsymbol{r}_3$. A KOBRA simulation is run for a single time step, $\Delta t = 0.1$ ps. Because s, the functional form of the steric potential, and the points of intersection along the element centrelines are all known, the expected node forces, $\boldsymbol{F}_i$, can be calculated. The test passes if the following criteria are met for both rods:

1. *Intra*-element forces are equal: $\boldsymbol{F}_2^a = \boldsymbol{F}_3^a$ and $\boldsymbol{F}_2^b = \boldsymbol{F}_3^b$.

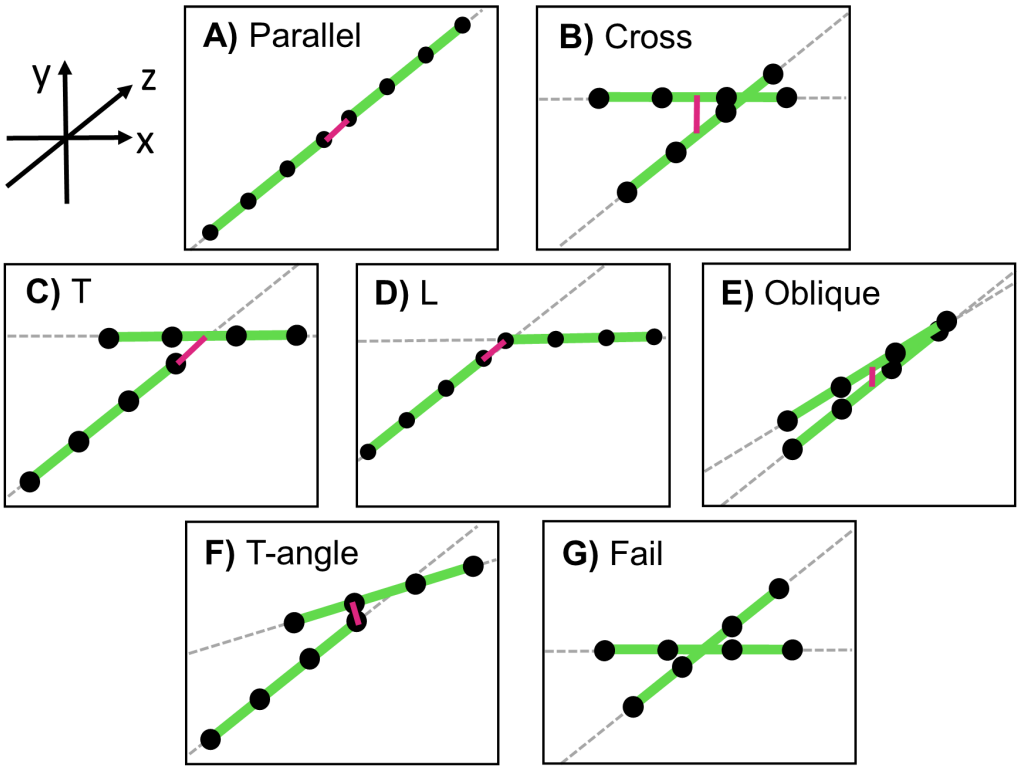


Figure 3.8: Schematic of unique configurations of rod pairs used in the **steric_fully_separate** integration test. There are many variations of the above, but the most representative ones are shown here. Magenta lines indicate separation between the rod centrelines, $s = -R$ in most cases. **A**) Parallel. Rods have identical orientation. **B**) Cross. Rods are perpendicular, with an interaction point at their centroids. **C**) T-shape. Rods are perpendicular, co-planar, and interact between the tip and centroid. **D**) L-shape. Rods are perpendicular, co-planar, and interact end-to-end. **E**) Oblique. Rods have small and large angular deviations, usually interacting at their centroids. **F**) T-angle. Similar to **C**, but the rods are oblique. **G**) Fail. Similar to the above configurations, but with the centrelines purposefully intersecting, such that an error is intentionally raised and the KOBRA simulation fails.

2. *Inter*-element forces have equal magnitudes and opposite signs: $\boldsymbol{F}_2^a = -\boldsymbol{F}_2^b$ and $\boldsymbol{F}_3^a = -\boldsymbol{F}_3^b$.
3. Force is only applied to nodes 2 and 3.
4. Forces match the expected values: $|\boldsymbol{F}_2^a + \boldsymbol{F}_3^a| = |\boldsymbol{F}_2^b + \boldsymbol{F}_3^b| = 2ks$, where k is the steric force constant.

steric_vdw_potential_minimum

In the absence of thermal noise, rod elements containing interacting VDW sites should eventually settle into the potential well minimum of the modified LJ profile (Figure 3.1).

Three KOBRA simulations are run for 10^4 steps with $\Delta t = 10$ ps. Each consists of a single pair of rods, a and b, in a perpendicular configuration similar to Figure 3.8B, but with different initial surface-surface distances: $s_0 = \{-R, 0.5s_{min}, 1.5s_{min}\}$, where s_{min} is the distance at which $U_{LJ} = -\epsilon$. Only nodes 2 and 3 interact.

The four pass criteria of **stericInteract_initForceSymmetric** are applied to the first step of each simulation, but #4 changes depending on s_0, and hence the initial potential regime that a and b occupy. A fifth criterion is introduced: in the final simulation step, a and b should occupy the minimum of the LJ potential, such that $s = s_{min}$ and $U^a = U^b \approx -\epsilon$. In the absence of thermal noise, the only dynamics are those driven by rod-rod interactions. The test passes if the five criteria are met across all simulations.

3.7 Performance and validation

Now that the rod-rod interaction algorithms have been defined and tested, its computational performance can be measured and its behaviour validated in a biological context. A main advantage of the CG representation of KOBRA is that it can run simulations with system sizes and time steps much larger than atomistic MD, so this section will focus on these aspects of the software.

A basic interacting rod model of fibrinogen is presented and the run time of KOBRA simulations containing large numbers of rods is quantified. Following this, a model of a fibrin protofibril is assembled from nineteen fibrinogen rods, and KOBRA simulations of individual protofibrils are run to estimate an appropriate interaction energy, ϵ, and bending stiffness, B.

3.7.1 Rudimentary rod model of fibrinogen

The rod model of fibrinogen (Figure 3.9) was created with $M = 4$ elements, physical dimensions $L = 45$ nm and $R = 2.5$ nm, with four VDW sites positioned according to the

approximate locations of A:a knob-hole interactions [61]. The two distal sites represent a-holes in the D nodules ($n_{site} = 0$, $L_{site} = 0, 1$), and the two central sites represent A-knobs in the E nodule ($n_{site} = 1$, $L_{site} = 0.4375, 0.5625$).

The LJ parameters were chosen as $\sigma_{01} = \sigma_{10} = 0.89$ nm and $\epsilon_{01} = \epsilon_{10} = 8.5$ k_BT, determined from a peak rupture force of 35 pN observed in single-molecule pulling experiments of A:a knob-hole interactions [110]. This is a significant approximation, as knob-hole interactions have been demonstrated to exhibit catch-slip behaviour, whereby increasing the rupture force on a bond causes it to strengthen. The LJ potential does not capture such behaviour by itself, and may require ϵ to be dynamic and adjustable as a simulation is running. This type of interaction is not yet represented in KOBRA or FFEA.

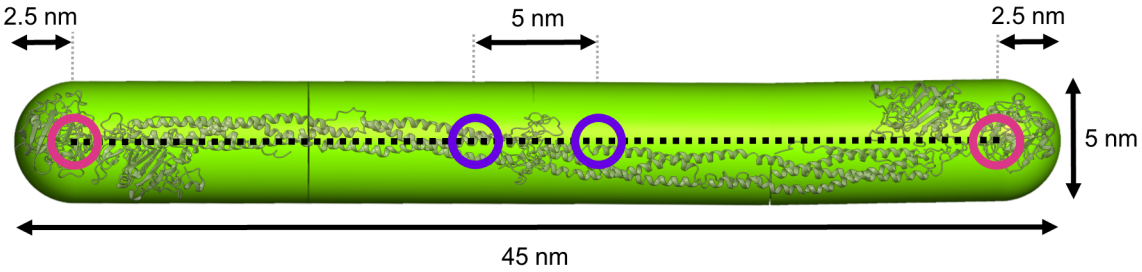


Figure 3.9: KOBRA rod model of fibrinogen, with an overlay of the crystal structure (PDB: 3GHG [95]). The rod has four elements and the steric volume is shown in green. The VDW sites are shown as circles, with A-knobs in the centre (purple) and a-holes at the ends (magenta).

Values for the material constants were chosen to approximate a generic protein (Table 3.1). In future it should be possible to extract these parameters from atomistic MD simulations, but this was not attempted here due to time constraints.

Parameter	Symbol	Value
Stretch	κ	3.5×10^{-11} N
Bend	B	3.5×10^{-29} m^4 Pa
Twist	β	5×10^{-29} N m^2

Table 3.1: Rod material parameters for KOBRA benchmark simulations. Values are homogeneous across the entire rod. The bending stiffness, B, is the value along the diagonal of the isotropic bending stiffness matrix, $\boldsymbol{B}$.

3.7.2 Closely packed configurations of rods

To capture a 'worst case' scenario of the maximum possible run time that a user might encounter with a KOBRA simulation, it is desirable to generate spatial configurations of rods that maximise the number of interactions.

For a solution of fibrinogen at a physiological molar concentration, $c = 6$ μM, equivalent to the experiments of Chapter 4, the number density of molecules is $n = 3.61 \times 10^{21}$ m^{-3} (§B.1). For a typical KOBRA simulation box width, $h = 500$ nm, this yields $nh^3 = 450$

rods. This solution is in the dilute concentration regime, so molecules do not tend to overlap and steric interactions are infrequent such that they do not contribute significantly to the overall dynamics [54]. Consequently, a solution of this concentration is not particularly useful for measuring the performance of the interaction algorithms.

To maximise the number of interactions, rods were generated in closely packed configurations (Figure 3.10). Rods were aligned in the z-axis and positioned such that every rod, excluding those on the periphery, was partially overlapped by 1 nm with its neighbours in the x, y and z axes. Thermal noise was enabled to encourage interactions at every time step, and to be further representative of a typical KOBRA simulation. Rod centroids were calculated such that their displacement from the origin was minimised.

Whilst this is not a biologically realistic environment for fibrinogen molecules in solution, fibrin fibres have been proposed to adopt a partially crystalline nanostructure [105, 156], although the half-staggered arrangement of monomers in the longitudinal axis, helical fibre structure, and physiological protofibril density, are not represented here.

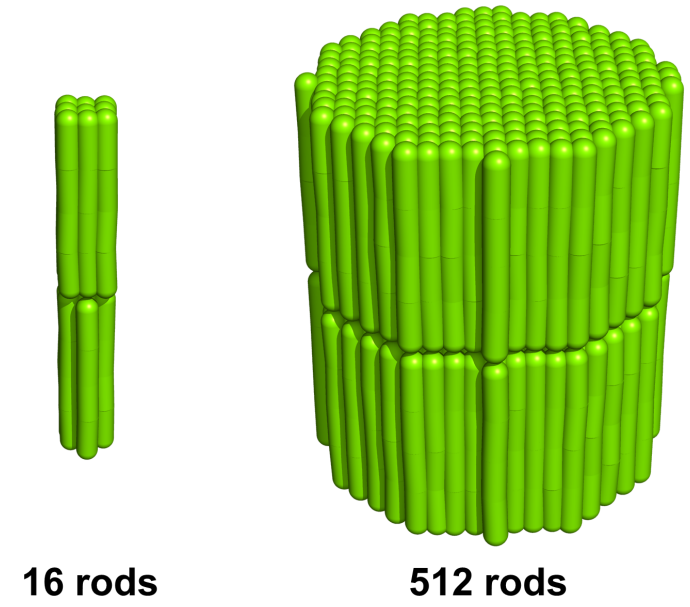


Figure 3.10: Two tightly packed configurations of 16 and 512 rods, as used in scaling simulations. Approximate steric volumes are shown in green. Visualised in PyMOL [157].

3.7.3 Parallel scalability

In order to use KOBRA to simulate large systems within a reasonable timeframe it is important that it runs efficiently across multiple CPU cores. Benchmarking the performance of the updated KOBRA software will be useful for developers that wish to target areas of the codebase for improvement, and for users in deciding what systems they wish to simulate. A commonly used metric for judging the parallel performance of programs *scalability*.

The *strong* scaling of a program is determined by increasing the number of cores, n, for a fixed system size, e.g. a fixed number of rod elements. In an ideal case, the run time, T, scales linearly with n, and the parallel speedup, S, is given by

$$S = \frac{T_1}{T_n} \ , \tag{3.16}$$

where the run time with n cores, T_n, is faster by a factor of n than that with 1 core, T_1. A linear scaling of S with n is rarely observed due to numerous factors that impact run time such as communication overhead between cores, load balancing, memory bottlenecks, etc. [158].

Amdahl's law proposes a more realistic expression for the maximum speedup, and is given by

$$S \leq \frac{1}{t_s + \frac{t_p}{n}} \ , \tag{3.17}$$

where t_s is the proportion of run time spent in serial operations (and hence cannot be parallelised), and t_p is the proportion spent in parallel operations. In the limit $n \to \infty$, Equation 3.17 reduces to $1/t_s$, which suggests the maximum speedup attainable by a program is limited by its serial run time [158, 159]. The *weak* scaling of a program involves increasing the system size proportionally with n such that the system size per core is constant [160], but this was not quantified.

3.7.4 KOBRA simulations of packed fibrinogen

Three sets of KOBRA simulations were run using packed configurations of the fibrinogen rod model shown in Figure 3.9. Each simulation was run three times with a different rod-rod interaction mode enabled: none (control), steric only, and steric with VDW (or stericVDW). Each set aims to quantify a different aspect of computational performance:

1. **Interaction behaviour:** 3 simulations of $N = 64$ rods run on $n = 4$ cores for a simulation time of $t = 10$ µs. For each interaction mode, the dynamics of the packed rods should be qualitatively different when visualised.

2. **Strong scaling:** 21 simulations of $N = 128$ rods with $n = 1, 2, 4, ...64$ cores for a simulation time of $t = 1$ µs. Quantifies the parallel performance of the software.

3. **Simulation time per day:** 30 simulations of $N = 1, 2, 4, ...512$ rods with $n = 4$ cores for a simulation time of $t = 1$ µs. Quantifies the speed of the software, allowing for comparison with other types of molecular simulation.

All simulations were run on an Intel Xeon Gold 6240R CPU. The time step, $\Delta t = 100$ ps, was chosen to be below the maximum stable value imposed by the steric interactions

(§B.2), because the numerical stability of the internal elastic energy algorithm at large time steps has not been extensively tested [127]. The 3GHG structure of fibrinogen was estimated to contain $\sim$ 16,000 atoms [95], so the largest system of $N = 512$ rods thus represents $\sim 8 \times 10^6$ atoms.

Snapshots were taken at $t = 10$ µs of the 64 rod simulations (Figure 3.11) to capture the general evolution of each system from its initial packed state. The control, steric, and stericVDW simulations had run times of 11.8, 49.1, 132 minutes, respectively.

These images qualitatively suggest that the interaction algorithms are producing the expected behaviour, in the absence of quantitative data such as the mean squared displacement.

The control rods are able to pass through each other, so their dynamics are governed solely by diffusion and the internal response to elastic deformation. The steric rods appear to have moved a greater distance relative to the control, forming a more rarefied configuration. Excluded volume interactions are known to cause flexible polymers to swell and exhibit enhanced diffusivity [54].

In contrast, the stericVDW rods remain in close proximity; the interaction energy of $\epsilon = 8.5$ k_BT is sufficiently larger than diffusion ($\sim k_BT$) so as to inhibit their separation. The structure formed by the rods is more akin to a cluster than a fibre, despite the initial configuration (Figure 3.10), but this is likely due to the rudimentary nature of the rod model. It could be that the generic material parameters (Table 3.1) caused each rod to exhibit greater conformational flexibility than actual fibrinogen molecules, which would hinder the formation of ordered, fibrous structures [54].

Figure 3.11: Final frames of three 10 µs simulations containing 64 rods with different interactions enabled: none (control), steric only, and steric with VDW (or stericVDW). Approximate steric volumes are shown in green. Visualised in PyMOL [157].

The strong scaling (Figure 3.12) of non-interacting rods peaks with a speedup of $S \sim 2.4$

at $n = 4$ and $n = 8$ cores, however, interacting rods exhibit close to serial performance, for both steric and stericVDW. This was entirely expected, as little time was spent attempting to parallelise both interaction algorithms. It would be simple to investigate this further with the use of a profiling tool such as `gprof`, to determine which routines occupy the largest proportion of simulation time.

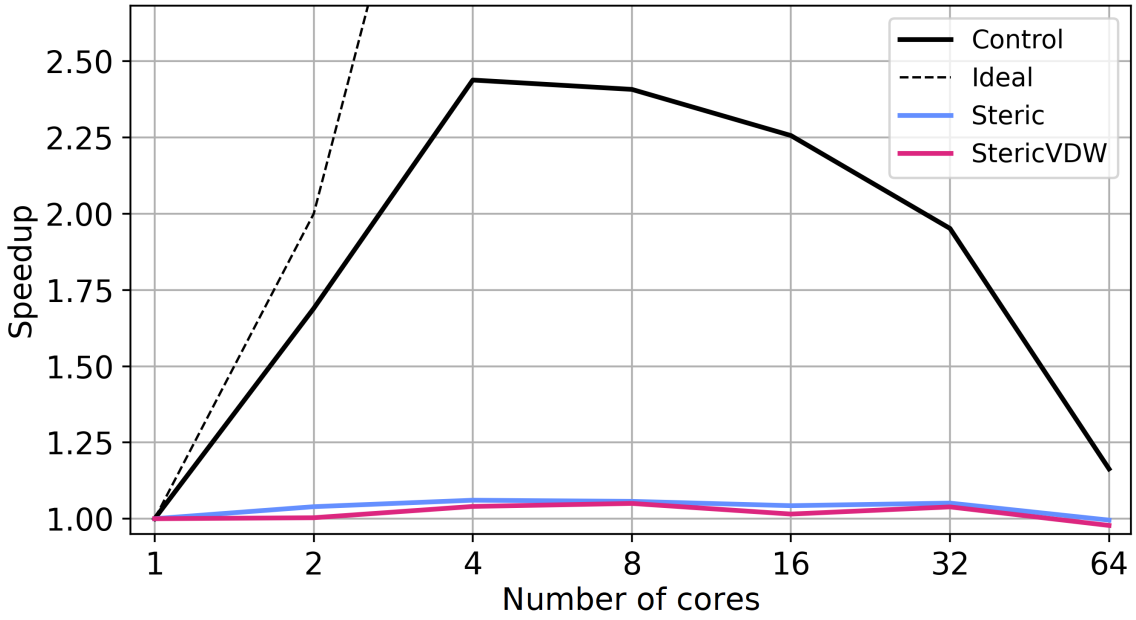


Figure 3.12: Strong scaling of $t = 1$ µs simulations containing $N = 128$ rods with different interactions enabled: none (solid black), steric only (blue), and steric with VDW (magenta).

Despite the rod-rod interactions exhibiting serial runtimes, Figure 3.13 illustrates that KOBRA is nonetheless able to access time scales that rival those obtainable by atomistic MD.

A stericVDW simulation of $N = 64$ rods ($\sim 10^6$ atoms) achieves simulation speeds of around 100 µs day^{-1} on a professional-grade CPU, with minimal effort taken to achieve good parallel performance. In contrast, atomistic simulations must be run on specialised supercomputer hardware to achieve similar speeds with million-atom systems [161, 162]. The $N = 512$ rod system ($\sim 8 \times 10^6$ atoms), had a simulation speed of 1.7 µs day^{-1}.

StericVDW runs twice as slow as steric for $N > 16$ rods: 1000 versus 500 µs day^{-1} with $N = 32$ rods, 80 versus 40 µs day^{-1} with $N = 128$ rods, and so on. For $N < 8$ rods both modes exhibit similar speeds, possibly because the rods were repelled before VDW interactions could be established.

With $N = 1$ rod the simulation speeds of all three interaction modes are equivalent at around 4×10^4 µs day^{-1}. Although trivial to run, it is good to confirm this because it shows that the interaction algorithms do not adversely affect the run time in other parts of the software.

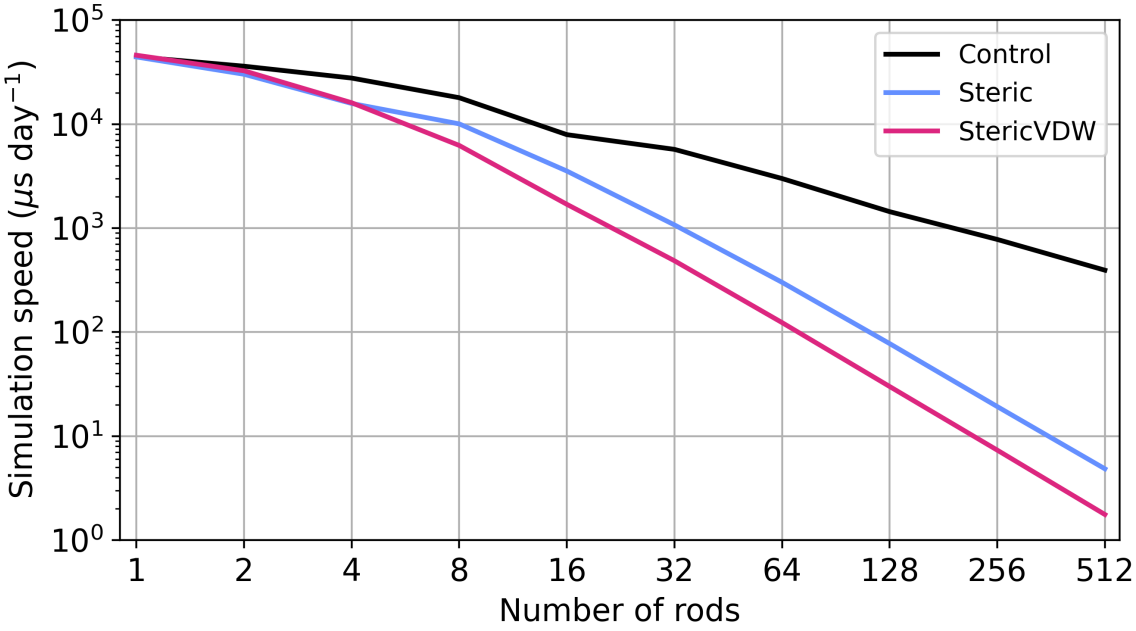


Figure 3.13: Simulation speed, measured in microseconds per day, of $t = 1$ µs simulations with different interactions enabled: none (solid black), steric only (blue), and steric with VDW (magenta).

It is difficult to directly compare the performance of KOBRA to other molecular simulations of fibrinogen, because most studies do not include the run time. Atomistic MD studies can be limited in their ability to sample the dynamics of a single fibrinogen molecule due to its large size, with relatively short simulation times of hundreds of nanoseconds and time steps on the order of femtoseconds [163, 164].

The CG simulations of a 19-monomer fibrin protofibril by Zhmurov et al. [106] used one bead to represent each amino acid, allowing them to compute the conformational dynamics of such a large structure. With software optimised to run on GPUs, they ran multiple simulations for 10 ms, but did not publish the time step.

Yesudasan et al. [148] published a CG model where each fibrinogen molecule was represented with nine beads. Intermolecular interactions were permitted between specific types of beads using a soft attractive potential; once the beads came into close proximity, they formed permanent harmonic bonds. After populating a simulation box with a physiological concentration of monomers, they were able to capture protofibril formation, lateral aggregation and fibre branching in only 900 ns of simulation time, with a 450 fs time step.

3.7.5 Rod model of a fibrin protofibril

A 19-monomer fibrin protofibril (Figure 3.14) was created by positioning fibrinogen rods, as described in §3.7.1, in a half-staggered formation that has been observed by high res-

olution AFM experiments (Figure 1.6C). Each rod was overlapped onto its neighbours by 2.5 nm in order to bring the VDW interaction sites, representing A:a knob-hole interactions, into close proximity with each other (Figure 3.14B), such that monomers would interact immediately at the start of a KOBRA simulation. This grouping of VDW sites represents the D-E-D interface, where three fibrinogen nodules interact during polymerisation (Figure 1.6C). The overall structure of the protofibril rod model is held together solely by interactions at the D-E-D interfaces.

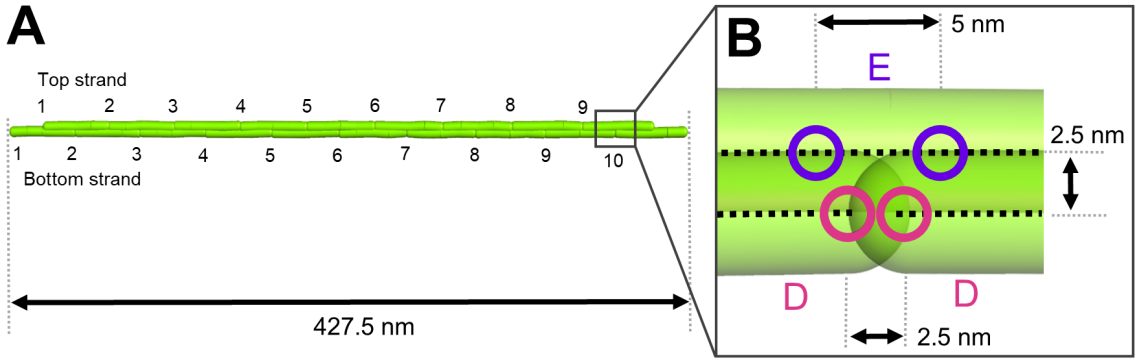


Figure 3.14: KOBRA rod model of a fibrin protofibril, assembled from nineteen fibrinogen rods. **A)** The full length of the protofibril. The numbers of component rods in the top and bottom strands are labelled at their starting positions. Visualised in PyMOL [157]. **B)** Close-up of a D-E-D interface, composed of four VDW interaction sites representing fibrinogen A-knobs in the E nodule (purple circles) and a-holes in the D nodules (magenta circles).

3.7.6 KOBRA simulations of a fibrin protofibril

If KOBRA simulations can preserve the structure of a protofibril throughout an entire simulation, it would serve as a straightforward validation of the interaction algorithms in a biologically relevant setting. An initial simulation of the 19-mer model in Figure 3.14 was run for a duration of $t_{max} = 5$ ms, using the same rod material parameters as previous fibrinogen simulations (Table 3.1) and $\Delta t = 100$ ps.

Six snapshots of the simulation up to $t = 2$ ms are shown in Figure 3.15. The protofibril breaks into two sections towards the start of the simulation ($t = 10$ µs), whereupon it coils up to form two compact, globular structures ($20 < t < 500$ µs) that persist for the remainder of the simulation ($t > 500$ µs). Clustering was also observed in previous KOBRA simulations of dense systems of interacting rods (Figure 3.11). Such behaviour is not typically observed during the polymerisation of fibrin (Figure 1.6).

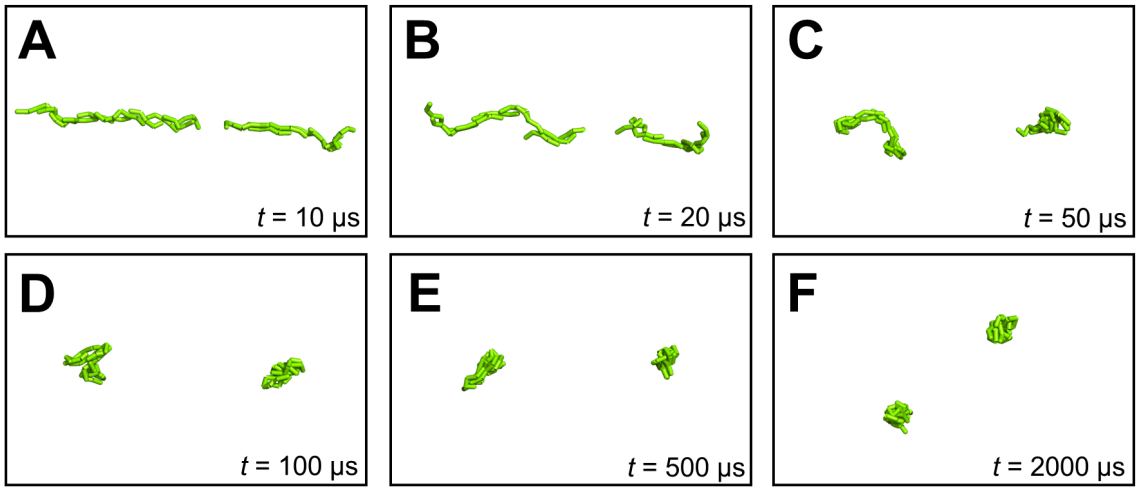


Figure 3.15: Snapshots of the initial KOBRA protofibril simulation, with $\epsilon = 8.5$ k_BT and $B = 3.5 \times 10^{-29}$ m^4 Pa. **A**) Protofibril breaks in two. **B-C**) Coiling and self-interaction. **D-F**) Cluster formation. Visualised in PyMOL [157].

The clustering seen in Figure 3.15 involves sections of protofibril bending onto themselves and self-interacting, suggesting that the component rods may be too flexible. Furthermore, the attractive rod-rod interactions described in §3.4 are able to occur from any direction, with no upper limit on the number of simultaneous interactions at a given site. The combination of these two factors could be the cause of cluster formation by allowing interactions between many rods within a small region of space.

The interaction energy, $\epsilon = 8.5$ k_BT, was derived from experimental data concerning A:a knob-hole interactions [110], but the material parameters were obtained from educated guesses instead of being more accurately parameterised from atomistic MD trajectories. Time constraints did not permit the construction and simulation of a complete atomistic model of fibrinogen, so seven KOBRA simulations were run as part of a parameter sweep in which ϵ and the bending stiffness, B, were varied through the values shown in Table 3.2, with $\Delta t = 100$ ps and $t_{max} = 100$ μs.

