PAPER

Knotting and weak knotting in confined, open random walks using virtual knots

To cite this article: Keith Alexander *et al* 2020 *J. Phys. A: Math. Theor.* **53** 045001

View the article online for updates and enhancements.

Knotting and weak knotting in confined, open random walks using virtual knots

Keith Alexander[1]**, Alexander J Taylor**[1]
and Mark R Dennis[1,2]

[1] H H Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL,
United Kingdom
[2] School of Physics and Astronomy, University of Birmingham, Birmingham, B15
2TT, United Kingdom

E-mail: mark.dennis@physics.org

Abstract

We probe the character of knotting in open, confined polymers, assigning knot types to open curves by identifying their projections as virtual knots. In this sense, virtual knots are transitional, lying in between classical knot types, which are useful to classify the ambiguous nature of knotting in open curves. Modelling confined polymers using both lattice walks and ideal chains, we find an ensemble of random, tangled open curves whose knotting is not dominated by any single knot type, a behaviour we call weakly knotted. We compare cubically confined lattice walks and spherically confined ideal chains, finding the weak knotting probability in both families is quite similar and growing with length, despite the overall knotting probability being quite different. In contrast, the probability of weak knotting in unconfined walks is small at all lengths investigated. For spherically confined ideal chains, weak knotting is strongly correlated with the degree of confinement but is almost entirely independent of length. For ideal chains confined to tubes and slits, weak knotting is correlated with an adjusted degree of confinement, again with length having negligible effect.

Keywords: confined random walks, knot recognition, knot statistics, open knots

(Some figures may appear in colour only in the online journal)

1. Introduction

Random walks are natural models of flexible polymers, and statistical quantities can be probed by their computer simulation where experiments would be challenging [1, 2]. An intriguing feature of polymers, and the random walks which model them, is the possibility they might be knotted [3–9]. It has long been established that as the chain length of polymers increases, chains which do not contain a knot become exponentially rare, and the knots which do occur become increasingly complex [5, 6]. However, length is not the sole property which affects the probability of knotting. Unsurprisingly, polymers which are compact (i.e. enclosed within small volumes), whether through solvent conditions or external confinement, are also more likely to be knotted, with knotting more likely to be complex [10–17]. The presence of knots in these systems has physical consequences, for example in the ejection of viral DNA from a bacteriophage capsid [18], or in merely restricting the available space of conformations which can be explored in a given window of time. This has potential ramifications for microfluidic systems which use a variety of confined environments [19–21].

Many previous investigations of this knot proliferation have dealt with ring polymers [4, 11, 15, 22–27]. This has the advantage that any knots in the polymer are invariant (except through cutting and glueing processes such as the action of DNA topoisomerases and recombinases). However, most physical polymers are linear and open-ended, which presents a problem, as knots are only defined mathematically in closed curves, up to ambient isotopy. While an open curve is always topologically trivial, its spatial conformation can possess geometric resemblance to a knotted closed curve: the difference between an untied and tied shoelace is evident from inspection. If the endpoints of the polymer are on strands extended away from its bulk, it is relatively easy to close the ends unambiguously, trapping the knotted area for identification. However, when the endpoints lie within the polymer tangle (which is especially common in compact curves), this operation is much more difficult to define without ambiguity; there might be several possible knot types captured depending on the chosen closure path through the tangle [28]. Therefore there is an inherent ambiguity in determining whether an open curve may be described as knotted, and if so, identifying the type of 'open knot'.

In such situations, a standard method is to take many closures according to a statistical rule, and look for the most common knot type; this might involve extending the endpoints on straight lines to a point on the surface of a large sphere enclosing and centred on the curve, thus giving a closed curve with a definite knot type for each point on the sphere [29–31]. We call this method *sphere closure*. Upon averaging over area on the sphere closures, usually the most common knot—i.e. the modal average—is taken to represent the knot type of the open curve. This is not the only means of identifying knots in proteins, and indeed one of the earliest observed, deeply knotted proteins was identified by a non-statistical, geometric relaxation algorithm [32].

Complementing the sphere closure approach, we presented an alternative closure method based on identifying projections of the open curve as *virtual knots* [33], a generalization of the familiar, classical knots which can be interpreted as lying 'in-between' the standard knot types [34]. Our method, which we call *virtual closure*, involves projecting the open 3D space curve to an open knot diagram, and we consider projections in all directions over the sphere. The open knot diagrams are interpreted either as non-classical virtual knots, or classical knots; in the following, to avoid confusion, we refer to a knot only as virtual if it is not of a classical type, and these non-classical virtual knots types have been partially classified [35]. By taking many projections from uniform directions, we again build up a picture of the knotting of the open curve. Including the possibility that an open curve projection is identified as a virtual knot increases the sensitivity to identifying how an open space curve may be knotted—quantifying

the inherent ambiguity—with a subtlety missed by sphere closure. In [33] we mainly applied this method to knotted proteins; here we extend it to more general open random walks.

Recently, consideration of open curves has led to the study of *knotoids* [36]. These are essentially open knot diagrams of the sort that result from projections of open curves, and they behave very similarly to the virtual knots which occur as open knot diagrams. Indeed, while knotoids in principle can distinguish open curve projections better than virtual knots, in practice suitably powerful knotoid invariants have not been used to our knowledge. By considering knotoid diagrams as drawn in the plane (rather than the surface of a sphere, like virtual knot diagrams), open curve projections can be further distinguished with a simple extension of the Jones polynomial [37]. The important common feature of virtual closure and knotoid methods is that the problem of open curve knot recognition is mapped to distinguishing ensembles of open knot diagrams, instead of ensembles of closed curves as in sphere closure.

In trying to understand how an open curve might be identified as (virtually) knotted, we consider the statistics of the knot types that occur over different projections. As mentioned above, in earlier sphere closure studies, an open curve is identified with the most common knot type occurring over different closures, even if this does not cover over 50% of closures. For example, the trivial unknot may be the single most common knot type, occurring in 40% of closures. Should we say this curve is unknotted, despite 60% of closures being knotted (albeit in different ways)?

We attempted to resolve this in [33] by *defining* an open space curve to be knotted if at least 50% of closures result in a non-trivial knot (classical or virtual). Situations where the single most common knot type is non-trivial and covers at least 50% of closures we call *strongly knotted*. This leaves situations where the most common knot type occurs in less than 50% of closure directions, interpreted as significant ambiguity in the knot type of the open curve. Such space curves are called called *weakly knotted*: they are definitely tangled, but not in a way that can be recognised as a single knot type, classical or virtual. It is weak knotting that confounds the classification of knotting in open curves.

Unconfined, open polymers typically appear as space curves whose endpoints are extended away from the bulk [22, 38], and are therefore likely to be either unknotted or strongly knotted. Knotting ambiguity is not of much concern in such systems, as observed in early studies of sphere closure, although not quite in these terms [29]. On the other hand, *confined, open polymers* tend to have more compact and complex conformations, and we expect each curve to project to different knot types in different directions, making these good candidates for weak knotting. There have already been indications of this from a study comparing knot recognition methods [28], in which the different methods tended to disagree on knot type increasingly often with tighter confinement.

While we do not investigate it directly here, there is an important link between weak knotting and shallow knotting, in which the knot type of an open curve can be changed (often to the unknot), by removing a short length from one of the endpoints. Such shallow knotted configurations are common in protein knots and are also present as transition states in polymer dynamics, for example during an untying process, making weak knotting an important indicator of this. We expect virtual closure to be a better detector of weak knotting than sphere closure, as knot projections might be identified as virtual as well as classical knot types.

Here we quantitatively investigate how confinement affects the weak knotting of open random walks modelling confined open polymers. We generate random walks of three different types: cubically confined lattice walks, spherically-confined off-lattice walks and unconfined off-lattice walks, and compare the results for each. We focus our attention on spherically-confined off-lattice walks, and then generalise to different confining geometries, namely tubes
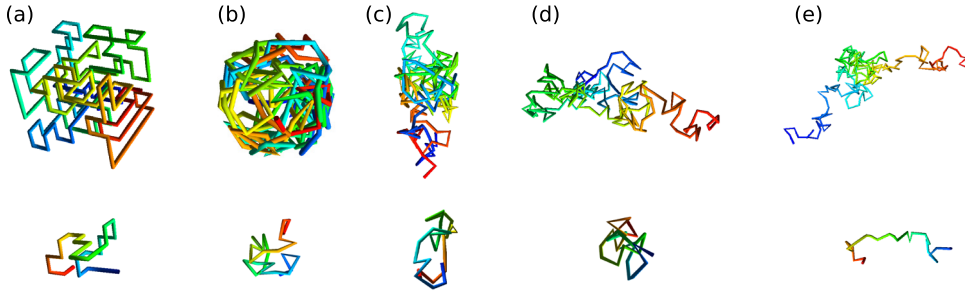
Figure 1. Examples of the classes of open random walks investigated. (a) Cubically confined lattice walks. Off-lattice walks (b) confined to spheres; (c) confined to tubes; (d) confined to slits; (e) unconfined.

and slits. We will see that while overall knotting probability is highly dependent on walk length—as is well known for ring polymers and closed random walks—this is not the case for weak knotting. Instead, the walk's degree of confinement, determined by the relative radius of gyration of unconfined and confined walks of the same length, determines the weak knotting probability with no further contribution from length. Some examples of random walks in these different classes are shown in figure 1.

