Eur. Phys. J. E (2011) **34**: 5
DOI 10.1140/epje/i2011-11005-0

THE EUROPEAN PHYSICAL JOURNAL E

Regular Article

Computer simulations of the growth of synthetic peptide fibres

T.P. Stedall[1], M.F. Butler[2], D.N. Woolfson[3,4], and S. Hanna[1,a]

[1] H.H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, UK
[2] Unilever Research Colworth, Colworth Park, Sharnbrook, Bedfordshire, MK44 1LQ, UK
[3] The School of Chemistry, University of Bristol, Bristol, BS8 1TS, UK
[4] Department of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD, UK

Received 2 April 2010 and Received in final form 1 October 2010
Published online: 10 January 2011 –

Abstract. We present a coarse-grained computer model designed to study the growth of fibres in a synthetic self-assembling peptide system. The system consists of two 28 residue α-helical sequences, denoted AB and CD, in which the interactions between the half peptides, A, B, C and D, may be tuned individually to promote different types of growth behaviour. In the model, AB and CD are represented by double ended rods, with interaction sites distributed along their lengths. Monte Carlo simulations are performed to follow fibre growth. It is found that lateral and longitudinal growth of the fibre are governed by different mechanisms —the former is diffusion limited with a very small activation energy for the addition of units, whereas the latter occurs via a process of secondary nucleation at the fibre ends. As a result, longitudinal growth generally proceeds more slowly than lateral growth. Furthermore, it is shown that the aspect ratio of the growing fibre may be controlled by adjusting the temperature and the relative strengths of the interactions. The predictions of the model are discussed in the context of published data from real peptide systems.

1 Introduction

The field of peptide self-assembly attracts considerable interest. It has many potential applications including stem-cell differentiation [1], drug delivery [2] and tissue regeneration [3]. Furthermore protein aggregation, in particular of the peptide Aβ, is implicated in degenerative diseases such as Alzheimer's disease and is extensively studied experimentally [4] and computationally [5].

Computer modelling is of importance for the understanding of peptide self-assembly. For example, the micelle forming amphiphillic peptides of Stupp and co-workers [6] have also been extensively studied by coarse-grained modelling [7–9]. In peptides, two secondary structures commonly arise: α-helices and β-strands. The latter have been studied generically via coarse-grained modelling [10] and statistical mechanics [11]. Self-assembling systems involving α-helices are less common and are the focus of the present work.

The Self-Assembling peptide Fibres (SAFs) [12–15] form a synthetic system designed *de novo* by the Woolfson group. A number of SAF variants have been described. We consider a system consisting of two 28 residue peptides, denoted AB and CD (see fig. 1a). Each peptide may be regarded as consisting of two half-peptides with amino acid sequences shown in the figure. Each peptide sequence consists of four heptads with the repeating motif HPPHPPP, where H denotes a hydrophobic residue and P a polar one. The peptides fold into α-helices, which associate to form coiled coils [16].

The heptad positions in the coiled coil are standardly denoted *abcdefg*. The hydrophobic residues isoleucine and leucine situated in the a and d positions occur on the inside of the coiled coil, which is both enthalpically and entropically favourable. The coiled coil is further stabilised by salt bridging between the oppositely charged lysine and glutamic acid residues in the e and g positions. A single asparagine is substituted for an isoleucine in sequences B and C. The interaction of asparagine with isoleucine is highly unfavourable and its interaction with itself is preferential. This imparts a specificity to the assembly; that is, B binds preferentially to C and A binds to D. The result is the formation of offset dimers that promote subsequent assembly into fibrils (see fig. 1b). The lateral assembly of fibrils is promoted by charged arginine and aspartic acid residues in positions b, c and f. The SAF peptides system self-assembles into fibres of several microns in length and 50–100 nm in width. An example of such fibres is given in fig. 2.

The pathway of SAF fibre formation has recently been elucidated experimentally [17]. It is found that a short, and possibly wide nucleus forms from hetero-dimers which

[a] e-mail: s.hanna@bristol.ac.uk

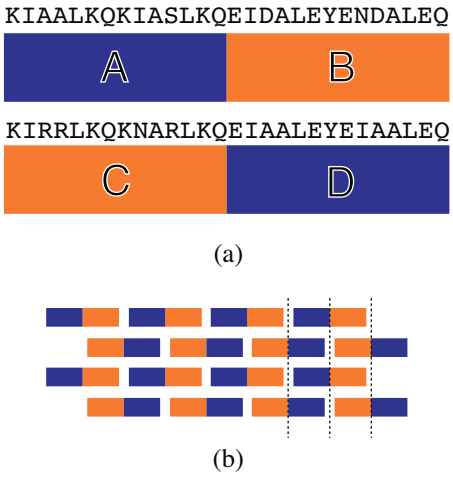

Fig. 1. (Colour on-line) (a) The peptide species AB and CD, each composed of two half-peptides. The amino acid sequences are also shown. (b) Fibre formation via the offset assembly of AB and CD units. The A and D half-peptides are shaded dark grey (blue), while the B and C half-peptides are shaded light grey (orange). The dotted lines indicate the half-units used to measure the lengths of the growing fibres.

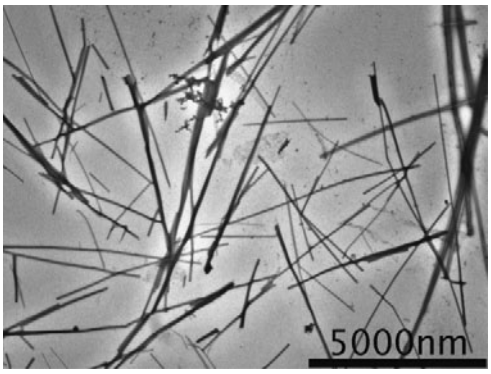

Fig. 2. Transmission electron microscopy image of SAF fibres. Image courtesy of Patrick King (University of Bristol).

are created via the A-D interaction. The nucleus initially grows both laterally and longitudinally, but at later times ceases to grow laterally while continuing to extend longitudinally to form a fibre. It is the aim of the present work to contribute further to this understanding using coarse-grained computer simulation, both via the simple model described here and a more detailed model, which will be the subject of a future publication.

2 Methods

As outlined above, SAF assembly is complex and involves processes at several length and time scales. We address the problem of modelling the system by constructing the simplest possible coarse-grained representation of the SAF peptides, which nonetheless retain their essential features.

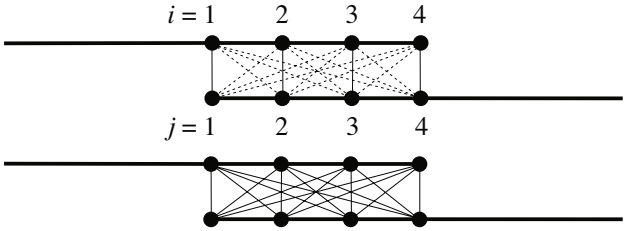

Fig. 3. Top: a specific bond, solid line interactions have parameter k_{PQ}, dotted line interactions have parameter k'_{PQ}. Bottom: an unspecific bond, all interactions have parameter $k^{\mathrm{u}}_{\mathrm{PQ}}$. In these examples, $n = 4$ (see text).

2.1 The n-site model

We model each peptide sequence as a rigid rod of unit length. Each half peptide is represented by a linear array of n interaction sites. Hence, the peptide species AB and CD possess $2n$ interaction sites distributed evenly along their lengths. The half peptides will then interact through pairwise interactions; there will be n^2 of these between two half peptides, of which $n(n+1)/2$ are non-redundant and could potentially be of different strengths (see fig. 3). Denoting the n interaction sites of a half-unit P and Q as P_i and Q_j respectively, where P and Q belong to {A,B,C,D} and $1 \leq i, j \leq n$, we define two types of interaction, illustrated in fig. 3, which we describe as "specific" and "unspecific".

In a specific interaction, we employ just two relative strength parameters: k_{PQ} when $i = j$, and k'_{PQ} otherwise. We refer to these as adjacent and non-adjacent components of the interaction, respectively. Setting $k_{\mathrm{PQ}} \gg k'_{\mathrm{PQ}}$ makes directional alignment of the half-units, P and Q, more likely. In an unspecific interaction, a single relative strength parameter, $k^{\mathrm{u}}_{\mathrm{PQ}}$, is used for all i, j. It can be seen that an unspecific interaction places far less constraint on the relative orientations of P and Q than the specific case.

The parameter n must be chosen carefully. Simulation times increase as n^2, so n should be kept as small as possible. However, for very small n, fibre growth is not observed. Preliminary simulations indicated that $n = 4$ is the smallest value for which fibre growth is obtained within reasonable waiting times, and therefore this value was used throughout the paper. An example of fibre growth with $n = 4$ is shown in fig. 4.

Each pair of interaction sites interacts via a 6-3 potential:

$$\begin{aligned} V(r) &= \phi(r) - \phi(r_c), \\ \phi(r) &= 4k\left[(\sigma/r)^6 - (\sigma/r)^3\right], \end{aligned} \tag{1}$$

where k is the appropriate half unit interaction strength parameter *i.e.* k_{PQ}, k'_{PQ} or $k^{\mathrm{u}}_{\mathrm{PQ}}$, and σ sets the position of the potential well. k and V are expressed in units of k_BT, with k_B the Boltzmann constant, set to unity for convenience. σ is expressed as a fraction of the length of a rod, $\sigma = 0.08$, giving a potential minimum at 0.1 length units. r_c is the radial cutoff used in energy evaluation and is typically two simulation length units. Potentials are cut and shifted at r_c so that they are continuous.

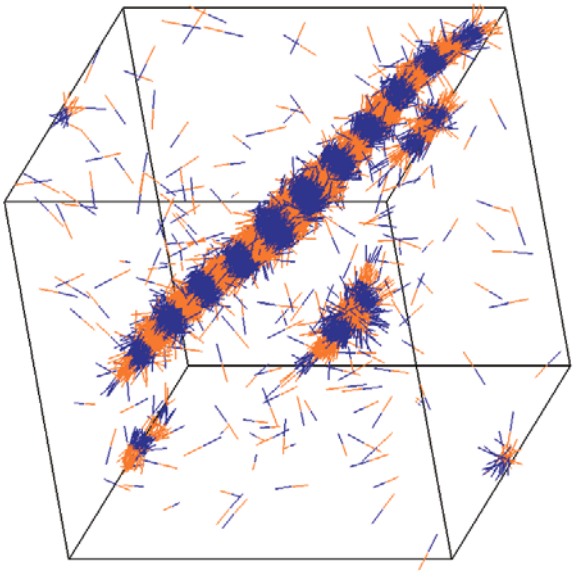

Fig. 4. (Colour on-line) Example of fibre formation by homogeneous nucleation from dilute solution in a constant-μVT simulation ($T = 10$; see text).