The objective remained the same as before: to determine whether a protofibril structure could, at the very least, maintain its initial configuration throughout an entire KOBRA simulation. Snapshots at simulation times of 20, 50 and 100 μs are shown in Figure 3.16 for varying ϵ and Figure 3.17 for varying B. The energy sweep simulations had a fixed bending stiffness of $B = 3.5 \times 10^{-29}$ m^4 Pa, whereas the bending sweep had a fixed interaction energy of $\epsilon = 10$ k_BT.

ϵ (k_BT)	B (10^{-28} m^4 Pa)	t_{max} (µs)	Note
8.5	0.35	5000	Initial attempt
0	0.35	100	Energy sweep 1
5	0.35	100	Energy sweep 2
10	0.35	100	Energy sweep 3
15	0.35	100	Energy sweep 4
10	1	100	Bending sweep 1
10	5	100	Bending sweep 2
10	10	100	Bending sweep 3

Table 3.2: Parameters for KOBRA protofibril simulations: interaction energy, ϵ, bending stiffness, B, and duration, t_{max}. The quoted bending stiffness is the value along the diagonal of the isotropic bending stiffness matrix, $\boldsymbol{B}$.

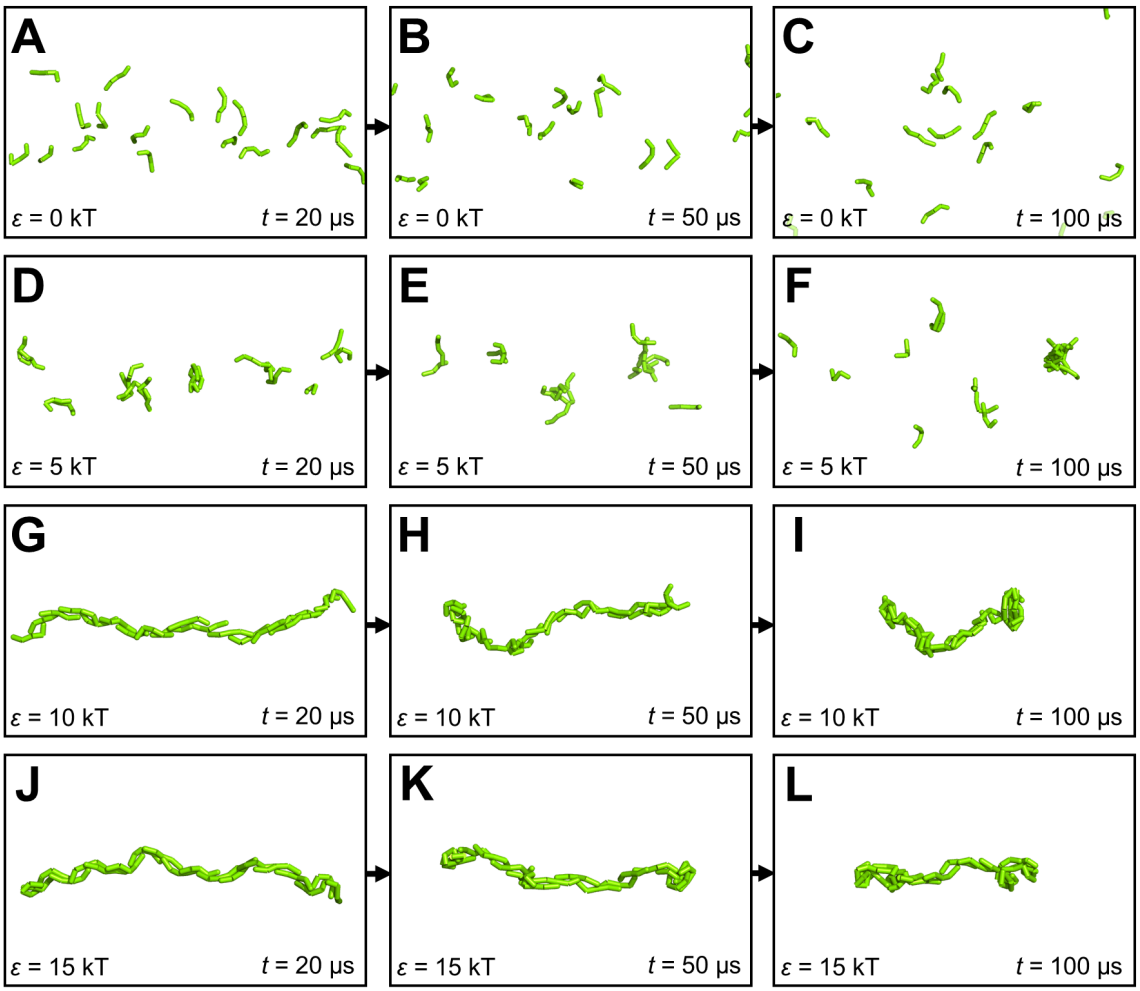


Figure 3.16: Snapshots of four KOBRA protofibril simulations at times of 20, 50 and 100 µs, that are part of a parameter sweep over the interaction energy. **A-C**) $\epsilon = 0$ k_BT, immediate diffusion of component rods. **D-F**) $\epsilon = 5$ k_BT, diffusion of rods into some small clusters. **G-I**) $\epsilon = 10$ k_BT, intact protofibril with some coiling at both tips. **J-L**) $\epsilon = 15$ k_BT, intact protofibril with some coiling at both tips. Visualised in PyMOL [157].

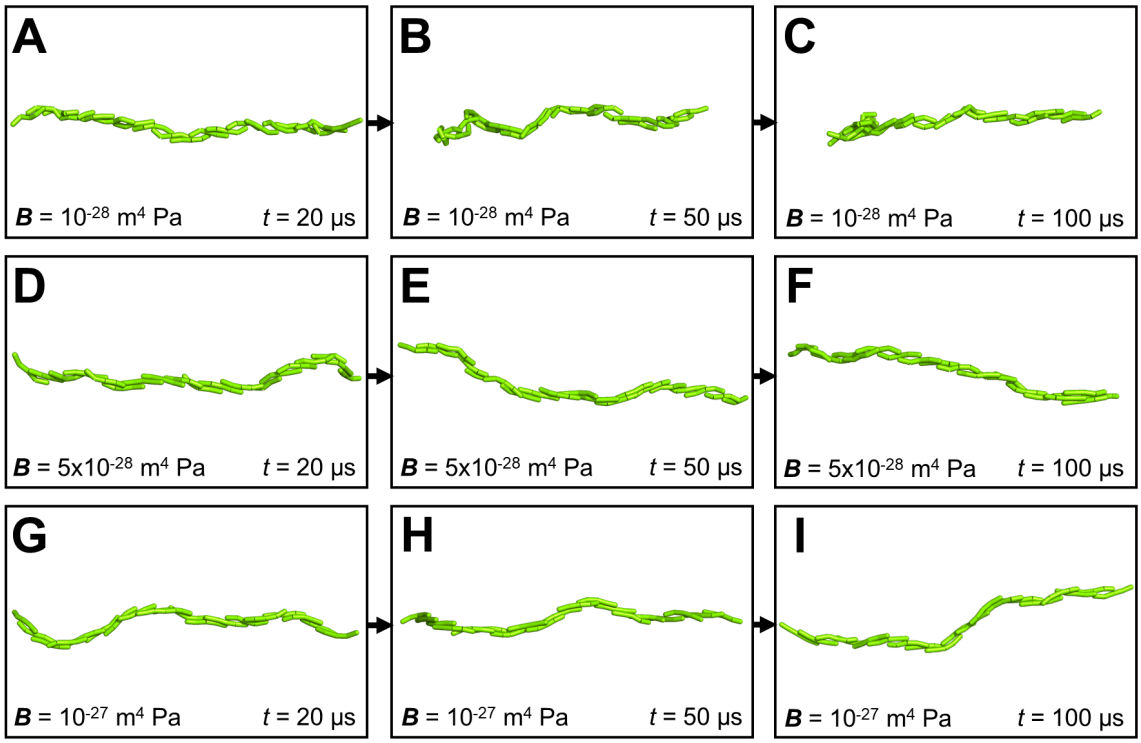


Figure 3.17: Snapshots of three KOBRA protofibril simulations at times of 20, 50 and 100 µs, that are part of a parameter sweep over the bending stiffness. **A-C**) $B = 10^{-28}$ m^4 Pa, intact protofibril with some coiling at the left tip. **D-F**) $B = 5 \times 10^{-28}$ m^4 Pa, intact protofibril with marginal coiling at the right tip. **G-I**) $B = 10^{-27}$ m^4 Pa, intact protofibril with no discernible coiling. Visualised in PyMOL [157].

In Figure 3.16A-C the attractive interactions were disabled as a control, causing the rods to diffuse outwards in an expected manner, previously observed in Figure 3.11. Figure 3.16D-F with $\epsilon = 5$ k$_B$T shows the formation of some small clusters, but the protofibril does not maintain its shape at all. Figure 3.16G-I with $\epsilon = 10$ k$_B$T shows an intact protofibril at $t = 20$ µs, but for $t > 50$ µs the tips of the fibril begin to self-interact and its structure is noticeably more compact at $t = 100$ µs. Similar behaviour is seen in Figure 3.16J-L with $\epsilon = 15$ k$_B$T. Modifying ϵ alone is insufficient to model a protofibril if the component rods are too flexible.

Now considering the bending stiffness parameter sweep, Figure 3.17A-C with $B = 10^{-28}$ m^4 Pa shows an intact protofibril, with a small amount of coiling visible at the left tip for $t > 50$ µs, albeit to a lesser extent than in Figure 3.16G-L. Figure 3.17D-F also shows an intact structure, with only a marginal amount of coiling visible at the right tip for $t = 100$ µs (three rods are stacked on top of each other). No coiling is apparent for the full duration of the simulation in Figure 3.17G-I with $B = 10^{-27}$ m^4 Pa.

The snapshots in Figures 3.16 and 3.17 provide qualitative evidence that KOBRA is capable of simulating an intact fibrin protofibril, with a potential candidate for longer simulations having $\epsilon = 10$ k$_B$T and $B = 10^{-27}$ m^4 Pa. As a quantitative counterpart, the

mean squared displacement (MSD) and mean end-to-end distance, d, were plotted for each trajectory. The mean squared displacement (MSD) at a specific lag time, τ, is given by

$$\mathrm{MSD}(\tau) = \left\langle ||\boldsymbol{r}(t+\tau) - \boldsymbol{r}(t)||^2 \right\rangle \ , \tag{3.18}$$

where $\langle\rangle$ denotes an average over all nodes in the system and all possible starting times [140, 165]. The MSD is plotted against lag time in Figure 3.18A. The usage of a lag time to step through trajectories means that a greater number of samples can be used in the MSD calculation. The mean end-to-end distance, d, of each component fibrinogen rod at a given simulation time, t, is given by

$$d(t) = \left\langle ||\boldsymbol{r}_4(t) - \boldsymbol{r}_0(t)|| \right\rangle \ , \tag{3.19}$$

where $\langle\rangle$ now denotes an average over all rods in the system. Due to $\boldsymbol{r}_4$ and $\boldsymbol{r}_0$ existing at either end of the rod centreline, d is about 5 nm short of the 'true' end-to-end distance because it excludes the steric radius. The mean end-to-end distance is plotted against simulation time in Figure 3.18B.

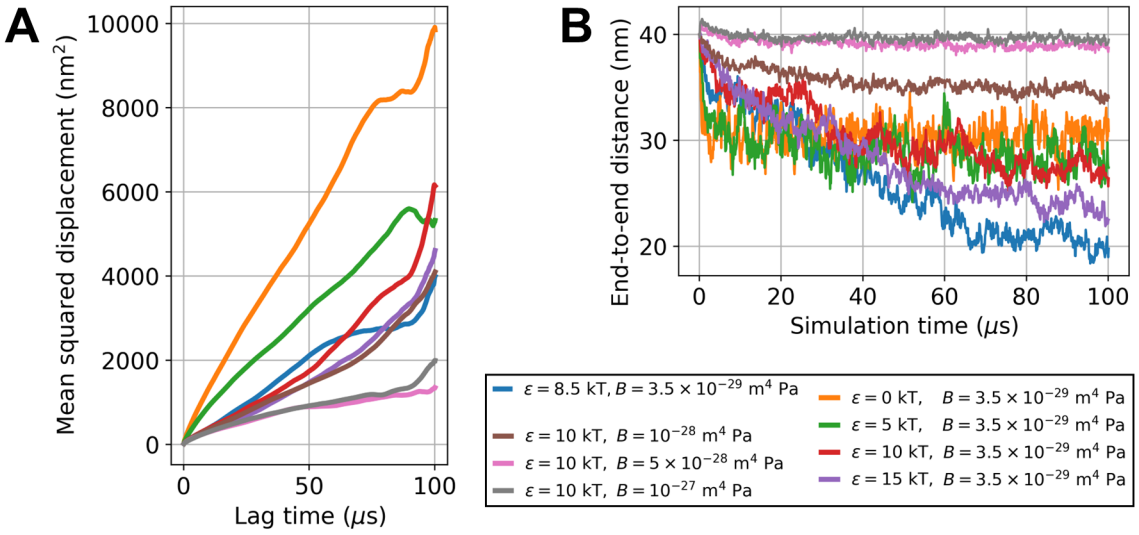


Figure 3.18: Average quantities of protofibril simulations plotted against time. **A**) Mean squared displacement, averaged over all nodes and starting times. **B**) Mean end-to-end distance, averaged over all rods.

Figure 3.18 is broadly in agreement with the behaviour shown in Figures 3.15, 3.16 and 3.17. The control simulation with zero attractive interactions exhibits the largest MSD of $\sim 10^4$ nm^2 (Figure 3.18A), due to its diffusion being obstructed solely by steric interactions with other rods. At the lower range are the stiffest protofibrils with $B = 10^{-27}$ and 5×10^{-28} m^4 Pa, with MSDs about one-fifth smaller than the control at $\tau = 100$ μs. The

remaining simulations exhibit a range of $4000 <$ MSD < 6000 nm^2.

The value of d for the two stiffest protofibrils oscillates around the starting value, $d = 40$ nm, for the full duration of their trajectories (Figure 3.18B). The third-stiffest protofibril with $B = 10^{-28}$ m^4 Pa exhibited some compacting in Figure 3.17B-C, which is also visible in Figure 3.18B due to a reduction of d to 35 nm at $t = 100$ µs. The most drastic compacting is visible in the initial simulations, with $\epsilon = 8.5$ k$_B$T and $B = 3.5 \times 10^{-29}$ m^4 Pa, as its end-to-end distance is halved at $t = 100$ µs.

The simple model of the component fibrinogen rods means that each node in the system has the same value of B, including the parameter sweep over B in Figure 3.17. The bending stiffness is therefore isotropic across each rod and the entire protofibril. In reality, the flexibility of fibrinogen is heterogeneous across the molecule, with large bending motions being expressed through molecular hinges in both coiled-coil regions [163]. A user that wished to parameterise a KOBRA rod model of fibrinogen from atomistic MD would need to ensure that the rod has a sufficient number of nodes to resolve both hinges [127].

The αC regions of fibrinogen have not been studied in this chapter, yet they are thought to be critical to the lateral aggregation of protofibrils during the polymerisation of fibrin [102]. This would, however, require additional physics to be implemented in the KOBRA software in order to represent the intrinsically disordered αC-connector that precedes the globular αC-domain (Figure 1.5).

3.8 Summary

An updated version of the KOBRA software has been presented, including algorithms and tests for repulsive steric interaction, attractive VDW interactions, and periodic boundary conditions. A simplified version of the program logic is shown in the steps below:

- Initialise system
- Time loop
  - Pairwise rod loop
    - Compute distances between elements
    - Update steric neighbour lists of both rods
    - Compute distances between VDW sites
    - Update VDW neighbour lists of both rods
  - Rod loop
    - Compute elastic energies and forces of all nodes
    - Element loop

            - Access steric neighbour list
            - Compute energy and force sum from interactions with steric neighbours
            - Interpolate force onto both nodes of element
        * VDW site loop
            - Access VDW neighbour list
            - Compute energy and force sum from interactions with VDW neighbours
            - Interpolate force onto both nodes of parent element
        * Node loop
            - Sum elastic, thermal and interaction forces on node
            - Update node position
        * Check for PBC wrap
- Finish

The parallel scalability of the program is poor, with the speedup of both steric and VDW interactions barely increasing above serial performance, however, this was expected, as minimal effort was devoted to parallelising the algorithms; the primary focus during development was on incorporating the additional physics and ensuring its implementation was correct. An additional positive outcome of this approach was that premature optimisation was largely avoided, and there are likely large gains in speedup to be made in the future without requiring a considerable time investment.

The time step size accessible by KOBRA simulations remains relatively large at $\Delta t = 100$ ps for the fibrinogen rod model.

Snapshots were taken of the longest simulations, with $t = 10$ μs and $N = 64$ rods, which served as qualitative evidence that the interaction algorithms were producing the expected behaviour. Where control rods diffused around the starting configuration and passed through each other, steric rods moved a greater distance due to their initial repulsion. The VDW rods immediately formed a cluster and did not move apart, indicating that the 8.5 k_BT knob-hole interactions were sufficient to overcome the energies of diffusion.

Simulation speeds, recorded in μs day^{-1}, were quantified for ten systems containing 1 to 512 rods. These demonstrated that the updated KOBRA software is highly capable of simulating the length and time scales that characterise the biological mesoscale, despite the increased computational load required to model biomolecular interactions. For a system of $N = 64$ rods, the equivalent of $\sim 10^6$ atoms, VDW interactions could be run at speeds of

100 μs day^{-1}. This speed decreases as N is increased, with the largest system of $N = 512$ rods having run at a speed of 1.7 μs day^{-1}, although this represents a huge system of $\sim 8 \times 10^6$ atoms.

It was demonstrated that a larger rod complex can be successfully assembled from smaller, interacting rods such that it maintains its overall structure throughout a 100 μs simulation. In lieu of a full parameterisation from atomistic MD trajectories, the KOBRA simulations and parameter sweep of a fibrin protofibril suggested that an interaction energy, $\epsilon = 10$ k_BT, and bending stiffness, $B = 10^{-27}$ m^4 Pa, are good parameter estimates for further simulations of protofibrils that may be done in the future. Snapshots of all eight simulations were in good agreement with the mean squared displacement and mean end-to-end distance of the component fibrinogen rods.

Chapter 4

Flow-induced structural remodelling of fibrinogen quantified by site-specific labelling

4.1 Introduction

4.1.1 Hydrodynamic flow can perturb the structure of proteins

Hydrodynamic flow is ubiqutious in biological systems. However, the relationships between types of flow, protein sequence, structure and stability, and the potential pathological consequences are not fully understood [166, 167]. The presence of shear flow during fibrin network polymerisation has a direct effect on the structure and stability of the resulting blood clot [109, 117, 123, 168]. The unintended aggregation of proteins[1] caused by hydrodynamic forces poses a significant hindrance in biopharmaceutical manufacturing, requiring additional money and time to eliminate them from the batch. Such measures are crucial to prevent adverse side effects that may occur in the resulting drug [169].

Hydrodynamic drag forces in bulk fluid flow can induce proteins to undergo partial mechanical unfolding, which is thought to be a key stage in the pathway of flow-induced aggregation (Figure 4.1). Complete unfolding of a protein isn't necessary for insoluble aggregate formation; sufficient exposure of aggregation-prone regions to the solvent, such as previously sequestered hydrophobic cores, allows them to self-associate via protein-protein interactions [166]. Surface-mediated effects are also thought to be involved in this pathway, but are not discussed in this thesis [170–172]. A consensus has yet to emerge in the literature as to the precise molecular mechanism by which proteins aggregate under flow [167].

[1] Not to be confused with enzyme-catalysed polymerisation, like that which forms the fibrin network.

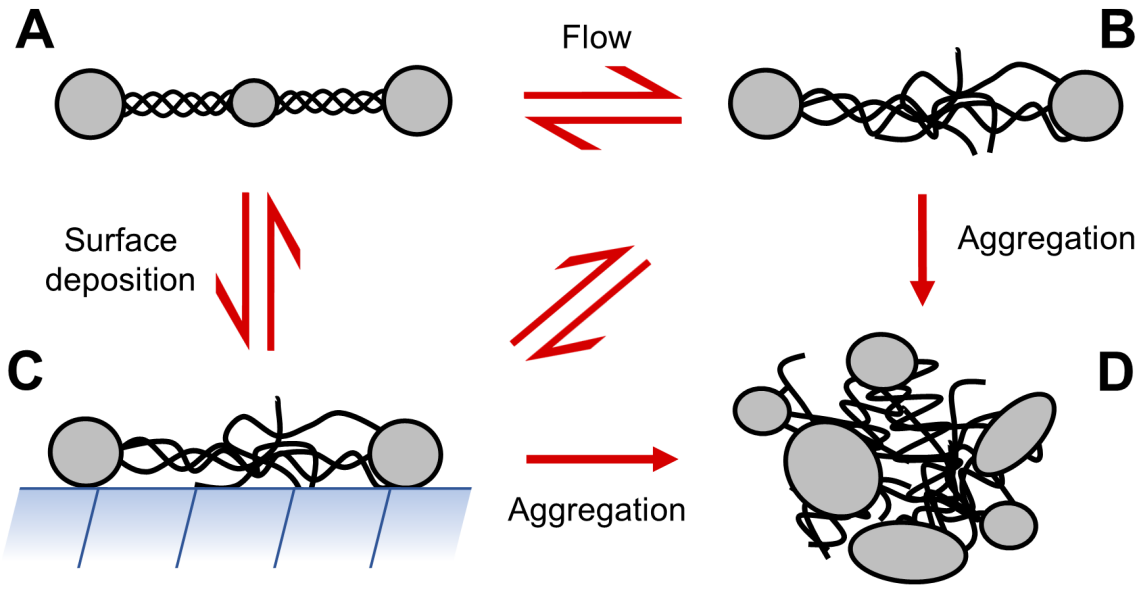


Figure 4.1: Proposed kinetic model of the flow-induced aggregation pathway. **A**) Soluble, folded protein in bulk fluid. **B**) Partial unfolding due to hydrodynamic flow. **C**) Partial unfolding due to deposition at a fluid-surface interface. **D**) Interaction of perturbed proteins to form insoluble aggregates. Proteins may re-fold or release from a surface back into the bulk fluid, but aggregation is irreversible.

The Brockwell research group devised a reciprocating extensional flow device (EFD) to study the aggregation propensity of small, globular proteins and monoclonal antibodies, with a particular focus on replicating stresses encountered during industrial processes [166, 170, 173, 174]. More recently, it has been employed in tandem with a fluorescent labelling assay in buffers that promote or inhibit aggregation, to investigate the impact of flow-induced conformational changes on the ribonuclease H enzyme [175].

The EFD consists of two syringes (diameter = 4.61 mm) connected by a single capillary (diameter = 0.3 mm, length = 75 mm), where the syringe-capillary interface features a geometric contraction with sharp corners (Figure 4.2A). The apparatus geometry serves as a simplified representation of human blood vessels, such as the schematic in Figure 1.7.

In-vivo, vessels have elastic walls and are heterogeneous in size and structure, with diameters ranging from micrometres in the smallest capillaries [176] to centimetres in the aorta [177]. An idealised yet commonly observed relationship in vascular systems is that, when a trunk vessel bifurcates into two symmetric branch vessels, the trunk diameter is typically 1.26 times the diameter of each branch [178]. By comparison, the EFD syringe diameter is 15.4 times that of the capillary.

At the start of a flow experiment, one syringe is loaded with protein sample and the plungers are alternately depressed by a linear actuator to force the sample through the capillary for a set duration. Shearing flows dominate in the syringes and capillary body, but the immediate vicinity of the geometric contraction is dominated by extensional flow

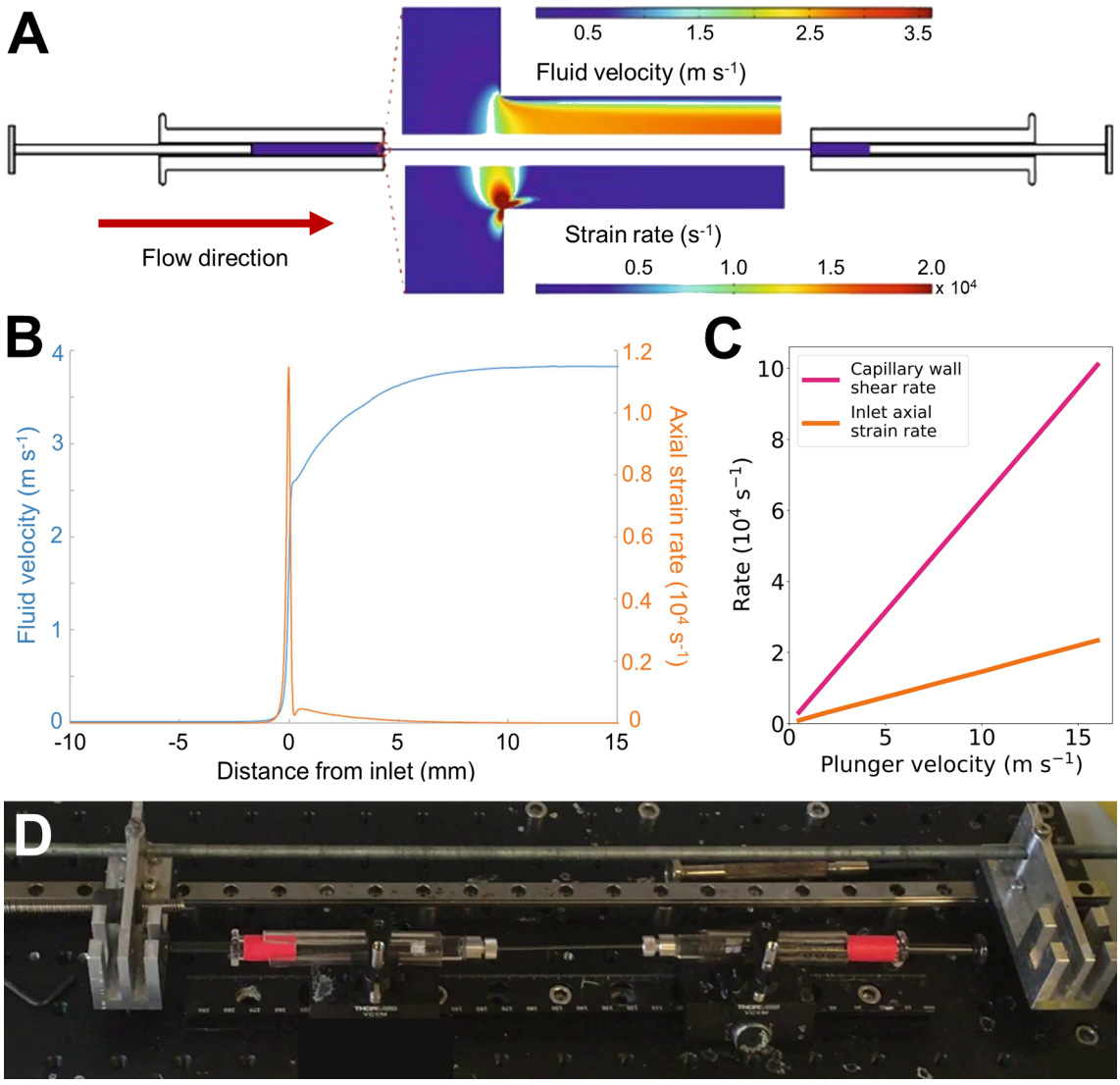


Figure 4.2: **A**) 2D schematic of the reciprocal extensional flow device (EFD), showing two syringes connected by a single capillary. Fluid velocity and strain rate fields were computed from axisymmetric CFD simulations, plunger velocity 8 mm s^{-1}. Adapted from [173]. **B**) Fluid velocity and axial strain rate in the vicinity of the inlet to the geometric contraction, plunger velocity 8 mm s^{-1}. Adapted from [179]. **C**) The axial strain rate at the inlet and wall shear rate in the capillary body depend linearly on the plunger velocity. Data from [173]. **D** Photograph of the EFD. The left syringe is being depressed by a metal plate driven by a linear actuator.

that acts longitudinally across the protein, due to a rapid 238-fold increase in velocity within a 2 mm region (Figure 4.2B). The wall shear rate of the capillary and the extensional strain rate along the centreline, computed from computational fluid dynamics (CFD) simulations, depend linearly on the plunger speed (Figure 4.2C and Table A.1).

Proteins that differ in structure by only a few residues, such as the IgG1 antibodies STT and WFL, have been shown to exhibit markedly different aggregation propensities when exposed to flow in the EFD, even with their similar hydrodynamic radii of 6 nm [173]. The aggregation propensity of a protein depends on some combination of sequence, pass number, plunger speed and concentration to overcome a force threshold and initiate local unfolding [166, 173], but the choice of buffer is also important [170, 175].

The capillary length of the EFD can be reduced to cause a corresponding reduction in the residence time of a protein in the high shear region, and hence assess the relative significance of extensional versus shear flow in the hydrodynamic response of a protein. Considering STT and WFL once more, the aggregation of STT was found to be independent of the capillary length, suggesting extensional flow as the dominant regime, whereas WFL aggregation was significantly reduced, suggesting shear was dominant [173].

All proteins studied in the EFD to date have exhibited globular aspect ratios, with hydrodynamic diameters on the order of 10 nm or less [166, 170, 173, 175]. The flow reponse of particularly large or fibrous proteins, such as fibrinogen (MW $\sim$ 340 kDa, length $\sim$ 45 nm), has not been investigated in the EFD until now. Furthermore, despite the clear presence of extensional flow fields in the human circulatory system at sites of vasoconstriction and pathological blockages, there are few examples in the literature that study its effects on fibrinogen structure. This is despite numerous experimental studies that expose fibrinogen to shear flow [109, 117, 123, 168], and single-molecule pulling experiments that apply force longitudinally [180, 181]. Another fibrous blood protein, von Willebrand factor, exists natively in an irregularly coiled state that unfolds under flow, but is more susceptible to extension than shear [51, 58].

4.1.2 Thiol-specific labelling of fibrinogen under flow provides structural insight into its mechanical response

The folded hexameric structure of fibrinogen (Figure 1.5A) is held together by 17 inter-chain and 12 intra-chain disulfide bonds [182] formed from a total of 58 cysteine residues, of which 50 are resolved in the X-ray crystal structure (Figure 4.3A) from Kollman et al. [95].

