Computational and Mathematical Methods in Medicine
Vol. 00, No. 00, March 2008, 1–14

RESEARCH ARTICLE

DNA Knotting in Spooling Like Conformations in Bacteriophages

J. Arsuaga[ab*] and Y. Diao[c†]

[a]*Department of Mathematics, San Francisco State University, 1600 Holloway Ave., San Francisco, CA 94132, USA*; [b]*Institute for Mathematics and Its Applications, University of Minnesota, 207 Church Street S.E., Minneapolis, MN 55455, USA*; [c]*Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA*

(*Received Feb. 2008*)

A number of idealized models have been proposed to explain the long range organization of the DNA in bacteriophages. However none of these models can account for the distributions of complex knots found when examining DNA extracted from bacteriophage P4 capsids. Furthermore these models do not consider possible chirality biases in the arrangement of the DNA molecule inside the capsid. In this paper we address these two issues by proposing a randomized version of one of the most popular models: the coaxially spooled model. We present analytical and numerical results for the properties of the random polygons (knots) generated using this model. We show that such model can easily generate complex knotted conformations and although it accounts for some chirality of the organization of the DNA molecules inside bacteriophage capsids does ot fully explain the experimental data.

Keywords: random polygons, random knots, bacteriophage, DNA packing, spooling model.

AMS Subject Classification: 57M25, 92Bxx

1. Introduction

All icosahedral bacteriophages with double-stranded DNA genomes pack their chromosomes in a similar manner [13] and are believed to to hold similar DNA arrangements as those found in adeno- and herpes viruses [9, 34] as well as in lipo-DNA complexes used in gene therapy [35]. Most bacteriophages consist of an icosahedral capsid that contains the nucleic acid (i.e. DNA or RNA), a tail with tail fibers, and a connector, that allows the assembly of the capsid and the tail [33]. During phage morphogenesis a proteinic procapsid is first assembled [44]. This is followed by the transfer of one copy of the viral genome inside the capsid through the connector (e.g. [37]). At the end of the packing process the volume of the DNA molecule is reduced 10^2 times [19] reaching concentration levels of 800mg/ml [20] and an osmotic pressure of 70 atm [14, 41]. A number of models have been proposed to explain the global organization of DNA inside phage capsids. These include the ball of string model [32], the coaxial and concentric spooling models [3, 10, 13, 32, 36],

*Corresponding author. Email: jarsuaga@sfsu.edu Partially supported by NIH grant NIH-2S06GM52588-12.

†Partially supported by NSF grant DMS-0712958.

ISSN: 1748-670X print/ISSN 1748-6718 online

DOI: 10.1080/1748670YYxxxxxxxx
http://www.informaworld.com

the spiral-fold model [7], the folded toroidal model [18], the twisted toroidal model [30] and liquid crystalline models (e.g. [22]). Spooling and toroidal models are the most accepted in the biological community [3, 10, 13, 32, 33, 36] however liquid crystalline models are somewhat more consistent with local properties of the DNA molecule under extreme conditions of pressure [38] and have been proposed for chromosome organization in other viruses [9, 34]. Here we investigate if spooling like conformations are consistent with data obtained from bacteriophage P4 capsids.

Bacteriophage P4 has an icosahedral capsid and its genome is a double-stranded DNA genome, 10 to 11.5 Kb in length, flanked by 16 bp single-stranded ends that are complementary in sequence [44]. The packaged genome is preserved as a linear molecule [31] by keeping at least one of its ends attached to the connector region[11]. However a high percentage of DNA molecules extracted from bacteriophage P4 are closed circular knots. Furthermore most of these knots have an elevated knot complexity. This phenomenon is further magnified in P4 tailless mutants [4, 23, 24] and in naturally occurring deletion mutants [45]. In [4] it was shown that the majority of the knots are formed inside the viral capsid and therefore may be used as reporters for the DNA organization inside the capsid. In fact, analysis of knots of up to 7 crossings, revealed a very distinct distribution of knots. Some of the characteristics include the absence of the four crossing knot (called the figure eight knot in knot theory) and a prevalence of toroidal knots (i.e. knots that can be constructed on the surface of a torus) over twist knots [5, 39]. These findings suggested a chiral organization of the viral genome [5].

Current idealized models have rarely been confronted with these data and if so they have been unable to reproduce such knot distributions. For instance in [3, 21] molecular dynamics simulations of the coaxially and concentrically spooling models were unsuccessful in obtaining knotted conformations. On the other hand, models that explain the knotting under these conditions of confinement [2, 4, 5, 27, 28] show very unrealistic conformations for the viral genome and are far from those models proposed by other groups. This discrepancy between DNA packing models and knotting of DNA in P4 phages motivates this study.

In this paper, we present a randomized spooling model and investigate the knotting and chirality of the generated configurations. First we introduce the essential mathematical concepts needed for understanding this paper and briefly describe the previously published experimental results. Second we describe the *spooling random polygon*. Next we provide some analytical estimates of the complexity and of the relationship between the spooling of a random polygon with its knotting probability. Finally we perform numerical studies to compute the knotting probability and complexity.

2. Methods

In this section we will first provide some basic mathematical background on knot theory. We will then give a detailed description of our model. After that, we will talk about what mathematical quantities we would like to estimate and how we will carry out our simulations to compute these quantities.

2.1. *Basic Mathematical Background*

The most relevant topological concepts used in this paper are given in this subsection. For a more detailed exposition of the basics of knot theory please refer to a standard text such as [1, 25].