The 6-3 potential is softer than the Lennard-Jones 12-6 potential, being designed to mimic generic molecular rather than atomistic interactions. As with the conventional Lennard-Jones form, the short-range repulsive term prevents overlap of the peptides, while the long-range attractive term permits assembly of the units; the use of a 6-3 potential is advantageous because it has a longer range attraction than the 12-6 potential.

2.2 Monte Carlo simulations

Monte Carlo simulations are performed at constant temperature, T, in a simulation box with fixed dimensions (constant volume, V) with periodic boundary conditions. The simulations are divided into two groups. In the first group, the total number of peptides, N, is kept fixed during the simulation (constant-NVT) while, in the second group, the concentration of peptides *in the solution phase*, as defined by the chemical potential, μ, is kept fixed (constant-μVT). In both cases, trial moves consist of stochastic translations and rotations of the peptide units (rods). In the μVT simulations unit insertions and removals are also performed, in order to maintain a constant concentration during fibre growth. Insertions and removals are performed in the solution (free unit) phase only.

In discussing the different types of simulation we avoid the use of the word "ensemble" as this implies thermodynamic equilibrium, whereas fibre growth is a non-equilibrium process. In fact, when fibre growth is sufficiently slow, we regard the solution phase as being approximately at equilibrium. A similar Monte Carlo approach to crystal or fibre growth has been used by several previous authors (see, *e.g.*, refs. [18–21] for studies of peptide aggregation and ref. [22] for a more classical crystallisation study).

For translations and rotations the standard Metropolis criterion [23] is used. The size of each move is selected from a uniformly distributed random deviate with maximum value set by appropriate scaling factors for translations and rotations. The scaling factors are chosen to be sufficiently small that the acceptance rate does not fall below 50% even after all units have joined the fibre phase.

The acceptance rate discussed here is the total acceptance rate for the whole simulation *i.e.* for fibre and solution phase together. A consequence of this is that moves of peptides within the growing fibres are less likely to be accepted than those of peptides at the surface of the fibre or in the solution phase. As a result, the insides of each fibre will tend to evolve more slowly than the other parts of the simulation. However, since fibre growth will be determined by the transport of material to, and deposition on, the surface of the fibre, this is unlikely to have a major influence on the growth kinetics.

In the μVT simulations, the acceptance probabilities for particle insertions and removals are given by Yao *et al.* [24] (see appendix A). In the simulations, insertions and removals constitute 0.1% of all moves, a frequency chosen to maintain the required concentration without introducing instability. Units are only exchanged with the free unit (solution) phase, as determined by a special phase distinction algorithm. This algorithm maintains lists of the units included in each fibre, generated on the basis of the proximities of the ends and mid-points of each peptide to its nearest neighbours. By this means, the concentration of free units, as set by the chemical potential, remains approximately constant.

The simulations are initiated by placing peptide units at random in the simulation box. Bad contacts are removed by equilibrating at high temperature. In some simulations, fibre growth is seeded by placing a model fibre nucleus in the centre of the simulation box. These nuclei mimic fibres in that they consist of units offset in the configuration illustrated in fig. 1b. They are however more ordered than the fibres that form during simulations.

In μVT simulations, the initial number of free units, N_{F}, is chosen to be close to the equilibrium value; unit insertions and removals are not performed during the initial equilibration.

2.3 Analysis

During the simulations, an algorithm is used periodically to monitor and analyse fibre formation. In particular, fibre size, length and width are determined in terms of the number and type of constituent units.

The algorithm operates by considering the proximities of the centres and ends of the peptide units. In fibres, longitudinally adjacent units from different species will have close ends and centres of mass. Therefore, for each unit the proximity of its centre of mass to the ends of all units of the other species is determined. If this distance is smaller than twice the potential well width, the target unit is appended to a list of the selected unit's neighbours. Each of these lists is expanded recursively, until its length

Table 1. A summary of the NVT simulations performed exclusively with unspecific interactions. $k^{\mathrm{u}}_{\mathrm{EE}}$ is the interaction strength parameter for all unspecific interactions not otherwise given.

#	$k^{\mathrm{u}}_{\mathrm{AB}}$	$k^{\mathrm{u}}_{\mathrm{CD}}$	$k^{\mathrm{u}}_{\mathrm{AC}}$	$k^{\mathrm{u}}_{\mathrm{BD}}$	$k^{\mathrm{u}}_{\mathrm{AD}}$	$k^{\mathrm{u}}_{\mathrm{BC}}$	$k^{\mathrm{u}}_{\mathrm{EE}}$	T
1	–	–	–	–	–	–	1.0	2, 5
2	10.0	10.0	–	–	–	–	1.0	5, 10
3	–	–	10.0	10.0	–	–	1.0	5, 10
4	–	–	–	–	10.0	10.0	1.0	5, 10

remains constant. The resulting lists each represents either a free unit or a fibre. Free units and very small fibres (containing fewer than five units) are discarded and the remaining set of lists gives the distribution of fibre sizes in terms of the number of constituent units.

Additionally, a number of thermodynamic parameters are calculated periodically during the simulation. The configurational temperature, pressure, heat capacity and chemical potential are calculated for the free unit phase only; the accuracy of these calculations is dependent on the reliability of the phase distinction algorithm.

The configurational temperature is calculated from the gradients of the potential energy terms in the system [25]. This value may be compared to the input temperature to ensure the simulation is sampling the appropriate ensemble, *i.e.* it gives a useful indication of how closely the solution phase in the simulation approximates to equilibrium. As such, it is a valuable diagnostic tool [26]. We find that the configurational temperature tracks the input temperature to within 10% when fibre formation occurs, and within 1% when it does not. The chemical potential in the free unit phase is calculated using Widom's particle insertion method [27] in the μVT simulations. It is found that the calculated chemical potential accurately tracks the required value, to within 1%. Any discrepancies that occur in the configurational temperature and chemical potential values result from occasional errors in the phase distinction algorithm, due to disorder in partially formed fibres.

2.4 Simulations performed

The simulations are performed in a cubic box of dimension 20 simulation length units. Periodic boundary conditions are used with the minimum image convention. The NVT simulations are listed in tables 1 and 2. The equilibration period consists of 1×10^6 iterations at $T = 10$. 1000 units of each species are present (2000 units in total) and the production runs consist of 5×10^7 iterations, which take approximately 50 hours on a 3.2 GHz xeon processor. In the tables, the parameter $k^{\mathrm{u}}_{\mathrm{EE}}$ refers to all unspecific interaction strength parameters that are not otherwise specified.

The final column of table 2 lists a reduced temperature for each simulation. When a set of simulations is performed with some or all of the interaction strengths varied, this is equivalent to running the simulations at different temperatures, because the trial move acceptance probabilities depend only on the ratio of energy and temperature.

Table 2. A summary of the NVT simulations performed containing specific interactions. $k^{\mathrm{u}}_{\mathrm{EE}}$ is the interaction strength parameter for all unspecific interactions.

#	k_{AD}	k'_{AD}	k_{BC}	k'_{BC}	$k^{\mathrm{u}}_{\mathrm{EE}}$	T	T^*
5	10.0	1.0	10.0	1.0	1.0	3.0	0.12
						5.0	0.19
						7.0	0.27
6	5.0	1.0	15.0	1.0	1.0	6.5	0.25
	7.0		13.0			6.5	0.25
	9.0		11.0			6.5	0.25
	11.0		9.0			6.5	0.25
	13.0		7.0			6.5	0.25
	15.0		5.0			6.5	0.25
7	10.0	0.1	10.0	0.1	0.1	5.15	0.25
		0.5		0.5		5.35	0.25
		1.0		1.0		5.6	0.25
		5.0		5.0		7.6	0.25
8	10.0	0.1	10.0	0.1	1.0	6.05	0.25
		0.5		0.5		6.25	0.25
		1.0		1.0		6.5	0.25
		5.0		5.0		8.5	0.25
9	10.0	0.1	10.0	0.1	0.1	5.15	0.25
					0.5	5.55	0.25
					1.0	6.05	0.25
					5.0	10.05	0.25
10	10.0	1.0	10.0	1.0	0.1	5.6	0.25
					0.5	6.0	0.25
					1.0	6.5	0.25
					5.0	10.5	0.25

In fact, we wish to be able to run the simulations with different parameters but at the same *effective* temperature, so that they may be compared. We achieve this by defining a reduced temperature, T^*, as the ratio of the absolute temperature, T, to the energy of a growing fibre, E

$$T^* = T/E, \tag{2}$$

where

$$E = k_{\mathrm{AD}} + k'_{\mathrm{AD}} + k_{\mathrm{BC}} + k'_{\mathrm{BC}} + k^{\mathrm{u}}_{\mathrm{AA}} + k^{\mathrm{u}}_{\mathrm{BB}} + k^{\mathrm{u}}_{\mathrm{CC}} + k^{\mathrm{u}}_{\mathrm{DD}}. \tag{3}$$

The expression for E is obtained by considering the close contacts occurring in a fibre growing via the offset assembly model (fig. 1b). A detailed justification for this definition is given in appendix B. Since offset assembly does not occur in the absence of specific interactions (simulations 1-4, table 1), eq. (3) does not apply and T^* is not

Table 3. A summary of the μVT simulations performed. In all simulations, the specific interactions are given by $k_{\mathrm{AD}} = 10$, $k'_{\mathrm{AD}} = 0.1$, $k_{\mathrm{BC}} = 10$ and $k'_{\mathrm{BC}} = 0.1$, and the unspecific interactions by $k^{\mathrm{u}}_{\mathrm{EE}} = 0.1$. In each case, $T^* = T/20.6$, the nucleus cross-section is 25 units and the nucleus length is 6 fibre segments (see text). Values of μ are determined from eq. (A.4).

#	T	N_{F}	μ
A	10.0, 12.5, 15.0, 17.5, 20.0, 22.5, 25.0, 27.5, 30.0, 32.5 & 35.0	750	−82.8 to −23.7
B	10.0, 20.0	250, 500, 750, 1000, 1250	−69.3 to −37.1
C	40.0, 42.0, 44.0, 46.0, 48.0, 50.0, 52.0 & 54.0	1000	−112.3 to −83.2

quoted in those cases. In scaling the temperature by the total energy, the average move energy is also scaled by the temperature. This was the main aim in defining a reduced temperature. If the reduced temperature had not been used, the results for simulations 6 to 10 would have been dominated by the tendency of the model to nucleate, which is a function of the model parameters at any given absolute temperature. As such, the choice of T^* was not based on equilibrium properties.

