


Cite this: *RSC Adv.*, 2020, 10, 38275

Received 16th July 2020
Accepted 12th October 2020

DOI: 10.1039/d0ra06205g

rsc.li/rsc-advances

Simplified computational model for generating biological networks†

Matthew H. J. Bailey,  ‡ David Ormrod Morley  ‡ and Mark Wilson  ‡*

A method to generate and simulate biological networks is discussed. An expanded Wooten–Winer–Weaire bond switching methods is proposed which allows for a distribution of node degrees in the network while conserving the mean average node degree. The networks are characterised in terms of their polygon structure and assortativities (a measure of local ordering). A wide range of experimental images are analysed and the underlying networks quantified in an analogous manner. Limitations in obtaining the network structure are discussed. A “network landscape” of the experimentally observed and simulated networks is constructed from the underlying metrics. The enhanced bond switching algorithm is able to generate networks spanning the full range of experimental observations.

Two dimensional random networks are observed in a range of contexts across considerably different length scales in nature: from nanometres, in the form of amorphous graphene; to metres, in the form of the Giant’s causeway; to tens of kilometres, in the form of geopolitical borders.^{1–3} A framework for describing these continuous random networks for chemical systems was first introduced by Zachariassen to describe silica-like glasses, and has proved to be extremely versatile in the years since,⁴ being used to describe biological networks, including actin networks and basement membranes made of collagen and laminin.^{5–8}

Despite the physical diversity of continuous random networks, they share some common properties which arise from the fact that all planar 2D networks can be represented as a tiled arrangement of polygons, joined at their edges.⁹ Viewing a network as a collection of polygons explains some restrictions on the nature of the network. For example, the angles at each node must add up to 360° and the mean number of edges connected to each node affects the mean number of edges [or ring size] of each polygon. Traditionally, semi-empirical laws provided a theoretical understanding of the complex behaviour of real networks, most commonly for inorganic networks such as MgO grains, silica, or epithelial cells.^{10–12} Some examples of these laws include: Lewis’ law,¹² governing the distribution of polygon areas; Lemaître’s law,¹³ relating the number of hexagons to the width of the polygon distribution; and the Aboav–Weaire law,¹⁴ governing the preference of polygons to be

surrounded by similar or dissimilar polygons. Recent work has helped clarify the latter by re-casting the problem in terms of the assortativity, a metric designed for network theorists to describe social networks.^{15,16}

Studying generic 2D networks in biology is a potentially difficult task. There are a variety of different techniques used, including fluorescence light-, atomic force and electron-microscopy.^{17–19} The resolution of light microscopy is limited by diffusion,¹⁸ and the power of electron or laser beams have to be limited to avoid damaging the delicate networks.¹⁷ However, even when a high-quality image is obtained, there is further difficulty in analysis. For example, each edge in a network is a complex molecule made up of tens of thousands of atoms which may not lie strictly in a single plane. In addition, the most interesting dynamic behaviour often occurs over very long timescales – potentially decades. To make this complexity scientifically digestible, different simplified models have been developed for biological networks, as well as semi-empirical laws. A simplified model is useful for exploring the complex behaviour of these biological networks, and to provide greater understanding of experimental data by eliminating the effects of finite sample size and providing precise control over the conditions. These models can then be used to understand macroscopic or mechanical properties of the networks, either directly or to parameterise multiscale approaches.^{6,20} Examples of existing simplified graph-based models include Erdős–Rényi random graphs in which edges are randomly placed connecting nodes, Mikado networks for many-layered systems in which edges are randomly placed across one another, and bond switching in which edges in an ordered graph are randomly exchanged.^{21,22}

In this work, an extension of a bond switching method, known as the Wooten–Winer–Weaire (WWW) algorithm, is presented to generate 2D networks which align with those

Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK. E-mail: mark.wilson@chem.ox.ac.uk

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0ra06205g

‡ These authors contributed equally to this work.



observed in biological systems.^{23,24} Bond switching methods offer good control over physical properties in the networks generated, and are sufficiently flexible to reproduce a wide range of networks. To rigorously compare networks generated by bond switching to real biological networks, key metrics are extracted from a range of experimental microscopy images of networks made of collagen IV or laminin.

