Self Assembly of Model Polymers into Biological Random Networks

Matthew H J Bailey1, Mark Wilson1

Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford, OX1 3QZ

Abstract

The properties of biological networks, such as those found in the ocular lens capsule, are difficult to study without simplified models. Model polymers are developed, inspired by “worm-like” curve models, that are shown to spontaneously self assemble to form networks similar to those observed experimentally in biological systems. These highly simplified coarse-grained models allow the self assembly process to be studied on near-realistic time-scales. Metrics are developed (using a polygon-based framework) which are useful for describing simulated networks and can also be applied to images of real networks. These metrics are used to show the range of control that the computational polymer model has over the networks, including the polygon structure and short range order. The structure of the simulated networks is compared to previous simulation work and microscope images of real networks. The network structure is shown to be a function of the interaction strengths, cooling rates and external pressure. In addition, “pre-tangled” network structures are introduced and shown to significantly influence the subsequent network structure. The network structures obtained fit into a region of the network landscape effectively inaccessible to random (entropically-driven) networks but which are occupied by experimentally-derived configurations.

Keywords: Continuous Random Network, self assembly, polygon statistics, polygon distributions, collagen network

1. Introduction

Two-dimensional (2D) networks are critically important in biology — from basement membranes surrounding muscles, to the lens capsule of the eye. Furthermore, some three-dimensional (3D) networks are composed of stacked layers of 2D networks in which the intra-layer interactions are significantly larger than those of the inter-layer. The ubiquity of these systems makes understanding the formation and ageing of the 2D networks a key research interest, often with the aim that an understanding of the structure and properties of biological networks can lead to the development of synthetic materials that mimic natural biological networks. By understanding the millions of years of design experience nature has applied, stronger or stretchier materials can be developed allowing, for example, prosthetic replacements for parts of the body to be built[1].

One example of a 2D biological network is the collagen IV network in the ocular lens capsule. Artificial lens capsules can be implanted after cataract surgery, which can improve the ability to accommodate focal depths in elderly people[2]. The loss of focal accommodation in human senescence has been linked to collagen IV networks becoming less stretchy as they age[3]. For this reason, artificial intraocular lenses can often be superior to the removed biological lens[4]. The mechanism of this network ageing is a poorly-understood process, despite being a topic of scientific interest for 100 years[5].

There is a clear need for simplified computational models to study this network, as individual collagen molecules have backbones of over 1000 amino acids[6], and the ageing process occurs over a human lifetime[5]. These properties make it difficult to observe the relevant pathway, because it cannot be directly observed on laboratory timescales and the complexity of the individual molecules makes large-scale atomistic simulation computationally unfeasible. A further issue arises because the 2D networks are most interesting when they are part of living creatures. The removal of biological networks (and their preparation for study) can often disturb their delicate structure, with significant knock-on effects for their properties[7]. Previous work has used animal models, such as mice, cows, and monkeys, to understand the ageing of human lens capsules; or coarse-grained finite-element models to treat the whole lens as an engineering problem[8, 9, 1]. Previous computational models have been based on graph theory; these include Erdős-Rényi random graphs, Mikado networks and bond-switching of ordered graphs[10, 11, 12]. These graph approaches are parameterised and based on experimental data and small-scale simulations, but provide an incomplete link between detailed molecule simulations and the wider network behaviour.

Collagen IV molecules can spontaneously self-assemble into networks in vitro, which can occur at more computationally accessible timescales than aging does[13, 14].

Email address: mark.wilson@chem.ox.ac.uk (Mark Wilson)

Preprint submitted to Elsevier October 26, 2020
These self-assembly studies have shown that the network structures formed are strongly dependent on the environmental conditions, for example on the concentration of dissolved salts [15, 16]. Such analysis provides some insight into which network arrangements are favourable, and how they self-assemble.

This work presents a highly simplified computational model for collagen-like molecules that self assemble into 2D network structures. The models are based on information obtained from experimental images of collagen IV networks, both self-assembled and in the ocular lens. The model is simple enough to be able to capture long timescales as well as to incorporate different interaction types.

First, a method of analysing 2D networks is established, including a way to unambiguously assign polygon structures. Next, biological random networks (both experimental images and simulated networks) are compared and contrasted to inorganic chemical networks, which have been studied using similar frameworks and which are generally more well-defined in terms of supporting a single atom local coordination number. The regularity of the polygons in the networks are a key point of comparison. Following that, the available range of control over the generated network structure is demonstrated by varying physically meaningful parameters, including pressure, cooling rates and the strength of attractions (a proxy for physical effects such as salt concentration). Finally, the generated networks are compared to data obtained from microscope images of 2D collagen IV networks, and are shown to better reproduce these networks than do previous continuous random (entropically-driven) networks.

