

journal homepage: [www.elsevier.com/locate/csbj](http://www.elsevier.com/locate/csbj)

# Self assembly of model polymers into biological random networks

Matthew H.J. Bailey, Mark Wilson\*

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



## ARTICLE INFO

### Article history:

Received 28 October 2020  
Received in revised form 1 February 2021  
Accepted 1 February 2021  
Available online 12 February 2021

### Keywords:

Continuous random network  
Self assembly  
Polygon statistics  
Polygon distributions  
Collagen network

## 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 are 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.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 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 [10–12]. These graph approaches are parameterised and based on experimental data and small-scale simulations, but provide an incomplete link

\* Corresponding author.

E-mail address: [mark.wilson@chem.ox.ac.uk](mailto:mark.wilson@chem.ox.ac.uk) (M. Wilson).

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]. 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 placed in the context of the range of experimental data available from microscope images of 2D collagen IV networks, and are shown to be more appropriate for reproducing these experimental networks when compared with previous continuous random (entropically-driven) networks.

## 2. Methods.

### 2.1. Network structures

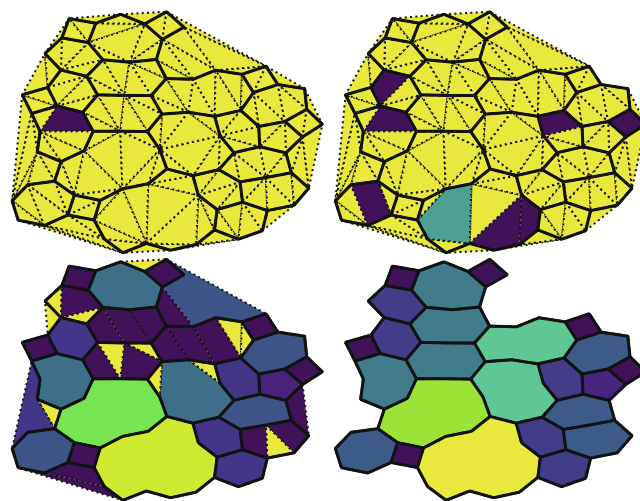
Network theory has been long used to analyse biological structures ranging from hexagonal arrangements of cells to graph models of biopolymers [17,10]. The existence of such a wide range of network models indicates that it is first helpful to establish a method to describe 2D networks rigorously, here 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 [18]. 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 [19,20]. Viewing a 2D network as being constructed of polygons allows use of consistent metrics, applicable across many different types of network including those generated via a self-assembly model, graph theory or experimental microscope images.

For example, the preference of polygons to be adjacent to similar or dissimilar polygons (described, for example, by the assortativity, and first discussed by [21] to describe social networks [21]) describes the short-range structure of a network and improves on the Aboav-Weaire parameter traditionally used to describe chemical 2D networks [22,20]. 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 [23], 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 [24], and also in Le Roux and Jund [25]. 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 [24]. 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 [Fig. 1](#). If edges that are not in the original graph are systematically removed, updating the polygons at each stage, it is possible to construct a set of unique polygons for the network that covers the entire area and maps neatly to an intuitive definition of polygons.



**Fig. 1.** Snapshots from the process of constructing the polygon graph. Dashed edges are those that only exist in the Delaunay triangulation (top left), which are removed one-by-one (top right then bottom left). Removing a dashed edge merges two polygons together, and the final network structure (bottom right) is left after all dashed edges are removed. The polygons are coloured by their number of edges.

## 2.2. The simplified model

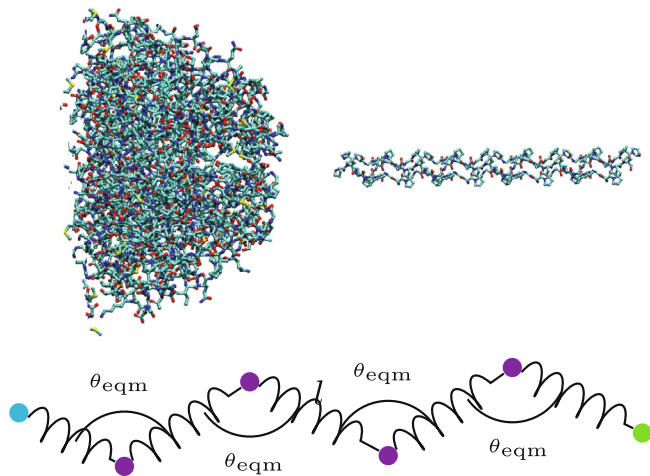
The assembly of collagen-like molecules was simulated using a model inspired by a combination of “worm-like” curve polymers and patchy particles. Patchy particles are colloids with attractive regions on their surface which have been shown experimentally and computationally to self-assemble [26,27], and mimic the crystallisation properties of proteins [28]. 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 Fig. 2. This is a coarse-graining of a worm-like curve model, which have been used successfully in modelling the stiffness of biopolymers [29,30]. 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  $\epsilon_{BB}$  and a range  $\sigma_{BB}$

$$U_{BB}(r) = \begin{cases} 4\epsilon_{BB} \left[ \left( \frac{\sigma_{BB}}{r} \right)^{12} - \left( \frac{\sigma_{BB}}{r} \right)^6 \right] & r < \sigma \\ 0 & r \geq \sigma. \end{cases} \quad (1)$$

Two head groups interact with one another, according to a Lennard-Jones potential with an energy scale  $\epsilon_{HH}$  and a range  $\sigma_{HH}$  (representing an effective size of the bead),

$$U_{HH}(r) = 4\epsilon_{HH} \left[ \left( \frac{\sigma_{HH}}{r} \right)^{12} - \left( \frac{\sigma_{HH}}{r} \right)^6 \right]. \quad (2)$$

$\sigma_{BB}$  and  $\sigma_{HH}$  are chosen to prevent molecules from interpenetrating and to favour  $k = 3$  coordination. Body-head interactions are set to zero throughout ( $U_{BH} = 0$ ). 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



**Fig. 2.** Two schematic parts of a collagen IV molecule, representing the simplification in the coarse graining process. The left-hand image represents a trimer in an NC1 domain, approximately 5000 atoms which are simplified into a single green head bead in the polymer representation [31]. The right hand image (shown on a different scale) shows one representative section of a triple helix joining the two ends of the collagen IV molecule, which is represented by two purple body beads and a joining spring [32]. The blue head bead at the other end of the polymer represents a 7S domain, which is not pictured here.