Bond switching is a stochastic sampling (often described as a Monte Carlo) method, in which an initially regular network is systematically modified by a series of edge deletion and addition moves (corresponding to bond-breaking and bond-creation in chemical networks). In the WWW method,²⁴ the moves are chosen randomly to “switch” the coordination arrangements around two nodes. These switches preserve the coordination number k , of each node. For a more detailed discussion of the original method and figures, see Wooten *et al.*²⁴ However, in the extended bond switching method, the mean coordination number $\langle k \rangle$ is fixed at an initial value whilst the k of individual nodes may change. A schematic image of this is available as Fig. SI 1 in the ESI.† The changes in k are affected by the potential model and the moves chosen. The relaxation of the move selection criterion is especially useful for generating biological continuous random networks as it can encompass a range of different behaviours, for example cross-linking along the backbone of biopolymers or lateral interactions of head groups.

The steps in our algorithm are:

- Initialise using a network with known polygon structure, calculating the energy according to a potential model.
- Select one edge between two nodes and make a trial move by deleting it, and adding a new edge between the first node and a neighbour of the second.
- Locally optimise the geometry around the switching site according to the potential model and recalculate the energy (giving the energy difference ΔE).
- Accept or reject the trial move randomly with probability p , determined by the Metropolis criterion:²⁵

$$p = \min[1, \exp(-\Delta E/k_B T)] \quad (1)$$

- If the trial move was rejected, return to the previous accepted state. If it was accepted, replace the network with the trial network and repeat the process.

This algorithm is quick to compute, because the geometry only needs to be optimised in a small region around the switching site. The algorithm works best for potential models with explicit bonds and well-defined neighbours. To this end, a simplified Keating potential is applied which consists of harmonic bond (of length r) and angular (of angle θ) terms:²⁶

$$U(r, \theta) = \frac{1}{2}k_r(r - r_{\text{eqm}})^2 + \frac{1}{2}K_\theta(\cos(\theta) - \cos(\theta_{\text{eqm}}))^2, \quad (2)$$

with the equilibrium bond length r_{eqm} fixed at 1, and the equilibrium angle θ_{eqm} chosen to be $\frac{2\pi}{k}$ with k the coordination number of the central node. The force constants k_r and K_θ are both free parameters. When the coordination number of two

nodes changes in the simulation, the θ_{eqm} is recalculated for both of them. Fixing θ_{eqm} (*i.e.* not changing with k) led to the final networks exhibiting more artefacts and often showing significant demixing of nodes.

This simple, but physically motivated, potential model is useful to explore the range of networks that can be generated. Future developments will show how more detailed and biologically accurate potential models, such as the exponential-quadratic form used by Burd,²⁷ may further improve agreement with experimental observation. The simplicity of the potential model could be useful as a framework to be combined with more complex requirements. For example, a strong energetic penalty could be applied to configurations exhibiting undesirable artefacts. In this work this technique is used to reject all configurations in which any node has coordination number $k \geq 8$ or a polygon with side count $n \geq 20$. Ormrod Morley and Wilson extend this idea further; by using Monte Carlo to minimise a cost function instead of energy directly they demonstrate how key structural metrics (there the fraction of 6-membered rings in amorphous graphene) can be targeted.²³ It is, however, useful to develop an understanding of the structures formed by simplistic models prior to considering how to effectively constrain them to eliminate unphysical properties, rather than pre-judging what is required.

