ARTICLE TYPE www. rsc.org/xxxxxx | XXXXXXXX

Folding Kinetics of a Polymer

Štěpán Růžička, David Quigley, and Michael P. Allen*

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

We present the results of computer simulations giving a kinetic insight into the liquid-to-solid transition of a homopolymer chain with short-range interactions. By calculating the absolute rates in each direction of the transition, using molecular dynamics employing the forward flux sampling scheme, we provide the phase diagram based on purely kinetic data, and compare it with the results from Monte Carlo simulations. Additionally, we present and discuss a remarkably simple and general relation between the polymer topology and the folding pathway, and show that the eigenvalue spectrum of a matrix defined by non-bonded contacts (the Laplacian matrix) provides an insight into the nonequilibrium ensembles of these trajectories. In particular, the Laplacian matrix seems to identify a large fraction of configurations on the folding pathway at the free energy maximum that have a very low probability of reaching the crystallized state. This implies that the eigenvalues of this matrix may be suitable additional reaction coordinates to describe the folding transition of chain molecules.

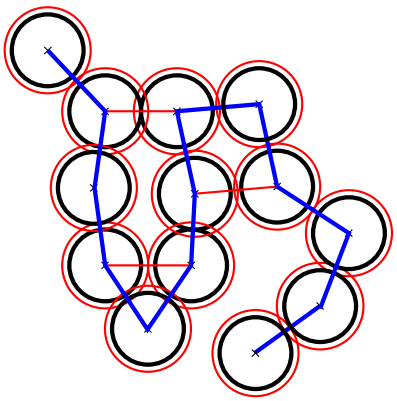

Fig. 1 Two-dimensional illustration of the flexible homopolymer chain. The hard-sphere atoms are indicated by black circles (diameter σ), and the range of the nonbonded interactions ($\chi\sigma$ where $\chi > 1$) by red circles. Bonds (with neighbour separation $r \leq \chi_b\sigma$) are illustrated by blue lines, and nonbonded interactions (separation $r \leq \chi\sigma$) by red lines.

1 Introduction

A very simple and general way to examine the underlying microscopic folding mechanisms of polymers and proteins is to analyse the behaviour of a flexible polymer chain using computer simulations. A minimalistic model of this kind[1–8], is composed of identical bonded hard spheres with square-well non-bonded attractions. This model is illustrated in Fig. 1, and will be defined in detail in section 2.1.

Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom; E-mail: m.p.allen@warwick.ac.uk

The attractive hard-sphere chain has become well established, and has been intensively studied, because of its surprisingly rich phase behaviour, including multi-stage[1] or direct[5] folding and spontaneous symmetry breaking[3] to give chiral structures. Although molecular dynamics simulations of this model[1,2] have been carried out, there has been no quantitative study as yet of the kinetics associated with these transformations. With the bonds between successive atoms restricted such that the spheres are touching (or very nearly so), the only relevant variables are the temperature (relative to the depth of the attractive square well), the width or range of the square well, and the length of the chain. (One can add terms that influence the chain stiffness, but this is not considered here). In particular long chains with sufficiently long-ranged attractive interactions (i.e. extending more than a few percent beyond the hard-sphere diameter itself) exhibit a distinct two-stage folding sequence. On decreasing the temperature, the folding is initiated by a continuous transition from an extended "random coil" conformation to a compact "globule". This transition is analogous to the vapour-liquid phase transition in simple, non-polymeric, systems. (We use the terminology "phase transition" loosely here, as it strictly applies only in the thermodynamic limit of long polymers, or in a macroscopic system of large numbers of finite-length polymers; the term "pseudo-transition" is technically more appropriate for finite sized systems). At a lower temperature, called the freezing temperature, this gradual collapse is followed by a strong first-order transition from the globule to the crystalline state. This transition is characterized by a peak in the heat capacity, and a doubly-peaked energy distribution function or (equivalently) a free energy landscape with two energy minima separated by a barrier. It is closely analogous to the freezing transition in

simple systems. For the attractive hard-sphere chain, it has been studied in computer simulations [1,2]; for the parameters used, the free energy barriers were relatively small (compared with thermal energies) and the transitions could be observed to occur spontaneously on lowering the temperature. Consequently, simple molecular dynamics (in which collisions occur between the particles, as will be described in section 2.3) were sufficient to explore the relevant phenomena; no special simulation techniques were necessary.

Much more recently, polymer chain models with short-range interactions (extending just a few percent beyond the hard-sphere diameter) have been studied by Taylor *et al.* [5,6]. In contrast to the situation described above, the collapse occurs in a single step, from the random coil state to the crystal. This is more characteristic of proteins than polymers. A similar phenomenon is observed in colloidal suspensions of spherical particles, whose attraction range may be much smaller than the characteristic particle size. For non-polymeric spheres interacting via the square-well potential, the liquid-vapour coexistence curve is known from simulation to become metastable with respect to the solid-vapour coexistence line when the well width is below 25% of the hard-sphere diameter [9,10]. Moreover, for the polymer chain, the transition was found [5,6] to be strongly first-order, with a very high barrier between the two phases. The study of the equilibrium properties in the vicinity of this strong transition was only possible due to the use of advanced Monte Carlo (MC) simulation techniques: bond-rebridging MC moves, which altered the connectivity of the chain, combined with so-called "flat histogram" sampling using the Wang-Landau technique [11]. More details of these techniques will be given in section 2.2. These simulations provided precise information about the equilibrium thermodynamics; however, they did not tell anything about the realistic kinetics and the kinetic hindering effects associated with the transition between low and high density states. One of the main motivations to study polymer chains with short-range attractions is that some features of their crystallisation may mimic, in the fashion of a toy model, the crystallisation of proteins, and a study of the dynamics of the final, structure-forming, stages of this process is of great interest. Of course, adopting a homopolymer as a starting point is a gross oversimplification of the heterogeneous interactions between amino acids in a protein, and the next logical step would be to replace the set of uniform, short-ranged, non-bonded interactions with a mixture of attractive and repulsive interactions tailored to particular amino acid sequences. We note that considerable recent progress has been made to represent proteins using hard-sphere/square-well models of this kind [12].

There is a long history of computer simulations studying the dynamics of polymer collapse, with much of the emphasis lying on the influence of solvent hydrodynamics [13,14]. However, this usually involves a non-equilibrium situation, initiated by a quench (either a sudden change in temperature or solvent quality). Even a 'slow' quench on the simulation timescale is unphysically fast in reality. The competition between crystallisation and glassification as a function of quench rate has been studied recently by molecular dynamics [15], using a variety of order parameters, to characterise the different states. The dynamics of the chain *at equilibrium* may be very different, and this is the object of study here, in the vicinity of the transition to the crystalline phase itself. Because of the height of the free energy barrier for short-ranged attractions, the events of interest will be rare, and this requires special dynamical simulation methods, which weight the sampling towards those paths that cross the barrier. Examples of these types of study are much less common: for example, we are aware of studies of the hydrophobicity-driven collapse of short chains using transition path sampling and the string method [16,17]. In the present study, we use Forward Flux Sampling [18–20], which will be described in section 2.3.

A variety of properties may be used to analyse the progress of collapse and crystallisation, including the potential energy and the radius of gyration of the polymer chain. However, an interesting alternative, reflecting the geometry of the interactions within the chain, is presented by the matrix of nearest-neighbour contacts. The contact matrix is a minimalist approach to describing protein structure: its elements are defined by contacts between pairs of amino acids. There have been attempts to examine the statistical properties of contact matrices [21,22], estimate its influence on the folding rate [23], or even use it to construct a model of folding [24]. Closely related to the present study, the contact matrix has been used in molecular dynamics simulations of competition between glassy and crystalline states of polymers [15] and in exact enumeration studies of their low-energy configurations [25]. The attractive square-well model employed here allows us to define the nonbonded contacts in an unambiguous way, and examine how the matrix is related to the folding dynamics in a very well characterised system. We give the appropriate definitions in section 2.4.