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

$$\vec{F} = \underbrace{\vec{F}_{\text{Bond}} + \vec{F}_{\text{Angle}} + \vec{F}_{\text{LJ}}}_{\text{Conservative Forces}} - \underbrace{\frac{m\vec{v}}{\gamma}}_{\text{Drag}} + \underbrace{\xi \sqrt{\frac{k_B T m}{\gamma dt}}}_{\text{Brownian}} \quad (3)$$

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. In Section 3.4.2, the thermostat desired temperature  $T$  is changed over a period of time denoted  $t_{\text{cool}}$ .

## 2.3. Simulation protocol

This work used LAMMPS to perform Molecular Dynamics simulations due to the ease of use and ready support for 2D systems [33]. Molecular Dynamics methods are well suited for studying self assembly methods, as they repeatedly numerically solve Newton's equations of motions and thus naturally capture dynamic behaviour of a complex system. Unless otherwise mentioned, each set of parameters was sampled in ten different configurations of  $20 \times 20$  molecules in a periodic simulation cell, initially placed on a square grid. The grid arrangement was equilibrated by simulation at  $T = 30$  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 simulation cell size was relaxed using an  $NpT$  ensemble (i.e. constant number of particles, pressure, and temperature with the temperature controlled by a Nosé-Hoover thermostat instead of the Langevin thermostat) at  $T = 30$  K with  $p = 20$  Pa for a further  $t = 40$   $\mu\text{s}$ . This pressure was chosen based on parameter scans (discussed more in Section 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 cell 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 300K by adjusting  $T$  in Eq. (3) over 100  $\mu\text{s}$ , 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  $\mu\text{s}$  to 30 K. The total time from  $T = 300$  K to 30 K was recorded as the value  $t_{\text{cool}}$ . 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 ini-

tial scan of the effective parameter space). These default parameters were bond and angle energy scales as  $k_l = 1.657 \times 10^{-4} \text{ Nm}^{-1}$ ,  $K_\theta = 200 \times 10^{-21} \text{ J}$ , a cooling time of  $t_{\text{cool}} = 100 \mu\text{s}$ , Lennard-Jones head-head and body-body energy scales were  $\epsilon_{\text{HH}} = 4.142 \times 10^{-21} \text{ J}$ ,  $\epsilon_{\text{BB}} = 16.142 \times 10^{-21} \text{ J}$ . The length scales in the simulation were such that the equilibrium polymer length was  $l_{\text{eqm}} = 300 \text{ nm}$ , and the Lennard-Jones length scales were  $\sigma_{\text{HH}} = 50 \text{ nm}$  and  $\sigma_{\text{BB}} = 137.5 \text{ nm}$ . The polymer length is similar to the observed length of collagen IV molecules, which has been reported to be in the range 300nm to 400nm [34,35]. The body bead interactions are considerably larger than the actual width of a collagen IV triple helix, 1.5nm, as they also take into account the excluded volume effect of polymers where close coordination is entropically unfavourable. The Lennard-Jones length scales were chosen to favour  $k = 3$  coordination geometrically akin to the design of patchy particles. These parameters were chosen to promote the assembly of the coarse grained units into networks that best matched observed images, while avoiding “over-fitting” to the relatively sparse experimentally observed macroscopic properties. The parameters used in this work are generic by choice, and aim to reproduce as wide a variety of network types as possible, reflecting the variance in microscope images of biological networks – for example, images of collagen IV networks can vary dramatically depending on the environmental conditions or self assembly conditions [34,36,15,14]. Future refinements of this model could more specifically reproduce properties of interest of a given biological network; for example, choosing  $k_l$  to reproduce mechanical properties or to assign the stiffness of the angular potential based on experimental data as has been highlighted for chromosomes [37,38].

### 3. Results and discussion

The polymer model described has been developed so as to be deliberately generic in order to capture as wide a range of biological network structure as possible. Current images of biological networks show an amazing diversity of form, structure and function. A number of previous approaches have been used to explore biological networks, including entropically driven random network methods and modelling random networks as being idealised structures. We compare the polymer network model to those different approaches before presenting the degree of control and improvements available in the proposed model.

#### 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 polygon 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 [39,20]:

$$\sum_n p_n = 1 \quad (4a)$$

$$\sum_n n p_n = \langle n \rangle \quad (4b)$$

$$\sum_n \frac{p_n}{n} = \text{constant} \quad (4c)$$

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.

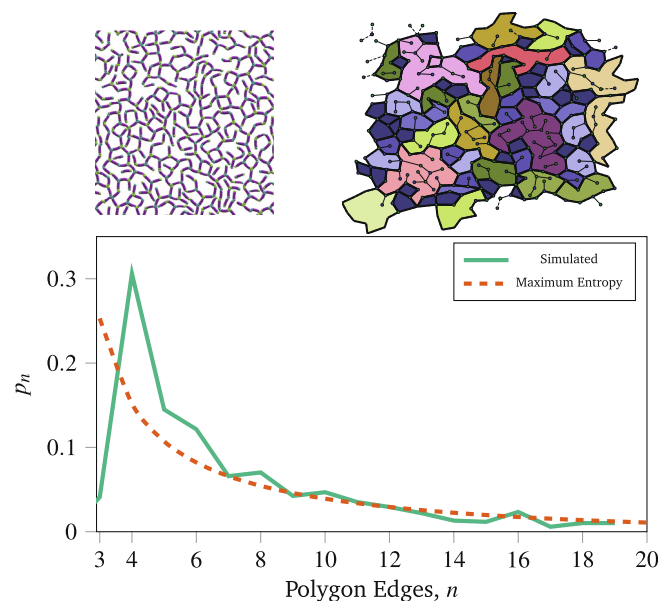
The self-assembled networks show polygon edge count distributions which do not resemble the numerical maximum entropy model; this is visible in Fig. 3, which compares polygon edge counts from a polymer simulated and the polygon edge counts from maximum entropy model distribution. This was based on simulations described in Section 3.4.1 at  $k_l = 1.5 \times 10^{-4} \text{ Nm}$ . 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 coordina-