However, it would be useful, if possible, to relate T^* to the equilibrium melting temperature of the system, T^0_{m}. In practice this is difficult, for two reasons. Firstly, because the fibres are small, surface free-energy contributions are significant and the melting temperatures, T_{m}, are depressed. Finding T^0_{m} involves determining T_{m} for fibres of varying sizes, and then extrapolating to the infinite case. Individual melting point determinations are made more difficult by the fact that T_{m} changes as the fibres grows. The second difficulty is related to the anisotropy of the fibres. The surface free energy for the ends of the growing fibres is quite different from that of the sides resulting in different melting temperatures for the two faces. The consequences of this for fibre morphology will be discussed further in sect. 3.6.

Notwithstanding the above considerations, we have determined T^0_{m} for one system, *i.e.* simulation 9 in table 2 with $k^{\mathrm{u}}_{\mathrm{EE}} = 0.1$. In this case we find that the equilibrium melting temperature occurs when $T^* \simeq 2$, compared to a simulation temperature of $T^* = 0.25$.

Table 3 details the μVT simulations performed. The equilibration period consists of 5×10^6 iterations at $T = 10$ (for simulations C the equilibration period consists of 2.5×10^7 iterations to allow a larger fibre to grow). The production runs use $k_{\mathrm{AD}} = 10$, $k'_{\mathrm{AD}} = 0.1$, $k_{\mathrm{BC}} = 10$, $k'_{\mathrm{BC}} = 0.1$ and $k^{\mathrm{u}}_{\mathrm{EE}} = 0.1$ and again consist of 5×10^7 iterations. These simulations are nucleated; the nuclei contain 25 units in cross-section, and are 6 fibre segments in length. A fibre segment is equivalent to half the length of a unit. It is used to specify length because of the overlapping nature of the offset assembly. Fibre segments appear as bands of colour in simulation snapshots, *e.g.* the large fibre in fig. 4 is 24 fibre segments in length.

The number of units in the simulation is initially equal to the total specified equilibrium free unit number, N_{F}, plus the total number of units in the nucleus. After the equilibration period, the free units are replenished from a supply so that their number remains approximately constant. For computational reasons, this supply is finite and set at 4000 units including those initially present. Once the supply is exhausted, the simulation enters the terminal regime. The chemical potential is calculated from eq. (A.4) (see appendix A). Since eq. (A.4) is approximate, the equilibrium free unit concentration differs slightly from the specified number, as noted in some of the graphs below.

Where error bars are shown in the figures, these are determined by repeating simulations 3 or more times.

3 Results

In this section we describe the general behaviour of the n-site model considering, first, the case where all of the interactions between the AB and CD peptides are unspecific (table 1). This is followed by a detailed analysis of fibre formation in the presence of specific interactions (table 2). Here, and in subsequent sections, we use the labelling convention indicated in fig. 1a, *i.e.* A and D units are depicted in dark grey (blue) while B and C units are depicted in light grey (orange).

3.1 Self-assembly without specificity in the interactions

Figure 5 illustrates the case where all interactions are unspecific and have equal strength (simulation 1 in table 1). At $T = 2$ (fig. 5a), bundles of fully adjacent units form, *i.e.* there is no indication of the offset assembly illustrated in fig. 1b. At higher temperatures ($T = 5$, fig. 5b), nucleation has occurred at one point and a globular structure can be seen beginning to grow. There is no evidence of fibre formation in either case.

In simulations 2-4 in table 1, certain pairs of interactions are favoured, *i.e.* interactions {A-B,C-D}, {A-C,B-D} and {A-D,B-C}, respectively. All pairs produce bundles of units similar in appearance to those in fig. 5a and for this reason they are not illustrated. Most of the units align in an adjacent configuration, with only a small proportion offset as illustrated in fig. 1b. Clearly, once a bundle has formed, units may stick to it via any interaction but the favoured interactions will dominate the process. Specific interactions are based on parallel alignment so, under such cases, the pair of interactions {A-D,B-C} will tend to produce offset assembly while the pair {A-C,B-D} will produce non-offset clusters. The more interesting former case is explored in the following sections, with the addition of specific interactions.

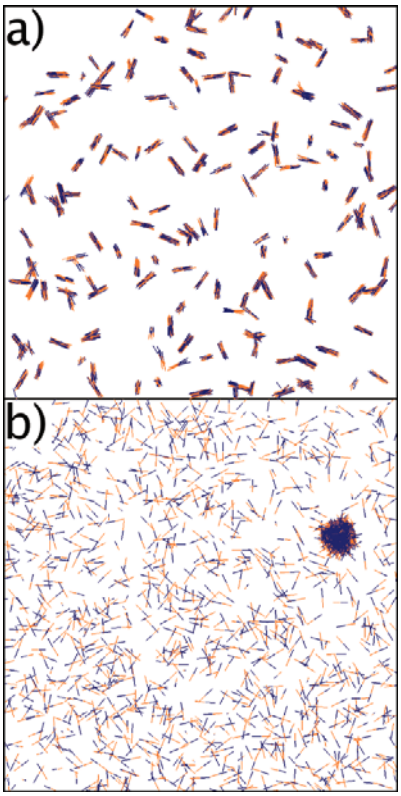

Fig. 5. (Colour on-line) Snapshots of systems in which no specificity is used and all interactions are equally favoured (simulation 1): (a) $T = 2$ and (b) $T = 5$.

3.2 Fibre formation in the offset assembly case

Here the case where the pair of interactions A-D and B-C are both specific and favoured is studied. We consider the case when the adjacent parts of these interactions, k_{AD} and k_{BC}, are ten times stronger than the non-adjacent parts and the unspecific interactions (simulation 5, table 2). The simulations in this section are performed in the NVT ensemble. Figure 6 shows homogeneously nucleated fibre formation for $T = 3$, 5 and 7. The waiting time for homogeneous nucleation increases rapidly with temperature and nucleation is not observed for $T > 7$ within the length of simulation used. As expected, at higher temperatures fewer, larger fibres form. The results from fig. 6 are analysed further in figs. 7 and 8. Figure 7a shows the normalised distribution of fibre sizes in terms of constituent units, fig. 7b the normalised distribution of fibre lengths in terms of segments and fig. 7c the normalised distribution of fibre cross-sections in terms of the average number of units per fibre segment for the three values of temperature. The transition from many small fibres at lower temperatures to a few large ones at high temperatures is clearly seen. The experimental length distribution is found to be approximately exponential [28], in common with other protein fibre forming systems. However, in the simulations reported here, the numbers of fibres is quite small, so that it is not possible to draw any firm conclusions about the shapes of their length distributions.

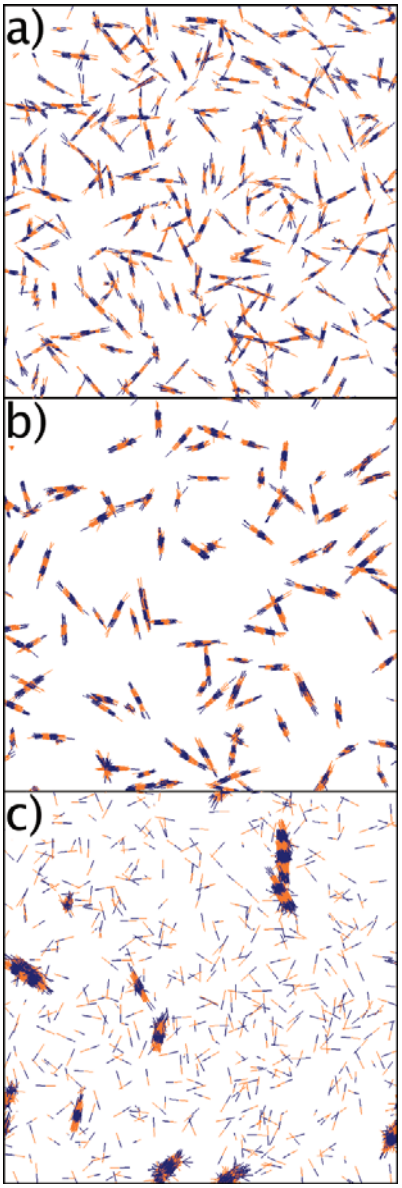

Fig. 6. (Colour on-line) Snapshots showing fibre formation at (a) $T = 3$, (b) $T = 5$ and (c) $T = 7$, for simulation 5 in table 2.

Figure 8a shows the radial distribution functions, based on centre of mass separation, for the simulations shown in fig. 6. For $T = 3$ and $T = 5$ the peaks correspond to short-range order only *i.e.* the nearest-neighbour unit spacings within individual fibres. Because the fibres are small the simulation box is relatively homogeneous and there is no long-range order. For $T = 7$ there are relatively few fibres which are much larger than at lower temperatures. As a result, the short-range peaks are more pronounced and there are also long-range peaks corresponding to inhomogeneity in the simulation box.

The radial distribution function at $T = 5$ is analysed further in fig. 8b and the marked peaks are explained as follows. The first set of peaks is due to laterally adjacent units, with the first peak corresponding approximately

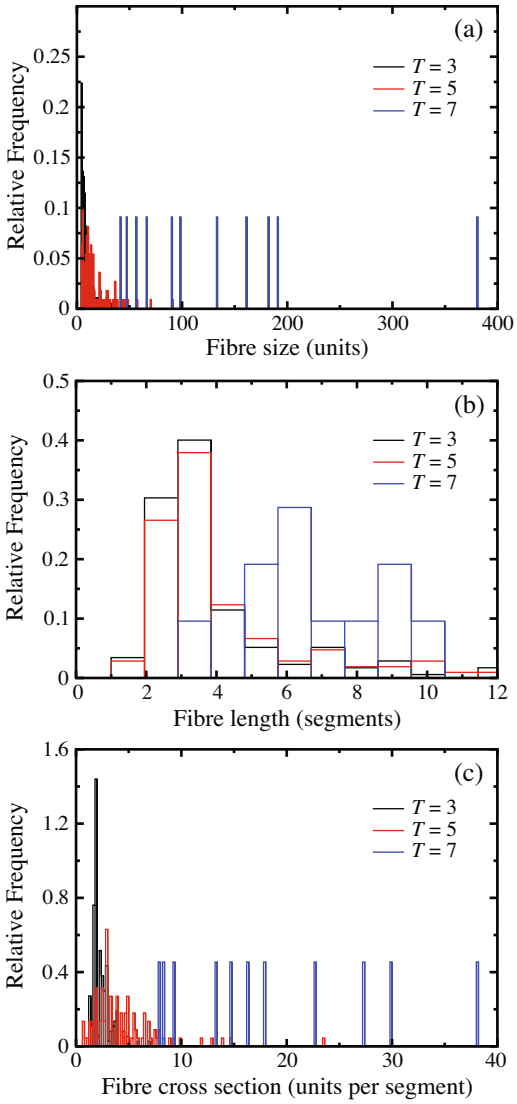

Fig. 7. (Colour on-line) Frequency analysis of fibres at $T = 3$, 5 and 7 (table 2, simulation 5) in terms of (a) number of units in fibres, (b) number of segments in fibres and (c) number of units in a fibre segment (cross-section).