The novel feature of the current work is that we use simulation to calculate rate constants for the polymer crystallisation/melting transition, at the transition point itself, including the regime in which these rates are exceptionally low, rather than simply observing the response to a non-equilibrium perturbation in a brute-force simulation. Moreover, we check these quantitative predictions of forward and backward rates against the equilibrium thermodynamics, allowing us to identify any discrepancies with kinetic effects that are genuinely due to the finite timescale of the trajectories themselves, rather than being caused by an external driving force. In common with other work, we examine the contact matrix as a method for probing the geometry of the folding polymer, but we go further in suggesting that it may be a way to discriminate between the fates of configurations near the top of the barrier.

Following the description of our methods in section 2, we present our results in section 3, and conclude with a discussion in section 4.

2 Methods

2.1 Simulation Model

A range of simple simulation models have been useful in discussing polymer collapse. These include chains of atoms which attract each other through a Lennard-Jones potential, surrounded by an explicit solvent[26]. Also of interest are models in which the monomers and solvent have only repulsive excluded volume interactions, both on a lattice[27,28] and off-lattice[29], meaning that the collapse is driven solely by entropic effects. Collapse of semiflexible polymers has been studied using off-lattice Brownian Dynamics simulations with an effective solvent[30], and an interaction between the monomers intended to represent, in an average way, the effects of solvent quality. A comparison of models with, and without, explicit solvent[31] has highlighted some deficiencies associated with replacing the solvent by effective interactions, although this conclusion will depend on the details of the physical system and the nature of the simplified model. As mentioned earlier, the inclusion of hydrodynamic interactions may also change the picture[13].

In this study, we follow Taylor *et al.*[5,6] and adopt the simplest model that exhibits the effect of interest. The polymer is represented by $N = 128$ hard spheres, with successive atoms connected by bonds. The bonding potential is defined by a narrow infinite square well

$$u_{\rm b}(r_{ij}) = \begin{cases} +\infty & 0 < r_{ij} < \sigma \\ 0 & \sigma < r_{ij} < \chi_{\rm b}\sigma \\ +\infty & \chi_{\rm b}\sigma < r_{ij} \end{cases} , \quad \text{for } |i-j| = 1 \qquad (1)$$

where r_{ij} is the distance between centers of two monomers i and $j = i \pm 1$, σ is the diameter of the bead, and $\chi_{\rm b}$ is the relative width of the nearest-neighbour bond. In the previous MC studies[6] $\chi_{\rm b} = 1$, but for simple hard sphere dynamics, we choose a slightly larger value $\chi_{\rm b} = 1.04$; we have verified that this makes very little difference to the equilibrium properties.

The nonbonded interaction between pairs of monomers is given by the finite square well potential

$$u_{\rm w}(r_{ij}) = \begin{cases} +\infty & 0 < r_{ij} < \sigma \\ -\varepsilon & \sigma < r_{ij} < \chi\sigma \\ 0 & \chi\sigma < r_{ij} \end{cases} , \quad \text{for } |i-j| > 1 \qquad (2)$$

where ε is the depth, and χ is the relative width, of the square well. In other words, all non-adjacent pairs of atoms in the chain interact with this attractive potential. Taylor *et al.*[5,6] have determined that an all-in-one 'protein-like' crystallisation from the expanded state occurs for $\chi \lesssim 1.06$, and a two-step 'polymer-like' mechanism via a liquid-like globule, for $\chi \gtrsim 1.06$; we study the globule-crystal stage of the two-step process for χ values in this vicinity. All beads have equal mass m which we take equal to unity. Throughout, we work in reduced units: $\sigma = 1$, $\varepsilon = 1$ and $k_{\rm B} = 1$ (Boltzmann's constant) so T should be understood as the combination $k_{\rm B}T/\varepsilon$ etc. Corresponding real values, for monomer beads corresponding to amino acids of the kind found in proteins, would be $m \approx 2 \times 10^{-25}$ kg, $\sigma \approx 6 \times 10^{-10}$ m, $\varepsilon \approx 7 \times 10^{-22}$ J, and a time unit $\approx 10^{-11}$ s. We note that the idea of this model is to simulate a polymer in a surrounding liquid (rather than a vacuum), but with the explicit degrees of freedom associated with the solvent atoms "integrated out". The attractive wells between monomers represent not only the direct interatomic forces in a real polymer, but also the attraction of monomers to each other *relative to* the solvent. In other words, $k_{\rm B}T/\varepsilon$ represents the solvent quality, with a high value corresponding to a self-avoiding walk configuration in a good solvent, and a low value generating collapsed states as seen in a poor solvent. This approach (and an alternative one) to tuning the solvent quality, have been used in simulations of the same model in the presence of an explicit solvent of hard spheres[32]. As we shall see, the dynamical algorithm includes the effects of random collisions which mimic the effect of solvent atom impacts as well as providing a means of thermalizing the system.

2.2 Wang-Landau Sampling

As explained in section 1, standard Monte Carlo sampling from, for example, the Boltzmann distribution, is not sufficient to properly equilibrate the system in the presence of a large barrier between the different phases. Some kind of weighted ensemble must be sampled, to "flatten out" this barrier, and the results are typically reweighed afterwards to calculate properties at any desired temperature. Moreover, it is important to use a set of MC moves that allow efficient exploration of configuration space in both the compact crystalline state, and the disordered state. Accordingly, the statistical thermodynamics of the above chain model are determined by the Wang-Landau (WL) Monte Carlo (MC) method, using a move set consisting of crankshaft, pivot, end-bridging, and regrowth moves; the latter two being connectivity altering[33,34]. The regrowth move consists of deleting and then regrowing up to 3 beads at either end of the chain, using a configuration-bias algorithm to select non-overlapping bead positions; this includes the possibility of reversing the chain. This MC move set is used with the WL algorithm[11], after the manner of Ref. 5, to iteratively approximate the density-of-states function $W(E)$, giving a well-sampled set of configurations across the whole energy range. This involves maintaining a table of values $W(E)$, and accept-

ing proposed moves using the criterion

$$P_{\rm acc} = \min\bigl(1, W(E_{\rm old})/W(E_{\rm new})\bigr)\,.$$

At each step a "visit histogram" for the new state $H(E)$ is updated $H(E) \to H(E)+1$, and the corresponding density-of-states table entry is changed according to the standard prescription $W(E) \to fW(E)$. This has the effect of down-weighting the likelihood of visiting the same energy in future. When the histogram of visited energies $H(E)$ is sufficiently flat, the modification factor f is reduced from its initial value $f = {\rm e} = 2.71828$ towards unity by replacing $f \to \sqrt{f}$, and the histogram $H(E)$ is reset to zero. In this case, the flatness criterion was chosen to be $\min(H(E)) > 0.8\overline{H(E)}$ where the latter is the average over the sampled energies. This process is continued, progressively reducing the value of f until it reaches a prescribed limit: typically 20–30 iterations, amounting to more than 10^{10} MC moves for each run in total, are needed. By construction, this method converges towards a probability distribution of states which is proportional to the inverse of $W(E)$, and the function $W(E)$ itself is calculated along the way. After convergence, thermodynamic quantities such as the entropy $S(E)$ and temperature $T(E)$ can be obtained from the microcanonical ensemble expressions

$$S(E) = \ln W(E)\,, \quad T(E)^{-1} = \partial S(E)/\partial E \tag{3}$$

while the Helmholtz free energy and energy distribution function at any desired temperature T can be calculated from

$$F(E) = E - T\ln W(E)\,, \quad P(E) \propto W(E)\exp(-E/T)\,. \tag{4}$$

From these measurements, the thermodynamic freezing temperatures, $T_{\rm f}$, for a range of χ values, can be estimated, as described in section 3.1.