**Fig. 3.** The process of applying network statistics to a network generated by the coarse graining procedure, here with the metric being polygon edge counts. First, a snapshot of a coarse grained simulation is taken (top left panel) and the polygons are identified (top right panel). Finally, the fraction of polygons with  $n$  edges (denoted  $p_n$ ) is compared to the distribution predicted by maximum entropy (red dashed line).

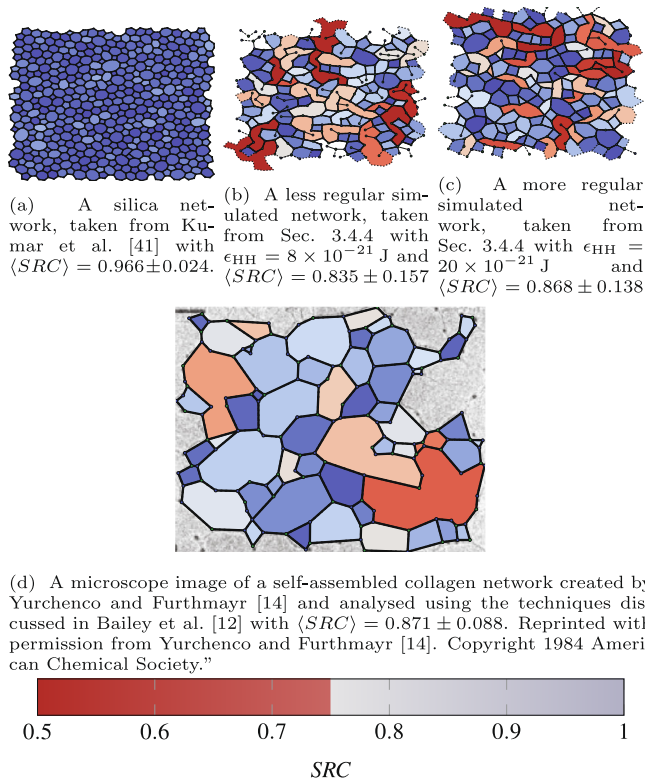
tion numbers, more flexible angular potentials and even curved edges. To quantify this difference in shape regularity, we make use of a metric originally developed for computer graphics [40]. The shape regularity coefficient ( $SRC$ ) of a shape  $\mathcal{S}$ , is defined as

$$SRC(\mathcal{S}) = SO(\mathcal{S})V_{xy}(\mathcal{S})CO(\mathcal{S}) \quad (5)$$

in which  $SO(\mathcal{S})$  is the solidity, defined as the ratio of the polygon area to the area of its convex hull;  $V_{xy}(\mathcal{S})$  is the balanced repartition of the shape, defined as  $\sqrt{\frac{\min(\sigma_x, \sigma_y)}{\max(\sigma_x, \sigma_y)}}$  where  $\sigma_{x(y)}$  is the standard deviation of the  $x(y)$  coordinate of the vertex positions; and finally  $CO(\mathcal{S})$  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  $SRC = 1$ . Less-regular shapes, such as ellipses or rectangles, have  $0 \leq SRC < 1$ . Fig. 4 shows four example networks with varying degrees of regularity as defined by the  $SRC$ , with each polygon coloured by  $SRC$ . The silica network in Fig. 4a [41] shows highly regular polygons, corresponding to  $\langle SRC \rangle = 0.966 \pm 0.024$ , 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  $\langle SRC \rangle \approx 0.96$  with relatively narrow distributions (small standard deviations). By comparison biological networks show  $\langle SRC \rangle$  values which are significantly smaller. For example, networks generated in the present work show  $\langle SRC \rangle$  in the range  $0.747 \leq \langle SRC \rangle \leq 0.928$  (see Table SI 2). In a previous publication a number of experimental images were analysed [12]. The microscope images of collagen IV networks showed that the polygons formed were often irregular.

Table SI 1 lists the values of  $\langle SRC \rangle$  obtained from 21 such images which show values in the range  $0.69 < \langle SRC \rangle < 0.91$ . As a visual example, Fig. 4d shows a section of a network from Yurchenco and Furthmayr [14] which corresponds to  $\langle SRC \rangle = 0.871 \pm 0.088$ .



**Fig. 4.** A comparison of an inorganic network Fig. 4a and two simulated biological networks [Fig. 4b and Fig. 4c]. The color ramp of red to blue represents the range of irregular ( $SRC = 0.5$ ) to regular ( $SRC = 1$ ) polygons.

The greater flexibility of the biological networks, compared to the silica, is highlighted by the broader range of  $SRC$  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  $SRC$  values are more flexible angular terms around nodes in the graph which can lead to concave shapes. These concave shapes can be seen in Fig. 4b and Fig. 4d as the long shapes, coloured pale. In comparison, all polygons in Fig. 4a are convex and approximately isotropic. A full table of data with  $\langle SRC \rangle$  values is available in the Supplementary Information as Table SI 2.

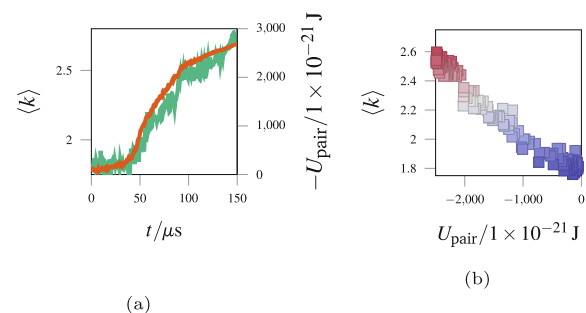
### 3.2.2. Coordination number and internal energy

The energetics of the networks generated by the polymer self-assembly model, while interesting, are highly dependent on the potential energy of interaction between polymers. The short-range nature of the Lennard-Jones potential used in this work means that it is possible to use the node coordination numbers 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.