Covalent disulfide bonds[2] are usually formed when two or more highly reactive thiol groups (R-SH) are oxidised (Figure 4.4). Conversely, when a disulfide is reduced, it is

[2]The structure of two cysteine residues linked by a disulfide bond is known as *cystine*, but this term is not used in this thesis for the sake of clarity.

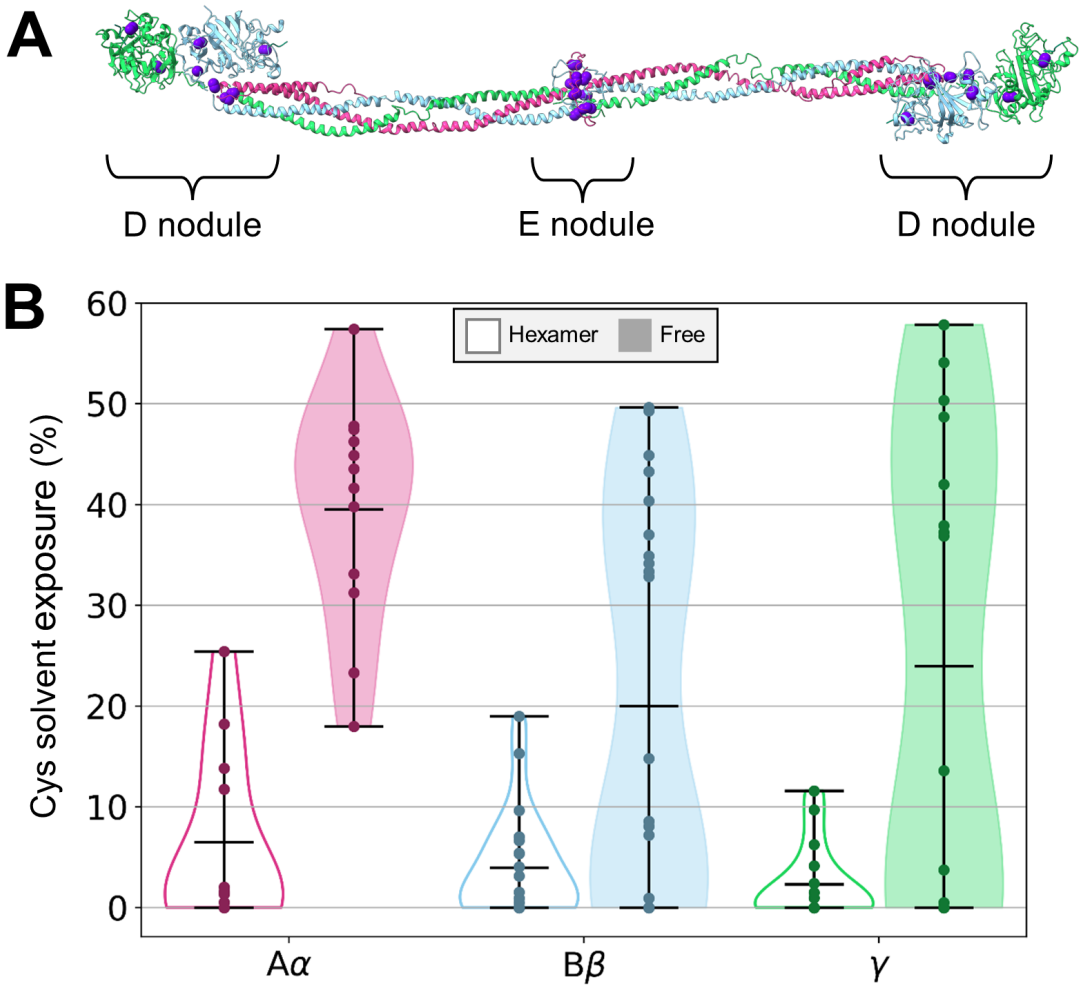


Figure 4.3: A) Ribbon representation of crystal structure of human fibrinogen (PDB: 3GHG [95]). Thiol groups in 50 cysteine residues are shown as purple spheres. Chain colours: Aα in magenta, Bβ in light blue and γ in green. **B**) Relative solvent exposure of structurally resolved cysteines, computed from chains in the folded hexameric structure (transparent) and free in solution (shaded). Outer bars denote range, central bars denote an average over the cysteine count in the dimer (Aα: 12, Bβ: 22, γ: 16).

cleft into its constituent thiols. Butera and Hogg [124] conducted mass spectrometry analysis of 13 disulfides in fibrinogen extracted from healthy human plasma, revealing that 10-50% of the bonds were in a reduced state.

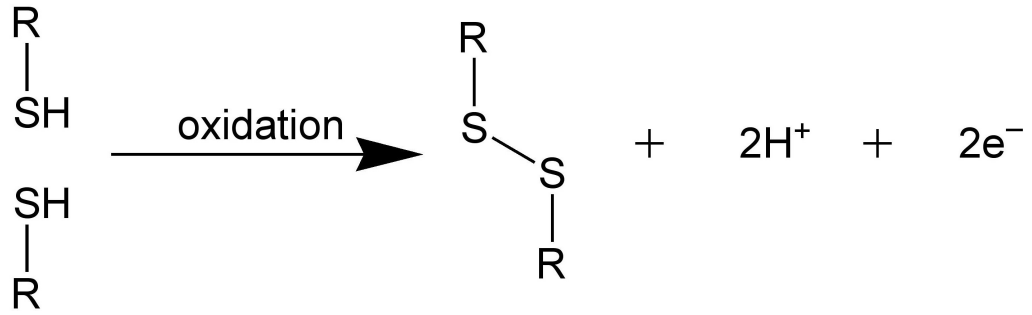


Figure 4.4: Two thiol groups (left) form a disulfide bond (right) when oxidised.

Cysteine is the only amino acid to contain a side-chain thiol [183], allowing it to be specifically labelled with a thiol-specific dye such as IAEDANS (Figure 4.5), which forms a stable thioether linkage (R-S-R′) and fluoresces at a wavelength of 336 nm [184]. The reactivity of IAEDANS is high, such that any cysteine sufficiently exposed to the solvent has a high probability of being labelled, provided IAEDANS is in a sufficient molar excess.

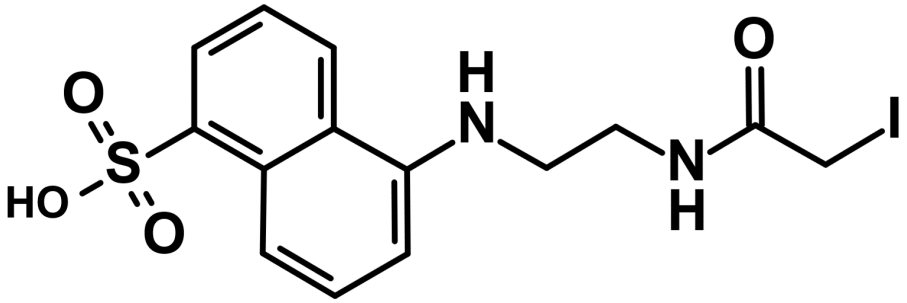


Figure 4.5: Chemical structure of the thiol-reactive IAEDANS dye.

By applying hydrodynamic flow to fibrinogen in the presence of IAEDANS, its structural resistance to flow can be assessed. Flow-induced partial unfolding can expose previously sequestered cysteine residues to the solvent[3], which are subsequently labelled by IAEDANS. Following the cessation of flow, the fluorescence emitted from bound IAEDANS can be quantified to provide structural insights; generally, a greater amount of fluorescence in the presence of flow would suggest that more unfolding has occurred. Thiol-specific labels have been employed to study a diverse range of proteins under flow, from bovine serum albumin (BSA) [166] and monoclonal antibodies [186] to fibrous cytoskeletal components [187] and fibrinogen [124].

[3]This exposure of buried cysteines is *not* due to mechanical rupture of disulfides, which would require hydrodynamic forces on the order of 1400 pN [185] that are not addressed in this thesis.

Figure 4.3B compares the solvent exposure of cysteines in the fibrinogen Aα, Bβ and γ chains between their native arrangement within the hexameric structure and when the chains are freely suspended in solution. This serves as an approximation of how flow-induced partial unfolding might affect the exposure of cysteines, and hence their likelihood of being labelled by IAEDANS.

Relative solvent exposure is a measure of the solvent accessible surface area (SASA) of a cysteine within a chain, given as a percentage of the total SASA of a single cysteine free in solution. Figure 4.3B shows that Aα has the largest range of cysteine exposure in the folded structure of up to 25%. Furthermore, when free in solution it has the largest proportion of highly exposed cysteines with a range of 20-40%, compared to Bβ and γ which possess deeply buried cysteines with less than 10% exposure.

These data suggest that Aα may have the highest chance of being labelled by IAEDANS, however, both Aα and γ contain cysteines in flexible portions of the molecule that were not accounted for due to being absent from the crystal structure.

4.1.3 Controlling for the variable redox state of fibrinogen disulfides

The variable redox state of fibrinogen disulfide bonds [124] presents a problem for labelling with thiol-specific probes. To quantify the extent of flow-induced remodelling, the intensity of IAEDANS fluorescence must be compared between quiescent (control) and flow-stressed samples. However, if the number of cysteines available for labelling is not equal across fibrinogen molecules, it is difficult to decipher if an increase in intensity is caused by flow-induced remodelling, or just native differences in redox state. A consistent baseline of fluorescence must be established in order to make valid comparisons across experimental replicates.

A chemical reductant such as Tris[2-carboxyethyl]phosphine (TCEP) could be introduced into the samples before flow is applied to reduce disulfide bonds that are most readily exposed to the solvent. This would normalise the number of available cysteines in control and pre-stressed samples, and hence the baseline fluorescence. Furthermore, TCEP has an advantage over dithiothreitol (DTT), another commonly used reductant, in that it does not compete with thiols in the IAEDANS reaction.

The presence of TCEP *during* the application of flow will additionally ensure that, if previously sequestered disulfides become exposed to the solvent, they will be reduced into cysteines and become available for labelling [186]. Given that this lowers the structural stability of fibrinogen relative to its native state, these experiments should be viewed as characterising the *susceptibility* of fibrinogen to flow-induced remodelling, rather than an *absolute* measure. A schematic of the interactions in the flow-induced remodelling assay is shown in Figure 4.6.

4.1.4 Chapter outline

The results in this chapter concern a series of experiments in which fibrinogen is exposed to flow inside the EFD and labelled with IAEDANS, with the aim of investigating whether its structure partially unfolds in response to hydrodynamic flow. Samples containing fibrinogen, IAEDANS and TCEP in Tris-buffered saline (TBS) were stressed at different plunger speeds and a low pass number to minimise aggregation. The extent of IAEDANS fluorescent labelling was quantified post-stress by reverse-phase high performance liquid chromatography (HPLC) and sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

This experimental workflow was previously devised by laboratory colleagues Leon Willis and Alex Page [175], thus saving time that would otherwise have been spent designing experimental procedures. Unless otherwise stated, all experiments presented in this chapter were conducted by the author, with assistance from Leon Willis and Alex Page.

A preliminary experiment is also presented, wherein the appropriate concentration of TCEP was determined at which the exterior disulfides of quiescent fibrinogen are reduced without significant disruption to its quaternary structure, monitored by SDS-PAGE. Samples containing fibrinogen and TCEP in TBS were run in the EFD at a single plunger speed and pass number to assess the effects of the selected TCEP concentration on protein aggregation, quantified by HPLC. Supplemental parameters relating to fibrinogen sequence and cysteine solvent accessibility were determined computationally.

The physiological relevance of fibrinogen under the conditions inside the EFD, with the presence of IAEDANS and TCEP, is questionable. The experiments in this chapter are part of a proof-of-concept study that aims to adapt a proven workflow to a large, fibrous protein and investigate whether it behaves abnormally inside the EFD when compared to previous studies.

With the exception of the rare inherited disease of fibrinogen Aα-chain amyloidosis [93, 188], fibrinogen has not been reported to adversely aggregate within the circulatory system, as in the pathway proposed by Figure 4.1. Therefore, these experiments should ideally be conducted in an aggregation-minimising environment, e.g. low pass numbers or plunger velocities [170, 173].

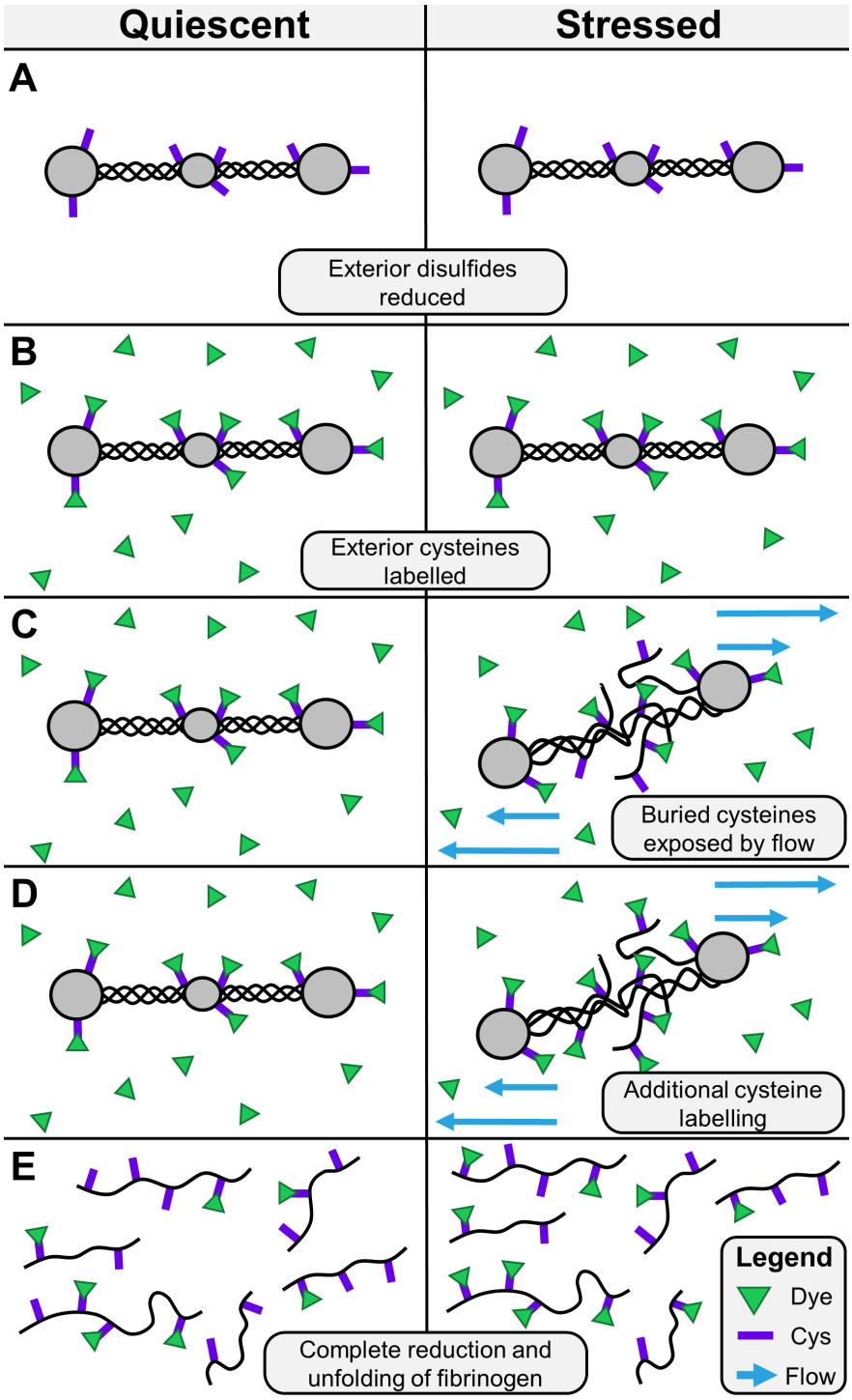


Figure 4.6: Cartoon of a fibrinogen monomer during the fluorescent labelling assay. **A**) Solvent-exposed disulfide bonds on the fibrinogen exterior are reduced by TCEP. **B**) Exterior cysteines are labelled by IAEDANS dye. **C**) Sample is stressed by flow (shear is shown here). Control is kept quiescent. Flow-induced remodelling of fibrinogen may expose buried cysteines. **D**) Exposed cysteines are subsequently labelled. **E**) Sample and control both quenched in DTT to halt labelling and completely reduce fibrinogen. Chains kept unfolded in guanidine. Cysteines, dye molecules and fibrinogen chains not to scale.

4.2 Materials and methods

4.2.1 Fibrinogen preparation

Plasminogen-depleted fibrinogen (Fg) purified from human plasma (Merck 341578) was prepared by Tímea Feller. The lysophilised powder was dissolved in high-purity (MilliQ) water and spun at 2000 rpm for 10 min to remove aggregates. The solution was dialysed in a Pur-A-lyzer Dialysis kit (Merck PURX60100, MWCO 12-14 kDa) against Tris-buffered saline (TBS, Tris 50 mM, NaCl 100 mM, pH 7.4, 0.22 μm filtered) to remove any proteins <12 kDa and set the required buffer conditions. Concentrations were measured after the dialysis using a NanoDrop spectrophotometer, generally ranging from 17 to 22 mg mL^{-1}; the extinction coefficient of fibrinogen is $\epsilon_{280}^{Fg} = 5.12 \times 10^5$ $M^{-1}cm^{-1}$. Concentrated fibrinogen was aliquoted and stored at -80°C. Unless otherwise stated, all fibrinogen samples used in experiments were diluted in TBS to physiological concentration of 2 mg mL^{-1} or 6 μM. [86].

4.2.2 TCEP serial dilution and non-reducing SDS-PAGE

TCEP was serially diluted two-fold in ddH_2O and prepared to eight concentrations from 50mM to 0.4 mM with fibrinogen at 2 mg mL^{-1}. Samples were kept quiescent at room temperature for 80 min, then diluted four-fold in TBS and two-fold in non-reducing loading buffer, boiled and analysed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE); the gel composition is given in Table 4.1. Final fibrinogen concentration in the gel was 0.25 mg mL^{-1}. Gels were held at a current of 50 mA for 20 min followed by 80 mA for 55 min, stained overnight in InstantBlue Coomassie, then imaged in white light on the Uvitec Alliance Q9 Advanced imaging system.

Component	**Resolving gel volume (mL)**	**Stacking gel volume (mL)**
30% (w/v) Acrylamide: 0.8% (w/v) bis-acrylamide	7.50	0.83
3 M Tris.HCl, 0.3% (w/v) SDS pH 8.45	5.00	1.55
ddH_2O	0.44	3.72
Glycerol	2.00	-
10% (w/v) ammonium persulfate	0.10	0.10
Tetramethylethylenediamine (TEMED)	0.01	0.01

Table 4.1: Components of tris-tricine buffered SDS-PAGE gel. Makes two square 8 cm x 10 cm x 1.5 mm gels.

The optimal TCEP concentration for further experiments was identified by observing in which lane a group of three bands in the 50-75 kDa range (Aα, Bβ and γ chains) disappeared, being replaced by a single doublet at 250 kDa at the top of the gel (whole fibrinogen). The titration was repeated in a narrower dilution range from 2 mM to 0 mM and

a final concentration of 0.4 mM was selected. At such a low concentration, TCEP did not significantly alter the pH of the fibrinogen solution.

4.2.3 Extensional flow device

Two 4.61 mm diameter, 1 mL syringes (Hamilton 81331) were connected by a 0.3 mm diameter, 75 mm long glass capillary with a screw cap, ferrule (Hamilton 55750-01) and o-ring at each end. The Reynolds numbers in the syringe and capillary were 37 and 570, respectively.

For each flow experiment one syringe was loaded with 0.5 mL of fibrinogen sample and run for a number of passes at a set plunger speed. One pass constituted the full depression of a plunger by a linear actuator, transferring the entirety of the loaded sample through the capillary into the opposite syringe; on the next pass the actuator reversed direction and depressed the opposite plunger. Samples were decanted immediately once the required pass number had been reached. Control samples of equivalent volume were kept on the bench in a quiescent state for the same duration as the stressed sample. Both syringes were placed in rubber tubing and clamped in place to prevent slippage during compression. Experimental setup adapted from Dobson et al. [166]. Strain and shear rates are given in Table A.1.

4.2.4 Fluorescent labelling of fibrinogen with IAEDANS

Five flow experiments were conducted at five different plunger speeds: 2, 3, 4, 8 and 16 mm s^{-1}. For each speed, a solution of 2 mg mL^{-1} fibrinogen + 0.4 mM TCEP + 0.2 mM IAEDANS (Invitrogen 11594886) + TBS, was prepared to a total volume of 1100 µL, of which 500 µL was stressed for 50 passes, alongside the remaining 600 µL that acted as the quiescent control for the same duration. The apparatus was covered with boards and aluminium foil to minimise the interference of daylight with labelling. After 50 passes, IAEDANS labelling was quenched by adding 50 µL of 1 M DTT to both the control and sample. The five experiments were run on the same day and were repeated once, one week later.

For plunger speeds of 3-16 mm s^{-1} the actuator was paused every two passes such that the duration of each experiment was equal to that of the slowest speed, 2 mm s^{-1}. At this speed the actuator moved 50 mm in 25 s for a single pass, so the duration of each experiment was $50 \times 25 = 1250$ s, or 21 min. Pause durations were 16, 25, 37 and 43 s for speeds of 3, 4, 8 and 16 mm s^{-1}, respectively.

After quenching, samples were divided into 50 µL and 300 µL volumes for analysis by SDS-PAGE and reverse-phase high-performance liquid chromatography (HPLC), respectively.

A fully unfolded standard was produced on the same day as each set of experiments with a

similar preparation method: 2 mg mL^{-1} fibrinogen + 0.4 mM TCEP + 0.2 mM IAEDANS + 6 M GdnHCl for a total volume of 1100 µL. These were left in the dark at room temperature overnight to ensure that IAEDANS had completely labelled the unfolded fibrinogen, then quenched the following day in 50 µL of 1 M DTT

4.2.5 Analysis of soluble protein and IAEDANS labelling by non-reducing SDS-PAGE

The 50 µL aliquots of samples decanted following an IAEDANS flow labelling experiment were diluted to 200 µL in TBS, then to 400 µL in non-reducing loading buffer and frozen at -20 °C. These were later thawed and analysed by SDS-PAGE with the same gel components and imaging system as in §4.2.2. Fluorescent labelling of bands was visualised by excitation with light from a built-in UV transilluminator. The gel was subsequently stained with InstantBlue Coomassie and imaged in white light. The location and pixel intensities of UV and white light bands were quantified using built-in gel documentation software.

A total of 120 pixel intensities were recorded across sixty bands imaged by both white light and UV in all experiments: three fibrinogen Aα, Bβ and γ chains, for five EFD plunger speeds, in quiescent control and stressed samples, over two biological replicates. Values for whole fibrinogen were obtained by summing over the chains. Intensities were averaged over both biological replicates ($n = 2$, for comparison between plunger speeds in a single chain) and over plunger speed ($n = 10$, for comparison between chains) then plotted as bar charts.

4.2.6 Analysis of soluble protein and IAEDANS labelling by HPLC

The 300 µL aliquots of samples decanted following an IAEDANS flow labelling experiment were diluted to 600 µL in 6 M guanidine hydrochloride (GdnHCl) and left in the dark at room temperature overnight; time constraints meant analysis could not occur the same day, so GdnHCl was necessary to prevent fibrinogen from precipitating. The following day, samples and the corresponding unfolded standard were buffer-exchanged with 0.1% trifluoroacetic acid (TFA) using Zeba spin desalting columns (Thermo Fisher 89882, MWCO 7 kDa), following the manufacturer instructions.

Samples were injected in 2 µL volumes onto a Shimadzu Nexera LC-40 HPLC system connected to a Nucleosil 300 C4 column (5 µm, 250×4.6 mm, Chromex CMF-0255) and separated by gradient elution of 5-80% acetonitrile (0.1% (v/v) TFA) in ddH_2O (0.1% (v/v) TFA) from 0 to 7.5 min, followed by 95-5% from 8 to 15 min, at 1 mL min^{-1}. Standards were injected in volumes of 0.5, 1, 2, 4, 8 µL and subjected to the same gradient. Modules in the HPLC system included a RF-20A fluorimeter that measured IAEDANS emission at 336 nm, and a Pure Distribution Analysis UV/vis detector that measured absorbance of

protein and dye at 280 nm and 336 nm, respectively.

A group of three peaks correspondeding to the Aα, Bβ and γ chains were detected in every chromatogram for protein absorbance at 280 nm, dye absorbance at 336 nm, and dye emission at 336 nm. These peaks were automatically detected and their areas integrated using built-in HPLC software. As with SDS-PAGE, the area for whole fibrinogen was obtained by summing over each chain. Peak areas were averaged over both biological replicates ($n = 2$, for comparison between plunger speeds in a single chain) and over plunger speed ($n = 10$, for comparison between chains) then plotted as bar charts.

Data for the unfolded standard was extracted in the same way, but the IAEDANS absorbance at 336 nm could not be quantified for masses of fibrinogen under 8 μg, due to measurements being below the threshold of the detector.

4.2.7 Dye labelling efficiency

The molar concentration of protein in a sample was calculated through application of the Beer-Lambert law

$$[\text{Protein}] = \frac{A_{280}^{\text{prot}} - \left(0.57 \times A_{336}^{\text{dye}}\right)}{\epsilon_{280}^{\text{prot}}} \tag{4.1}$$

where A_{280}^{prot} is the protein absorbance at 280 nm, A_{336}^{dye} is the dye absorbance at 336 nm, and $\epsilon_{280}^{\text{prot}}$ is the molar extinction coefficient of the protein (fibrinogen: $\epsilon_{280}^{\text{prot}} = 5.12 \times 10^5$ $\text{M}^{-1}\text{cm}^{-1}$). The optical path length was 1 cm. The dye absorbance at 280 nm was accounted for by a correction factor of 0.57, defined as $A_{280}^{\text{free dye}}/A_{336}^{\text{free dye}}$; this was previously verified by Leon Willis and Alex Page [175], and is analogous to corrections for other dyes by Kim et al. [189].

The dye to protein ratio is given by

$$[\text{Dye}]{:}[\text{Protein}] = \frac{A_{336}^{\text{dye}}}{[\text{Protein}] \times \epsilon_{336}^{\text{dye}}} \tag{4.2}$$

where $\epsilon_{336}^{\text{dye}}$ is the molar extinction coefficient of the dye (IAEDANS: $\epsilon_{336}^{\text{dye}} = 5700\ \text{M}^{-1}\text{cm}^{-1}$). Equation 4.2 is a measure of the labelling efficiency of the dye to the protein [190], where a value of 1 indicates that one IAEDANS molecules has bound to each fibrinogen molecule.

4.2.8 Fibrinogen parameters computed from amino acid sequences and crystal structure

The molar extinction coefficient of whole fibrinogen was calculated as $\epsilon_{280}^{\text{Fg}} = 5.12 \times 10^5$ $\text{M}^{-1}\text{cm}^{-1}$ using the ProtParam tool of the Expert Protein Analysis System (Expasy) [191],

Protein	Sequence length (AA)	Mature length (AA)	Cysteine count
Aα	644	610	8
Bβ	491	461	11
γ	437	411	10
Fg	-	2964	58

Table 4.2: Computed sequence parameters of fibrinogen and its component chains: mRNA transcript sequence length, mature protein length and cysteine residue count. UniProt Knowledgebase IDs: P02671-2 (Aα), P02675 (Bβ) and P02679-2 (γ).

which was in agreement with the literature at a difference of <0.5% [192]. mRNA transcript sequences of chains were obtained from the UniProt Knowledgebase [193] and truncated to remove signal peptides of length 19, 30 and 26 residues from Aα, Bβ and γ, respectively, from the start of each sequence. An excess 15 residues were additionally removed from the end of Aα so that it matched the published mature protein sequence [94]. The total number of cysteine residues in each chain was counted (Table 4.2). ProtParam can compute parameters for multimeric proteins by receiving a concatenation of multiple sequences as input, so an extended sequence of 2(Aα + Bβ + γ) was used as a representation of whole fibrinogen. The FASTA file format was used to store and process sequence data. UniProt Knowledgebase IDs for the chains were P02671-2 (Aα), P02675 (Bβ) and P02679-2 (γ).

The crystal structure of hexameric fibrinogen (PDB: 3GHG) [95] was used to calculate the solvent accessible surface area (SASA) of cysteine residues. The PDB structure contains twelve components: A, D (Aα); B, E (Bβ); C, F (γ); M, N, O, P (synthetic peptides); and U, V (sugars); only A-F were considered for computation, as other components did not contain cysteines. Flexible regions are absent from the structure: Aα 1-26 (fibrinopeptide A and knob A motif) and 201-610 (C-terminus connector and domain); Bβ 1-57 (fibrinopeptide B and knob B motif) and 459-461 (C-terminus); γ 1-13 (N-terminus) and 395-411 (C-terminus including γ-γ cross-linking site) [97]. The A-F components were extracted and saved as six separate PDB files to represent the fibrinogen chains free in solution, for a total of seven PDB files including 3GHG.

The Shrake-Rupley algorithm was used to calculate SASA [194]. The residues of the chain(s) in the seven files were looped over, the SASA of each cysteine was summed, then averaged over the dimer (each like pair of chains), resulting in six values for total SASA (three chains, hexameric or free in solution). The relative solvent accessibility was calculated for the cysteines in each chain,

$$\mathrm{RSA}_i^{\mathrm{Cys}} = \frac{\mathrm{SASA}_i^{\mathrm{Cys}}}{\mathrm{SASA}_F^{\mathrm{Cys}}} \ , \tag{4.3}$$

where i is the cysteine index and $\mathrm{SASA}_F^{\mathrm{Cys}} = 240.5 \pm 5.68$ Å^2 is a single cysteine amino

acid free in solution [195].