2.3 Forward Flux Sampling

Now we turn to the dynamical simulations. In collision dynamics (CD), free flight of the spheres occurs between elastic collisions in standard fashion[35,36]. Collisions occur at each discontinuity in eqns (1), (2): repulsive impulses at a separation equal to the hard-sphere diameter σ, attractive impulses between bonded atoms at $\chi_{\rm b}\sigma$, and attractive impulses between non-bonded atoms at $\chi\sigma$. All of these collisions conserve total energy and momentum. Additionally, thermal jolts reselect the velocities of individual atoms from the Boltzmann distribution at a prescribed temperature T, and introduce a stochasticity into the dynamics[1,2]. The time separation of the jolts has a Poisson distribution with mean time τ giving the strength of the coupling (typically $\tau = 0.1$ in reduced units). The dynamics of each individual trajectory therefore corresponds to an equilibrium canonical ensemble at temperature T; the non-equilibrium nature of the ensemble of trajectories results solely from the way that they are selected (or discarded), according to the evolution of a chosen order parameter. The thermal jolts also serve the purpose of mimicking the effect of collisions with surrounding solvent molecules, so that the dynamics is in some way representative of the motion of a polymer chain in a liquid, rather than in vacuum.

We choose, as the primary order parameter describing the qualitative difference between the globule and crystal state, the potential energy E of the chain. To study the ensemble of trajectories crossing the barrier between these two states, we use Forward Flux Sampling (FFS)[18–20]. This method defines a set of hyperplanes, or "interfaces", dividing phase space orthogonal to this order parameter E. We denote these interfaces by a symbol such as E_i^A where the superscript labels the original state (A for the crystal, B for the globule) and the subscript is the interface number $0 \le i \le n$.

FFS works in the following fashion, for each chosen value of χ and T. Recall that, in the vicinity of the transition temperature T between the globular and crystalline states (as determined by WL simulations), the equilibrium probability distribution function for the energy will have two peaks, of comparable weight, one characterising state A and the other state B, separated by an energy range with extremely low probability. A set of configurations, characteristic of the equilibrium ensemble at temperature T in state A, is used as a starting point. State A is formally defined as all those configurations with a potential energy $E < E^A$, a value chosen such that only a very small fraction of the corresponding peak in the probability distribution lies outside this range. Dynamical trajectories are initiated from these configurations. The "zeroth" energy hypersurface E_0^A is chosen such that $E^A < E_0^A$, and that trajectories originating from A only occasionally reach the energy E_0^A. A sample of these configurations at E_0^A (travelling in the sense $A \to B$) is taken for future use. A very long series of independent, equilibrium, simulations thereby generates a large set of configurations at this energy (typically 8192), and at the same time enables the calculation of the mean probability flux through E_0^A, which we call $\Phi^A(E_0^A)$. If P^A denotes the probability of being most recently in state A (i.e. the weight in the corresponding peak of the equilibrium probability distribution) then the fraction $\Phi^A(E_0^A)/P^A$ simply represents the inverse of the average time between crossings of E_0^A (in the direction $A \to B$) from simulations most recently in state A. The subsequent stages of FFS generate trajectories leading from each hypersurface, E_i^A, to the next, E_{i+1}^A. Each trajectory starts from a state previously stored at E_i^A, selected at random, and is evolved in time according to the equilibrium dynamics described at the start of this section, until it either reaches E_{i+1}^A or falls back to $E < E^A$. This procedure builds up a sample of new configurations (again, typically, 8192 in our work) at E_{i+1}^A, while simultaneously allowing an estimate to be made of

the fraction of partial pathways started from E_i^A which reach E_{i+1}^A before they fall back to an energy $E < E^A$. We denote this by $P(E_{i+1}^A \mid E_i^A)$: the conditional probability of reaching the interface E_{i+1}^A from E_i^A. It is important that the equilibrium collision dynamics includes a stochastic element, so that the repeated sampling of starting points at each stage does not result in identical trajectories, but instead a set of equally valid, but divergent, segments.

In this fashion, complete trajectories from A to B are built up from equilibrium (but highly selected) trajectory segments, and this allows us to calculate the rate at which this process would occur in an equilibrium simulation:

$$P^A k_{A\to B} = \Phi^A(E_0^A) \prod_{i=0}^{n-1} P(E_{i+1}^A \mid E_i^A)\,. \qquad (5)$$

All the quantities in this equation are known, except for $k_{A\to B}$, the rate constant, which is the main object of the study. Equation (5) therefore computes the rate at which trajectories leave state A, conditional on them subsequently reaching state B. This is an appropriate measure of the rate constant when the time taken to traverse the phase space between E_0^A and E_n^A is small in comparison to the average residence time in each state. FFS is performed in both directions, i.e. studying trajectories that take the system from globule to crystal, and from crystal to globule, independently. This approach allows us to focus on kinetic effects associated with each direction separately, and is readily implemented in tandem with our stochastic dynamics. The rate $k_{B\to A}$ from globule to crystal is given by an equation analogous to eqn (5). The above expression for the rate is known to be relatively insensitive to the choice of order parameter and also forms the basis for Transition Interface Sampling (TIS), the two methods differing in how the probabilities $P(E_{i+1}^A \mid E_i^A)$ are calculated[37].

The interfaces in FFS are positioned using an empirical approach as follows. Energies E_0^A and E_0^B are chosen such that the ranges $E < E_0^A$ and $E > E_0^B$ capture 99.9% of the corresponding integrated densities P^A and P^B. The boundaries of A and B are then defined as $E^A \equiv E_0^A - 50$ and $E^B \equiv E_0^B + 50$. In order to reconstruct the full pathways we define an extra plane as $E_n^A \equiv E_0^B$. Interface E_{n-1}^A is placed such that $P(E^B \mid E_{n-1}^A) = 0.9$, i.e. to ensure that the last step in the trajectory is associated with a high, but not overwhelming, probability. The remaining interfaces $E_1^A, ..., E_{n-2}^A$ are chosen according to the optimization scheme of Ref. 20. An analogous procedure applies to interfaces $E_1^B, ..., E_m^B$. Typical examples of the interface positions actually used in practice are given in section 3.

A particular concern in FFS is to sample the starting configurations in both the A and B states as thoroughly as possible. To avoid undersampling of the globule state B, 32 random disordered chains are equilibrated, and the energy E_0^B is sampled in parallel starting from these different chains. For the crystallised state, it should be borne in mind that the starting state, A, is not a perfect crystal, but an ensemble of states at the phase transition temperature T_f. Sampling of the crystallized states A, for the $A \to B$ trajectories, was achieved as follows. CD simulations started from different terminal points of $B \to A$ trajectories, E_m^B, gave a set of distributions $P_i^A(E)$, $i = 1, ..., 256$. E_0^A is then sampled in parallel starting from a configuration belonging to the distribution $P^A(E)$ having the lowest mean energy, and from 31 other configurations which are separated from this state by a large number of MC moves, including connectivity-altering moves, conducted at constant temperature. This gives a total of 32 starting configurations for sampling the first energy surface E_0^A. Section 3.2 will discuss the definition of state A consistent with this approach.

For values of χ at the upper end of the range of interest to us, the barrier between states A and B is sufficiently low that brute force (BF) collision dynamics gives good estimates of the forward and backward rates. This is achieved by conducting equilibrium CD simulations at a chosen temperature, and measuring residence times in the two states; the simulations are typically long enough that 8192 barrier crossings are observed. The distribution of residence times is then fitted to a Poisson distribution, with the rate constants as parameters.

2.4 Laplacian Matrix

Further insight into the folding and unfolding process may be obtained by studying structural details of the ensembles of polymer configurations generated along the corresponding FFS trajectories. The idea is to investigate a possible link between the topology of the interactions in the chain around the transition state of crystallisation and the dynamical effects associated with folding and unfolding.