A *knot* K is a simple closed curve in $\mathbf{R}^3$. In this paper we assume that such a curve is a piece-wise smooth curve (this includes a space polygon without self-intersections). For a fixed knot K, a *regular projection* of K is a projection of the simple closed curve onto a plane such that no more than two segments of K cross at the same point in the projection. If the over/under information of the strands at each intersection is kept, then the projection is called a *knot diagram*. A fundamental theorem in knot theory states that two knots K_1 and K_2 are *topologically equivalent* if and only if any regular diagram of K_1 can be changed to a regular diagram of K_2 through a sequence of *Reidemeister moves*, as illustrated in Figure 1. An intersection in a knot diagram is called a *crossing*. The minimum number of

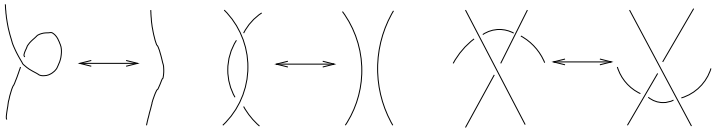

Figure 1. Reidemeister moves

crossings among all possible knot diagrams that are topologically equivalent to K is called the *crossing number* of K and is denoted by $Cr(K)$. The average number of crossings over all possible regular projections of K (as a fixed space curve) is called the *average crossing number* of K and is denoted by $\mathrm{ACN}(K)$. Note that $Cr(K)$ is a knot invariant since it is the minimum number of crossings among all possible knot diagrams that are topologically equivalent to K, but not $\mathrm{ACN}(K)$, as it is a quantity that would be affected by a geometric (not topological) change of K.

An important concept in this paper is that of chirality. A knot K is said to be *achiral* if it is topologically equivalent to its mirror image, otherwise it is said to be *chiral*. For example, the right-handed trefoil is topologically distinct from its mirror image (which is the left-handed trefoil). On the other hand, the only four crossing knot 4_1 (also called the figure eight knot) is achiral. If each crossing in a diagram of K is assigned a $+1$ or -1 following the sign convention shown in Figure 2, then the summation of these ± 1's is called the *writhe* of this diagram. For example, the writhe of the minimum projection of the knot 4_1 is zero, whereas

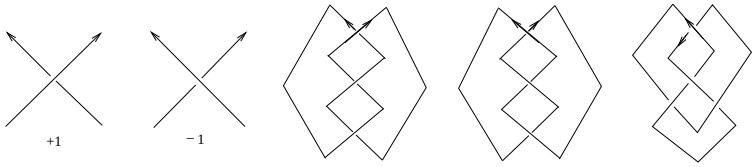

Figure 2. The crossing sign convention and the projections of the right-handed, left-handed trefoils and the figure eight knot whose writhes are 3, −3 and 0 respectively.

the writhe of the minimum projection of the right-handed trefoil is $+3$ (see Figure 2). The *average writhe* of the knot K (as a space curve), on the other hand, is the average of the writhe over all its regular projection diagrams. In the literature the average writhe is often just called writhe. However in this paper we will make a distinction to avoid the possible confusion. Therefore, the writhe always refers to

a single knot projection and the average writhe of K is the average of the writhe over all regular diagrams of K. If the writhe were a knot invariant, one would easily conclude that the writhe of an achiral knot projection must be zero since the writhe of a knot diagram changes sign upon reflection. Unfortunately this is not the case. For example the trivial knot can have a writhe as high as one wants since a simple super-coil is an unknot and can have as many twists as one wants (which gives a writhe as high as one wants). However, the writhe still provides a measure for the chirality of a knot in the average sense: if two random knots were to be generated by the same generating method but with fixed knot types so that one of them is chiral and the other is achiral, then the mean of the average writhe of the achiral random knots would tend to be closer to zero. In this sense one may argue that the chirality of a (random) knot can be quantified by its writhe. Furthermore, the impact of writhe on the chirality of a knot is more visible for knots with certain structure, as we shall see from Theorem 3.3 and its consequences. An important concept connecting the writhe with the knottedness of K (which is used in Theorem 3.3) is that of *Seifert circles*. If every crossing in a knot diagram is split as shown at the left side of Figure 3, then the result is a collection of disjoint (topological) circles called the *Seifert circles* of the knot diagram, see Figure 3.

Figure 3. Splitting a knot diagram into Seifert circles.

2.2. *Existing Experimental Data*

This study is motivated by the data previously published in

[4, 5, 23, 24, 39, 40, 45]. In these experiments bacteriophages were grown in E-coli cells lysogenic for P2. Tailless mutants were grown on a second strain, also lysogenic for P2, deficient for the genes encoding for the tail and tail fibers. Tailless mutant were purified and DNA was phenol extracted and chloroform purified. DNA samples were run in high resolution two dimensional gel electrophoresis revealing a large quantities of knots (95%). Quantification of the knot populations of up to seven crossings revealed the absence of four crossings knots, as well as a (yet unidentified) population of six and seven crossing knots. The data also showed a prevalence of toroidal knots for five and seven crossing knots and the absence of five crossing twist knots.