4.3 Results

4.3.1 Fibrinogen in a reducing buffer (0.4 mM TCEP) displays minimal disruption to its quaternary structure

Before commencing the main IAEDANS labelling experiments, a preliminary experiment was conducted to determine an appropriate concentration of TCEP, and to assess its impact on the structure of fibrinogen in the presence of flow. There exists an optimum TCEP concentration at which the exterior disulfides of fibrinogen are consistently reduced between experiments, but its quaternary structure is not broken apart into its six component chains.

A serial dilution of TCEP was added to 2 mg mL^{-1} (6 μM) fibrinogen and analysed by non-reducing SDS-PAGE (Figure 4.7). The three bands visible in SDS-PAGE at TCEP concentrations of 50 to 1.6 mM are characteristic of fibrinogen, denoting the Aα (66.1 kDa), Bβ (54.4 kDa) and γ (48.5 kDa) chains [94]. As the concentration decreases the three bands fade and a single doublet band (250 kDa) increases in visibility from [TCEP] = 6.3 mM, until only the doublet is present at 0.8 and 0.4 mM, suggesting the majority of fibrinogen in the sample is intact. The transition from the three bands to the single doublet indicates the correct TCEP concentration, which was selected as 0.4 mM after additional SDS-PAGE gels were run.

Whole fibrinogen appears as a doublet in non-reducing SDS-PAGE due to two variants that are typically present in human blood plasma [196]. *Low-molecular-weight* fibrinogen (305 kDa) lacks an αC region due to proteolytic cleavage and accounts for ~25% of total plasma fibrinogen, versus 70% for the normal *high-molecular-weight* fibrinogen (340 kDa) [103, 104].

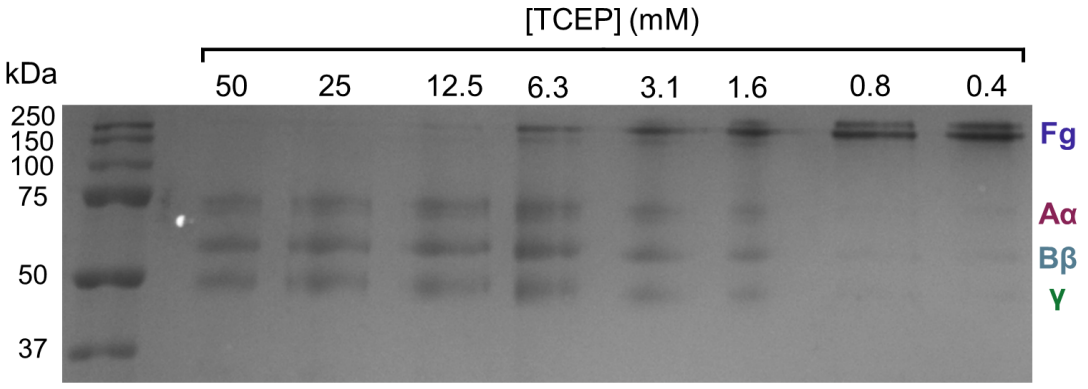


Figure 4.7: SDS-PAGE of serial dilution of TCEP in 2 mg mL^{-1} fibrinogen (white light). Three bands in the molecular weight range of 50-75 kDa are each labelled with the corresponding fibrinogen chain.

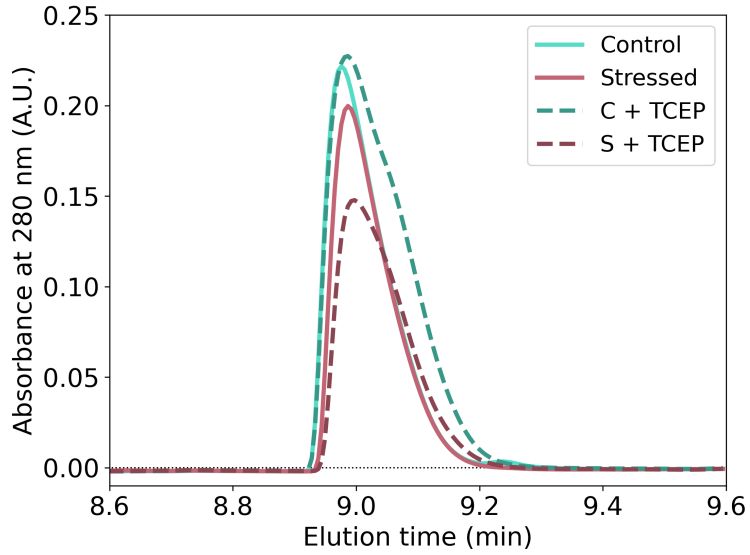


Figure 4.8: HPLC protein absorbance at 280 nm, of 2 mg mL^{-1} fibrinogen $\pm$ 0.4 mM TCEP. Stressed at 4 mm s^{-1} for 500 passes. Four samples are plotted: control - TCEP (solid, blue), stressed - TCEP (solid, pink), control + TCEP (dashed, blue), stressed + TCEP (dashed, pink).

Samples containing 2 mg mL^{-1} fibrinogen diluted in Tris-buffered saline (TBS) in the presence or absence of 0.4 mM TCEP were stressed in the EFD at 4 mm s^{-1} for 500 passes, centrifuged at 30,000 RPM for 30 min to remove aggregates, and quantified by HPLC; a quiescent control was kept on the bench for the same duration as each stressed run (Figure 4.8). The protein absorbance at 280 nm, Abs_{280}, decreased in the presence of flow, suggesting that soluble protein had been lost and aggregation had ocurred, with the aggregation being enhanced by 20% in the presence of TCEP.

A single chromatogram peak at $t \sim 9$ min, corresponding to the elution of a whole fibrinogen monomer, supports the SDS-PAGE results (Figure 4.7) in that fibrinogen has maintained its quaternary structure in the presence of 0.4 mM TCEP; complete reduction would have resulted in three distinct peaks. Both ($\pm$ TCEP) quiescent control samples had similar peak heights of $\text{Abs}_{280} \sim 0.2$, indicating they had equal concentrations of soluble fibrinogen, and that TCEP had therefore not disrupted the quaternary structure. The presence of TCEP appeared to cause peak broadening, so integrated peak area was not used in this instance.

The enhancement of aggregation in the presence of TCEP suggests that the structure of fibrinogen has been weakened, which was expected given that disulfide bonds are being chemically reduced. In order to minimise aggregation in the following IAEDANS labelling experiments, the pass number was reduced from 500 to 50.

The EFD runs were repeated multiple times and the general relationship between the four peaks in Figure 4.8 was consistently observed. However, EFD parameters such as plunger speed and pass number were changed each time, and hence precluded further analysis

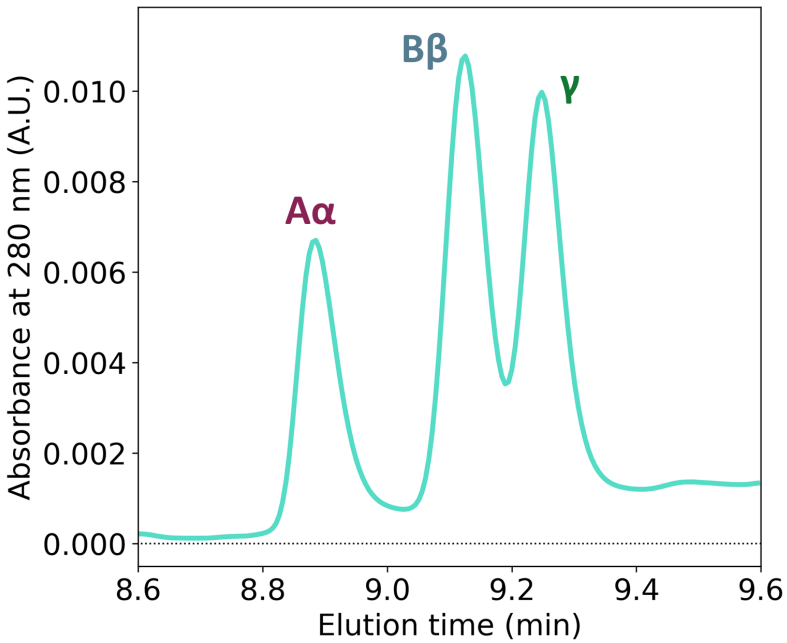


Figure 4.9: HPLC protein absorbance at 280 nm, of a quiescent control of the IAEDANS labelling assay. Reduction and denaturing of fibrinogen post-stress results in three peaks, labelled here with the corresponding fibrinogen chain.

and the calculation of errors.

4.3.2 Fibrinogen aggregation is statistically insignificant and independent of plunger speed

The remainder of the results section concerns the main experiments of this chapter: the labelling of fibrinogen by IAEDANS under flow. This subsection focusses on the quantification of the amount of soluble fibrinogen in the samples in order to check for aggregate formation, even with a reduction in pass number from 500 to 50, because this influences how the fluorescence data from IAEDANS labelling is handled.

Samples containing 2 mg mL^{-1} (6 µM) fibrinogen, 0.4 mM TCEP and 0.2 mM IAEDANS diluted in TBS were stressed in the EFD at five plunger speeds from 2 to 16 mm s^{-1} for 50 passes. Post-stress, samples were quenched with DTT to halt the IAEDANS reaction and analysed by SDS-PAGE and HPLC; the latter method had samples additionally denatured in guanidine hydrochloride (GdnHCl) to solubilise any aggregates that may have formed. Samples were not centrifuged post-stress.

The DTT and GdnHCl in the HPLC samples resulted in three distinct peaks in the Abs_{280} chromatogram that corresponded to the Aα, Bβ and γ chains, with mean elution times of, 8.85 ± 0.01, 9.07 ± 0.02 and 9.22 ± 0.01 min, respectively (Figure 4.9). The process used to assign HPLC peaks to their respective fibrinogen chains is described in §4.3.3.

The area underneath Abs_{280} peak of each chain was integrated and summed to obtain

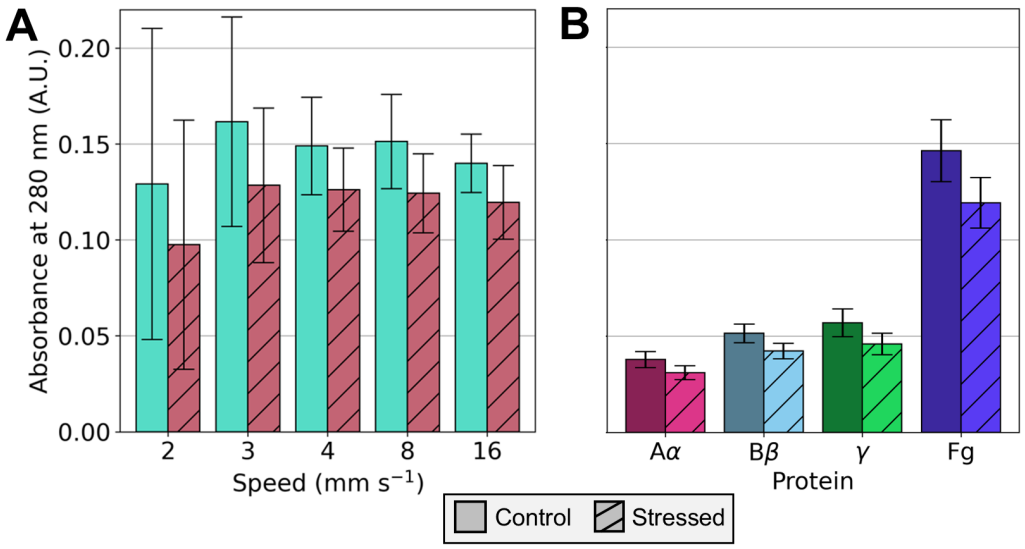


Figure 4.10: HPLC peak area of protein absorbance at 280 nm, IAEDANS labelling assay. **A)** Whole fibrinogen against plunger speed, averaged over two biological repeats ($n = 2$). Error bars denote range. **B)** Constituent chains and whole fibrinogen averaged over five plunger speeds and two biological repeats ($n = 10$). Error bars denote the standard error.

the Abs_{280} area of whole fibrinogen, which was plotted against plunger speed (Figure 4.10A). The consistent decrease in Abs_{280} when comparing the quiescent control to the stressed sample across every plunger speed is suggestive of soluble protein loss and hence aggregation, as discussed for Figure 4.8. Furthermore, there does not appear to be a clear dependence of aggregation on plunger speed, and it is observed to a similar extent in each chain when averaging over plunger speed (Figure 4.10B).

The error bars tend to overlap in both the control and stressed Abs_{280}, especially for 2 and 3 mm s^{-1} in Figure 4.10A where the errors are larger than the measurements themselves. The quantified aggregation is therefore statistically insignificant.

The HPLC results were corroborated by Coomassie-stained SDS-PAGE. The white light image in Figure 4.11 shows three bands at 50-75 kDa that correspond to the Aα, Bβ and γ chains at every speed, similar to Figure 4.7, but with no bands visible for MW > 150 kDa due to the addition of DTT post-stress. Faint, blurred bands above the 50-75 kDa range are suggestive of oligomers, which may suggest the presence of aggregates in the sample, although these were not investigated in detail.

The pixel intensity of each gel band was integrated and plotted against plunger speed (Figure 4.12A). With the exception of a large loss of protein at 16 mm s^{-1} (likely due to a data processing error), the difference between the quiescent and stressed samples at each given speed is statistically insignificant due to the large error bars. When averaged over plunger speed, the extent of aggregation in each chain (Figure 4.12B) is smaller than that of Figure 4.10B, but some error overlap persists.

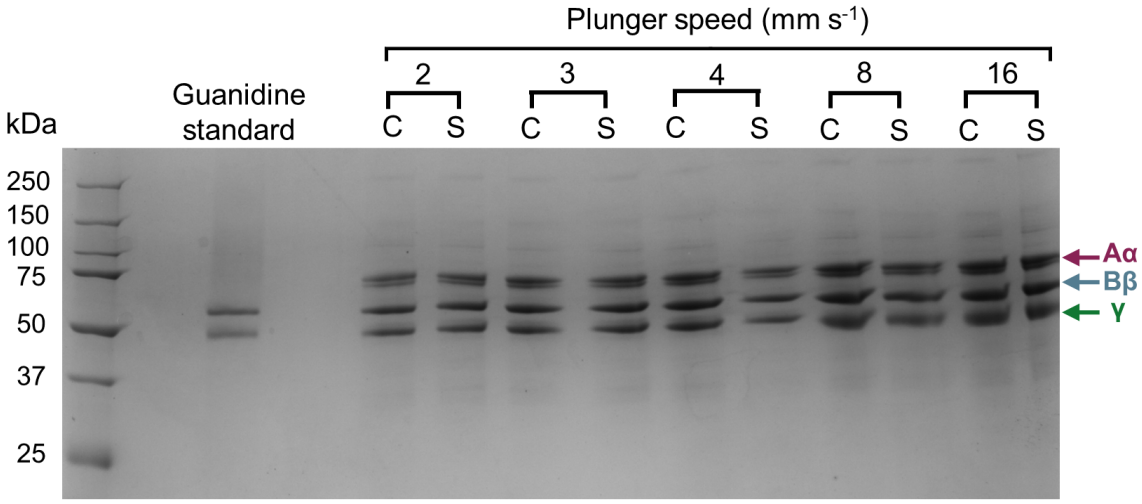


Figure 4.11: SDS-PAGE white light image, IAEDANS labelling assay. Three bands in the molecular weight range of 50-75 kDa are each labelled with the corresponding fibrinogen chain. An unfolded standard was loaded unsuccessfully in the third lane. C = quiescent control, S = stressed samples.

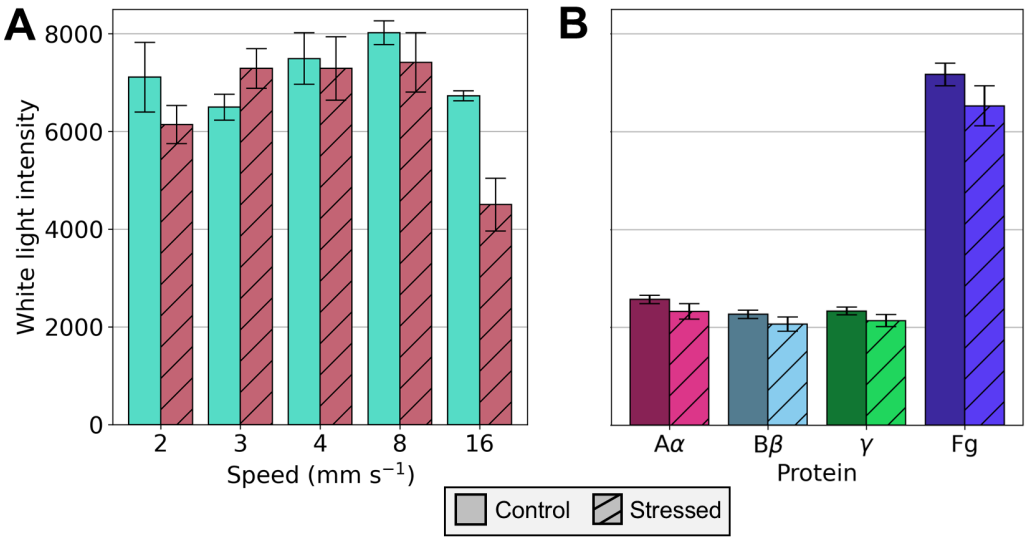


Figure 4.12: SDS-PAGE white light pixel intensity, IAEDANS labelling assay. **A**) Whole fibrinogen against plunger speed, averaged over two biological repeats ($n = 2$). Error bars denote range. **B**) Constituent chains and whole fibrinogen averaged over five plunger speeds and two biological repeats ($n = 10$). Error bars denote the standard error.

The fraction of soluble protein in the stressed sample relative to the quiescent control is tabulated for both HPLC and SDS-PAGE (Table 4.3), corresponding to the data plotted in Figures 4.10 and 4.12. The SDS-PAGE data has greater variation than the HPLC data, which was expected due to SDS-PAGE being the less precise method of the two.

Plunger speed (mm s)	**Protein absorbance at 280 nm**			
	Stressed ÷ Control			
	Aα	Bβ	γ	Fg
2	0.732	0.742	0.708	0.725
3	0.816	0.808	0.789	0.802
4	0.849	0.853	0.841	0.847
8	0.820	0.821	0.821	0.821
16	0.842	0.847	0.856	0.849
Mean	0.812	0.814	0.803	0.809
	White light intensity			
2	0.830	0.921	0.857	0.867
3	1.150	1.166	1.050	1.121
4	0.964	0.989	0.967	0.972
8	0.899	0.935	0.940	0.923
16	0.678	0.572	0.756	0.668
Mean	0.904	0.917	0.914	0.910

Table 4.3: HPLC protein absorbance at 280 nm of stressed sample as a fraction of the quiescent control, and the corresponding percentage change (i.e. loss of soluble protein due to aggregation).

Aggregation was not expected to be quantifiable via HPLC due to the highly reducing and denaturing environment in the sample post-stress, with a combination of TCEP, DTT and GdnHCl, in addition to a low pass number of 50. It was expected that any aggregates would have been solubilised into the component Aα, Bβ and γ chains, and that the resulting Abs_{280} would have been constant across all samples. The mechanism of aggregation and the ultimate fate of the aggregated material is unclear. Despite the aggregation being statistically insignificant, even a small loss of protein may result in the additional loss of bound IAEDANS, which directly impacts the analysis of the fluorescence data, as discussed in §4.3.3.

4.3.3 Quantification of IAEDANS labelling of fibrinogen is obfuscated by a loss of soluble protein

By analysing fluorescence emitted by IAEDANS labels bound to fibrinogen cysteines, information about the its structural remodelling in response to hydrodynamic flow can be determined. This is best done in an aggregation-minimising environment so that the fluorescence may be accurately quantified between the quiescent and stressed samples, however, there has been a small loss of protein from the samples despite efforts to prevent this. This subsection presents the HPLC and SDS-PAGE data that quantifies fluorsence of IAEDANS labels, without any corrections to adjust for soluble protein loss.

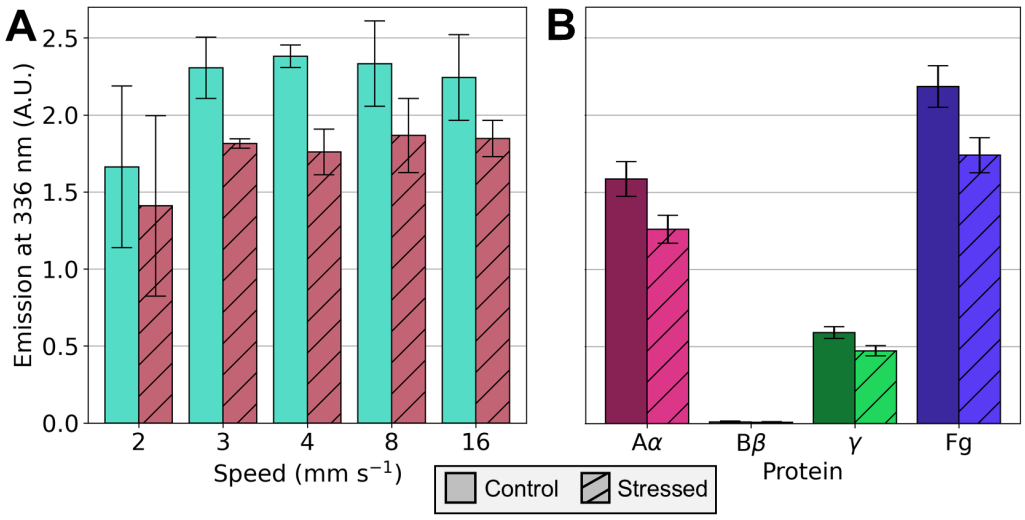


Figure 4.13: HPLC peak area of IAEDANS fluorescence emission at 336 nm, IAEDANS labelling assay. **A)** Whole fibrinogen against plunger speed, averaged over two biological repeats ($n = 2$). Error bars denote range. **B)** Constituent chains and whole fibrinogen averaged over five plunger speeds and two biological repeats ($n = 10$). Error bars denote the standard error.

The peak area of fluorescence emission at 336 nm of IAEDANS conjugated to fibrinogen was plotted against plunger speed (Figure 4.13A) and averaged over plunger speed for each fibrinogen chain (Figure 4.13B). Fluorescence was reduced in the presence of flow for all speeds, and there were significant differences between the fibrinogen chains, with Bβ having a negligible signal and stressed Aα (1.3 AU) fluorescing ×2.6 more than stressed γ (0.5 AU). There is a larger drop in fluorescence for Aα than γ.

The absorbance of IAEDANS at 336 nm, Abs_{336}, despite being measured close to the sensitivity threshold of the detector, appears to be correlated with the fluorescence results; absorbance drops in the presence of flow at every speed (Figure 4.14A), no signal was detected for Bβ, and Aα has a larger drop than γ (Figure 4.14B). If there are fewer fibrinogen molecules that have been labelled by IAEDANS, or some IAEDANS-bound fibrinogen aggregated and was lost from the sample, then the resulting fluorescence will be lower. The stressed samples showed similar amounts of absorbance (0.0035 AU, Figure 4.14A) and fluorescence (1.7 AU, Figure 4.13A) for speeds above 2 mm s^{-1}.

Despite Aα and γ both losing a similar amount of soluble protein as measured by Abs_{280} (Figure 4.10B), Aα appears to lose more IAEDANS labels than γ in the presence of flow, as shown by the difference between the chains in the drop of Abs_{336} (Figure 4.14B).

The triple bands at 50-75 kDa corresponding to the fibrinogen chains are once again visible when imaging the SDS-PAGE gel in UV (Figure 4.15), although Bβ bands are only visible for speeds above 4 mm s^{-1}. A gradual increase of band intensity with plunger speed is faintly visible in the image and the quantified data (Figure 4.16A). The intensities of both the control and stressed samples increase at speeds 2 to 8 mm s^{-1}, which may suggest it is

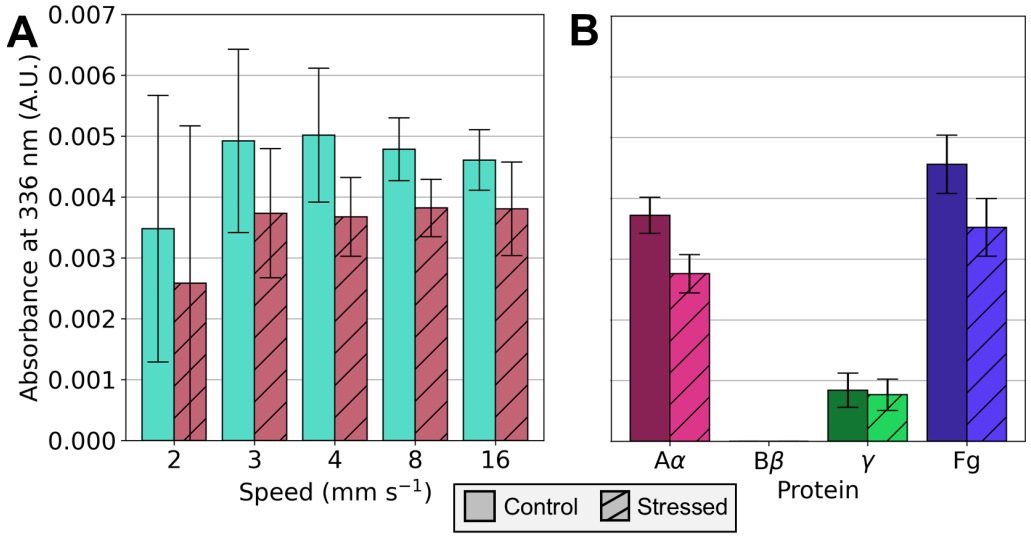


Figure 4.14: HPLC peak area of IAEDANS absorbance at 336 nm, IAEDANS labelling assay. **A)** Whole fibrinogen against plunger speed, averaged over two biological repeats ($n = 2$). Error bars denote range. **B)** Constituent chains and whole fibrinogen averaged over five plunger speeds and two biological repeats ($n = 10$). Error bars denote the standard error.

an artifact of the analysis rather than enhanced labelling; the control bands should be the same intensity regardless of speed.

Stark differences in brightness between bands within a given lane on the gel (Figure 4.15) enabled the retrospective identification of the three peaks in the HPLC chromatograms (Figure 4.9). The molecular weight and positions of the chains in the gel were known from white light SDS-PAGE (Figure 4.11). Furthermore, in the HPLC chromatograms the chains eluted in the same order and at the same respective times regardless of the detector used (Figure 4.9). By matching the magnitude of the fluorescence chromatogram peaks to the corresponding intensities of bands in the UV gel, each peak was assigned to its respective fibrinogen chain.

4.3.4 Labelling of fibrinogen Aα chain is enhanced in the presence of flow

The fluorescence of IAEDANS molecules bound to fibrinogen is highly dependent upon the amount of fibrinogen present in the sample, so the loss of protein caused by aggregation, although minimal and statistically insignificant, led to a corresponding drop in fluorescence. This subsection discusses an attempt to correct the fluorescence data to minimise the confounding effects of aggregation, and examines what the resulting fluorescence data suggests about the structural remodeling of fibrinogen under flow. Although some protein was lost in the presence of hydrodynamic flow, it is not clear if this was directly caused by flow. Two types of corrections may be applied to the fluorescence data that each attempt to address different potential sources of aggregation.

The first correction adjusts for protein loss in the presence of flow, as was observed in the

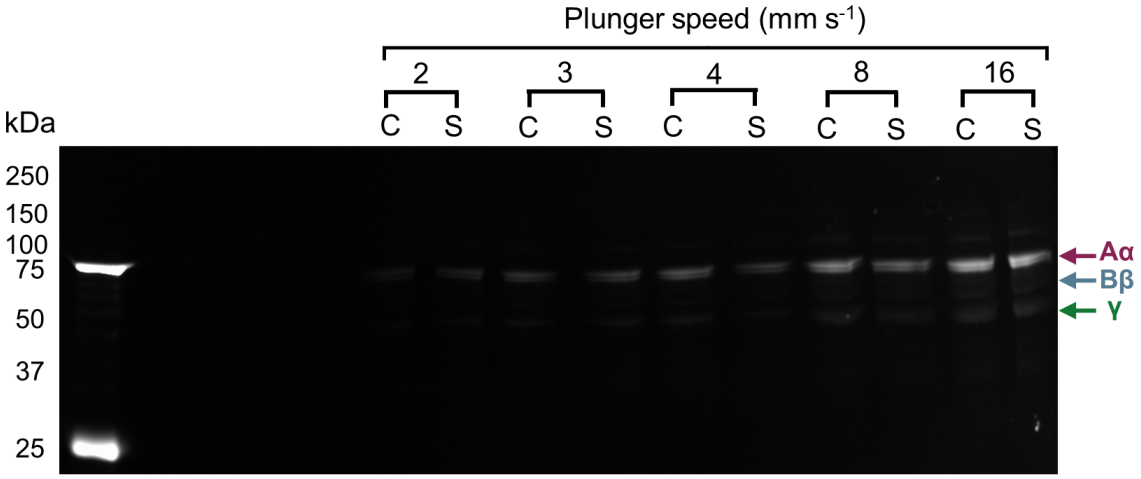


Figure 4.15: SDS-PAGE UV image, IAEDANS fluorescent labelling assay. Three bands in the molecular weight range of 50-75 kDa are each labelled with the corresponding fibrinogen chain. C = quiescent control, S = stressed samples.