The simplest starting point is a matrix that describes the contacts between polymer beads. The elements of the contact (or adjacency) matrix are unity for atom pairs within interaction range, and zero otherwise. Contact matrices have been used to describe the equilibrium structure of proteins in terms of amino acid contacts[21,22] and as a generator of order parameters for metadynamics simulations of atomic clusters[38]. In addition, the evolution of a crystallising polymer system has been monitored through the time correlation function of the contact matrix[15]. An extension of this idea is the Laplacian (or Kirchoff) matrix, which appears in algebraic graph theory, particularly in discussions of the connectivity of graphs and their spanning trees, and we adopt this definition here. As we shall see later, this matrix has been used to discuss the dynamical motions that give rise to protein folding[39,40], and vibrational modes in colloidal systems[41], but we simply adopt it here as a geometrical quantity.

For general interaction potentials, there is some subjectivity

in defining what is meant by a "contact", but in the current model the situation is simpler: adjacent atoms in the chain are permanently bonded together, while non-adjacent atoms may, or may not, be within the interaction range of the nonbonded pair potential, $u_{\rm w}(r_{ij})$ defined in eqn (2). Here we focus on *nonbonded* contacts, in a sense combining the ideas discussed in the previous paragraph. Our nonbonded Laplacian matrix $\mathbb{G}$ is defined as follows:

$$G_{ij} = \begin{cases} -1 & \text{if } |i-j|>1 \text{ and } r_{ij} \le \chi\sigma, \\ 0 & \text{if } |i-j|>1 \text{ and } r_{ij} > \chi\sigma, \\ 0 & \text{if } |i-j|=1, \\ -\sum_{k,k\neq j} G_{kj} & \text{if } |i-j|=0. \end{cases} \quad (6)$$

The non-zero off-diagonal entries are restricted to those atoms that are not permanently bonded together, but are within interaction range of $u_{\rm w}(r_{ij})$. In this matrix the diagonal elements are defined to make the column and row sums zero; each diagonal element is equal to the number of nonbonded contacts made by the corresponding atom. The matrix $\mathbb{G}$ is invariant to any global translation or rotation of coordinates, and we focus on its eigenvalues, which are invariant to relabelling of the atoms. This makes the eigenvalue spectrum of $\mathbb{G}$ also suitable to describe the geometry of atomic clusters in which there are no permanent bonds (noting that invariance to relabelling is not a requirement for a chain molecule, in which the labels denote the sequence number along the chain). The smallest eigenvalue of $\mathbb{G}$ is zero, the largest eigenvalue will be denoted γ, and the sum of all the eigenvalues is exactly twice the number of interacting pairs of atoms, i.e. $-2E$.

It is convenient to define $P^{A\to B}(\gamma \mid E)$ as the probability distribution of γ at an interface E sampled by pathways started from A, and $P^{B\to A}(\gamma \mid E)$ as the same quantity but sampled by pathways started from B. Following the FFS simulations for each χ, these distributions are studied, on the energy hypersurfaces of interest.

3 Results

3.1 Wang-Landau Simulations

Twenty independent WL simulations were conducted at each χ. The inverse temperature, defined by eqn (3), was determined as a function of energy for each case, and averaged. Typical results are shown in Fig. 2. In the vicinity of the phase transition the curves adopt the characteristic "small-system loop" form, and the freezing temperature $T_{\rm f}^{\rm WL}$ may be estimated by a Maxwell construction, in which a horizontal line defines equal areas in the two sections of the loop. Thermodynamically, the equal-area construction is equivalent to the condition that the two phases have equal free energies, and is one of the standard methods for accurate location of the transition temperature in finite-sized systems.

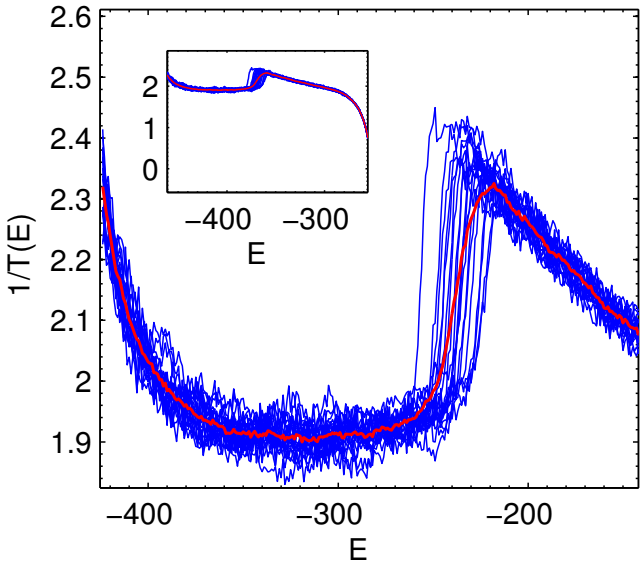

Fig. 2 Inverse temperature, from eqn (3), vs energy in the vicinity of the phase transition for $\chi = 1.07$. The graph forms a Maxwell loop and indicates a first-order transition. Blue lines show the results of twenty independent WL MC runs, while the red line represents the average. The inset shows the results over the full range of energies.

An equivalent procedure is to observe doubly-peaked canonical ensemble energy distributions $P(E)$, defined in eqn (4), as a function of T, as typified by Figure 3. Away from the phase transition, these distributions would consist of a single peak, approximately Gaussian in form, around the value of the average energy. Very close to the phase transition temperature, the system spends time in both phases, giving a doubly-peaked structure. For temperatures just below, and just above, the precise transition temperature, one or other of these peaks would dominate; the transition temperature $T_{\rm f}^{\rm WL}$ corresponds to equal weight in the two peaks. An advantage of the Wang-Landau method is that it allows us to construct these distributions, for any desired temperature, after the simulations are completed, from the density of states, according to eqn (4). Also shown in Figure 3 is the free energy curve as a function of energy, defined by $F(E) = -T\ln P(E)$; this gives a clear picture of the barrier to be crossed in the process of interconverting the crystalline phase and the globule phase. The phase diagram of $T_{\rm f}^{\rm WL}$ vs χ, determined from these results, is shown in Fig. 4; it is quite similar to that of Ref. 6.

3.2 Barrier Crossing Rates

Figure 3 also shows a typical set of interface positions for the FFS CD simulations in both directions, and the energy distributions given by the CD simulations in the states A and B; these graphs must be scaled according to the folding and un-

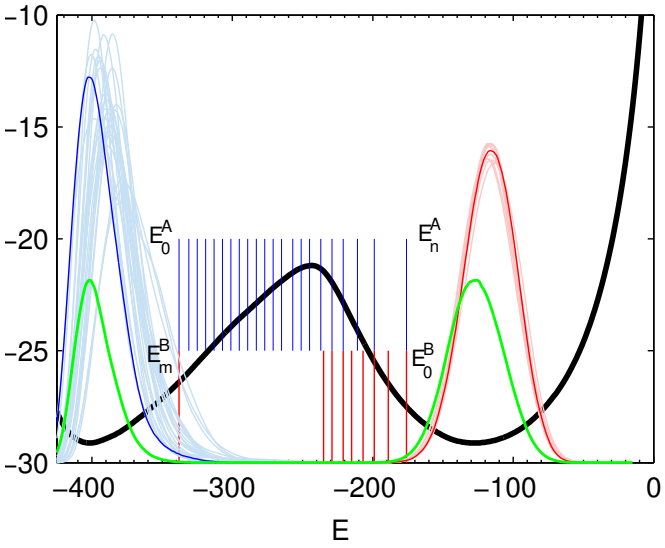

Fig. 3 Free energy $F(E)$ (black) and energy distribution function $P(E)$ (green), obtained from WL simulations for the chain with $\chi = 1.07$, and calculated for $T = T_f^{WL} = 0.498$ from eqn (4). Vertical lines schematize the energies of energy hypersurfaces in the configurational space used to generate folding (red) and unfolding (blue) trajectories. Displayed also are energy distributions $P^A(E)$ and $P^B(E)$ of states in the crystalline (blue) and globule (red) phases, respectively. The probability axis is not shown.

