Random Knotting: Theorems, Simulations and Applications

De Witt Sumners[1]

Department of Mathematics, Florida State University, Tallahassee, FL 32303-4510, USA summers@math.fsu.edu

1 Introduction

A fundamental problem in science is the description and quantification of topological entanglement complexity of filaments — embedded arcs, circles and graphs in 3-space. This problem is the province of the mathematical subject of *knot theory*, and it has a fascinating history going back to the work of Gauss and Maxwell [Sv]. In more recent times, physicists and mathematicians (stimulated by the behavior of real filament systems), have become interested in measuring entanglement of random systems and entanglement changes with length and/or density (length per unit volume), and in understanding the physical ramifications of entanglement. Random entanglement increases with filament length and/or density; physical intuition is absolutely clear on this point. As (almost) everyone has experienced, a 100 foot electrical extension cord (when carelessly bunched up and put in a corner of the garage by your teenager) is difficult to untangle when you get ready to use it — when you find the ends of the cord and pull them apart, the cord is invariably highly entangled, and has at least one knot in it. In order to use the extension cord, you must snake a cord end through the tangle to resolve it. By comparison, a 25 foot extension cord (stored in the same careless fashion) untangles much more easily — you need only to find the ends and pull them apart, often outstretching the cord with little or no entanglement. Can one quantify observations like this, and/or prove theorems about entanglement? The answer is YES, and this article will describe some of the theoretical and simulation results on random entanglement, and give a few scientific applications. This article is by no means exhaustive; I will only describe some of the work I am most familiar with. I will, for example, prove that, on the simple cubic lattice Z^3, the probability that a randomly chosen n-edge polygon in Z^3 is knotted goes to one exponentially rapidly with length n; in other words, all but exponentially few polygons of length n in Z^3 are knotted. I once had the experience of presenting this proof (Murphy's Law of entanglement) at a research conference; one of the questions I was asked after my talk (by a non-

mathematician!) was "Why does one need to prove that the longer a random polygon, the more likely it is to be knotted? This phenomenon is obvious, and needs no proof!" Applying that same line of reasoning, the Jordan-Schoenflies Curve Theorem (which says that every circle in the XY-plane bounds a compact region (the inside) homeomorphic to the 2-disk) is also obvious and needs no proof. As every mathematician knows, not every "obvious" statement is true, and a proof is the intellectual laboratory where mathematical truth can be verified. An interesting example of something which is "obvious" but in fact false is the 3-dimensional analogue of the Jordan-Schoenflies Theorem — that every 2-sphere in 3-space bounds a 3-disk. This was thought to be true by mathematicians, and Alexander even claimed a proof, but took it back when he found a mistake in his proof, and later a counterexample — the infamous Alexander Horned Sphere [Ax1]. The horned sphere is very badly embedded in 3-space — it has a Cantor set of points where the embedding is "wild". By restricting to finite polyhedral 2-spheres (which cannot have any "wild" points), Alexander was able to prove the 3-dimensional version of the Jordan-Schonflies Theorem [Ax2].

What are the effects of entanglement in real physical systems? Leaving aside your frustration at resolving the entanglement in the spaghetti-like mass of computer wires under your desk, unresolved entanglement of DNA in cells is a death sentence for that cell [BZ, LZC]. Most drugs for the treatment of bacterial infections or cancer work by inhibiting cellular enzymes, which resolve molecular entanglement in the cell, and the target cell (a pathogen or cancer cell) dies as a result. In polymer science, macroscopic properties of polymer systems often depend on microscopic intermolecular entanglement [RBHS]; entanglement determines whether or not the polymer system is a gel or a polymer fluid, and if a solid, the entanglement has consequences in the strength of the material. In fluid dynamics, plasma and superfliud physics [Mof,Ric1,Ric2], entanglement of magnetic and vortex filaments have important consequences for the energy of the system.

2 The Frisch-Wasserman-Delbruck Conjecture

The specific problem of occurrence of knots in a long linear polymer chain was first addressed independently by Frisch and Wasserman in 1961 [FW,Was], and by Delbruck in 1962 [Del]. Both groups formulated questions about the probability that a closed circular polymer chain with degree of polymerization n (n monomers) would contain a knot. More specifically, if one performs a cyclization reaction (random closure) on a dilute solution containing linear polymers with polymerization degree n, what would be the yield of this reaction? Neglecting dimers, trimers, etc., and focusing only on the circular reaction products with n monomers, what is the product spectrum (histogram of knot types)? Frisch and Wasserman and Delbruck conjectured (no surprise here!):

Conjecture 1 (Frisch-Wasserman Delbruck (FWD) Conjecture). The probability that a randomly embedded circle of length n in R^3 is knotted tends to one as n tends to infinity.

Numerical evidence pointing to the truth of the FWD conjecture abounds: The first Monte Carlo simulation of knotting in random polygons was done by Vologodskii et al [VLF]. They generated a random sample of polygons of length n, and used $\Delta(-1)$, the *order of the knot* (the Alexander polynomial evaluated at $t = -1$) to detect knotting. One of the models they investigated was a random walk model on the simple cubic lattice Z^3. One starts at the origin, and performs a random walk; when a self-intersection of the walk is encountered, perturb the entire lattice a small amount in a random direction to remove the intersection and keep walking. As one continues on the walk, bias the walk to return to the origin after n steps. The found that the knot probability rose with increasing length n, and obtained quite high knot probabilities (60% for $n = 300$). Other groups [CM,MW] performed similar studies 20-30 years ago, with similar results — the knot probability grows with n, tending toward unity as n tends to infinity.

In 1986 I gave a talk at a Canadian Chemical Society meeting in Saskatoon where I presented the FWD conjecture. Stu Whittingon was in the audience, and after my talk came up and asked "Have you heard of the Kesten Pattern Theorem?" At that point Stu and I began to work on the FWD conjecture. The model we chose to use was self-avoiding walks (SAW) and self-avoiding polygons (SAP) on Z^3. The simple cubic lattice is useful for describing excluded volume effects in polymers in dilute solution, and allows one to do both rigorous asymptotic proofs as n goes to infinity, and numerical simulations for small values of n, allowing us to complete a proof of the FWD conjecture [SW].

On the simple cubic lattice Z^3, a step is a directed edge joining two adjacent lattice points. An *n-step self-avoiding walk* (n-SAW) beginning at lattice point x_0 is an $(n+1)$-tuple of distinct lattice points $\{x_1, \ldots, x_n\}$ where x_i and x_{i+1} are adjacent in the lattice for $0 \leq i < n$. An n-step self-avoiding polygon (n-SAP) is an n-SAW whose first and last vertices are adjacent in the lattice. An n-SAW (n-SAP) is *rooted* if $x_0 = 0$. Since there are finitely many rooted n-SAWs (n-SAPs) for each n, then the probability that a randomly chosen n-SAP is knotted is simply the ratio of the number of rooted, knotted n-SAP divided by the number of rooted n-SAP.

The key idea in the proof of the FWD conjecture is the Kesten Pattern Theorem [Kes]. A Kesten pattern X is a SAW which has a way in and a way out, a walk that can be concatenated so that it can appear many times in a long SAW. For an example of a SAW in Z^3 which is not a Kesten pattern, consider a "crab trap", a lattice cube in which the boundary sphere of the cube is saturated by the walk, which then enters the interior of the cube. Once inside the cube, the walk cannot exit the cube because all the boundary vertices are already occupied, and the walk must terminate inside the cube. Kesten proved

that given any Kesten Pattern X there is a positive density $D_X > 0$ associated to this pattern such that, for sufficiently large n, X appears (up to translation) at least $\lfloor D_X n \rfloor$ times in all but exponentially few self-avoiding walks of length n. The first step was to extended Kestens result to cover patterns in SAP as well as SAW on Z^3 [SW].

So a Kesten pattern appears at least once in all but exponentially few sufficiently long SAW and SAP. In order to prove the FWD on Z^3, one needs to produce a Kesten pattern T such that if T appears in a given SAP, then that SAP is guaranteed to be knotted. A *tight knot* is such a pattern, for example the one specified by the SAW T given below. Suppose that we have a right-handed coordinate system in Z^3, and let i, j, k be the unit vectors in the X, Y, Z directions, respectively. Beginning at the origin in Z^3, take the following 18-step walk:

$$T : \{j, j, -i, k, k, i, i, -k, -j, -k, -k, -i, -i, k, k, i, j, j\}.$$

How can we be guaranteed that if the SAW T appears in a SAP, then that SAP is knotted? Ordinarily local patterns such as T in a long circle to not guarantee that the circle is knotted, because the circle can snake back through the local entanglement T and undo the local knot (just as one resolves the entanglement in a long extension cord). However, the self-avoiding condition prevents any such snaking back through the entanglement, and any SAP that contains the pattern T must be knotted. More precisely, each occupied vertex in a SAW (SAP) sits in the middle (barycenter) of a dual 3-cube, and one can think of this dual 3-cube as the excluded volume generated by that occupied lattice site. Let T be the 16-step sub walk of T obtained by deleting the first and last steps. Let $N(T')$ denote the *lattice neighborhood* of T' — the union of the 16 dual 3-cubes which surround the vertices of T'. $N(T')$ is a 3-ball, and T enters and exits $N(T')$ transversely in its first and last steps. Suppose now that K is any SAP that contains T. The 2-sphere boundary of $N(T')$ separates K into the connected sum of two knots, one of which is the trefoil formed by the intersection of K and $N(T')$. In order to prove that K is knotted, we compute the *genus* of K. Every knot K is spanned by many orientable surfaces (called *Seifert surfaces*). If one takes the minimum genus over all the Seifert surfaces spanning the knot, one obtains the *genus* of the knot $g(K)$. K is unknotted if and only if K spans a 2-disk, so K is unknotted if and only if $g(K) = 0$, and K is knotted if and only if $g(K) \geq 1$. Since the knot genus is additive on connected sums [Adm], and the trefoil is a summand of K of genus one, then the genus of K is at least one, hence K is knotted. A similar proof of the FWD conjecture (also based on Kesten's pattern theorem) was found independently by Nick Pippenger [Pip].

The number p_n of rooted SAP of length n in Z^3 behaves as [RSW]

$$p_n = e^{\kappa n + o(n)}$$

and the number of unknotted polygons p_n^0 behaves as

$$p_n^0 = e^{\lambda n + o(n)}$$

with $0 < \lambda < \kappa$ so that the knot probability $P(n)$ behaves as

$$P(n) = 1 - e^{-\alpha n + o(n)}$$

for some positive constant α.

Theorem 1 (SW,Pip). *The probability that an n-SAP in Z^3 is knotted goes to one exponentially rapidly as n tends to infinity.*

Although we have proved that there exists a positive constant that describes the knotting probability $P(n)$, no rigorous analytic method for computing is known, and the value of must be determined by simulation. For $n < 24$, there are no knotted SAPs, and for $n = 24$, there are exactly 3496 knotted 24-SAPs (all trefoils) [Do1], out of a total of something of the order of 10^{13} 24-SAPs, so $P(n) = 0$ for $n < 24$ and $P(24)$ is positive but vanishingly small. Simulation results on Z^3 produce something on the order of 1% knots for $n = 1000$ [RW]. Knot-type specific estimates of knot probability parameters have also been made [DT1].