Fig. 5 shows one example of the correlation between the node coordination number and the internal energy. Fig. 5a shows the evolution of the internal energy and mean node coordination number for a simulation taken from Section 3.4.2, cooled over 100  $\mu$ s. Fig. 5b further highlights the correlation between  $U_{pair}$  and  $\langle k \rangle$ , with scatter points being the positions of  $\langle k \rangle$  and  $U_{pair}$  shown at different time points in Fig. 5a. Similar correlations exist across all the simulations we performed. The linear link between  $U_{pair}$  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. [16,34] favour a “chain-link fence” network, effectively a square-net structure. Burd [42], has alternatively suggested a primarily hexagonal network. Yurchenco and Furthmayr suggest a disordered collagen IV network interacting with an ordered scaffold [14,43]. Imaging experiments have shed some light on the structure of the networks. However, complicating factors such as the biological origins of the networks, the pres-



**Fig. 5.** Two figures to highlight the strong correlation between the mean coordination number  $\langle k \rangle$  and the pairwise head-group interaction energy  $U_{pair}$ . Fig. 5a shows  $\langle k \rangle$  as a function of time for a system undergoing network self assembly. Fig. 5b shows a scatter plot of  $\langle k \rangle$  against  $U_{pair}$  to emphasize their correlation. Since  $\langle k \rangle$  dominates the internal energy, we can use  $\langle k \rangle$  as a simple proxy for the energy.

ence 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 Fig. 6a and Fig. 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  $k_i$  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 the parameters of underlying coarse grained polymer model 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 [30]. Here the impact of both the molecular stiffness, controlled by varying the energy scale of the harmonic bonds  $k_i$ , 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 Section 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  $k_i$ , 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. [44]. 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  $k_i$  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 [34]. 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 formation. These data are available in Table 1, and justify the importance of a worm-like curve model to network formation.

#### 3.4.2. Dynamical cooling time

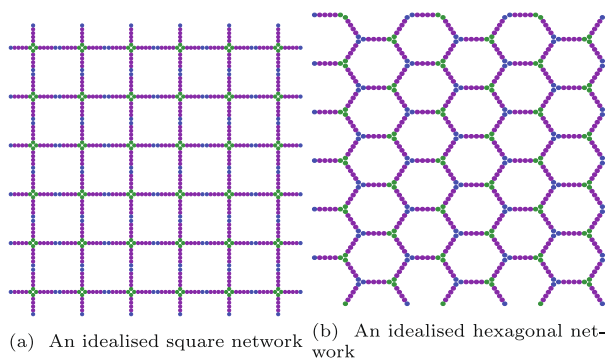
The timescales of self assembly are difficult to control in a laboratory setting, but relatively simple to control computationally. The freezing rate of the simplified polymers as they formed into networks was controlled by varying the time the polymers had to cool (referred to as  $t_{\text{cool}}$  in Section 2.3) from  $t_{\text{cool}} = 54 \mu\text{s}$  to  $t_{\text{cool}} = 150 \mu\text{s}$ . When frozen rapidly, many fewer polygons were formed, varying from 31.6 polygons on average when cooled over  $t_{\text{cool}} = 54 \mu\text{s}$ , to 105.0 on average when cooled over  $t_{\text{cool}} = 150 \mu\text{s}$ . Fig. 7 shows the total number of polygons formed as a function of the cooling time, with all other parameters as discussed in Section 2.3. In addition, the figure shows two example networks to highlight the origin of the differences. The cooling solvent locks polymers into place before they can form full polygons, and many polymers end up “dangling”. As the cooling rate slows, the polymers have more time to slot into place in the evolving network.

When the polymer networks are cooled for longer, the average edge length and polygon areas remain similar — however, 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 Fig. 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.

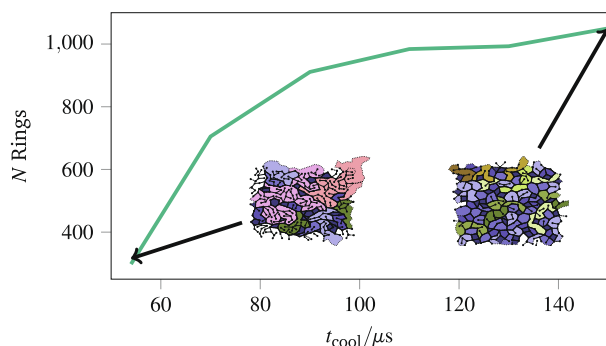
**Table 1**

The mean edge length  $\langle l \rangle$ , number of polygons  $N_{\text{polygon}}$ , and the standard deviation of the number of polygons  $\sigma_{N_{\text{polygon}}}$  counted across 10 simulations as the angular strength  $K_\theta$  changes.

$K_\theta / 1 \times 10^{-21} \text{ J}$	$\langle l \rangle / \text{nm}$	$N_{\text{polygon}}$	$\sigma_{N_{\text{polygon}}}$
5	294.62	79.2	8.64
50	311.65	78.3	4.76
100	314.64	71.1	6.06
150	316.96	68.2	5.83
200	318.27	71.6	4.48
250	318.11	69.0	7.04
300	318.43	69.9	7.58
400	318.51	66.0	3.65



**Fig. 6.** A comparison of two idealised networks, one square net (left panel — where each edge is two molecules long) akin to a chain-link fence, and hexagons akin to a honeycomb (right panel — where each edge is one molecule long).