A honeycomb network of hexagons with $\langle k \rangle = 3$ is used as the starting lattice, to allow comparison with previous networks.^{28,29} The use of a well-defined initial polygon structure, and monitoring the changes at each step, allows the algorithm to ensure that the structure generated is physically acceptable. By constraining the mean node coordination, the mean number of edges per polygon is also constrained because of Euler's characteristic equation for polyhedra. For a 2D network which is either periodic or infinitely large, $\langle k \rangle = 3$ corresponds to $\langle n \rangle = 6$. This does not hold for finite-size aperiodic networks, such as those typically imaged experimentally. However, the approximation for aperiodic systems is small (for example, for 100 rings, $\langle n \rangle = 5.94$).¹⁶

The biological networks considered are visually extremely diverse. To demonstrate that diversity, three networks are shown in Fig. 1, which are accompanied by snapshots of the simulated networks, showing that these can qualitatively replicate the range of observed structures. The more regular networks in the figure are characteristic of those formed by inorganic materials.¹⁶ The figure also highlights associated structural metrics, the second moment of the ring distribution, $\mu_2(k)$, and the assortativity, r , both of which are discussed below. The difference in accompanying structural metrics in the experimental and simulated configurations highlights the need for a quantitative comparison instead of qualitative similarity. The structural metrics were extracted by a combination of image analysis tools and a mathematical approach based on graph theory as described below. A microscope image can be converted into an abstract graph by extracting the positions of edges and nodes, here using the Ridge Detection plugin for the image processing package ImageJ.^{30,31} The image analysis algorithms are discussed in detail in the ESI.† Obtaining a set of



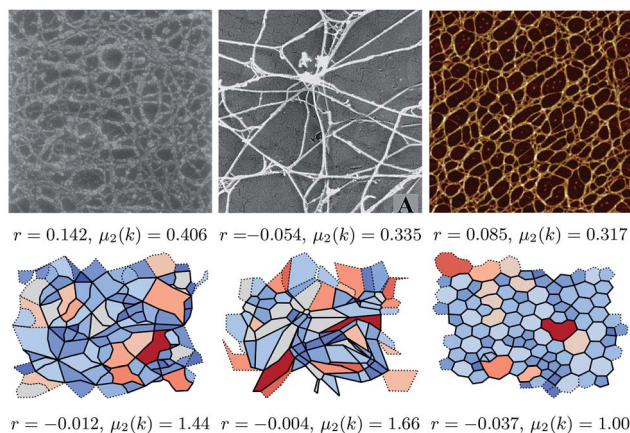


Fig. 1 Images of biological network from (left to right) Barnard *et al.*,³² Bos *et al.*,³³ and Wang *et al.*³⁴ respectively. The lower panels show networks generated by our method that appear similar, from left to right with $\{k_i, K_0\} = \{0.01, 0.01\}$, $\{10, 0.01\}$ and $\{100, 100\}$. The simulated networks were chosen on grounds of visual similarity; the difference in accompanying metrics (assortativity r and second moment of the k distribution $\mu_2(k)$) demonstrates why a more rigorous method of comparison is needed. The structural metrics are discussed in more detail in the body of the text. Reproduced from Barnard *et al.*³² and Bos *et al.*,³³ Copyright (1992) and (2001), with permission from Elsevier. Reproduced from Wang *et al.*,³⁴ Copyright (2017) with permission.

nodes and edges for each image allows the same analysis methods to be applied for the simulated networks.

Images of biological networks produced by Barnard *et al.*, Bos *et al.*, Wang *et al.*, and Yurchenco *et al.* taken between 1984 and 2017 were analysed.^{32–36} A subset of the results are shown in Table 1, and the complete data can be found in the ESI† The experimental networks show a mean node coordination $\langle k \rangle = 2.78$, with a standard deviation of 0.237. The average number of polygon sides $\langle n \rangle$ is 6.88, with a standard deviation of 1.42. The data seen in Table 1 show slightly lower $\langle k \rangle$ and larger $\langle n \rangle$ than for randomised hexagonal networks (where $\langle k \rangle = 3$ and $\langle n \rangle = 6$).

