Properties of Topological Networks of Flexible Polygonal Chains

Research Article

J. Arsuaga[1*], Y. Diao[2†], M. Klingbeil[3‡], V. Rodriguez[4§]

1 Department of Mathematics
San Francisco State University
San Francisco, CA 94116, USA

2 Department of Mathematics and Statistics
University of North Carolina at Charlotte
Charlotte, NC 28223, USA

3 Department of Microbiology
University of Massachusetts
Amherst, MA 01003, USA

4 Department of Physics and Astronomy
San Francisco State University
San Francisco, CA 94116, USA

Abstract: Trypanosomatida parasites, such as trypanosoma and leishmania, are the cause of deadly diseases in many third world countries. The three dimensional structure of their mitochondrial DNA, known as kinetoplast DNA (kDNA), is unique since it is organized into several thousands of minicircles that are topologically linked. How and why the minicircles form such a network have remained unanswered questions. In our previous work we have presented a model of network formation that hypothesizes that the network is solely driven by the confinement of minicircles. Our model shows that upon confinement a percolating network forms. This network grows into a space filling network, called saturating network, upon further confinement of minicircles. Our model also shows, in agreement with experimental data, that the mean valence of the network (that is, the average number of minicircles topologically linked to any minicircle in the network) grows linearly with minicircle density. In our previous studies we disregarded DNA flexibility and used rigid minicircles to model DNA minicircles, here we address this limitation by allowing minicircles to be flexible. Our numerical results show that the topological characteristics that describe the growth and topology of the minicircle networks have similar values to those observed in the case of rigid minicircles suggesting that these properties are robust and therefore a potentially adequate description of the networks observed in trypanosomatid parasites.

MSC: 57M25, 92B99

Keywords: kDNA • minicircles • minicircle networks • random minicircle networks • linking

* *E-mail: jarsuaga@sfsu.edu*
† *E-mail: ydiao@uncc.edu*
‡ *E-mail: klingbeil@microbio.umass.edu*
§ *E-mail: torrodri@mail.sfsu.edu*

1. Introduction

Trypanosomes are parasites that cause fatal diseases in humans and livestock. African trypanosomiasis, caused by *Trypanosoma brucei (T. brucei)*, alone kills over 30,000 people in Africa every year [1]. One important characteristic of trypanosomes is the three dimensional organization of their mitochondrial DNA, known as kinetoplast DNA (kDNA), into several thousands of minicircles that are topologically linked forming a gigantic chainmail-like network (reviewed in [2]). Some topological properties of kDNA have been determined for *Crithidia fasciculata*, a model organism whose network is believed to mirror that of *T. brucei*. For example, DNA minicircles in *C. fasciculata* are confined in a cylindrically shaped region of the mitochondrion called kinetoplast disk [3] where the concentration of DNA is similar to that found in the bacterial nucleoid [2]. Surprisingly, DNA minicircles are not supercoiled but relaxed [4] and topologically linked by a single interlock (called the *Hopf link*) [4]. Finally, the average number of minicircles linked to any given minicircle, called the mean minicircle valence, is cell cycle dependent. The mean minicircle valence is three before replication and grows linearly the concentration of minicircles during replication[5, 6]. How and why the minicircles form a network in these organisms and not in other evolutionary related species remain unanswered questions.

In order to understand the principles behind this complex organization DNA molecules, we proposed a new theoretical approach to study the formation of topological networks (networks of topologically linked circles) under the assumption that confinement of minicircles in the kinetoplast disk is the main driving force of network formation [7–10]. In our model, rigid geometrical minicircles are placed on a two dimensional lattice where the density of minicircles is an adjustable parameter. Our studies consistently show the existence of two key minicircle densities in the formation of the network: a critical percolation density D_{perc} and a mean saturation density D_{sat}; furthermore they showed a linear relationship between the density of minicircles and their mean valence.

In this study we extend the model presented in [8] (where rigid geometrical circles were used to model minicircles) by modeling minicircles using freely joined polygonal chains and determine how the relevant densities and the mean valence of the network depend on chain flexibility. First we estimate the critical percolation and the mean saturation densities and second we determine the relationship between minicircle density and mean minicircle valence. As one would expect, given the random structure of the minicircles, both the critical percolation density and the mean saturation density are higher than those obtained when rigid geometrical minicircles were used. However these values are still fairly small compared with those observed in trypanosomes. Our findings help us quantify the effects of chain flexibility on network formation and support the robustness of our previously reported results.