2.3. *Definition of the coaxially spooled random polygon model*

In this model the DNA is modeled as an equilateral random polygon and the capsid is modeled by a sphere S of radius R (in bond length units). In this sphere there is a cylinder of radius r running through the center. Vertices of the polygon are not allowed to enter the cylinder. The addition of this cylinder in our model is motivated by the core proteins in the capsid (e.g. [30]) and it is also used as an "ad-hoc" parameter to control the rigidity of the polygons when generating spooling conformations. In order to generate toroidal conformations two spherical caps of height $R - \sqrt{R^2 - (r+1)^2}$ are also defined. Polygons are not allowed to enter these caps. See Figure 4 for an illustration of of the model and of the spherical

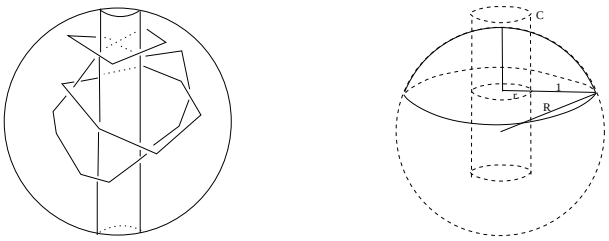

Figure 4. Left: A random spooling DNA packing model. Right: The cap to be avoided by the vertices of the random polygon.

caps. In order to generate a random polygon satisfying the above conditions the following steps are followed. A coordinate system is set so that the original point is the center of S and the z-axis is the axis of T. For simplicity, we may always choose the starting point whose projection is on the x-axis; and without loss of generality, we will use anticlockwise spin packing. Name the vertices of the polygon to be X_0, X_1, X_2, ..., X_n. Once X_k has been generated, we will follow the following procedure to generate X_{k+1}. Let the z-coordinate of X_k be z_k and consider the plane parallel to the xy-plane. It intersects S at a circle C_k whose radius r' is at least $r+1$. Let $r_k = \sqrt{x_k^2 + y_k^2}$ where $X_k = (x_k, y_k, z_k)$. Let us try to determine $r_{k+1} = \sqrt{x_{k+1}^2 + y_{k+1}^2}$ first. Here we will introduce the third parameter h, which is a small positive number. For the time being, let us assume that it is a number set between 0.1 and 0.5. Basically, a larger h allows more variation in the polygon in its distance to the center of the cylinder as well as in the difference of the z-coordinates between the end points of each edge. We consider the following three cases.

Case 1. $r + h \le r_k \le r' - h$. In this case, we will choose r_{k+1} from the interval $[r_k - h, r_k + h]$ uniformly.

Case 2. $r_k < r + h$. In this case, we will choose r_{k+1} from the interval $[r, r + 2h]$ uniformly.

Case 3. $r_k > r' - h$. In this case, we will choose r_{k+1} from the interval $[r' - 2h, r']$ uniformly.

Once r_{k+1} is determined, we will go ahead to determine z_{k+1}. This is done similarly to the above in three different cases. First set

$$t_k = \min\{\sqrt{R^2 - (r+1)^2}, \sqrt{R^2 - r_k^2}\}.$$

Case 1′. $-t_k + h \le z_k \le t_k - h$. In this case, we will choose z_{k+1} from the interval $[z_k - h, z_k + h]$ uniformly.

Case 2′. $z_k > t_k - h$. In this case, we will choose z_{k+1} from the interval $[t_k - 2h, t_k]$ uniformly.

Case 3′. $z_k < -t_k + h$. In this case, we will choose r_{k+1} from the interval $[-t_k, -t_k + 2h]$ uniformly.

Now we just need to find how much rotation we need to take from the ray $\overrightarrow{O_k X_k}$ (where $O_k = (0, 0, z_k)$) in the anticlockwise direction (which will then completely determine the coordinates of X_{k+1}). Under the condition that $|X_k X_{k+1}| = 1$, that angle is given by

$$\theta_k = \cos^{-1}\left(\frac{r_k^2 + r_{k+1}^2 + z^2 - 1}{2r_k r_{k+1}}\right),$$

where $z = z_{k+1} - z_k$. See Figure 5 for an illustration of this process.

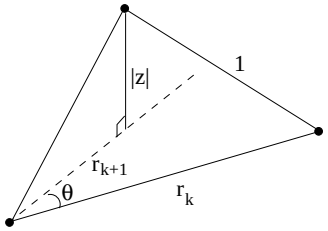

Figure 5. Generating X_{k+1} given X_k.

And we have

$$x_{k+1} = r_{k+1}\cos(\sum_{1\le j\le k}\theta_j),\ y_{k+1} = r_{k+1}\sin(\sum_{1\le j\le k}\theta_j).$$

(Keep in mind that X_1 started on the x-axis.) Figure 6 shows a random polygon generated by this algorithm. A polygon with a few segments is shown to illustrate the natural entanglement of such configurations.

We call the polygons generated in the above process *spooling random polygons*. A spooling random polygon with n edges is denoted by P_n^s.

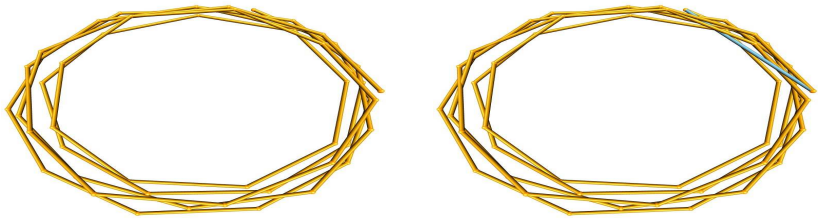

Figure 6. Left: A coaxially spooled equilateral random walk. Right: A forced closure of the random walk at left by a straight line segment. This figure is generated using Knotplot.

2.4. *Determination of the Rounding Number*

In this model, we can easily control the number of times the polygon will go around the z-axis (the rounding number). Say we want to generate a random polygon that goes around the z-axis m times, then we will stop the generating process at the first n such that $\sum_{1\le j\le n}\theta_j \ge 2m\pi$ and replace that last generated point X_{n+1} with X_1, resulting therefore in a polygon. The last edge of the polygon is no longer of unit length. Although it is conceivable that for a large polygon this probably won't affect its topological property that much, a biologically meaningful way to close the two open ends may still be needed for a more in depth study in the future. In this paper, we will trade off this imperfection for simplicity and speed.