We have now proved that all but exponentially few sufficiently long SAP contain a tight trefoil. What about other knots? In fact, the above argument can be used to show that, as n tends to infinity, every knot eventually appears as a summand — that is, if K^* is a fixed knot type, then all but exponentially few sufficiently long SAP contain a copy of K^* as a summand [SSW]. Let K^* denote a knot type in S^3, and let K^* be a polygonal representative of K^* in S^3. Insert a new vertex in the interior of an edge of K^*, and regard this new vertex as the point at infinity in S^3. By removing this point, we get a knotted polygonal arc in R^3 with the property that the ends of the knotted arc go off to infinity along the X-axis. Take a regular projection of this arc on the XY ($z = 0$) plane; by forgetting the crossover information (over-under at each crossing), and using isotopy in the XY plane and subdivision when necessary, we obtain an immersion of the arc in the square lattice XY plane, and the ends go off to infinity along the X axis. The non-trivial part of the immersed arc (all of the immersion except for parts of the ends which go off to infinity) is contained in a square in the XY plane with vertices $(\pm d, \pm d, 0)$ for some even integer d. This square intersects the immersion in two points where the straight ends of the immersed arc go off to infinity. We can now recover the knotted arc by remembering the crossing information. Replace each two lattice steps which represent an underpass by a path which detours one step into the $z = -1$ plane, continues two steps as the underpass in the $z = -1$ plane, and then returns in one step to the $z = 0$ plane. We now have an infinite SAW contained in two parallel planes, $z = 0$ and $z = -1$. This arc exits the square with vertices $(\pm d, \pm d)$ at two points: $(-d, 0, 0)$ and $(d, 0, 0)$. By removing the infinite ends (remove all vertices $(x, 0, 0)$ with $x \leq -d$ and $x \geq d$), one obtains the finite SAW α where α has captured the knot type

K^*. The lattice neighborhood $N(\alpha)$ is homeomorphic to a punctured 2-disk cross an interval — a piece of Swiss cheese. We need to fill in the holes by extending α to α^* such that α^* still represents the knot K^* and $N(\alpha^*)$ is a 3-ball. We do this by draping the ends of α over the entire disk, analogous to pouring syrup on a pancake, filling up all the holes. Start at the vertex $(d, 0, 0)$, the right-hand endpoint of α. Add the vertex $(d, 0, 1)$, then traverse upward through $(d, 2, 1)$, $(d, 3, 1), \ldots, (d, d, 1)$. Proceed left to vertex $(d-1, d, 1)$, then traverse down through $(d-1, d-1, 1), \ldots, (d-1, -d, 1)$; then left to vertex $(d-2, -d, 1)$, then up through $(d-2, -d+1, 1)$, etc. Proceed to zigzag up and down until vertex $(1, d, 1)$ is reached (on an upward traversal), then add an endpoint vertex $(1, d+1, 1)$. These steps which have been added to α drape the SAW over the right-hand half of the enclosing square, and the knot type has been preserved because the added steps can be vertically pulled up and off of the "knotted" part of the arc. Perform the analogous procedure on the left-hand half of the square, starting with the vertex $(-d, 0, 0)$, moving one step up to the $z = 1$ plane, then proceeding up the left-hand edge of the square until the top is reached, then moving one step to the right, going straight down until the bottom of the square is reached, then up, etc. On the last downward traversal, the vertex $(0, -d, 1)$ is reached, and one then adds an endpoint $(0, d, -1)$. The SAW α^* so produced is the desired Kesten pattern representing the knot type K^*.

We have shown the following:

Theorem 2 (SSW). *Let K denote a given knot type; then there exists a positive density D_K such that, for sufficiently large n, the knot K appears at least $\lfloor nD_K \rfloor$ times as a summand of all but exponentially few SAPs of length n on Z^3.*

There are two other models in which the FWD conjecture has been proven. One model is the Gaussian Random Polygon (GRP) model. An n-GRP is a piecewise linear circle in R^3 with n edges in which the edge lengths form a Gaussian distribution. In [DPS] a continuum version of the Kesten pattern theorem was proved, and then used to prove that the probability that a randomly chosen n-GRP is knotted tends to one exponentially rapidly as n tends to infinity. The other is the Equilateral Polygon (EP) model. An n-EP is an equilateral polygon in R^3 with n edges. In [Do2] Diao proved the FWD conjecture for equilateral polygons. All of the proofs of the FWD discussed to this point have relied on local (tight) knots to force knotting of sufficiently long polygons. Can it be shown that almost all sufficiently long random polygons have global (non-local) knots in them? The answer is YES. Jungreis [Jun] proved global knotting in the GRP model; with high probability a long randomly chosen polygon is a satellite knot (it is an essential loop in a knotted solid torus), so the knot cannot be unknotted by small perturbations which could kill local knots). Diao *et al.* [DNS] also proved global knotting in the EP model. As far as I know, global knotting for SAP on Z^3 is almost certainly true but has not yet been proved.

The codimension two phenomenon of knotting occurs in all dimensions; p-spheres can be knotted in $(p+2)$ space [Rol] for all $p \geq 1$. Consider rooted p-spheres in Z^{p+2}. Consider p-spheres in Z^{p+2} that are the union of *unit* p-cells, where a *unit* p-cell in Z^{p+2} is one spanned by 2^p vertices that are adjacent in Z^{p+2}. Let $n - S^p$ denote a p-sphere with n unit p-cells. The integer n is the n-dimensional area measure of the p-sphere. There are finitely many rooted $n - S^p$ for each n. The generalization of the FWD to higher-dimensional knots is:

Conjecture 2 (Generalized FWD Conjecture). The probability that a randomly chosen $n - S^p$ in Z^{p+2} $(p \geq 2)$ is knotted tends to one as n tends to infinity.

In the above conjecture, the number n represents the p-dimensional volume measure (the size) of the p-sphere in Euclidean $(p + 2)$-space. The above conjecture of the inevitability of knotting of randomly embedded spheres as the volume of the sphere tends to infinity can be made in any codimension 2 context in Euclidean space; smooth codimension 2 spheres, piecewise linear codimension 2 spheres, locally flat codimension 2 spheres, etc. Efforts to prove this conjecture for $p = 2$ (2-spheres in 4-space) have been made, without success. One problem is that there is no known analogue of a Kesten Pattern Theorem for 2-disks in Z^4, so a new idea may be required in order to prove this conjecture. To my knowledge, although almost certainly true, no numerical simulation evidence for this conjecture exists.

3 Entanglement Complexity of Random Knots and Random Arcs

Intuition tells us that the entanglement of random knots grows with length — the longer it is, the more entangled. We are now positioned to measure entanglement complexity, and prove that the complexity grows at least linearly with the length. The reason for this growth in complexity with length is that long random knots are highly composite (have many summands), and most measures of knot complexity are additive under connected sum. It is instructive at this point to recall the work of Kendall [Ken] on the complexity of Brownian motion. Kendall proved that given any (smooth or finite polygonal) arc in R^3, and given any closed tube neighborhood of this arc, the Brownian motion eventually enters one end of the tube for the last time, traverses around the interior of the tube, then exits the other end of the tube. This happens no matter how complicated the arc and how small the tube diameter. So, Kendall has shown that Brownian motion exhibits *all knots at all scales.*

Let $\mathcal{K}$ denote the set of knot types in R^3. A *good measure of knot complexity* is a function $F : \mathcal{K} \to [0, \infty)$ that satisfies the following:

(a) $F(\text{unknot}) = 0$

(b) There exists a knot type $K \in \mathcal{K}$ such that $F(nK \sharp L) \geq nF(K) > 0$ for all $L \in \mathcal{K}$, where $\sharp$ denotes connected sum of knots.

Good measures of knot complexity are designed so that the complexity of any knot that contains pK as a summand is bounded below by p times the complexity of K; hence any good measure of knot complexity diverges to infinity at least linearly with length. More precisely, we have the following lemma:

Lemma 1 (SSW). *For any F (a good measure of knot complexity), let K be a knot that satisfies part (b) of the definition above. Then, there exists a positive integer n_K such that for sufficiently large $n > n_K$, all but exponentially few n-SAPs have F-complexity which exceeds $F(K)\ ((n/n_K) - 1)$.*

Proof. Choose D_K as in Theorem 2 above and choose n_K such that $\lfloor (n_K - 1)D_K \rfloor = 0$ and $\lfloor n_K D_K \rfloor = 1$. For sufficiently large $n > n_K$, all but exponentially few n-SAPs K' are of the form $K' = \lfloor n_K D_K \rfloor K \sharp L$ for some $L \in \mathcal{K}$. This means that $F(K') \geq \lfloor n_K D_K \rfloor F(K) > F(K)((n/n_K) - 1)$. □

Theorem 3 (SSW). *The following are good measures of knot complexity:*

(a) number of prime factors
(b) genus
(c) bridge number -1
(d) span of any non-trivial knot polynomial
(e) log(order)
(f) crossing number
(g) unknotting number
(h) minor index
(i) braid index -1

Proof. All of the above are non-trivial non-negative integer knot invariants, and some are known to be additive on connected sum. In any event, each of them is additive on trefoil summands of a given knot, so the fact that random knots tend to have many trefoil summands means that their complexity must grow at least linearly with length. I will give the argument for one of the more interesting entanglement invariants, the unknotting number. For a knot K, the *unknotting number* $\mu(K)$ is the minimum number of times that the knot must be passed through itself in order to unknot it. Unknotting number is believed to be additive on connected sums:

Conjecture 3 ((Additivity of Unknotting Numbers)). $\mu(K_1 \sharp K_2) = \mu(K1) + \mu(K_2)$.

For any knot K, let X denote the bounded knot complement in S^3, and X^* denote the infinite cyclic covering space of X. $H_1(X^*; Z)$ is presented as a module over the ring $\Lambda = Z[t, t^{-1}]$ (the integral group ring of the infinite cyclic multiplicative group) by a square matrix with entries in Λ, called

the *Alexander matrix*. Let $m(K)$ denote the *minor index* of K, that is, the minimum size for any square Alexander matrix that presents the Λ-module $H_1(X^*; Z)$. □

We now use the following theorem of Nakanishi:

Theorem 4 (Nak). *For any* $K \in \mathcal{K}, 0 \leq m(K) \leq \mu(K)$.

Hence, if $K = 3_1$ (the trefoil knot), then $0 \leq n \leq m(nK\sharp L) \leq \mu(nK\sharp L)$ for any $L \in \mathcal{K}$, so the minor index and the unknotting number are good measures of knot complexity. □

Scientists continue to be interested in measuring entanglement of arcs in 3-space — they would like to be able to measure exactly where an arc is knotted, for example. In all of 3-space, every arc is unknotted — one can thread an end through any entanglement in order to resolve it. However, if the ends are constrained (to lie in a 2-sphere for example), the arc can be knotted. In the proof of the FWD conjecture above, the pair $(N(T'), N(T') \cap T)$ is a *knotted ball pair* — a knotted arc in an enclosing 3-ball. On the other hand, if one were to take a closed tube neighborhood B of the SAW T (with the ends of T in the ends of B), then the pair (B, T) would be an unknotted ball pair — whether or not the arc T is unknotted depends on the surrounding 3-ball one chooses [SW, Fig. 1].