In the next section, we will provide the necessary definitions and terminology, as well as some basic mathematical background. In Sections 3, we present numerical methods and results and in Section 4, we discuss our results and future directions of our work.

2. Definitions, terminology and background

In our previous work we introduced the Square Lattice Minicircle (SLM) model [8, 9] to measure the effects of volume confinement on minicircle network formation and on the topological properties of the network. In this model the kinetoplast disk is represented as a region of the simple square lattice (although other lattices have been used [9]) of size $d \times d$ called $d \times d$ *lattice grid*. DNA minicircles are modeled by rigid geometrical circles; this simplification is based on the experimental observation that minicircles are linked by a single interlock (i.e. by a *Hopf link*) and this is the only possible link with rigid minicircles (See Figure 1). In the SLM model minicircles are randomly oriented, that is with their normal vector uniformly chosen over the unit sphere and their center of masses placed at each lattice point of the grid. We call a lattice grid with minicircles at its vertices *minicircle grid*. If a is the distance between two adjacent lattice points in the minicircle grid, then the density of minicircles in the network is given by $1/a^2$, since minicircle density is the number of minicircles per unit area.

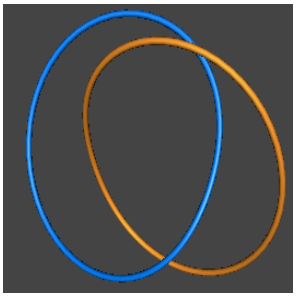

Figure 1. **A Hopf link is the only possible link for two rigid minicircles**

A SLM minicircle network may contain thousands or even tens of thousands of minicircles. The topology is simple if one only examines the relationship between any pair of minicircles: they either form a trivial link or a Hopf link. However, how can we characterize the topology of the entire network? To do so, we introduce the concept of a *minicircle linked cluster*. We say that a set of minicircles form a linked cluster if none of these minicircles can be separated from the others without by breaking at least one of the minicircles. Figure 2 shows an example of a minicircle linked cluster and an example of minicircles not forming a linked cluster.

Intuitively a minicircle network with most or even all of the minicircles in a single linked cluster is more stable structurally than a minicircle network in which only small linked clusters exist (since these unlinked clusters can be separated more easily leading to the disintegration of the network). Here we introduce a mathematical concept that is not as intuitive, but is more robust. A linked cluster in a minicircle network is said to be a *percolating cluster* if the cluster connects opposite boundaries of the minicircle grid. We say that a minicircle network is *stable* if it contains a percolating cluster. Loosely speaking, a percolating cluster serves as a backbone that supports the overall structure of the network. Of course, the denser the minicircles are in a network, the higher the chance for the network to be stable. In our previous studies based on the SLM model we have shown that there exists

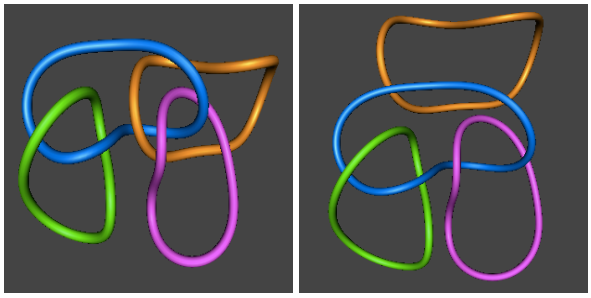

Figure 2. **Left: A linked cluster; Right: Minicircles that do not form a linked cluster.**

a critical density (called *critical percolation density*) such that once the minicircle density exceeds this density, a minicircle network will become stable with a probability that quickly approaches one as the density increases. Similarly below this critical density, the probability of getting a stable network quickly approaches zero as the density decreases. The critical percolation density is an important topological characteristic of the network and can be estimated computationally due to the rapid phase change of the network stability around it.

Figure 3 shows the critical percolation densities for different grid sizes. The critical percolation density D_{perc} is estimated as the limit of those measured for finite lattices.