2.5. *Computation of the Knotting Probability and Complexity*

Each polygon is projected along the z-axis. The Dowker code of the resulting diagram is computed and it is simplified using Reidemeister I and II moves. From the simplified Dowker code the unknotted trajectories are identified by computing the Alexander polynomial evaluated at $t = -1$ (denoted by $\Delta(-1)$). This identification is based on the fact that the trivial knot has $\Delta(-1) = 1$. Doing so can cause some error: it is known that $\Delta(-1)$ does not identify all knotted polygons (i.e., $\Delta(-1) = 1$ for some non-trivial knots). However, this error is not large since relatively few knots have the property $\Delta(-1) = 1$. The advantage of using $\Delta(-1)$ (or more generally the Alexander polynomial $\Delta(t)$ itself) is that the computation of the polynomial is equivalent to the computation of the determinant of certain matrix, which can be done in polynomial time on the number of crossings. Identification of the knots with low crossing number was done by calculating $\Delta(-2)$ and $\Delta(-3)$.

2.6. *Computation of the Average Writhe and Average Crossing Number*

The average writhe is a geometric measure of self-entanglement which measures non-planarity of a spatial conformation. The average writhe of a closed curve C can be defined using the Gauss double integral

$$\frac{1}{4\pi}\int_0^L\int_0^L \frac{(\dot\gamma(t),\dot\gamma(s),\gamma(t)-\gamma(s))}{|\gamma(t)-\gamma(s)|^3}dtds, \tag{1}$$

where γ is an equation of C parameterized by arclength, L is the length of C and $(\dot\gamma(t),\dot\gamma(s),\gamma(t)-\gamma(s))$ is the triple scalar product of $\dot\gamma(t)$, $\dot\gamma(s)$, and $\gamma(t)-\gamma(s)$. For a given projection, the sign of the crossing is determined using the convention as illustrated in Figure 2. At a positive crossing, one can think of the integrand as the measure of the solid angle created by those projections from which the tangent vectors appear to cross. If the crossing is a negative one, the integrand is equal to the measure of this solid angle multiplied by -1. This interpretation of the integrand allows for an easy calculation of the writhe when C is a polygon of n segments of unit length. In this case, the Gauss integral in (1) can be written as a double sum over the segments of the polygon:

$$\sum_{1\le i\ne j\le n}\frac{1}{4\pi}\int_0^1\int_0^1 \frac{(\dot\gamma_i(t),\dot\gamma_j(s),\gamma_i(t)-\gamma_j(s))}{|\gamma_i(t)-\gamma_j(s)|^3}dtds, \tag{2}$$

where γ_i and γ_j are arclength parameterized equations of the edges ℓ_i and ℓ_j of C. Similarly, the average crossing number (ACN) of C is given by [16]

$$\frac{1}{4\pi}\int_0^L\int_0^L \frac{|(\dot\gamma(t),\dot\gamma(s),\gamma(t)-\gamma(s))|}{|\gamma(t)-\gamma(s)|^3}dtds. \tag{3}$$

And again, in the case that C is a polygon of n segments of unit length, its ACN can be expressed as

$$\sum_{1\le i\ne j\le n}\frac{1}{4\pi}\int_0^1\int_0^1 \frac{|(\dot\gamma_i(t),\dot\gamma_j(s),\gamma_i(t)-\gamma_j(s))|}{|\gamma_i(t)-\gamma_j(s)|^3}dtds. \tag{4}$$

3. Theoretical Results

The following theorem concerns the mean crossing number of a spooling random polygon.

Theorem 3.1: *Let P_n^s be a spooled random polygon of n edges, then the average number of crossings in its projection to the xy-plane perpendicular to its center axis is of the order of $O(n^2)$.*

Proof: Here is an outline of the proof of this lemma. Let ℓ_1, ℓ_2, ..., ℓ_n be the consecutive edges of P_n^s and let $a_{ij} = a(\ell_i, \ell_j)$ be the average crossing number between ℓ_i and ℓ_j, then the ACN of P_n (written as χ_n) is simply the sum of the $a(\ell_i, \ell_j)$'s: $\chi_n = \sum_{1\le i<j\le n} a_{ij} = \frac{1}{2}\sum_{1\le i,j\le n} a_{ij}$, and $E(\chi_n) = \sum_{1\le i<j\le n} E(a_{ij}) = \frac{1}{2}\sum_{1\le i,j\le n} E(a_{ij})$, where $a_{ij}/2$ has the integral form as given in (4). For any $j > i+1$ and any fixed ℓ_i, there is a positive probability that the starting point of ℓ_j is of a small (constant) distance away from ℓ_i by the way ℓ_j is generated. Since ℓ_i and ℓ_j are confined in the same sphere of radius R, ℓ_j cannot be too far away from ℓ_i either. So the denominator in the integrand of (4) is bounded above by a constant. Since the end point of ℓ_j is independent of the position of ℓ_i, there is a positive probability that the numerator of the integrand of (4) is bigger than a constant (however small it may be). It follows that $E(a_{ij} \ge b$ for some positive constant b as long as ℓ_i and ℓ_j are not adjacent edges. This implies that $E(\chi_n) = \frac{1}{2}\sum_{1\le i,j\le n} E(a_{ij}) \ge bn(n-3)/2 = O(n^2)$. □

Remark. It must be pointed out that the positive constant b mentioned in the above proof can vary greatly depending on the parameters as shown in Figure 9 in the case of $h = 0.1$ and $h = 0.3$. But the $O(n^2)$ behavior is clearly seen in the numerical study as shown in Figure 9.