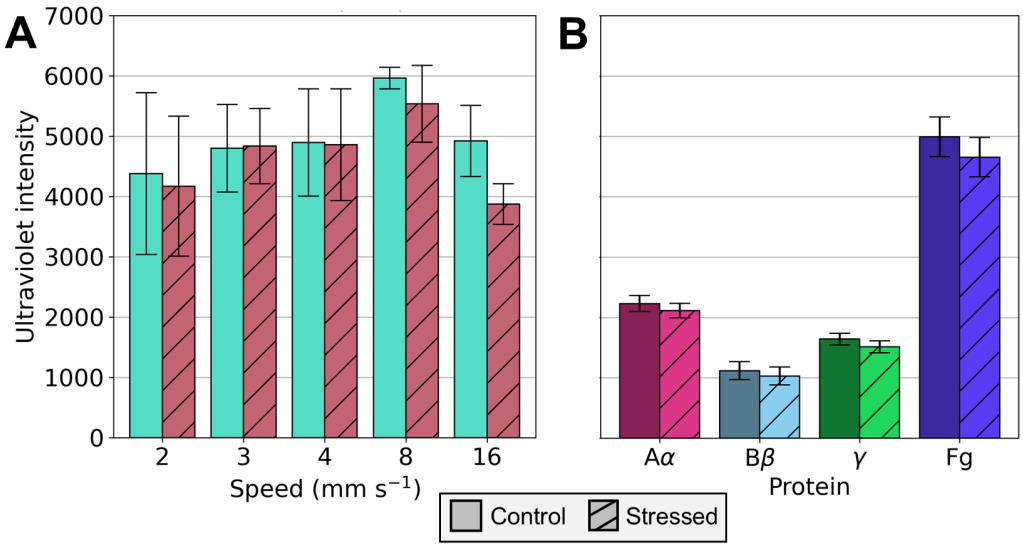


Figure 4.16: SDS-PAGE UV pixel intensity of IAEDANS fluorescence emission at 336 nm, IAEDANS labelling assay. **A**) Whole fibrinogen against plunger speed, averaged over two biological repeats ($n = 2$). Error bars denote range. **B**) Constituent chains and whole fibrinogen averaged over five plunger speeds and two biological repeats ($n = 10$). Error bars denote the standard error.

HPLC Abs_{280} data (Figure 4.10). The correction factor, K_1, is obtained by dividing the absorbance of the control by the absorbance of the stressed sample,

$$K_1 = \frac{\text{Abs}_{280}^{C}}{\text{Abs}_{280}^{S}} \ . \tag{4.4}$$

The correction is applied by multiplying the fluorescence of the stressed sample by K_1.

The second correction is more subtle. It adjusts for protein losses that are not attributable to flow-induced unfolding, e.g. precipitation in ambient conditions, getting stuck on the HPLC column, etc. The correction factors, K_2^C and K_2^S, are given by,

$$K_2^C = \frac{\text{Abs}_{280}^{\text{Std}}}{\text{Abs}_{280}^{C}} \qquad K_2^S = \frac{\text{Abs}_{280}^{\text{Std}}}{\text{Abs}_{280}^{S}} \ , \tag{4.5}$$

where $\text{Abs}_{280}^{\text{Std}}$ is the protein absorbance in a fully unfolded fibrinogen standard. The correction is applied by multiplying the fluorescence of the control and stressed samples by K_2^C and K_2^S, respectively.

The fully unfolded standard consists of 2 mg mL^{-1} fibrinogen, 0.4 mM TCEP and 0.2 mM IAEDANS diluted in 6 M GdnHCl to 1100 μL, such that it is identical to the samples prepared for an EFD run in the IAEDANS labelling assay, but with GdnHCl in place of TBS. The standard was incubated overnight in quiescent conditions and no evidence of aggregation was observed, so in theory it contains the amount of soluble fibrinogen that would be expected in EFD experiments conducted in aggregation-inhibiting conditions. Multiple masses of standard were analysed via HPLC (Figure A.1), but the 2 μg standard matches the mass of fibrinogen analysed in the flow labelling assay.

The correction factors, K_1, K_2^C and K_2^S, were calculated for whole fibrinogen across both biological replicates and are given in Table A.2. A worked example of a correction is given in §A.3.

Applying the corrections described above revealed that IAEDANS fluorescence was enhanced in the stressed sample relative to the control for every plunger speed (Figure 4.17A), although this enhancement was not always statistically significant. The magnitude of the stressed fluorescence between the fibrinogen chains did not change significantly (Figure 4.17B), with Bβ still exhibiting a negligible signal and Aα (2.1 AU) fluorescing ×2.6 more than γ (0.8 AU).

The fraction of the stressed signal relative to the control (Table 4.4) showed that Aα, γ and whole fibrinogen each have an average gain in fluorescence signal between 1.1 and 1.3; undetectable signals for much of Bβ precluded a comparable gain estimate. The gain in the γ chain is not significant due to error overlap. No trend with plunger speed in the fluorescence magnitude or gain was observed in Figure 4.17A. The corrected UV intensity

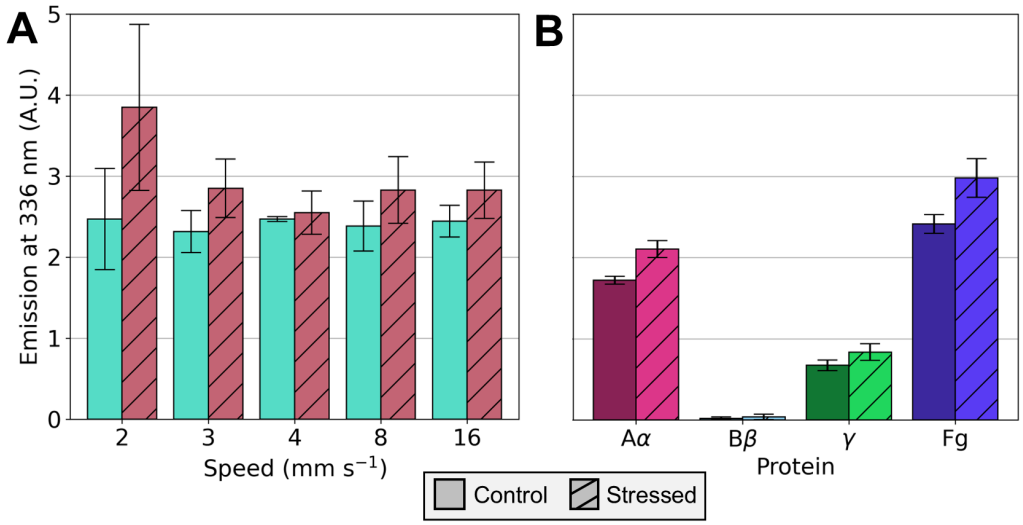


Figure 4.17: Corrected HPLC peak area of IAEDANS emission at 336 nm, IAEDANS labelling assay. **A**) Whole fibrinogen against plunger speed, averaged over two biological repeats ($n = 2$). Error bars denote range. **B**) Constituent chains and whole fibrinogen averaged over five plunger speeds and two biological repeats ($n = 10$). Error bars denote the standard error.

Plunger speed (mm s^{-1})	**Corrected IAEDANS emission at 336 nm**							
	Stressed ÷ Control							
	Aα		Bβ		γ		Fg	
2	1.609	1.460	1.962	1.263	1.464	1.754	1.576	1.530
3	1.221	1.187	1.152	-	1.315	1.297	1.247	1.209
4	1.141	0.977	1.184	-	1.191	0.756	1.155	0.912
8	1.208	1.156	1.429	-	1.196	1.192	1.205	1.165
16	1.207	1.102	1.458	-	1.191	1.101	1.203	1.102
Mean	1.277	1.176	1.437	-	1.272	1.220	1.277	1.183

Table 4.4: Corrected IAEDANS fluorescence emission at 336 nm, stressed sample as a fraction of the quiescent control. Stressed samples were corrected for protein loss due to flow. Control and stressed samples were corrected for protein loss due to other mechanisms. Data for both replicates is shown. Four Bβ samples had no measurable signal, so the mean was not calculated.

from SDS-PAGE could not be quantified as an unfolded standard was not successfully created.

An attempt was made to calculate the labelling efficiencies of each fibrinogen chain, which would have revealed how many IAEDANS molecules bound to each chain, on average. However, due to inconsistencies between the fluorescence data of the samples and the unfolded standard, this was unsuccessful. This was successfully quantified for ribonuclease H in the previous experimental workflow that was adapted for use in this chapter [175].

Plunger speed ($\mathbf{mm\ s^{-1}}$)	**Inlet speed** ($\mathbf{mm\ s^{-1}}$)	$\dot{\gamma}_{\mathbf{wall}}$ ($\mathbf{s^{-1}}$)	$\dot{\epsilon}_{\mathbf{inlet}}$ ($\mathbf{s^{-1}}$)	**Pe**	$F_{\mathbf{wall}}$ (**pN**)	$F_{\mathbf{inlet}}$ (**pN**)	**Pe′**	$F'_{\mathbf{wall}}$ (**pN**)	$F'_{\mathbf{inlet}}$ (**pN**)
2	480	13 000	3200	1	0.5	0.1	7	1.9	0.5
16	3800	100 000	23 000	10	3.9	0.9	58	15.4	3.6

Table 4.5: Péclet number, Pe, and Stokes drag force, F, on fibrinogen with its αC domains in closed ($L \sim 45$ nm) or axially extended ($L' \sim 90$ nm) conformations, due to shear in the EFD capillary or centreline extensional strain at the inlet, at minimum and maximum plunger speeds.

4.4 Discussion

4.4.1 Mechanical response of fibrinogen

The Péclet number, Pe, is the ratio of advective to diffusive forces in a system (Equation 1.4), such that for $\mathrm{Pe} > 1$ advective forces are the dominant contribution to molecule dynamics. Rigid, rod-like polymers in shear flow tend to align with the flow direction at $\mathrm{Pe} > 1$, with some tumbling [197], but it is well-established that extensional flow is highly effective at inducing alignment [58, 115]. For flexible, coiled polymers beyond some critical strain rate, this alignment may be accompanied by an increase in end-to-end length that enhances the drag force on the polymer [76] as it undergoes a 'coil-stretch transition' from a bundled coil to an extended chain [65, 198]. At every plunger speed, the Péclet number inside the EFD capillary was in the range $1 < \mathrm{Pe} < 60$ (Table 4.5).

The central, folded structure of fibrinogen in its native state has an end-to-end length of 45 nm in a straight conformation, or 35 nm in an S-shape bent conformation observed in semi-dilute concentrations [97]. The largely unstructured αC region usually forms noncovalent interactions with the central E nodule via its folded αC domain, but it also exists in open conformations that extend out from the main structure [99, 199].

The αC regions may effectively double the end-to-end length of fibrinogen if both become elongated. Assuming a fibrinogen length of 45 nm, the αC connectors branch off just before the 5 nm diameter Bβ and γ nodules in the distal D nodule. Length distributions from AFM have shown αC to have a peak end-to-end length of 21 ± 6 nm [100], for an additional maximum fibrinogen length of $(21{+}6{-}5){\times}2 = 44$ nm and a total of $44{+}45 = 89$ nm. The αC-E interaction forces are infrequent and fairly weak, typically on the order of 10s of piconewtons [200], although alignment and extensional force may reduce the probability of interaction by pulling the αC domains away from the central E nodules. Simulations of quiescent semidilute fibrinogen have shown that αC domains can interact across different molecules and sometimes entangle [97], which may enhance the effective particle size and drag force.

The flow forces inside the EFD are too weak to result in major unfolding of fibrinogen, regardless of αC extension. Atomic force microscopy (AFM) experiments and simulations of end-to-end pulling of fibrinogen showed that tensile force propagates primarily through the γ chain, beginning with reversible extension of the coiled-coil connectors at

<20 pN, then proceeding through multiple regimes of γ chain unfolding up to 75-150 pN [180, 201].

A mathematical dumbell model comparable in size to a fibrinogen molecule estimated that extensional forces in the EFD contraction were on the order of tens of femtonewtons [179, 202]. By approximating αC-extended fibrinogen as a Brownian sphere in Stokes flow, the drag force in the EFD capillary was estimated to range from 0.5 to 16 pN (Table 4.5). The structural changes most likely to occur in fibrinogen during these experiments are extension of the αC connectors and partial extension of the coiled-coil connectors, although the γ nodules may be susceptible to marginal amounts of unfolding [180].

The lack of correlation between plunger speed and aggregation extent (Figure 4.10A and 4.12A) suggests that fibrinogen is below an energy barrier required to initiate further aggregation events. Protein aggregation has been broadly observed to increase with speed and pass number [173], but can be modulated by concentration, sequence, and buffer conditions [170, 175].

Only the plunger speed was varied in the experiments in this chapter, thus representing a one-dimensional section of the aggregation landscape of a protein [173]. Given that fibrinogen is the precursor to the fibrin network and is well known for its elasticity and structural strength [203], it could be generally aggregation-resistant.

4.4.2 Effects of EFD geometry

The velocity gradients in the EFD are an order of magnitude higher from shear in the capillary compared with the extension rates in the contraction. However, the slender shape of fibrinogen means the hydrodynamic drag forces in both regions may be comparable due to the alignment of the molecule with the velocity gradient in extensional flow (Table 4.5).

The corners to the inlet of the contraction have not yet been investigated as a possible third source of force, in addition to the contraction and the capillary. Axisymmetric CFD simulations showed very high strain rates in the vicinity of the corners that extended radially into the capillary and the syringe (Figure 4.18) [179].

The flow around an infinitely sharp 270° corner (i.e. pointing into the fluid; re-entrant) is mathematically singular [204] as the velocity gradient becomes infinite as you approach the corner, which can lead to numerical difficulties in CFD calculations [205]. This does not occur in practical scenarios due to all corners possessing some degree of smoothness, although the nearby strain rates will still be much higher compared to other regions of flow.

Molecules that pass close to the EFD corners will experience much higher strain rates than those passing through the centreline. In a Newtonian fluid, the flow velocity near a 270°

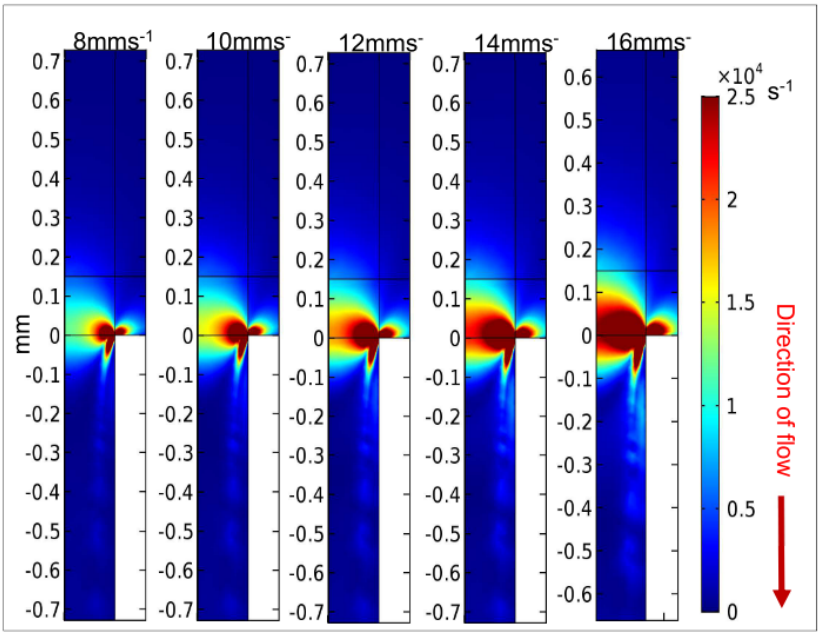


Figure 4.18: CFD simulations of axisymmetric EFD geometry showing strain rate profiles at various plunger speeds. Regions of high strain rate are localised to the corners. Note that the colour scale is restricted to a maximum strain rate of 2.5×10^4 s^{-1} and the calculated strain rates are significantly higher. Adapted from work by John Dobson [179].

corner scales as $r^{0.55}$ and the strain rate decays as $r^{-0.45}$, where r is the distance from the corner [204, 206], so a portion of molecules will experience an enhanced strain rate. The geometry of the EFD was not altered in the experiments in this chapter, so it remains unclear whether the contraction, capillary, or corners were the regions responsible for the minimal aggregation observed in fibrinogen.

4.4.3 Fluorescent labelling of fibrinogen

The magnitude of the quiescent fluorescence signals (Figure 4.17B) are significantly different between the chains, in a way that does not correlate with the solvent accessibility (Figure 4.3B). The Bβ chain does not label at, whereas Aα labels twice as much as γ. This is additionally reflected in the dye absorbance (Figure 4.14B). It could be that the observed fluorescence is a result of IAEDANS labelling cysteines that are not contained within the 3GHG structure.

Calculations of SASA from a PDB file are limited in that they only represent a static structure. In reality SASA is constantly fluctuating with time due to thermal noise, modulated by the flexibility of the structure in the vicinity of cysteine residues. Additionally, shear forces have been shown to greatly enhance the SASA of disulfide bonds in antibodies [186].

Each chain has four cysteines at either side of both coiled-coil connectors, localised to four

disulfide-forming regions (Figure 1.5B and 4.3A). Given the high reactivity of IAEDANS and the presence of TCEP in the samples, it is reasonable to assume that if Aα or γ were being labelled at these sites, then the close proximity of Bβ cysteines would result in a fluorescence signal corresponding to Bβ. Although the forces at the maximum plunger speed fall within the range that may extend the coiled-coil connectors, this structural motif may not be a good candidate for IAEDANS labelling.

Unlike Bβ, the Aα and γ chains both possess two cysteines in regions of fibrinogen that are highly mobile: Aα442 and Aα472 in one of two subdomains in the compact αC domain [207], attached to the intrinsically disordered αC connector, and γ8 and γ9 in the γ chain N-terminus that forms part of the central E nodule. Both structures are unresolved in the 3GHG structure and hence SASA could not be obtained, although the bovine αC domain has been characterised by nuclear magnetic resonance (NMR) spectroscopy (PDB: 2JOR) [96].

The enhanced ability of these cysteines to explore the solvent [208] increases the probability of interactions with TCEP and IAEDANS. Furthermore, less force may be required to expose these cysteines to the solvent due to intrinsic disorder of γ-N, and the variable stability of the αC domain [96, 207].

The Aα chain fluorescence was enhanced by ∼20% in the presence of flow, whereas γ was not. Despite being statistically significant in the case of Aα, the relatively small increase in only one chain suggests the dominant factor affecting the IAEDANS labelling was the quiescent redox state of fibrinogen, modified by TCEP.

Fluorescence enhancement did not correlate with aggregation after corrections, so the partial unfolding responsible for additional Aα labelling was not instigated by aggregation.

The concentration of fibrinogen used (2 mg mL^{-1} or 6 μM) falls within the 0.5 to 5 mg mL^{-1} range hypothesised to involve a switch from surface-mediated to bulk-mediated aggregation [170], although because the concentration and pass number were not varied, and aggregation was insignificant, this cannot be confirmed using the results in this chapter.

Work by colleagues has highlighted that interactions with EFD surfaces can influence aggregation [170, 172]. In the presence of aggregation-permitting (phosphate) buffer, aggregates of ribonuclease H deposited at the fluid-plunger interface were very highly labelled by IAEDANS, yielding an exponential relationship between labelling extent and plunger speed [175].

4.5 Summary

Fibrinogen at a physiological concentration of 2 mg mL^{-1} was stressed in a reciprocal extensional flow device (EFD) for five plunger speeds from 2 to 16 mm s^{-1} and 50 passes;

wall shear rates in the capillary exceeded those reported for pathologically contracted blood vessels [118]. Proteins were exposed to extensional flow at a geometric contraction and shear flow in the syringe and capillary. A chemical reductant, TCEP, and a thiol-reactive fluorescent dye, IAEDANS, were present in the samples during experiments. Post-stress, the labelling reaction was quenched in DTT and samples were analysed separately by reverse-phase high performance liquid chromatography (HPLC) and sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

A statistically insigificant amount of aggregation was observed in the presence of flow for each of the fibrinogen Aα, Bβ and γ chains. Although there was a weak trend in that the amount of soluble protein in each chain decreased slightly when stressed, which usually indicates aggregation, this was ultimately inconclusive due to overlapping of error bars. This was observed despite post-stress HPLC samples being mixed with a reductant (DTT) and denaturant (GdnHCl) that were expected to solubilise any aggregates. The presence of additional, faint bands in SDS-PAGE white light and UV images, above the 50-75 kDa molecular weight range expected for reduced fibrinogen, were not investigated in detail but nonetheless hint at the presence of oligomers in the samples; aggregates tend to become significantly labelled in the presence of IAEDANS [175].

A preliminary experiment showed that fibrinogen does aggregate at a higher pass number of 500, but this was not rigorously tested. Without further EFD experiments and a reduction in the sources of error, there is no strong evidence that fibrinogen aggregates under pathological flow.

The Bβ chain was not labelled by IAEDANS at all, whereas Aα labelled over twice as much as γ. This could be due to Aα and γ possessing cysteines whose solvent accessibility is enhanced by highly flexible structural motifs: in Aα, the globular αC domain connected to the flexible αC connector; in γ, the flexible N-terminal tail in the central E-nodule. The total fluorescence quantified in each chain is largely a result of the quiescent solvent accessibility of their cysteine residues, and by implication the redox state of their disulfide bonds.

Only the Aα chain labelled more in the presence of flow. This was likely due to partial unfolding and not aggregation, although it is unclear whether said unfolding was bulk- or surface-mediated, or whether it was caused primarily by extensional or shear flow. The precise location of the Aα cysteine(s) that were exposed by partial unfolding is unknown.

The limited scope of the experiments in this chapter, coupled with the structural complexity of fibrinogen, make it difficult to formulate a detailed description of its response to flow. Despite this, as an exploratory proof-of-concept investigation into the stressing of fibrous proteins in the EFD, it has been successful. A previously developed methodology was adapted to a completely different type of protein, with minimal technical difficulties, by a researcher with zero prior experience in a structural biology laboratory. The protein

absorbance and dye fluorescence data were consistent across experimental techniques, being reproduced by HPLC and SDS-PAGE.

4.6 Future work

4.6.1 Improving the experimental workflow

The most obvious way to reduce the size of the errors in the HPLC data is to perform more biological replicates. Most of the data presented in this chapter was averaged from two such replicates conducted one week apart from each other, but three should be the minimum required. A quantitative method should be adopted to determine whether the difference between two HPLC measurements, such as the pairs of bars shown in Figure 4.17B, is statistically significant. One example of such a method is an analysis of variance test, in which the F-statistic is computed for groups of data and used to assess significance [209].

Another approach to error reduction would be to take advantage of a newer model of the EFD, which is able to run three replicates simultaneously, unlike the model in Figure 4.2D that can only handle a single syringe-capillary setup. Under ideal conditions, a complete EFD experiment of five different plunger speeds from 2 to 16 mm s^{-1} takes $21 \times 5 = 105$ min or 1 hr 45 min, however, in reality this takes much longer to complete due to downtime between experiments, failed samples, additional procedures and analysis, etc. Performing three replicates on the same day would save time, as well as minimise variance due to environmental factors, such as ambient light levels interfering with IAEDANS labelling as samples are prepared and decanted.

The small reduction in protein absorbance between quiescent and stressed samples, even though it was statistically insigificant, proved to be a confounding factor when attempting to quantify the fluorescence emission data via HPLC. This should be addressed immediately if these experiments are to be repeated. It could simply be that the DTT and GdnHCl added post-stress were not in a sufficient molar excess with fibrinogen to completely solubilise the aggregates (if any). Another explanation is that some aggregates were deposited on the syringe surface and become lodged behind the plunger, in which case they would not have been decanted with the sample and therefore lost without the opportunity to be solubilised; this material can be recovered by rinsing the syringe with the DTT + GdnHCl mixture [175].

The low pass number of 50 was purposefully chosen to minimise aggregation, although proteins have previously been stressed in the EFD for up to 2000 passes [166]. Aggregation could instead be inhibited by use of an alternative buffer to TBS, such as arginine succinate, which was used by Alex Page to study the flow-induced unfolding of ribonuclease H [175]; a preliminary test would be required to ensure the stability of fibrinogen

in this chemical environment.

The plunger speed was the only EFD parameter to be varied during these experiments, with pass number, fibrinogen concentration, and buffer conditions remaining constant. This resulted in a limited sampling of the potential parameter space [170, 173], as a complete exploration was out of scope for this project.

Aggregation-inhibiting conditions would mean the pass number could be increased without the danger of incurring additional aggregation. Fibrinogen has a long catabolic half-life of around two days [210] and blood takes around 30 seconds to completely circulate the human body [121, 211]. Therefore, a fibrinogen molecule may pass through the same region of a blood vessel on the order of 5000 times, which might be the point at which the vessel bifurcates, the site of a pathological blockage, etc. An EFD pass number on the order of 1000 is hence a plausible upper limit for physio- or pathological conditions.

Achieving physiological strain and shear rates observed in blood flow on the order of 1200 s^{-1} [211] and 2000 s^{-1} [119], respectively, would require plunger speeds below 1 mm s^{-1} (Table A.1). This is trivial to set in the EFD by reprogramming its Arduino microcontroller interface, although it might require retrofitting with a slower linear actuator. For pathological shear rates, speeds of 2 to 7 mm s^{-1} are suitable [118].

4.6.2 Using mass spectrometry to obtain residue-level spatial information on partial unfolding

The mass spectrometry study from Butera and Hogg [124] quantified the redox state of fibrinogen disulfides in quiescent and sheared conditions. Under quiescent and physiological (2000 s^{-1}) shear flow, fibrinogen has a variable redox state, but a greater number of disulfide bonds are reduced in pathological (10,000 s^{-1}) flow; the latter is thought to enhance the rate of chemical reduction of disulfides, e.g. by thiol-disulfide exchange [212]. In brief, their methodology can be summarized as

1. Apply shear flow to blood plasma samples extracted from healthy human donors, leaving some quiescent.
2. Extract fibrinogen from plasma and alkylate unbonded cysteine thiols with 2-iodo-N-phenylacetamide (^{12}C-IPA, Figure 4.19), thereby fixing its redox state.
3. Isolate fibrinogen by SDS-PAGE and incubate with DTT to reduce disulfide bonds.
4. Alkylate remaining cysteines thiols with a stable carbon-13 isotope of the previous label, ^{13}C-IPA.
5. Analyse cysteine-containing peptide fragments by HPLC and mass spectrometry.

The mass difference between cysteines labelled with ^{12}C-IPA or ^{13}C-IPA is 6.02 Da, so by looking for this difference in the mass spectra of a given cysteine-containing peptide frag-

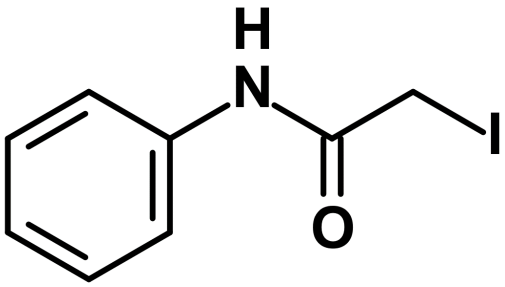


Figure 4.19: Chemical structure of 2-iodo-N-phenylacetamide (^{12}C-IPA), the thiol-specific label used by Butera and Hogg [124].

ment, it can be determined whether a particular cysteine was either natively unbonded or part of a disulfide.

This type of labelling can be adapted for use in the experiments presented in this chapter by being used in place of IAEDANS. Previously, IAEDANS was added to fibrinogen samples immediately before the application of flow, then afterwards the total emitted fluorescence was compared between quiescent and flowed states to quantify the mechanical response.

Instead, the quiescent and flowed samples could be incubated with ^{12}C-IPA some time before setting up the apparatus, which will fix the redox state of fibrinogen. Immediately before flow application, both samples are loaded with ^{13}C-IPA, such that cysteines that became exposed to the solvent by partial unfolding are subsequently labelled by ^{13}C-IPA. By comparing the mass spectra of peptides in both samples, the cysteines of fibrinogen that are most susceptible to flow-induced unfolding can be determined. This provides residue-level detail of the regions of fibrinogen that are resistant or vulnerable to hydrodynamic flow.

Chapter 5

Conclusion

5.1 Summary

The original version of the KirchhOff Biological Rod Algorithm (KOBRA) software circa 2020 was summarised in Chapter 2. Slender biological macromolecules are discretised into chains of interconnected elements. These elements exhibit three elastic degrees of freedom - stretch, bend and twist - that allow the rod structure can deform in response to thermal noise. The material parameters that define the elastic reponse can be parameterised from atomistic molecular dynamics (MD) simulations, or defined manually with experimentally determined values.

Whilst interfaces can be defined between KOBRA rods and Fluctuating Finite Element Analysis (FFEA) tetrahedral meshes, intermolecular interactions were not represented in the software. If two rods made contact, they passed through each other in a nonphysical manner. This limited the application of KOBRA to sparse systems that are not representative of the crowded nature of molecular biology.

The work presented in Chapter 3 addressed this lack of rod-rod interactions by developing two algorithms that capture repulsive steric interactions and attractive van der Waals (VDW) forces. The Lennard-Jones (LJ) potential was modified to impart a soft repulsive force between rod elements that partially overlap. Van der Waals sites, positioned by the user on the centrelines of rods, interact with each other via a modified form of the LJ potential. Periodic boundary conditions were also implemented for rods.

The functionality of rod-rod interactions were verified with a series of unit and integration tests, and performance benchmark simulations of a fibrinogen rod model. Minimal effort was devoted to ensuring the interactions were parallelised, so the interaction algorithms exhibited serial run time scaling. However, this should be seen as an opportunity for the next KOBRA developer to achieve large gains in performance with only a modest amount of development time.

Despite its serial performance, KOBRA simulations of densely packed interacting rods, representing one million atoms, demonstrated excellent speed, achieving hundreds of microseconds simulated per day. For systems of 512 rods, equivalent to eight million atoms, significantly larger than any previously simulated with KOBRA, speeds of 2 microseconds per day were attained. The KOBRA software is now much better equipped to simulate a diverse range of crowded macromolecular systems at the biological mesoscale.

Fibrin protofibrils were assembled from nineteen interacting fibrinogen rods, and KOBRA simulations were run as part of a parameter sweep over the interaction energy and bending stiffness. A set of parameters was found that successfully preserved the overall structure of the protofibril throughout the duration of a simulation, despite it being held together solely through attractive rod-rod interactions. This provided additional validation of the KOBRA interaction algorithms in a biologically relevant context.

Chapter 4 detailed proof-of-concept wet lab experiments that were conducted to quantify the *in-vitro* mechanical response of fibrinogen to shear and extensional flow. These experiments were primarily intended to test whether a large, fibrous protein could be used in the reciprocal extensional flow device (EFD), rather than being an accurate study of fibrinogen under flow. Physiological concentrations of fibrinogen were passed through the EFD for a number of plunger speeds and a fixed residence time. The thiol-specific fluorescent dye, IAEDANS, was employed to label cysteine residues that might become exposed to the solvent via flow-induced partial unfolding of fibrinogen.