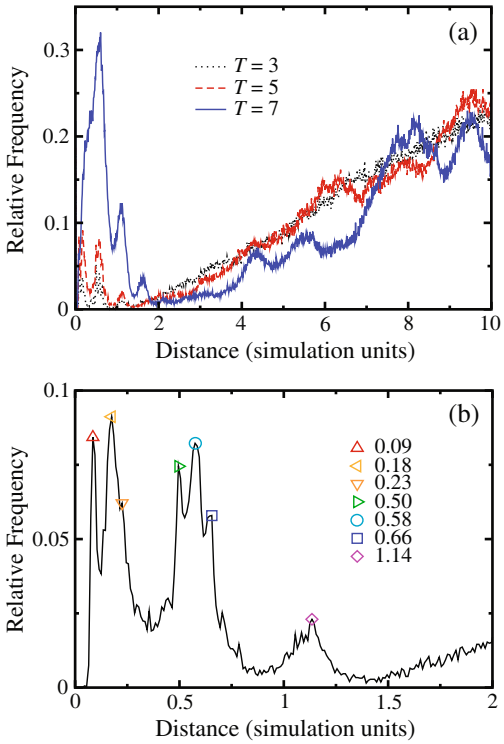

Fig. 8. (Colour on-line) (a) Normalised radial distribution functions at $T = 3$, $T = 5$ and $T = 7$; (b) analysis of the $T = 5$ radial distribution function, for simulation 5, table 2.

to the potential width of 0.1 and the second two peaks corresponding to multiples of this peak by $\sqrt{3}$ and 2. This strongly suggests hexagonal packing, as is indeed observed in X-ray studies of the experimental system [15]. The second set of peaks corresponds to units adjacent longitudinally but overlapping by approximately half a unit, *i.e.* offset assembly. The peak splitting that is apparent arises because of a modification to the packing of fig. 3 in which each of the interaction sites on one unit sit in two potential minima on the adjacent unit. This is discussed further in appendix B. As a result, the centres of mass may be separated by $n \pm 1/2$ site separations, longitudinally, and the first and third peaks of the second set correspond to these distances for $n = 4$. The third peak in the figure corresponds to two non-overlapping, longitudinally adjacent units.

Finally, an Avrami analysis [23] is performed on the simulations with $T = 5$ and $T = 7$. The Avrami equation for crystal growth is

$$1 - v_c = \exp(-Kt^m), \tag{4}$$

where v_c is the volume fraction of the crystalline material, t is time and K and m are constants typical of the nucleation and growth mechanism. Here we interpret v_c as the number fraction of units included in fibres. The Avrami exponent m is particularly significant, and includes contributions from both nucleation and growth processes [29]

$$m = qd + r, \tag{5}$$

where d is the dimensionality of growth, *i.e.* 1, 2 or 3-dimensional, and r describes the "order" of the nucleation: $r = 1$ if nuclei form at a constant rate, while $r = 0$ if all crystals are nucleated simultaneously at the beginning of the growth process. q describes the time dependence of the growth: $q = 1$ for linear crystal growth and $q = 0.5$ for diffusion limited growth.

The number fraction of units included in fibres is fitted to eq. (4) by mapping the Monte Carlo iterations to time. The raw data and lines of best fit are shown in fig. 9. For $T = 5$ the Avrami exponent, m, is found to be 1.15 and for $T = 7$ it is 1.75. The discrepancy in the fit at early times for $T = 7$ is attributed to poor statistics due to the small number of fibres growing in this case (there are 112 fibres in the $T = 5$ simulation compared to only 11 at $T = 7$).

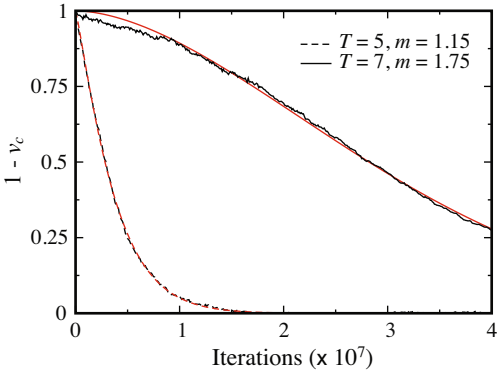

Fig. 9. (Colour on-line) Plots of the number fraction of uncrystallised units, $1 - v_c$, *versus* the number of Monte Carlo iterations, for simulation 5, table 1, $T = 5$ and 7, including fits to the Avrami equation (eq. (4), smooth curves).

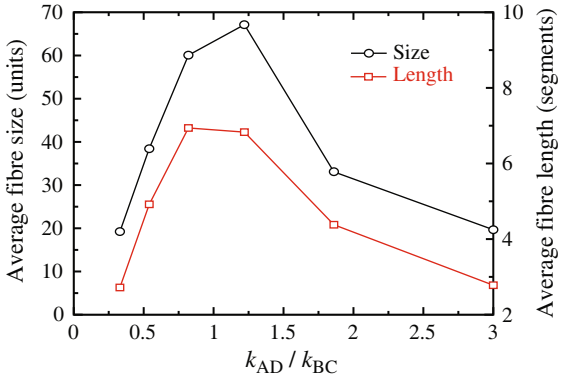

Fig. 10. (Colour on-line) Average fibre size and length *versus* k_{AD}/k_{BC}, for systems with $k'_{AD} = k'_{BC} = k^u_{EE} = 1.0$ (simulation 6, table 2).

At first sight, the Avrami exponent for $T = 5$ would appear to indicate 1-dimensional, *i.e.* fibre-like, growth. However, a careful inspection of the simulations indicates that this is not the case. In fact, growth occurs principally by the lateral addition of units, *i.e.* the growing fibres broaden, but little lengthening is observed. Therefore, we conclude that growth is actually 2-dimensional, but diffusion limited *i.e.* $d = 2$ and $q = 0.5$ in eq. (5). That growth should be diffusion limited is reasonable, given the low concentration of units in the simulation, and the relatively high undercooling. Further consideration will be given to this point, and to the shapes of the growing fibres, later in sect. 3.4.

It is a little surprising that m increases with temperature. Generally, we would expect m to be independent of temperature, unless there is a change in nucleation mechanism [30]. This does, in fact, appear to be the case. On quenching to $T = 5$ at the beginning of fibre growth, we observe the spontaneous and simultaneous formation of a large number of nuclei, from which we deduce that $r \simeq 0$ in eq. (5). However, at $T = 7$ far fewer nuclei are formed, and these appear sporadically throughout the simulation. If the number of nuclei were to increase linearly with time, we would expect $r = 1$. However, it appears that the dependence is less than linear, which we explain by noting that there will be a concentration dependence in the nucleation rate which will disfavour the formation of nuclei during the later stages of the simulation.

3.3 Influence of model parameters on fibre formation

Here we investigate the effect of varying model parameters on the growth of fibres. All simulations are performed at $T^* = 0.25$, which is equivalent to $T = 6.5$ for the model in the previous section.

Figure 10 shows the effect of varying the ratio of the adjacent parts of the specific interactions, k_{AD}/k_{BC}, on the resulting fibre dimensions (simulation 6, table 2). It can be seen that the largest fibres form when $k_{AD}/k_{BC} \simeq 1$. When either interaction is favoured, it appears that growth is inhibited at both ends of the fibre. For example, when $k_{AD}/k_{BC} < 1$, growth is inhibited at fibre ends terminated with A and D units whereas, when $k_{AD}/k_{BC} > 1$, it is inhibited at the ends terminated with B and C units. The result is a proliferation of very small fibres in which the more favourable interactions dominate.

Simulations 7–10 (table 2) explore the effect of varying the non-adjacent parts of the specific interactions, k'_{AD} and k'_{BC}, and the unspecific interactions, k^u_{EE}, while keeping $k_{AD} = k_{BC} = 10.0$. The results are summarised in figs. 11 and 12. It should be noted that $k'_{AD} = k'_{BC}$ throughout.

Figure 11 shows the average fibre size, length and cross-section when k'_{AD} and k'_{BC} are varied, for $k^u_{EE} = 0.1$ (fig. 11a) and $k^u_{EE} = 1.0$ (fig. 11b), *i.e.* simulations 7 and 8, respectively. Also shown for each simulation is the number of iterations, τ, taken for the number of free units to fall to $1/e$ of its initial value (fig. 11c). τ may be regarded as a crude measure of the rate of fibre growth.

Two general trends are apparent. Firstly, all dimensions of the fibres appear to decrease with increasing k'_{AD} and k'_{BC}. This is in line with expectations, as it is the adjacent parts of the specific interactions that play the major role in promoting offset assembly; increasing the non-adjacent part will only serve to disrupt this. The rate of growth increases (τ decreases) as well, indicating that larger values of k'_{AD} and k'_{BC} promote the growth of large numbers of small fibres. Growth is faster for $k^u_{EE} = 1.0$ than for $k^u_{EE} = 0.1$, which is consistent with this picture. The second observation is that the fibre sizes and lengths for $k^u_{EE} = 0.1$ (fig. 11a) are generally greater than those for $k^u_{EE} = 1.0$ (fig. 11b), while the fibre cross-sections are smaller, *i.e.* larger values of k^u_{EE} appear to encourage the formation of shorter, fatter fibres.