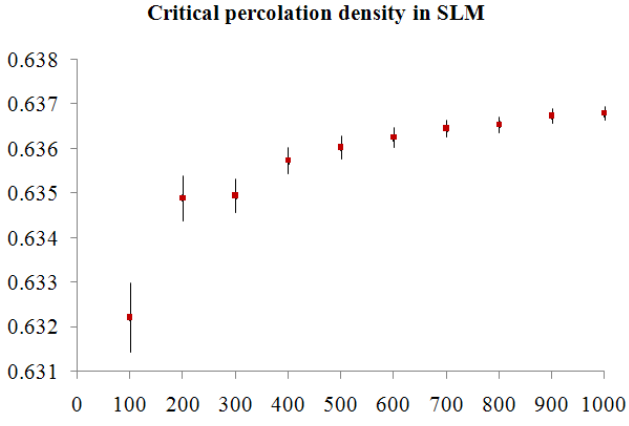

Figure 3. **Estimation of critical percolation densities as a function of the lattice size in the SLM model.**

The structure of kDNA suggests that the minicircles that fill the kinetoplast disk actually all belong to the same linked cluster. We call a minicircle network in which all minicircles belong to the same linked cluster a *saturating minicircle network*. The average density at which the network becomes saturated is called the mean *saturating density*. In this paper and our previous papers, we only require that 99% or more of minicircle to belong to the same linked cluster for a network to be considered saturated in order to avoid some computation complications

along the boundaries of the networks that would drive the mean saturation density estimation artificially high. Figure 4 shows the saturation density as a function of the network grid size.

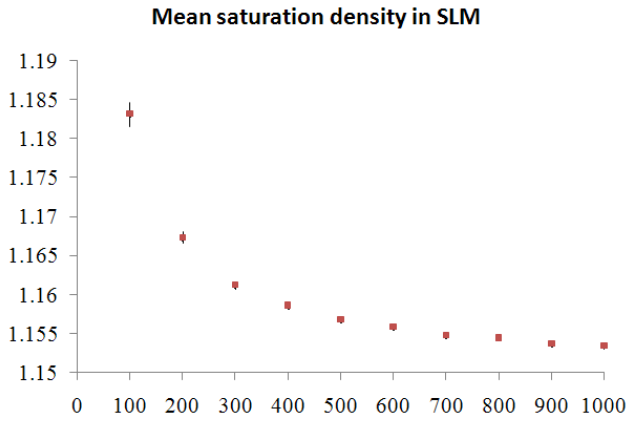

Figure 4. **Estimation of average saturation densities as a function of the lattice size in the SLM model.**

A third parameter of interest is the *mean minicircle valence* which is defined by the average number of minicircles linked to a given minicircle in the network. This value has been experimentally measured and shown to be equal to 3 in *C. fasciculata* before DNA replication and 6 after replication [5, 6]. Our results are in agreement with these experimental as shown in Figure 5 were a linear relation between minicircle density and valence is shown.

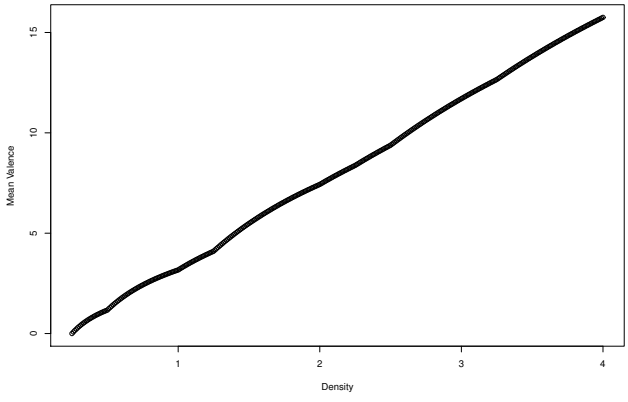

Figure 5. **The SLM model predicts a linear relationship between minicircle density and mean valence**

In [5] the authors suggested that chain flexibility should be a key factor determining the properties of the kDNA network. This problem has, in fact, been extensively investigated in DNA replication and segregation by Vologodskii and Cozzarelli [11, 12] but never in network formation. In the work presented here we allow minicircles to be flexible and study how the properties of the network change. Since flexible closed chains (scaled to be of

total length 2π) have a mean radius of gyration less than 1 (the radius of unit circle), one expects that the linking probability between two such closed chains whose centers of mass are of a distance d away is less that the linking probability between two rigid geometrical minicircles (unit circles) whose centers are of the same distance d apart. This lower linking probability (verified by our simulation) may have significant effects on the properties of the network and is the main subject of study in this paper.