Given any (smooth, finite polygonal) arc (circle) in 3-space, it is possible to measure the (geometric) complexity of the embedding as follows: given any regular projection of the arc (circle), compute the crossing number, and then average this quantity over all projections, to get the *average crossing number* of the arc (circle). It is possible to alter the trefoil pattern T by adding edges to obtain a slightly longer trefoil pattern S such that one sees at least 3 crossings in every regular projection of S. The Kesten pattern theorem guarantees that all but exponentially few sufficiently long n-SAWs (n-SAPs) contains at least $\lfloor nD_S \rfloor$ copies of the trefoil pattern S, so the average crossing number of any such n-SAW (n-SAP) is bounded below by $3\lfloor nD_S \rfloor$. This means that if $\mathcal{X}_n$ denotes the expected value of average crossing number over all n-SAWs (n-SAPs), we have the following result:

Theorem 5 (RSW). $\mathcal{X}_n \to \infty$ *with* n*, and the divergence is at least linear.*

In order to detect true topological entanglement in an arc in 3-space, one needs to join the ends up to form a circle, and then use topological measures of entanglement for circles. In [RSW] this closing operation was performed on n-SAWs in Z^3 using two methods: (a) Choose a direction at random, and construct two parallel rays from the endpoints of the SAW to infinity. One can assume that the rays meet at the point infinity in S^3, so this operation produces a circle. (b) Choose a direction at random, and add two short (length less than one) parallel line segments to the ends of the n-SAW, then close up by adding a straight line segment connecting the ends of the added segments. This almost always produces an embedded circle. It can be shown [RSW] that either

of these closure methods produces circles whose (topological) entanglement complexity (as in Theorem 4) diverges to infinity at least linearly with n. In [RSW] the results of Monte Carlo simulations are discussed which show that for $n \leq 2000$, the entanglement measure expected log(order) gives almost identical results for both methods of n-SAW closure, and that this measure grows with increasing n.

4 Writhe, Signature and Chirality of Random Knots

There are other interesting measures of geometric and topological entanglement complexity which are additive when applied to connected sums of knots, but can take on negative as well as positive values, so it is possible for the measure to yield zero when evaluated on a nontrivial connected sum. For measures such as this, we will show that a lower bound for the growth of the absolute value of one of these measures is $\sqrt{n}$ (instead of n itself).

Given a regular planar projection in the direction ξ of the (smooth, finite polygonal) oriented circle K in R^3, assign the integer ± 1 to each crossing as determined by the right-hand rule (*oriented skew lines sign convention*) in Figure 1. Note that the sign of a given crossing in the projection is in-

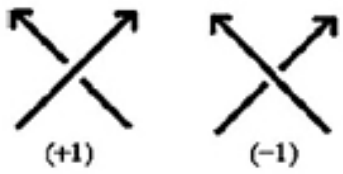

Fig. 1. Oriented skew lines sign convention

dependent of orientation of the circle; reversing the orientation of the circle in turn reverses the orientation of each arrow in a crossing, leaving the sign of the crossing invariant. Adding up the signs for each crossing in the regular projection, one obtains the *projected writhe* $\omega_\xi(K)$; by averaging over all directions $\xi \in S^2$, we obtain the *writhe* $\omega(K)$. The writhe is a real-valued geometric measure of non-planarity of K; if K is a subset of a plane in R^3, then $\omega(K) = 0$ [Ful]. Note that $\omega(K)$ is not a topological invariant — the writhe changes when one twists or bends a knot.

In order to approximate the writhe for a given knot K, one needs to choose a finite (but usually large) set of directions, compute the projected writhe for each direction, then average the results. For SAPs on Z^3, however, one can compute the writhe exactly using only 4 directions because of the symmetry of Z^3 [LS,Cim,LaS]. Moreover, the (first) proof of this fact [LS] used linking numbers to characterize the writhe; the topological invariance

of linking numbers and the fact that one need use only 4 directions is useful both in proving rigorous results about the writhe of random n-SAPs and in doing Monte Carlo simulations to compute average properties of the writhe for n-SAPs [ROS].

Suppose that we are using the projection direction ξ (ξ is a unit vector); let K_ξ denote the pushoff of K in the direction ξ; $K_\xi = K + s\xi$, where s is a small positive number. The pushoff K_ξ inherits its orientation from K. Let $Lk(K, K_\xi)$ denote the *linking number* of K with K_ξ.

Lemma 2 (LS). $\omega_\xi(K) = Lk(K, K_\xi)$.

Proof. The linking number $Lk(K, K_\xi)$ is computed from any regular projection of the pair of oriented curves (K, K_ξ). Given any regular planar projection, ignore self-crossings of each curve, and assign the integer ± 1 to each crossing involving the top strand from K and the bottom strand from K_ξ. The sum of these signed crossings is the linking number. We would like to use the projection in direction of unit vector ξ to determine the linking number; of course, since K_ξ is the pushoff of K in direction ξ, K_ξ lies in the shadow of K and we cannot see it at all in the projection. However, in the plane of the projection, as one walks around the knot diagram in the orientation direction, one can construct the *planar pushoff* of the projection of K; push K to the right to obtain a parallel copy K^* of K. K and K^* form a pair of railroad tracks; each original crossing of K becomes 4 crossings where the railroad crosses itself. By putting a small upward bump in the overcrossing track, we see that there is a small embedded annulus (horizontal curtain) connecting K to K^*. Moreover, each original overcrossing of K gives rise to a single overcrossing of K over its planar pushoff K^*. Since K^* is oriented in parallel to K, the sign of the overcrossing of K over K^* is identical to the sign of K overcrossing itself. We conclude that $\omega_\xi(K) = Lk(K, K^*)$. Now allow the force of gravity to pull the horizontal annulus curtain to vertical position. This downward rotation of the annulus is an isotopy in the complement of K which takes the planar pushoff K^* to the pushoff K_ξ. Hence, $Lk(K, K^*) = Lk(K, K_\xi)$, and we conclude that $\omega_\xi(K) = Lk(K, K_\xi)$. □

Consider now the 2-sphere S of projection (pushoff) directions in R^3, and take S to have radius $1/2$, centered at the origin. The 3 coordinate planes in R^3 separate S into 8 connected regions (octants) specified by constancy of sign in each coordinate. The interior of an octant consists of points with no coordinate $= 0$. Let K be a rooted SAP in Z^3. For the vector direction $\xi \in S$, we let $K_\xi = K + \xi$ denote the pushoff of K in direction ξ.

Claim (1). If ξ lies in the interior of any octant, then K and K_ξ are disjoint. Without loss of generality, assume that ξ lies in the interior of the octant where all 3 coordinates are positive. Then $\xi = (\xi_1, \xi_2, \xi_3)$ where $0 < \xi_i < 1/2$ for $1 \leq i \leq 3$. Points in the SAW K have the property that at least two of the coordinates are integers. The points in the pushoff K_ξ are obtained by adding the vector $\xi = (\xi_1, \xi_2, \xi_3)$ to each of the points of K. Suppose now that

$(x, y, z) \in K$, and that $(x + \xi_1, y + \xi_2, z + \xi_3) \in K \cap K_\xi$. By the pigeonhole principle, at least one of the following is true: both x and $x + \xi_1$ are integers; both y and $y + \xi_2$ are integers; both z and $z + \xi_3$ are integers. This means that at least one of $\{\xi_1, \xi_2, \xi_3\}$ is a non-zero integer, which is impossible. The validity of claim 1 means that $Lk(K, K_\xi)$ is defined for any pushoff in the interior of any of the 8 octants of S.

Claim (2). If ξ and ζ lie in the interior of the same octant, then $Lk(K, K_\xi) = Lk(K, K_\xi)$. Consider the great circle on S that goes through both ξ and ζ. The points ξ and ζ separate the great circle into two arcs; the shorter of these arcs is contained in the interior of the octant containing both ξ and ζ. The points along this shorter arc define a 1-parameter family of pushoffs, starting with ξ and ending with ζ. Thus the curve K_ξ can be isotoped to K_ξ in the complement of the curve K, so we conclude that $Lk(K, K_\xi) = Lk(K, K_\zeta)$. Since the area on S of each of the octants is $\pi/8$, we conclude that $\omega(K)$ can be computed as the average of 8 directional writhes, one direction chosen from the interior of each of the 8 octants. We can however do better than this — we only need to average 4 directional writhes, since the directional writhe on any octant is the same as the directional writhe on its antipodal octant, as shown in the next claim.

Claim (3). If ξ is in the interior of an octant, then $Lk(K, K_\xi) = Lk(K, K_{-\xi})$. For the parameter t with $-1 \le t \le 0$, let $K_{t\xi}$ be obtained by adding the vector $t\xi$ to each point of K. As in claim 1, for each value of t, the curves $K_{t\xi}$ and K_ξ are disjoint. To see this, suppose that $(x + \xi_1, y + \xi_2, z + \xi_3) = (x' + \xi_1, y' + \xi_2, z' + \xi_3)$ for (x, y, z) and $(x', y', z') \in K$. Suppose also that both x and x' are integers. Then $(x' - x) = (1 - t)\xi_1$. But $(1 - t)\xi_1$ cannot be an integer because $1 \le (1-t) \le 2$ and $0 < \xi_1 < 1/2$. Therefore we can isotop K to $K_{-\xi}$ in the complement of K_ξ, so $Lk(K_\xi, K) = Lk(K_\xi, K_{-\xi})$. Likewise, we can isotop K_ξ to K in the complement of $K_{-\xi}$, so $Lk(K_{-\xi}, K_\xi) = Lk(K_{-\xi}, K)$. By symmetry of linking numbers in R, we conclude $Lk(K, K_\xi) = Lk(K, K_{-\xi})$. □

Theorem 6 (LS,Cim,LaS). *For any SAP K, the writhe $\omega(K)$ is the average of 4 directional writhes, with the directions chosen in any 4 mutually nonantipodal octants of S, and $4\omega(K)$ is an integer.*

Now we would like to investigate the properties of the ensemble of rooted n-SAPs. If K is an n-SAP, let K^* denote the mirror image of K. Writhe has the property that $\omega(K) = -\omega(K^*)$; so if we were average the writhe over all n-SAPs, the expected value of the writhe $\langle\omega\rangle_n = 0$ by symmetry. Consequently, we are interested in the expected value of the absolute value of the writhe $\langle|\omega|\rangle_n$, or the square of the writhe $\langle\omega^2\rangle_n$. More generally, we are interested in the distribution of $\langle|\omega|\rangle_n$ over the set of n-SAPs.

Let $P = (0, 0, 0)$ and $Q = (0, 1, 0)$ in Z^3. Both P and Q are on the boundary of a solid cube C of size $2 \times 2 \times 2$ whose corners are:

$$\{(0,-1,0),(0,1,-1),(0,-1,-1),(0,1,1),$$
$$(2,-1,0),(2,1,-1),(2,-1,-1),(2,1,1)\}$$

The cube C is symmetric about the plane $z = 0$. Let $\{i, j, k\}$ denote unit vectors in a right-handed coordinate system for Z^3 . Consider the 10-SAP B contained in C and described as follows: begin at P, and follow the sequence of steps $\{i, i, -k, -j, -i, j, j, k, -i, -j\}$. Let $\omega(B)$ denote the writhe of B. In order to use Theorem 8 above, we choose pushoff directions as follows: $v_1 = (1, 1, 1)$, $v_2 = (-1, 1, 1)$, $v_3 = (-1, -1, 1)$, $v_4 = (1, -1, 1)$. Let $\{B_1, B_2, B_3, B_4\}$ denote pushoffs of B in the 4 pushoff directions $\{v_1, v_2, v_3, v_4\}$. For each pushoff direction we choose a small enough scalar to ensure that adding that scalar multiple of the pushoff direction to B to create the pushoff creates no intersections between B and its pushoff. By inspection, we have that $Lk(B, B_1) = Lk(B, B_3) = Lk(B, B_4) = +1$ and $Lk(B, B_2) = -1$. Hence $\omega(B) = +1/2$. If B^* denotes the mirror image of B (reflected in the plane $z = 0$), then $\omega(B^*) = -1/2$.