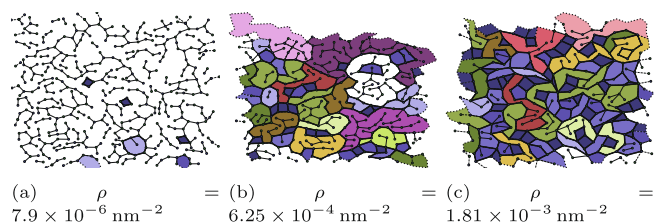
**Fig. 7.** The number of polygons formed,  $N$ , against the time taken to cool the networks from 300K to 30K, denoted  $t_{\text{cool}}$  with two networks cooled at different rates inset as examples. Different colours are used to highlight polygons with the same number of edges.

### 3.4.3. Molecular density

The huge diversity in the structures of biological networks is matched by a large diversity in network densities. To explore the effects of that diversity on the network structure, the molecule density (molecules per unit area) was varied whilst they were forming.

However, instead of assigning a potentially non-physical density *a priori*, we introduced an isothermal-isobaric ( $NpT$ ) step in the simulation protocol, which allowed the periodic simulation cell size to be physically-determined rather than assigned based in relatively sparse data. After the simulation cell size was set in the  $NpT$  ensemble, it was thereafter fixed at the final value and the pressure coupling removed. The pressure in this section should not be interpreted directly, but instead as a proxy for the molecule density, for two reasons. The first is that nature of the simulation means that the position of particles are described by standard 3D Cartesian coordinates but images in the  $z$  direction are effectively infinitely separated. This means that pressure is a more convenient physical measure than stress, which would be strictly accurate for 2D systems. The second is that effects of the implicit solvent are neglected when performing the  $NpT$  step as the Langevin thermostat is disabled, and a Nosé-Hoover thermostat is applied in its place as described in Section 2.3.

Below a pressure of  $p = 2$  Pa at the lowest molecule densities no networks form. At low molecule densities 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 Fig. 8. Networks formed at intermediate densities often had dangling edges, and networks at high densities 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



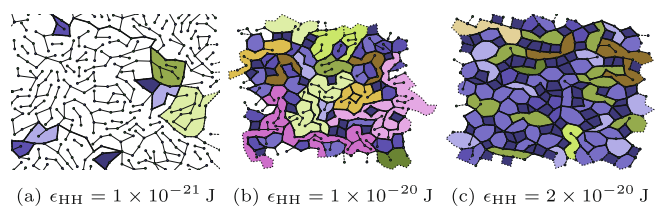
**Fig. 8.** A comparison of three networks cooled at low, intermediate, and high molecule densities. Different colours are used to highlight polygons with the same number of edges, and these networks are drawn at different scales.

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 [45], 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 of self assembling polymers, 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 Fig. 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



**Fig. 9.** A comparison of three networks formed at low, intermediate, and high Lennard-Jones well depths. Polygons that have the same number of edges are coloured the same.

**Table 2**

Average molecules per unit area ( $\rho$ ), average coordination number ( $\langle k \rangle$ ), average edges per polygon ( $\langle n \rangle$ ), and the percentage of dangling edges (D.E.) which do not form parts of polygons, as a function of molecular density.

$p/\text{Pa}$	$\langle \rho \rangle / \text{nm}^{-2}$	$\langle k \rangle$	$\langle n \rangle$	D.E. /%
5	$4.60 \times 10^{-4}$	2.588	6.789	62.6
10	$6.25 \times 10^{-4}$	2.691	7.039	47.5
20	$1.01 \times 10^{-3}$	2.799	6.823	36.55
50	$1.81 \times 10^{-3}$	3.062	6.034	23.875

**Table 3**

Average polygon area, average edges per polygon ( $\langle n \rangle$ ), average coordination number ( $\langle k \rangle$ ) of networks as a function of Lennard-Jones well depth.

$\epsilon_{\text{HH}} / 1 \times 10^{-21} \text{ J}$	$\langle \text{Area} \rangle / \text{nm}^2$	$\langle n \rangle$	$\langle k \rangle$
2	720321	6.422	2.69
4	386607	6.892	2.82
6	332060	6.923	2.87
8	282086	6.564	2.95
10	261117	6.409	2.98
12	257307	6.375	2.98
14	241528	6.273	3.01
16	231528	6.095	3.03
20	217023	5.925	3.05

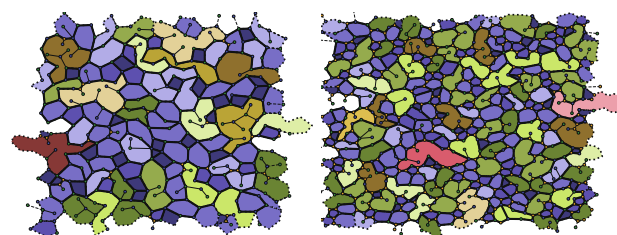
reward for head groups interacting in this regime meant the average coordination number approached 3, and more small polygons were formed. Data on these networks are available in Table 3 which bear a significant resemblance to the results in Table 2. This is because both  $p$  and  $\epsilon_{HH}$  affect the ratio between the kinetic energy of monomers and the potential energy gained in forming a network. By changing this ratio, the balance shifts between enthalpy guiding network formation towards a thermodynamically rewarding arrangement, and the kinetics locking a network into place as it forms. Studies of real collagen IV networks suggest that the coordination of 7S tetramers cannot be thermally reversed, and that thermal irreversibility is important to network formation [14].

### 3.4.5. Pre-seeding lateral “Tangling” interactions

One reason biopolymers exhibit a richness in network structure is that their interactions are not limited to the terminal interactions of head groups. For example, as 3D objects, collagen IV molecules may tangle around one another. Yurchenco and Ruben argue that these lateral tangling interactions are key to forming an amorphous polygon network [43,46]. In a 2D simulation it is possible 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 Fig. 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 interactions with one another and springs being harmonic potentials between beads in a molecule.

The simulation procedure was followed as described above, but with between 0% and 50% of the linear polymers substituted for their tangled counterparts. Two example networks can be seen in Fig. 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 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 Fig. 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 Section 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



(a) A network with 20 pre-tangled molecules out of 400. (b) A network with 200 pre-tangled molecules out of 400.

