UC Davis
UC Davis Electronic Theses and Dissertations

Title

Statistical Biophysics in Global Disease:Liquid-Crystal Energetics in Trypanosome Mitochondrial DNA Knotting andBinding Fitness for the S477N COVID-19 Mutant

Permalink

https://escholarship.org/uc/item/6x6596h6

Author

Polischuk, Ryan Dolan

Publication Date

2021

Peer reviewed|Thesis/dissertation

Statistical Biophysics in Global Disease:
Liquid-Crystal Energetics in Trypanosome Mitochondrial DNA Linking and
Binding Fitness for the S477N COVID-19 Mutant

By

RYAN DOLAN POLISCHUK
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Dr. Rajiv Singh

_____

Dr. Daniel L. Cox

_____

Dr. Javier Arsuaga

Committee in Charge

2021

i

Contents

Statistical Biophysics in Global Disease

Abstract

This dissertation develops physical and mathematical models for the trypanosome parasite's mitochondrial DNA and for SARS-CoV-2 mutagenesis. In the case of the trypanosome parasite, we introduce a liquid-crystal based knotting model to improve upon two earlier, purely mathematical models for minicircle topology that were proposed by Chen, Arsuaga, and their colleagues. In chaps. 1 and 3, we provide a detailed review of these models and their extensions, explaining the biological evidence for the model assumptions and discussing their performance in predicting topological quantities like linking probability and mean valence. Then, in chap. 2, we introduce an orientational energy—the Lebwohl-Lasher model—that reproduces the order of its kinetoplast "minicircles" (mitochondrial DNA plasmids).

We develop a mean-field theory (MFT) for this model, which predicts a field-parallel pseudo-transition in the quadrupolar $\langle Q_0 \rangle$ order parameter. We then expand on this analysis with Markov-chain Monte-Carlo simulations. These results match the MFT predictions for most temperatures and they predict pseudo-transitions at $T_{pc,0}^* = 0.5$ and $T_{pc,2}^* = 1.1$ (for $\langle Q_2 \rangle$). The latter crossover is particularly interesting because it breaks the systems symmetry while remaining consistent with the Hohenberg-Mermin-Wagner theorem.

Our study of SARS-CoV-2 spike (S) mutagenesis addresses the question of whether or not the S477N substitution in the protein's important "receptor-binding domain" provides a binding improvement consistent with its empirically observed fitness advantage in worldwide genomic data. We use 3-D modeling, molecular dynamics simulations, and binding free-energy calculations to show that the S477N substitution likely improves fitness, consistent with the population behavior.

Preface & Dedication

What is worth a grown man's or woman's time?

In 2015 in my last semester of Boston College's Arts & Sciences Honors Program, Professor Mark O'Connor asked our class to read Michel de Montaigne's "On the Education of Children." Prof. O'Connor told me that it was particularly important that I read this essay as an aspiring physicist, because he wanted me to develop an appreciation for intellectual breadth and interdisciplinary just at the moment when I was to begin diving deeply into the technical knowledge of physics.

> *"I was a researcher, and am one still, but I no longer*
>
> *look in the stars or in books. I am starting to listen*
>
> *to what murmers in my own blood."*
>
> —*Demian*, Herman Hesse

Prof. O'Connor's primary message was that our personal development should occur hand-in-hand with the intellectual kind. As a scientist my learning from papers and seminars will continue throughout my career—as will my hopeless attempt to check off every title on his "lifetime reading list," (those Great Books of philosophy, literature, and religion). In line with Prof. O'Connor's advice, even as I have become a Doctor of Philosophy in physics, so too have I taken steps to "learn to live" in this life, as Hemingway puts it in *The Sun Also Rises*. This dissertation is dedicated to those role models who taught me the questions to ask in order to lead a "life well lived," and, indeed, to the same persons, who demonstrated what such lives look like: **Mark O'Connor** and **Jim Keenan**.