Suppose now that A is a SAP which (given one of its two orientations) intersects the cube C only in the SAW B' which begins at P and ends at Q and traverses all the steps of B except the last one: that is, B' starts at P and consists of the 9 steps $\{i, i, -k, -j. - i, j, j, k, -i\}$. We can truncate the polygon A by deleting the 9 steps of B' and adding in the step i which connects P to Q in the boundary of C, generating the new polygon A'. The polygon A is obtained by concatenating A' and B.

Lemma 3 (ROS). $\omega\langle A\rangle = \omega(A') + \omega(B)$.

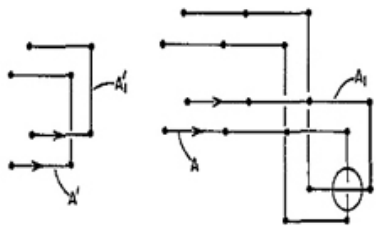

Fig. 2. [ROS]: Additivity of writhe

Proof. Consider the following pushoffs A_1 of A and A'_1 of A' in direction v_1. Fig. 2 shows the projection down the Z-axis of these curves near the cube C; the remainder of the projection of the polygon has been suppressed. In Fig. 2, the (+) crossing in the circle is where A_1 crosses over A in the interior of cube C. By a small move (isototpy in the interior of C), one can move curve A straight up, crossing through curve A_1 until A now goes over A_1 in the circle, and no other crossings have been changed. This gives a pair of curves

which are isotopic (by an isotopy in the interior of C) to the pair $\{A', A_1'\}$. This proves that

$$Lk(A, A_1) = Lk(A', A_1') + 1 = Lk(A', A_1') + Lk(B, B_1).$$

A similar calculation for each of the other 3 pushoff directions gives:

$$Lk(A, A_i) = Lk(A', A_i') + 1 = Lk(A', A_i') + Lk(B, B_i) \qquad i = 3, 4$$

and

$$Lk(A, A_1) = Lk(A', A_1') + 1 = Lk(A', A_1') + Lk(B, B_1) \qquad i = 2.$$

Averaging these 4 equations completes the proof of Lemma 9. □

Theorem 7 (ROS). *For every function $f(n) = o(\sqrt{n})$, the probability that $\langle|\omega|\rangle_n < f(n)$ goes to zero as n goes to infinity.*

Proof. We use the Kesten Pattern Theorem [Kes] and a coin-tossing argument [DPS,ROS]. We call the $(3, 1)$ ball pair consisting of any translate of the SAW B' and the surrounding cube C a *pattern* $P = \{C, B'\}$. Let the pattern P^* be the ball pair $\{C, B'^*\}$ where B'^* is the mirror image of B' (reflected in the plane $z = 0$). Kesten's pattern theorem implies that there is a positive number ε such that for all except exponentially few sufficiently long n-SAPs, there are at least $\lfloor \varepsilon n \rfloor$ pairwise disjoint translates of C, each of which intersects the SAW in a translate of B' or B'^*. The distribution of patterns among the copies of C is analogous to tossing a coin, since each of the patterns B' and B'^* occur independently with probability $1/2$ in each of the $\lfloor \varepsilon n \rfloor$ translations of the cube C. Consequently the probability that B' occurs exactly k times among the $\lfloor \varepsilon n \rfloor$ occurrences of either B' or B'^* is less than $(1/\sqrt{\lfloor \varepsilon n \rfloor})$ for every $k \leq \lfloor \varepsilon n \rfloor$. (This can be seen by using Stirling's approximation to the binomial distribution.) The fraction of polygons with at least $\lfloor \varepsilon n \rfloor$ occurrences of either P or P^* is at least $(1 - e^{-\gamma n})$ for some positive γ. For each of these SAPs, the writhe is the sum of two terms (Lemma 9). The first term is from the polygon formed by truncating $\lfloor \varepsilon n \rfloor$ times, and the second term is from the $\lfloor \varepsilon n \rfloor$ copies of B or B^* formed in these truncations. If the total writhe of that SAP is less than $f(n)$, then the contribution to the writhe from the $\lfloor \varepsilon n \rfloor$ occurrences of the patterns must be one of at most $\lceil 2f(n) + 1 \rceil$ different values. Hence

$$\text{Prob}(\langle|\omega|\rangle_n < f(n)) \leq \frac{(1 - e^{-\gamma n})\lceil 2f(n) + 1 \rceil}{\sqrt{\varepsilon_n}}$$

which goes to zero as n goes to infinity if $f(n) = o(\sqrt{n})$. □

Theorem 10 strongly suggests that $\langle|\omega|\rangle \sim n^{\alpha}$. Monte Carlo simulation [ROS] for values of n between 400 and 1100 give a writhe distribution that is symmetric about the origin and sharply peaked at the origin ($n = 400$), and

less sharply peaked at the origin as n increases. When values of $\log(\langle|\omega|\rangle)$ are plotted against $\log n$, the evidence [ROS,Fig. 4] for linear behavior is excellent, and produces the following estimate: $\alpha = 0.522 \pm 0.004$.

Are almost all long randomly chosen SAP's chiral? The answer is yes. One invariant that detects chirality (inequivalence of an unoriented knot and its mirror image) is the *signature* $\sigma(K)$ of the knot K [BZ]. If K^* denotes the mirror image of K, then $\sigma(K^*) = -\sigma(K)$, so if K is *achiral* ($K = K^*$), then $\sigma(K) = 0$. The signature is additive on connected sums, so $\sigma(K \sharp L) = \sigma(K) + \sigma(L)$. One can adapt the writhe argument above with the trefoil pattern T replacing the curl pattern B, and a suitable rectangular parallelepiped replacing the cube C. Let $\langle|\sigma|\rangle_n$ denote the average of the absolute value of the signature, averaged over all n-SAPs.

Corollary 1 (DPS). *For every function $f(n) = o(\sqrt{n})$, the probability that $\langle|\sigma|\rangle_n < f(n)$ goes to zero as n goes to infinity.*

Hence, we conclude that the average of the absolute value of the signature of random n-SAPs grows at least as fast as $\sqrt{n}$, and that most long random knots are chiral. This result was first proved for Gaussian random polygons [DPS] by the same method.

5 Application of Random Knotting to Viral DNA Packing

Knots and links are of biological interest because they can detect and preserve topological information, especially information about DNA and the enzymes that act on DNA. Knotted DNA molecules can be well characterized experimentally by gel electrophoresis and microscopy (both transmission electron and atomic force microscopy), and therefore used as assays for different biochemical reactions. Characterization of knotted products formed by random cyclization of linear molecules has been used to quantify important biochemical properties of DNA such as its effective diameter [RCV,ShW]. DNA knots and catenanes obtained as the product of site-specific recombination have also been a key to unveiling the mechanism of enzymatic action [WC,SBS]. In both cases, the development of mathematical and computational tools has greatly enhanced analysis of the experimental results [FLA,ES]

Significant numbers of DNA knots are found also in biological systems: in *Escherichia coli* cells harboring mutations in the *GyrB* or *GyrA* genes [SKI], bacteriophages P2 and P4 [LDC,LPC], and cauliflower mosaic viruses [MML]. However, very little biological information about these systems has been inferred from the observed knots. In particular, interpretation of the experimental results for bacteriophages has been limited by the experimental difficulty in quantifying the complex spectrum of knotted products. These difficulties have paralleled those encountered in developing a theory for random knotting of ideal polymeric chains in cases where interactions with other

macromolecules and/or confinement in small volumes have a significant function [Man,TRO,MMO].

Bacteriophages are viruses that infect bacteria. They pack their double-stranded DNA genomes to near-crystalline density in viral capsids and achieve one of the highest levels of DNA condensation found in nature. When I was on sabbatical in Berkeley in 1989, Jim Wang described to me the problem of DNA packing in icosehedral viral capsids, and the high degree of knotting produced when the viral DNA is released from the capsids. Despite numerous studies some essential properties of the packaging geometry of the DNA inside the phage capsid are still unknown. Although viral DNA is linear double-stranded with sticky (cohesive) ends, the linear viral DNA quickly becomes cyclic when removed from the capsid, and for some viral DNA the observed knot probability is an astounding 95%. In the summer of 1998, my PhD students Javier Arsuaga and Mariel Vazquez spent 2 months in the laboratory of Joaquim Roca in Barcelona, supported by the Burroughs Wellcome Interfaces grant to the Program in Mathematics and Molecular Biology. In the Roca laboratory, they infected bacterial stock, harvested viral capsids and extracted and analyzed the viral DNA. They quantified the DNA knot spectrum produced in the experiment, and used Monte Carlo generation of knots in confined volumes to compare a random knot spectrum to the observed viral DNA knot spectrum. A series of papers were produced as a result of this collaboration [TAV,AVT,ArT,AVM], and I will describe some of the results from these papers, focusing on (and reproducing here) most of the discussion and analysis from the most recent PNAS paper [AVM].

All icosahedral bacteriophages with double-stranded DNA genomes are believed to pack their chromosomes in a similar manner [EC]. During phage morphogenesis, a procapsid is first assembled, and a linear DNA molecule is actively introduced inside it by the connector complex [RHA,STS]. At the end of this process, the DNA and its associated water molecules fill the entire capsid volume, where DNA reaches concentrations of 800 mg/ml [KCS]. Some animal viruses [SB] and lipoDNA complexes used in gene therapy [SDD] are postulated to hold similar DNA arrangements as those found in bacteriophages. Although numerous studies have investigated the DNA packing geometry inside phage capsids, some of its properties remain unknown. Biochemical and structural analyses have revealed that DNA is kept in its B form [ACT,EH,LDB] and that there are no specific DNAprotein interactions [HMC,Ser] or correlation between DNA sequences and their spatial location inside the capsid, with the exception of the cos ends in some viruses. Many studies have found that regions of the packed DNA form domains of parallel fibers, which in some cases have different orientations, suggesting a certain degree of randomness [EH,LDB]. The above observations have led to the proposal of several long-range organization models for DNA inside phage capsids: the ball of string model [RWC], the coaxial spooling model [EH,CCR,RWC,CCR], the spiral-fold model [BNB], and the folded toroidal model [Hud]. Liquid crystalline models, which take into account properties of DNA at high concentrations

and imply less global organization, have also been proposed [LDB]. Cryo EM and spatial symmetry averaging has recently been used to investigate the surface layers of DNA packing [JCJ]. In [AVM], the viral DNA knot spectrum is used to investigate the packing geometry of DNA inside phage capsids.