**Fig. 11.** Comparison of networks formed with pre-tangled polymers. The nodes are coloured by the same scheme as used in Fig. 10.

**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 Fig. 10.

$n_{\text{Fig. 10b}}/\%$	$n_{\text{Fig. 10c}}/\%$	$\mu_2(k)$	$r$	$\langle \text{Area} \rangle / \text{nm}^2$
0	2.5	0.443	-0.116	245 473
0	25	0.257	-0.082	173 958
2.5	0	0.454	-0.114	247 026
2.5	2.5	0.417	-0.104	235 151
2.5	25	0.250	-0.068	203 734
25	0	0.281	-0.070	205 269
25	2.5	0.295	-0.084	190 625
25	25	0.215	-0.083	145 209

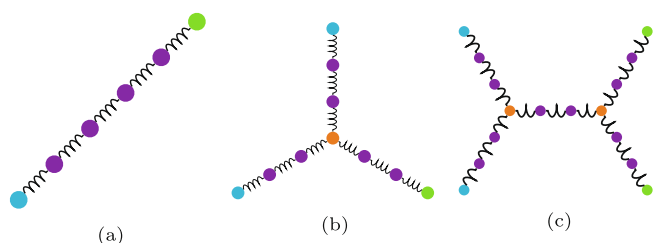
could better capture the properties of physical networks and allow for a potential model with head group preferences that better reflects real networks.

### 3.5. The network landscape

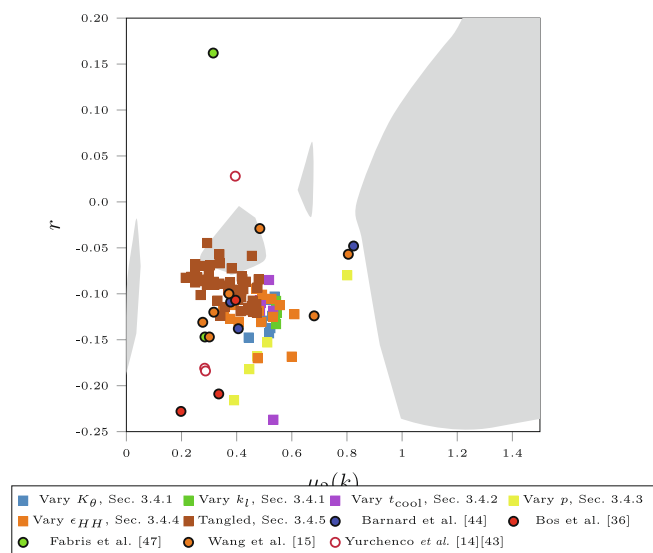
The simulation results discussed here can be viewed in the context of Bailey et al. [12]. In that work the formation of disordered collagen-like networks was studied starting from a regular hexagonal network and systematically switching the bonds, maintaining an average value of  $\langle k \rangle$ , effectively modelling 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 and distribution widths of  $k$ .

Fig. 12 shows a “network landscape” similar to that in Bailey et al. [12], capturing  $\mu_2(k)$  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. [44,36,47,15,14,43]). 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  $k$ . Fig. 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-



**Fig. 10.** Three example molecules used in the simulations, demonstrating the ability to seed a simulation with pre-tangled interactions. Purple circles represent body beads, contributing to  $U_{BB}$ , and green/blue beads are head beads that contribute to  $U_{HH}$ , similar to those shown in Fig. 6a. Finally, orange circles represent tangling sites (which can be seen in action in Fig. 11), but are otherwise identical to body beads. The gap from a blue/green head bead to an orange tangling site is 150nm, which matches a distance between interaction sites observed by Yurchenco and Furthmayr [14].



**Fig. 12.** A “network landscape”, showing the assortativity  $r$  (representing short range order of polygons) against the second moment of the node coordination number distribution  $\mu_2(k)$  (representing the range of node types). The simulated points in coloured squares match experimental data points (circles) better than previous work presented in Bailey et al. [12], which are represented as shaded regions.

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 disassortative) as  $\mu_2(k)$  increases. This is most obvious in the results from Section 3.4.2 and Section 3.4.3, and also holds for the results from Section 3.4.1. This relationship can be explained by reference to the assortativity of polygons formed by a random point process, which is  $r \approx -0.15$  [20]. 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 Section 3.4.4 and the number of tangled nodes as in Section 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.

#### 4. Conclusion

In conclusion, a highly simplified model for collagen IV has been constructed which is shown to self assemble to form well-defined biological networks. The “worm-like” polymer model features physically meaningful parameters that allow control over the networks formed. The polymer networks are significantly different from typical inorganic networks, for example easily forming convex, irregular polygon structures. There is a rich range of networks 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.

#### CRediT authorship contribution statement

**Matthew H.J. Bailey:** Data curation, Investigation, Software, Visualization, Writing - original draft. **Mark Wilson:** Conceptualization, Project administration, Supervision, Writing - review & editing.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### 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.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.csbj.2021.02.001>.