In this work we model flexible minicircles using freely jointed polygons (also known as equilateral random polygons). This model is somewhat an extreme case. Rather than providing an accurate picture of the biological problem, it provides reference values for the properties of topological networks whose components are highly flexible. It is well known that even freely jointed polygons with a small number of edges can be non-trivially knotted. In this study we do not reject knotted minicircles but instead consider an unbiased representation of polygons for any fixed polygon length. Also we expect links to be more complicated than simple Hopf links, such as those shown in Figure 6 and observed in previous studies [13].

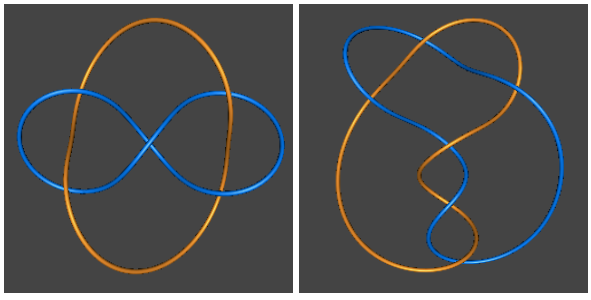

Figure 6. **From the left to the right: A Whitehead link and a more complicated link.**

3. Simulation methods

3.1. Generation of random freely jointed polygons

We generated freely jointed polygons using the generalized hedgehog algorithm [14] as implemented in Knotplot. The generalized hedgehog algorithm is a Monte Carlo method that generates independent polygons in a "stepwise uniform" manner. The algorithm is ergodic and its time complexity grows linearly with the length of the polygon. It was recently brought to the attention of the authors by J. Cantarella that the first step of the original algorithm description needs a small modification. To generate the initial polygon with four edges with vertices $X_0 = O$, X_1, X_2 and X_3, we will choose X_2 first according to its theoretical distribution. If we let $r = |X_2|$, then its density function is $1/2$ for $r \in [0, 2]$ and 0 otherwise; thus one can choose $r = |X_2|$ according to this density and then choose X_2 uniformly on the sphere centered at O with radius r. After that, X_1 and X_3 can be chosen uniformly on the intersection circle of the unit spheres centered at O and X_2. Using this algorithm a data set of

10,000 conformations for each analyzed polygonal length was generated.

3.2. Generation of minicircle grids and networks

We generated minicircle grids of size $d \times d$ with polygons of fixed number of edges n and total length 2π as follows. For each lattice point in the grid, an equilateral polygon with n edges is assigned. This polygon is randomly selected from the generated data set of conformations and its center of mass is superimposed to the lattice point. Once the minicircle is positioned, a random direction is selected and the conformation rotated about this direction by a random angle. Once all the $d \times d$ polygons are positioned, the topology of the minicircle grid is determined by calculating the linking number between minicircles that could potentially be linked to each polygon.

3.3. Link determination

As in previous studies we estimated the linking of two polygons by computing their linking number. If the linking number between two polygons is not zero, then the two polygons are non-trivially linked. In this case the two polygons are said to be homologically linked. There exist nontrivial links whose linking number is zero so there are polygons that are topologically linked but not homologically linked. Trying to determine topologically linked polygons would require the use of various knot invariant polynomials, which is beyond the scope of this paper. Instead, we focused our investigation to $P(Lk \neq 0)$, since $P(Lk \neq 0)$ is a lower bound for the actual linking probability. The linking number can be computed following the Gaussian integral formulation (see for example [15]).

4. Simulation Results

All calculations were performed for polygons with $n = 16$, 18 and 20 edges and for grids of dimensions 100×100, 200×200 and 300×300. Samples of 100 minicircle grids were generated for each grid dimension, density value and polygon edge number. The total length of each polygon is scaled to 2π so that results obtained can be compared to our previous results with rigid unit circles.