However, some caution is required in the interpretation of the results obtained with these methods. Fig. 2 demonstrates the difficulty in assigning edges to an image. The experimental image (left panel, from Yurchenco and Furthmayr³⁵) is analysed using two different manual interpretations of the network. As

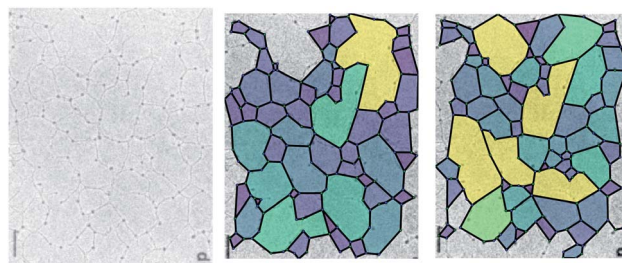


Fig. 2 The polygon structure of a collagen IV basement membrane as imaged by Yurchenco and Furthmayr.³⁵ The two images are different ways of highlighting the polygons in the same image, according to which ambiguous edges are included. Reprinted (adapted) with permission from Yurchenco and Furthmayr.³⁵ Copyright 1984 American Chemical Society.

the assignment of the edges can be ambiguous, there are many equally valid ways to select which edges are retained, and so different polygon structures arise. Here, for example, a small change in the algorithm identifies more nodes with a small associated increase in the mean node coordination, $\langle k \rangle$, and reduction in the mean polygon side count $\langle n \rangle$ (Table 1). Variations in $\langle k \rangle$ are emphasised by an ambiguity as to whether an edge in a biological network is best described as being one biomolecule or two, which makes describing the fraction of 2-coordinate nodes problematic. $\langle k \rangle$ is also affected by the finite size of an image, because where an edge is terminated by the image boundary, we remove it from the calculations, reducing k for edge nodes. Missing edges, either from low contrast or when terminated by an edge, leads polygons to be counted as too large or to not be counted at all respectively. These issues are more significant in smaller or lower-detail images, because the contrast leads to worse ring identification and a greater fraction of polygons are on the boundary.

The data extracted from the experimental images (Table SI 1 in the ESI†) indicate that the most useful metrics for describing these networks are the width of the node distribution (characterised by the second moment, $\mu_2(k)$), and the assortativity r , as described by Newman.¹⁵ For hexagonal inorganic networks, $\mu_2(k) = 0$, whereas all of the simulated and imaged networks have $0.3 \leq \mu_2(k) \leq 1.2$. The assortativity represents the similarity of neighbouring polygons, and is bounded as $-1 \leq r \leq 1$ with a negative value corresponding to polygons being neighbours to dissimilar polygons, and positive values indicating similar neighbours. An assortativity of $r = 0$ means that there is no preference for similar or dissimilar neighbours, and is seen in Erdős–Rényi random graphs. If the assortativity were ruled by geometrical constraints alone in a random point process, we would expect $r \approx -0.15$.¹⁶ Hard-disk simulations of colloids have shown positive assortativities of ~ 0.4 ,³⁸ networks of biological reactivity have small negative assortativities,¹⁵ whilst inorganic networks show negative values whose magnitudes depend on the details of the building blocks.¹⁶

Although the polygon side count distributions (characterised by the mean number of sides, $\langle n \rangle$ and its second moment, $\mu_2(n)$) have been extensively used to characterise more regular