2. Classical and virtual knotting of open walk closures

As in our survey of virtually knotted proteins [33], we employ both sphere closure and virtual closure to analyse the knotting of open random walks. For each open chain, we consider 100 choices of closure direction, uniformly distributed over the sphere of directions according to the generalised spiral points algorithm of [39]. In sphere closure, we project the space curve into the plane perpendicular to the projection direction, keeping crossing information. We then join the ends of the projections, adding over-crossings where the closure arc crosses existing strands of the diagram, equivalent to joining the ends to a point on a sphere of arbitrarily large radius. Any projection of this closed curve, with crossings signed appropriately, gives a closed knot diagram of the same classical knot type. We find this type by calculating the Alexander polynomial of the knot diagram [4, 40]. While this may not be the strongest knot invariant, it distinguishes enough of the different knot types that occur in the analysis of a single curve that we can reliably judge the weakness of knotting. With a perfect knot invariant, we would expect to see a very marginal increase in weak knotting detections.

Under virtual closure, the endpoints of the projected diagram are joined making *virtual crossings* with existing strands, which are not 'classical' over- or under-crossings, and indeed should not be interpreted as physical crossings. This procedure gives a closed (virtual) knot diagram whose topological information is only that of the classical crossings in the original projection. We use the generalised Alexander polynomial to distinguish virtual knot types [41, 42]. As the generalised Alexander polynomial cannot distinguish certain simple virtual knots, we also employ the Jones polynomial when necessary [43]. All classical knot types (including those with virtual crossings removable by virtual Reidemeister moves) have a generalised Alexander polynomial of zero; the classical Alexander polynomial then distinguishes these.

Examples are useful to illustrate the different closure mechanisms. Figure 2(a) shows an open curve embedded in 3-space, with (b) showing the corresponding projected diagram (the
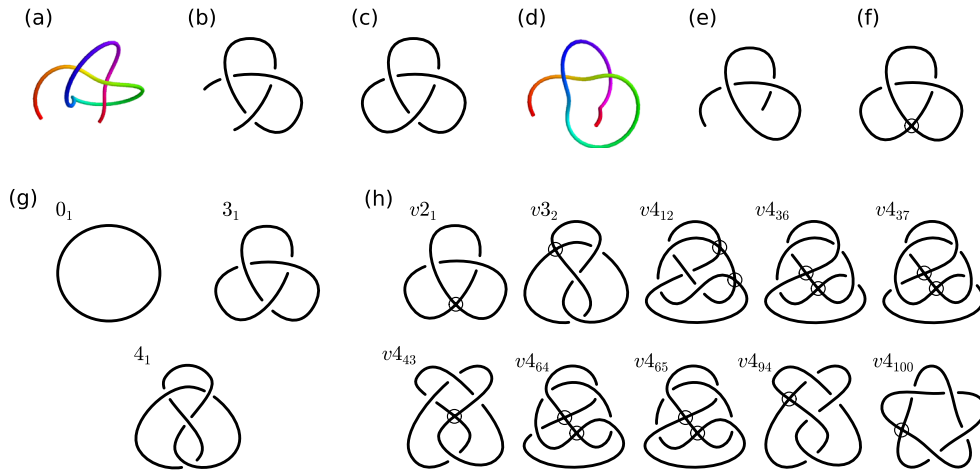
Figure 2. Examples of classical and virtual knot projections. (a) represents an open space curve, (b) its projected diagram and (c) the resulting closed knot diagram upon closing the endpoints unambiguously. (d) represents the same curve viewed from a different direction, (e) its projected diagram and (f) the resulting closed knot diagram on virtual closure. Sphere closure (including an over-crossing) results in the same knot as (c). (g) gives standard projections of the three classical knots of up to four crossings: the topologically trivial unknot, 0_1, the trefoil knot 3_1 and the figure 8 knot 4_1. (h) shows virtual knots possible from projections of open curves of up to four crossings, using the notation of [35].

diagram has been smoothed, but no crossings have changed). Figure 2(c) shows the closure of the projected diagram. Since no strands separate the endpoints, the closure introduces no additional crossings: both sphere and virtual closure identify this projection as trefoil knotted (i.e. classical knot type 3_1). Figure 2(d) shows the same open curve viewed from a different direction, with (e) the corresponding projected diagram. Here, joining the strands requires an extra crossing. Under sphere closure, this is an over-crossing, leading to a final diagram identified as a trefoil knot, as in (c). On the other hand, making this a virtual crossing instead gives the virtual knot shown in (f), the *virtual trefoil* $v2_1$. Identifying the projection as a virtual knot captures the fact that this diagram is intermediate between the classical unknot and the trefoil knot.

The possible classical knot types up to four crossings are shown in figure 2(g) with their usual labels [44]. The full set of virtual knots includes many more members than the virtual closures of projected knot diagrams which are our interest; those virtual knots which occur under virtual closure of an open space curve, of up to four classical crossings, are shown in figure 2(h). The labels here are from the virtual knot table [35], where we add the prefix 'v' to distinguish from the classical knots. All of these virtual knots except $v4_{12}$ are prime virtual knots, and appeared in the classification of more general prime virtual knots of genus 1 [45]. $v4_{12}$ in fact is a composite of two virtual trefoils, $v2_1 \# v2_1$, and is not prime as stated in [33]. Notably, that composite virtual knots appear with only four crossings, unlike the six crossing minimum seen in classical knots.

Evidently, virtual closure allows for many more possible knot types than sphere closure, for a curve or its projection of the same minimum classical crossing number (which we take as a measure of complexity). As in the example considered above, the virtual types occur in between classical knot types, and their identification sharpens the sense of ambiguity an open-chain projection can have between classical knot types. In particular, $v2_1$ is simpler—has

fewer classical crossings—than the simplest non-trivial classical knot 3_1, yet is still (virtually) knotted.

3. Generating random walks

We investigate both lattice and off-lattice walks. All walks are equilateral and have a step length of 1. In each model, for a given chain length and degree of confinement, we perform statistics based on 10 000 walks. Our lattice walks, following [46], are segments of Hamiltonian walks on a cubic lattice of $L \times L \times L$ nodes, which are automatically self-avoiding. The off-lattice walks we investigate are ideal chains, often used to model polymers in theta conditions [1]. For unconfined walks, begin at the spatial origin without loss of generality and step one unit in a uniformly randomly chosen direction. This process is repeated until the walk reaches the desired length. Each step is independent of the walk's history, and is not self-avoiding though generically does not cross itself.

It is more complicated to generate confined ideal chains. A simple algorithm is an accept-reject method, where the walk is generated like an unconfined ideal chain, starting at a random point within the confining volume, but whenever a step would take the walk outside the confined volume, it is rejected, and a new step sampled until a point inside the volume is returned. Boundaries implemented in this way are called absorbing boundaries—vertices of walks generating this way tend to avoid the region within a step length of the boundary.

A better sampling of walks confined within a sphere, very close to a uniform distribution, was proposed in [47]. Here, walks are generated like ideal chains, unless they are under a step length from the boundary, where each step is sampled from a distribution which increases towards the boundary. The radial density of walk vertices from each method is given in figure 3(a), compared to a uniform distribution.

It is easy to adapt this method for walks in slits by treating each parallel wall as an arbitrarily large sphere. More difficult is adapting to tubes, which requires steps from within a step length of the boundary to be sampled from a differently shaped distribution than that of the sphere. We outline how we do this in appendix B. While the results are not as close to uniform as the spherical case, they are close enough that any effect on knotting statistics is likely to be small. The radial vertex densities of walks in tubes using absorbing boundary conditions, the original distribution shape and our corrected distribution compared to uniform are given in figure 3(b).