Let us now concentrate on the projection of P_n^s onto the xy-plane. Call this projection D_n. Essentially D_n is treated as a knot diagram and we can compute $w(D_n)$, the writhe of D_n. Because of the special structure of P_n^s, we have the following result as well. The proof of this is elementary. An interested reader can convince him/her self by working out a few examples.

Theorem 3.2: *Let D_n be the projection of a P_n^s onto the xy-plane and let $\sigma(P_n^s)$ be the rounding number of P_n^s (namely the number of times wraps around its central axis), then the number of Seifert circles in D_n is exactly $\sigma(P_n^s)$.*

The following important theorem is due to Morton. The original theorem was stated for a general diagram using the number of Seifert circles in the place of $\sigma(P_n^s)$. For the original version of the theorem and a proof of it, please refer to [29].

Theorem 3.3: *[29] Let $w(D_n)$ be the writhe of D_n, e be the smallest power of v in the HOMFLY polynomial $P_{D_n}(v,z)$ of D_n (which is the same as the HOMFLY polynomial of P_n^s, of course), and E be the largest power of v in $P_{D_n}(v,z)$, then we have $w(D_n) - (\sigma(P_n^s) - 1) \le e \le E \le w(D_n) + (\sigma(P_n^s) - 1)$.*

An immediate consequence of this theorem is the following important corollary. It reveals a very nice relation between the writhe of D_n and the chirality of P_n^s.

Corollary 3.4: *[29] If $|w(D_n)| \ge \sigma(P_n^s)$, then P_n^s is a non-trivial knot. Furthermore, in this case P_n^s cannot be an achiral knot.*

Proof: We have $e = 0 = E$ for the unknot (by the definition of the HOMFLY polynomial). Yet if $|w(D_n)| \ge \sigma(P_n^s)$, then either $e \ge 1$ or $E \le -1$. On the other hand, if P_n^s is an achiral knot, then it is equivalent to its mirror image.

Since $P_K(v,z) = P_{K'}(v^{-1},z)$ when K' is the mirror image of K, it follows that $P_{D_n}(v,z) = P_{D_n}(v^{-1},z)$. Therefore $e = -E$ so we must have $e \leq 0 \leq E$. But again, if $w(D_n) \geq \sigma(P_n^s)$, we have $1 \leq e$ and if $w(D_n) \leq -\sigma(P_n^s)$, we have $E \leq -1$, leading to contradictions in both cases. Thus P_n^s has to be a chiral knot. □

This corollary clear shows that if P_n^s is knotted because of its high writhe (as initially proposed in [5]), then it simply cannot be an achiral knot. So if the DNA in a bacteriophage capsid is packed in a coaxially spooled way with certain writhe bias, then one would observe a large proportion of knots and none of these knots would be achiral. In extreme cases one would expect the figure eight knot to be missing.

Theorem 3.3 can be strengthened to the following theorem, which relates the writhe of D_n to the complexity of P_n^s.

Theorem 3.5 : *Let $w(D_n)$ be the writhe of D_n such that $|w(D_n)|-\sigma(P_n^s) = q \geq 0$, then P_n^s is a non-trivial knot with minimum crossing number at least $(q+2)/2$.*

Proof : Let us consider the case $w(D_n) \geq \sigma(P_n^s) + q$. In this case, P_n^s must be a non-trivial knot by Corollary 3.4. Let us look a minimum diagram D (which of course is likely to be different from D_n). It is obvious that $w(D) \leq Cr(D)$. Furthermore, we must have $s(D) \leq Cr(D)$ where $s(D)$ is the number of Seifert circles of D. Otherwise D would be a simple diagram with only twists that can be untwisted using Reidemeister I move repeatedly. Applying Theorem 3.3 to D_n and D separately, we have $w(D_n) - \sigma(P_n^s) + 1 \leq e \leq E \leq w(D) + s(D) - 1$. It follows that $q+1 \leq 2Cr(D)-1$. That is, $Cr(D) \geq (q+2)/2$ where $Cr(D)$ is the minimum crossing number of P_n^s since D is a minimal projection of P_n^s by our choice. The case of $w(D_n) \leq -(n+q)$ can be similarly handled and is left to the reader as an exercise. □

Remark. It is important for us to point out here that the writhe we talked about in the above discussion refers to the writhe of the diagram of the random polygon (which is the projection of the polygon along the z-axis), not the average writhe of that polygon. Theorem 3.3 does not apply if we use the average writhe instead, because if we project the polygon into a plane that is not parallel to the xy-plane, then the number of Seifert circles in that diagram may no longer be the same as the rounding number of the polygon. Therefore, in our simulation study, we would only study the mean of the writhe of the diagram of the polygons into the xy-plane, not the mean average writhe. But our result for the crossing number is for the mean average crossing number (mean ACN).

4. Numerical Results

As mentioned in the methods section our model has three parameters R, r and h, where R and r are the radii of the confining sphere and cylinder respectively and h is the parameter that controls the degree of randomness of the polygon at each step and therefore the rigidity of the chain. In order to illustrate the results of the simulations two values of h are used (i.e. $h = 0.1$ and $h = 0.3$). The values of R and r are also fixed and results for $R = 10$ and $r = 1$ are shown. For simplicity the starting point for all polygons is set to be $(1.5, 0, 0)$. These studies are only our initial efforts to try to understand the model and how the parameters affect our results. Routine studies with systematic testings on these parameters will be performed in the future.