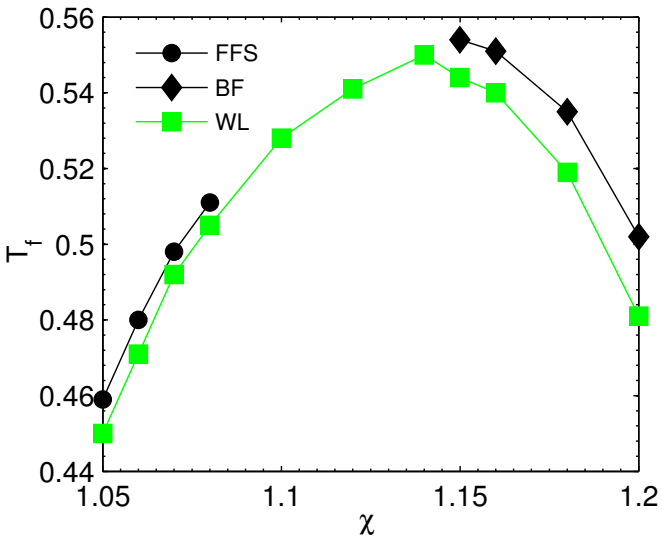

Fig. 4 Phase diagram with freezing temperatures T_f^{WL}, determined from WL simulations by Maxwell construction, and T_f^{FFS}, T_f^{BF}, determined from CD simulations by constructing chevron plots. The errors are smaller than the symbol sizes.

folding rates[42] to compare with $P(E)$. It is noticeable that the CD simulations confirm a unique globule phase, B, and we may identify a single energy distribution $P^B(E)$. However, CD simulations initialized in different realisations of the crystal phase did not explore as wide an energy range as the corresponding WL simulations. From this we infer that the crystal phase (at the freezing temperature T_f) consists of a large number of basins with slightly different mean energies, separated by kinetic barriers which cannot be overcome (at this temperature) on a simulation timescale without the use of unphysical MC moves. We therefore identify the *crystal state* A as the state, reachable from other states within the distribution function $P^A(E)$ via constant temperature MC (including unphysical moves), which has the lowest mean energy. This is the rationale behind the selection of starting configurations for the $A \to B$ trajectories, described in section 2.3. We should re-emphasise that the aim here is to study the crystal phase at the equilibrium transition temperature and not, for example, to use the minimum-energy perfect crystal as our starting point for unfolding trajectories. This inevitably complicates the kinetic analysis and, incidentally, violates the assumptions of simple models such as the Zwanzig two-state picture of protein folding[43]. We return to this in section 4.

The rate constants obtained from FFS simulations in both directions are plotted against inverse temperature on a so-called chevron plot, Fig. 5 (based on the assumption of Arrhenius-like behaviour). Transition temperatures T_f^{FFS}, defined dynamically by the equation $k_{A\to B} = k_{B\to A}$, are shown in Fig. 4. These were slightly, but systematically, higher than the corresponding WL transition temperatures by a small amount $\Delta T = T_f^{FFS} - T_f^{WL} = 0.005$–$0.012$.

For the higher values $\chi \geq 1.15$ studied in this paper, barrier crossing could be observed directly, and rate constants were calculated by brute force (BF) simulation. Transition temperatures T_f^{BF} obtained in this way are also shown in Fig. 4 and were again comparable to, but slightly and systematically higher, than the temperatures obtained in MC simulation. This suggests that the discrepancy is due to a real dynamical (kinetic) effect rather than any deficiency of the FFS algorithm itself. In Fig. 5 we indicate schematically the extent to which the unfolding rate constant must be shifted to give the observed shift ΔT, and we return to this in the Discussion of section 4.

3.3 Laplacian Matrix

Here we examine the distribution of top eigenvalues γ of the nonbonded Laplacian matrix $\mathbb{G}$ at specific energies, along the folding and unfolding FFS trajectories. We focus on the chain with $\chi = 1.07$ and $T = T_f^{WL} = 0.498$, but similar results were obtained for other attractive well widths, becoming even more distinct with decreasing χ. For the purposes of discussion, we denote the energy at the top of the free energy barrier as E_{max}, and we note that the penultimate interface in the folding

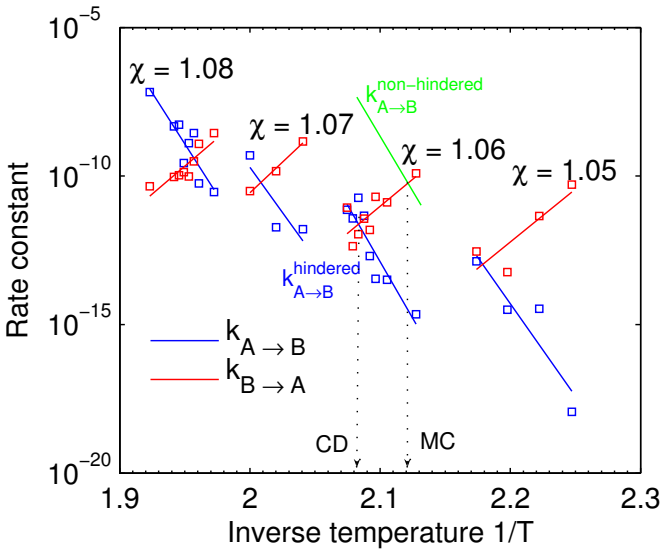

Fig. 5 Chevron plots with unfolding $A \to B$ and folding $B \to A$ rate constants computed by FFS. The intersections give the transition temperatures T_f^{FFS}. The schematic for $\chi = 1.06$ indicates how much the unfolding rate would have to be shifted to bring T_f^{FFS} into agreement with T_f^{WL} (see discussion in section 4).

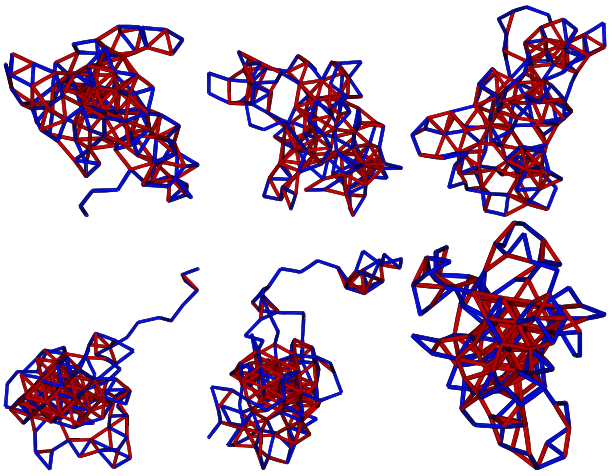

Fig. 6 Configurations for the system with attractive well range $\chi = 1.07$, $T = T_f^{WL} = 0.498$, at the globule-like energy $E = -176$. Blue: permanent bonds, red: nonbonded contacts with $r_{ij} < \chi\sigma$. Top three configurations are from folding $B \to A$ trajectories, with eigenvalue $\gamma = 9.6$; bottom three are from unfolding $A \to B$ trajectories with eigenvalue $\gamma = 12.1$. Close inspection of the unfolding configurations reveals a better-ordered crystalline core.

direction $B \to A$ was typically chosen to be $E^B_{m-1} \approx E_{max}$ (see Fig. 3). For the case discussed here, $E_{max} = -235$

Along unfolding trajectories, we found that $P^{A\to B}(\gamma \mid E)$ is unimodal (approximately Gaussian) at all interfaces E, with the mean value of γ growing with E, and that the conformations at any E with large γ have a more crystalline appearance. Along folding trajectories, the distribution $P^{B\to A}(\gamma \mid E)$ at interfaces far enough from E_{max} is also unimodal with similar properties. Fig. 6 shows three configurations taken from both folding and unfolding trajectories, at the same energy $E = -176$, which lies well on the globule side of the transition. The chosen configurations are representative of the ensemble, in the sense that their largest eigenvalues γ are close to the maximum of the corresponding distribution. Unsurprisingly, the unfolding trajectories retain more structure characteristic of the crystalline starting state, and the distribution $P^{A\to B}(\gamma \mid E)$ is shifted to higher γ than $P^{B\to A}(\gamma \mid E)$. Inspection of the configurations shows that a near-crystalline core can often be identified within the unfolding configurations, typically with attached polymer tails or loops, whereas the folding configurations are more disordered. Recall that, since the energies are identical, the total number of contacts is the same in all these cases, but they are arranged differently.