4. Results

4.1. Knotting in three-dimensionally confined random walks

We begin by investigating the dependence of knotting probability on chain length, comparing unconfined ideal chains, spherically confined ideal chains and walks on the finite cubic lattices of $L \times L \times L$ nodes where $L = 6$ or 7. For the best comparison between ideal chains and lattice walks, the ideal chains are confined to spheres of equivalent volume as the lattices, with radius $R = \left(\frac{3}{4\pi}\right)^{1/3}(L-1) \approx 3.1$ and 3.72 respectively.

Figure 4(a) shows the probability that walks of N steps are knotted, i.e. if 50% or fewer closures are unknotted. The solid line gives the knotting probability under virtual closure and the dashed line knotting under sphere closure. For each walk model, knotting probability increases with length, and is strongly encouraged by confinement. The reduced flexibility of the lattice walks lowers the knotting probability compared to the ideal chains at shorter
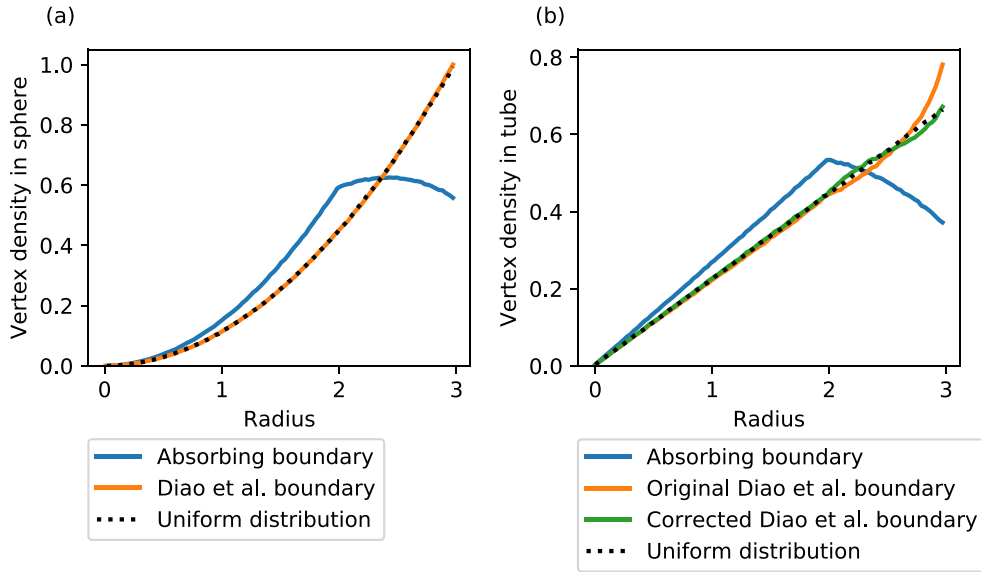
Figure 3. Distributions of vertices for spherically-confined random walks. (a) Radial distribution of vertices in a sphere with absorbing boundaries and boundaries using the model of [47] compared to a uniform distribution. (b) Radial distribution of vertices in a tube with absorbing boundaries and boundaries using the model of [47], and our modification of it, compared to a uniform distribution.

lengths, but their knotting probability begins to overtake that of unconfined ideal chains at longer lengths. For the ideal chains, the knotting probability is very similar under sphere closure and virtual closure. There is a greater discrepancy in the lattice walk results, as the knots found are likely to be simpler (and hence more cases of $\nu 2_1$ versus the unknot). For complex ideal chains, the two methods might disagree in the knot which occurs, but not whether it is knotted; on a lattice there are more conformations which are 'only just' knotted.

We now examine the conditional probability that, given a knotted walk, the knotting is weak, i.e. no single knot type occurs in 50% or more closures. Note that this is distinct from the probability of weak knotting; we want to capture the character of the knotted walks. The results for the same chains are plotted in figure 4(b); unlike the knotting considered above, unconfined walks display almost no weak knotting at any length. On the other hand, all confined walks show an increase in weak knotting with length. Strikingly, under virtual closure the trends for ideal chains and lattice walks, for the same volume size, are very close. Sphere closure gives significantly less weak knotting, primarily due to the reduced number of knot types possible under sphere closure, which reduces the competition for the most common knot. Lattice walks show the greatest reduction, where the simplicity of the curves means the number of knot types accessible is less than in the geometrically complex ideal chains.

From these results we see that confinement drives knotting from almost entirely strong knotting to almost entirely weak knotting (as a measure of knot type ambiguity). To probe this effect directly, we consider ideal chains of fixed length confined in spheres of different radius, under virtual closure (which is more sensitive to weak knotting). The dependence of (weak) knotting of fixed-length chains with respect to reciprocal confining radius $1/R$ is shown in figures 4(c) and (d), where the unconfined walks are at zero. For all lengths investigated the knotting probabilities increase as the confining volume decreases until the probability
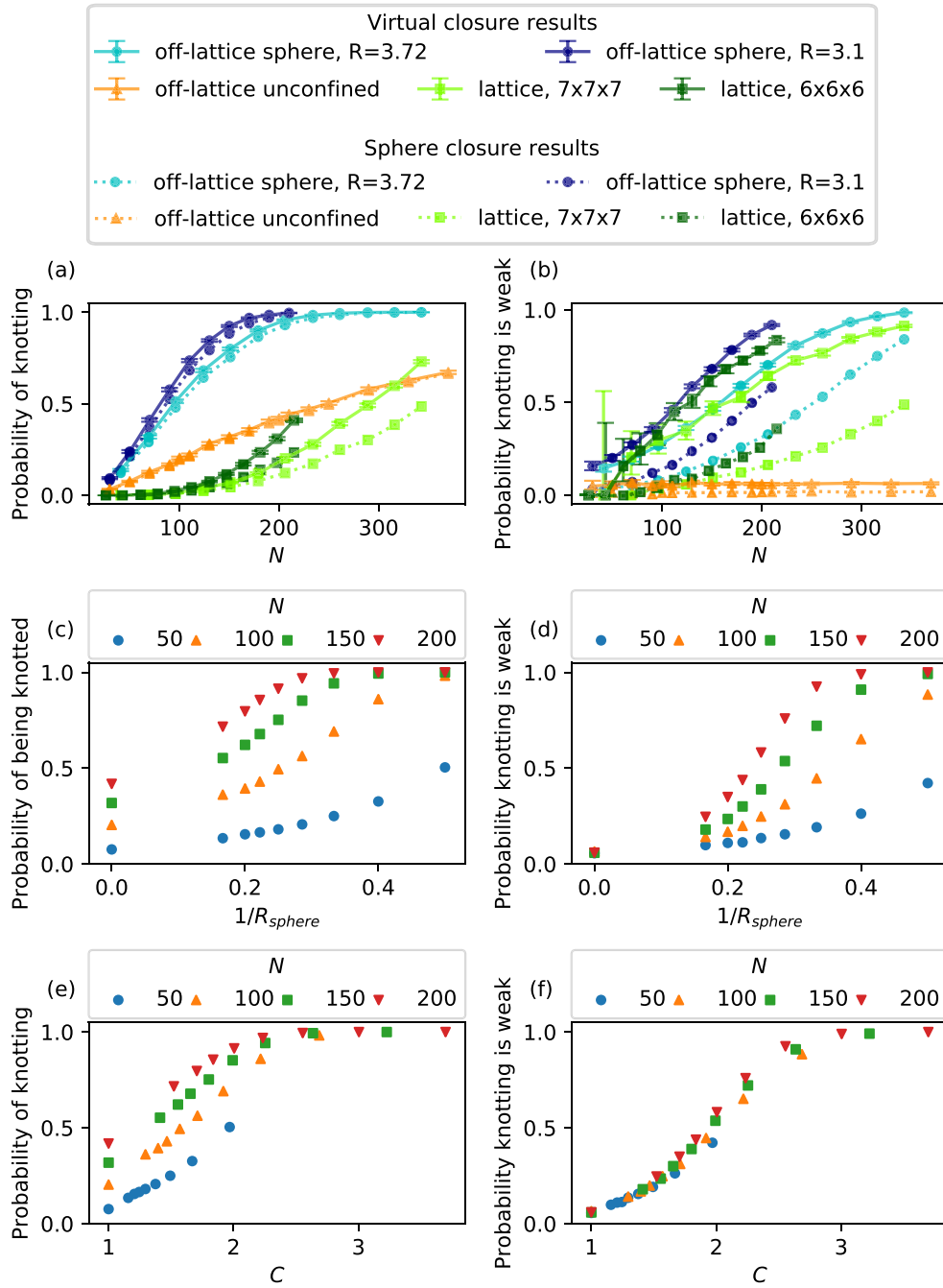
Figure 4. Knotting and weak knotting probabilities for the different random chain models. Plots on the left are for knotting probability, on the right conditional probability that knotting is weak. (a) Knotting and (b) probability knotting is weak, with different colours representing different chains as in the key. (c) and (d) show (weak) knotting in spherically confined ideal chains as the confining radius is varied. (e) and (f) show the same data as (c) and (d), plotted against the degree of confinement C as defined in (1).

saturates, with longer chains more likely to be knotted. As expected, with tighter confinement the knotting is more likely to be weak, saturating with all knotted walks being weakly knotted.