Acknowledgments

I received some excellent advice when I was an undergraduate—"Pick the professor not the course."–and I think the same can be said of a Ph.D. advsior and of a lab. I am very fortunate to have had the opportunity to work with Javier Arsuaga for the last five years. He is a kind and supportive man, who is dedicated to his work, while also being well-rounded personally and encouraging the same behavior in his students. Javier has built a lab with co-P.I. Mariel Vazquez in which the complex biology of genome organization and knotting is studied with an appropriately broad set of quantitative tools and full technical rigor. Working with Javier, I have had the opportunity to model myself as "a scientist first; a physicist second," as another of my committee members, Dan Cox, once described himself to me. I am grateful to Javier, Dan, Mariel, and my committee chair, Rajiv Singh, for their mentorship and research guidance.

I have had the pleasure of working with some deeply talented colleagues throughout my degree. At the risk of missing many people, I would like to specifically mention a few folks: Keith Fraga, Anirudha Menon, and Tanat Kissikov for their friendship and excellent scientific discussions; lab-mates like Tamara Christiani, Maxime Pouokam, Robert Stolz, Lara Ibrahim, Allison Moore, Gina Gonzalez, and Michelle Flanner for their day-to-day conversation and material contributions to my work; and, briefly, my good friend Giuseppe Barbalinardo, for conversations on topics technical and personal over an Amtrak coffee or a weekend beer.

CHAPTER 1

Introduction

1.1. An Overview of Kinetoplast Biology

As disease-causing parasites, trypanosomes attracted significant interest among biologists well before their complex mitochondrial DNA was discovered. The invention of transmission-electron microscopy in the 1930s and the development of effective ultra-microtomes in the early 1950s [**PB53**] provided biologists with adequate tools to study the sub-cellular structures of trypanosomes. In 1958 Meyer and colleagues reported the first thin-section images of *Trypanosoma cruzi* [**RN58**], the parasite responsible for Chagas disease. Although the kinetoplast had already been identified via light microscopy, its purpose was still unknown. Meyer and colleagues showed that the kinetoplast contains electron-dense material; combined with a study by Steinert and colleagues in the same year showing the incorporation of ^{3}H thymidine [**SHM58**], biologists concluded that the kinetoplast contains DNA. In the 1960s, Delain showed that kDNA is comprised of circular DNA molecules, which could be topologically linked, or "catenated" [**DG69**, **DP67**].

In *Trypanosoma brucei* and *Trypanosoma cruzi*, kDNA is partitioned into thousands of circular DNA molecules called *minicircles* and several larger *maxicircles*. Maxicircles contain genes associated with mitochondrial function (reviewed in [**JSE08**]) and minicircles transcribe RNA that edits the transcription products of the maxicircles (reviewed in [**Gor08**]). Both minicircles and maxicircles vary in size among organisms and, sometimes, within an organism. In most trypanosomes minicircles range in size from 0.5 kb to 2.5 kb and maxicircles from 20 kb to 40 kb [**SE95**]. Together, minicircles and maxicricles are spatially clustered in a

small cylindrical volume—the kinetoplast discussed earlier—whose dimensions are correlated with the size of the minicircles (reviewed in [**Lea08**]).

1.1.1. Experimental Evidence for Minicircle Topology.

In this section we will discuss the experimental evidence for several key aspects of our model of kDNA topology. These include: *C. fasciculata* minicircles are linked on average to three other circles ("mean valence"); catenated minicircle dimers are limited to simple Hopf links—where the two strands cross over and under just one time; minicircle DNA is relaxed (non-supercoiled); the minicircle network is two-dimensional; and the kinetoplast has no visible structural changes in the absence of maxicircles. These terms

Diao and his colleagues made an important modeling choice by deciding to treat minicircles as rigid, geometrical circles [**DHK$^+$12**]. They believed that this simplification would still allow the model to capture the topological properties of the system. Their choice is supported by the experimental results produced by Rauch and her colleagues in 1993, which indicated that *C. fasciculata* minicircles are not supercoiled and that topological linkages between two circles are of the simplest variety—Hopf links [**Rea93**]. We note that a Hopf link is the only topological linkage available to a pair of rigid circles and that supercoiled plasmids are more likely to follow complex trajectories.

Minicircles are linked via Hopf links.

Rauch and her colleagues determined that minicircles are linked only once by partially digesting kinetoplasts with the restriction enzyme SstII, and then studying the resultant catenated minicircle dimers under electron microscopy. SstII cleaves minicircles a single time, and networks were analyzed after treatment using gel electrophoresis. The authors used RecA-coating to visualize the DNA, and they note that this allows "rigorous evaluation of

strand overlay at crossover points." Ibrahim and colleagues produced results that were consistent with those found by Rauch et al.—a singly linked minicircle dimer from Ibrahim's study is shown in Figure 1.1 [**ILK$^+$18**].