2. Methods.

2.1. Network Structures.

Prior to discussing the results of the simulations, it is first helpful to establish a method to describe 2D networks rigorously, here using the gaps between molecules instead of the molecules themselves. Intuitively, a 2D network can be considered as being assembled from a collection of edge-sharing polygons — for example, a honeycomb is a network of hexagons, and chain-link fences may form a network of squares [17]. This concept can be extended to cover generic 2D networks across many length scales, from the atomic structure of glasses to the cantons of Switzerland [18, 19]. Viewing a 2D network as being constructed of polygons allows use of consistent metrics, applicable across many different types of network.

For example, the preference of polygons to be adjacent to similar or dissimilar polygons (described, for example, by the assortativity, and first discussed by Newman to describe social networks [20]) describes the short-range structure of a network and improves on the Aboav-Weaire parameter traditionally used to describe chemical 2D networks [21, 19]. Alternatively, the number of edges of a polygon (often referred to as their size) or their area is useful to quantify the voids in a network. In a biological context these voids could be affected by the presence of a scaffold, such as laminin [22], or by repulsive cations that the network forms around akin to a Voronoi partition [16]. A significant feature of biological networks (compared with inorganic glasses) is the ability to support a distribution of local coordination environments. The coordination number (called the “degree” in network theory) $k$ of nodes in the network can be used to analyse the preference for different types of bonding between molecular head groups or lateral interactions.

There are numerous ways to describe the polygon structure of a network, and each has its own benefits and shortcomings. These include polygons chosen by shortest path criteria, primitive rings that cannot be decomposed into smaller rings, and a series of stronger criteria. Polygon (or “ring”) assignment criteria are discussed in great detail in Yuan and Cormack [23], and also in Le Roux and Jund [24]. The latter discusses their implementation in the R.I.N.G.S. code, a popular package for obtaining ring size distributions. For example some methods of assigning polygons to a network ensure each edge is belongs to one polygon, or the area is covered by a layer of polygons with no overlaps, or all shortest paths between nodes are included [23]. We have found these methods to be difficult to interpret, especially in the presence of nodes with $k = 2$ or in the case of small periodic systems.

To this end, we have developed a method to describe the polygon structure utilising a Delaunay triangulation, a common technique to break large polygons into triangles often used in computer graphics applications. First, a network graph is created, in which small clusters of head groups are nodes, joined by edges representing molecules. An algorithmic description of this process can be found in the Supplementary Information Section I. The Delaunay triangulation of a set of nodes creates a set of triangles, whose edges connect neighbouring nodes and cover the entire area. We restricted our study to only networks where the Delaunay triangulation is a superset of the edges in the network; this criterion was met for all imaged and simulated networks we encountered. It is computationally simple to identify a triangle as a simple polygon. From there polygons with more edges can be built up by connecting smaller polygons, as shown in Figure 1. If edges that are not in the original graph are systematically removed, updating the polygons at each stage, it is possible construct a set of unique polygons for the network that covers the entire area and maps neatly to an intuitive definition of polygons.

2.2. The Simplified Model.

The assembly of collagen-like molecules was simulated using a model inspired by a combination of “worm-like” curve polymers and patchy particles. Patchy particles are colloids with attractive regions on their surface which have been shown experimentally and computationally to self-assemble [25, 26], and mimic the crystallisation properties
of proteins[27]. The simplified collagen like molecules are represented by a series of beads joined by springs. Each pair of beads is joined by a stiff harmonic potential, and each adjacent trio of beads has their angle constrained by an angular potential as shown schematically in Figure 2. The energy terms are controlled by the force constants, $k_l$ and $K_\theta$, respectively. The beads may have two (or more) types, here body (B) beads and head (H) beads. To mimic the effective excluded volume of a polymer biomolecule, the body beads repel each other according to a cut-off Lennard-Jones potential described by an energy scale $\epsilon_{\text{BB}}$ and a range $\sigma_{\text{BB}}$

$$U_{\text{BB}}(r) = \begin{cases} 4\epsilon_{\text{BB}} \left( \frac{2\sigma_{\text{BB}}}{r} \right)^{12} - \left( \frac{2\sigma_{\text{BB}}}{r} \right)^{6} & r < \sigma \\ 0 & r \geq \sigma. \end{cases}$$  

Two head groups interact with one another, according to a Lennard-Jones potential with an energy scale $\epsilon_{\text{HH}}$ and a range $\sigma_{\text{HH}}$.

$$U_{\text{HH}}(r) = 4\epsilon_{\text{HH}} \left[ \left( \frac{\sigma_{\text{HH}}}{r} \right)^{12} - \left( \frac{\sigma_{\text{HH}}}{r} \right)^{6} \right].$$

$\sigma_{\text{BB}}$ and $\sigma_{\text{HH}}$ are chosen to prevent molecules from interpenetrating and to favour $k = 3$ coordination. Body-head interactions are set to zero throughout ($U_{\text{HH}} = 0$). There is only one type of head group interaction in this model, but real collagen IV molecules have been observed to have two different head groups: 7S and NC1. The 7S and NC1 domains interact only with other domains of the same type, and favour different coordination numbers, $k$. However, initial simulations with two types of head groups lead to networks with only even sized polygons forming; this is similar to the geometrical frustration observed in ferromagnetic systems. By changing the attractive part and range of the Lennard-Jones potentials, the screening effect of salts in the self assembly solvent can effectively be mimicked without explicitly taking them into account (which greatly reduces computational expense).