When walks are unconfined ($1/R \to 0$), the probability is small for knotting to be weak, for all walk lengths. Now, tightening the confinement means the probability of weak knotting rises faster for longer walks, as these are more constrained by a given confinement volume. We quantify this degree of confinement C as

$$C \equiv \frac{\langle R_g(N, R = \infty)\rangle}{\langle R_g(N, R)\rangle}, \tag{1}$$

where $\langle R_g(N,R)\rangle$ is the average radius of gyration of the walk of length N confined to a sphere of radius R ($\langle R_g(N, R = \infty)\rangle$ is the average radius of gyration of unconfined walks of length N). We expect $C \geqslant 1$, with $C = 1$ suggesting no confinement.

The probability of knotting with respect to C is shown in figure 4(e). Clearly knotting probability increases with C, and also with chain length. However, this length dependence does not seem to be present for the probability knotting is weak against C, plotted in figure 4(f)—the points all lie close to the same curve, over the whole range from no weak knotting at $C = 1$ to a probability of unity at $C \approx 3$. Thus the degree of confinement as parametrised by C has a crucial correlation with weak knotting, with unconfined walks not exhibiting weak knotting, and strongly confined knots almost surely weakly knotted, with the most common (virtual) knot type not covering a majority of closures.

Our definition of (weak) knotting places great emphasis on using 50% as a cutoff. However, it is unlikely that curves whose most common knot type covers 51% of closures and those with only 49% coverage are very different conformationally. For walks of a given length and confining radius, a more natural question is, what is the mean coverage of the most common knot? This proportion of closure directions covered by the most common knot for knotted curves is plotted against C in figure 5(a). Again, the dependence on C appears to be effectively independent of chain length, with a mean coverage of almost the whole sphere for $C = 1$ (unconfined), decreasing to about 25% at $C = 2.5$. The precision in the results for larger C is limited by the number of closures (100 here) considered for each space curve, as in most cases there is a different knot type for each closure direction. This supports the previous result that, as the degree of confinement of a chain increases, the knotting is more likely to be weak, as the most common knot type drops below 50% at $C \approx 2$. Plotted also is the coverage of virtual knots with C, which shows a similar N independence, rising as the mean coverage of the commonest knot falls. This suggests that with weaker knotting, a higher fraction of closures yields virtual knots.

More insight can be gained by considering the distribution of the coverage fraction of the most common knot, plotted in figure 5(b), where the different curves represent walks of different lengths with a similar value of C. For walks with a low degree of confinement ($C \lesssim 1.8$) the distribution is weighted towards a large fractional coverage, i.e. knotting is often very strong. As C increases, the distribution shifts to lower coverage fractions, such that for most tightly confined walks ($C \gtrsim 3$) the distribution is sharply peaked at low fractional coverage. It is easy to see how plot (a) results from the means of the distributions in (b).

In the Introduction, we stated that weak knotting mainly stems from the endpoints of the open random chains being close to or inside the tangled bulk of the curve. The plot in figure 5(c) verifies this statement, where the probability knotting is weak is plotted against the smaller displacement of the two endpoints from the walk centroid, normalised with respect to radius of gyration, once again binning according to C. For each value of C, the probability knotting is weak is constant up to a distance of about $0.7 R_g$ (though dependent on C), then
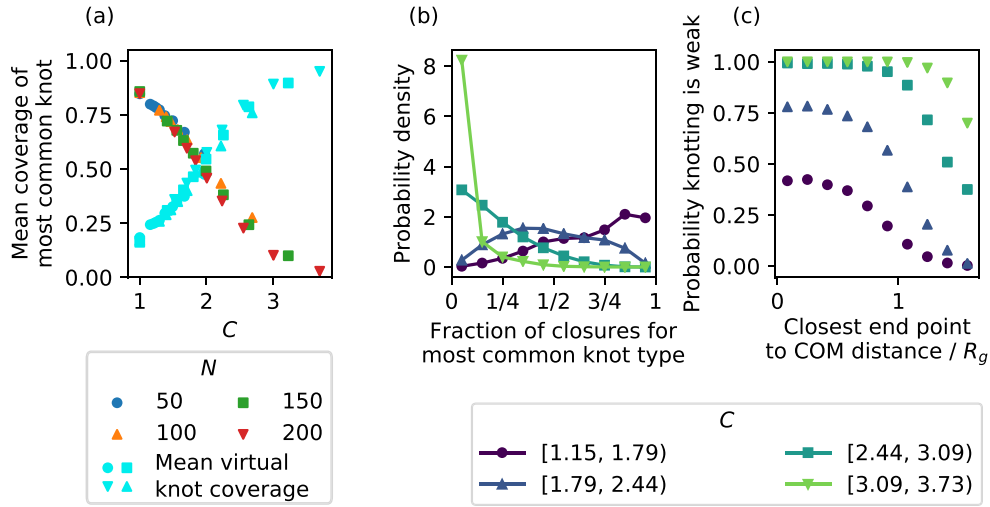
Figure 5. Further results for the virtual closure analysis of spherically confined ideal chains. (a) Mean coverage of the most common knot against the degree of confinement C (see (1)). In cyan is shown the mean coverage of virtual knots for the same walks. (b) The probability distribution of fraction of closures of the commonest knot type. Results for walks have been binned according to C. (c) Probability knotting is weak against the displacement from the centroid of the closer endpoint, divided by radius of gyration. Again, walks are binned by degree of confinement.

decreases to 0, with the distance the weak knotting probability reaches zero being dependent on C. An end need only be buried slightly within a very complex walk to be weak. Of course, the radius of gyration of a strongly spherically confined walk is fixed at the confining radius.

4.2. Knotting in off-lattice walks confined to tubes and slits

We now investigate the effect of different forms of confinement on knotting and weak knotting, varying the number of confined dimensions from three in the sphere, to two in the tube, one in the slit and none when unconfined. The knotting probability against chain length for these different cases is plotted in figure 6(a), for two choices of confining radius, 2 (relatively tight) and 5 (loose). The radius of slits is half the distance between the walls. As before, knotting probability increases with chain length, and as the confining radius decreases. There are fewer knots in tubes compared to spheres of equal radius and fewer yet in slits, as we would expect. Changing the confining radius has less effect on knotting in lower-dimensional confinement: there is little change in slits and less effect in tubes than spheres. We do not investigate sufficiently tight tubes and slits to see the non-monotonic change in knot probability found in [48].

Knotting is much more commonly weak in spheres than tubes and slits, shown in figure 6(b). The growth of weak knotting with length, compared to the previously seen effect in spheres, is modest in tubes and almost non-existent in slits, and the effect of tightening the confining radius in slits is again negligible.