4.1. *The Mean Radius of Gyration of the Spooling Random Polygons: the Interplay between h and R*

In order to test the effects of the parameter h (as defined in section 2.3) on the dimensions of the knot we calculated the mean radius of gyration of spooling random polygons. Figure 7 below shows our results. Not surprisingly, polygons generated with $h = 0.1$ have conformations tighter than those generated with $h = 0.3$. More importantly even for $n = 200$, the mean radius of gyration for the $h = 0.1$ case is still far away from $R = 10$ (below 3), indicating that most of these polygons are still tightly wrapped around the central cylinder and that the confining sphere did not really play an important role here. However, in the case of $h = 0.3$ the mean radius of gyration is more than 8.5 when $n = 200$, indicating that the polygons could have gone beyond $R = 10$ if not confined by the sphere.

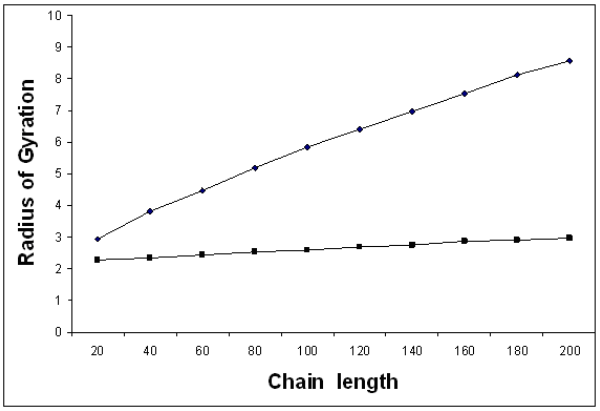

Figure 7. The mean radius of gyration of the spooling random polygons: The bottom curve is for $h = 0.1$ and the top one is for $h = 0.3$.

4.2. *The Knotting Probability of the Spooling Random Polygons*

As expected from the tighter structures found for $h = 0.1$, the knotting probability increased faster in the case $h = 0.1$ than in the case $h = 0.3$. Remarkably and in high contrast with other random polygon models the knotting probability rapidly saturates (as it is observed experimentally). However, in contrast with the results published in [8, 45] larger polygons show higher knotting probability than shorter ones. When $w(D_n)$ is large, but not over $\sigma(P_n^s)$, how does it affect the knotting probability of P_n^s? We suspect that the knotting probability would be much higher for a spooling random polygon P_n^s with a fixed large writhe $w(D_n)$. A detailed numerical study on this is planned for the future.

4.3. *The Mean Average Crossing Number of P_n^s*

The mean average crossing number helps to estimate the complexity of the knots in the average sense [42]. Knots extracted from bacteriophage P4 have been estimated to have an average of 30 crossings [4]. Here we computed the mean of the average crossing number before Reidemeister moves were removed, therefore a much higher number is expected. Figure 9 shows the result of our numerical study on the mean crossing number of P_n^s (as a function of n). Polynomials of degree two are fitted to the data as it is suggested by Theorem 3.1. Interestingly the mean average crossing number is heavily dependent on the parameter h, suggesting that more complex knots are found for smaller h when the rest of the parameters are held fixed. This is consistent with our numerical result on knotting probability shown in Figure 8.

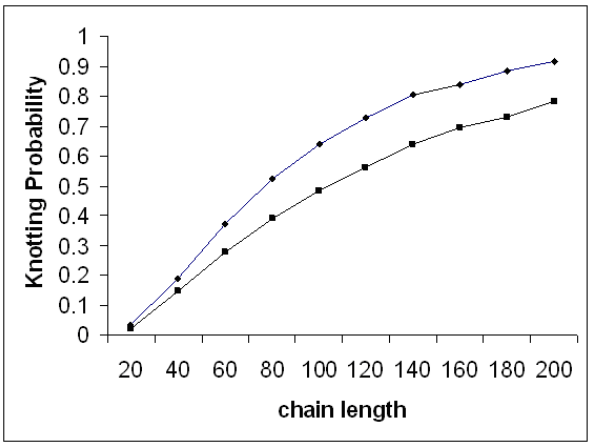

Figure 8. Knot Probability of the random spooling model.

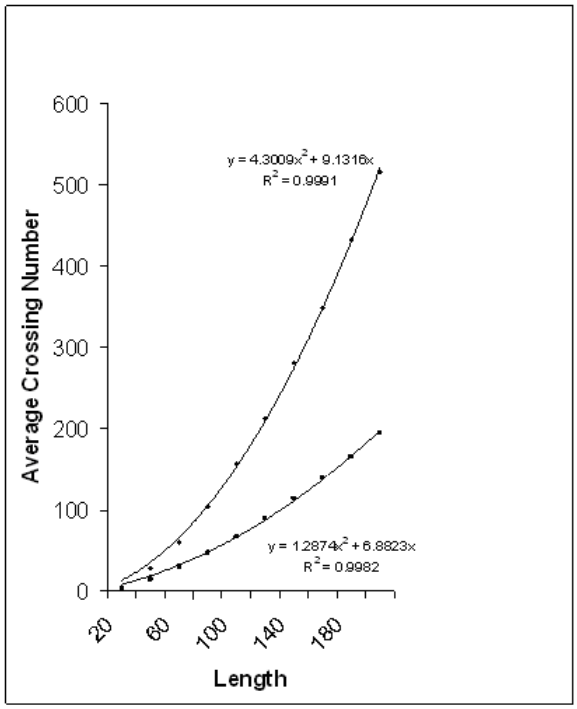

Figure 9. The mean ACN of P_s^n with n up to 200.