The solvent was taken into account implicitly using a Langevin thermostat, which sets the force $\vec{F}$ on the molecules to

$$\vec{F} = \frac{\vec{F}_{\text{Bond}} + \vec{F}_{\text{Angle}} + \vec{F}_{\text{LJ}}}{m} \frac{m \vec{v}}{\gamma} + \xi \frac{k_B T m}{\gamma dt} \text{Brownian}$$

This thermostat has the traditional conservative forces including bond forces $\vec{F}_{\text{Bond}}$, intra-molecular angle forces $\vec{F}_{\text{Angle}}$ and head group forces $\vec{F}_{\text{LJ}}$, but adds a drag term featuring the mass $m$ and velocity $\vec{v}$ of each bead, divided by a damping factor $\gamma$ representing the viscosity of the solvent. Finally, the thermostat features a Brownian dynamics term with $dt$ the timestep, $T$ the desired temperature, and $\xi$ a random number recalculated every step.

### 2.3. Simulation Protocol.

This work used LAMMPS to perform Molecular Dynamics [31]. Unless otherwise mentioned, each set of parameters was sampled in ten different configurations of $20 \times 20$ molecules in a periodic box, initially placed on a square grid. The grid arrangement was equilibrated by simulation at $T = 30 \, \text{K}$ for $t = 40 \, \mu\text{s}$, with an initial $t = 4 \, \mu\text{s}$ limiting the maximum motion per timestep to prevent the simulation exploding. The drive to form a network was so strong that an initial network was commonly formed during equilibration. Next, the box size was relaxed by simulating a $NpT$ ensemble (i.e. constant number of particles, pressure, and temperature) at $T = 30 \, \text{K}$ with $p = 20 \, \text{Pa}$ for a further $t = 40 \, \mu\text{s}$. This pressure was chosen based on parameter scans (discussed more in Sec. 3.4.3) as it best reproduced the biological networks of interest. The use of pressure is a shorthand, as the corresponding physical property is more accurately stress. However, LAMMPS treats 2D systems as being in a 3D...
box with the z component of position and velocity set to 0. Following equilibration of the pressure, the simulation cell size was fixed and the simulation returned to a NVT (i.e. constant number of particles, volume and temperature) ensemble. The proto-network was heated to 300 K by adjusting T in Eq. (3) over 100 µs, which melted it. This is a relatively low melting temperature for computational efficiency, and all energy values in the simulation could be arbitrarily rescaled if necessary. After melting was complete (as evidenced through equilibration of the structural metrics, i.e. k = 1 for all molecules), the collagen polymer liquid was cooled over a final 100 µs to 30 K. The total time from T = 300 K to 30 K was recorded as the value t_{cool}. Snapshots were extracted and analysed at the end of each simulation.

In general, only one parameter was varied across each set of simulations, with the remaining parameters fixed at the default values that most reliably produced networks (chosen after an initial scan of the effective parameter space). These default parameters were bond and angle energy scales as $k_l = 1.657 \times 10^{-4} \text{ N m}^{-1}$, $K_\theta = 200 \times 10^{-21} \text{ J}$, a cooling time of $t_{cool} = 100 \mu s$, Lennard-Jones head-head and body-body energy scales were $\epsilon_{HH} = 4.142 \times 10^{-21} \text{ J}$, $\epsilon_{BB} = 16.142 \times 10^{-21} \text{ J}$. The length scales in the simulation were such that the equilibrium polymer length was $l_{eqn} = 300 \text{ nm}$, and the Lennard-Jones length scales were $\sigma_{HH} = 50 \text{ nm}$ and $\sigma_{BB} = 137.5 \text{ nm}$. The polymer length is similar to the observed length of collagen IV molecules, which has been reported to be in the range 300 nm to 400 nm[32, 33]. The body bead interactions are considerably larger than the actual width of a collagen IV triple helix, 1.5 nm, as they also take into account the excluded volume effect of polymers where close coordination in entropically unfavourable. The Lennard-Jones length scales were chosen to favour $k = 3$ coordination geometrically akin to the design of patchy particles.