These results should not be surprising since confinement in one and two dimensions, realised by slits and tubes respectively, has a smaller impact than confinement in three dimensions in spheres. To see this directly, we plot the probability of weak knotting in these partially confined geometries against the degree of confinement C (defined in (1)) in figure 6(c). We
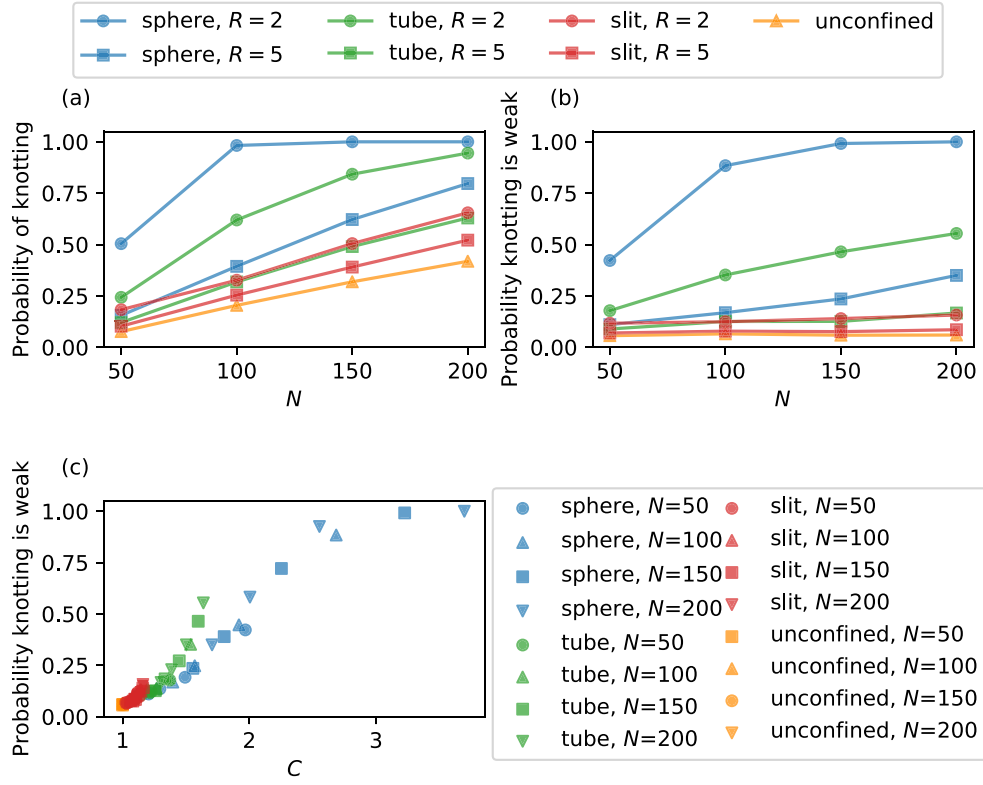
Figure 6. Knotting and weak knotting in random walks confined to spheres, tubes, slits and unconfined. Data for $R = 2$ and $R = 5$ are shown, exemplifying tight and loose confinement. (a) Probability of knotting against chain length. (b) Probability of weak knotting against chain length. (c) Probability of weak knotting against C, the degree of confinement relative to the radius of gyration.

investigate walks in tubes and slits over a much smaller range of degree of confinement than previously for spheres; nevertheless, the data indicates that probability knotting is weak against C is independent of confinement radius and walk length for both tube confinement and slit confinement, with knotting more likely to be weak, for fixed C, for slit confinement than tube, and tube than sphere. This is because partial confinement has a more significant effect on weak knotting than on the radius of gyration with respect to which it is normalised.

The radius of gyration fails to quantify the anisotropy of partially confined walks: compared to the prolate, cigar-like shape expected for unconfined walks, walks in slits tend to be less prolate and walks in tubes more prolate. We therefore consider the rank two, symmetric gyration tensor $\mathbf{S}$ with components S_{ij},

$$S_{ij} = \frac{1}{N} \sum_{n=1}^{N} (r_i^{(n)} - \bar{r}_i)(r_j^{(n)} - \bar{r}_j), \qquad i, j = 1, 2, 3, \tag{2}$$

where $\boldsymbol{r}^{(n)}$ is the nth walk vertex, $n = 1, \ldots, N$, and $\bar{\boldsymbol{r}} = N^{-1} \sum_{n=1}^{N} \boldsymbol{r}^{(n)}$ is the walk centroid. With this terminology, the radius of gyration $R_g = \sqrt{\mathrm{tr}\mathbf{S}}$. A suitable orthogonal transformation $\mathbf{O}$ (i.e. choice of Cartesian coordinate system) diagonalises this tensor,
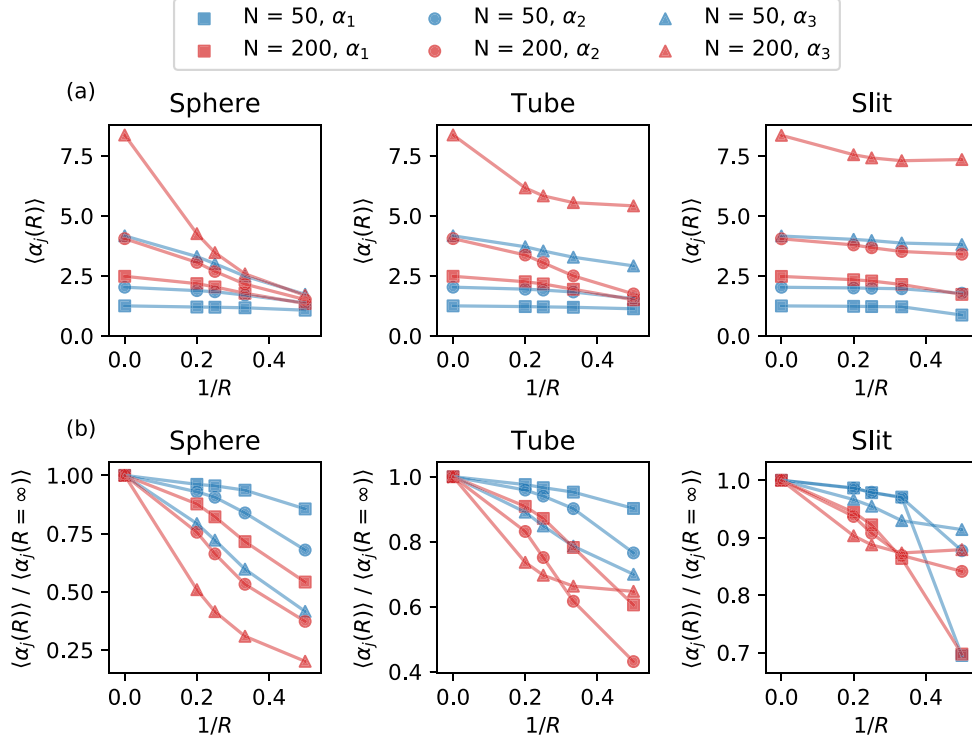
Figure 7. The variation of semi-axes of the inertia ellipsoid of open walks confined to spheres, tubes and slits. (a) The absolute mean size of each semi-axis in confined walks of length $N = 50$ and 200. (b) The mean size of each semi-axis in confined walks, relative to unconfined walks.

$$\mathbf{O S O}^T = \mathrm{diag}[\lambda_1, \lambda_2, \lambda_3], \tag{3}$$

where the axes are ordered so $0 < \lambda_1 \leqslant \lambda_2 \leqslant \lambda_3$, and clearly $R_g^2 = \lambda_1 + \lambda_2 + \lambda_3$. Therefore $\alpha_j = \sqrt{3\lambda_j}$, $j = 1, 2, 3$ are the semi-axis lengths of the associated *ellipsoid of inertia*, indicating the shape of the walk (with notation following [49], where the α_i were used to characterise the shape of knotted ring polymers).

The mean absolute value of each α_j against $1/R$ for each kind of confined walk is plotted in figure 7(a). All walks start from the prolate shape of unconfined walks at $1/R \to 0$. The confinement is isotropic in spheres, so all α_j are distributed like R_g, which for longer walks tends to the confining radius R. In tubes, α_1 and α_2 approach R, while the effect on the largest semi-axis α_3 is more limited. In slits, each semi-axis reduces as $1/R$ increases, but not to a great extent (unsurprising as, for the chains considered, α_1 is often shorter than half the distance between the walls).