The effect of varying k^u_{EE} is explored further in fig. 12, for constant k'_{AD} and k'_{BC}. These results are derived from simulations 9 and 10. It should be noted that, when $k^u_{EE} = 5.0$, offset assembly does not occur, so the data in fig. 12a and b only cover the range $0.1 \leq k^u_{EE} \leq 1.0$. With so few data points it is difficult to establish trends. Nevertheless, it appears clear that the fibre width increases, and length decreases, with increasing k^u_{EE}, in line with the ob-

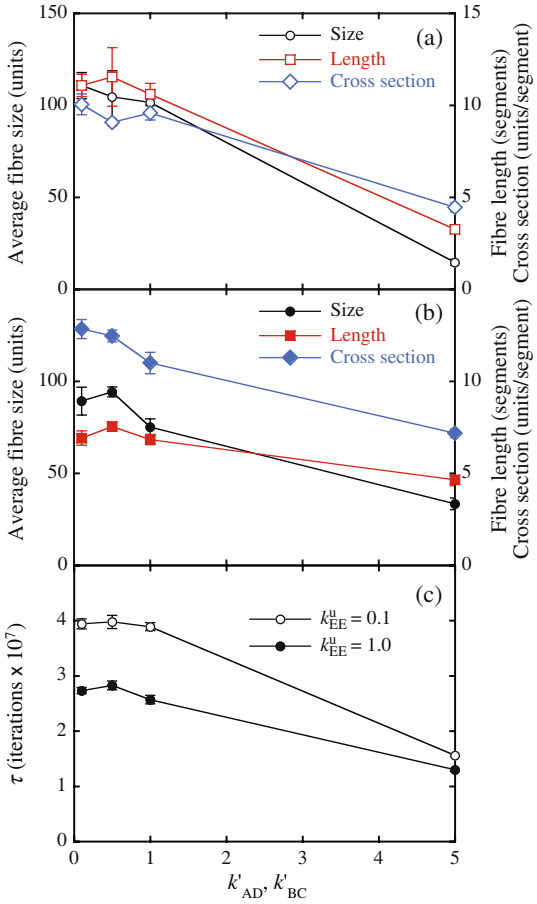

Fig. 11. (Colour on-line) Average fibre size, length and cross-section *versus* the non-adjacent part of the specific interactions k'_{AD} ($= k'_{BC}$) for (a) $k^u_{EE} = 0.1$ and (b) $k^u_{EE} = 1.0$. (c) The time constant, τ, for consumption of free units (data from simulations 7 and 8, table 2).

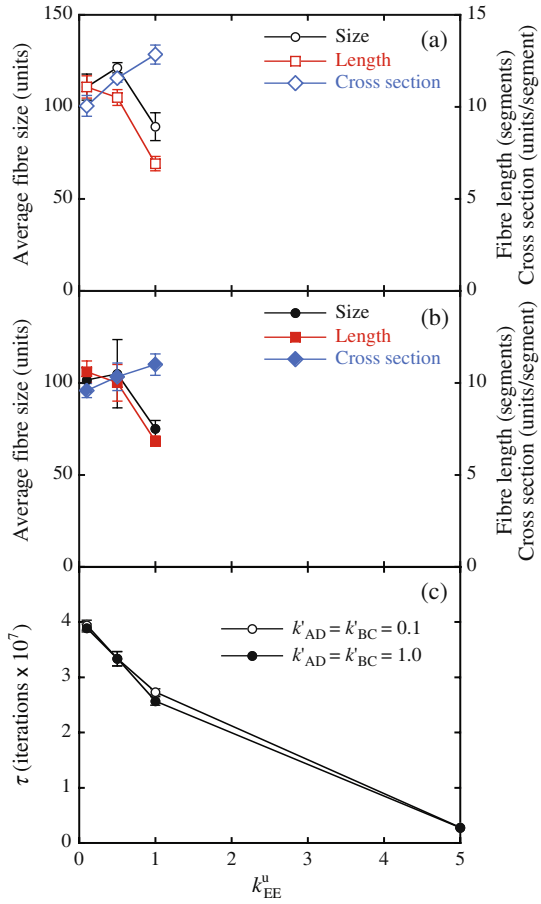

Fig. 12. (Colour on-line) Average fibre size, length and cross-section *versus* the unspecific interaction k^u_{EE} for (a) $k'_{AD} = k'_{BC} = 0.1$ and (b) $k'_{AD} = k'_{BC} = 1.0$. (c) The time constant, τ, for consumption of free units (data from simulations 9 and 10, table 2).

servations in fig. 11. On the other hand, changing k'_{AD} and k'_{BC} makes little discernible difference to the dimensions or to the growth rate.

When $k^u_{EE} = 5.0$, its effect is sufficient to cause the assembly of non-offset bundles, as observed in simulations where specificity was not used (sect. 3.1). Although in simulations 7 and 8 similarly large values of k'_{AD} and k'_{BC} were not sufficient to give rise to such assembly, the k^u_{EE} parameter represents many more interactions, which accounts for its greater effect.

The overall trends are clear. As k'_{AD}, k'_{BC} or k^u_{EE} decrease, so the growth rate also decreases, but the size of fibres formed increases. The main difference between figs. 11 and 12 is that decreasing k^u_{EE} also produces narrower fibres. To explain these observations, the following is proposed. Lateral addition can occur via all interactions and it will be favourable due to the large number of units already present in the fibre. It will be favoured, therefore, by large values of k^u_{EE}. On the other hand, longitudinal growth occurs by the offset assembly imparted by the specific interactions and is less likely to be stable until a substantial segment has formed. Hence when k^u_{EE} is large lateral growth dominates and growth is faster. Only when k^u_{EE} is small can the slower process of longitudinal growth dominate. This idea is discussed further in sect. 3.4.

3.4 Nominal growth rate vs. temperature

Fibre growth is analysed at a range of temperatures for a constant concentration of free units (simulations A in table 3). In each simulation a single fibre was grown from a seed nucleus, as described in sect. 2.2. Multiple simulations were performed and averages taken. Fibre size, measured by the total number of units in the fibre, fibre length (in segments) and mean fibre cross-section (units per segment), are plotted in fig. 13. Each simulation is equilibrated for 5×10^6 iterations in the NVT ensemble at $T = 10$ to eliminate bad contacts before switching to the μVT ensemble and, in most cases, increasing the temperature. The change of ensemble leads to a temporary cessation of longitudinal growth and, at the higher temperatures studied, a partial melting of the fibre (see fig. 13c). A second period of equilibration is required, therefore,

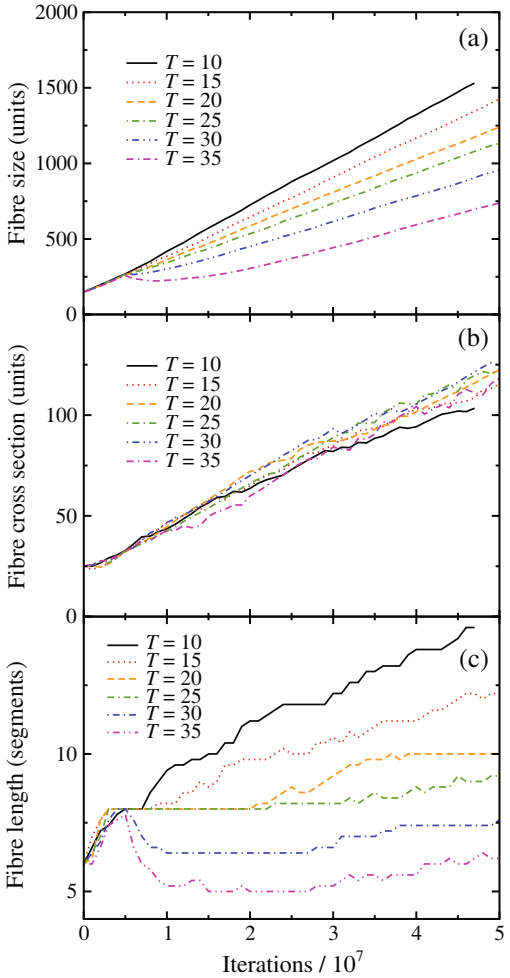

Fig. 13. (Colour on-line) (a) Total fibre size, (b) mean fibre cross-section and (c) fibre length, *versus* number of iterations, for μVT simulations performed with a constant concentration of free units ($N_F = 750$, simulation A in table 3).

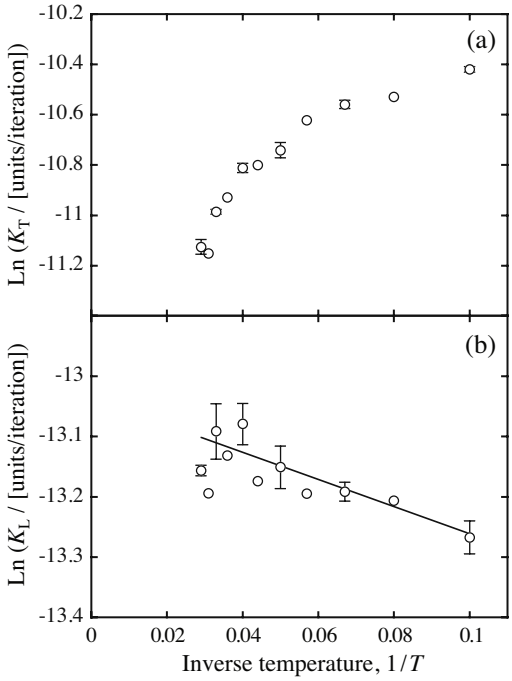

Fig. 14. (a) Overall fibre growth rate, K_T, and (b) lateral growth rate, K_L, plotted as natural logarithms *versus* the inverse temperature (data from simulation A, table 3).

before a steady-state growth regime is reached. Lateral growth continues throughout this second equilibration period (fig. 13b), with the result that the aspect ratio of the fibre decreases with increasing temperature. Because of the time taken to reach the steady state, growth rates are calculated only for the latter parts of the simulations, after 2×10^7 iterations.

Nominal growth rates for the fibres are shown in fig. 14. These are described as "nominal" because it is assumed that there is a direct mapping between Monte Carlo iterations and time. The logarithm of the overall fibre growth rate, K_T, is plotted against inverse temperature in fig. 14a, and the rate of increase of fibre cross-section, K_L, is similarly plotted in fig. 14b. It is clear that the overall growth rate of the fibres decreases with increasing temperature. However, the lateral growth rate shows the opposite trend, with a slight increase with temperature, suggesting that the overall growth rate is dominated by the longitudinal component.

Longitudinal growth data are not plotted in fig. 14, because of the noisy nature of the curves in fig. 13c, which display a sporadic growth interspersed with plateaux. However, we estimate the range to decrease from 1.4×10^{-7} segments/iteration at $T = 10$, to 0.5×10^{-7} segments/iteration at $T = 35$.

The step-like curves in fig. 13c suggest the sequential addition of new segments. In fact, these graphs are averages and, on graphs from individual simulations, the addition of single segments can clearly be seen. In animations of these simulations, the creation of new segments can be observed directly, and it is apparent that longitudinal growth is a process requiring secondary nucleation at the fibre ends. The fact that the waiting time for the fibre to start lengthening increases with temperature supports this notion.