3. Results and Discussion.

3.1. Comparison to the Ideal (Maximum Entropy) Networks.

A key feature of the polymer model is that it can take enthalpic effects into account naturally. The importance of enthalpy to biological networks can be demonstrated by comparing with polymer networks which are dominated purely by entropy, such as those studied using bond switching methods[12]. A simple, numerical model is the maximum entropy distribution. This maximises the entropy of polygon edge counts according to the following constraints[34, 19]:

$$\sum_n p_n = 1 \quad (4a)$$
$$\sum_n np_n = \langle n \rangle \quad (4b)$$
$$\sum_n \frac{p_n}{n} = \text{constant} \quad (4c)$$

with $p_n$ the fraction of polygons with $n$ edges. This system of equations can be solved numerically with Lagrange’s method of undetermined multipliers.

The self-assembled networks show polygon edge count distributions which do not resemble the numerical maximum entropy model; this is visible in Figure 3, which compares polygon edge counts from a polymer simulated and the polygon edge counts from maximum entropy model distribution. This was based on simulations described in Sec. 3.4.1 at $k_l = 1.5 \times 10^{-4} \text{ N m}$. Similar differences are observed across all simulations, and the difference between a maximum entropy model and the distribution of edge counts from polymer simulations is sufficiently pronounced that the numerical fitting procedure often fails. For polygons with many edges, their population is accurately predicted by a maximum entropy approach because they are sufficiently flexible to minimise bond and angular strain. However, for polygons with few edges, such as triangles and squares, this angular strain is unavoidable. There are therefore very few triangles in the generated networks because of the strain involved in forming them, but an overpopulation of squares compared to the maximum entropy population.

3.2. Comparison to Previous Network Studies.

Prior to considering how different controllable variables may affect the network structure, a clear set of metrics is required. These metrics must capture the key differences that biopolymer networks exhibit compared with inorganic networks, describe the effects of enthalpy and entropy, and provide simple proxies for complex physical phenomena. These metrics build on those previously applied to characterise 2D networks[12].

3.2.1. Polygon Convexity.

One interesting difference that shown by biopolymer networks, compared with inorganic networks is that the polygons are more likely to be concave or distorted. The strong angular potentials and fixed coordination numbers...
of atoms in inorganic glasses, such as silica, lead to a strong preference for convex polygons. This is not the case for polymer networks, which can have variable coordination numbers, more flexible angular potentials and even curved edges. To quantify this difference in shape regularity, we make use of a metric originally developed for computer graphics\cite{35}. The shape regularity coefficient (SRC) of a shape $\mathcal{S}$, is defined as

$$\text{SRC}(\mathcal{S}) = \frac{\text{SO}(\mathcal{S})V_{xy}(\mathcal{S})\text{CO}(\mathcal{S})}{\text{in which } \text{SO}(\mathcal{S}) \text{ is the solidity, defined as the ratio of the polygon area to the area of its convex hull; } V_{xy}(\mathcal{S}) \text{ is the balanced repartition of the shape, defined as } \sqrt{\frac{\min(x_i, x_j)}{\max(x_i, x_j)}} \text{ where } x_i(y) \text{ is the standard deviation of the } x(y) \text{ coordinate of the vertex positions; and finally } \text{CO}(\mathcal{S}) \text{ is the convexity, defined as the ratio of the convex hull perimeter to the polygon perimeter.}

Regular shapes, like a square, pentagon or circle are characterised by $\text{SRC} = 1$. Less-regular shapes, such as ellipses or rectangles, have $0 \leq \text{SRC} < 1$. Figure 4 shows four example networks with varying degrees of regularity as defined by the $\text{SRC}$, with each polygon coloured by $\text{SRC}$. The silica network in Figure 4a\cite{36} shows highly regular polygons, corresponding to $\langle \text{SRC} \rangle = 0.966 \pm 0.024$, the latter figure being the standard deviation. A more complete table of data can be seen in Table SI 1 of the Supplementary Information. The silica networks show $\langle \text{SRC} \rangle \approx 0.96$ with relatively narrow distributions (small standard deviations). By comparison biological networks show $\langle \text{SRC} \rangle$ values which are significantly smaller. For example, networks generated in the present work show $\langle \text{SRC} \rangle$ in the range $0.747 \leq \langle \text{SRC} \rangle \leq 0.928$ (see Table SI 2). In a previous publication a number of experimental images were analysed\cite{12}. The microscope images of collagen IV networks showed that the polygons formed were often irregular.

Table SI 1 lists the values of $\langle \text{SRC} \rangle$ obtained from 21 such images which show values in the range $0.69 < \langle \text{SRC} \rangle < 0.91$. As a visual example, Figure 4d shows a section of a network from Yurchenco and Ruben\cite{37} which corresponds to $\langle \text{SRC} \rangle = 0.871 \pm 0.088$. The greater flexibility of the biological networks, compared to the silica, is highlighted by the broader range of $\text{SRC}$ values observed in the former.

Overall, therefore, the flexibility of biopolymers leads to less regular polygons in a 2D network, and this matches observation from microscope images. Other factors that lead to lower $\text{SRC}$ values are more flexible angular terms around nodes in the graph compared to inorganic networks, and the presence of $k = 2$ sites which can lead to concave shapes. These concave shapes can be seen in Figure 4b and Figure 4d as the long shapes, coloured pale. In comparison, all polygons in Figure 4a are convex and approximately isotropic. A full table of data with $\langle \text{SRC} \rangle$ values is available in the Supplementary Information as Table SI 2.