#### References

- [1] Pedrigi RM, Humphrey JD. Computational model of evolving lens capsule biomechanics following cataract-like surgery. *Ann Biomed Eng* 2011;39(1):537–48. <https://doi.org/10.1007/s10439-010-0133-0>. ISSN 0090-6964.
- [2] Christie BA, McMaster BM, Blaker JW. Accommodating Intraocular Lens 1990.
- [3] Duncan G, Michael Wormstone I, Davies PD. The aging human lens: Structure, growth, and physiological behaviour. *Br J Ophthalmol* 1997;81(10):818–23. <https://doi.org/10.1136/bjo.81.10.818>. ISSN 00071161.
- [4] Cavallotti CAP, Cerulli L, Editors. Age-Related Changes of the Human Eye, Humana Press, Totowa, NJ, ISBN 978-1-934115-55-8, 2008. <https://doi.org/10.1007/978-1-59745-507-7..>
- [5] Duane A. Studies in monocular and binocular accommodation with their clinical applications. *Am J Ophthalmol* 1922;5(11):865–77. [https://doi.org/10.1016/S0002-9394\(22\)90793-7](https://doi.org/10.1016/S0002-9394(22)90793-7). ISSN 00029394.
- [6] Bodian DL, Radmer RJ, Holbert S, Klein TE. 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. [https://doi.org/10.1142/9789814335058\\_0021..](https://doi.org/10.1142/9789814335058_0021..)
- [7] Boerboom RA, Krahn KN, Megens RTA, van Zandvoort MAMJ, Merckx M, Bouten CVC. High resolution imaging of collagen organisation and synthesis using a versatile collagen specific probe. *J Struct Biol* 2007;159(3):392–9. <https://doi.org/10.1016/j.jsb.2007.04.008>. ISSN 10478477.
- [8] Sueiras VM, Moy VT, Ziebarth NM. Lens capsule structure assessed with atomic force microscopy. *Mol Vis* 21(July 2014) (2015) 316–23. ISSN 1090-0535..