4.1. Distribution of Linking Numbers

Since flexible chains may produce links different from the Hopf link, our first study was aimed at quantifying the distribution of linking numbers for different densities and polygon lengths within our range of study. Our results show that relatively few adjacent polygon pairs have absolute linking number greater than 1. For example, in the case that the grid dimension is 100×100 and the polygons have $n = 16$ edges, for densities ranging from 3.08 to 3.84 (around the critical percolation density at that grid size), we found that less than 1 percent of adjacent polygon pairs have absolute linking number larger than 1 (details are shown in Figure 7). These percentages

only increased slightly when the density increases to values around the mean saturating density. This implies that most of the linking pairs we encounter are indeed Hopf links. Notice that for the ranges of densities we considered, polygons that are not immediate neighbors on the lattice can be linked and the data provided in Figure 7 is obtained over all polygon pairs where linking is possible. These percentages would be much higher if only neighboring polygons are considered.

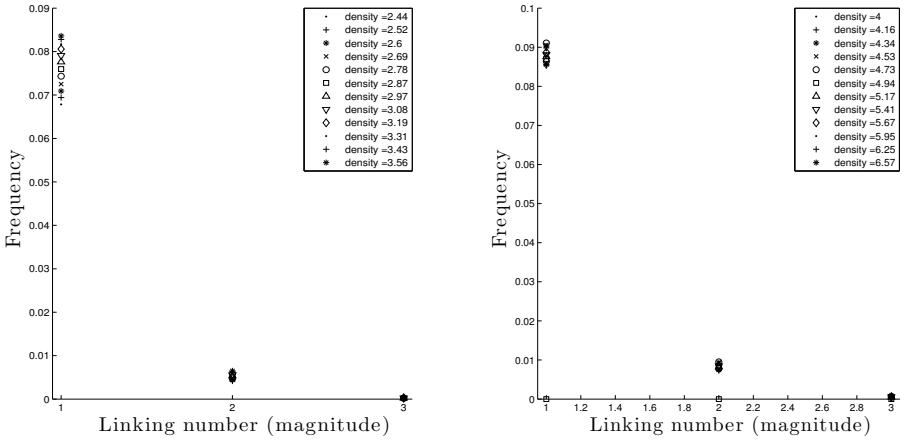

Figure 7. Distributions of absolute linking number for densities around the critical percolation density (left) and the mean saturating density (right) for SLEP with $n = 16$ edges and grid dimension 100×100.

4.2. Estimation of Critical Percolation Densities

Next we estimated the critical percolation density. Table 1 shows the estimated critical percolation densities for polygons with 16, 18 and 20 edges at grid dimension of 100×100, 200×200 and 300×300. As expected all cases showed a rapid transition from non-stable to stable networks confirming the existence of a percolation density. The case of $n = 16$ edges is shown in Figure 8 and the other two cases are very similar.

Table 1. **Estimated critical percolation densities for various numbers of edges and grid dimensions.**

	100×100	200×200	300×300
16 edges	3.04 ± 0.06	3.05 ± 0.05	3.10 ± 0.06
18 edges	3.25 ± 0.06	3.26 ± 0.06	3.26 ± 0.06
20 edges	3.40 ± 0.05	3.41 ± 0.04	3.42 ± 0.02

As observed in our earlier studies, the critical percolation density increases (although slowly) with the dimension of the grid. It also increases with the number of edges of the polygon as one would have expected (polygons with more edges have smaller radius of gyration when scaled to a fixed length). Notice that the effects due to the flexibility of the chain (hence smaller overall radii) are very visible in the critical percolation values since that corresponding value in the SLM model (where rigid circles are used for minicircles) at grid dimension 100×100

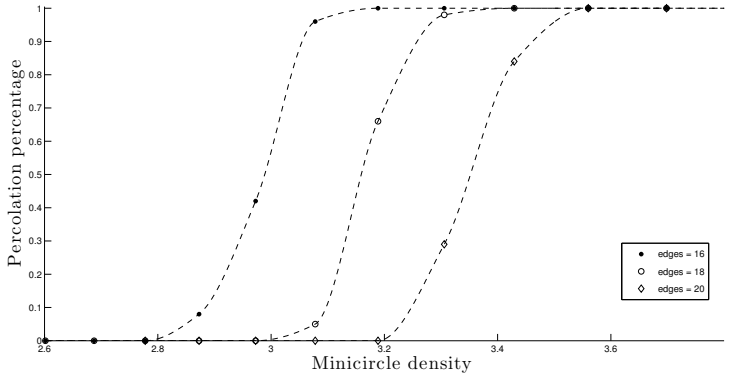

Figure 8. **Estimated critical percolation densities at grid with dimension** 100×100 **with 16, 18 and 20 edges respectively:** 3.04 ± 0.06**,** 3.25 ± 0.06 **and** 3.40 ± 0.05.

is estimated to be around 1.18 [8, 9].