3.2.2. Coordination Number and Internal Energy.

The energetics of the networks, while interesting, are highly dependent on the potential model used to describe them. The short-range nature of the Lennard-Jones potential used in this work means that it is possible to use the node coordination numbers $k$ as a proxy for the internal energy in the network. This removes the complicating effects of bond stretches and angular strain, and solely represents the energy gained in forming the network. This makes it possible to study the energetics of the network easily, using only the final polygon structure.

Figure 5 shows one example of the correlation between the node coordination number and the internal energy. Figure 5a shows the evolution of the internal energy and mean node coordination number for a simulation taken from Sec. 3.4.2, cooled over 100 ps. Figure 5b further highlights the correlation between $U_{\text{pair}}$ and $\langle k \rangle$, with scatter points being the positions of $\langle k \rangle$ and $U_{\text{pair}}$ shown at different time points in Figure 5a. Similar correlations exist across all the simulations we performed. The linear link between $U_{\text{pair}}$ and $\langle k \rangle$ makes $\langle k \rangle$ useful, because it is independent of simulation parameters and smooths out thermal noise.

3.3. Energetics of Idealised Network Structures.

The actual network structures adopted in collagenous networks has been a topic of some discussion. Timpl et al. and more recently Cummings et al.\cite{32, 16} favour a “chain-link fence” network, effectively a square-net structure. Burd\cite{39}, has alternatively suggested a primarily hexagonal network. Yurchenco and Furthmayr suggest a disordered collagen...
IV network interacting with an ordered scaffold[14, 37]. Imaging experiments have shed some light on the structure of the networks. However, complicating factors such as the biological origins of the networks, the presence of surfaces, or dissolved salts, mean that the matter has not been conclusively resolved.

Using our current polymer model, the energies of both square and hexagonal networks were compared. The initial networks are shown in Figure 6a and Figure 6b. The idealised networks were constructed at their energy minima (corresponding to all bonds remaining unstretched), and scaled in the xy plane. The stretching simulations revealed that the simple potential model used here always favours the hexagonal structure over the square net structure, regardless of the choice of $k_l$ and $\epsilon_{HH}$. However, an improved potential model that better takes the preference of 7S and NC1 domains for different coordination numbers ($k = 2$ and $k = 4$ instead of $k = 3$ for both types) may reverse this stabilisation order and favour a square net. Equally, environmental conditions or scaffolds may also reverse the stabilisation order.

Data on the stretching simulations can be found in Section SI IV of the Supplementary Information.

### 3.4. Control of the Network Structure.

Having established a highly simplified model and metrics to describe the results, this work next investigates the sensitivity of network structure to model parameters and simulation conditions.

#### 3.4.1. Polymer Potential Parameters.

Collective behaviour, such as network assembly, can be strongly affected by the properties of the individual assembling units — here, the simplified model polymer. With any model, it is important to capture as much of the critical physics of a system while retaining computationally affordability. Two such important physical properties of collagen IV are how stiff or elastic a molecule is, because of the need for elasticity in the lens capsule, and the flexibility of the collagen IV molecule, which has been shown to be key in forming networks[40]. Here the impact of both the molecular stiffness, controlled by varying the energy scale of the harmonic bonds $k_b$, and the non-linear flexibility governed by changing the energy of the angular bonds $K_\theta$, are investigated, while all other parameters used are the defaults discussed in Sec. 2.3. There is a brief discussion of the effects of intrinsic curvature (by changing the equilibrium angle between three beads) in Section SI III, which was found to only have minor effects.

The bond strength was found to have little effect on the nature of the polygons formed, and the mean length of an edge in the network did not change with $k_l$, remaining at $\langle l \rangle = 317 \text{ nm} \pm 53 \text{ nm}$. This matches the spacing between interaction sites for untangled molecules observed by Yurchenco and Furthmayr [14], although it is longer than the distance between tangled sites in the more complex networks observed by Barnard et al. [41]. This indicates that polymers successfully forming a network do not deviate significantly from their equilibrium length in order to accommodate network formation. This near-invariance to $k_l$ allows the selection of convenient energy and timescales for simulations such that relatively long time-scale properties (such as network ageing) can be accessed while retaining relatively rigid polymers.