The bacteriophage P4 has a linear, double-stranded DNA genome that is 10–11.5 kb in length and flanked by 16-bp cohesive *cos* ends [WMC]. It has long been known that extraction of DNA from P4 phage heads results in a large proportion of highly knotted, nicked DNA circles [LDC,LPC]. DNA knotting probability is enhanced in P4 derivatives containing genome deletions [WMH] and in tailless mutants [IJC]. Most DNA molecules extracted from P4 phages are circles that result from the cohesive-end joining of the viral genome. Previous studies have shown that such circles have a knotting probability of about 20% when DNA is extracted from mature P4 phages [LDC]]. This high value is increased more than 4-fold when DNA is extracted from incomplete P4 phage particles (which we refer to as capsids) or from noninfective P4 mutants that lack the phage tail (which we refer to as tailless mutants [LDC]). Knotting of DNA in P4 deletion mutants is even greater. The larger the P4 genome deletion, the higher the knotting probability [WMH]. For P4 vir1 del22, containing P4s largest known deletion (1.6 kb deleted [RDM]), knotting probability is more than 80% [IJC]. These values contrast with the knotting probability of 3% (all trefoil knots) observed when identical P4 DNA molecules undergo cyclization in dilute free solution [RCV,RUV]. These differences are still more striking when the variance in distribution of knot complexity is included. Although knots formed by random cyclization of 10-kb linear DNA in free solution have an average crossing number of three [RCV,RUV], knots from phage particles have a knotting probability of 95% and appear to have very large crossing numbers, averaging about 26. [LDC,WMH,IJC].

The reasons for the high knotting probability and knot complexity of bacteriophage DNA have been investigated. Experimental measurements of the knotting probability and distribution of knotted molecules for P4 vir1 del22 mature phages, capsids, and tailless mutants was performed by 1- and 2-dimensional gel electrophoresis, followed by densitometer analysis. We will describe the Monte Carlo simulations to determine the effects that the confinement of DNA molecules inside small volumes have on knotting probability and complexity. We conclude from our results that for tailless mutants a significant amount of DNA knots must be formed before the disruption of the phage particle, with both increased knotting probability and knot complexity driven by confinement of the DNA inside the capsid.

In [AVM] it is shown that the DNA knots provide information about the global arrangement of the viral DNA inside the capsid. The distribution of the viral DNA knots is analyzed by high-resolution gel electrophoresis. Monte Carlo computer simulations of random knotting for freely jointed polygons confined to spherical volumes is performed. The knot distribution produced by simulation is compared to the observed experimental DNA knot spectrum. The simulations indicate that the experimentally observed scarcity of the achi-

ral knot 4_1 and the predominance of the torus knot 5_1 over the twist knot 52 are not caused by confinement alone but must include writhe bias in the packing geometry. Our results indicate that the packaging geometry of the DNA inside the viral capsid is non-random and writhe-directed.

5.1 Knot Type Probabilities for P4 DNA in Free Solution

The probability that a DNA knot K of n statistical lengths and diameter d is formed by random closure in free solution is given by $P_K(n,d) = P_K(n,0)e^{-rd/n}$, where r depends on the knot type and equals 22 for the trefoil knot 3_1 and 31 for the figure 8 knot 4_1 [RCV,ShW]. The knotting probability of a 10-kb DNA molecule cyclized in free solution is 0.03, which implies an effective DNA diameter near 35 Å. Because $P_{3_1}(34,0) = 0.06$ and $P_{4_1}(34,0) = 0.009$, then $P_{3_1}(34,35) = 0.027$ (1/36 times that of the unknot) and $P_{4_1}(34,35) = 0.003$ (1/323 times that of the unknot). These values were used to estimate the fractions of the knot 3_1 and the knot 4_1 generated for P4 DNA in free solution.

5.2 Monte Carlo Simulation

Knotting probabilities of equilateral polygons confined into spherical volumes were calculated by means of Markov-chain Monte Carlo simulations followed by rejection criteria. Freely jointed closed chains, composed of n equilateral segments, were confined inside spheres of fixed radius, r, and sampled: values of n ranged from 14 to 200 segments; r values, measured as multiples of the polygonal edge length, ranged from 2 to infinity. Excluded volume effects were not taken into account. Markov chains were generated by using the Metropolis algorithm [MRR]. The temperature, a computational parameter, was held at $T = 300$ K to improve the efficiency of the sampling algorithm. Other values of T produced similar results, thus indicating that the computation is robust with respect to this parameter. Chains contained inside the sphere were assigned zero energy. Chains lying partly or totally outside the confining sphere were assigned an energy given by the maximum of the distances of the vertices of the chain to the origin. Only chains with zero energy were sampled. A random ensemble of polygons was generated by the crankshaft algorithm as follows: (i) two vertices of the chain were selected at random, dividing the polygon into two subchains, and (ii) one of the two subchains was selected at random (with equal probabilities for each subchain), and the selected subchain rotated through a random angle around the axis connecting the two vertices. This algorithm is known to generate an ergodic Markov chain in the set of polygons of fixed length [Mil]. Correlation along the subchains was computed by using time-series analysis methods as described by Madras and Slade [MS]. Identification of the knotted polygons was achieved by computing the Alexander polynomial Δ (t) [BZ,Rol,Adm] evaluated at $t = -1$. It is known that $\Delta(-1)$ does not identify all knotted chains; however, for polygonal

chains not confined to a spherical volume, nontrivial knots with trivial $\Delta(-1)$ values rarely occur. This circumstance has been observed by using knot invariants, such as the HOMFLY polynomial, that distinguish between knotted and unknotted chains with higher accuracy than the Alexander polynomial. Computer simulations for small polygons (< 55 segments) show that the knotting probabilities obtained by using $\Delta(-1)$ agree with those obtained by using the HOMFLY polynomial. Furthermore, Deguchi and Tsurusaki have reported that the value of $\Delta(-1)$ can almost always determine whether a given Gaussian polygon is unknotted for lengths ranging from 30 to 2,400 segments [DT2]. Each selected knotted polygon was further identified by evaluating its Alexander polynomial at $t = -2$ and $t = -3$. Although the Alexander polynomial is an excellent discriminator among knots of low crossing number and its computation is fast, it does not distinguish completely among some knotted chains [for example, composite knots $3_1\sharp 3_1$ and $3_1\sharp 4_1$ have polynomials identical to those of prime knots 8_{20} and 8_{21}, respectively [RW]. Evaluation of the polynomial at $t = -2$ and t = -3 is also ambiguous because the Alexander polynomial is defined up to units (power or t) in $Z[t^{-1}, t]$. To deal with this uncertainty, we followed van Rensburg and Whittington [RW] and chose the largest exponent k such that the product $n^{\pm k}\Delta(-n)$ is an odd integer with $n = 2$ or 3. This value was taken as the knot invariant. To compute the writhe, we generated > 300 regular projections and resulting knot diagrams for each selected polygon. To each of the projected crossings a sign was assigned by the skew lines convention (Fig. 1). The directional writhe for each diagram was computed by summing these values. The writhe was then determined by averaging the directional writhe over a large number of randomly chosen projections. To generate writhe-directed random distributions of polygons, we used a rejection method in which polygons whose writhe was below a positive value were not sampled.

5.3 Results and Discussion Knot Complexity of DNA Molecules Extracted from Phage P4

We extracted the 10-kb DNA from the tailless mutant of phage P4 vir1 del22, which produces 95% knotted molecules [AVT], and analyzed it by a high-resolution two-dimensional gel electrophoresis [TAV] (Fig. 3A). This technique allowed us to separate DNA knot populations according to their crossing number (i.e., the minimal number of crossings over all projections of a knot), as well as to separate some knot populations of the same crossing number [SKB,VCL]. In the first dimension (at low voltage), individual gel bands corresponding to knot populations having crossing numbers between three and nine were discernible; knots with higher crossing numbers were embedded in a long tail (denoted as K in Fig. 3 A). The second dimension (at high voltage) further resolved individual gel bands corresponding to knot populations with crossing numbers between six and nine. Although knot populations containing three, four, and five crossings (denoted as 3–5 in Fig. 3B) migrated as sin-

gle bands in a main arch of low gel velocity, knot populations containing six and more crossings split into two subpopulations (denoted as 6–9 and 6′–9′ in Fig. 3B), creating a second arch of greater gel velocity. Fig. 3(A) shows

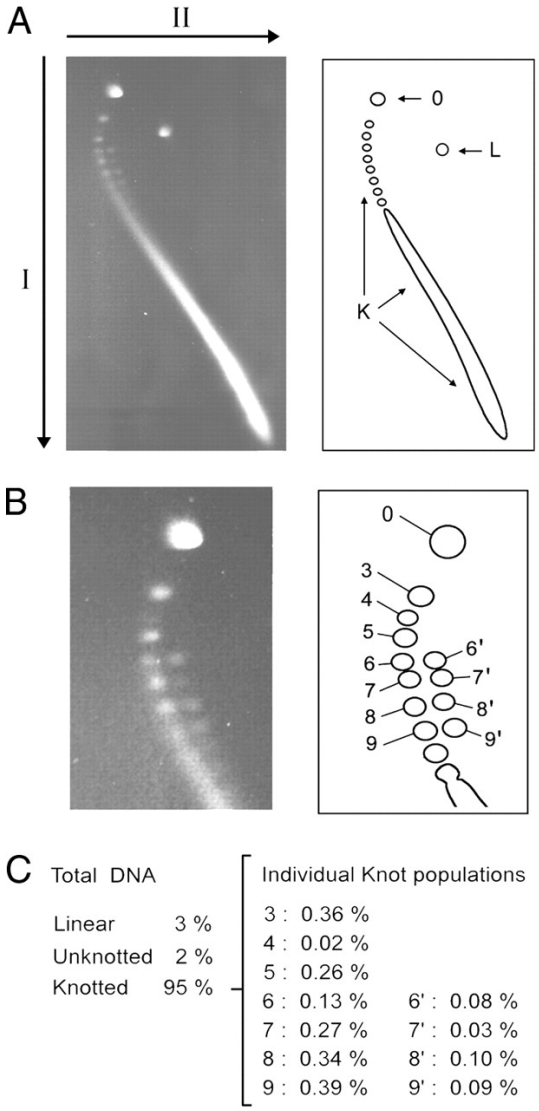

Fig. 3. [AVM]: Analysis of knotted DNA by gel electrophoresis.