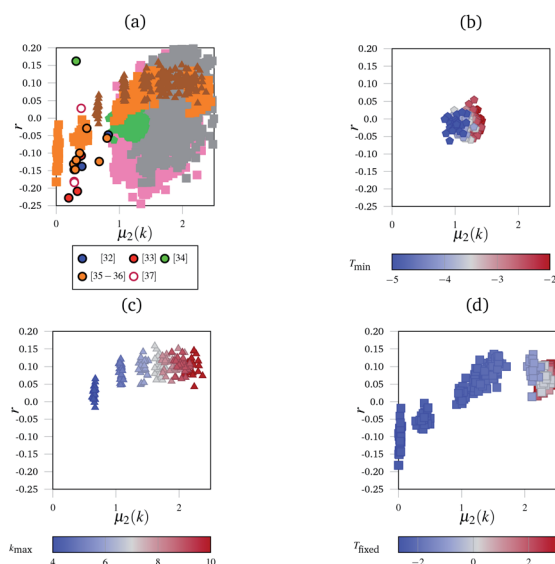
Table 1 Statistics collected for a number of images of biological networks (imaged under different conditions) showing the mean node coordination, $\langle k \rangle$, the second moment, $\mu_2(k)$, the mean number of polygon edges, $\langle n \rangle$, the number of polygons in each configuration, N , and the assortativity, r . Two different analyses of the networks imaged by Yurchenco *et al.*³⁵ are shown in order to highlight the sensitivity of the obtained metrics. A more detailed table is available in the ESI

Image	$\langle k \rangle$	$\mu_2(k)$	$\langle n \rangle$	r	N
Barnard <i>et al.</i> ³²	2.961	0.378	5.574	−0.109	135
Bos <i>et al.</i> ³³	2.780	0.335	5.825	−0.209	40
Wang <i>et al.</i> ³⁴	2.945	0.805	5.735	−0.057	498
Yurchenco and Furthmayr ³⁵	2.721	0.284	6.254	−0.181	61
Yurchenco and Furthmayr ³⁵	2.667	0.287	6.593	−0.184	54



Fig. 3(a) shows the combined metrics from Fig. 3(b)–(d) along with the full sets of results obtained by varying k_1 and K_0 .

Fig. 3 The “network landscape” showing both the simulated and experimental networks. In subfigure (a) different markers represent sets of simulated results, and outlined circles represent experimental data from the references indicated.^{32–37} Grey and pink squares represent the networks generated by the simulated annealing method with different bond and angle strength parameters. The subfigures highlight subsections of that landscape under different conditions with consistent markers to subfigure (a)—for example, triangles in subfigure (a) represent the same data as the triangles in subfigure (c). The subfigures are coloured according to which parameter was varied in the simulation. In (b), this is the minimum temperature $T_{\min}$ the simulation reached. In (c), this is the maximum coordination number of an individual node $k_{\max}$. In (d), it was the fixed temperature T_{fixed} the networks were heated at.



The experimental data, taken from ref. 32–37, lies at relatively low $\mu_2(k)$ and negative r . This low $\mu_2(k)$ and negative r region is occupied by relatively few simulated networks, with most simulated networks showing a larger $\mu_2(k)$ value. This is a combination of multiple factors that were explored in the panels of Fig. 3. Highly coordinated tangles such as those seen experimentally in Fig. 1(b) are relatively rare, and node coordinations above 4 are geometrically unstable with respect to splitting into two 3-coordinate nodes. This naturally limits the width of the node distribution histograms, and hence $\mu_2(k)$. Next, many biological networks are constructed in an ordered manner, or enthalpically driven in their assembly. For example, collagen IV in basement membranes has been hypothesised to have a square or a regular hexagonal structure.^{27,40} The images analysed here could be of regular networks with defects, either from aging or from preparation for imaging. Finally, the position of experimental data shows that real networks are not close to the random limit. This matches up with our intuition, as maximised entropy limits the amount of useful work that can be extracted from a system, which is an unhelpful property for biological networks. We anticipate that future network models which take the enthalpic factors into account more accurately (such as having terms in the potential that disfavour demixing) will fill in the “valley” observed in the landscape.