4.4. *Knot Distribution of the Spooling Random Polygons*

Next we estimated the distribution of knots for up to six crossings. The result for chains of 50 segments and $h = 0.1$ are shown in Table 1. Interestingly and in contrast with completely random polygons the distribution shows an elevated number of five crossing knots with respect to the four crossing knots. It is also important to highlight that (also in contrast with random models) the knot 5_1 is more probable than its twist counterpart. Therefore the spooling model presents features that are observed experimentally. However the match between these results and the results observed experimentally is not perfect. For example, the figure eight knots do occur (although with a relatively low percentage). But this should not surprise us, since we did not take out the polygons with small writhe values and Corollary 3.4 only guarantees us that there will be no achiral knots (such as the

figure eight knot) if the absolute writhe of the polygon is large than its rounding number. In our case, the mean absolute writhe is calculated to be about 1.6 which is much smaller than the mean rounding number, which is larger than 5.

Knot	0_1	3_1	4_1	5_1	5_2	6_1	6_2	6_3	≥ 7
Probability	0.64	0.157	0.012	0.017	0.016	0.001	0.002	0.001	0.07

Table 1. Knot distribution up to 6 crossings in the case of $h = 0.1$, $n = 50$.

5. Conclusions

In this paper, we have introduced a new computational model to investigate DNA knots extracted from bacteriophage P4. This model assumes an initial biased conformation suggested by current models of DNA packing in bacteriophages, and it incorporates a degree of randomness. How much randomness there is in the packing of the genome is a matter of discussion [12, 26]. This is an effort to bring two highly deviated approaches: total randomness or complete organization. As we pointed out in the introduction section, although both types of idealized models (total randomness or complete organization) have their own merit, they fail to produce numerical results that accurately reflect all aspects of DNA knots extracted from bacteriophage P4. Therefore, it is important for us to explore new models such as the one proposed here.

Our preliminary study of the random spooling polygons showed that this model can indeed produce numerical results that can partially describe the knotting probability and the knot distribution observed experimentally. Furthermore, because of the special structure of these polygons we are able to carry out some rigorous mathematical analysis. These analytical results give an important support and meaningful interpretation of our numerical results.

Further explorations of the random spooling polygon model will be needed in order for us to understand it better. More in depth numerical studies will be needed. As we pointed out earlier, we will need to explore the effects of the parameters used in the model in more detail. A big task will be to determine the effect of excessive writhe on the knotting probability.

Additionally, it may also be important for us to impose further conditions on the model due to new geometric or topological restrictions observed experimentally. For example, our current model lacks volume exclusion, however a number of studies have shown that electrostatics play an important role in the packing reaction and in the organization of the chromosome (e.g. [17]). Also, in this study we made several assumptions that will need to be investigated in the future. First we have introduced an internal cylinder as an 'ad-hoc' parameter and also as a model for the core proteins. Second our polygons are obtained by (artificially) adding a new segment to close the open equilateral chains generated by our method, this may produce artifacts. More realistic rejoining of the chain ends should be investigated.

While some existing theoretical results in knot theory can be nicely applied to the random spooling polygon model as we have demonstrated in Section 3, the model remains a challenge for further mathematical analysis. For instance, if we remove the requirement that the random polygons spool around a central cylinder in the confining capsid, then how can we generalize Theorem 3.3 and its corollaries (in Section 3)? On the other hand, can we use the average writhe of the polygon

instead of the writhe of a particular projection? The reason is that in practice, it is hard to determine what projection plane is the one we need. These are the mathematical issues that we will explore in the future.

References

[1] C. Adams, *The Knot Book: An elementary introduction to the mathematical theory of knots*. Revised reprint of the 1994 original, American Mathematical Society, Providence, RI, 2004.

[2] J. Arsuaga et al., *Sampling Large Random Knots in a Confined Space*, J. Physics A **40** (2007), pp. 11697–11711.

[3] J. Arsuaga et al., *Investigation of viral DNA packaging using molecular mechanics models*, Biophys. Chem. **101** (2002), pp. 475–484.

[4] J. Arsuaga et al., *Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids.* Proc. Natl. Acad. Sci. USA **99** (2002), pp. 5373–5377.

[5] J. Arsuaga et al., *DNA knots reveal a chiral organization of DNA in phage capsids*, Proc. Natl. Acad. Sci. USA **102** (2005), pp. 9165–9169.

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

[7] L. Black et al., *Ion etching bacteriophage T4: support for a spiral-fold model of packaged DNA*, Proc. Natl. Acad. Sci. USA **82** (1985), pp. 7960–7964.

[8] T. Blackstone et al., *The role of writhe in DNA condensation*, Proceedings of International Workshop on Knot Theory for Scientific Objects. OCAMI Studies Volume **1** (2007). Osaka Municipal Universities Press; pp. 239–250.

[9] F.P. Booy et al., *Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus.* Cell. textbf64(1991), pp.1007–1015.

[10] K. Cerritelli et al., *Encapsidated conformation of bacteriophage T7 DNA*, Cell **91** (1997), pp. 271–280.

[11] D.K. Chattoraj and R.B. Inman, *Location of DNA ends in P2, 186, P4 and lambda bacteriophage heads*, J. Mol. Biol. **87** (1974), pp. 11–22.

[12] L. R. Comolli, A. J. Spakowitz, C. E. Siegerist, P. J. Jardine, S. Grimes, D. L. Anderson, C. Bustamante and K. H. Downing, *Three-dimensional architecture of the bacteriophage phi29 packaged genome and elucidation of its packaging process*, Virology, in press.

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

[14] A. Evilevitch et al., *Osmotic pressure inhibition of DNA ejection from phage*, Proc. Natl. Acad. Sci. USA **100** (2003), pp. 9292–9295.