We also plot the relative size of the mean confined semi-axes compared to their mean unconfined sizes in figure 7(b). In spheres, as expected, the reduction with $1/R$ is largest for the largest semi-axis α_3. However, in tubes for the longest walks, we find that α_3 is the least relatively reduced, despite being the most reduced in absolute terms. We see this inversion in slits also in both long and short walks.
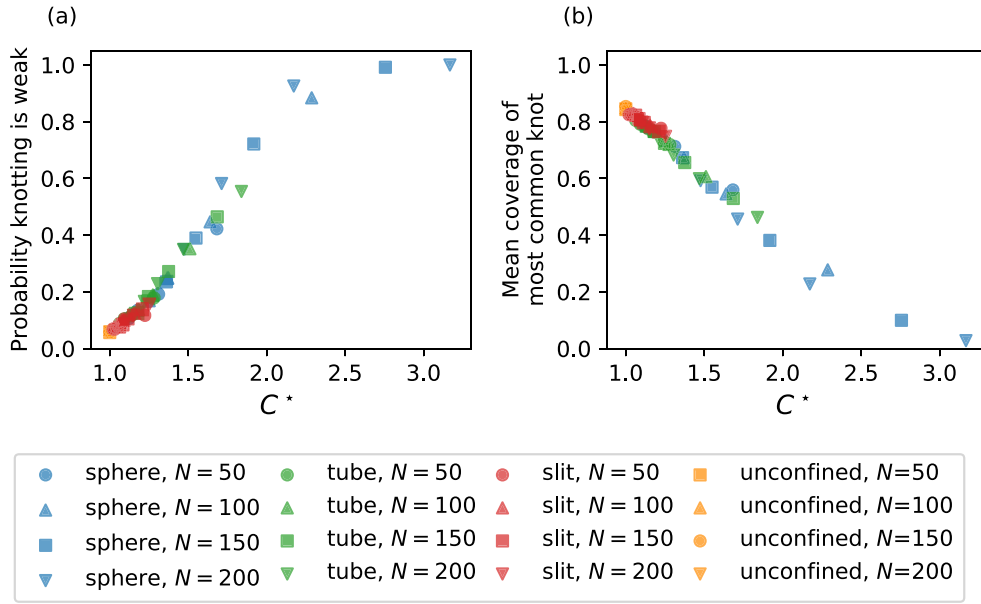
Figure 8. (a) probability that knotting is weak and (b) the mean coverage of the most common knot in ideal chains confined to different geometries, plotted against the adjusted degree of confinement, $C^\star$ (as in (4)).
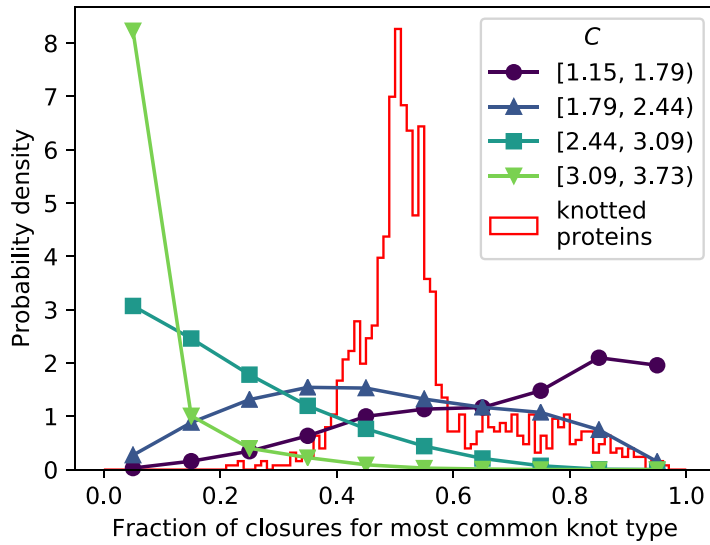

Figure 9. Probability distribution for the coverage of the most common knot on virtual closure for knotted ideal chains (as in figure 5(b)), compared to all knotted proteins in the Protein Data Bank.

This explains why the range of degree of confinement C (defined in (1)) is so much smaller in tubes and slits than in spheres. The radius of gyration is derived from the sum of squares of the semi-axis lengths, and, unconfined, we expect walks shaped such that $\alpha_3 \gg \alpha_1, \alpha_2$, therefore the radius of gyration is dominated by α_3. In spheres, α_3 greatly decreases with radius,

and so C varies considerably. In tubes and slits, α_3 is less affected by the confinement than α_1 and α_2, despite being somewhat reduced. This detail is lost in the radius of gyration, suggesting we may get a better predictor of weak knotting if we put each semi-axis on the same footing. We propose the following *adjusted degree of confinement*, $C^\star$:

$$C^\star = \frac{1}{3} \sum_{i=1}^{3} \frac{\langle \alpha_i(N, R = \infty) \rangle}{\langle \alpha_i(N, R) \rangle}, \tag{4}$$

summing the mean size of each semi-axis relative to those of unconfined walks, and dividing by 3 to ensure $C^\star = 1$ for unconfined walks.

The probability knotting is weak against this adjusted degree of confinement for walks in each confinement shape is plotted in figure 8(a). We see that now all the values fall onto the same curve, regardless of the shape of confinement. Spheres probe the largest range of adjusted degree of confinement, followed by tubes and lastly by slits. Knotting is more likely to be weak for larger $C^\star$. Furthermore, we plot the mean coverage of the most common knot against the adjusted degree of confinement in figure 8(b). We see again that all the values for all walks lie on the same curve. With larger $C^\star$, the mean coverage of the most common knot is lower, i.e. the weaker the knotting. In summary, $C^\star$ is strongly correlated with the weak character of knotting in confined walks, without further dependency on length of walk, or shape or radius of confinement.

5. Discussion

We have performed a detailed numerical investigation into the cases when assigning a knot type to an open tangled curve is ambiguous, which we call 'weak knotting'. While unconfined open random walks can often be assigned a clear knot type, as walks are confined more and more tightly (i.e. within a progressively smaller radius), closures in different directions become assigned to more different knot types, with no one type dominant. We were able to quantify this ambiguity using a knot recognition method which used multiple closures to give a detailed picture of knotting. Identifying projections using virtual knots is more sensitive to knotting, and to the ambiguity of knotting over previous sphere closure schemes. Lattice and off-lattice walks were investigated, and while each showed different overall knotting, with lattice walks less likely to be knotted, the ambiguity in their knotting under confinement was very similar when measured using virtual closure.

In particular, we found that the relative radius of gyration of unconfined walks to confined walks is well correlated with the probability of weak knotting, independent of chain length. We also found that weak knotting is strongly correlated to walk endpoints being close to the walk centroid, i.e. within the bulk of the tangle. In other words, the knotting spectrum of open random walks grows more complex with confinement, and when their endpoints are more buried in the tangle. By varying the confining geometry, we found that a better predictor of weak knotting is using the relative reduction of each semi-axis of the characteristic ellipsoid of unconfined and confined walks.

This analysis shows that the distinction between weak and strong knotting is important when open chains are randomly tangled and confined within volumes smaller than the unconfined radius of gyration. The methods introduced here should be useful in more extensive statistical analyses of the shapes of open macromolecule chains. Examples where this may prove important include systems of polymers which are compact through poor solvent conditions, or external confinement due to microfluidic environments or viral capsids.

The original motivation of our initial investigation of virtual closure analysis of open chains was to study knotted proteins [33], which are examples of macromolecules which are compact and not completely flexible. Our results revealed that, while proteins in the Protein Data Bank are rarely knotted, almost 40% are weakly knotted. The combination of a low knotting probability and a relatively high proportion of weak knotting is unlike any of the random walk models investigated here. Figure 9 provides more detail, showing a comparison between the distribution of most common knot coverage in knotted proteins versus knotted, spherically confined, off-lattice walks. Evidently, the protein case is very different from these walks, indicating the subtlety in modelling protein conformations as random chains. The distribution is relatively sharp, with a peak in most common knot coverage only slightly above 50%. If a method even more sensitive to weak knotting were used such as knotoids considered on the plane [37], now implemented by the knotted proteins database KnotProt [50], it would not be surprising to find most knotted proteins classified as weakly knotted. Much of the shape of this distribution can be attributed to the large number of knotted carbonic anhydrase structures which dominate the data set and are notable for the shallowness of their knotting. As it is highly likely that a shallowly knotted chain is also weakly knotted, this accounts in part for the large number of protein chains with low fractional coverage for their commonest knot.

The distinction between weak and strong knotting is likely to be useful in the analysis of open chain dynamics. For instance, during the process of knot formation, a flexible open chain passes through one or more weakly knotted conformations before reaching a final deeply knotted state. Similar transitions should be observable in unknotting pathways. In these situations, the weakly knotted state is likely to be unstable, quickly untying or securing a knot with some perturbation to the endpoints. In controlling chain dynamics, confinements or effective sizes might be manipulated to enhance the chance of a weakly knotted configuration, which might then make a further transition to a strongly knotted configuration. Further study of the weak-strong transitions may reveal further structure in these distinctions of knot types for open curves. An important extension to our work in this regard would be to investigate walks in tighter tubes and slits. In such extremely confined environments, the geometrical complexity of the walks is likely to be reduced, resulting in a falling proportion of weak knots which we predict to be captured by a decreasing adjusted degree of confinement (defined in (4)) as confining radius is decreased. This is in line with the non-monotonic knotting behaviour previously seen [48]. This work cannot be fully carried out using the algorithms for generating walks described here, however, as these require a radius of at least 1.