DNA was extracted from tailless mutants of phage P4 vir1 del22 and analyzed by two-dimensional agarose gel electrophoresis. The first dimension at

low voltage (top to bottom) separated DNA knot populations according to their crossing number. The unknotted DNA circle or trivial knot (0) has the slowest gel velocity, whereas knotted DNA populations (K) have gel velocity proportional to their crossing number. The second dimension at high voltage (left to right) segregated the linear DNA molecules (L) from the arched distribution of knotted molecules and further resolved some gel bands corresponding to knot subpopulations. (B) Upper area of the gel picture showing knot populations of low crossing number. Individual gel bands corresponding to knot populations containing three to nine crossings are indicated (labeled 3–9) in the main arch of the gel. A second arch of higher gel speed containing knot subpopulations of six and more crossings is generated by the second dimension of the electrophoresis. Individual gel bands of knot subpopulations of six to nine crossings (labeled 6′–9′) are indicated. (C) Quantification of the individual knot populations of six to nine crossings (3–9 and 6′–9′). Both densitometric and phosphorimaging reading of three independent samples of DNA extracted from tailless mutants of phage P4 vir1 del22 produced nearly identical results. The indicated percentage values are relative to the total amount of knotted molecules. Fig. 4 shows the gel velocity at low voltage of individual

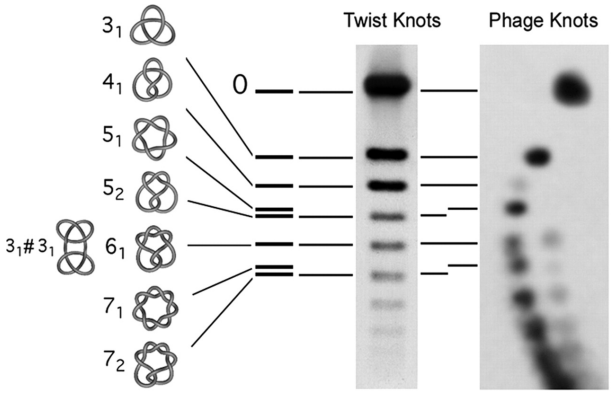

Fig. 4. [AVM]: Identification of specific knot types by their position in the gel.

knot populations resolved by two-dimensional electrophoresis (Right) is compared with the gel velocity at low voltage of twist knots (3_1, 4_1, 5_2, 6_1, and 7_2) of a 10-kb nicked plasmid (Center) and with known relative migration distances of some knot types [SKB,VCL] (Left). Geometrical representations of the prime knots 3_1, 4_1, 5_1, 5_2, 6_1, 7_1 and 7_2 and of the composite knot $3_1\sharp 3_1$ are shown. The unknotted DNA circle or trivial knot (0) is also indicated. Note that in the main arch of the two-dimensional gel and below the knots 3_1 and 4_1, the knot population of five crossings matches the migration of the torus knot 5_1, which migrates closer to the knot 4_1 than to the knots of six crossings. The other possible five-crossing knot, the twist knot 5_2, appears to

be negligible or absent in the viral distribution. Note also that the knot population of seven crossings matches the migration of the torus knot 7_1 rather than the twist knot 7_2. In the secondary arch of the two-dimensional gel, the first knot population of six crossings has low-voltage migration similar to that of the composite knot $3_1\sharp 3_1$.

We quantified the individual knot populations of three to nine crossings, which represented 2.2% of the total amount of knotted molecules (Fig. 4C). Densitometer readings confirmed the apparent scarcity of the figure 8 knot 4_1 relative to the other knot populations in the main arch of the gel (denoted by 4 in Fig. 4B). It also made evident the shortage of the knot subpopulation of seven crossings in the second arch of the gel (denoted by $7'$ in Fig. 3B). The scarcity of the knot 4_1 relative to the knot 3_1 and to other knot populations is enhanced if we make the correction for DNA molecules plausibly knotted outside the viral capsid. Namely, if a fraction of the observed knots were formed by random cyclization of DNA outside the capsid, then, in the worst-case scenario, all observed unknotted molecules (no more than 5% of the total molecules extracted) would be formed in free solution. In such a case, one can predict that 38% of the total number of observed 3_1 knots and 75% of the observed 4_1 knots are formed by random knotting in free solution [RCV,SW]. If all of the knots plausibly formed outside the capsid were removed from the observed knot distribution, the experimental values for knots 4_1 and 3_1 (1:18 ratio) would be corrected, resulting in a 1:44 ratio.

5.4 Identification of Specific Knot Types by Their Location on the Gel

Gel electrophoresis can distinguish some knot types with the same crossing number. For example, at low voltage, torus knots (such as 5_1 and 7_1) migrate slightly slower than their corresponding twist knots (5_2 and 7_2) [SKB,VCL]. We used this knowledge in conjunction with a marker ladder for twist knots (3_1, 4_1, 5_2, 6_1, and 7_2) to identify several gel bands of the phage DNA matching the migration of known knot types (Fig. 4). In the main arch of the gel, in addition to the unambiguous knots 3_1 and 4_1, the knot population of five crossings matched the migration of the torus knot 5_1. The other possible five-crossing knot, the twist knot 5_2 that migrates between and equidistant to the four-and six-crossing knot populations, appeared to be negligible or absent. The knot population of seven crossings matched the migration of the torus knot 7_1 rather than the twist knot 7_2, which has slightly higher gel velocity. Yet, we cannot identify this gel band as the knot 7_1, because other possible knot types of seven crossings cannot be excluded.

Several indicators led us to believe that the second arch of the gel consists of mainly composite knots. First, the arch starts at knot populations containing six crossings, and no composite knots of fewer than six crossings exist. Second, the population of six crossings matched the migration at low voltage of the granny knot $3_1\sharp 3_1$ [KKS], although the square knot $3_1\sharp -3_1$ cannot be

excluded. Third, consistent with the low amount of 4_1 knots, the size of the seven-crossing subpopulation is also reduced: any composite seven-crossing knot is either $3_1\sharp 4_1$ or $-3_1\sharp 4_1$. The increased gel velocity at high voltage (second gel dimension) of composite knots relative to prime knots of the same crossing number likely reflects distinct flexibility properties of the composites during electrophoresis [WSR].

5.5 Monte Carlo Simulations of Random Knot Distributions in Confined Volumes

Next, we asked whether the observed distribution of DNA knots could be compatible with a random embedding of the DNA inside the phage capsid. We used Monte Carlo simulations to model knotting of randomly embedded, freely jointed polygons confined to spherical volumes. Because the persistence length of the duplex DNA is not applicable in confined volumes (it is applicable in unbounded three-dimensional space), we considered freely jointed polygons as the zeroth approximation of the packed DNA molecule. Then, the flexibility of the chain is given by the ratio R/N, where N is the number of edges in the polygon and R is the sphere radius in edge-length units. When we computed random knot distributions for a range of chain lengths confined to spheres with a fixed radius, the probabilities of the knots 3_1, 4_1, 5_1, and 5_2 produced nonintersecting distributions, with simpler knots being more probable (Fig. 5A). That is, the knot 3_1 is more probable than the knot 4_1, and both are more probable than any five-crossing knot. In addition, the probability of the twist knot 5_2 is higher than that of the torus knot 5_1. Similar results had been observed for other random polymer models with/without volume exclusion and with/without confinement [DT2,Man], indicating that this phenomenon is model-independent. All of the simulated distributions, showing the monotonically decreasing amounts of knotted products with increasing crossing number, highly contrasted with our experimental distribution, in which the probability of the knot 4_1 is markedly reduced and in which the knot probability of the knot 5_1 prevails over that of the knot 5_2 (Fig. 5B). These differences provide a compelling proof that the embedding of the DNA molecule inside the phage capsid is not random. Fig. 5(A) shows the distribution probabilities $P(k)$ obtained by Monte Carlo simulations of the prime knots 3_1, 4_1, 5_1, and 5_2 for closed ideal polymers of variable chain lengths (n = number of edges) confined to a spherical volume of fixed radius ($R = 4$ edge lengths). Error bars represent standard deviations. (B) Comparison of the computed probabilities of the knots 3_1, 4_1, 5_1, and 5_2 (for polymers of length $n = 90$ randomly embedded into a sphere of radius $R = 4$) with the experimental distribution of knots. The relative amount of each knot type is plotted. Note that fractions of knots 3_1 and 4_1 plausibly formed in free solution are not subtracted from the experimental distribution. If these corrections are considered, the relative amount of knot 4_1 is further reduced. Fig. 6 shows the writhe of polygons of length $n = 90$ randomly embedded into a sphere

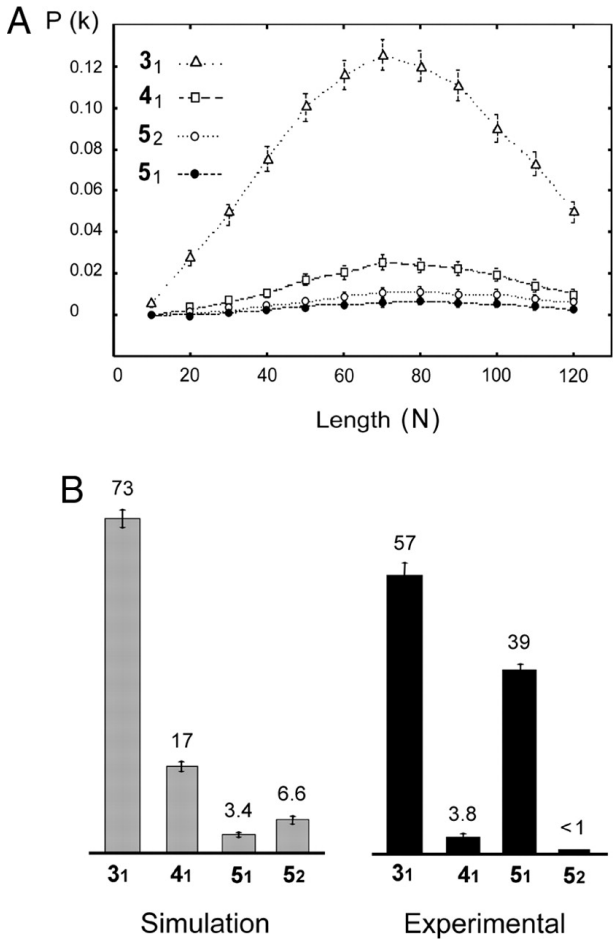

Fig. 5. [AVM]: Comparison of experimental and computer-simulated distributions of knots.

of radius $R = 4$ were computed, and only conformations whose writhe values were higher than a fixed value ($Wr = 4, 6, or 8$) were sampled. The computed mean writhe value ($\langle Wr \rangle$) of each sampled population is indicated. The ratios of the probabilities of the knots 4_1, 5_1, and 5_2 relative to that of the knot 3_1 for each writhe-biased sampling are plotted (P).

How can we explain the scarcity of 4_1 in the spectrum of viral knots? The knot 4_1 is achiral (equivalent to its mirror image). Random polygonal realizations of the 4_1 knot in free space and in confined volumes produce a family of polygons whose writhe distribution for any polygonal length is a Gaussian curve with zero mean (the writhe is a geometrical quantity measuring the signed spatial deviation from planarity of a closed curve) and whose variance grows as the square root of the length (Theorem 7). Therefore, we argue that

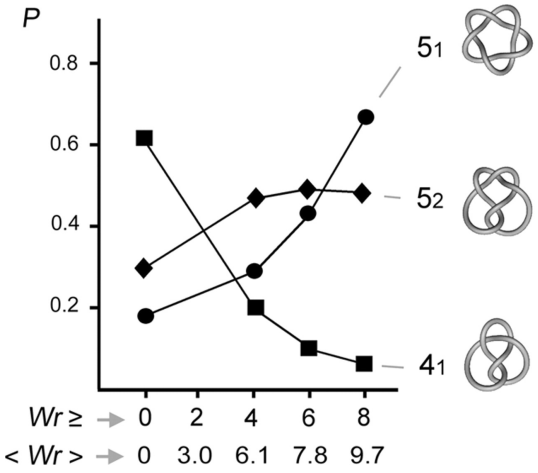

Fig. 6. [AVM]: Effect of a writhe-biased sampling on the probability of knots 4_1, 5_1, and 5_2.

the main reason for the scarcity of the knot 4_1 is a writhe bias imposed on the DNA inside the phage capsid. To test this hypothesis, we simulated polygons randomly embedded in spheres whose mean writhe value was gradually increased. To induce writhe in the sampling, we used a rejection method in which polygons of writhe below a cutoff value were not sampled. Then, we calculated the probabilities of the prime knots 4_1, 5_1, and 5_1 for each writhe-biased sampling. The results shown in Fig. 6 were computed with a freely jointed chain of 90 edges confined in a sphere of radius of 4 edge-length units. A drop of the probability of the knot 4_1, as well as an exponential increase of the probability of the torus knot 5_1 but not of the twist knot 5_2, readily emerged by increasing the writhe rejection value. The same results, but with knots of opposite sign, were obtained for knot distributions with the corresponding negative writhe values. These writhe-induced changes in the knot probability distribution are independent of the number of edges in the equilateral polygon and the sphere radius length. Accordingly, previous studies had shown that the mean writhe value of random conformations of a given knot does not depend on the length of the chain but only on the knot type and that these values are model-independent [RSW]. Because the writhe-directed simulated distributions approach the observed experimental spectrum of knots, we conclude that a high writhe of the DNA inside the phage is the most likely factor responsible for the observed experimental knot spectrum.