4.3. Estimation of Saturation Densities

Next, we studied the valuea t which saturation was reached. As in the previous case the grid size was $d = 100$ and polygons had 16, 18 or 20 edges. Results are shown in Table 2. As can be observed in the table the estimated saturation densities also increased with polygon edge number.

Table 2. **Mean saturating densities for various numbers of edges and grid dimensions.**

	100×100	200×200	300×300
16 edges	5.66 ± 0.05	5.8 ± 0.1	?
18 edges	6.20 ± 0.1	6.1 ± 0.1	?
20 edges	6.50 ± 0.1	6.5 ± 0.1	?

4.4. Estimation of Mean Minicircle Valence

Figure 9 shows the mean valence of a minicircle as the function of the minicircle density for the three polygon edge numbers 16, 18 and 20. The purpose is to see whether a near linear relationship still exists as in the case of the SLM model, and to see what is the chain flexibility effect on the magnitude of the mean valence. The data is obtained by averaging the valences of the center minicircles in a sample of 1000 7×7 minicircle grids at each density. Notice that overall the near linear relationship still exists but there are a few details that are different from the SLM model. The first one is that for the SLM model, the mean valence is 0 for density below 0.5, since below that density, two adjacent minicircles (which are rigid unit circles) are too far away to be linked. However in the case of the SLEP, two polygons can be linked at a much lower density (though the probability of that would be very small). This explains why the left end of the graph did not simply end at 0 for densities below

0.5 (rather it showed a gradual decay to 0 for densities below 0.5). On the other hand, when density reaches 4, under the SLM model, only the neighboring minicircles can be linked. However under the SLEP model many more non-neighboring minicircle pairs can be linked at that density and the overall effect of this caused the mean valence (more than 20) to surpass the mean valence (around 15) under the SLM model at minicircle density 4.

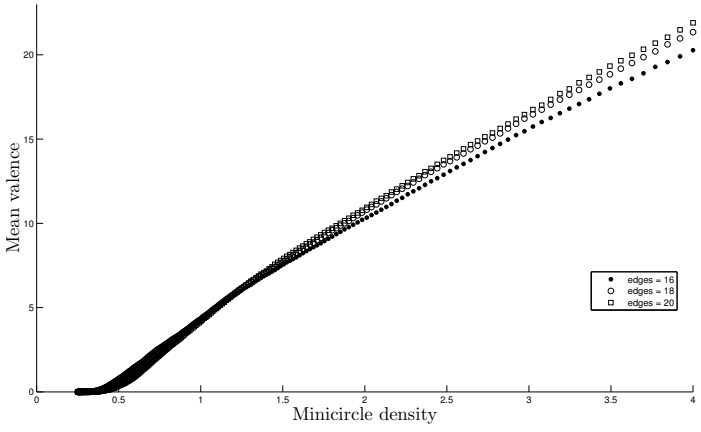

Figure 9. **Mean valence of a minicircle as the function of the minicircle density.**

5. Discussion

Trypanosomes are parasitic organisms that cause diseases, most of them in developing countries. A characteristic feature of these organisms is that part of their mitochondrial DNA is organized into a chain-mail like structure. The origin of this structure has remained elusive to experimental biologists. In this and in our previous we have proposed a mathematical and computational model that helps quantify the effects of confinement on the structure of the network. In our previous studies we had assumed that minicircles were rigid objects. Here we have extended our previous models by allowing minicircles to have some degree of flexibility and for that purpose we have modeled minicircles by freely jointed polygonal chains. Our results are in overall agreement with our previous results. First, upon confinement minicircles form clusters some of which merge into a percolating cluster upon reaching a critical percolation density. As expected and due to chain flexibility percolation values are higher than those estimated for rigid minicircles. Interestingly we also observe that percolation density increases with polygonal length. As in previous studies and motivated by the observed structure of the kDNA in trypanosomes we also estimated the density at which all minicircles are part of the same cluster. In this case we also observe that the densities are higher than those of rigid minicircles (due to chain flexibility). In this case and as one would expect longer chains reach saturation earlier than shorter ones. Finally, in the case of the mean valence, the overall near linear relationship between the mean valence and the minicircle density is still clearly present,