Acknowledgments

Appendix A. The virtual knot $v4_{12}$ is the composite of two virtual trefoils $v2_1$

We here briefly describe a sequence of classical and virtual Reidemeister moves, depicted in figure A1, demonstrating that the 4-crossing virtual knot $v4_{12}$ from the tabulation in [35] is not a prime virtual knot (in the sense of [45]), but in fact is the composite of two virtual trefoil knots $v2_1$. We indicated that this knot was prime in [33], although it does not appear in the classification of such virtual knots in [45].
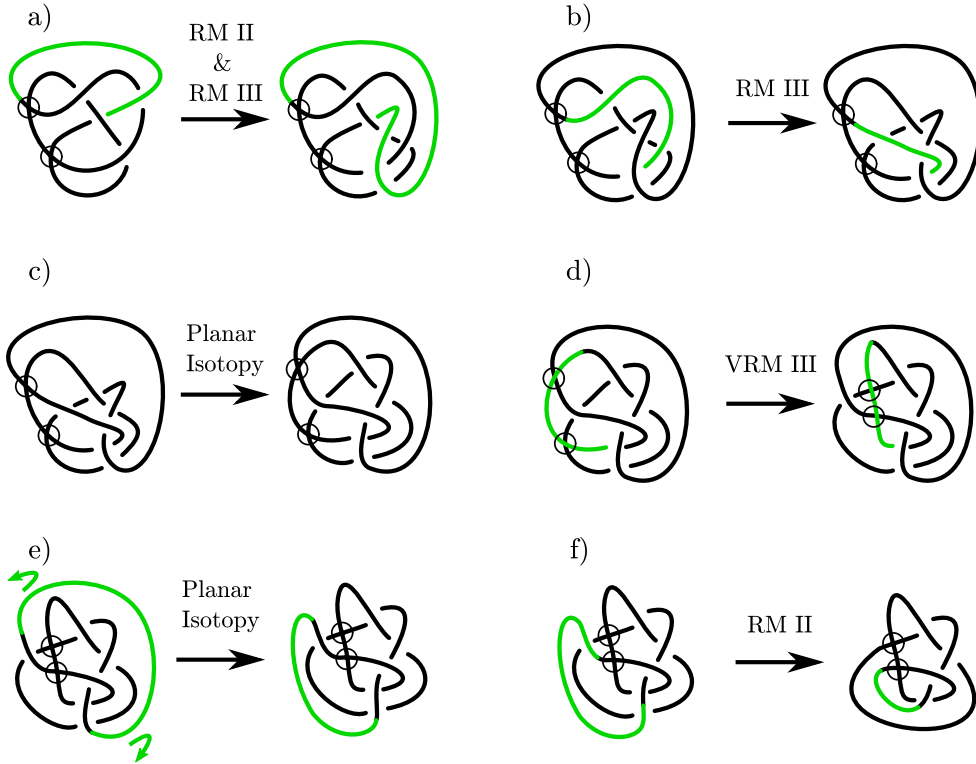
Figure A1. Sequence of classical Reidemeister moves (indicated by RM II, RM III) and virtual Reidemeister moves (indicated VRM II, VRM III) and isotopies, demonstrating how $v4_{12}$, from the form given in [35], is a composite of two copies of $v2_1$, as discussed in more detail in the text.

Beginning with a diagram of $v4_{12}$ isotopic to the form in [35], a sequence of classical Reidemeister moves leads to the diagram on the right of figure A1(c), with six classical crossings and two virtual crossings, which is followed by a type III virtual Reidemeister move giving the form in figure A1(d). The indicated outside arc is now deformed through 'the point at infinity' (i.e. the diagram is properly considered as drawn on the sphere, and the arc is passed around the back of the sphere with respect to the diagram). A Reidemeister move of type II gives a form which is clearly composite, as required.

Appendix B. Generating ideal chains in spheres, tubes and slits

To generate ideal chains confined to the sphere, we follow the algorithm of Diao *et al* [47] to ensure that walk vertices are uniformly distributed throughout the volume. Here we summarise this model, with R representing the radius of the confining sphere centred at the origin, and X the previous endpoint of the walk whose next step is being considered, a distance $r = |X|$ from the centre of the sphere. If X is further than a unit from the spherical boundary, the next point is chosen by stepping a unit in a uniformly random direction. If X is within a unit of the boundary, the next step's direction is sampled from a distribution weighted towards the boundary: this depends on the choice of an azimuthal angle ϕ and polar angle θ with respect to an axis given by the direction of X with respect to the centre of the sphere. As the problem

is axisymmetric, ϕ is uniformly distributed between 0 and 2π. The value of $-\cos\theta$ is sampled differently in each of the following three ranges: angles which, on a unit sphere about X, lie further than a unit from the boundary, angles within a unit of the boundary, and outside the boundary. The values of $-\cos\theta$ which lie within these categories are:

$$-\cos\theta = \begin{cases} \text{further than a unit from the boundary,} & -1 \leqslant -\cos\theta \leqslant a(r,R); \\ \text{within a unit of the boundary,} & a(r,R) < -\cos\theta \leqslant b(r,R); \\ \text{outside the boundary,} & b(r,R) < -\cos\theta, \end{cases}$$
(B.1)

where

$$a(r,R) = \begin{cases} \frac{1}{2r}(R^2 - r^2 - 2R), & r > R - 1 \text{ and } R - 1 > \min(r, |r-1|); \\ -1, & r > R - 1 \text{ and } R - 1 \leqslant \min(r, |r-1|); \\ 1, & r \leqslant R - 1, \end{cases}$$
(B.2)

marking the angle which would step a unit from the boundary, and

$$b(r,R) = \begin{cases} \frac{1}{2r}(R^2 - r^2 - 1), & r > R - 1; \\ 1, & r \leqslant R - 1, \end{cases}$$
(B.3)

marking the angle which would step to the boundary.

The probability density function (PDF) proposed by [47] from which $-\cos\theta$ is sampled is uniform for angles further than a unit from the boundary, rises linearly within a unit of the boundary, and is zero outside of the boundary. Explicitly, this is

$$\text{PDF}(-\cos\theta) = \begin{cases} \frac{1}{2}, & -1 \leqslant -\cos\theta \leqslant a(r,R); \\ \frac{1}{2}(1 + c[-\cos\theta - a(r,R)]), & a(r,R) < -\cos\theta \leqslant b(r,R); \\ 0, & b(r,R) < -\cos\theta \end{cases}$$
(B.4)

where

$$c = \frac{4r((r+1)^2 - R^2)}{(2R-1)^2}$$
(B.5)

for normalisation. The radial distribution of vertices for walks generated in this way is shown in figure 3(a), compared with both the desired uniform distribution and vertices for walks in absorbing boundaries.

We adapt this method to find distributions of vertices of chains confined in other geometries. The case of confinement in slits, or between two parallel walls, is relatively straightforward by approximating each wall as a sphere with large radius, in practice 10 000. The same PDF gives almost uniformly distributed vertices in the slit.

Confinement in a tube is more challenging since the problem is no longer axisymmetric in the same way as the sphere. The following modification of the model of [47] gives a distribution of vertices close to uniform, although not quite as good as the spherical case. Align the tube along the z-axis, and take ϕ now as the polar angle with respect to this direction, with $\cos\phi$ chosen uniformly between 0 and π. θ is now the azimuthal angle, and it again determines how close the step is to the boundary, which also now depends on ϕ. Once ϕ is chosen, the next step can only lie on a section of a circle situated within the tube. Unlike with spherical confinement, the choice of ϕ alters the radius of this circle, from 1 at $\phi = \pi/2$ to 0 at $\phi = 0$ or π. To account for this when sampling θ, we scale the problem to give an effective r of

$$r_{\text{eff}} = r_X / \sin \phi, \tag{B.6}$$

where r_X is the actual radial distance of X, and an effective R of

$$R_{\text{eff}} = R_{\text{tube}} / \sin \phi, \tag{B.7}$$

where R_{tube} is the actual radius of the tube.

The tube-confined θ is sampled with respect to cylindrical radius between 0 and π, clockwise or anticlockwise at random. The PDF (B.4), removing the cosine, overcompensates the boundary-avoiding behaviour. In order to limit this rise in the region within a unit of the boundary to be slower than linear, we used the distribution

$$\text{PDF}(\theta) = \begin{cases} \frac{1}{\pi}, & -\pi \leqslant \theta \leqslant \alpha; \\ \frac{1}{\pi}(1 + c(\theta - \alpha)^n), & \alpha < \theta \leqslant \beta; \\ 0, & \beta < \theta \end{cases} \tag{B.8}$$

where the angles $\alpha = \arccos(-a)$ and $\beta = \arccos(-b)$ are equivalent to the boundaries used in the case of the sphere case, and

$$c = -\frac{\beta(n + 1)}{(\beta - \alpha)^{n+1}} \tag{B.9}$$

for normalisation. The resulting radial distribution of walk vertices using this corrected PDF is plotted in figure 3(b) together with the distribution using an absorbing boundary, and the original spherical PDF. This used a value of $n = 0.6$, which we found gave the closet fit to uniform. Evidently, the fit is an improvement, although not ideal. The small deviation is unlikely to have a large effect on the knotting statistics.