- [9] Burd HJ, Regueiro RA. Finite element implementation of a multiscale model of the human lens capsule. *Biomech Model Mechanobiol* 2015;14(6):1363–78. <https://doi.org/10.1007/s10237-015-0680-2>. ISSN 1617-7959.
- [10] Broedersz CP, Mackintosh FC. Modeling semiflexible polymer networks. *Rev Mod Phys* 2014;86(3):995–1036. <https://doi.org/10.1103/RevModPhys.86.995>. ISSN 15390756.
- [11] Picu RC. Mechanics of random fiber networks – a review. *Soft Matter* 2011;7(15):6768–85. <https://doi.org/10.1039/c1sm05022b>. ISSN 1744683X.
- [12] Bailey MHJ, Ormrod Morley D, Wilson M. Simplified computational model for generating biological networks. *RSC Adv* 2020;10(63):38275–80. <https://doi.org/10.1039/D0RA06205C>. ISSN 2046–2069.
- [13] Kalluri R, Cosgrove D. Assembly of Type IV collagen Insights from  $\alpha 3(\text{IV})$  collagen-deficient mice. *J Biol Chem* 2000;275(17):12719–24. <https://doi.org/10.1074/jbc.275.17.12719>. ISSN 00219258.
- [14] Yurchenco PD, Furthmayr H. Self-assembly of basement membrane collagen. *Biochemistry* 1984;23(8):1839–50. <https://doi.org/10.1021/bi00303a040>. ISSN 0006–2960.
- [15] Wang Z, Xiao Q, Song X, Wan Y, Zhu J. Cation-specific effects on the self-assembly of collagen molecules mediated by acetate on mica surface observed with atomic force microscopy. *J Food Qual* 2017;20177:1–10. <https://doi.org/10.1155/2017/1692975>. ISSN 0146–9428.
- [16] Cummings CF, Pedchenko V, Brown KL, Colon S, Rafi M, Jones-Paris C, Pokydeshaeva E, Liu M, Pastor-Pareja JC, Stothers C, Ero-Tolliver IA, McCall AS, Vanacore R, Bhavé G, Santoro S, Blackwell TS, Zent R, Pozzi A, Hudson BG, et al. Extracellular chloride signals collagen IV network assembly during basement membrane formation. *J Cell Biol* 2016;213(4):479–94. <https://doi.org/10.1083/jcb.201510065>. ISSN 0021–9525.
- [17] Lewis FT. The effect of cell division on the shape and size of hexagonal cells. *Anat Rec* 1926;33(5):331–55. <https://doi.org/10.1002/ar.1090330502>. ISSN 0003–276X.
- [18] Tóth LF. What the bees know and what they do not know. *Bull Am Math Soc* 1964;70(4):468–82. <https://doi.org/10.1090/s0002-9904-1964-11155-1>. ISSN 0002–9904.
- [19] Zachariasen WH. The atomic arrangement in glass. *J Am Chem Soc* 1932;54(10):3841–51. <https://doi.org/10.1021/ja01349a006>. ISSN 15205126.
- [20] Ormrod Morley D, Thorneywork A, Dullens R, Wilson M. Generalised network theory of physical two-dimensional systems. *Phys Rev E* 2020;1–13. <https://doi.org/10.1103/PhysRevE.101.042309>.
- [21] Newman MEJ. Assortative Mixing in Networks. *Phys Rev Lett* 2002;89(20):208701. <https://doi.org/10.1103/PhysRevLett.89.208701>. ISSN 0031–9007.
- [22] Aboav DA. The arrangement of cells in a net. *Metallography* 1980;13(1):43–58. [https://doi.org/10.1016/0026-0800\(80\)90021-X](https://doi.org/10.1016/0026-0800(80)90021-X). ISSN 00260800.
- [23] Charonis AS, Tsilibary EC, Yurchenco PD, Furthmayr H, Coritz A. Binding of laminin to type IV collagen: a morphological study. *J Cell Biol* 1985;100(6):1848–53. <https://doi.org/10.1083/jcb.100.6.1848>. ISSN 15408140.
- [24] Yuan X, Cormack AN. Efficient algorithm for primitive ring statistics in topological networks. *Comput Mater Sci* 2002;24(3):343–60. [https://doi.org/10.1016/S0927-0256\(01\)00256-7](https://doi.org/10.1016/S0927-0256(01)00256-7). ISSN 09270256.
- [25] Le Roux S, Jund P. Ring statistics analysis of topological networks: new approach and application to amorphous GeS<sub>2</sub> and SiO<sub>2</sub> systems. *Comput Mater Sci* 2010;49(1):70–83. <https://doi.org/10.1016/j.commatsci.2010.04.023>. ISSN 09270256.
- [26] Zhang Z, Glotzer SC. Self-assembly of patchy particles. *Nano Lett* 2004;4(8):1407–13. <https://doi.org/10.1021/nl0493500>. ISSN 15306984.
- [27] Zhao Y, Berger R, Landfester K, Crespy D. Polymer patchy colloids with sticky patches. *Polym Chem* 2014;5(2):365–71. <https://doi.org/10.1039/C3PY01096A>. ISSN 1759–9954.
- [28] Fusco D, Charbonneau P. Crystallization of asymmetric patchy models for globular proteins in solution. *Soft Matter Phys* 2013;88(1):1–9. <https://doi.org/10.1103/PhysRevE.88.012721>. ISSN 15393755.
- [29] Underhill PT, Doyle PS. On the coarse-graining of polymers into bead-spring chains. *J Nonnewton Fluid Mech* 2004;122(1–3):3–31. <https://doi.org/10.1016/j.jnnfm.2003.10.006>. ISSN 03770257.
- [30] Kirkness MWH, Lehmann K, Forde NR. Mechanics and structural stability of the collagen triple helix. *Curr Opin Chem Biol* 2019;53:98–105. <https://doi.org/10.1016/j.cbpa.2019.08.001>. ISSN 18790402.
- [31] Vanacore RM, Shanmugasundararaj S, Friedman DB, Bondar O, Hudson BG, Sundaramoorthy M. The  $\alpha 1\alpha 2$  network of collagen IV reinforced stabilization of the noncollagenous domain-1 by noncovalent forces and the absence of met-lys cross-links. *J Biol Chem* 2004;279(43):44723–30. <https://doi.org/10.1074/jbc.M406344200>. ISSN 00219258.
- [32] Berisio R, Vitagliano L. Crystal structure of the collagen triple helix model 3. *Protein Sci* 2002;11:262–70. <https://doi.org/10.1110/ps.32602.the>.
- [33] Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys* 1995;117(1):1–19. <https://doi.org/10.1006/jcph.1995.1039>. ISSN 00219991.
- [34] Timpl R, Wiedemann H, Van Delden V, Furthmayr H, Kühn K. A Network Model for the Organization of Type IV Collagen Molecules in Basement Membranes. *Eur J Biochem* 1981;120(2):203–11. <https://doi.org/10.1111/j.1432-1033.1981.tb05690.x>. ISSN 0014–2956.
- [35] Dölz R, Engel J, Kühn K. Folding of collagen IV. *Eur J Biochem* 1988;178(2):357–66. <https://doi.org/10.1111/j.1432-1033.1988.tb14458.x>. ISSN 0014–2956.
- [36] Bos KJ, Holmes DF, Meadows RS, Kadler KE, McLeod D, Bishop PN. Collagen fibril organisation in mammalian vitreous by freeze etch/rotary shadowing electron microscopy. *Micron* 2001;32(3):301–6. [https://doi.org/10.1016/S0968-4328\(00\)00035-4](https://doi.org/10.1016/S0968-4328(00)00035-4). ISSN 09684328.
- [37] Shinkai S, Nakagawa M, Sugawara T, Togashi Y, Ochiai H, Nakato R, Taniguchi Y, Onami S. PHI-C: deciphering Hi-C data into polymer dynamics. *NAR Genomics Bioinform* 2020;2(2):1–10. <https://doi.org/10.1093/nargab/lqaa020>.
- [38] Le Treut G, Képès F, Orland H. A polymer model for the quantitative reconstruction of chromosome architecture from HiC and GAM data. *Biophys J* 2018;115(12):2286–94. <https://doi.org/10.1016/j.bpj.2018.10.032>. ISSN 15420086.
- [39] Weaire D, Rivier NY. Soap cells and statistics-random patterns in two dimensions. *Contemp Phys* 1984;25(1):59–99. <https://doi.org/10.1080/00107518408210979>. ISSN 0010–7514.
- [40] Giraud R, Ta VT, Papadakis N. Robust shape regularity criteria for superpixel evaluation. *Proc. – Int. Conf. Image Process. ICIP 2017-Septe (2018)* 3455–3459, ISSN 15224880, <https://doi.org/10.1109/ICIP.2017.8296924>.
- [41] Kumar A, Sherrington D, Wilson M, Thorpe MF. Ring statistics of silica bilayers. *J Phys Condens Matter* 2014;26(39):1–6. <https://doi.org/10.1088/0953-8984/26/39/395401>. ISSN 1361648X.
- [42] Burd HJ. A structural constitutive model for the human lens capsule. *Biomech Model Mechanobiol* 2009;8(3):217–31. <https://doi.org/10.1007/s10237-008-0130-5>. ISSN 1617–7959.
- [43] Yurchenco PD, Ruben GC. Basement membrane structure in situ: evidence for lateral associations in the type IV collagen network. *J Cell Biol* 1987;105(6):2559–68. <https://doi.org/10.1083/jcb.105.6.2559>. ISSN 0021–9525.
- [44] Barnard K, Burgess SA, Carter DA, Woolley DM. Three-dimensional structure of type IV collagen in the mammalian lens capsule. *J Struct Biol* 1992;108(1):6–13. [https://doi.org/10.1016/1047-8477\(92\)90002-R](https://doi.org/10.1016/1047-8477(92)90002-R). ISSN 10958657.
- [45] Rezaei N, Lyons A, Forde NR. Environmentally controlled curvature of single collagen proteins. *Biophys J* 2018;115(8):1457–69. <https://doi.org/10.1016/j.bpj.2018.09.003>. ISSN 15420086.
- [46] Yurchenco PD, Ruben GC. Type IV collagen lateral associations in the EHS tumor matrix Comparison with amniotic and in vitro networks. *Am J Pathol* 1988;132(2):278–91. ISSN 0002-9440.
- [47] Fabris G, Lucantonio A, Hampe N, Noetzel E, Hoffmann B, DeSimone A, et al. Nanoscale topography and poroelastic properties of model tissue breast gland basement membranes. *Biophys J* 2018;115(9):1770–82. <https://doi.org/10.1016/j.bpj.2018.09.020>. ISSN 15420086.