At higher temperatures, segment additions become less likely so the rate of fibre extension decreases and, for $T \geq 30$, segments may initially melt off. In this latter case, longitudinal growth resumes once the lateral size has increased, *i.e.* the aspect ratio of the stable growing fibres decreases with temperature. The explanation for this behaviour is that the secondary nucleus needs to attain a certain width before it is stable. This is only possible if the end of the growing fibre has itself attained such a width. Hence, the fibre must first thicken before longitudinal growth is possible. It should be noted that the fibre morphology is tapered (see fig. 4). Therefore, the end cross-section required to nucleate a new segment will be rather less than the mean values given in fig. 14b.

As noted above, lateral growth does not show a strong temperature dependence, although there is a slight increase in growth rate with temperature, suggesting the

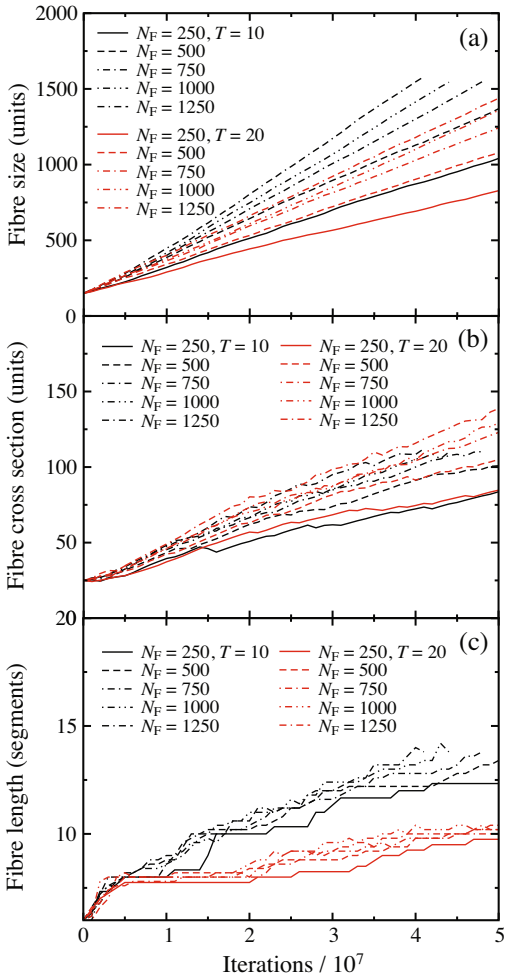

Fig. 15. (Colour on-line) (a) Total fibre size, (b) mean fibre cross-section and (c) fibre length, *versus* number of iterations, for a range of nominal free unit concentrations, N_F, at $T = 10$ (dark lines) and $T = 20$ (grey/red lines) (data from simulation B, table 3).

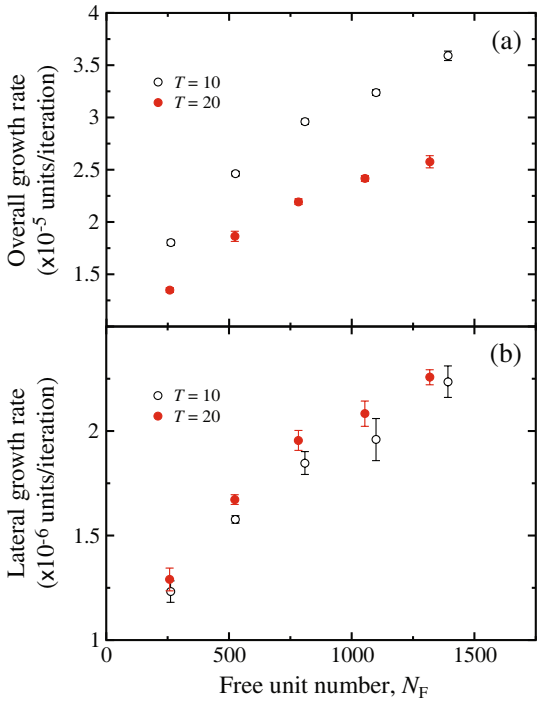

Fig. 16. (Colour on-line) (a) The overall fibre growth rate and (b) the lateral growth rate, *versus* free unit concentration, N_F, for $T = 10$ and $T = 20$. N_F is given by the mean number of free units in the simulation box, as measured during the simulation (data from simulation B, table 3).

process is thermally activated. We use the Arrhenius equation to estimate the activation energy via a least squares fit to fig. 14b. This yields a value of 2.3±0.8 in simulation energy units. Such a small activation energy for the addition of free units is as expected, given the values of the interaction parameters ($k_{AD} = k_{BC} = 10.0, k'_{AD} = k'_{BC} = 0.1$). The weak temperature dependence of the lateral growth supports the idea that lateral unit addition is stabilised by the large number of units already present and does not require nucleation. In fact, the temperatures studied here are so low that lateral addition is not significantly reversible. Interestingly, in the range of temperatures where lateral growth would become reversible, fibres would tend to melt out longitudinally. It should also be noted that there is no contribution to the growth rate temperature dependence from the rate of particle diffusion in the model, because this is determined only by the trial move scaling parameters (sect. 2.2), which are the same for all simulations.

3.5 Nominal growth rate vs. free unit concentration

Fibre growth is next analysed at a range of free unit concentrations for $T = 10$ and $T = 20$ (simulations B in table 3) using nucleated growth, as discussed previously. The total fibre size, cross-section and length are plotted against number of iterations in fig. 15, and the growth rates are plotted against free unit number, N_F, in fig. 16.

Both the overall growth rate and the lateral growth rate increase with free unit concentration. In fact, the lateral growth rate increases approximately linearly, as might be expected from a collision-based argument, *i.e.* lateral growth rate is proportional to the number of units available to bind to the fibre since lateral binding is nearly always favourable. On the other hand, the longitudinal growth rate shows only a slight concentration dependence at each temperature, which is difficult to quantify (see fig. 15c). Although lengthening is slightly preferred at higher concentrations, as would be expected, the results imply that segment addition is not limited by the supply of units but by the difficulty of formation of a new segment. That is, these results are coincident with the idea, noted in the previous section, that lateral growth is diffusion limited but that longitudinal growth is rate limited, requiring the nucleation of each segment on the end of a growing fibre.

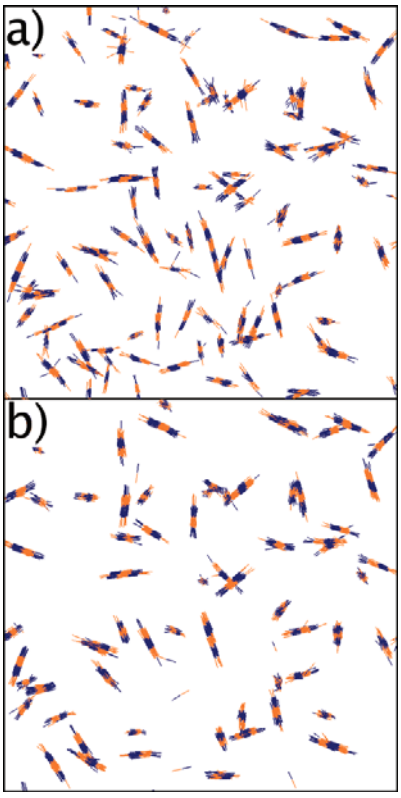

Fig. 17. (Colour on-line) Snapshots after (a) 2.5×10^7 iterations at $T = 5$ and (b) a further 2.5×10^7 iterations at $T = 10$, from a repeat of simulation 5, table 2.

3.6 Melting fibres

In addition to growing fibres, it is possible to melt pre-existing ones. In fig. 17, simulation 5 is repeated with a temperature of 5 for 2.5×10^7 iterations. The temperature is then increased to 10 for another 2.5×10^7 iterations. Figure 17a shows the simulation at the end of the low-temperature period. As previously, many small fibres can be seen. Figure 17b shows the simulation at the end of the high-temperature period. Here the smaller fibres have melted and the larger ones have thickened.

As indicated earlier (sect. 2.4), the change in shape of the fibres arises from differences in the surface free energy for the sides compared with the ends. We might model this analytically, by assuming the fibre to be a square prism, of length l and cross-section x^2. Following a standard Thompson-Gibbs analysis [31], the free-energy change on melting the fibre at temperature T_m may be written as

$$\Delta G = 2x^2\sigma_e + 4xl\sigma_s - \frac{x^2 l \Delta H}{T_\mathrm{m}^0}(T_\mathrm{m}^0 - T_\mathrm{m}) \equiv 0, \qquad (6)$$

where σ_e and σ_s are the surface free energies per unit area of the ends and sides, respectively, and ΔH is the enthalpy per unit volume of the fibre. From this expression we obtain, for T_m

$$T_\mathrm{m} = T_\mathrm{m}^0 \left(1 - \frac{2\sigma_e}{l\Delta H} - \frac{4\sigma_s}{x\Delta H}\right). \qquad (7)$$

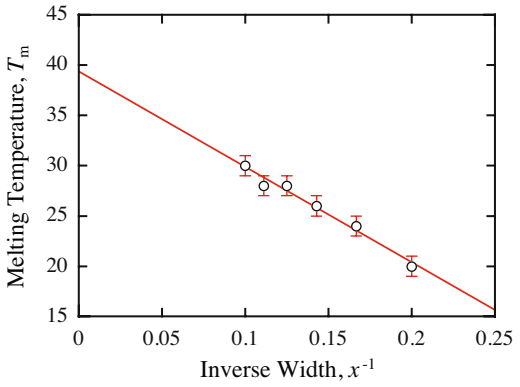

Fig. 18. (Colour on-line) The variation of melting temperature, T_m, with the inverse width of fibre, for a system with $k_\mathrm{AD} = k_\mathrm{BC} = 10$, $k'_\mathrm{AD} = k'_\mathrm{BC} = 0.1$ and $k^\mathrm{u}_\mathrm{EE} = 0.1$. The simulations were performed with a fibre length, $l = 14$.

By simulating the melting of fibre nuclei with different dimensions, it is possible, on plotting T_m *vs.* $1/l$ for constant x, and T_m *vs.* $1/x$ for constant l, to extract values for $\sigma_e/\Delta H$, $\sigma_s/\Delta H$ and T_m^0.