The simulated networks show interesting physical properties, including demixing, which are not observed in the experimental images. Similar polygons tended to cluster together and become more convex as the networks were cooled for longer periods of time. This demixing is driven enthalpically, as it appears stronger at low temperatures. The enthalpic drive arises from accommodating angular strain, for example, four-coordinate nodes are least strained when in a region of squares, and two-coordinate nodes are least strained when they are part of a long boundary between two large rings. After demixing, the networks show a positive assortativity because of this strain-driven preference for similar shapes to be adjacent. The potential slow evolution of structure has significant implications for the network aging process. This could be important, for example, in the collagen-IV network in the eye.

In conclusion, the bond switching method developed here can generate 2D polygonal networks similar to those found in biology. It can generate networks with a wide range of polygon side counts and node coordination distributions, mirroring the diversity exhibited by biological networks. A number of experimental biological networks were studied using methods previously used for inorganic chemical networks, and techniques from network theory were applied to understand their short range structure. The biologically-inspired networks showed the importance of letting the node distribution vary, and were almost all disassortative. However, the simulated networks show a range of assortativities approximately in the range $-0.2 \leq r \leq 0.2$. This is an interesting result, as the extra degree of freedom allowed in the simulations by varying k allows the polygon networks to get closer to the maximum entropy solution (with $r \approx 0$), which is not possible in traditional bond switching approaches.¹⁶ The enthalpic drive towards disassortative mixing in real networks is an interesting topic, and

could be studied in future by improvements to the simplified potential model presented here. Analysing these biological networks presented challenges because of the ambiguity in identifying edges, and it is anticipated that the simulated networks can be used to help resolve this. The simulated networks can bridge the gap between the hexagonal networks favoured by inorganic materials and more disordered biological networks, as well as showing interesting physical properties of their own, such as demixing. The additional degree of freedom allows for more entropically-dominated polygon networks, which allows the assortativity to approach zero. Future work will build on these methods utilising improved potentials, and exploring further the role of entropy, in order to understand the factors which underpin the observed network landscape.

Conflicts of interest

There are no conflicts to declare.

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. This paper conforms to the RCUK data management requirements. We thank Professors Elspeth Garman and Harvey Burd for valuable discussions.

Notes and references

- 1 D. Ormrod Morley and M. Wilson, *Mol. Phys.*, 2019, **117**, 3148–3157.
- 2 D. Weaire and N. Y. Rivier, *Contemp. Phys.*, 2009, **50**, 199–239.
- 3 G. Le Caër and R. Delannay, *J. Phys. A: Math. Gen.*, 1993, 1777–1800.
- 4 W. H. Zachariasen, *Phys. Rev.*, 1932, **40**, 917–922.
- 5 J. H. Hartwig and H. L. Yin, *Cell Motil. Cytoskeleton*, 1988, **10**, 117–125.
- 6 H. J. Burd, *Biomech. Model. Mechanobiol.*, 2009, **8**, 217–231.
- 7 K. J. Hamill, K. Kligys, S. B. Hopkinson and J. C. Jones, *J. Cell Sci.*, 2009, **122**, 4409–4417.
- 8 B. Kang, S. Jo, J. Baek, F. Nakamura, W. Hwang and H. Lee, *Acta Biomater.*, 2019, **90**, 217–224.
- 9 G. Schliecker and S. Klapp, *Europhys. Lett.*, 1999, **48**, 122–128.
- 10 D. A. Aboav and T. G. Langdon, *Metallography*, 1969, **2**, 171–178.
- 11 A. C. Wright and M. F. Thorpe, *Phys. Status Solidi B*, 2013, **250**, 931–936.
- 12 F. T. Lewis, *Anat. Rec.*, 1926, **33**, 331–355.
- 13 A. Gervois, J. P. Troadec and J. Lemaitre, *J. Phys. A: Math. Gen.*, 1992, **25**, 6169–6177.
- 14 D. A. Aboav, *Metallography*, 1980, **13**, 43–58.
- 15 M. E. J. Newman, *Phys. Rev. Lett.*, 2002, **89**, 208701.
- 16 D. Ormrod Morley, A. Thornework, R. Dullens and M. Wilson, *Phys. Rev. E*, 2020, 1–13.