ORCID iDs

Alexander J Taylor ⓘ https://orcid.org/0000-0003-3986-7114
Mark R Dennis ⓘ https://orcid.org/0000-0003-1147-1804

References

[1] Flory P J 1953 *Principles of Polymer Chemistry* (Ithaca, NY: Cornell University Press)
[2] Rubinstein M and Colby R H 2003 *Polymer Physics* (Oxford: Oxford University Press)
[3] Tezuka Y and Oike H 2002 Topological polymer chemistry *Prog. Polym. Sci.* **27** 1069–122
[4] Orlandini E and Whittington S G 2007 Statistical topology of closed curves: some applications in polymer physics *Rev. Mod. Phys.* **79** 611–42
[5] Sumners D W and Whittington S G 1988 Knots in self-avoiding walks *J. Phys. A: Math. Gen.* **21** 1689–94
[6] Pippenger N 1989 Knots in random walks *Discrete Appl. Math.* **25** 273–8
[7] Millett K C 1994 Knotting of regular polygons in 3-space *JKTR* **3** 263–78
[8] Orlandini E 2017 Statics and dynamics of DNA knotting *J. Phys. A: Math. Theor.* **51** 053001
[9] Orlandini E 2017 Statistical topology and knotting of fluctuating filaments *Physica* A **504** 155–75
[10] Van Rensburg E J J and Whittington S G 1990 The knot probability in lattice polygons *J. Phys. A: Math. Gen.* **23** 3573–90
[11] Tesi M C, van Rensburg E J J, Orlandini E and Whittington S G 1994 Knot probability for lattice polygons in confined geometries *J. Phys. A: Math. Gen.* **27** 347–60
[12] Arsuaga J, Vázquez M, Trigueros S, Sumners D W and Roca J 2002 Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids *Proc. Natl Acad. Sci. USA* **99** 5373–7

[13] Micheletti C, Marenduzzo D, Orlandini E and Sumners D W 2006 Knotting of random ring polymers in confined spaces *J. Chem. Phys.* **124** 124–33

[14] Micheletti C, Marenduzzo D, Orlandini E and Sumners D W 2008 Simulations of knotting in confined circular DNA *Biophys. J.* **95** 3591–9

[15] Tubiana L, Orlandini E and Micheletti C 2011 Multiscale entanglement in ring polymers under spherical confinement *Phys. Rev. Lett.* **107** 188302

[16] Poier P, Likos C N and Matthews R 2014 Influence of rigidity and knot complexity on the knotting of confined polymers *Macromolecules* **47** 3394–400

[17] Diao Y, Ernst C, Montemayor A and Ziegler U 2014 The knot spectrum of confined random equilateral polygons *Comput. Math. Biophys.* **2** 19–33

[18] Marenduzzo D, Micheletti C, Orlandini E and Sumners D W 2013 Topological friction strongly affects viral DNA ejection *Proc. Natl Acad. Sci. USA* **110** 20081–6

[19] Whitesides G M 2006 The origins and the future of microfluidics *Nature* **442** 368–73

[20] Theberge A B, Courtois F, Schaerli Y, Fischlechner M, Abell C, Hollfelder F and Huck W T S 2010 Microdroplets in microfluidics: an evolving platform for discoveries in chemistry and biology *Angew. Chem., Int. Ed. Engl.* **49** 5846–68

[21] Sackmann E K, Fulton A L and Beebe D J 2014 The present and future role of microfluidics in biomedical research *Nature* **507** 181–9

[22] Grosberg A Y, Feigel A and Rabin Y 1996 Flory-type theory of a knotted ring polymer *Phys. Rev. E* **54** 6618–22

[23] Marcone B, Orlandini E, Stella A L and Zonta F 2007 Size of knots in ring polymers *Phys. Rev. E* **75** 041105

[24] Mansfield M L and Douglas J F 2010 Properties of knotted ring polymers. I. Equilibrium dimensions *J. Chem. Phys.* **133** 044903

[25] Dai L, van der Maarel J R C and Doyle P S 2012 Effect of nanoslit confinement on the knotting probability of circular DNA *ACS Macro Lett.* **1** 732–6

[26] Micheletti C and Orlandini E 2012 Knotting and metric scaling properties of DNA confined in nano-channels: a Monte Carlo study *Soft Matter* **8** 10959–68

[27] Micheletti C and Orlandini E 2012 Numerical study of linear and circular model DNA chains confined in a slit: metric and topological properties *Macromolecules* **45** 2113–21

[28] Tubiana L, Orlandini E and Micheletti C 2011 Probing the entanglement and locating knots in ring polymers: a comparative study of different arc closure schemes *Prog. Theor. Phys. Suppl.* **191** 192–204

[29] Millett K, Dobay A and Stasiak A 2005 Linear random knots and their scaling behavior *Macromolecules* **38** 601–6

[30] Millett K C and Sheldon B M 2005 Tying down open knots: a statistical method for identifying open knots with applications to proteins *Physical and Numerical Models in Knot Theory: Including Applications to the Life Sciences* (Singapore: World Scientific) pp 203–17

[31] Millett K C, Rawdon E J, Stasiak A and Sulkowska J I 2013 Identifying knots in proteins *Biochem. Soc. Trans.* **41** 533–7

[32] Taylor W R 2000 A deeply knotted protein structure and how it might fold *Nature* **406** 916–9

[33] Alexander K, Taylor A J and Dennis M R 2017 Proteins analysed as virtual knots *Sci. Rep.* **7** 42300

[34] Kauffman L H 1999 Virtual knot theory *Eur. J. Comb.* **20** 663–90

[35] Green J and Bar-Natan D 2004 A table of virtual knots (www.math.toronto.edu/drorbn/Students/GreenJ/) accessed 16 Apr 2018

[36] Goundaroulis D, Dorier J, Benedetti F and Stasiak A 2017 Studies of global and local entanglements of individual protein chains using the concept of knotoids *Sci. Rep.* **7** 6309

[37] Goundaroulis D, Gügümcü N, Lambropoulou S, Dorier J, Stasiak A and Kauffman L 2017 Topological models for open-knotted protein chains using the concepts of knotoids and bonded knotoids *Polymers* **9** 444

[38] Bao X R, Lee H J and Quake S R 2003 Behavior of complex knots in single DNA molecules *Phys. Rev. Lett.* **91** 265506

[39] Rakhmanov E A, Saff E B and Zhou Y M 1994 Minimal discrete energy on the sphere *Math. Res. Lett.* **1** 647–62

[40] Alexander J W 1928 Topological invariants of knots and links *Trans. AMS* **30** 275–306

[41] Kauffman L H and Radford D E 2003 Bioriented quantum algebras and a generalized Alexander polynomial for virtual links *Contemp. Math.* **318** 113–40

J. Phys. A: Math. Theor. **53** (2020) 045001

K Alexander *et al*

[42] Sawollek J 1999 On Alexander–Conway polynomials for virtual knots and links (arXiv:math/9912173v2)

[43] Jones V F R 1985 A polynomial invariant for knots and links via Von Neumann algebras *Bull. AMS* **12** 103–11

[44] Rolfsen D D 1976 *Knots and Links* (Providence, RI: AMS Chelsea Publishing)

[45] Akimova A A and Matveev S V 2014 Classification of genus 1 virtual knots having at most five classical crossings *J. Knot Theor. Ramifications* **23** 1450031

[46] Lua R, Borovinskiy A L and Grosberg A Y 2004 Fractal and statistical properties of large compact polymers: a computational study *Polymer* **45** 717–31

[47] Diao Y, Ernst C, Montemayor A and Ziegler U 2012 Generating equilateral random polygons in confinement III *J. Phys. A: Math. Theor.* **45** 465003

[48] Orlandini E and Micheletti C 2013 Knotting of linear DNA in nano-slits and nano-channels: a numerical study *J. Biophys.* **39** 267–75

[49] Rawdon E J, Kern J C, Piatek M, Plunkett P, Stasiak A and Millett K C 2008 Effect of knotting on the shape of polymers *Macromolecules* **41** 8281–7

[50] Jamroz M, Niemyska W, Rawdon E J, Stasiak A, Millett K C, Sulkowski P and Sulkowska J L 2014 Knotprot: a database of proteins with knots and slipknots *Nucleic Acids Res.* **43** D306–14