The distributions $P^{A\to B}(\gamma \mid E)$ and $P^{B\to A}(\gamma \mid E)$ are qualitatively different in the region $E \approx E_{max}$, as shown in Fig. 7: the latter becomes *bimodal*. The critical value separating these two modes is denoted as γ^c; in this case $\gamma^c = 11.9$. The inset Fig. 7(a) shows that pathways started from A, and reaching these energies, do not sample the population of low-γ states seen in Fig. 7(b). The microscopic reversibility of our dynamics then implies that those folding $B \to A$ transition pathways destined to go on to the crystalline state must cross E_{max} at $\gamma > \gamma^c$. The main part of Fig. 7 presents a probability analysis of all the 8192 $B \to A$ configurations at this interface: each such configuration may be ascribed a probability of reaching the crystalline state A, based on the ultimate fate of the ongoing FFS trajectories spawned from it. In agreement with the preceding analysis, this probability analysis shows that folding pathways crossing E_{max} with $\gamma < \gamma^c$ have almost no chance to reach A. (Also, of course, the configurations with $\gamma > \gamma^c$ have a crystallization probability much less than 1, but their chances are overwhelmingly higher than those with $\gamma < \gamma^c$). Incidentally, the *equilibrium* distribution of γ at this energy, obtained by WL, shown in Fig. 7(c), is very similar to the nonreactive $\gamma < \gamma^c$ portion of Fig. 7(b). The FFS process leading up to the energy E_{max} clearly enhances the weight of those trajectories with high eigenvalues γ. Similar observations apply to other eigenvalues near the top of the spectrum of $\mathbb{G}$. It is important to emphasize that these differences in distributions are expected. Although the individual trajectories are completely microscopically reversible, and the *starting* configurations are at equilibrium, the selection process of FFS at successive interfaces makes the distribution of configurations further along the trajectories a *nonequilibrium* one. The foregoing analysis suggests that the eigenvalue γ of the nonbonded Laplacian

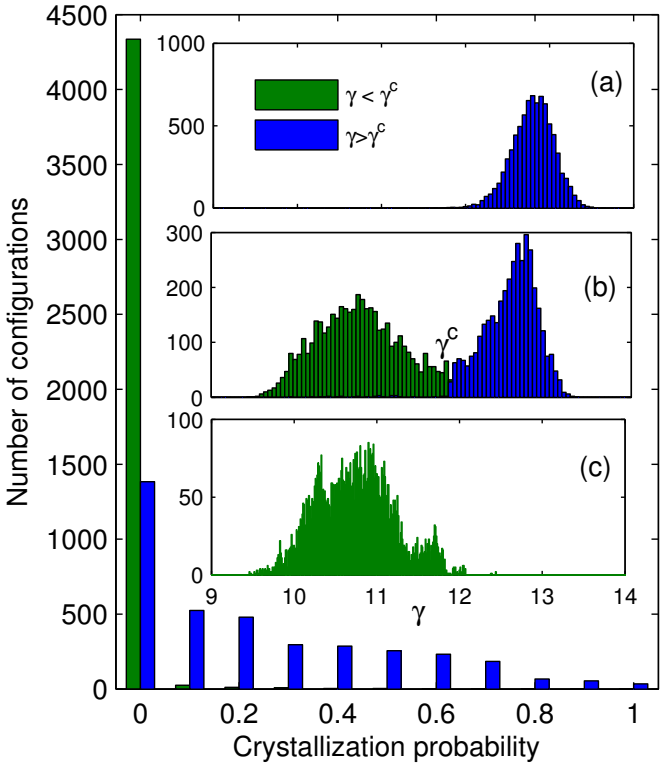

Fig. 7 Crystallization probabilities for configurations on the folding trajectory $B \to A$ at $E = E_{\max}$ with $\gamma < \gamma^c$ (green) and $\gamma > \gamma^c$ (blue). Insets show eigenvalue distributions: (a) $P^{A\to B}(\gamma \mid E_{\max})$ sampled by pathways started in A; (b) $P^{B\to A}(\gamma \mid E_{\max})$ sampled by pathways started in B (the critical value $\gamma^c = 11.9$ is indicated); (c) $P^{\mathrm{WL}}(\gamma \mid E_{\max})$, the equilibrium distribution, determined by Wang-Landau Monte Carlo.

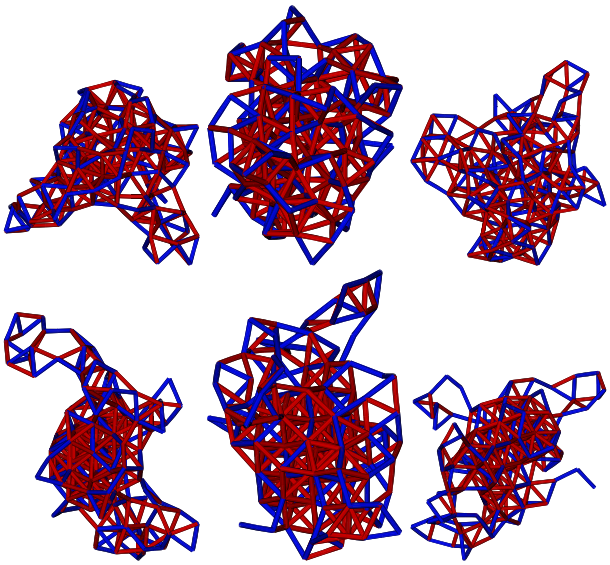

Fig. 8 Configurations for the system with attractive well range $\chi = 1.07$, $T = T_{\mathrm{f}}^{\mathrm{WL}} = 0.498$, all from folding $B \to A$ trajectories, near the free energy maximum $E = E_{\max}$. Blue: permanent bonds, red: nonbonded contacts with $r_{ij} < \chi\sigma$. Top three configurations have eigenvalue $\gamma = 10.8$; bottom three have eigenvalue $\gamma = 12.8$. The top three configurations have almost no chance to form a crystal; close inspection of the bottom three configurations reveals a better-ordered crystalline core.

matrix $\mathbb{G}$ is a good discriminator between the different ensembles of trajectories, in other words a good measure of incipient crystallinity.

Figure 8 shows typical configurations taken on the $E = E_{\max}$ hypersurface, all from folding $B \to A$ trajectories, representative of the two peaks in the bimodal $P^{B\to A}(\gamma \mid E_{\max})$ distribution. As for Fig. 6, it is significant that typical $\gamma > \gamma^c$ configurations appear to have a compact crystal nucleus with attached chains or loops, while for $\gamma < \gamma^c$ the same number of interactions are typically arranged in a less well ordered cluster. We note that similar configurations were observed by Zhou *et al.* [2] in the vicinity of the phase transition.

4 Discussion and Conclusions

We have carried out a set of simulations of the crystallisation of a single polymer chain, at and around the transition temperature, to determine the forward and reverse rates, using the forward flux sampling (FFS) method and, for suitably low barri-

ers, direct molecular dynamics. The latter method determined the rates directly, from average residence times in folded and unfolded states, while the FFS results were obtained by combining the initial flux with a product of conditional probabilities for the reactive trajectories. A family of models was investigated, with different ranges of the attractive square-well interaction between hard-sphere monomers. Rate constants spanning more than 10 orders of magnitude were successfully determined in this way.