Consistent with the involvement of writhe in the DNA packing geometry, it is also the reduced amount of prime knots of six crossings visible in the main arch of the gels (Fig. 3C). All prime knots of six crossings have a lower $\langle Wr \rangle$ ($\langle Wr \rangle$ of $6_1 = 1.23$, $\langle Wr \rangle$ of $6_2 = 2.70$, and $\langle Wr \rangle$ of $6_3 = 0.16$) than the torus knots of five and seven crossings ($\langle Wr \rangle$ of $5_1 = 6.26$ and ($\langle Wr \rangle$

of $7_1 = 9.15$). In contrast, the negligible amount of the twist knot of five crossings ($\langle Wr \rangle$ of $5_2 = 4.54$) in the experimental distributions is striking. The apparent predominance of torus knots (5_1 and 7_2) over twist knots (5_2 and 7_2) in the experimental distribution suggests that writhe emerges from a toroidal or spool-like conformation of the packed DNA. Consistent with our findings, theoretical calculations of long-range organization of DNA by Monte Carlo [MM] and molecular dynamics methods [MMT,ArT,TK,LL favor toroidal and spool-like arrangements for DNA packed inside the phage capsids. Calculations of optimal spool-like conformations of DNA in phage P4 already predicted a large nonzero writhe [ArT]. These studies gave an estimated writhe of 45 for the 10-kb DNA, which closely corresponds with the level of supercoiling density typically found in bacterial chromosomes [ArT].

The actual writhe value of the DNA packaged in the phage P4 capsid cannot be estimated in the present study. The phage P4 capsid has a diameter of 38 nm. If the parameters used to compute writhe-biased ensembles as in Fig. 6 ($n = 90$ and $R = 4$) were applied to a 10-kb DNA molecule, they would translate into 90 segments of 35 nm confined in a model capsid of radius 140 nm. Likewise, our study cannot argue for or against recent models that suggest that to minimize DNA bending energy, a spool conformation might be concentric rather than coaxial [LL]. Therefore, beyond the main conclusion of this work that the distribution of viral knots requires the mean writhe of the confined DNA be nonzero, the applicability of our simulations to other aspects of the DNA packaging in phage P4 is limited. We argue that further identification of the knotted DNA populations will provide more critical information for the packing geometry of DNA inside the phage.

Knots can be seen as discrete measuring units of the organizational complexity of filaments and fibers. Here, we show that knot distributions of DNA molecules can provide information on the long-range organization of DNA in a biological structure. We chose the problem of DNA packing in an icosahedral phage capsid and addressed the questions of randomness and chirality by comparing experimental knot distributions with simulated knot distributions. The scarcity of the achiral knot 4_1 and the predominance of the torus knot 5_1 in the experimental distribution highly contrasted with simulated distributions of random knots in confined volumes, in which the knot 4_1 is more probable than any five-crossing knot, and the knot 5_2 is more probable the knot 5_1. To our knowledge, these results produce the first topological proof of nonrandom packaging of DNA inside a phage capsid. Our simulations also show that a reduction of the knot 4_1 cannot be obtained by confinement alone but must include writhe bias in the conformation sampling. Moreover, in contrast to the knot 5_2, the probability of the torus knot 5_1 rapidly increases in a writhe-biased sampling. Given that there is no evidence for any other biological factor that could introduce all of the above deviations from randomness, we conclude that a high writhe of the DNA inside the phage capsid is responsible for the observed knot spectrum and that the cyclization reaction captures that information.

References

[Ax1] Alexander, J.W.: An Example of a Simply Connected Surface Bounding a Region which is not Simply Connected. Proceedings of the National Academy of Sciences USA, **10**, 8–10 (1924)

[Ax2] Alexander, J.W.: On the Subdivision of 3-Space by a Polyhedron. Proceedings of the National Academy of Sciences USA, **10**, 6–8 (1924).

[Sv] Silver, D.S.: Knot Theory's Odd Origins. American Scientist, **94**, 58–66 (2006).

[Mof] Moffatt, H.K.: Knots and Fluid Dynamics. In Ideal Knots, Series on Knots and Everything, eds. Stasiak, A., Katritch, V., Kauffman, L.H. (World Scientific, Singapore), Vol. **19**, 223–233 (1999).

[Ric1] Ricca, R.L.: New Developments in Topological Fluid Mechanics: From Kelvin's Vortex Knots to Magnetic Knots. In Ideal Knots, Series on Knots and Everything, eds. Stasiak, A., Katritch, V., Kauffman, L.H. (World Scientific, Singapore), Vol. **19**, 255–273 (1999).

[Ric2] Ricca, R.L.: Tropicity and Complexity Measures for Vortex Tangles, Lecture Notes in Physics v. 571. Springer, Berlin Heidelberg New York, 366–372 (2001).

[LZC] Liu, Z, Zechiedrich, E.L., Chan, H.S.: Inferring global topology from local juxtaposition geometry: interlinking polymer rings and ramifications for topoisomerase action. Biophys. J., **90**, 2344–2355 (2006).

[BZ] Buck, G, Zechiedrich, E.L.: DNA disentangling by type-2 topoisomerases. J. Mol Biol. 2004 Jul 23;340(5):933–9.

[FW] Frisch, H.L., Wasserman, E.: Chemical Topology. J. Am. Chem. Soc. **83**, 3789–3795 (1961).

[Was] Wasserman, E.: Chemical Topology. Scientific American **207**, 94–102 (1962)

[Del] Delbruck, M: in Mathematical Problems in the Biological Sciences, Proceedings of Symposia in Applied Mathematics **14**, 327– (1962).

[VLF] Vologodskii, A.V., Lukashin, A.V., Frank-Kemenetkii, M.D., Anshelevich, V.V.: The Knot Problem in Statistical Mechanics of Polymer Chains. Sov. Phys.-JETP **39**, 1059– (1974).

[CM] des Cloizeaux, J., Mehta, M.L.: J. de Physique **40**, 665– (1979).

[MW] Michels, J.P.J., Wiegel, F.W.: Probability of Knots in a Polymer Ring. Phys. Lett. **90A**, 381–384 (1984).

[SW] Sumners, D.W., Whittington, S.G.: Knots in Self-Avoiding Walks. J. Phys. A: Math. Gen. **21**, 1689–1694 (1988).

[Kes] Kesten, H.: On the Number of Self-Avoiding Walks. J. Math. Phys. **4**, 960–969 (1963).

[Pip] Pippenger, N.: Knots in Random Walks. Discrete Appl. Math. **25**, 273-278 (1989).

[Adm] Adams, C.C.: The Knot Book. W.H. Freeman and Co., New York (1991).

[RSW] van Rensburg, E.J.J., Sumners, D.W., Wasserman, E., Whittington, S.G.: Entanglement Complexity of Self-Avoiding Walks. J. Phys. A.: Math. Gen. **25**, 6557–6566 (1992).

[RBHS] Lacher R.C., Bryant J.L., Howard L., Sumners D.W. Linking phenomena in the amorphous phase of semicrystalline polymers. Macromolecules **19**, 2639–2643 (1986).

[DNS] Diao, Y.,Nardo, J.C., Sun Y.: Global Knotting in Equilateral Polygons. J. of Knot Theory and Its Ramifications **10**, 597–607 (2001).

[Do2] Diao, Y.: The Knotting of Equilateral Polygons in R^3. J. of Knot Theory and Its Ramifications **4**, 189–196 (1995).

[DPS] Diao, Y., Pippenger, N., Sumners, D.W.: On Random Knots. J. of Knot Theory and Its Ramifications **3**, 419–424 (1994).

[Do1] Diao, Y.: Minimal Knotted Polygons on the Simple Cubic Lattice. J. of Knot Theory and Its Ramifications **2**, 413–425 (1993).

[RW] van Rensburg, E.J.J., Whittington, S.G.: The Knot Probability in Lattice Polygons. J. Phys. A.: Math. Gen. **23**, 3573–3590 (1990).

[SSW] Soteros C., Sumners D.W., Whittington S.G.: Entanglement complexity of graphs in Z^3, Math. Proc Camb. Phil. Soc. **111**, 75–91 (1992).

[Jun] Jungreis, D.: Gaussian Random Polygons are Globally Knotted. J. of Knot Theory and Its Ramifications **4**, 455–464 (1994).

[Ken] Kendall:The knotting of brownian motion in 3-space. J. Lon. Math. Soc. **19**, 378– (1979).

[DT1] Deguchi, T., Tsurasaki, K.: Universality of Random Knotting. Phys. Rev. E. **55**, 6245–6248 (1997).

[Nak] Nakanishi, Y.: A Note on Unknotting Number. Math. Sem. Notes Kobe Univ. **9**, 99–108 (1981).

[Ful] Fuller, B.: The Writhing Number of a Space Curve. Proc. Nat. Acad. Sci. USA **68**, 815–819 (1971).C

[LS] Lacher, R.C., Sumners D.W.: Data structures and algorithms for computation of topological invariants of entanglements: link, twist and writhe. In Computer Simulation of Polymers, Prentice-Hall, Roe, R.J., ed. Englewood Cliffs, NJ, 365–373 (1991).

[Cim] Cimasoni, D.: Computing the Writhe of a Knot. J. Knot Theory and Its Ramifications **10**, 387–395 (2001).

[LaS] Laing, C, Sumners D.W.: Computing the Writhe on Lattices. J. Phys A, Math. Gen. **39**, 3535–3543 (2006).

[ROS] van Rensburg, E.J.J., Orlandini, E., Sumners, D.W., Tesi. M.C., Whittington, S.G.: The writhe of a self-avoiding polygon. J. Phys. A Math. Gen. **26**, L981–L986 (1993).

[BZ] Burde, G., Zieschang, H.: Knots. De Gruyter, Berlin, New York (1985).

[Rol] Rolfsen, D.: Knots and Links. Publish or Perish (1976).

[RCV] Rybenkov, V. V., Cozzarelli, N. R. & Vologodskii, A. V.: Probability of DNA Knotting and the Effective Diameter of the DNA Double Helix. Proc. Natl. Acad. Sci. USA **90**, 5307–5311 (1993).

[ShW] Shaw, S. Y., Wang, J. C.: Knotting of a DNA chain during ring closure. Science **260**, 533–536 (1993).