FIGURE 1.1. A singly linked minicircle dimer

Minicircles are not supercoiled.

Prokaryotic DNA and mitochondrial DNA in eukaryotic cells is negatively supercoiled due to excess torsional strain in the double helix. The typical superhelical density of a DNA molecule is about +0.06 (right-handed over-twisting) [**Rea93**]. This means that the molecule wraps about its backbone more than is energetically favorable in a counter-clockwise direction with respect to the backbone. In Rauch's study, the authors estimated the degree of supercoiling in *C. fasciculata* minicircles by comparing minicircles released from the network by sonification with free, ^{3}H-labeled minicircles that had been previously relaxed by topoisomerase-I. The authors separated the products by gel electrophoresis and quantified the amount of kDNA belonging to each oligomer band. The authors used the linking number between the minicircles that were released from the network and the relaxed ^{3}H-labelled minicircles as a control.

3

They found that more than 90% of the unlabeled minicircles (*i.e*, those treated with topoisomerase-I) were either (1) nicked–*i.e*, a single-strand break—or (2) released from the network (as intact circular monomers or linear molecules). They found that the average superhelical density is +0.002, which represents a nearly relaxed right-handed over-twisting.

The mean valence of C. fasciculata minicircles is three.

As we mentioned earlier, Chen and his colleagues determined that the mean valence of *C. fasciculata* minicircles is approximately three. This is a key result for our model's quantitative predictions. The authors used three different analysis methods to determine this property of the kinetoplast, and they determined that the average value is $\angle V \rangle = 2.96$.

First, Chen and colleagues partially digested isolated (purified) *C. fasciculata* kDNA networks with the restriction enzyme Xhol, which cuts each minicircle at a single location. The authors partially digested minicircles into a Tris-borate-EDTA (TBE) gel, which allowed them to fraction the minicircle oligomers. They then extracted the kDNA bands corresponding to dimers and trimers from the gel and visualized the DNA via electron microscopy.

In addition, Chen and his colleagues used two other techniques that are similar to one another to confirm this result. These graphs include the square, hexagonal, triangular planar lattices and several others. Both techniques rely on the seven regular graphs that the authors propose as candidate kDNA topology models. They use these models to make quantitative predictions for the number of minicircle dimers and trimers that would be released from the network as a function of restriction-enzyme concentration. The plots on which the authors compare their experimental data to their theoretical predictions are shown in Fig. 1.2. Results for the fractional release of minicircle dimers agree with the predictions of three of the authors' four valence-three graphs (and with none of the valence-four or valence-six graphs). The trimer data is consistent with all four valence-three graphs, and possibly with one of the valence-four graphs.
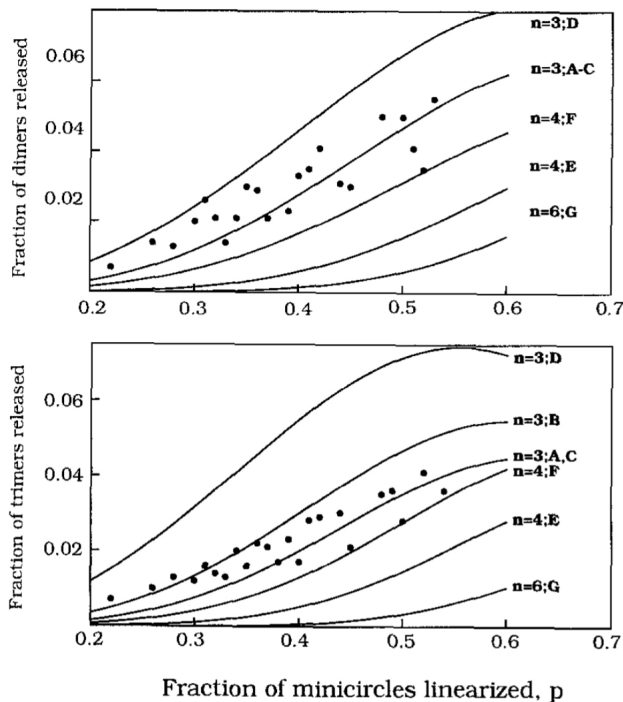
FIGURE 1.2. Mean valence and the fractional release of minicircle oligomers (adapted from [**CRW⁺95**])