A principal result of the study is that the top eigenvalue of the Laplacian matrix $\mathbb{G}$, based on non-bonded interactions or 'contacts' between monomers, is a good order parameter for crystallinity in such studies. For the folding trajectories, a clearly bimodal distribution of the top eigenvalue γ of $\mathbb{G}$ has been observed at the energy $E \approx E_{\max}$ corresponding to the maximum in the free energy curve $F(E)$, and this has been directly correlated with the propensity to crystallize. The branch of this distribution corresponding to higher values of γ has a very good overlap with the distribution obtained from unfolding trajectories and, correspondingly, the branch corresponding to lower values of γ has almost zero probability of crystallizing. In fact, we have found similar results if the permanently bonded pairs are included in the definition of $\mathbb{G}$; also, in practice, it makes little difference whether we define the diagonal elements as in eqn (6), or set them to zero. Also, the distributions of some of the other eigenvalues of $\mathbb{G}$ appear to show similar behaviour.

To compare with the thermodynamically-determined transition temperatures, from Monte Carlo simulations conducted using connectivity-altering moves and the Wang-Landau density of states sampling method, the independently-determined forward and reverse rate constants were compared on a chevron plot. The kinetic transition temperatures were in qualitative agreement with those determined by Monte Carlo. However, the small systematic discrepancy between them corresponds to a substantial difference in the rate constants, amounting to several orders of magnitude. There are two possible explanations. One is that the physics of collision dynamics, hindered by metastable basins acting as kinetic traps, yields significantly poorer sampling than the MC simulation, which escapes the traps with the aid of connectivity-altering moves cutting and reconnecting different segments of the chain. This explanation would suggest the definition of two transition temperatures, the effective kinetic and the thermodynamic transition temperature. Fig. 5 indicates schematically the extent to which the unfolding regression line must be shifted to give the observed shift ΔT; this could provide a macroscopic quantitative measure of what we call here the kinetic hindering in the unfolding process. The other possibility is that the FFS method itself suffers from a serious deficiency in the way that forward and backward trajectories are sampled. We cannot discount this, although the fact that brute force simulations in which long equilibrium simulations are carried out, involving a large number of forward and backward transitions, give similar discrepancies, would suggest that a physical explanation of kinetic hindering is more likely. It is worth noting that that the discrepancy between the dynamical and MC estimate of the transition temperature was also observed by Zhou *et al.* [2]. The authors attributed this difference to the (slightly) different models used in both of those simulations. By comparing our results with those of Taylor *et al.* [6], we indirectly verified this hypothesis by MC simulation, and we showed that $T_{\rm f}^{\rm WL}$ is higher for $\chi_{\rm b} = 1.04$ than for $\chi_{\rm b} = 1.0$. However, our CD results suggest that the different transition temperatures might also be caused by other factors. Hysteresis of this kind is consistent with the rugged landscape of the folded basin, and in protein folding it is typically taken as an indication of the breakdown of the simple 'two-state' model. This is evidently the case here, since the basic assumption of that model, namely "the folding transition is well described by a single folded state and a large ensemble of unfolded states"[43] is not correct for the systems that we are studying here.

We emphasize that, although it is interesting to compare the formation of glassy states with perfect crystals on quenching polymer systems at various rates[15], the systems being simulated here are intrinsically at equilibrium. Our study is, of course, limited by the finite time taken in the individual trajectories, and by the characterisation of the process by just two order parameters (the energy, and the top eigenvalue of $\mathbb{G}$). To thoroughly test the origins of the difference between the thermodynamic and kinetic phase diagrams would require a more extensive investigation using a variety of different methods, which should shed more light on how other phenomena including finite-time[44], non-stationarity[45], or inertia[46] effects, or choice of reaction coordinate, are related to the systematic discrepancy violating ergodicity. One can imagine a programme of research to identify all the free energy basins, and transition pathways between them, which could be used to construct a much more detailed kinetic model[47]. It might also be interesting to connect the phenomenon seen here with examples of kinetic arrest in loop dynamics of DNA[48] and breathing modes of colloidal particles[49]. These would all be suitable topics for future study.

The contact matrix is not the only viable choice to discuss polymer folding.. It is worth mentioning that two classes of structure on surfaces $E \approx E_{\max}$ have also been identified by Taylor *et al.* [8] using the radius of gyration (R_g) as a second reaction coordinate. We have also confirmed this in our own simulations. The correlation between R_g and γ was tested and found to be only weak in our case. An analysis similar to that in Fig. 7 suggests that γ has significantly better predictive properties than R_g. Another alternative, of course, is the standard bond-orientational order parameter[50] used in earlier studies of crystal nucleation rates[51].

Is there a deeper reason to study $\mathbb{G}$, or its relatives? It is tempting to try to associate $\mathbb{G}$ with dynamical modes that lead to polymer folding and crystallisation. The Laplacian matrix that includes *all* neighbour contacts (both bonded and non-bonded), gives a rather general connection between the topology defined by the interactions within a chain configuration and its dynamical evolution, in the (somewhat crude) approximation of an elastic network model. Indeed, in Gaussian (Rouse-like) models of polymers, there is a very close connection between the spectrum of eigenvalues of the Laplacian matrix and the structure and dynamics of the polymer[52,53]. This has been used in the discussion of proteins to identify vibrational modes of oscillations[39,54], and more recently in a newly emerging field, looking at the low-frequency vibrational modes in colloidal systems[41]. Sadoc[40] tried to associate the spectrum of $\mathbb{G}$ with dynamical modes that lead to folding and crystallisation in more general non-Gaussian contact representations of proteins. Attempts have been made to correlate the contact geometry with the folding rate[23] or the folding process itself[24]. In this picture, the largest eigenvalue would be associated with the highest frequency of vibration. However, as pointed out above, we must not assume that ideas from protein folding will translate straightforwardly to polymer crystallization, even in the 'all-in-one' limit of very short-ranged attractions: the low-energy landscape may still consist of several, kinetically separated, basins. The shortcomings of the elastic network model must also be borne in mind. One may hope to relate the Laplacian matrix to the vibrational modes described by the Hessian or an analogous dynamical matrix, which is commonly used to explore energy landscapes[47], but this would involve significant assumptions about the interactions. Such an interpretation would not apply to the present model, because of the discontinuous nature of the potentials. In addition, for a more realistic model, it may be important to distinguish between the strong interatomic bonds and weaker non-bonded interactions. Even within the Gaussian model, it has been shown that the inclusion of effects such as semi-flexibility may make a significant difference in the properties of molecules which have an identical Laplacian eigenvalue spectrum based on the harmonic bond network[55]. Nonetheless, our results clearly reinforce the view that this type of matrix is a simple object capturing successfully important topological features of the polymer on its route to crystallization. An extremely speculative interpretation of this, accepting that the Laplacian acts as an analogue for the Hessian, would be that the chain must approach the transition state in a conformation which can explore higher frequency modes in order to reach the collapsed/crystalline state. The corresponding eigenvector describes this mode in "contact space". The low-frequency folding motion in itself is not sufficient, requiring faster modes to be active such that local adoption of a crystalline structure can occur rapidly in response to the slower, gross conformational collapse. Further investigation of the possibility of a coarse-grained description of polymer folding, based on this approach, and using continuous potentials, would be highly desirable.

To summarize, the transition of the homopolymer chain from the disordered globule to the crystal state has been simulated by dynamical forward flux sampling and brute force simulation. The forward and reverse rate constants, which span more than ten orders of magnitude, have been determined. To our knowledge this study represents the first quantitative calculation of equilibrium collapse rates in a regime inaccessible to brute force simulations. We have analysed, for the first time, the eigenvalue spectrum of the Laplacian matrix corresponding to the nonbonded interactions, at the top of the free energy barrier. The top eigenvalue of this matrix yields important information regarding the forward and reverse trajectories in the folding transition, which complements the potential energy as an order parameter: specifically, it seems sensitive to the presence of a regular core, which acts as a nucleus in the crystallisation process. Although the crystal state, at equilibrium with the globule phase, consists of many metastable basins which act as kinetic traps, the net effect can be described as a relatively small shift in the effective transition temperature relative to the true thermodynamic value.