Changing the angular strength parameter, $K_\theta$ in the range $5 \times 10^{-21} \text{ J} \leq K_\theta \leq 4 \times 10^{-19} \text{ J}$ had two major effects. First, it reduced the average length of an edge in the network from 295 nm for “loose” angles (low $K_\theta$) to 319 nm (high $K_\theta$) for “stiff” angles. This is because the head-to-head distance within one molecule is shorter if the backbone curves, and looser molecules can bend more easily to accommodate better head group interactions. The longer edge lengths better reproduce the lengths observed in biological networks[32]. Second, it decreased the number of polygons successfully formed in the networks, from 81.0 on average in the loosest case to 66.0 on average in the stiffest case. This demonstrates that the bending of the polymers is important in order to allow network for-
When the networks are cooled for longer, the average node coordination $\langle k \rangle$ increases to 2.996 for the slowest cooled network from 2.638 for the fastest cooled network. This can be inferred from Figure 7, as the polygons in the fast-cooled networks feature more polygons with dangling edges, and a network with many fewer dangling edges in the slower-cooled network. The slow-cooled networks better resemble those seen in biology, and reinforce the value of a computational model simple enough to access long timescales.

### 3.4.3. External Pressure

Since the external pressure on a real ocular lens capsule (where biological collagen IV networks are of interest) changes in the process of accommodation, the external pressure on the simulated networks was varied whilst they were forming. The isothermal-isobaric ($NpT$) step in the simulation protocol served a useful purpose in setting the periodic box size to be physically-determined rather than an a priori assigned size. The nature of the simulation means that the $z$ coordinate is present but effectively infinite, and molecular positions have no $z$ component. This means that pressure is a more convenient physical measure than stress, which would be strictly accurate for 2D systems.

Below a pressure of $p = 2\text{ Pa}$, no networks form. At low pressures, the available volume for each molecule is so large that the mean intermolecular interaction energies are relatively small. As pressures increase, the coordination number per node increases and the average area and number of edges of polygons decreases. These data can be seen in Table 2, and examples are shown in Figure 8. Networks formed at intermediate pressures often had dangling edges, and networks at high pressures had many small polygons forced together, surrounded by large polygons. The number of dangling edges is seen in Table 2, which counts the number of edges which are not involved in any polygon. For a full description of how the dangling nodes are quantified, see the polygon finding algorithm discussion in Bailey et al. [12].

### 3.4.4. Lennard-Jones Interaction Strength

One well-studied aspect of collagen network formation is the effect of salts dissolved in the solvent, which can affect the stiffness of collagen molecules[42], the rate of network formation[16] and the structure of the network[15]. These different effects are difficult to deconvolute from one
another experimentally; for example one cannot easily tell how the stiffness of monomers changes the final network without simultaneously changing the rate and energetics of network formation. In a simulation, these parameters are more easily separable and can be individually varied which can shed light on how multiple salting effects contribute.

We tested the effect of energetics of head-group interaction by varying the well-depth for Lennard-Jones interactions in the range $1 \times 10^{-21} \text{J} \leq \epsilon_{\text{HH}} \leq 2 \times 10^{-20} \text{J}$. Some example networks are shown in Figure 9 at different values of $\epsilon_{\text{HH}}$, and highlight the dramatic effect on network structure. Relatively weak Lennard-Jones interactions lead to irregular networks that do not show any significant polygonal character. In this weak attraction regime there is sufficient thermal energy to disconnect nodes in the network for a longer period. This meant that the networks did not build up highly-coordinated sites, and remained as connected chains. Strong Lennard-Jones attractions encouraged more effective polygon network formation, with more hexagons and squares. The energetic reward for head groups interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime meant the average coordination number approached 3, and more small polygons interacting in this regime mean
ted to capture some of these 3D lateral interactions by pre-forming tangled molecules and introducing them to the network. Some example molecules are shown in Figure 10, chosen as the simplest possible tangles that match observed images[14]. These are schematic molecules, with coloured circles representing beads that interact by Lennard-Jones potentials between beads in a molecule.

The simulation procedure was followed as described above, but with between 0 % and 50 % of the linear polygons substituted for their tangled counterparts. Two example networks can be seen in Figure 11, showing the propensity of tangling sites to lead to a greater number of small polygons. The addition of tangling sites also reduces the width of the $k$ distribution, measured by $\mu_2(k)$, as each tangle point had a fixed value of $k = 3$. Data on $\mu_2(k)$ can

| $\epsilon_{\text{HH}}/1 \times 10^{-21} \text{J}$ | $\langle \text{Area} \rangle/\text{nm}^2$ | $\langle n \rangle$ | $\langle k \rangle$ |
|----------------|----------------|----------|----------|
| 2              | 720321         | 6.42     | 2.69     |
| 4              | 386607         | 6.89     | 2.82     |
| 6              | 332060         | 6.92     | 2.87     |
| 8              | 282066         | 6.56     | 2.95     |
| 10             | 261117         | 6.41     | 2.98     |
| 12             | 257307         | 6.38     | 2.98     |
| 14             | 241528         | 6.27     | 3.01     |
| 16             | 231528         | 6.10     | 3.03     |
| 20             | 217023         | 5.93     | 3.05     |
be seen in Table 4, which show a clear decrease in $\mu_2(k)$ as a greater percentage of straight polymers are replaced with tangled polymers. The Figure 10c tangled molecules had a greater effect on $\mu_2(k)$ because they had two sites of fixed $k$ instead of one or zero. The decreased $\mu_2(k)$ lead to a better fit to experimental networks, as discussed further in Sec. 3.5. The networks with different fractions of tangling sites had similar short-range orderings, as represented by the assortativity $r$ in Table 4, showing a slight trend towards a more random short range order (with $r$ closer to 0) when there are more tangling sites.