C. fasciculata minicircles form a 2-D topological network.

As in Chen's and Diao's models, we treat the kinetoplast in our model as a two-dimensional system. In other words, the centers of all of the minicircles fall within a plane. This assumption is largely accepted and is supported by several pieces of biological evidence. *In-vivo* studies using electron microscopy have determined that the *C. fasciculata* kinetoplast is approximately 0.35 μm in thickness. In one of these studies, Ferguson and his colleagues used thin-section confocal microscopy to measure kinetoplasts stained with acridine orange [**FTWE92**]. Now, the contour length of a *C. fasciculata* minicircle is approximately 833 nm. A geometrical circle of circumference 833 nm has diameter 0.27 μm. Similar calculations led Delain and Riou to predict that the kinetoplast contains a single layer of minicircles (i.e, their centers are co-planar), which are stretched like rubber bands [**DG69**]. However, these calculations do not rule out the possibility of a kinetoplast with a few topologically

5

linked, compressed layers (i.e, where a minicircle is linked to circles whose centers are in planes above and below it—not just to circles in its plane).

In order to rule out this non-two-dimensional case, Pérez-Morga and Englund used electron microscopy to visualize isolated networks [**IME93**]. The authors performed single-angle metallic shadowing with a Pt-Pd alloy, and then coated grids with carbon in the presence of ethidium bromide. After this, they coated the grids via photographic emulsion and imaged them with a Zeiss 10 A/B EM. The isolated networks are visible as spread single-layer sheets of linked plasmids (see Figure 1.3).
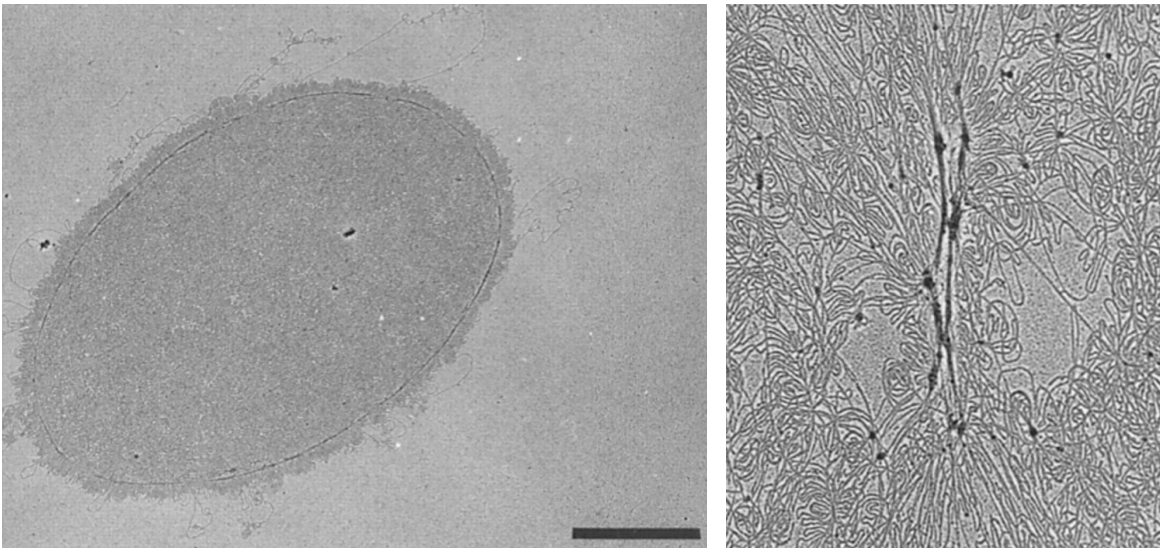
FIGURE 1.3. (Left) Electron micrograph of an isolated *C. fasciculata* kinetoplast (scale bar 2 μm). The right image is an unspecified magnification of a region of the first image [**IME93**].