- 17 R. A. Boerboom, K. N. Krahn, R. T. Megens, M. A. van Zandvoort, M. Merks and C. V. Bouten, *J. Struct. Biol.*, 2007, **159**, 392–399.
- 18 A. M. Barlow, L. B. Mostaço-Guidolin, E. T. Osei, S. Booth and T.-L. Hackett, *PLoS One*, 2020, **15**, e0229278.
- 19 S. P. Reese, N. Farhang, R. Poulson, G. Parkman and J. A. Weiss, *Biophys. J.*, 2016, **111**, 1797–1804.
- 20 E. A. Sander, T. Stylianopoulos, R. T. Tranquillo and V. H. Barocas, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 17675–17680.
- 21 C. P. Broedersz and F. C. Mackintosh, *Rev. Mod. Phys.*, 2014, **86**, 995–1036.
- 22 R. C. Picu, *Soft Matter*, 2011, **7**, 6768–6785.
- 23 D. Ormrod Morley and M. Wilson, *J. Phys.: Condens. Matter*, 2018, **30**, 50LT02.
- 24 F. O. Wooten, K. Winer and D. Weaire, *Phys. Rev. Lett.*, 1985, **54**, 1392–1395.
- 25 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.*, 1953, **21**, 1087–1092.
- 26 A. Kuronen, K. Kaski and S. von Althaus, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2003, 1–4.
- 27 H. J. Burd and R. A. Regueiro, *Biomech. Model. Mechanobiol.*, 2015, **14**, 1363–1378.
- 28 G. T. Barkema and N. Mousseau, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2000, **62**, 4985–4990.
- 29 D. K. Limbu, R. Atta-Fynn and P. Biswas, *MRS Adv.*, 2019, **4**, 87–93.
- 30 C. A. Schneider, W. S. Rasband and K. W. Eliceiri, *Nat. Methods*, 2012, **9**, 671–675.
- 31 T. Wagner, M. Hiner and X. Raynaud, *Ridge Detection*, 2018, <https://github.com/thorstenwagner/ij-ridgedetection>.
- 32 K. Barnard, S. A. Burgess, D. A. Carter and D. M. Woolley, *J. Struct. Biol.*, 1992, **108**, 6–13.
- 33 K. J. Bos, D. F. Holmes, R. S. Meadows, K. E. Kadler, D. McLeod and P. N. Bishop, *Micron*, 2001, **32**, 301–306.
- 34 Z. Wang, Q. Xiao, X. Song, Y. Wan and J. Zhu, *J. Food Qual.*, 2017, **2017**, 1–10.
- 35 P. D. Yurchenco and H. Furthmayr, *Biochemistry*, 1984, **23**, 1839–1850.
- 36 P. D. Yurchenco and G. C. Ruben, *J. Cell Biol.*, 1987, **105**, 2559–2568.
- 37 G. Fabris, A. Lucantonio, N. Hampe, E. Noetzel, B. Hoffmann, A. DeSimone and R. Merkel, *Biophys. J.*, 2018, **115**, 1770–1782.
- 38 A. Chremos and P. J. Camp, *Phys. Rev. E*, 2007, **76**, 056108.
- 39 A. J. Stone and D. J. Wales, *Chem. Phys. Lett.*, 1986, **128**, 501–503.
- 40 C. F. Cummings, V. Pedchenko, K. L. Brown, S. Colon, M. Rafi, C. Jones-Paris, E. Pokydesheva, M. Liu, J. C. Pastor-Pareja, C. Stothers, I. A. Ero-Tolliver, A. S. McCall, R. Vanacore, G. Bhavé, S. Santoro, T. S. Blackwell, R. Zent, A. Pozzi and B. G. Hudson, *J. Cell Biol.*, 2016, **213**, 479–494.