Although the experimental parameters were chosen to try and minimise the adverse aggregation of fibrinogen, a small yet statistically insignificant amount occurred nonetheless. The mechanism that caused the aggregation was not determined. This complicated the subsequent analysis of the IAEDANS fluoresence and required corrections to be applied to the data. The Aα chain of fibrinogen appeared to label more under flow, but the cysteine responsible for this increase in labelling could not be identified. The overall fluoresence quantified from IAEDANS bound to fibrinogen was largely a result of the quiescent solvent accessibility of its cysteine residues. In the absence of a broader experimental scope and detailed analysis via mass spectrometry, it is difficult to comment on the flow response of fibrinogen.

5.2 Future work

5.2.1 Background flow

The translational and rotational motion of rod nodes in KOBRA, given respectively by Equations 2.3 and 2.4, apply to the case where the fluid medium is stationary. To investigate the effects of flow, a ubiquitous feature of biological systems [49, 213], these should be modified to take account of viscous drag against a moving background. For an arbitrary

background fluid velocity, $\boldsymbol{U}(\boldsymbol{r}_i, t)$, Equation 2.3 becomes

$$d\boldsymbol{r}_i = \frac{dt}{\zeta}(\boldsymbol{F}_i + \boldsymbol{f}_i) + dt\boldsymbol{U}(\boldsymbol{r}_i, t) \ , \tag{5.1}$$

and Equation 2.4 becomes

$$d\theta_i = \frac{dt}{\zeta_\theta}(\tau_i + g_i) + \frac{dt}{2}\Omega_i \ , \tag{5.2}$$

where Ω_i is the vorticity component acting along element i,

$$\Omega_i = \left[\boldsymbol{\nabla} \times \boldsymbol{U}(\boldsymbol{r}_i, t)\right] \cdot \boldsymbol{l}_i \ . \tag{5.3}$$

Ridley modelled bulk shear flow in FFEA simulations of packed spherical colloids [142] using a modified form of PBCs known as Lees-Edwards boundary conditions (LEBCs) [141]. Assuming shear flow is applied in the x-axis and there is a velocity gradient in the y-axis, then every node in the system experiences an additional drag force in the x-axis,

$$F_x = \dot{\gamma} h \zeta \left(\frac{y_{\text{node}}}{h} - \frac{1}{2} \right) \ , \tag{5.4}$$

where $\dot{\gamma}$ is the shear rate, h is the width of the simulation box, y_{node} is the y position of a node, and ζ is the viscous drag coefficient.

The velocity gradient causes the periodic cells in the y-axis to move whilst the central simulation box remains fixed (Figure 5.1). When a node crosses a periodic boundary in y it is shifted in x by an amount proportional to the shear rate,

$$\Delta x = \dot{\gamma} h t \ , \tag{5.5}$$

where t is the elapsed simulation time. A node crossing a boundary in the x-axis is unaffected by the shift and behaves according to normal PBCs.

The software implementation of LEBCs has its own branch in the FFEA source code repository, separate to the main release. The lack of flow physics from the current version of FFEA prevents a valid comparison between the rod simulations in Chapter 3 and the labelling experiments in Chapter 4.

5.2.2 Hydrodynamic interactions

The movement of a body through a viscous fluid causes perturbations in the local flow velocity field that decay as L/r, where L is the largest linear dimension of the body and r is

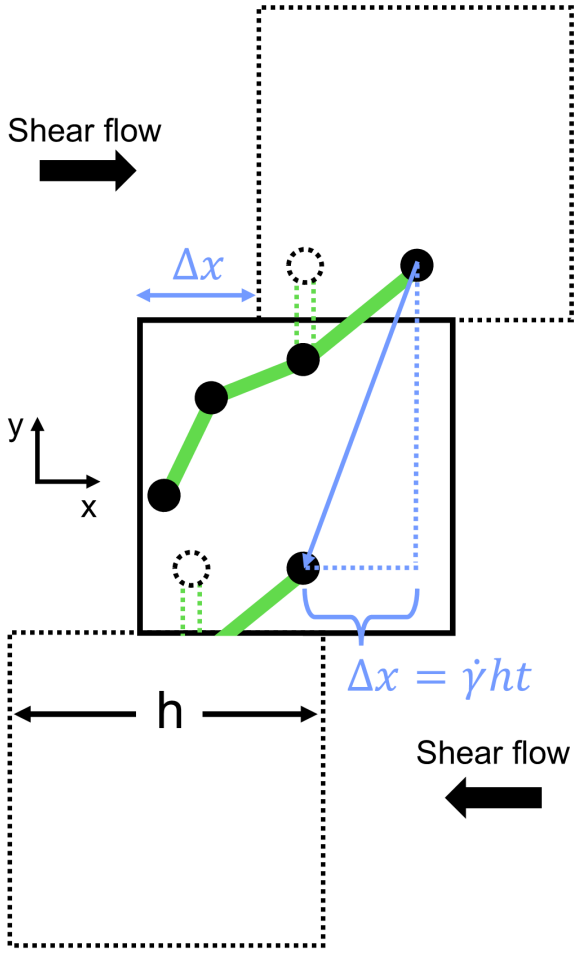


Figure 5.1: Schematic of Lees-Edwards boundary conditions. Periodic cells in the y-axis move due to shear flow applied in the x-axis. A rod node and its image are separated by a shift, Δx, proportional to the shear rate. Additional periodic images are not shown.

the distance from the body to some point in the fluid [76]. Any other bodies within range will move due to this flow, and will additionally perturb their local fluid environment as a result (Figure 5.2). These fluid-mediated correlations in macromolecular dynamics are known as *hydrodynamic interactions* (HI) [54, 214].

Hydrodynamic interactions have been shown to cause synchronised motion in closely packed systems of filaments such as cilia arrays [77, 216] and bacterial flagella bundles [217], accelerate the folding rates of proteins [218], and are thought to be involved in generating fast, large-scale flows in the cytoplasm [219]. The long range and complex nature of HI makes them computationally expensive to model for large systems, with N coupled particles simulated by Brownian dynamics [220] scaling as $\mathcal{O}(N^3)$ [214], although there are numerous hydrodynamic models of proteins [77–79, 82].

A complete mathematical description of HI is out of scope for this section, but a brief look at point particles is useful for understanding viscous coupling. The hydrodynamic framework developed by Batchelor [76] defined slender bodies as a linear series of point

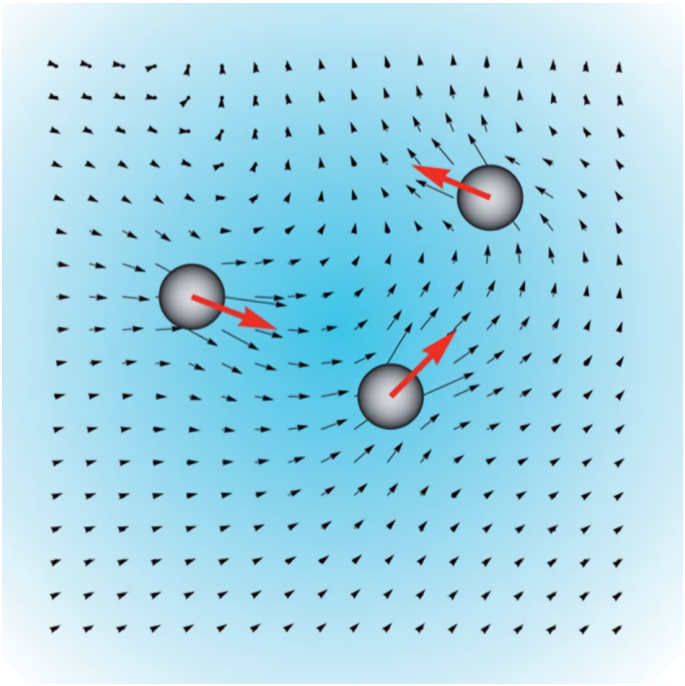


Figure 5.2: Schematic of hydrodynamic interactions. The movement of solute particles (red arrows) in a viscous fluid causes perturbations in the background flow field (black arrows) that influence the motion of nearby particles. Adapted from [215].

particles that influenced the surrounding fluid.

When considering a system of N spherical particles suspended in a fluid, the velocity of a particle, i, due to the force acting on all other particles, j, in the system, is given by

$$\boldsymbol{u}_i = \sum_j^N \boldsymbol{M}_{ij} \cdot \boldsymbol{F}_j \ , \tag{5.6}$$

where $\boldsymbol{M}_{ij}$ is the 3×3 mobility matrix, and $\boldsymbol{F}_j$ is the external force acting on particle j [221]. In the absence of HI the self-mobility is given by

$$\boldsymbol{M}_{ij} = \frac{1}{6\pi\eta R}\boldsymbol{I} \ , \quad i = j \ , \tag{5.7}$$

where η is the fluid viscosity, R is the hydrodynamic radius of the particles and $\boldsymbol{I}$ is the identity matrix. The simplest model of HI between different particles is given by the Oseen approximation to the mobility,

$$\boldsymbol{M}_{ij} = \frac{1}{8\pi\eta}\left(\frac{\boldsymbol{I} + \hat{\boldsymbol{r}}_{ij} \otimes \hat{\boldsymbol{r}}_{ij}}{r_{ij}}\right) \ , \quad i \neq j \ , \tag{5.8}$$

where particles i and j are separated by a distance, r_{ij}, and $\hat{\boldsymbol{r}}_{ij} \otimes \hat{\boldsymbol{r}}_{ij}$ is the outer product

of the unit vector, $\hat{\boldsymbol{r}}_{ij} = \boldsymbol{r}_{ij}/r_{ij} = (\hat{r}_x, \hat{r}_y, \hat{r}_z)$, with itself [220]. Combining Equations 5.7 and 5.8, the mobility may be written as

$$\boldsymbol{M}_{ij} = \frac{1}{\pi\eta}\left[\frac{\delta_{ij}}{6R}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} + \frac{1-\delta_{ij}}{8r_{ij}}\begin{pmatrix}1+\hat{r}_x^2 & \hat{r}_x\hat{r}_y & \hat{r}_x\hat{r}_z\\ \hat{r}_x\hat{r}_y & 1+\hat{r}_y^2 & \hat{r}_y\hat{r}_z\\ \hat{r}_x\hat{r}_z & \hat{r}_y\hat{r}_z & 1+\hat{r}_z^2\end{pmatrix}\right], \tag{5.9}$$

where δ_{ij} is the Kronecker delta [26]. The component, $\boldsymbol{M}_{ij}$, is a 3×3 matrix, so the complete Oseen tensor for N particles, $\boldsymbol{M}$, is of size $3N \times 3N$.

The Oseen tensor is able to account for far-field HI, where $r_{ij} \gg R$, with its effective range decaying as $1/r_{ij}$, but it significantly overestimates the magnitude of HI between nearby particles. The Rotne-Prager-Yamakawa tensor is a more sophisticated approach that includes terms of order $1/r_{ij}^3$ [222, 223].

Coarse-grained molecular models that include steric interactions and an implicit solvent may need to include HI in addition to local viscous drag. Where filaments sterically interact they are also likely fall within the L/r effective range of HI, unless the environment is sufficiently packed so as to screen the interaction and shorten its range [54]. Mesoscale models such as KOBRA and FFEA are well suited to reach the larger timescales at which flow is important and to compensate for the computationally complexity of HI.

5.2.3 Generalised interactions between structures

The rods of KOBRA and tetrahedral meshes of FFEA are markedly different in terms of their geometry and viscoelastic models. If additional objects are planned for the future, such as elastic sheets to represent membranes [73], a generalised interaction framework that treats objects in an abstract manner will be essential to avoid increased complexity and divergence in the codebase.

Designing and programming such a framework ideally requires a professional research software engineer, as previously suggested by Welch [127], who could focus solely on software stability and provide several years of continuous oversight. These are significant advantages over a PhD student, who has the additional pressures of generating results and writing a thesis, and is therefore incentivised to only program exactly what they require for their project. To date, KOBRA and FFEA have been developed exclusively by PhD students.

The rod-rod interactions described in Chapter 3 were not designed to interact with tetrahedral elements. The rod-tetrahedra interfaces developed by Welch [127] transmit force between the two objects, which precludes the need for steric interactions at the point of attachment. However, if an attached rod is sufficiently flexible to curl back on itself, or unattached rods are likely to diffuse towards the surface of tetrahedra, then the objects

will pass through each other. Furthermore, dynamic binding events are not currently represented between rods and tetrahedra.

It is significantly easier in the short term to write an algorithm for a specific case than it is to generalise it for many possible cases, however, the latter approach can significantly improve software maintainability and reduce technical debt in the long term. As the number of cases increases, the generalised approach will pay for itself in terms of hours spent programming.

5.2.4 Orientation-dependent rod-rod interactions

The VDW interactions in KOBRA are isotropic, in the sense that forces between elements do not depend upon their radial orientation. However, if a rod is particularly coarse relative to the protein it is based on, the interactions lose accuracy with regards to the biology. For example, the protein may have a binding site located in a specific sector of its cross-section; interactions in KOBRA would behave as if the site spanned the entire rod diameter. Ideally, a user should have the option to define the size and angular dependence of attractive regions on rods.

The current implementation treats VDW sites as point attractive potentials, but the Monte Carlo model from Vácha and Frenkel [224] features rods with attractive surface patches that have a set orientation (Figure 5.3A). These 'patchy' rods are rigid and are not connected in a chain, but their geometry is very similar to KOBRA rod elements, so a brief summary of the attractive potential may prove useful for future development[1].

The net attractive potential between two rods, i and j, is given by

$$U_{attr} = U_{dist}\frac{V_i + V_j}{2d} f_1 f_2 \ , \tag{5.10}$$

where U_{dist} is the distance-dependent component of the potential and d is the rod diameter. The terms V_i and V_j define sections of each rod axis that are facing an attractive patch, and fall within the cutoff radius of the potential (Figure 5.3B). The functions, f_1 and f_2, scale the potential based on the relative orientation of the surface patches on i and j, respectively, and are given by

$$f_1 = \begin{cases} 0 \ , & \boldsymbol{c}_{ij}^{\perp} \cdot \boldsymbol{m}_i \leq C_1 \\ \dfrac{1}{2} - \dfrac{\boldsymbol{c}_{ij}^{\perp} \cdot \boldsymbol{m}_i - (C_2 + C_1)/2}{C_2 - C_1} \ , & C_1 < \boldsymbol{c}_{ij}^{\perp} \cdot \boldsymbol{m}_i \leq C_2 \\ 1 \ , & \boldsymbol{c}_{ij}^{\perp} \cdot \boldsymbol{m}_i > C_2 \ , \end{cases} \tag{5.11}$$

[1]Some mathematical symbols in this section have been altered from the original paper to be consistent with KOBRA whilst avoiding clashes.

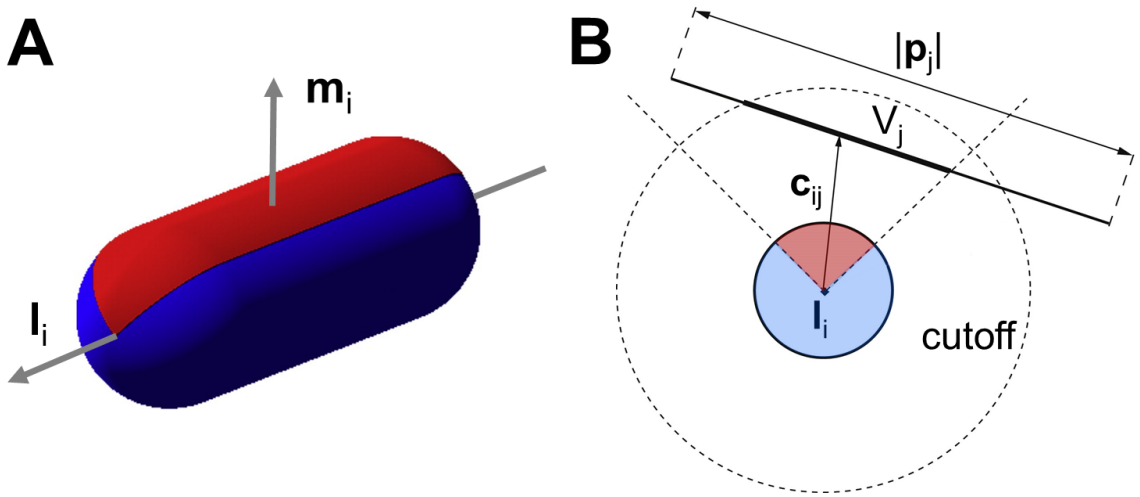


Figure 5.3: Model of rod interactions from Vácha and Frenkel. **A**) A rigid rod (blue) with an attractive patch (red). The patch orientation, $\boldsymbol{m}_i$, is perpendicular to the unit rod axis, $\boldsymbol{l}_i$. **B**) Schematic of a rod-rod interaction, where rod i is facing into the page. The interaction occurs between the patch on i and the section of axis, V_j, that faces the patch and lies within the cutoff radius. Adapted from [224].

where $\boldsymbol{c}_{ij}$ is the centreline displacement calculated between the sections V_i and V_j (Figure 5.3B). The unit vector, $\boldsymbol{m}_i$, defines the patch orientation and is perpendicular to the unit rod axis, $\boldsymbol{l}_i$ (Figure 5.3A). The unit vector, $\boldsymbol{c}_{ij}^{\perp}$, is formed from the projection of $\boldsymbol{c}_{ij}$ perpendicular to $\boldsymbol{l}_i$. The constants, C_1 and C_2, define the angular dependence of the patches. In general, the attractive potential is at a maximum when two patches are directly facing each other. A diverse range of clustered, isotropic and fibrous aggregate structures were observed in simulations by tweaking the patch parameters, rod aspect ratio and system temperature [224].

Appendix A

Experiments

A.1 Shear and strain rates in the EFD

Plunger speed ($mm\ s^{-1}$)	Axial strain rate (s^{-1})	Wall shear rate (s^{-1})
0.1	*3.56×10^2*	*5.98×10^2*
0.2	*5.00×10^2*	*1.23×10^3*
0.3	*6.45×10^2*	*1.86×10^3*
0.4	*7.90×10^2*	*2.49×10^3*
0.5	*9.35×10^2*	3.12×10^3
0.6	*1.08×10^3*	3.75×10^3
0.7	*1.22×10^3*	4.38×10^3
0.8	1.37×10^3	5.01×10^3
0.9	1.51×10^3	5.64×10^3
1	1.66×10^3	6.27×10^3
2	3.11×10^3	**1.26×10^4**
3	4.55×10^3	**1.89×10^4**
4	6.00×10^3	**2.52×10^4**
5	7.45×10^3	**3.15×10^4**
6	8.90×10^3	**3.78×10^4**
7	1.03×10^4	**4.41×10^4**
8	1.18×10^4	5.04×10^4
16	2.34×10^4	1.01×10^5

Table A.1: Plunger speed, extensional strain rate at the geometric contraction, and capillary wall shear rate in the EFD. Approximate ranges for *physiological* and **pathological** flows in blood vessels are based on shear rates [118, 119, 121] and extensional strain rates [211, 225, 226] reported in the literature. Extrapolated from supplementary data in [173].

A.2 Unfolded fibrinogen standard

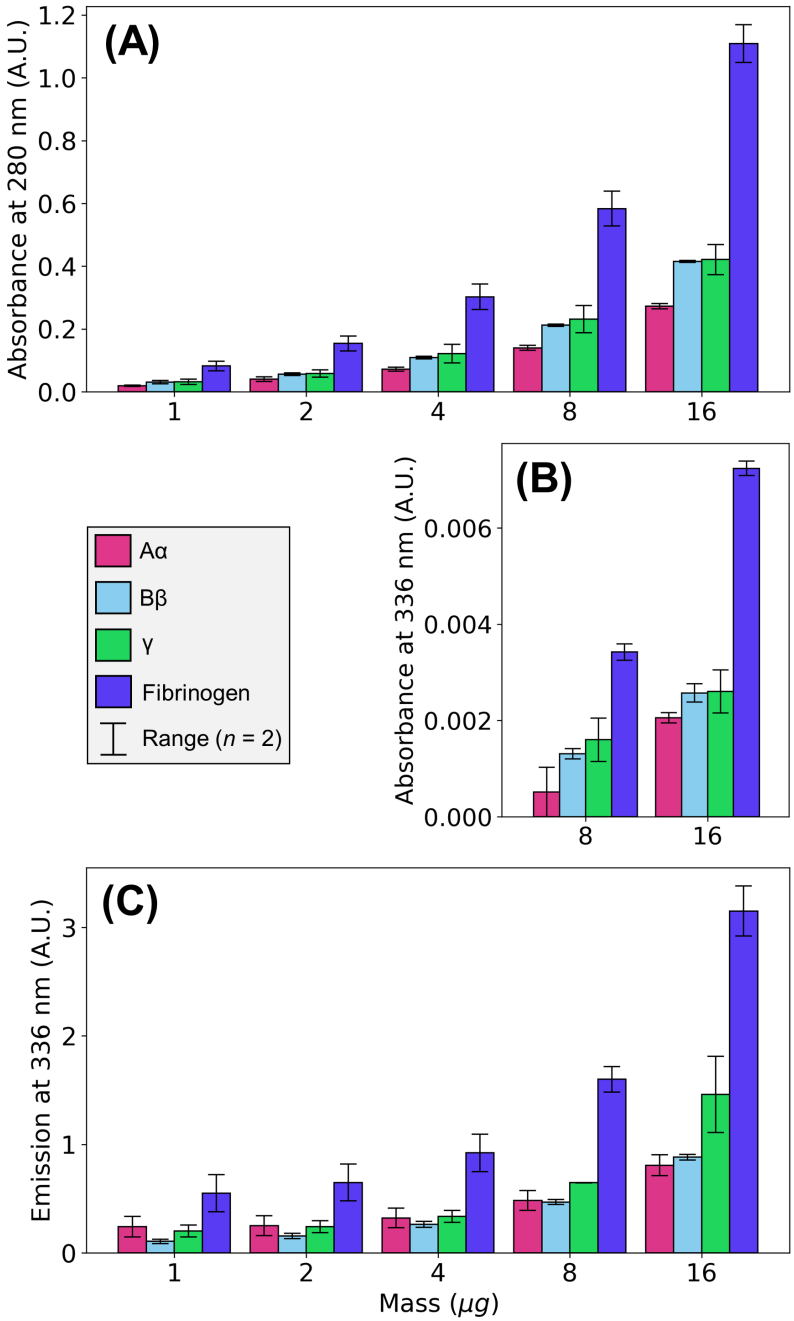


Figure A.1: HPLC peak areas of unfolded standards. **A**) Protein absorbance at 280 nm. **B**) Dye absorbance at 336 nm (many measurements were below the detector threshold, hence why only masses of 8 and 16 are shown). **C**) Dye emission at 336 nm. Averaged over biological repeats ($n = 2$).

Plunger speed (mm s^{-1})	Replicate 1: 28/09/2023		
	Flow loss correction	Standard loss correction	
		Control	Stressed
2	1.476	2.716	4.009
3	1.214	1.221	1.482
4	1.182	1.056	1.248
8	1.222	1.030	1.259
16	1.242	1.047	1.300
Replicate 2: 05/10/2023			
2	1.295	0.844	1.093
3	1.281	0.821	1.052
4	1.179	1.018	1.200
8	1.214	1.009	1.225
16	1.119	1.144	1.280

Table A.2: Correction factors that adjust the IAEDANS fluoresence emission at 336 nm based on the loss of protein during experiments. These are applied to each replicate before any averages are taken.

A.3 Correcting the IAEDANS fluoresence: a worked example

The following is a worked example of the corrections applied to IAEDANS fluorescence signals measured at a plunger speed of 16 mm s^{-1} for the biological replicate on 05/10/2023. The protein absorbances at 280 nm were measured as

$$\mathrm{Abs}_{280}^{C} = 0.155 \,, \qquad \mathrm{Abs}_{280}^{S} = 0.139 \,, \qquad \mathrm{Abs}_{280}^{\mathrm{Std}} = 0.177 \,,$$

and the initial IAEDANS emission at 336 nm as

$$\mathrm{Em}_{336}^{C} = 1.97 \,, \qquad \mathrm{Em}_{336}^{S} = 1.73 \ .$$

The correction factor for protein loss due to flow was calculated as

$$K_1 = \frac{0.155}{0.139} = 1.119 \ ,$$

and the two correction factors from the unfolded standards as

$$K_2^C = \frac{0.177}{0.155} = 1.144 \,, \qquad K_2^S = \frac{0.177}{0.139} = 1.280 \ ,$$

which are recorded in Table A.2.

The corrected fluorescence signals are therefore given by

$$\begin{aligned} \mathrm{Em}_{336}^{C*} &= K_2^C \times \mathrm{Em}_{336}^{C} \\ &= 1.144 \times 1.97 = 2.25 \\ \mathrm{Em}_{336}^{S*} &= K_1 \times K_2^S \times \mathrm{Em}_{336}^{S} \\ &= 1.119 \times 1.280 \times 1.73 = 2.48 \ , \end{aligned}$$

indicating in a 10% increase of fluorescence in the presence of flow.

Appendix B

Mathematics

B.1 Concentration regime and volume fraction of fibrinogen

The number density, n, of rod-like polymers in a solution is given by

$$n = cN_A \ ,$$

where c is the concentration (units mol m^{-3}) and $N_A = 6.02 \times 10^{23}$ is Avogadro's number (units mol^{-1}). From the experiments of Chapter 4, $c = 6$ μM $= 0.006$ mol m^{-3}, which gives $n = 3.61 \times 10^{21}$ m^{-3}.

Based on n and the physical dimensions of individual polymers, the concentration regime of the solution can be determined. Doi and Edwards §9.1 describes four regimes [54], but the two most relevant for fibrinogen are: *dilute*, where polymers are sufficiently far apart that they rotate freely with little to no collisions and *semi-dilute*, where polymers begin to interact such that their dynamics are affected. In general, a polymer solution is semi-dilute if

$$nL^3 > 1 \ ,$$

where L is the polymer length. The approximate dimensions of fibrinogen are $L = 45$ nm with cross-sectional radius, $R = 2.5$ nm, giving $nL^3 = 0.33$, placing it within the dilute concentration regime; for a semi-dilute solution, $n > 1 \times 10^{22}$ m^{-3}.

From Doi and Edwards §10.2, it can be further shown that fibrinogen molecules are not macroscopically aligned, i.e. the solution is *isotropic* [54]. For slender rod-like polymers, where $L \gg 2R$, the volume fraction is given by

$$\phi = n\frac{\pi(2R)^2 L}{4} \ ,$$

which for fibrinogen is small with $\phi = 0.013$. Above a critical volume fraction, given by

$$\phi^* \simeq 4\frac{2R}{L} \ ,$$

a polymer solution may undergo a phase transition into a *nematic* state, characterised by macroscopic alignment along a given axis. For fibrinogen, $\phi^* = 0.44$, hence $\phi \ll \phi^*$ and the solution is isotropic.

B.2 Maximum time step of a KOBRA simulation

The 3D mean-squared displacement of a particle experiencing thermal fluctuations is given by

$$\Delta r^2 = 6D\Delta t \ , \tag{B.1}$$

where Δr is some distance, D is the translational diffusion coefficient and Δt is the time step [227]. In KOBRA, rod dynamics evolve by applying forces to the nodes, and the diffusion coefficient is equivalent to that of a sphere,

$$D_{sphere} = \frac{k_B T}{6\pi\eta R} \ , \tag{B.2}$$

where k_B is Boltzmann's constant, T is the temperature, η is the viscosity, and R is the rod element radius [85].

In general, the smallest length scale in a computer simulation determines the largest viable time step, above which numerical instabilities may arise. For KOBRA rods, the smallest length is the element radius, and the upper limit on the time step is defined such that an element cannot randomly diffuse a distance greater than $\Delta r = \pm R$ in a single step.

If two elements are touching with surface-surface distance $s = 0$, then a simultaneous movement of R by both elements towards each other causes the element axis vectors, $\boldsymbol{p}$, to intersect. This scenario forces the simulation to crash, because further steric repulsion becomes impossible due to one element being 'trapped' inside the other.

Combining Equations B.1 and B.2 and setting $\Delta r^2 = R^2$ yields the condition for a stable time step,

$$\Delta t < \frac{\pi\eta}{k_B T} R^3 \ , \tag{B.3}$$

where $\eta = 10^{-3}$ Pa s and $T = 300$ K. For simulations of fibrinogen, $R = 2.5$ nm and $\Delta t < 11.84$ ns. For the lower length scale of the biological mesoscale and FFEA tetrahedral meshes [39], $R = 0.5$ nm and $\Delta t < 0.09$ ns. Equation B.3 is steric-limited and may be unnecessary for systems that are sparse, or where rods are less likely to diffuse towards each other (e.g. single-molecule pulling [180]).

B.3 Inertia may be neglected in diffusion at the mesoscale

In a system with a low Reynolds number, viscous forces dominate over inertial forces [48]. Consider a basic example of a spherical particle diffusing through homogeneous cytoplasm, subject to a viscous drag force. Rearranging Equation B.1 gives the characteristic (average) diffusion time for travelling a linear distance L of

$$\tau = \frac{L^2}{6D} \ . \tag{B.4}$$

So, comparing a small, globular protein such as ubiquitin ($R_{Ub} \sim 2$ nm [228]) with a vesicle ($R_{Ve} \sim 100$ nm), for a temperature $T = 310$ K and fluid viscosity $\eta = 10^{-3}$ Pa s [229] the diffusivities are $D_{Ub} = 100$ μm^2 s^{-1} and $D_{Ve} = 2$ μm^2 s^{-1} respectively, meaning they diffuse a distance of 1 μm in around $\tau_{Ub} = 5$ ms and $\tau_{Ve} = 250$ ms. To confirm this inertialess description of thermal diffusion, the timescale, τ_I, over which the inertia of a particle decays, is calculated as

$$\tau_I = \frac{m}{\zeta} = \frac{\frac{4}{3}\pi R^3 \rho}{6\pi\eta R} = \frac{2R^2\rho}{9\eta} \ , \tag{B.5}$$

where m is the particle mass and $\rho \sim 10^3$ kg m^{-3} is the mass density of a typical protein [230]. This yields $\tau_{I,Ub} = 1 \times 10^{-12}$ s and $\tau_{I,Ve} = 1 \times 10^{-9}$ s, which when used in Equation B.4 gives the length scales over which inertia decays as $L_{I,Ub} = 10$ pm and $L_{I,Ve} = 60$ pm, comparable to atomic bond lengths [146]. The effects of macromolecular inertia can hence be neglected for diffusion at the biological mesoscale.