Examination of the system studied in simulation 9, table 2, as well as in simulation A, table 3 ($k_\mathrm{AD} = k_\mathrm{BC} = 10$; $k'_\mathrm{AD} = k'_\mathrm{BC} = 0.1$; $k^\mathrm{u}_\mathrm{EE} = 0.1$) for nuclei with $5 \leq x \leq 10$ and $6 \leq l \leq 30$, indicated that, in fact, there is no obvious variation in T_m with l, *i.e.* $\sigma_e \simeq 0$. This suggests that there is little difference in the free energy of a peptide unit when it is in the bulk of the fibre compared with its sides. On the other hand, for the side free-energy term, we find $(\sigma_s/\Delta H) \simeq 0.5$ (length units) and finally, that $T_\mathrm{m}^0 \simeq 40$ (*i.e.* $T_\mathrm{m}^{0,*} \simeq 2$). The variation in T_m with fibre width is shown in fig. 18.

If we re-examine fig. 13, we note that $T < T_\mathrm{m}^0$ for all of the temperatures considered. However, at the end of equilibration, we note that the fibre has dimensions, from the graphs, of $l = 8$ and $x \simeq 5.5$. Substituting these values into eq. (7) gives $T_\mathrm{m} = 25.5$. Therefore, we expect to observe partial melting, through shortening of the fibre, for $T = 30$ and $T = 35$, as apparent in fig. 13c. Ultimately, the fibres attain sufficient width that longitudinal growth is able to resume at all temperatures. For $T = 35$, this occurs when $x \gtrsim 8$, *i.e.* after about 2×10^7 iterations (see fig. 13b). It can be seen that this point does, indeed, correspond to a resumption in longitudinal growth, as observed in fig. 13c.

To study melting further, a single fibre is nucleated and grown to large dimensions in a μVT simulation (simulations C in table 3). An equilibration period of 2.5×10^7 iterations at $T = 10$ is used to grow the fibre to dimensions of approximately $l = 11$ and $x = 8$, after which the temperature is raised to the values given in table 3. Figure 19a shows fibre size as a function of number of iterations. The fibre increases in size for $T \leq 48$, and decreases in size for $T \geq 50$. This is also true of the fibre cross-section, seen in fig. 19b. The fibre length, however, (fig. 19c), decreases for all values of temperature. Hence, for most of the temperature range studied, fibres are able

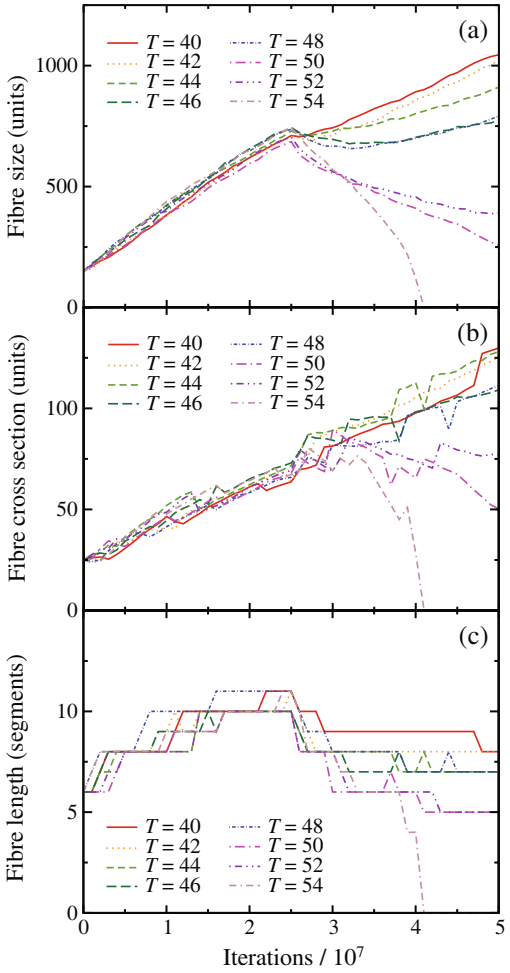

Fig. 19. (Colour on-line) (a) Total fibre size, (b) mean fibre cross-section and (c) fibre length, *versus* number of iterations at a range of final temperatures. For the growth period in the first half of these simulations $T = 10$ (data from simulation C, table 3).

to continue to grow laterally while melting longitudinally. Only at sufficiently high temperature does lateral melting also occur. This final study reinforces the conclusion that lateral addition is more stable than longitudinal addition and therefore that the latter is the limiting process in fibre formation.

Ultimately, despite the lateral growth, we would expect all fibres in fig. 19 to melt out, because $T \geq T_{\mathrm{m}}^{0}$ in all cases; it is just a matter of time before the reduction in fibre length outweighs any gains through lateral growth.

4 Discussion

The main finding of the n-site rigid-rod model for SAF crystallisation is that, under the correct conditions, the peptide sequences may be induced to form fibres, with very different growth rates for the lateral and longitudinal growth. The lateral growth of the model is comparatively rapid, with a slight dependence on temperature suggestive of a small activation energy. This is probably related to the need to orient the rod correctly before it may attach to the growth front. However, the main limitation on lateral growth appears to be the diffusion rate of the crystallisable units, *i.e.* lateral growth is diffusion limited. Longitudinal growth, on the other hand, is comparatively slow and displays a strong temperature dependence. The growth rate decreases with increasing temperature, consistent with a model of secondary nucleation, *i.e.* it is necessary to nucleate each new segment of the fibre before growth can continue. The net result is that longitudinal growth is much slower than lateral growth, unless high undercoolings are employed.

The overall picture, therefore, is one in which longitudinal growth is slow and requires nucleation, whereas lateral growth is rapid and limited only by the supply of crystallisable material to the growth front. Fibre growth is possible; however, increasing the temperature reduces the length of the growing fibre and results in a broader structure.

Similar results have been observed experimentally [32] and in computer simulations of colloidal hard-rod systems [33]. In those cases it was noted that longitudinal fibre growth was essentially self-poisoned by uncrystallised units lying flat on the crystal surface. The need to reorient such units before further crystallisation could occur presented a significant free-energy barrier to growth. In the present system, offset assembly means that the end surfaces of the growing fibre are not flat, but consist of an array of half-units spaced out by neighbours in the adjacent layer, rather like spines on a hedgehog. Longitudinal growth requires the insertion of new material between the spines, rendering the process even more difficult.

There are a number of points in which the n-site model deviates from the experimental system. Perhaps the most obvious is the treatment of the peptide sequences. In the model, all chemical detail of the peptides is replaced by n sites interacting via soft Lennard-Jones-type potentials. Efforts are underway to incorporate specific hydrophobic and electrostatic interactions into the model, and these will be discussed in a future publication.

A second issue concerning the representation of the peptides is the helix-coil transition. In the model, the peptides are treated as rods in solution. However, in the SAF system there is evidence, from circular dichroism spectroscopy, that the two peptides are largely unfolded in isolation, folding to become approximately half α-helical upon mixing [17]. This conformational change implies a free-energy barrier that is not included in the model.

A third difference between the n-site model and the SAF system concerns the formation of coiled coils. In the SAF system, adjacent α-helices twist around each other to form coiled coils. This configuration is controlled by the arrangement of functional groups on the outsides of the α-helices. It is the coiled coils that ultimately pack together to form fibres. However, in the n-site model, each peptide is a cylindrically symmetric rod that interacts equally with all neighbours. Therefore, the lateral packing is quite different from that occurring experimentally.

Despite these simplifications, the n-site model has been successful both in demonstrating the types and strengths of interactions required to promote fibre formation, and in highlighting possible mechanisms for lateral and longitudinal growth. However, the mode of growth of the SAF system contains features which are not reproduced in the current model. In particular, the SAF fibres are very long, typically many μm in length, and generally of uniform thickness, typically a few tens of nm, as illustrated in fig. 2. Given that the n-site model suggests that longitudinal growth is inhibited compared with lateral growth, the question remains as to what other mechanisms might be brought into play, which might limit the lateral growth and produce fibres of such high aspect ratio.

It is first worth noting that the dimensions of the fibres shown in fig. 2 are considerably greater than those in the model. The typical fibre in fig. 2 contains about 2500 peptide units in cross-section, whereas the simulations contain a maximum of around 150 units per cross-section. It is clear, therefore, that we are modelling the early stages of SAF assembly, and it is possible that a new growth limiting mechanism might be observed if we were to run much larger simulations.

One possible inhibitor of late-stage lateral growth could be a build-up of defects in the structure, resulting in such a disordered interfacial region that further growth is disfavoured. In the simulations, the growing fibres are already quite poorly ordered from the start of the growth process, so it is unlikely that this effect will be seen. However, we should note that, in the real systems, the degree of order required for crystal growth is higher [15], with the packing of the peptide sequences dependent on the formation of coiled coils. Therefore the real systems may be more susceptible to the accumulation of packing defects or lattice strain than the model, and this proposed mechanism may be of some importance.

Other factors that might be responsible for limiting the width of the SAF fibres include many which are important in conventional crystal growth. For example, growth might be limited by the rate at which units can be transported to the growth front, *i.e.* there could be localised depletion of crystallisable material. Related to this is the possibility of constitutional supercooling, in which compositional changes at the growth front resulting from depletion of the crystallisable units can lead to a reduction in the effective undercooling of the solution and hence in the driving force for crystallisation. This normally results in growth instabilities, however, which are not observed in individual SAF fibres. Growth might also be inhibited by the release of latent heat during crystallisation which would also affect the undercooling, if transiently, or by the poisoning effect of the accumulation of impurities. This latter effect could be electrostatic in origin, given that the peptide sequences are charged, and could arise from the formation of an electric double layer at the surface of the fibre. A further possibility is that size limiting results from a competition between interactions with different length scales. For example, the energy of a fibre might depend on a short-ranged attractive van der Waals contribution, which will scale approximately with the volume of the fibre, and long-ranged electrostatic repulsions, which will scale as a higher power. In this scenario, it is possible to envisage a maximum fibre size, above which the destabilizing repulsions will dominate. Such an effect has been demonstrated recently in polyelectrolyte bundles [34–37], and may also play a role in determining the SAF fibre dimensions.

It can be seen that there are many factors which might contribute to the shapes of the SAF fibres. Of those discussed, the kinetic effects, *i.e.* those relating to the transport or accumulation of molecular species or energy, are likely to influence the fibre width more than the length because, as suggested by the simulations, lateral growth occurs more rapidly than longitudinal growth. However, the current level of modelling is not capable of distinguishing between the various possible explanations. For this reason, we are developing a more detailed model which will include more realistic potential energy functions with parameters derived from experiment. This will be the subject of a future publication.