These *in-vivo* size measurements and the *in-vitro* imaging of expanded networks still leave open the possibility that networks that are 2-D when isolated are folded into several layers inside the organism, reproducing the observed thickness. To rule out this last possibility, Ferguson and colleagues performed *in-situ* hybridization with a fluorescently labelled A-track probe (this binds to AT-rich regions of DNA) to identify nicked and gapped minicircles

(with single- and double-stranded breaks, respectively) [**FTWE92**]. This probe hybridizes efficiently only with non-intact minicircles.

Now, previous studies had established that non-replicating kinetoplasts consist solely of closed minicircles and that partially replicated, isolated kinetoplasts contain nicked or gapped minicircles around their peripheries only [**Eng78**]. So, Ferguson and his colleagues showed that *in-situ* kinetoplasts could not be folded in such a way that their boundaries would not be the same as the boundaries of isolated networks. In other words, the *in-situ* networks that they observed were necessarily monolayer structures like the isolated networks.

We model the kDNA networks in the absence of maxicircles.

For our model we choose to neglect maxicircles, focusing solely on the kinetoplast's minicircles. As with the other assumptions discussed in this section, this choice is consistent with Chen's and Diao's models. In *C. fasciculata* the number of maxicircles is only 0.5% of the number of minicircles. Although no one has studied the topological properties of kinetoplasts containing both minicircles and maxicircles, two teams—one led by Pérez-Morga and the other by Hoeijmakers—have shown that isolated kinetoplasts deficient in maxicircles do not exhibit any visible structural alterations.

Hoeijmakers and Weijers studied the species *Trypanosoma brucei evansi*, which naturally lacks maxicircles [**HW80**]. The authors imaged the kDNA using negative-staining electron microscopy technique.Briefly, Hoeijmakers and Weijers prepared a DNA-cytochrome C complex in the presence of formaldehyde vapor, and then shadowed it with a Pt-Pd alloy. They took images using a Philips EM300 microscope. The authors were unable to distinguish isolated *T. evansi* networks from those of *T. brucei*, and the *T. evansi* networks had the same sizes and shapes as the *T. brucei* networks.

Pérez-Morga and Englund performed electron microscopy on isolated *C. fasciculata* kinetoplasts after treating the networks with the restriction enzyme SstI, which cuts maxicircles

only—not minicircles [**IME93**]. In a similar fashion to Hoeijmakers and Weijers, these latter authors observed no structural differences in these SstI-treated networks when compared to *C. fasciculata* networks that were prepared in the same way but not treated with SstI.

1.1.2. The Tripartite Attachment Complex and the Energetics Model. This project aims to answer one of the important outstanding questions that Jensen and Englund identified in their 2012 review: "How is the complicated and compact structure of the kinetoplast and its associated machinery maintained? Does the TAC play a role? [**JSE08**] As shown in Figure 1.4—and as implied by its name—the tripartite attachment complex consists of three components. The first part binds the kinetoplast to the trypanosome's inner mitochondrial membrane (the "*unilateral filaments*"). The other two parts connect the inner mitochondrial membrane to the outer membrane, and from there to the basal body, which is the structure at the base of the organism's flagellum. Together, the inner and outer membranes are said to comprise the *differentiated mitochondrial membranes.*

Our direct knowledge about the TAC is limited. We have some information from microscopy studies [**ILK**+**18**], and we have a limited amount of knowledge with respect to TAC biochemistry. Fortunately, most of this biochemical information pertains to our species of interest, *T. brucei*, via the RNAi gene-disruption studies referenced in the remainder of this section. In RNAi ("RNA interference"), a specific gene is silenced after translation via the action of a cellular machine called the RNA-induced silencing complex (RISC), which specifically binds and degrades the target mRNA transcript. We discuss the known protein constituents of the *T. brucei* TAC in the following paragraphs.