References

[1] R. Rojas and U. Hashagen, Eds., *The First Computers: History and Architectures* (History of Computing). London, England: MIT Press, 2002.

[2] J. Shalf, "The future of computing beyond Moore's law," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 378, no. 2166, p. 20 190 061, 2020. DOI: 10.1098/rsta.2019.0061.

[3] T. Haigh, M. Priestley, and C. Rope, "Los Alamos bets on ENIAC: Nuclear Monte Carlo simulations, 1947-1948," *IEEE Annals of the History of Computing*, vol. 36, no. 3, pp. 42–63, 2014. DOI: 10.1109/mahc.2014.40.

[4] N. Metropolis and S. Ulam, "The Monte Carlo method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949. DOI: 10.1080/01621459.1949.10483310.

[5] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, no. 5612, pp. 585–590, 1977. DOI: 10.1038/267585a0.

[6] S. A. Adcock and J. A. McCammon, "Molecular dynamics: Survey of methods for simulating the activity of proteins," *Chemical Reviews*, vol. 106, no. 5, pp. 1589–1615, 2006. DOI: 10.1021/cr040426m.

[7] J. Jung *et al.*, "New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems," *Journal of Computational Chemistry*, vol. 42, no. 4, pp. 231–241, 2021. DOI: 10.1002/jcc.26450.

[8] J. Jung *et al.*, "Scaling molecular dynamics beyond 100, 000 processor cores for large-scale biophysical simulations," *Journal of Computational Chemistry*, vol. 40, no. 21, pp. 1919–1930, 2019. DOI: 10.1002/jcc.25840.

[9] A. Singharoy *et al.*, "Atoms to phenotypes: Molecular design principles of cellular energy metabolism," *Cell*, vol. 179, no. 5, 1098–1111.e23, 2019. DOI: 10.1016/j.cell.2019.10.021.

[10] S. A. Hollingsworth and R. O. Dror, "Molecular dynamics simulation for all," *Neuron*, vol. 99, no. 6, pp. 1129–1143, 2018. DOI: 10.1016/j.neuron.2018.08.011.

[11] P. Scheltens, B. De Strooper, M. Kivipelto, H. Holstege, G. Chételat, C. E. Teunissen, J. Cummings, and W. M. van der Flier, "Alzheimer's disease," *The Lancet*, vol. 397, no. 10284, pp. 1577–1590, 2021. DOI: 10.1016/s0140-6736(20)32205-4.

[12] M. J. Solares and D. F. Kelly, "Harnessing the power of structural oncology," *Microscopy Today*, vol. 30, no. 2, pp. 10–17, 2022. DOI: 10.1017/s1551929522000426.

[13] M. Grabowski *et al.*, "Rapid response to emerging biomedical challenges and threats," *IUCrJ*, vol. 8, no. 3, pp. 395–407, 2021. DOI: 10.1107/s2052252521003018.

[14] Z. Fu, L. Ouyang, R. Xu, Y. Yang, and W. Sun, "Responsive biomaterials for 3D bioprinting: A review," *Materials Today*, vol. 52, pp. 112–132, 2022. DOI: 10.1016/j.mattod.2022.01.001.

[15] T. Schlick and S. Portillo-Ledesma, "Biomolecular modeling thrives in the age of technology," *Nature Computational Science*, vol. 1, no. 5, pp. 321–331, 2021. DOI: 10.1038/s43588-021-00060-9.

[16] A. R. Leach, *Molecular modelling: Principles and applications*, 2nd ed. Pearson Education, 2001.

[17] T. Schlick *et al.*, "Biomolecular modeling and simulation: A prospering multidisciplinary field," *Annual Review of Biophysics*, vol. 50, no. 1, pp. 267–301, 2021. DOI: 10.1146/annurev-biophys-091720-102019.

[18] D. A. Case *et al.*, "AmberTools," *Journal of Chemical Information and Modeling*, vol. 63, no. 20, pp. 6183–6191, 2023. DOI: 10.1021/acs.jcim.3c01153.

[19] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1–2, pp. 19–25, 2015. DOI: 10.1016/j.softx.2015.06.001.

[20] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, "Scalable molecular dynamics with NAMD," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1781–1802, 2005. DOI: 10.1002/jcc.20289.

[21] P. Eastman *et al.*, "OpenMM 8: Molecular dynamics simulation with machine learning potentials," *The Journal of Physical Chemistry B*, vol. 128, no. 1, pp. 109–116, 2023. DOI: 10.1021/acs.jpcb.3c06662.

[22] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. van Gunsteren, "Definition and testing of the GROMOS force-field versions 54A7 and 54B7," *European Biophysics Journal*, vol. 40, no. 7, pp. 843–856, 2011. DOI: 10.1007/s00249-011-0700-9.

[23] B. R. Brooks *et al.*, "CHARMM: The biomolecular simulation program," *Journal of Computational Chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009. DOI: 10.1002/jcc.21287.

[24] W. L. Jorgensen and J. Tirado-Rives, "The OPLS potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," *Journal of the American Chemical Society*, vol. 110, no. 6, pp. 1657–1666, 1988. DOI: 10.1021/ja00214a001.

[25] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, "An overview of the Amber biomolecular simulation package," *WIREs Computational Molecular Science*, vol. 3, no. 2, pp. 198–210, 2012. DOI: 10.1002/wcms.1121.

[26] P. Szymczak and M. Cieplak, "Hydrodynamic effects in proteins," *Journal of Physics Condensed Matter*, vol. 23, no. 3, 2011. DOI: 10.1088/0953-8984/23/3/033102.

[27] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI force field: Coarse grained model for biomolecular simulations," *The Journal of Physical Chemistry B*, vol. 111, no. 27, pp. 7812–7824, 2007. DOI: 10.1021/jp071097f.

[28] A. Kolinski, "Protein modeling and structure prediction with a reduced representation," *Acta Biochimica Polonica*, vol. 51, no. 2, pp. 349–371, 2004.

[29] A. Liwo *et al.*, "A unified coarse-grained model of biological macromolecules based on mean-field multipole–multipole interactions," *Journal of Molecular Modeling*, vol. 20, no. 8, 2014. DOI: 10.1007/s00894-014-2306-5.

[30] A. E. Dawid, D. Gront, and A. Kolinski, "SURPASS low-resolution coarse-grained protein modeling," *Journal of Chemical Theory and Computation*, vol. 13, no. 11, pp. 5766–5779, 2017. DOI: 10.1021/acs.jctc.7b00642.

[31] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, "Coarse-grained protein models and their applications," *Chemical Reviews*, vol. 116, no. 14, pp. 7898–7936, 2016. DOI: 10.1021/acs.chemrev.6b00163.

[32] M. M. Teeter, "Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin," *Proceedings of the National Academy of Sciences*, vol. 81, no. 19, pp. 6014–6018, 1984. DOI: 10.1073/pnas.81.19.6014.

[33] S. Kmiecik, M. Kouza, A. Badaczewska-Dawid, A. Kloczkowski, and A. Kolinski, "Modeling of protein structural flexibility and large-scale dynamics: Coarse-grained simulations and elastic network models," *International Journal of Molecular Sciences*, vol. 19, no. 11, p. 3496, 2018. DOI: 10.3390/ijms19113496.

[34] W. M. Lai, D. Rubin, and E. Krempl, *Introduction to Continuum Mechanics*, 4th ed. Elsevier, 2010.

[35] M. R. Shaebani, A. Wysocki, R. G. Winkler, G. Gompper, and H. Rieger, "Computational models for active matter," *Nature Reviews Physics*, vol. 2, no. 4, pp. 181–199, 2020. DOI: 10.1038/s42254-020-0152-1.

[36] M. Gravett *et al.*, "Moving in the mesoscale: Understanding the mechanics of cytoskeletal molecular motors by combining mesoscale simulations with imaging," *WIREs Computational Molecular Science*, 2021. DOI: 10.1002/wcms.1570.

[37] A. Solernou, B. S. Hanson, R. A. Richardson, R. Welch, D. J. Read, O. G. Harlen, and S. A. Harris, "Fluctuating Finite Element Analysis (FFEA): A continuum mechanics software tool for mesoscale simulation of biomolecules," *PLoS Computational Biology*, vol. 14, no. 3, pp. 1–29, 2018. DOI: 10.1371/journal.pcbi.1005897.

[38] B. Hanson, R. Richardson, R. Oliver, D. J. Read, O. Harlen, and S. Harris, "Modelling biomacromolecular assemblies with continuum mechanics," *Biochemical Society Transactions*, vol. 43, pp. 186–192, 2015. DOI: 10.1042/BST20140294.

[39] B. S. Hanson, "Mesoscale modelling of cytoplasmic dynein using fluctuating finite element analysis," PhD thesis, University of Leeds, 2018.

[40] L. Maveyraud and L. Mourey, "Protein x-ray crystallography and drug discovery," *Molecules*, vol. 25, no. 5, p. 1030, 2020. DOI: 10.3390/molecules25051030.

[41] S. K. Burley *et al.*, "Electron microscopy holdings of the protein data bank: The impact of the resolution revolution, new validation tools, and implications for the

future," *Biophysical Reviews*, vol. 14, no. 6, pp. 1281–1301, 2022. DOI: 10.1007/s12551-022-01013-w.

[42] D. J. Huggins *et al.*, "Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 9, no. 3, pp. 1–23, 2019. DOI: 10.1002/wcms.1393.

[43] A. J. Roberts, T. Kon, P. J. Knight, K. Sutoh, and S. A. Burgess, "Functions and mechanics of dynein motor proteins," *Nature Reviews Molecular Cell Biology*, vol. 14, no. 11, pp. 713–726, 2013. DOI: 10.1038/nrm3667.

[44] L. Li, J. Alper, and E. Alexov, "Cytoplasmic dynein binding, run length, and velocity are guided by long-range electrostatic interactions," *Scientific Reports*, vol. 6, 2016. DOI: 10.1038/srep31523.

[45] N. Kamiya, T. Mashimo, Y. Takano, T. Kon, G. Kurisu, and H. Nakamura, "Elastic properties of dynein motor domain obtained from all-atom molecular dynamics simulations," *Protein Engineering Design and Selection*, vol. 29, no. 8, pp. 317–325, 2016. DOI: 10.1093/protein/gzw022.

[46] A. Gray, O. G. Harlen, S. A. Harris, S. Khalid, Y. M. Leung, R. Lonsdale, A. J. Mulholland, A. R. Pearson, D. J. Read, and R. A. Richardson, "In pursuit of an accurate spatial and temporal model of biomolecules at the atomistic level: a perspective on computer simulation," *Acta Crystallographica Section D Biological Crystallography*, vol. 71, no. 1, pp. 162–172, 2015. DOI: 10.1107/S1399004714026777.

[47] D. Needleman and M. Shelley, "The stormy fluid dynamics of the living cell," *Physics Today*, vol. 72, no. 9, pp. 33–38, 2019. DOI: 10.1063/PT.3.4292.

[48] E. M. Purcell, "Life at low Reynolds number," *American Journal of Physics*, vol. 45, no. 1, pp. 3–11, 1977. DOI: 10.1119/1.10903.

[49] L. F. Brass and S. L. Diamond, "Transport physics and biorheology in the setting of hemostasis and thrombosis," *Journal of Thrombosis and Haemostasis*, vol. 14, no. 5, pp. 906–917, 2016. DOI: 10.1111/jth.13280.

[50] V. Calabrese, S. Varchanis, S. J. Haward, and A. Q. Shen, "Alignment of colloidal rods in crowded environments," *Macromolecules*, vol. 55, no. 13, pp. 5610–5620, 2022. DOI: 10.1021/acs.macromol.2c00769.

[51] H. T. Bergal, Y. Jiang, D. Yang, T. A. Springer, and W. P. Wong, "Conformation of von Willebrand factor in shear flow revealed with stroboscopic single-molecule

imaging," *Blood*, vol. 140, no. 23, pp. 2490–2499, 2022. DOI: 10.1182/blood.2022016969.

[52] R. A. Richardson, B. S. Hanson, D. J. Read, O. G. Harlen, and S. A. Harris, "Exploring the dynamics of flagellar dynein within the axoneme with Fluctuating Finite Element Analysis," *Quarterly Reviews of Biophysics*, vol. 53, e9, 2020. DOI: 10.1017/S0033583520000062.

[53] R. Kubo, "The fluctuation-dissipation theorem," *Reports on Progress in Physics*, vol. 29, no. 1, p. 306, 1966. DOI: 10.1088/0034-4885/29/1/306.

[54] M. Doi and S. F. Edwards, *The theory of polymer dynamics*. Oxford: Clarendon Press, 1988.

[55] F. Hong, F. Zhang, Y. Liu, and H. Yan, "DNA origami: Scaffolds for creating higher order structures," *Chemical Reviews*, vol. 117, no. 20, pp. 12 584–12 640, 2017. DOI: 10.1021/acs.chemrev.6b00825.

[56] H. L. Sweeney and E. L. Holzbaur, "Motor proteins," *Cold Spring Harbor Perspectives in Biology*, vol. 10, no. 5, a021931, 2018. DOI: 10.1101/cshperspect.a021931.

[57] A. J. Gale, "Current understanding of hemostasis," *Toxicologic Pathology*, vol. 39, no. 1, pp. 273–280, 2010. DOI: 10.1177/0192623310389474.

[58] T. A. Springer, "Von Willebrand factor, Jedi knight of the bloodstream," *Blood*, vol. 124, no. 9, pp. 1412–1425, 2014. DOI: 10.1182/blood-2014-05-378638.

[59] B. S. Lazarus, C. Chadha, A. Velasco-Hogan, J. D. Barbosa, I. Jasiuk, and M. A. Meyers, "Engineering with keratin: A functional material and a source of bioinspiration," *iScience*, vol. 24, no. 8, p. 102 798, 2021. DOI: 10.1016/j.isci.2021.102798.

[60] M. D. Shoulders and R. T. Raines, "Collagen structure and stability," *Annual Review of Biochemistry*, vol. 78, no. 1, pp. 929–958, 2009. DOI: 10.1146/annurev.biochem.77.032207.120833.

[61] T. Feller, S. D. Connell, and R. A. Ariëns, "Why fibrin biomechanical properties matter for hemostasis and thrombosis," *Journal of Thrombosis and Haemostasis*, vol. 20, no. 1, pp. 6–16, 2022. DOI: 10.1111/jth.15531.

[62] Y. Habibi, L. A. Lucia, and O. J. Rojas, "Cellulose nanocrystals: Chemistry, self-assembly, and applications," *Chemical Reviews*, vol. 110, no. 6, pp. 3479–3500, 2010. DOI: 10.1021/cr900339w.

[63] J. Hou, B. E. Aydemir, and A. G. Dumanli, "Understanding the structural diversity of chitins as a versatile biomaterial," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2206, 2021. DOI: 10.1098/rsta.2020.0331.

[64] M. D. G. Hughes *et al.*, "Building block aspect ratio controls assembly, architecture, and mechanics of synthetic and natural protein networks," *Nature Communications*, vol. 14, no. 1, 2023. DOI: 10.1038/s41467-023-40921-7.

[65] P. G. De Gennes, "Coil-stretch transition of dilute flexible polymers under ultrahigh velocity gradients," *The Journal of Chemical Physics*, vol. 5042, no. 1974, pp. 5030–5042, 1974. DOI: 10.1063/1.1681018.

[66] B. S. Hanson, D. Head, and L. Dougan, "The hierarchical emergence of worm-like chain behaviour from globular domain polymer chains," *Soft Matter*, vol. 15, no. 43, pp. 8778–8789, 2019. DOI: 10.1039/c9sm01656b.

[67] G. Corazza and R. Singh, "Unraveling looping efficiency of stochastic cosserat polymers," *Physical Review Research*, vol. 4, no. 1, 2022. DOI: 10.1103/physrevresearch.4.013071.

[68] M. Bergou, M. Wardetzky, S. Robinson, B. Audoly, and E. Grinspun, "Discrete elastic rods," *ACM Transactions on Graphics*, vol. 27, no. 3, 2008. DOI: 10.1145/1360612.1360662.

[69] S. L. Freedman, C. Suarez, J. D. Winkelman, D. R. Kovar, G. A. Voth, A. R. Dinner, and G. M. Hocky, "Mechanical and kinetic factors drive sorting of F-actin cross-linkers on bundles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 33, pp. 16 192–16 197, 2019. DOI: 10.1073/pnas.1820814116.

[70] S. L. Freedman, S. Banerjee, G. M. Hocky, and A. R. Dinner, "A Versatile Framework for Simulating the Dynamic Mechanical Structure of Cytoskeletal Networks," *Biophysical Journal*, vol. 113, no. 2, pp. 448–460, 2017. DOI: 10.1016/j.bpj.2017.06.003.

[71] K. Popov, J. Komianos, and G. A. Papoian, "MEDYAN: Mechanochemical Simulations of Contraction and Polarity Alignment in Actomyosin Networks," *PLoS Computational Biology*, vol. 12, no. 4, pp. 1–35, 2016. DOI: 10.1371/journal.pcbi.1004877.

[72] J. Rickman, F. Nédélec, and T. Surrey, "Effects of spatial dimensionality and steric interactions on microtubule-motor self-organization," *Physical Biology*, vol. 16, no. 4, 2019. DOI: 10.1088/1478-3975/ab0fb1.

[73] H. Ni and G. A. Papoian, "Membrane-MEDYAN: Simulating deformable vesicles containing complex cytoskeletal networks," *The Journal of Physical Chemistry B*, vol. 125, no. 38, pp. 10 710–10 719, 2021. DOI: 10.1021/acs.jpcb.1c02336.

[74] X. Li, Q. Ni, X. He, J. Kong, S. M. Lim, G. A. Papoian, J. P. Trzeciakowski, A. Trache, and Y. Jiang, "Tensile force-induced cytoskeletal remodeling: Mechanics before chemistry," *PLoS Computational Biology*, vol. 16, no. 6, pp. 1–19, 2020. DOI: 10.1371/journal.pcbi.1007693.

[75] D. B. Cortes, M. Gordon, F. Nédélec, and A. S. Maddox, "Bond Type and Discretization of Nonmuscle Myosin II Are Critical for Simulated Contractile Dynamics," *Biophysical Journal*, vol. 118, no. 11, pp. 2703–2717, 2020. DOI: 10.1016/j.bpj.2020.03.033.

[76] G. K. Batchelor, "Slender-body theory for particles of arbitrary cross-section in Stokes flow," *Journal of Fluid Mechanics*, vol. 44, no. 3, pp. 419–440, 1970. DOI: 10.1017/S002211207000191X.

[77] Y. Man, L. Koens, and E. Lauga, "Hydrodynamic interactions between nearby slender filaments," *EPL*, vol. 116, no. 2, 2016. DOI: 10.1209/0295-5075/116/24002.

[78] O. Maxian, B. Sprinkle, C. S. Peskin, and A. Donev, "Hydrodynamics of a twisting, bending, inextensible fiber in stokes flow," *Physical Review Fluids*, vol. 7, no. 7, 2022. DOI: 10.1103/physrevfluids.7.074101.

[79] R. G. Winkler and G. Gompper, "The physics of active polymers and filaments," *Journal of Chemical Physics*, vol. 153, no. 4, 2020. DOI: 10.1063/5.0011466.

[80] L. J. Fauci and C. S. Peskin, "A computational model of aquatic animal locomotion," *Journal of Computational Physics*, vol. 77, no. 1, pp. 85–108, 1988. DOI: 10.1016/0021-9991(88)90158-1.

[81] W. Huang and M. Khalid Jawed, "Numerical simulation of bundling of helical elastic rods in a viscous fluid," *Computers & Fluids*, vol. 228, p. 105 038, 2021. DOI: 10.1016/j.compfluid.2021.105038.

[82] M. Gazzola, L. H. Dudte, A. G. McCormick, and L. Mahadevan, "Forward and inverse problems in the mechanics of soft filaments," *Royal Society Open Science*, vol. 5, no. 6, p. 171 628, 2018. DOI: 10.1098/rsos.171628.

[83] H. Du, Q. Jiang, and W. Xiong, "Computer-assisted assembly process planning for the installation of flexible cables modeled according to a viscoelastic cosserat rod model," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 237, no. 11, pp. 1737–1747, 2022. DOI: 10.1177/09544054221136000.

[84] Z. Wang, M. Fratarcangeli, A. Ruimi, and A. Srinivasa, "Real time simulation of inextensible surgical thread using a kirchhoff rod model with force output for haptic feedback applications," *International Journal of Solids and Structures*, vol. 113–114, pp. 192–208, 2017. DOI: 10.1016/j.ijsolstr.2017.02.017.

[85] R. Welch, S. A. Harris, O. G. Harlen, and D. J. Read, "KOBRA: A fluctuating elastic rod model for slender biological macromolecules," *Soft Matter*, vol. 16, no. 32, pp. 7544–7555, 2020. DOI: 10.1039/d0sm00491j.

[86] S. Stamboroski, A. Joshi, P. L. M. Noeske, S. Köppen, and D. Brüggemann, "Principles of Fibrinogen Fiber Assembly In Vitro," *Macromolecular Bioscience*, vol. 21, no. 5, 2021. DOI: 10.1002/mabi.202000412.

[87] J. W. Weisel and R. I. Litvinov, "Mechanisms of fibrin polymerization and clinical implications," *Blood*, vol. 121, no. 10, pp. 1712–1719, 2013. DOI: 10.1182/blood-2012-09-306639.

[88] D. Gregory and D. Marshall, *Scanning electron micrograph of blood clot*, https://wellcomecollection.org/works/pd54cr22, Accessed 01/04/2024.

[89] S. S. Virani *et al.*, "Heart disease and stroke statistics—2020 update: A report from the american heart association," *Circulation*, vol. 141, no. 9, 2020. DOI: 10.1161/cir.0000000000000757.

[90] V. P. Memtsas, D. R. Arachchillage, and D. A. Gorog, "Role, laboratory assessment and clinical relevance of fibrin, factor XIII and endogenous fibrinolysis in arterial and venous thrombosis," *International Journal of Molecular Sciences*, vol. 22, no. 3, pp. 1–19, 2021. DOI: 10.3390/ijms22031472.

[91] Kattula S, B. JR, and Wolberg AS, "Fibrinogen and fibrin in hemostasis and thrombosis," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 37, no. 3, e13–e21, 2017. DOI: 10.1161/ATVBAHA.117.308564.Fibrinogen.

[92] S. S. Acharya and D. M. Dimichele, "Rare inherited disorders of fibrinogen," *Haemophilia*, vol. 14, no. 6, pp. 1151–1158, 2008. DOI: 10.1111/j.1365-2516.2008.01831.x.

[93] L. C. Serpell, M. Benson, J. J. Liepnieks, and P. E. Fraser, "Structural analyses of fibrinogen amyloid fibrils," *Amyloid*, vol. 14, no. 3, pp. 199–203, 2007. DOI: 10.1080/13506120701461111.

[94] A. Henschen, F. Lottspeich, M. Kehl, and C. Southan, "Covalent structure of fibrinogen," *Ann. N. Y. Acad. Sci.*, vol. 408, no. 1, pp. 28–43, 1983.

[95] J. M. Kollman, L. Pandi, M. R. Sawaya, M. Riley, and R. F. Doolittle, "Crystal structure of human fibrinogen," *Biochemistry*, vol. 48, no. 18, pp. 3877–3886, 2009. DOI: 10.1021/bi802205g.

[96] R. A. Burton, G. Tsurupa, R. R. Hantgan, N. Tjandra, and L. Medved, "NMR solution structure, stability, and interaction of the recombinant bovine fibrinogen αC-domain fragment," *Biochemistry*, vol. 46, no. 29, pp. 8550–8560, 2007. DOI: 10.1021/bi700606v.

[97] Y. F. Zuev, R. I. Litvinov, A. E. Sitnitsky, B. Z. Idiyatullin, D. R. Bakirova, D. K. Galanakis, A. Zhmurov, V. Barsegov, and J. W. Weisel, "Conformational Flexibility and Self-Association of Fibrinogen in Concentrated Solutions," *Journal of Physical Chemistry B*, vol. 121, no. 33, pp. 7833–7843, 2017. DOI: 10.1021/acs.jpcb.7b05654.

[98] G. Tsurupa, L. Tsonev, and L. Medved, "Structural Organization of the Fibrin(ogen) αC-Domain," *Biochemistry*, vol. 41, no. 20, pp. 6449–6459, 2002. DOI: 10.1021/bi025584r.

[99] Y. I. Veklich, O. V. Gorkun, L. V. Medved', W. Nieuwenhuizen, and J. W. Weisel, "Carboxyl-terminal portions of the α chains of fibrinogen and fibrin: Localization by electron microscopy and the effects of isolated αC fragments on polymerization," *Journal of Biological Chemistry*, vol. 268, no. 18, pp. 13 577–13 585, 1993. DOI: 10.1016/s0021-9258(19)38688-0.

[100] A. D. Protopopova, N. A. Barinov, E. G. Zavyalova, A. M. Kopylov, V. I. Sergienko, and D. V. Klinov, "Visualization of fibrinogen αC regions and their arrangement during fibrin network formation by high-resolution AFM," *Journal of Thrombosis and Haemostasis*, vol. 13, no. 4, pp. 570–579, 2015. DOI: 10.1111/jth.12785.

[101] J. F. Ross, *Private communication*, University of Leeds, 2023.

[102] A. D. Protopopova, R. I. Litvinov, D. K. Galanakis, C. Nagaswami, N. A. Barinov, A. R. Mukhitov, D. V. Klinov, and J. W. Weisel, "Morphometric characterization of fibrinogen's αC regions and their role in fibrin self-assembly and molecular organization," *Nanoscale*, vol. 9, no. 36, pp. 13 707–13 716, 2017. DOI: 10.1039/c7nr04413e.

[103] J. J. de Vries *et al.*, "Levels of fibrinogen variants are altered in severe COVID-19," *TH Open*, vol. 07, no. 03, e217–e225, 2023. DOI: 10.1055/a-2102-4521.

[104] B. Holm, F. Brosstad, P. Kierulf, and H. C. Godal, "Polymerization properties of two normally circulating fibrinogens, HMW and LMW. evidence that the COOH-terminal end of the a-chain is of importance for fibrin polymerization," *Thrombosis Research*, vol. 39, pp. 595–606, 5 1985. DOI: 10.1016/0049-3848(85)90239-7.

[105] K. A. Jansen *et al.*, "Molecular packing structure of fibrin fibers resolved by X-ray scattering and molecular modeling," *Soft Matter*, vol. 16, no. 35, pp. 8272–8283, 2020. DOI: 10.1039/d0sm00916d.

[106] A. Zhmurov, A. D. Protopopova, R. I. Litvinov, P. Zhukov, J. W. Weisel, and V. Barsegov, "Atomic Structural Models of Fibrin Oligomers," *Structure*, vol. 26, no. 6, pp. 857–868, 2018. DOI: 10.1016/j.str.2018.04.005.

[107] I. N. Chernysh, C. Nagaswami, and J. W. Weisel, "Visualization and identification of the structures formed during early stages of fibrin polymerization," *Blood*, vol. 117, no. 17, pp. 4609–4614, 2011. DOI: 10.1182/blood-2010-07-297671.

[108] I. N. Chernysh and J. W. Weisel, "Dynamic imaging of fibrin network formation correlated with other measures of polymerization," *Blood*, vol. 111, no. 10, pp. 4854–4861, 2008. DOI: 10.1182/blood-2007-08-105247.

[109] N. Badiei *et al.*, "Effects of unidirectional flow shear stresses on the formation, fractal microstructure and rigidity of incipient whole blood clots and fibrin gels," *Clinical Hemorheology and Microcirculation*, vol. 60, no. 4, pp. 451–464, 2015. DOI: 10.3233/CH-151924.

[110] R. I. Litvinov, O. Kononova, A. Zhmurov, K. A. Marx, V. Barsegov, D. Thirumalai, and J. W. Weisel, "Regulatory element in fibrin triggers tension-activated transition from catch to slip bonds," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 34, pp. 8075–8080, 2018. DOI: 10.1073/pnas.1802576115.

[111] S. Kim and S. J. Karrila, *Microhydrodynamics: Principles and selected applications*. 1992. DOI: 10.1016/0955-7997(92)90082-i.

[112] N. A. Chebotareva, B. I. Kurganov, and N. B. Livanova, "Biochemical effects of molecular crowding," *Biochemistry (Moscow)*, vol. 69, no. 11, pp. 1239–1251, 2004. DOI: 10.1007/s10541-005-0070-y.

[113] P. C. Sousa, R. Vaz, A. Cerejo, M. S. N. Oliveira, M. A. Alves, and F. T. Pinho, "Rheological behavior of human blood in uniaxial extensional flow," *Journal of Rheology*, vol. 62, no. 2, pp. 447–456, 2018. DOI: 10.1122/1.4998704.

[114] S. Varchanis, Y. Dimakopoulos, C. Wagner, and J. Tsamopoulos, "How viscoelastic is human blood plasma?" *Soft Matter*, vol. 14, no. 21, pp. 4238–4251, 2018. DOI: 10.1039/c8sm00061a.

[115] E. F. Yeo, J. M. Oliver, N. Korin, and S. L. Waters, "A continuum model for the elongation and orientation of von willebrand factor with applications in arterial flow," *Biomechanics and Modeling in Mechanobiology*, vol. 23, no. 4, pp. 1299–1317, 2024. DOI: 10.1007/s10237-024-01840-8.

[116] I. B. Bekard, P. Asimakis, J. Bertolini, and D. E. Dunstan, "The effects of shear flow on protein structure and function," *Biopolymers*, vol. 95, no. 11, pp. 733–745, 2011. DOI: 10.1002/bip.21646.