with some minor differences in the details, again caused by the flexibility of the chains as we had explained in the last section. From this and previous studies we can conclude that the properties of minicircle networks are general across models and that it is only the specific values of the densities and valence that change. Since knotted polygons have not been excluded from the study and the flexibility of the chains is much higher than that of DNA we can not provide an accurate picture of the topology of the kDNA networks, this model however gives us values that can be used as reference when chain flexibility is taken into consideration. A more accurate model will be developed in future studies in which more accurate models of DNA will be incorporated and knotted molecules removed from the ensemble.

Acknowledgement

V. Rodriguez is supported by the MBRS-RISE Program 5R25-GM048972 of the National Institutes of Health (to Dr. F. Bayliss) and by NSF grant DMS 0920887, Y. Diao is partially supported by grants DMS-0920880 and DMS-1016460 from National Science Foundation of USA and J. Arsuaga is partially supported by grants DMS 0920887 and DMS 1217324 from National Science Foundation of USA, and by grant 1R01GM109457 from National Institutes of Health. Figures 1, 2 and 6 are generated using the software KnotPlot developed by R. Scharein.

References

[1] World Health Organization. Neglected tropical diseases: Hidden successes, emerging opportunities. In *WHO Library Cataloging-in-Publication Data*, pages 38–39. 2006.

[2] T. A. Shapiro and P. T. Englund. The structure and replication of kinetoplast dna. *Annu. Rev. Microbiol.*, 49:117–143, 1995.

[3] M. Ferguson, A. F. Torri, D. C. Ward, and P. T. Englund. *In situ* hybridization to the *crithidia fasciculata* kinetoplast reveals two antipodal sites involved in kinetoplast dna replication. *Cell*, 70:621–629, 1992.

[4] C. A. Rauch, D. Perez-Morga, N. R. Cozzarelli, and P. T. Englund. The absence of supercoiling in kinetoplast dna minicircles. *EMBO J.*, 12(2):403–411, 1993.

[5] J. Chen, C. A. Rauch, J. H. White, P. T. Englund, and N. R. Cozzarelli. The topology of the kinetoplast dna network. *Cell*, 80(1):61–69, 1995.

[6] J. Chen, P. T. Englund, and N. R. Cozzarelli. Changes in network topology during the replication of kinetoplast dna. *EMBO J.*, 14(24):6339–6347, 1995.

[7] J. Arsuaga, Y. Diao, and K. Hinson. The effect of angle restriction on the topological characteristics of minicircle networks. *J. Statist. Phys.*, 146(2):434–445, 2012.

[8] Y. Diao, K. Hinson, and J. Arsuaga. The growth of minicircle networks on regular lattices. *J. Phys. A: Math.*

Theor, 45:doi:10.1088/1751–8113/45/3/035004, 2012.

[9] Y. Diao, K. Hinson, R. Kaplan, M. Vazquez, and J. Arsuaga. The effects of minicircle density on the topological structure of the mitochondrial dna from trypanosomes. *J. Math. Biol.*, 64(6):1087–1108, 2012.

[10] V. Rodriguez, Y. Diao, and J. Arsuaga. Percolation phenomena in disordered topological networks. *J. Phys.: Conf. Ser.*, 454:doi:10.1088/1742–6596/454/1/012070, 1993.

[11] A. V. Vologodskii and N. R. Cozzarelli. Monte carlo analysis of the conformation of dna catenanes. *J. Mol. Biol.*, 232:1130–1140, 1993.

[12] A. V. Vologodskii. Dna supercoiling helps to unlink sister duplexes after replication. *BioEssays*, 32:9–12, 2010.

[13] A. V. Vologodskii and V. Rybenkov. Simulation of dna catenanes. *Phys. Chem. Chem. Phys.*, 11:10543–10552, 2009.

[14] R. Varela, K. Hinson, J. Arsuaga, and Y. Diao. A fast ergodic algorithm for generating ensembles of equilateral random polygons. *J. Phys. A*, 42(9):1–13, 2009.

[15] M. Spera. A survey on the differential and symplectic geometry of linking numbers. *Milan J. Math.*, 74:139–197, 2006.