Acknowledgments

Computer facilities were provided by the Centre for Scientific Computing at the University of Warwick. We gratefully acknowledge helpful discussions with Dr. Rosalind Allen, Prof. Kurt Binder, Dr. Ellak Somfai, and Dr. Adam Swetnam, and the comments of the anonymous referees.

References

1 Y. Q. Zhou, C. K. Hall and M. Karplus, *Phys. Rev. Lett.*, 1996, **77**, 2822–2825.
2 Y. Q. Zhou, M. Karplus, J. M. Wichert and C. K. Hall, *J. Chem. Phys.*, 1997, **107**, 10691–10708.
3 J. E. Magee, V. R. Vasquez and L. Lue, *Phys. Rev. Lett.*, 2006, **96**, 207802.
4 W. Paul, T. Strauch, F. Rampf and K. Binder, *Phys. Rev. E*, 2007, **75**, 060801.
5 M. P. Taylor, W. Paul and K. Binder, *Phys. Rev. E*, 2009, **79**, 050801.
6 M. P. Taylor, W. Paul and K. Binder, *J. Chem. Phys.*, 2009, **131**, 114907.
7 D. T. Seaton, T. Wüst and D. P. Landau, *Phys. Rev. E*, 2010, **81**, 011802.
8 M. P. Taylor, W. Paul and K. Binder, *Phys. Procedia*, 2010, **4**, 151.
9 D. L. Pagan and J. D. Gunton, *J. Chem. Phys.*, 2005, **122**, 184515/1–6.
10 H. Liu, S. Garde and S. Kumar, *J. Chem. Phys.*, 2005, **123**, 174505/1–4.
11 F. G. Wang and D. P. Landau, *Phys. Rev. Lett.*, 2001, **86**, 2050–2053.
12 M. Cheon, I. Chang and C. K. Hall, *Proteins: Structure, Function, and Bioinformatics*, 2010, **78**, 2950–2960.
13 N. Kikuchi, J. F. Ryder, C. M. Pooley and J. M. Yeomans, *Phys. Rev. E*, 2005, **71**, 061804/1–8.
14 T. T. Pham, M. Bajaj and J. R. Prakash, *Soft Matter*, 2008, **4**, 1196–1207.
15 R. S. Hoy and C. S. O'Hern, *Soft Matter*, 2012, **8**, 1215–1225.

16 P. R. ten Wolde and D. Chandler, *Proc. Nat. Acad. Sci.*, 2002, **99**, 6539–6543.
17 T. F. Miller, E. Vanden-Eijnden and D. Chandler, *Proc. Nat. Acad. Sci.*, 2007, **104**, 14559–14564.
18 R. J. Allen, P. B. Warren and P. R. ten Wolde, *Phys. Rev. Lett.*, 2005, **94**, 018104.
19 R. J. Allen, C. Valeriani and P. R. ten Wolde, *J. Phys. Condens. Matter*, 2009, **21**, 463102.
20 F. A. Escobedo, E. E. Borrero and J. C. Araque, *J. Phys. Condens. Matter*, 2009, **21**, 333101.
21 M. Vendruscolo, B. Subramanian, I. Kanter, E. Domany and J. Lebowitz, *Phys. Rev. E*, 1999, **59**, 977.
22 J. L. England and E. I. Shakhnovich, *Phys. Rev. Lett.*, 2003, **90**, 218101.
23 L. Mirny and E. Shakhnovich, *Annual Review of Biophysics and Biomolecular Structure*, 2001, **30**, 361–396.
24 G. Lois, J. Blawzdziewicz and C. S. O'Hern, *Phys. Rev. E*, 2010, **81**, 051907.
25 R. S. Hoy and C. S. O'Hern, *Phys. Rev. Lett.*, 2010, **105**, 068001.
26 J. M. Polson and M. J. Zuckermann, *J. Chem. Phys.*, 2002, **116**, 7244–7254.
27 M. Dijkstra, D. Frenkel and J. P. Hansen, *J. Chem. Phys.*, 1994, **101**, 3179–3189.
28 G. Luna-Barcenas, G. E. Bennett, I. C. Sanchez and K. P. Johnston, *J. Chem. Phys.*, 1996, **104**, 9971–9973.
29 J. K. C. Suen, F. A. Escobedo and J. J. de Pablo, *J. Chem. Phys.*, 1997, **106**, 1288–1290.
30 A. Montesi, M. Pasquali and F. C. MacKintosh, *Phys. Rev. E*, 2004, **69**, 021916.
31 R. Chang and A. Yethiraj, *J. Chem. Phys.*, 2001, **114**, 7688–7699.
32 S. B. Opps, J. M. Polson and N. A. Risk, *J. Chem. Phys.*, 2006, **125**, 194904.
33 D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Academic Pr, 2002.
34 D. P. Landau and K. Binder, *Guide To Monte Carlo Simulations In Statistical Physics*, Cambridge Univ Pr, 2009.
35 M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, Oxford University Press, 1987.
36 D. C. Rapaport, *The art of molecular dynamics simulation*, Cambridge Univ Pr, 2004.
37 P. G. Bolhuis, *J. Chem. Phys.*, 2008, **129**, 114108.
38 F. Pietrucci and W. Andreoni, *Phys. Rev. Lett.*, 2011, **107**, 085504.
39 I. Bahar, A. R. Atilgan, M. C. Demirel and B. Erman, *Phys. Rev. Lett.*, 1998, **80**, 2733.
40 J. F. Sadoc, *Eur. Phys. J. E*, 2005, **18**, 321.
41 K. Chen, W. G. Ellenbroek, Z. Zhang, D. T. N. Chen, P. J. Yunker, S. Henkes, C. Brito, O. Dauchot, W. van Saarloos, A. J. Liu and A. G. Yodh, *Phys. Rev. Lett.*, 2010, **105**, 025501.
42 C. Valeriani, R. J. Allen, M. J. Morelli, D. Frenkel and P. R. ten Wolde, *J. Chem. Phys.*, 2007, **127**, 114109.
43 R. Zwanzig, *Proc. Nat. Acad. Sci.*, 1997, **94**, 148–150.
44 P. Geiger and C. Dellago, *Phys. Rev. E*, 2010, **81**, 021127.
45 J. T. Berryman and T. Schilling, *Journal of Chemical Physics*, 2010, **133**, 244101.
46 T. van Erp, *Advances in Chemical Physics*, in press, http://arxiv.org/abs/1101.0927v2.
47 D. Wales, *Energy Landscapes*, Cambridge University Press, 2004.
48 A. Bar, Y. Kafri and D. Mukamel, *Phys. Rev. Lett.*, 2007, **98**, 038103.
49 A. Olivetti, J. Barre, B. Marcos, F. Bouchet and R. Kaiser, *Phys. Rev. Lett.*, 2009, **103**, 224301.
50 P. J. Steinhardt, D. R. Nelson and M. Ronchetti, *Phys. Rev. B*, 1983, **28**, 784–805.
51 P.-R. ten Wolde, M. J. Ruiz-Montero and D. Frenkel, *J. Chem. Phys.*, 1996, **104**, 9932–9947.
52 B. E. Eichinger, *Macromolecules*, 1980, **13**, 1–11.
53 K.-h. Nitta, *Entropy*, 2009, **11**, 907–916.
54 T. Haliloglu, I. Bahar and B. Erman, *Phys. Rev. Lett.*, 1997, **79**, 3090–3093.
55 M. Dolgushev, G. Berezovska and A. Blumen, *J. Chem. Phys.*, 2010, **133**, 154905.