[WC] Wasserman, S. A., Cozzarelli, N. R.: Biochemical topology: applications to DNA recombination and replication. Science **232**, 951–960 (1986).

[SBS] Stark, W.M., Boocock, M.R., Sherratt, D. J.: Site-specific recombination by Tn3 resolvase. Trends Genet. **5**, 304–309 (1989).

[FLA] Frank-Kamenetskii, M.D., Lukashin, A.V., Anshelevich, V.V., Vologodskii, A.V.: Torsional and bending rigidity of the double helix from data on small DNA rings. J. Biomol. Struct. Dyn. **2**, 1005–1012 (1985).

[ES] Ernst, C., Sumners, D.W.: A Calculus for Rational Tangles: Applications to DNA Recombination. Math. Proc. Camb. Phil. Soc. **108**, 489–515 (1990).

[SKI] Shishido, K., Komiyama, N., Ikawa, S.: Increased production of a knotted form of plasmid pBR322 DNA in *Escherichia coli* DNA topoisomerase mutants. J. Mol. Biol. **195**, 215–218 (1987).

[LDC] Liu, L. F., Davis, J. L., Calendar, R.: Novel topologically knotted DNA from bacteriophage P4 capsids: studies with DNA topoisomerases. Nucleic Acids Res. **9**, 3979–3989 (1981).

[LPC] Liu, L. F., Perkocha, L., Calendar, R.. Wang, J. C.: Knotted DNA from Bacteriophage Capsids. Proc. Natl. Acad. Sci. USA **78**, 5498–5502 (1981).

[MML] Menissier, J., de Murcia, G., Lebeurier, G., Hirth, L.: Electron microscopic studies of the different topological forms of the cauliflower mosaic virus DNA: knotted encapsidated DNA and nuclear minichromosome. EMBO J. **2**, 1067–1071 (1983).

[Man] Mansfield, M. L.: Knots in Hamilton Cycles. Macromolecules **27**, 5924–5926 (1994).

[TRO] Tesi, M. C., van Resburg, J.J.E., Orlandini, E., Whittington, S. G.: Knot probability for lattice polygons in confined geometries. J. Phys. A: Math. Gen **27**, 347–360 (1994)

[MMO] C. Micheletti, C, Marenduzzo, D., Orlandini, E., Sumners, D.W.: Knotting of Random Ring Polymers in Confined Spaces. J. Chem. Phys. **124**, 064903 (2006).

[TAV] Trigueros, S, Arsuaga, J., Vazquez, M.E., Sumners, D.W., Roca, J.: Novel display of knotted DNA molecules by two-dimensional gel electrophoresis. Nucleic Acids Research **29**, 67–71 (2001).

[AVT] J. Arsuaga, M. Vazquez, S. Trigueros, D.W. Sumners and J. Roca, Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids, Proc. National Academy of Sciences USA **99** (2002), 5373–5377.

[ArT] Arsuaga, J., R. Tan, K-Z, Vazquez, M.E., Sumners, D.W., Harvey, S.C.: Investigation of viral DNA packing using molecular mechanics models. Biophysical Chemistry **101–102**, 475–484 (2002).

[AVM] Arsuaga, J., Vazquez, M.E., McGuirk, P., Sumners, D.W., Roca, J.: DNA Knots Reveal Chiral Organization of DNA in Phage Capsids. Proc. National Academy of Sciences USA **102**, 9165–9169 (2005).

[EC] Earnshaw, W. C., Casjens, S. R.: DNA packaging by the double-stranded DNA bacteriophages. Cell **21**, 319–331 (1980).

[RHA] Rishov, S., Holzenburg, A., Johansen, B.V., Lindqvist, B.H.: Bacteriophage P2 and P4 Morphogenesis: Structure and Function of the Connector. Virology **245**, 11–17 (1998).

[STS] Smith, D.E., Tans, S.J., Smith, S.B., Grimes S., Anderson, D.L., Bustamante, C.: The bacteriophage $\varphi 29$ portal motor can package DNA against a large internal force. Nature **413**, 748–752 (2001).

[KCS] Kellenberger, E., Carlemalm, E., Sechaud, J., Ryter, A., Haller, G.: Considerations on the condensation and the degree of compactness in non-eukaryotic DNA: in Bacterial Chromatin, eds. Gualerzi, C. & Pon, C. L. (Springer, Berlin), p. 11 (1986).

[SB] San Martin, C., Burnett, R.: Structural studies on adenoviruses. Curr. Top. Microbiol. Immunol. **272**, 57–94 (2003).

[SDD] Schmutz, M., Durand, D., Debin, A., Palvadeau, Y., Eitienne, E.R., Thierry, A. R.: DNA packing in stable lipid complexes designed for gene

transfer imitates DNA compaction in bacteriophage. Proc. Natl. Acad. Sci. USA **96**, 12293–12298 (1999).

[ACT] Aubrey, K., Casjens, S., Thomas, G.: Secondary structure and interactions of the packaged dsDNA genome of bacteriophage P22 investigated by Raman difference spectroscopy. Biochemistry **31**, 11835–11842 (1992).

[EH] Earnshaw, W. C., Harrison, S.: DNA arrangement in isometric phage heads. Nature **268**, 598–602 (1977).

[LDB] Lepault, J., Dubochet, J., Baschong, W., Kellenberger, E.: Organization of double-stranded DNA in bacteriophages: a study by cryo-electron microscopy of vitrified samples. EMBO J. **6**, 1507–1512 (1987).

[HMC] Hass, R., Murphy, R.F., Cantor, C.R.: Testing models of the arrangement of DNA inside bacteriophage λ by crosslinking the packaged DNA. J. Mol. Biol. **159**, 71–92 (1982).

[Ser] Serwer, P.: Arrangement of double-stranded DNA packaged in bacteriophage capsids : An alternative model. J. Mol. Biol. **190**, 509–512 (1986).

[RWC] Richards, K., Williams, R., Calendar, R.: Mode of DNA packing within bacteriophage heads. J. Mol. Biol. **78**, 255–259 (1973).

[CCR] Cerritelli, M., Cheng, N., Rosenberg, A., McPherson, C., Booy, F., Steven, A.: Encapsidated Conformation of Bacteriophage T7 DNA. Cell **91**, 271–280 (1997).

[BNB] Black, L., Newcomb, W., Boring, J., Brown, J.: Ion Etching of Bacteriophage T4: Support for a Spiral-Fold Model of Packaged DNA. Proc. Natl. Acad. Sci. USA 82, 7960–7964 (1985).

[Hud] Hud, N.: Double-stranded DNA organization in bacteriophage heads: an alternative toroid-based model. Biophys. J. **69**, 1355–1362 (1995).

[JCJ] Jiang, W., Chang, J., Jakana, J., Weigele, P., King, J., Chiu, W.: Strucure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. Nature **439**, 612–616 (2006).

[WMC] Wang, J.C., Martin, K.V., Calendar, R.: Sequence similarity of the cohesive ends of coliphage P4, P2, and 186 deoxyribonucleic acid. Biochemistry **12**, 2119–2123 (1973).

[LDC] Liu, L.F., Davis, J.L., Calendar, R.: Novel topologically knotted DNA from bacteriophage P4 capsids: studies with DNA topoisomerases. Nucleic Acids Res. **9**, 3979–3989 (1981).

[WMH] Wolfson, J.S., McHugh, G.L., Hooper, D.C., Swartz, M.N.: Knotting of DNA molecules isolated from deletion mutants of intact bacteriophage P4. Nucleic Acids Res. **13**, 6695–6702 (1985).

[IJC] Isaken, M., Julien, B., Calendar, R., Lindgvist, B.H.: Isolation of knotted DNA from coliphage P4: in DNA Topoisomerase Protocols, DNA Topology, and Enzymes, eds. Bjornsti, M. A. & Osheroff, N. (Humana, Totowa, NJ), Vol. **94**, pp. 69–74 (1999).

[RDM] Raimondi, A. Donghi, R., Montaguti, A., Pessina, A., Deho, G.: Analysis of spontaneous deletion mutants of satellite bacteriophage P4. J. Virol. **54**, 233–235 (1985).

[RUV] Rybenkov, V.V., Ullsperger, C., Vologodskii, A.V., Cozarelli, N.R.: Simplification of DNA Topology Below Equilibrium Values by Type II Topoisomerases. Science **277**, 690–693 (1997).

[MRR] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of State Calculations by Fast Computing Machines. J. Chem. Phys. **21**, 1087–1092 (1953).

[Mil] Millet, K.: Knotting of Regular Polygons in 3-Space. In Series of Knots and Everything, eds. Sumners D.W. & Millet K.C. (World Scientific, Singapore), Vol. **7**, pp. 31–46 (1994).

[MS] Madras, N., Slade, G.: The Self-Avoiding Walk, Birkhauser, Boston (1993).

[DT2] Deguchi, T., Tsurusaki, K.: A Statistical Study of Random Knotting Using the Vassiliev Invariants. J. Knot Theor. Ramifications **3**, 321–353 (1994).

[SKB] Stasiak, A., Katrich, V., Bednar, J., Michoud, D., Dubochet, J.: Electrophoretic mobility of DNA knots. Nature **384**, 122 (1996).

[VCL] Vologodskii, A.V., Crisona, N.J., Laurie, B., Pieranski, P., Katritch, V., Dubochet, J., Stasiak, A.: Sedimentation and electrophoretic migration of DNA knots and catenanes. J. Mol. Biol. **278**, 1–3 (1998).

[KKS] Kanaar, R., Klippel, A., Shekhtman, E., Dungan, J., Kahmann, R., Cozzarelli, N.R.: Processive recombination by the phage Mu Gin system: Implications for the mechanisms of DNA strand exchange, DNA site alignment, and enhancer action. Cell **62**, 353–366 (1990).

[WSR] Weber, C., Stasiak, A., De Los Rios, P., Dietler, G.: Numerical Simulation of Gel Electrophoresis of DNA Knots in Weak and Strong Electric Fields. Biophys. J. **90**, 3100–3105 (2006).

[RSW] van Rensburg, J., Sumners, D.W., Whittington, S.G.: The Writhe of Knots and Links, In Ideal Knots, Series on Knots and Everything, eds. Stasiak, A., Katritch, V., Kauffman, L.H. (World Scientific, Singapore), Vol. **19**, 70–87 (1999).

[MM] Marenduzzo, D., Micheletti, C.: Thermodynamics of DNA packaging inside a viral capsid: The role of DNA intrinsic thickness. J. Mol. Biol. **330**, 485–492 (2003).

[MMT] Maritan, A., Micheletti, C., Trovato, A., Banavar, J.: Optimal shapes of compact strings. Nature **406**, 287–289 (2000).

[TK] Tzil, S., Kindt, J.T., Gelbart, W., Ben-Shaul, A.: Nucleic acid packaging of DNA viruses. Biophys. J. **84**, 1616–1627 (2003).

[LL] LaMarque, J.C., Le, T.L., Harvey, S.C.: Packaging double-helical DNA into viral capsids. Biopolymers **73**, 3480–355 (2004).

Reference for this paper:

D.W. Sumners. Random Knotting: Theorems, Simulations and Applications. In Topological Fluid Mechanics, Springer Lecture Notes in Mathematics, R. Ricca, ed. (to appear).