[15] G. Gouesbet, S. Meunier-Guttin-Cluzel and C. Letellier, *Computer evaluation of Homfly polynomials by using Gauss codes, with a skein-template algorithm*, Appl. Math. Comput. **105** (1999), pp. 271–289.

[16] M.H. Freedman and Z. He, *Divergence-free Fields: Energy and Asymptotic Crossing Number*, Annals of Math. **134** (1991), pp. 189–229.

[17] D. N. Fuller, J. P. Rickgauer, P. J. Jardine, S. Grimes, D. L. Anderson, D. E. Smith. Ionic effects on viral DNA packaging and portal motor function in bacteriophage phi 29. (2007) Proc Natl Acad Sci U S A. **104**, (2007) **11245-11250**.

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

[19] P. J. Jardine and D. L. Anderson *DNA packaging in double-stranded DNA phages* The bacteriophages (2006), Ed. Richard Calendar, Oxford University Press, pp. 49–65

[20] E. Kellenberger et al., *Considerations on the condensation and the degree of compactness in non-eukaryotic DNA-containing plasmas*, In Bacterial Chromatin: Proceedings of the Symposium "Selected Topics on Chromatin Structure and Function" (eds. C. Gualerzi and C. L. Pon), Springer, Berlin (1986), pp. 11–25.

[21] J.C. LaMarque, T.L. Le and S.C. Harvey, *Packaging double-helical DNA into viral capsids*, Biopolymers **73** (2004), pp. 348–355.

[22] J. Lepault et al., *Organization of double-stranded DNA in bacteriophages: a study by cryo-electron microscopy of vitrified samples.* EMBO J. **6** (1987), pp. 1507–1512.

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

[24] L.F. Liu et al., *Knotted DNA from bacteriophage capsids*, Proc. Natl. Acad. Sci. USA **78** (1981), pp. 5498–5502.

[25] C. Livingston, *Knot theory*, Carus Mathematical Monographs **24**, AMS (1993).

[26] C. R. Locker, S. D. Fuller, S. C. Harvey. *DNA organization and thermodynamics during viral packing*, Biophys J. **93** (2007) pp. 2861–2869.

[27] C. Micheletti et al., *Knotting of random ring polymers in confined spaces*, J. Chem. Phys. **124** (2006), pp. 064903.1 –64903.10.

[28] K. Millett, *Knotting of regular polygons in 3-space*, Random knotting and linking (Vancouver, BC, 1993), World Sci. Publishing, Singapore (1994), pp. 31–46.

[29] H.R. Morton, *Seifert circles and knot polynomials*, Math. Proc. Cambridge Phil. Soc. **99** (1986), pp. 107–109.

[30] A.S. Petrov, M.B. Boz and S.C. Harvey, *The conformation of double-stranded DNA inside bacteriophages depends on capsid size and shape*, J Struct Biol. **160** (2007) pp. 241–248.

[31] G.R. Pruss, J.C. Wang and R. Calendar, *In vitro packaging of satellite phage P4 DNA*. Proc. Natl. Acad. Sci USA **71** (1974) pp. 2367–2371.

[32] K. Richards, R. Williams and R. Calendar, *Mode of DNA packing within bacteriophage heads*, J. Mol. Biol. **78** (1973), pp. 255–259.
[33] S. Rishov et al., *Bacteriophage P2 and P4 morphogenesis: structure and function of the connector*, Virology **245** (1998) pp. 11–17.
[34] C. San Martin and R. Burnett, *Structural studies on adenoviruses*, Curr. Top. Microbiol. Immunol. **272** (2003), pp. 57–94.
[35] M. Schmutz et al., *DNA packing in stable lipid complexes designed for gene transfer imitates DNA compaction in bacteriophage*, Proc. Natl. Acad. Sci. USA **96** (1999), pp. 12293-12298.
[36] P. Serwer, *Arrangement of double-stranded DNA packaged in bacteriophage capsids: An alternative model*, J. Mol. Biol. **190** (1986), pp. 509–512.
[37] D.E. Smith et al., *The bacteriophage phi29 portal motor can package DNA against a large internal force* Nature **413** (2001), pp. 748–752.
[38] E. L. Starostin, *On the perfect hexagonal packing of rods*, J. Phys.: Condens. Matter **18** (2006), pp. S187–S204.
[39] S. Trigueros et al., *Novel display of knotted DNA molecules by two dimensional gel electrophoresis.* Nucleic Acids Research **29** (2001), e67.
[40] S. Trigueros and J. Roca, *Production of highly knotted DNA by means of cosmid circularization inside phage capsids*, BMC Biotechnol **7**(1) (2007), pp. 94.
[41] S. Tzill et al., *Forces and Pressures in DNA Packaging and Release from Viral Capsids*, Biophys. J. **84** (2003), pp. 1616–1627.
[42] A. V. Vologodskii, N. J. Crisona, B. Laurie, P. Pieranski, V. Katritch, J. Dubochet and A. Stasiak, *Sedimentation and electrophoretic migration of DNA knots and catenanes*, J. Mol. Biol. **278** (1998), pp. 1–3.
[43] S. Wang, J.R. Chang and T. Dokland *Assembly of bacteriophage P2 and P4 procapsids with internal scaffolding protein.* Virology **348** (2006) pp. 133–40.
[44] J.C. Wang, K.V. Martin and R. Calendar, *On the sequence similarity of the cohesive ends of coliphage P4, P2, and 186 deoxyribonucleic acid*, Biochemistry **12** (1973), pp. 2119–2123.
[45] J.S. Wolfson et al., *Knotting of DNA molecules isolated from deletion mutants of intact bacteriophage P4*, Nucleic Acids Res. **13** (1985), pp. 6695–6702.