5 Conclusions

It has been shown that a coarse-grained rigid-rod model can reproduce some of the growth characteristics of fibres formed from α-helical peptide sequences. In particular, it has been shown that the lateral and longitudinal growth proceed via different mechanisms, with lateral growth being diffusion controlled while longitudinal growth requires secondary nucleation events at the ends of the growing fibre. The aspect ratio of the resultant fibres may be adjusted by controlling the lateral and longitudinal growth rates, which may be effected either by adjusting the relative strengths of the interactions between the different peptide sequences, or by adjusting the system temperature. The model does not explain why, in the experimental peptide system, fibres are able to grow at constant width. Various explanations for this behaviour have been discussed, and are the subject of ongoing work.

The authors thank Unilever plc and the Engineering and Physical Sciences Research Council (EPSRC) for supporting this project by co-funding a studentship, and Mr. P.J.S. King and Dr. E.H.C. Bromley for frequent, helpful discussions. This work was carried out using the computational facilities of the Advanced Computing Research Centre, Bristol University: `http://www.bris.ac.uk/acrc/`.

Appendix A. Grand canonical Monte Carlo

The acceptance probabilities for particle insertion and removal in the grand canonical ensemble are given by [24]

$$p(N_F \to N_F{+}1) = \min\left(1, \frac{zV}{N_F{+}1}\exp\left(\frac{-\Delta E}{k_BT}\right)\right), \quad \text{(A.1)}$$

$$p(N_F \to N_F - 1) = \min\left(1, \frac{N_F}{Vz}\exp\left(\frac{-\Delta E}{k_BT}\right)\right), \quad \text{(A.2)}$$

$$z = \frac{\exp(\frac{\mu}{k_BT})}{\Lambda^3}, \quad \text{(A.3)}$$

where additionally z is the activity, μ is the chemical potential and Λ is the de Broglie wavelength. Equating the expressions for the insertion and removal probability yields an approximate relation between μ and the free unit concentration N_F/V

$$\mu \simeq k_BT \log \Lambda^3 (N_F/V). \quad \text{(A.4)}$$

Under such a scheme Λ cancels and does not need to be specified in the simulation. For the purpose of determining the values of μ, we take $\Lambda = 1$.

Appendix B. Reduced temperature expression

To derive an expression for the energy of the fibres, E, in terms of the interaction parameters, we proceed as follows. We assume the form of offset assembly in the fibres, as observed in sect. 3.2 onwards. The interactions A-D, B-C, A-A, B-B, C-C and D-D will usually occur with the half-units laterally adjacent, as may be deduced from fig. 1. However the remaining interactions, where they occur, will tend to have their half-units positioned end-to-end and hence will make a much smaller contribution to the fibre energy. For this reason they will be neglected.

The minimum energy configuration of two units occurs when their sites reciprocally lie in the double minimum formed by the potentials of two sites on the adjacent unit. This is illustrated in fig. 20, and represents a slight modification to the picture shown in fig. 3. Neglecting end effects, it can be seen that each site on one unit is adjacent to two sites on its neighbour. Evidence for this packing arrangement is shown in the radial distribution functions in fig. 8.

For an unspecific interaction the contributions from the two sites will be equal while, for specific interactions, one contribution will arise from the adjacent site component (parameter k_{PQ}) and the other will arise from the non-adjacent site component (parameter k'_{PQ}). Provided that the potential well width is less than or of the order of the site separation, sites separated by greater distances will make a negligible contribution. Hence we write specific interaction terms each containing an adjacent and a non-adjacent component for A-D and B-C and unspecific terms each containing two contributions for A-A, B-B, C-C and D-D.

In a fibre of uniform width, neglecting end effects, the interactions A-D and B-C will occur twice as frequently as the self-interactions A-A, B-B, C-C and D-D, for the following reason. Any cross-section of a perfect fibre will consist of either A and D half-units or B and C half-units. If

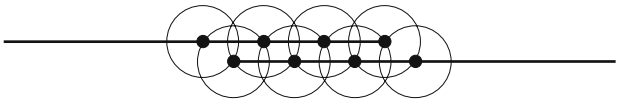

Fig. 20. The minimum energy configuration of the n-site model. Sites on one unit are located in the minima of two adjacent sites on the other unit.

there are N half-units of each species in the cross-section, consisting, say, of A and D, and N is large, there will be N^2 A-D interactions and approximately $\frac{1}{2}N^2$ each of A-A and D-D interactions. Hence we can write

$$\begin{aligned} E \sim\ & (k_{\mathrm{AD}} + k'_{\mathrm{AD}} + k_{\mathrm{BC}} + k'_{\mathrm{BC}}) \\ & + \frac{1}{2}(2k^{\mathrm{u}}_{\mathrm{AA}} + 2k^{\mathrm{u}}_{\mathrm{BB}} + 2k^{\mathrm{u}}_{\mathrm{CC}} + 2k^{\mathrm{u}}_{\mathrm{DD}}) \\ =\ & k_{\mathrm{AD}} + k'_{\mathrm{AD}} + k_{\mathrm{BC}} + k'_{\mathrm{BC}} \\ & + k^{\mathrm{u}}_{\mathrm{AA}} + k^{\mathrm{u}}_{\mathrm{BB}} + k^{\mathrm{u}}_{\mathrm{CC}} + k^{\mathrm{u}}_{\mathrm{DD}}, \end{aligned} \quad \text{(B.1)}$$

where E is approximately proportional to the energy of a uniform fibre.

Hence if a set of simulations is performed in which one or more interaction parameters are varied, the temperature can be selected for each simulation using eq. (2) such that the reduced temperature is constant for the entire set.

References

1. H. Hosseinkhani, M. Hosseinkhani, H. Kobayashi, Biomed. Mat. **1**, 8 (2006).
2. Y.R. Yoon, Y.B. Lim, E. Lee, M. Lee, Chem. Commun. **16**, 1892 (2080).
3. J. Guo, H. Su, Y. Zeng, Y.X. Liang, W.M. Wong, R.G. Ellis-Behnke, K.F. So, W. Wu, Nanomed.-Nanotechnol. **3**, 311 (2007).
4. W.F. Xue, S.W. Homans, S.E. Radford, Proc. Natl. Acad. Sci. U.S.A. **105**, 8926 (2008).
5. R. Pellarin, E. Guarnera, A. Caflisch, J. Mol. Biol. **374**, 917 (2007).
6. J.D. Hartgerink, E. Beniash, S.I. Stupp, Science **294**, 1684 (2001).
7. S. Tsonchev, K.L. Niece, G.C. Schatz, M.A. Ratner, S.I. Stupp, J. Phys. Chem. B **112**, 441 (2008).
8. Y.S. Velichko, S.I. Stupp, M.O. de la Cruz, J. Phys. Chem. B **112**, 2326 (2008).
9. R.Y. Chen, L.Y. Wu, J.M. Liao, C.L. Chen, J. Chin. Chem. Soc. **54**, 861 (2007).
10. S. Auer, C.M. Dobson, M. Vendruscolo, A. Maritan, Phys. Rev. Lett. **101**, 258101 (2008).
11. A. Aggeli, I.A. Nyrkova, M. Bell, R. Harding, L. Carrick, T.C.B. McLeish, A.N. Semenov, N. Boden, Proc. Natl. Acad. Sci. U.S.A. **98**, 11857 (2001).
12. M.J. Pandya, G.M. Spooner, M. Sunde, J.R. Thorpe, A. Rodger, D.N. Woolfson, Biochemistry **30**, 8728 (2000).
13. A.M. Smith, S.F.A. Acquah, N. Bone, H.W. Kroto, M.G. Ryadnov, M.S.P. Stevens, D.R.M. Walton, D.N. Woolfson, Angew. Chem. Int. Ed. **44**, 325 (2005).

14. A.M. Smith, E.F. Banwell, W.R. Edwards, M.J. Pandya, D.N. Woolfson, Adv. Funct. Mater. **16**, 1022 (2006).
15. D. Papapostolou, A.M. Smith, E.D.T. Atkins, S.J. Oliver, M.G. Ryadnov, L.C. Serpell, D.N. Woolfson, Proc. Natl. Acad. Sci. U.S.A. **104**, 10853 (2007).
16. J. Walshaw, D.N. Woolfson, J. Mol. Biol. **307**, 1427 (2001).
17. E.H.C. Bromley, K.J. Channon, P.J.S. King, Z.N. Mahmoud, E.F. Banwell, M.F. Butler, M.P. Crump, T.R. Dafforn, M.R. Hicks, J.D. Hirst *et al.*, Biophys. J. **98**, 1668 (2010).
18. A. Irbaeck, S. Mitternacht, Proteins **71**, 207 (2008).
19. J.H. Meinke, U.H.E. Hansmann, J. Chem. Phys. **126**, 014706 (2007).
20. J. Mondal, B.J. Sung, A. Yethiraj, J. Phys. Chem. B **113**, 9379 (2009).
21. S. Tsonchev, G. Schatz, M. Ratner, J. Phys. Chem. B **108**, 8817 (2004).
22. H. Müller-Krumbhaar, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder, 2nd edition (Springer-Verlag, Berlin, 1986) pp. 261–299.
23. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. **21**, 1087 (1953).
24. J. Yao, R.A. Greenkorn, K.C. Chao, Mol. Phys. **46**, 587 (1982).
25. H.H. Rugh, Phys. Rev. Lett. **78**, 772 (1997).
26. B.D. Butler, O. Ayton, O.G. Jepps, D.J. Evans, J. Chem. Phys. **109**, 6519 (1998).
27. B. Widom, J. Chem. Phys. **39**, 2808 (1963).
28. P.J.S. King, unpublished data (2010).
29. T. Pradell, D. Crespo, N. Clavaguera, M.T. Clavaguera-Mora, J. Phys.: Condens. Matter **10**, 3833 (1998).
30. D.A. Porter, K.E. Easterling, *Phase Transformations in Metals and Alloys* (Van Nostrand Reinhold, Wokingham, England, 1981) p. 290.
31. U.W. Gedde, *Polymer Physics* (Chapman & Hall, London, 1996) pp. 144–145.
32. H. Maeda, Y. Maeda, Phys. Rev. Lett. **90**, 018303 (2003).
33. T. Schilling, D. Frenkel, Phys. Rev. Lett. **92**, 085505 (2004).
34. H. Fazli, R. Golestanian, Phys. Rev. E **76**, 041801 (2007).
35. G.H. Lai, R. Coridan, O.V. Zribi, R. Golestanian, G.C.L. Wong, Phys. Rev. Lett. **98**, 187802 (2007).
36. M. Sayar, C. Holm, EPL **77**, 16001 (2007).
37. M. Sayar, C. Holm, Phys. Rev. E **82**, 031901 (2010).