The average polygon area decreased in the networks with more tangling sites. This is to be expected, as the tangled molecules were not a one-to-one replacement for single polymers and had a greater density of head group sites. The presence of tangling sites could better capture the properties of physical networks and allow for a potential model with head group preferences that better reflects real networks.

### 3.5. The Network Landscape

The simulation results discussed here can be viewed in the context of Bailey et al. [12]. In that work the formation of disordered collagen-like networks was studied starting from a regular hexagonal network and systematically switching the bonds, maintaining an average value of $\langle k \rangle$. This effectively models the network structure as entirely entropically-driven. The work also highlighted the role of the assortativity as an effective metric in characterising the network structure. Furthermore, the differences between networks could be best differentiated by considering a “network landscape” which shows the assortativity as a function of the width of the $k$ distribution, as characterised by $\mu_2(k)$. While this technique better simulated 2D biological networks than previous bond switching techniques, it identified a gap where it was not possible to produce networks with a certain range of polygon assortativities $r$ and distribution widths of $k$.

Figure 12 shows a “network landscape” similar to that in Bailey et al. [12], capturing $r$ and $\mu_2(k)$ for the networks discussed in the current work. The shaded regions highlight the regions of the landscape accessible to the entropic models. The figure also shows the results obtained from analysis of experimental images (from refs.[41, 43, 44, 15, 14, 37]). The values obtained from experiment show a wide range of values reflecting both the different conditions under which images are obtained and the difficulties in extracting detailed information. Experimental images occupy regions of the network landscape excluded from the entropically-driven models, which had studied the effects of temperature, cooling rates, and limits on $k$. Figure 12 shows the results from the present work, varying the system variables as discussed above.

Each set of simulations discussed earlier is present as a different coloured set of squares, where each square represents the result of a single simulation. Critically, the introduction of controlled inter- and intra-molecular interactions (enthalpic contributions) generates network configurations which significantly fill the regions of the network landscape occupied by the experimental configurations. The majority of parameter changes result in a clear relationship between $r$ and $\mu_2(k)$ with the former becoming less negative (less dissassortative) as $\mu_2(k)$ increases. This is most obvious in the results from Sec. 3.4.2 and Sec. 3.4.3, and also holds for the results from Sec. 3.4.1. This relationship can be explained by reference to the assortativity of polygons formed by a random point process, which is $r \approx -0.15[19]$. This disassortativity for an entirely random process comes from geometrical constraints; with $k$ and $\langle n \rangle$ fixed, small polygons must border large polygons[12]. When a range of values is allowed for $k$ (that is, $\mu_2(k) \neq 0$), these constraints no longer hold and the system can approach a more random organisation of polygons with $r \approx 0$. Varying the Lennard-Jones parameter $\epsilon_{HH}$ as in Sec. 3.4.4 and the number of tangled nodes as in Sec. 3.4.5 produces a more complicated behaviour. The simulations with tangled interactions lead to a cluster of points on the landscape with changing $\mu_2(k)$ but relatively similar $r$, demonstrating the control that is possible over simulated networks by changing the starting molecules. The effects of $\epsilon_{HH}$ lead to a similar clustering of landscape points in the region $r \approx -0.10$ and $0.25 \leq \mu_2(k) \leq 0.6$. This is because the energetics of network formation leave the short-range order relatively unaffected.

### Table 4: Width of the node coordination distribution as measured by its second moment $\mu_2(k)$ for simulations in which a certain percentage of linear molecules have been replaced with tangled molecules seen in Figure 10.

| $n_{\text{fig.10a}}$/% | $n_{\text{fig.10c}}$/% | $\mu_2(k)$ | $r$ | (Area)/nm$^2$ |
|----------------------|----------------------|------------|-----|---------------|
| 0                    | 2.5                  | 0.443      | -0.116 | 245743        |
| 0                    | 25                   | 0.257      | -0.082 | 173958        |
| 2.5                  | 0                    | 0.454      | -0.114 | 247026        |
| 2.5                  | 2.5                  | 0.417      | -0.104 | 235151        |
| 2.5                  | 25                   | 0.250      | -0.068 | 203734        |
| 25                   | 0                    | 0.281      | -0.070 | 205269        |
| 25                   | 2.5                  | 0.295      | -0.084 | 190625        |
| 25                   | 25                   | 0.215      | -0.083 | 145209        |
Varying the node coordination number distribution (representing short range order of polygons) against the second moment of the node in the network (which is thermally irreversible) and where two head groups encounter one another to create a node in the network. The rate of cooling, external pressure and strength of interaction are all important in controlling the nature of the final networks. By varying simple parameters allows for polymer flexibility, which encourages network formation as polymers can deform out of position to better coordinate into a network. The worm-like curve model works that can be formed, since varying simple parameters changes the balance between enthalpy (guiding the shape of polygons), entropy (affecting the ordering of polygons) and kinetics (locking networks into shape as they form).