[117] R. A. Campbell, M. M. Aleman, L. D. Gray, M. R. Falvo, and A. S. Wolberg, "Flow profoundly influences fibrin network structure: Implications for fibrin formation and clot stability in haemostasis," *Thrombosis and Haemostasis*, vol. 104, no. 12, pp. 1281–1284, 2010. DOI: 10.1160/TH10-07-0442.

[118] K. S. Sakariassen, L. Orning, and V. T. Turitto, "The impact of blood shear rate on arterial thrombus formation," *Future Science OA*, vol. 1, no. 4, 2015. DOI: 10.4155/fso.15.28.

[119] R. S. Reneman and A. P. G. Hoeks, "Wall shear stress as measured in vivo: Consequences for the design of the arterial system," *Medical and Biological Engineering and Computing*, vol. 46, no. 5, pp. 499–507, 2008. DOI: 10.1007/s11517-008-0330-2.

[120] A. Rana, E. Westein, B. Niego, and C. E. Hagemeyer, "Shear-dependent platelet aggregation: Mechanisms and therapeutic opportunities," *Frontiers in Cardiovascular Medicine*, vol. 6, 2019. DOI: 10.3389/fcvm.2019.00141.

[121] J. Strony, A. Beaudoin, D. Brands, and B. Adelman, "Analysis of shear stress and hemodynamic factors in a model of coronary artery stenosis and thrombosis," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 265, no. 5, H1787–H1796, 1993. DOI: 10.1152/ajpheart.1993.265.5.h1787.

[122] K. C. Gersh, K. E. Edmondson, and J. W. Weisel, "Flow rate and fibrin fiber alignment," *Journal of Thrombosis and Haemostasis*, vol. 8, no. 12, pp. 2826–2828, 2010. DOI: 10.1111/j.1538-7836.2010.04118.x.

[123] I. Varjú, P. Sótonyi, R. Machovich, L. Szabó, K. Tenekedjiev, M. M. Silva, C. Longstaff, and K. Kolev, "Hindered dissolution of fibrin formed under mechanical stress," *Journal of Thrombosis and Haemostasis*, vol. 9, no. 5, pp. 979–986, 2011. DOI: 10.1111/j.1538-7836.2011.04203.x.

[124] D. Butera and P. J. Hogg, "Fibrinogen function achieved through multiple covalent states," *Nature Communications*, vol. 11, no. 1, pp. 1–10, 2020. DOI: 10.1038/s41467-020-19295-7.

[125] M. S. Gravett, D. P. Klebl, O. G. Harlen, D. J. Read, S. A. Harris, S. P. Muench, and M. Peckham, "Exploiting cryo-EM structures of actomyosin-5a to reveal the physical properties of its lever," *bioRxiv*, 2023. DOI: 10.1101/2023.03.19.533260.

[126] M. S. C. Gravett, "Understanding the mechanics of myosin through molecular modelling and time-resolved cryoEM," PhD thesis, University of Leeds, 2022.

[127] R. Welch, "KOBRA: A fluctuating elastic rod model for slender biological macromolecules," PhD thesis, University of Leeds, 2020.

[128] A. Shkurti, R. Goni, P. Andrio, E. Breitmoser, I. Bethune, M. Orozco, and C. A. Laughton, "pyPcazip: A PCA-based toolkit for compression and analysis of molecular simulation data," *SoftwareX*, vol. 5, pp. 44–50, 2016. DOI: 10.1016/j.softx.2016.04.002.

[129] G. M. Alushin, V. H. Ramey, S. Pasqualato, D. A. Ball, N. Grigorieff, A. Musacchio, and E. Nogales, "The Ndc80 kinetochore complex forms oligomeric arrays along microtubules," *Nature*, vol. 467, no. 7317, pp. 805–810, 2010. DOI: 10.1038/nature09423.

[130] F. Lampert, C. Mieck, G. M. Alushin, E. Nogales, and S. Westermann, "Molecular requirements for the formation of a kinetochore-microtubule interface by Dam1 and Ndc80 complexes," *Journal of Cell Biology*, vol. 200, no. 1, pp. 21–30, 2013. DOI: 10.1083/jcb.201210091.

[131] V. A. Volkov, P. J. Huis In 'T Veld, M. Dogterom, and A. Musacchio, "Multivalency of Ndc80 in the outer kinetochore is essential to track shortening microtubules and generate forces," *eLife*, vol. 7, pp. 1–22, 2018. DOI: 10.7554/eLife.36764.

[132] C. Ciferri *et al.*, "Implications for Kinetochore-Microtubule Attachment from the Structure of an Engineered Ndc80 Complex," *Cell*, vol. 133, no. 3, pp. 427–439, 2008. DOI: 10.1016/j.cell.2008.03.020.

[133] Y. N. Dimitrova, S. Jenni, R. Valverde, Y. Khin, and S. C. Harrison, "Structure of the MIND Complex Defines a Regulatory Focus for Yeast Kinetochore Assembly," *Cell*, vol. 167, no. 4, pp. 1014–1027, 2016. DOI: 10.1016/j.cell.2016.10.011.

[134] S. Gonen, B. Akiyoshi, M. G. Iadanza, D. Shi, N. Duggan, S. Biggins, and T. Gonen, "The structure of purified kinetochores reveals multiple microtubule-attachment sites," *Nature Structural & Molecular Biology*, vol. 19, no. 9, pp. 925–929, 2012. DOI: 10.1038/nsmb.2358.

[135] M. J. Tyska and M. S. Mooseker, "Myosin-V motility: These levers were made for walking," *Trends in Cell Biology*, vol. 13, no. 9, pp. 447–451, 2003. DOI: 10.1016/S0962-8924(03)00172-7.

[136] A. T. Lombardo, S. R. Nelson, G. G. Kennedy, K. M. Trybus, S. Walcott, and D. M. Warshaw, "Myosin Va transport of liposomes in three-dimensional actin networks is modulated by actin filament density, position, and polarity," *Proceedings of the National Academy of Sciences*, vol. 116, no. 17, pp. 8326–8335, 2019. DOI: 10.1073/pnas.1901176116.

[137] J. A. Hammer and J. R. Sellers, "Walking to work: Roles for class V myosins as cargo transporters," *Nature Reviews Molecular Cell Biology*, vol. 13, no. 1, pp. 13–26, 2012. DOI: 10.1038/nrm3248.

[138] S. R. McGuffee and A. H. Elcock, "Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm," *PLoS Computational Biology*, vol. 6, no. 3, 2010. DOI: 10.1371/journal.pcbi.1000694.

[139] P. Zakharov, N. Gudimchuk, V. Voevodin, A. Tikhonravov, F. I. Ataullakhanov, and E. L. Grishchuk, "Molecular and Mechanical Causes of Microtubule Catastrophe and Aging," *Biophysical Journal*, vol. 109, no. 12, pp. 2574–2591, 2015. DOI: 10.1016/j.bpj.2015.10.048.

[140] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed. Academic Press, 2002. DOI: 10.1016/B978-012267351-1/50008-0.

[141] S. Bindgen, F. Weik, R. Weeber, E. Koos, and P. De Buyl, "Lees-Edwards boundary conditions for translation invariant shear flow: Implementation and transport properties," *Physics of Fluids*, vol. 33, no. 8, 2021. DOI: 10.1063/5.0055396.

[142] T. Ridley, "Computer simulations of dynamics, structure and rheology of packed soft colloids," PhD thesis, University of Leeds, 2020.

[143] D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford, "Pulling geometry defines the mechanical resistance of a β-sheet protein," *Nature Structural Biology*, vol. 10, no. 9, pp. 731–737, 2003. DOI: 10.1038/nsb968.

[144] S. E. Harding, G. Channell, and M. K. Phillips-Jones, "The discovery of hydrogen bonds in DNA and a re-evaluation of the 1948 Creeth two-chain model for its structure," *Biochemical Society Transactions*, vol. 46, no. 5, pp. 1171–1182, 2018. DOI: 10.1042/bst20180158.

[145] I. K. McDonald and J. M. Thornton, "Satisfying hydrogen bonding potential in proteins," *Journal of Molecular Biology*, vol. 238, no. 5, pp. 777–793, 1994. DOI: 10.1006/jmbi.1994.1334.

[146] E. Kraka and D. Cremer, "Characterization of CF bonds with multiple-bond character: Bond lengths, stretching force constants, and bond dissociation energies," *ChemPhysChem*, vol. 10, no. 4, pp. 686–698, 2009. DOI: 10.1002/cphc.200800699.

[147] C.-H. Chu *et al.*, "Beyond the Debye length in high ionic strength solution: Direct protein detection with field-effect transistors (FETs) in human serum," *Scientific Reports*, vol. 7, no. 1, 2017. DOI: 10.1038/s41598-017-05426-6.

[148] S. Yesudasan, X. Wang, and R. D. Averett, "Coarse-grained molecular dynamics simulations of fibrin polymerization: Effects of thrombin concentration on fibrin clot structure," *Journal of Molecular Modeling*, vol. 24, no. 5, 2018. DOI: 10.1007/s00894-018-3642-7.

[149] F. Ganovelli, F. Ponchio, and C. Rocchini, "Fast Tetrahedron-Tetrahedron Overlap Algorithm," *Journal of Graphics Tools*, vol. 7, no. 2, pp. 17–25, 2002. DOI: 10.1080/10867651.2002.10487557.

[150] Fnord and P. Dukhov, *Shortest distance between two line segments*, https://stackoverflow.com/questions/2824478/shortest-distance-between-two-line-segments, Accessed 21/02/2024.

[151] M. Mandelkern, J. G. Elias, D. Eden, and D. M. Crothers, "The dimensions of DNA in solution," *Journal of Molecular Biology*, vol. 152, no. 1, pp. 153–161, 1981. DOI: 10.1016/0022-2836(81)90099-1.

[152] U. K. Deiters, "Efficient coding of the minimum image convention," *Zeitschrift fur Physikalische Chemie*, vol. 227, no. 2-3, pp. 345–352, 2013. DOI: 10.1524/zpch.2013.0311.

[153] A. Dubey, "Good practices for high-quality scientific computing," *Computing in Science and Engineering*, vol. 24, no. 6, pp. 72–76, 2022. DOI: 10.1109/mcse.2023.3259259.

[154] G. J. Myers, C. Sandler, and T. Badgett, *The art of software testing*, 3rd ed. John Wiley and Sons, 2011.

[155] D. Thomas and A. Hunt, *The Pragmatic Programmer*, 2nd ed. Addison-Wesley, 2020.

[156] C. Yeromonahos, B. Polack, and F. Caton, "Nanostructure of the fibrin clot," *Biophysical Journal*, vol. 99, no. 7, pp. 2018–2027, 2010. DOI: 10.1016/j.bpj.2010.04.059.

[157] Schrödinger, LLC, "The PyMOL Molecular Graphics System, Version 2.5.0," 2021.

[158] G. Hager and G. Wellein, *Introduction to High Performance Computing for Scientists and Engineers*. Boca Raton, FL, USA: CRC Press, 2011.

[159] G. M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities," in *Proceedings of the April 18-20, 1967, spring joint computer conference on - AFIPS '67 (Spring)*, ser. AFIPS '67 (Spring), ACM Press, 1967. DOI: 10.1145/1465482.1465560.

[160] J. L. Gustafson, "Reevaluating Amdahl's law," *Commun. ACM*, vol. 31, no. 5, pp. 532–533, 1988. DOI: 10.1145/42411.42415.

[161] D. E. Shaw *et al.*, "Anton 3: Twenty microseconds of molecular dynamics simulation before lunch," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC '21, ACM, 2021. DOI: 10.1145/3458817.3487397.

[162] L. V. Bock, S. Gabrielli, M. H. Kolář, and H. Grubmüller, "Simulation of complex biomolecular systems: The ribosome challenge," *Annual Review of Biophysics*, vol. 52, no. 1, pp. 361–390, 2023. DOI: 10.1146/annurev-biophys-111622-091147.

[163] S. Köhler, F. Schmid, and G. Settanni, "The Internal Dynamics of Fibrinogen and Its Implications for Coagulation and Adsorption," *PLoS Computational Biology*, vol. 11, no. 9, pp. 1–19, 2015. DOI: 10.1371/journal.pcbi.1004346.

[164] Z. Sovova, J. Suttnar, and J. E. Dyr, "Molecular dynamic simulations suggest that metabolite-induced post-translational modifications alter the behavior of the fibrinogen coiled-coil domain," *Metabolites*, vol. 11, no. 5, 2021. DOI: 10.3390/metabo11050307.

[165] V. Calandrini, E. Pellegrini, P. Calligari, K. Hinsen, and G. Kneller, "nMoldyn - Interfacing spectroscopic experiments, molecular dynamics simulations and models for time correlation functions," *École thématique de la Société Française de la Neutronique*, vol. 12, pp. 201–232, 2011. DOI: 10.1051/sfn/201112010.

[166] J. Dobson *et al.*, "Inducing protein aggregation by extensional flow," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 18, pp. 4673–4678, 2017. DOI: 10.1073/pnas.1702724114.

[167] F. Grigolato and P. Arosio, "Synergistic effects of flow and interfaces on antibody aggregation," *Biotechnology and Bioengineering*, vol. 117, no. 2, pp. 417–428, 2020. DOI: 10.1002/bit.27212.

[168] R. I. Litvinov and J. W. Weisel, "Shear strengthens fibrin: the knob-hole interactions display 'catch-slip' kinetics," *Journal of Thrombosis and Haemostasis*, vol. 11, no. 10, pp. 1933–1935, 2013. DOI: 10.1111/jth.12374.

[169] A. S. Rosenberg, "Effects of protein aggregates: An immunologic perspective," *The AAPS Journal*, vol. 8, no. 3, E501–E507, 2006. DOI: 10.1208/aapsj080359.

[170] L. F. Willis, V. Toprani, S. Wijetunge, A. Sievers, L. Lin, J. Williams, T. J. Crowley, S. E. Radford, N. Kapur, and D. J. Brockwell, "Exploring a role for flow-induced aggregation assays in platform formulation optimisation for antibody-based proteins," *Journal of Pharmaceutical Sciences*, vol. 113, no. 3, pp. 625–636, 2024. DOI: 10.1016/j.xphs.2023.10.031.

[171] F. Grigolato and P. Arosio, "The role of surfaces on amyloid formation," *Biophysical Chemistry*, vol. 270, p. 106 533, 2021. DOI: 10.1016/j.bpc.2020.106533.

[172] I. Panagi, "The interaction of protein-based pharmaceuticals with interfaces under biopharmaceutical manufacturing," PhD thesis, University of Leeds, 2022.

[173] L. F. Willis, A. Kumar, J. Dobson, N. J. Bond, D. Lowe, R. Turner, S. E. Radford, N. Kapur, and D. J. Brockwell, "Using extensional flow to reveal diverse aggregation landscapes for three IgG1 molecules," *Biotechnology and Bioengineering*, vol. 115, no. 5, pp. 1216–1225, 2018. DOI: 10.1002/bit.26543.

[174] L. F. Willis, A. Kumar, T. Jain, I. Caffry, Y. Xu, S. E. Radford, N. Kapur, M. Vásquez, and D. J. Brockwell, "The uniqueness of flow in probing the aggregation behavior of clinically relevant antibodies," *Engineering Reports*, vol. 2, no. 5, pp. 1–13, 2020. DOI: 10.1002/eng2.12147.

[175] A. Page, *Private communication*, University of Leeds, 2024.

[176] R. Potter and A. Groom, "Capillary diameter and geometry in cardiac and skeletal muscle studied by means of corrosion casts," *Microvascular Research*, vol. 25, no. 1, pp. 68–84, 1983. DOI: 10.1016/0026-2862(83)90044-4.

[177] M. J. Roman, R. B. Devereux, R. Kramer-Fox, and J. O'Loughlin, "Two-dimensional echocardiographic aortic root dimensions in normal children and adults," *The American Journal of Cardiology*, vol. 64, no. 8, pp. 507–512, 1989. DOI: 10.1016/0002-9149(89)90430-x.

[178] T. F. Sherman, "On connecting large vessels to small: The meaning of Murray's law.," *The Journal of General Physiology*, vol. 78, no. 4, pp. 431–453, 1981. DOI: 10.1085/jgp.78.4.431.

[179] J. A. Dobson, "Biophysical characterisation of biopharmaceuticals under defined flow fields," PhD thesis, University of Leeds, 2017.

[180] A. Zhmurov, A. E. Brown, R. I. Litvinov, R. I. Dima, J. W. Weisel, and V. Barsegov, "Mechanism of fibrin(ogen) forced unfolding," *Structure*, vol. 19, no. 11, pp. 1615–1624, 2011. DOI: 10.1016/j.str.2011.08.013.

[181] A. E. Brown, R. I. Litvinov, D. E. Discher, and J. W. Weisel, "Forced unfolding of coiled-coils in fibrinogen by single-molecule AFM," *Biophysical Journal*, vol. 92, no. 5, pp. 39–41, 2007. DOI: 10.1529/biophysj.106.101261.

[182] J. Z. Zhang and C. Redman, "Fibrinogen assembly and secretion: Role of intrachain disulfide loops," *Journal of Biological Chemistry*, vol. 271, no. 47, pp. 30 083–30 088, 1996. DOI: 10.1074/jbc.271.47.30083.

[183] L. B. Poole, "The basics of thiols and cysteines in redox biology and chemistry," *Free radical biology and medicine*, vol. 80, pp. 148–57, 2015. DOI: 10.1016/j.freeradbiomed.2014.11.013.

[184] V. L. Mendoza and R. W. Vachet, "Probing protein structure by amino acid-specific covalent labeling and mass spectrometry," *Mass Spectrometry Reviews*, vol. 28, no. 5, pp. 785–815, 2009. DOI: 10.1002/mas.20203.

[185] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, and H. E. Gaub, "How strong is a covalent bond?" *Science*, vol. 283, no. 5408, pp. 1727–1730, 1999. DOI: 10.1126/science.283.5408.1727.

[186] T. A. Hakala *et al.*, "Accelerating Reaction Rates of Biomolecules by Using Shear Stress in Artificial Capillary Systems," 2021. DOI: 10.1021/jacs.1c03681.

[187] C. C. Krieger, X. An, H. Y. Tang, N. Mohandas, D. W. Speicher, and D. E. Discher, "Cysteine shotgun-mass spectrometry (CS-MS) reveals dynamic sequence of protein structure changes within mutant and stressed cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 20, pp. 8269–8274, 2011. DOI: 10.1073/pnas.1018887108.

[188] S. Jin, Z. Shen, J. Li, P. Lin, X. Xu, X. Ding, and H. Liu, "Fibrinogen A Alpha-chain amyloidosis associated with a novel variant in a chinese family," *Kidney International Reports*, vol. 6, no. 10, pp. 2726–2730, 2021. DOI: 10.1016/j.ekir.2021.07.014.

[189] Y. Kim, S. O. Ho, N. R. Gassman, Y. Korlann, E. V. Landorf, F. R. Collart, and S. Weiss, "Efficient site-specific labeling of proteins via cysteines," *Bioconjugate Chemistry*, vol. 19, no. 3, pp. 786–791, 2008. DOI: 10.1021/bc7002499.

[190] Thermo Scientific, *Tech tip 31: Calculate dye:protein (f/p) molar ratios*, https://assets.thermofisher.com/TFS-Assets/LSG/brochures/TR0031-Calc-FP-ratios.pdf, Accessed 21/02/2024.

[191] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch, "ExPASy: The proteomics server for in-depth protein knowledge and analysis," *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3784–3788, 2003. DOI: 10.1093/nar/gkg563.

[192] V. J. Marder, N. R. Shulman, and W. R. Carroll, "High molecular weight derivatives of human fibrinogen produced by plasmin. I. Physicochemical and immunological characterization.," *Journal of Biological Chemistry*, vol. 244, no. 8, pp. 2111–2119, 1969. DOI: 10.1016/S0021-9258(18)94373-5.

[193] The UniProt Consortium, "UniProt: the Universal Protein Knowledgebase in 2023," *Nucleic Acids Research*, vol. 51, no. D1, pp. D523–D531, 2022. DOI: 10.1093/nar/gkac1052.

[194] A. Shrake and J. Rupley, "Environment and exposure to solvent of protein atoms. lysozyme and insulin," *Journal of Molecular Biology*, vol. 79, no. 2, pp. 351–371, 1973. DOI: 10.1016/0022-2836(73)90011-9.

[195] E. Durham, B. Dorr, N. Woetzel, R. Staritzbichler, and J. Meiler, "Solvent accessible surface area approximations for rapid and accurate protein structure prediction," *Journal of Molecular Modeling*, vol. 15, no. 9, pp. 1093–1108, 2009. DOI: 10.1007/s00894-009-0454-9.

[196] T. Kamijo, T. Kaido, M. Yoda, S. Arai, K. Yamauchi, and N. Okumura, "Recombinant Gamma-Y278H fibrinogen showed normal secretion from cho cells, but a corresponding heterozygous patient showed hypofibrinogenemia," *International Journal of Molecular Sciences*, vol. 22, no. 10, 2021. DOI: 10.3390/ijms22105218.

[197] V. Calabrese, T. Porto Santos, C. G. Lopez, M. P. Lettinga, S. J. Haward, and A. Q. Shen, "Extensibility governs the flow-induced alignment of polymers and rod-like colloids," *Phys. Rev. Res.*, vol. 6, p. L012042, 1 2024. DOI: 10.1103/PhysRevResearch.6.L012042.

[198] B. W. Soh, V. Narsimhan, A. R. Klotz, and P. S. Doyle, "Knots modify the coil-stretch transition in linear DNA polymers," *Soft Matter*, vol. 14, no. 9, pp. 1689–1698, 2018. DOI: 10.1039/c7sm02195j.

[199] H. R. McPherson *et al.*, "Fibrinogen αC-subregions critically contribute blood clot fibre growth, mechanical stability, and resistance to fibrinolysis," *eLife*, vol. 10, pp. 1–20, 2021. DOI: 10.7554/eLife.68761.

[200] R. I. Litvinov, S. Yakovlev, G. Tsurupa, O. V. Gorkun, L. Medved, and J. W. Weisel, "Direct Evidence for Specific Interactions of the Fibrinogen αC-Domains with the Central E Region and with Each Other," *Biochemistry*, vol. 46, no. 31, pp. 9133–9142, 2007. DOI: 10.1021/bi700944j.

[201] A. Zhmurov, O. Kononova, R. I. Litvinov, R. I. Dima, V. Barsegov, and J. W. Weisel, "Mechanical transition from alpha-helical coiled coils to beta-sheets in fibrin(ogen)," *Journal of the American Chemical Society*, vol. 134, no. 50, pp. 20 396–20 402, 2012. DOI: 10.1021/ja3076428.

[202] J. Jaspe and S. J. Hagen, "Do protein molecules unfold in a simple shear flow?" *Biophysical Journal*, vol. 91, no. 9, pp. 3415–3424, 2006. DOI: 10.1529/biophysj.106.089367.

[203] R. I. Litvinov and J. W. Weisel, "Fibrin mechanical properties and their structural origins," *Matrix Biology*, vol. 60-61, pp. 110–123, 2017. DOI: 10.1016/j.matbio.2016.08.003.

[204] H. K. Moffatt, "Viscous and resistive eddies near a sharp corner," *Journal of Fluid Mechanics*, vol. 18, no. 1, pp. 1–18, 1964. DOI: 10.1017/S0022112064000015.

[205] G. B. Bishko, O. G. Harlen, T. C. McLeish, and T. M. Nicholson, "Numerical simulation of the transient flow of branched polymer melts through a planar contraction using the 'pom-pom' model," *Journal of Non-Newtonian Fluid Mechanics*, vol. 82, no. 2-3, pp. 255–273, 1999. DOI: 10.1016/S0377-0257(98)00165-7.

[206] W. R. Dean and P. E. Montagnon, "On the steady motion of viscous liquid in a corner," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 45, no. 3, pp. 389–394, 1949. DOI: 10.1017/S0305004100025019.

[207] G. Tsurupa, R. R. Hantgan, R. A. Burton, I. Pechik, N. Tjandra, and L. Medved, "Structure, stability, and interaction of the fibrin(ogen) αC-domains," *Biochemistry*, vol. 48, no. 51, pp. 12 191–12 201, 2009. DOI: 10.1021/bi901640e.

[208] K. Djinovic-Carugo and O. Carugo, "Missing strings of residues in protein crystal structures," *Intrinsically Disord. Proteins*, vol. 3, no. 1, e1095697, 2015.

[209] T. K. Kim, "Understanding one-way ANOVA using conceptual figures," *Korean Journal of Anesthesiology*, vol. 70, no. 1, p. 22, 2017. DOI: 10.4097/kjae.2017.70.1.22.

[210] P. DiBello and J. Shainoff, "The circulatory half-lives of alpha-profibrin and alpha-fibrin monomer, and comparisons with other fibrin(ogen) derivatives," *Thrombosis and Haemostasis*, vol. 89, no. 01, pp. 48–52, 2003. DOI: 10.1055/s-0037-1613542.

[211] S. Kania, A. Oztekin, X. Cheng, X. F. Zhang, and E. Webb, "Predicting pathological von Willebrand factor unraveling in elongational flow," *Biophysical Journal*, vol. 120, no. 10, pp. 1903–1915, 2021. DOI: 10.1016/j.bpj.2021.03.008.

[212] A. P. Wiita, S. R. K. Ainavarapu, H. H. Huang, and J. M. Fernandez, "Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques," *Proceedings of the National Academy of Sciences*, vol. 103, no. 19, pp. 7222–7227, 2006. DOI: 10.1073/pnas.0511035103.

[213] R. E. Goldstein and J. W. van de Meent, "A physical perspective on cytoplasmic streaming," *Interface Focus*, vol. 5, no. 4, 2015. DOI: 10.1098/rsfs.2015.0030.

[214] J. Skolnick, "Perspective: On the importance of hydrodynamic interactions in the subcellular dynamics of macromolecules," *Journal of Chemical Physics*, vol. 145, no. 10, 2016. DOI: 10.1063/1.4962258.

[215] M. Padgett and R. Di Leonardo, "Holographic optical tweezers and their relevance to lab on chip devices," *Lab on a Chip*, vol. 11, no. 7, pp. 1196–1205, 2011. DOI: 10.1039/c0lc00526f.

[216] J. Elgeti and G. Gompper, "Emergence of metachronal waves in cilia arrays," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 12, pp. 4470–4475, 2013. DOI: 10.1073/pnas.1218869110.

[217] E. Lauga and T. R. Powers, "The hydrodynamics of swimming microorganisms," *Reports on Progress in Physics*, vol. 72, no. 9, 2009. DOI: 10.1088/0034-4885/72/9/096601.

[218] T. Frembgen-Kesner and A. H. Elcock, "Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins," *Journal of Chemical Theory and Computation*, vol. 5, no. 2, pp. 242–256, 2009. DOI: 10.1021/ct800499p.

[219] C. E. Monteith, M. E. Brunner, I. Djagaeva, A. M. Bielecki, J. M. Deutsch, and W. M. Saxton, "A Mechanism for Cytoplasmic Streaming: Kinesin-Driven Alignment of Microtubules and Fast Fluid Flows," *Biophysical Journal*, vol. 110, no. 9, pp. 2053–2065, 2016. DOI: 10.1016/j.bpj.2016.03.036.

[220] D. L. Ermak and J. A. McCammon, "Brownian dynamics with hydrodynamic interactions," *The Journal of Chemical Physics*, vol. 69, no. 4, pp. 1352–1360, 1978. DOI: 10.1063/1.436761.

[221] B. Sprinkle, F. Balboa Usabiaga, N. A. Patankar, and A. Donev, "Large scale Brownian dynamics of confined suspensions of rigid particles," *Journal of Chemical Physics*, vol. 147, no. 24, 2017. DOI: 10.1063/1.5003833.

[222] J. Rotne and S. Prager, "Variational treatment of hydrodynamic interaction in polymers," *The Journal of Chemical Physics*, vol. 50, no. 11, pp. 4831–4837, 1969. DOI: 10.1063/1.1670977.

[223] E. Wajnryb, K. A. Mizerski, and P. J. Zuk, "Generalization of the Rotne–Prager–Yamakawa mobility and shear disturbance tensors," *Journal of Fluid Mechanics*, vol. 731, 2013. DOI: 10.1017/jfm.2013.402.

[224] R. Vácha and D. Frenkel, "Relation between molecular shape and the morphology of self-assembling aggregates: A simulation study," *Biophysical Journal*, vol. 101, no. 6, pp. 1432–1439, 2011. DOI: 10.1016/j.bpj.2011.07.046.

[225] P. Balogh and P. Bagchi, "Three-dimensional distribution of wall shear stress and its gradient in red cell-resolved computational modeling of blood flow in in-vivo-

like microvascular networks," *Physiological Reports*, vol. 7, no. 9, e14067, 2019. DOI: 10.14814/phy2.14067.

[226] A. Lauric, J. E. Hippelheuser, and A. M. Malek, "Induction of aneurysmogenic high positive wall shear stress gradient by wide angle at cerebral bifurcations, independent of flow rate," *Journal of Neurosurgery*, vol. 131, no. 2, pp. 442–452, 2019. DOI: 10.3171/2018.3.jns173128.

[227] H. C. Berg, *Random Walks in Biology*. Princeton, NJ, USA: Princeton University Press, 1993.

[228] I. Nir, D. Huttner, and A. Meller, "Direct sensing and discrimination among ubiquitin and ubiquitin chains using solid-state nanopores," *Biophysical Journal*, vol. 108, no. 9, pp. 2340–2349, 2015. DOI: 10.1016/j.bpj.2015.03.025.

[229] W. C. Parker, N. Chakraborty, R. Vrikkis, G. Elliott, S. Smith, and P. J. Moyer, "High-resolution intracellular viscosity measurement using time-dependent fluorescence anisotropy," *Optics Express*, vol. 18, no. 16, p. 16 607, 2010. DOI: 10.1364/oe.18.016607.

[230] H. Fischer, I. Polikarpov, and A. F. Craievich, "Average protein density is a molecular-weight-dependent function," *Protein Science*, vol. 13, no. 10, pp. 2825–2828, 2004. DOI: 10.1110/ps.04688204.