The properties of the polymers themselves can affect the networks that are formed. The worm-like curve model allows for polymer flexibility, which encourages network formation as polymers can deform out of position to better coordinate into a network. The rate of cooling, external pressure and strength of interaction are all important in controlling the nature of the final networks. By varying key variables it has been demonstrated that a key factor in network formation is the kinetics of node-forming events where two head groups encounter one another to create a node in the network (which is thermally irreversible) and locked forming networks into place.

Finally, the networks formed in this work have filled in a gap established in previous work in the network landscape, and again highlight the critical balance of enthalpic, entropic and kinetic factors.

5. Acknowledgements

We are grateful for support from the EPSRC Centre for Doctoral Training in Theory and Modelling in Chemical Sciences (TMCS), under grant EP/L015722/1. Thanks to David Ormrod Morley, Elspeth Garman for many interesting discussions. MHJB is grateful for a University College Radcliffe Scholarship. This paper conforms to the RCUK data management requirements.

References

[1] R. M. Pedrigi, J. D. Humphrey, Computational Model of Evolving Lens Capsule Biomechanics Following Cataract-Like Surgery, Ann. Biomed. Eng. 39 (2011) 537–548. URL: http://link.springer.com/10.1007/s10439-010-0133-0.

[2] B. A. Christie, B. M. McMaster, J. W. Blaker, Accommodating Intraocular Lens, 1990.

[3] G. Duncan, I. Michael Wormstone, P. D. Davies, The aging human lens: Structure, growth, and physiological behaviour, Br. J. Ophthalmol. 81 (1997) 818–823. doi:10.1136/bjo.81.10.818.

[4] C. A. P. Cavallotti, L. Cerulli (Eds.), Age-Related Changes of the Human Eye, Humana Press, Totowa, NJ, 2008. URL: http://link.springer.com/10.1007/978-1-59745-507-7. doi:10.1007/978-1-59745-507-7.

[5] A. Duane, Studies in monocular and binocular accommodation with their clinical applications, Am. J. Ophthalmol. 5 (1922) 865–877. URL: http://dx.doi.org/10.1016/S0002-9394(22)90793-7. doi:10.1016/S0002-9394(22)90793-7.

[6] D. L. Bodian, R. J. Radner, S. Holbert, T. E. Klein, Molecular dynamics simulations of the full triple helical region of collagen type i provide an atomic scale view of the protein’s regional heterogeneity, Pacific Symp. Biocomput. 2011, PSB 2011 (2011) 193–204.

[7] R. A. Boerboom, K. N. Krahn, R. T. A. Megens, M. A. M. J. van Zandvoort, M. Merkx, C. V. C. Bouten, High resolution imaging of collagen organisation and synthesis using a versatile collagen specific probe, J. Struct. Biol. 159 (2007) 392–399. doi:10.1016/j.jsb.2007.04.008.

[8] V. M. Sueiras, V. T. Moy, N. M. Ziebarth, Lens capsule structure assessed with atomic force microscopy., Mol. Vis. 21 (2015) 316–23. URL: http://www.ncbi.nlm.nih.gov/pubmed/25814829 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4361358.

[9] H. J. Burd, R. A. Regueiro, Finite element implementation of a multiscale model of the human lens capsule, Biomech. Model. Mechanobiol. 14 (2015) 1363–1378. URL: http://link.springer.com/10.1007/s10237-015-0680-2. doi:10.1007/s10237-015-0680-2.

[10] C. P. Broodersz, F. C. Mackintosh, Modeling semiflexible polymer networks, Rev. Mod. Phys. 86 (2014) 995–1036. doi:10.1103/RevModPhys.86.995. arXiv:1404.4332.

[11] R. C. Picu, Mechanics of random fiber networks - A review, Soft Matter 7 (2011) 6768–6785. doi:10.1039/c1sm05022b.

[12] M. H. J. Bailey, D. Ormrod Morley, M. Wilson, Simplified computational model for generating biologically realistic in vitro lens capsule structure assessed with atomic force microscopy., J. Struct. Biol. 159 (2007) 392–399. doi:10.1016/j.jsb.2007.04.008.

[13] K. Ellington, D. Cogrove, Assembly of Type IV collagen. Insights from a3(IV) collagen-deficient mice, J. Biol. Chem. 275 (2000) 12719–12724. doi:10.1074/jbc.275.17.12719.