The TAC component closest to the kinetoplast is **TAC102** [**ILK**+**18**]. TAC102 is a 102-kDa protein involved in kDNA segregation—Trikin and colleagues have shown that cells deficient in this protein produce offspring in which the majority of the kDNA stays with the parent basal body [**WHF**+**77**]. **p166** (166-kDa mass) was the first TAC protein identified, and its RNAi knockdown exhibits a similar kDNA-segregation defect to TAC102 [**WHF**+**77**].
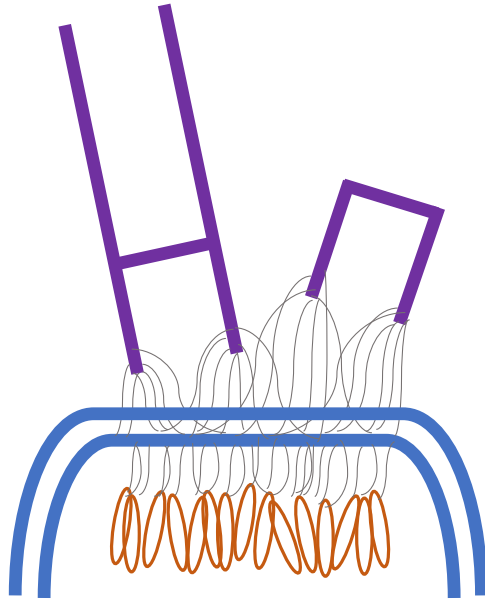
8

FIGURE 1.4. The tripartite-attachment complex (grey filaments) and its neighborhood: Minicircles are shown in dark orange, contained within the mitochondrial membrane (in blue). The flagellum (at left), the basal body (bottom portion of the flagellum), and the pro-basal body (on the right) are shown in purple. All three of these latter structures are cylindrical.

p166 is acidic and localizes to the region between the inner membrane and the kDNA. In the *exclusion zone* (the region outside the mitochondrion proximal to the basal body) there are three known TAC proteins: p197, TAC65, and Mab22 [**ZLRC⁺08**]. **p197** depletion leads to a defect in kDNA segregation [**GBZH13**], and **TAC65** also exhibits a kDNA-mis-segregation phenotype [**Kea16**]. It is not known, however, whether either of these proteins has a function beyond kDNA segregation. The final protein is known by the monoclonal antibody that associates with it, **Mab22**, and is part of the cytoskeleton.

There are two known TAC proteins associated with the differentiated mitochondrial membranes. **TAC40** functions in segregation with a similar RNAi phenotype to that of p166, although the kDNA structure shows no apparent changes under EM outside of a change in size [**Sea14**]. TAC40 is an outer mitochondrial-membrane porin protein (which facilitates intra-membrane transport), and it is closely related biochemically to another outer

membrane protein, **TAC60**, which also has a segregation function [**JSE08**]. Finally, there are two additional *T. brucei* proteins—the universal minicircle-sequence binding proteins (UMSBPs)—that localize to the region of the TAC [**JSE08**]. However, these proteins are not considered part of the TAC. Onn and colleagues have speculated that the UMSBPs play a role in the initiation of kDNA replication through their binding to two minicircle sequences. These proteins are tangential to the current study, and will not be discussed further.

The most natural protein to study from the perspective of our model is TAC102, because it is the closest protein to the kinetoplast. TAC102's impact on kDNA topology is unknown. However, Trikin and his colleagues used thin-section electron-microscopy to show, in 2016,

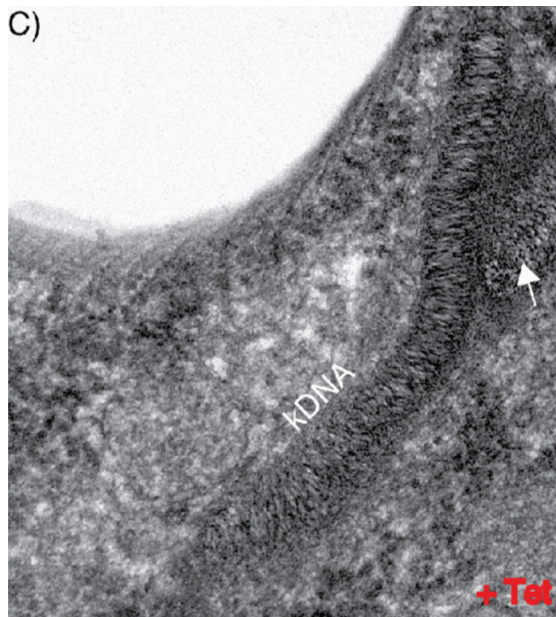FIGURE 1.5. TAC102-disrupted kinetoplast. The arrow indicates a non-2-D region of kDNA. This kinetoplast is approximately 1μm in diameter [**Tea16**].

that when TAC102 is disrupted in *T. brucei*, the parallel striation of kDNA strands with respect to the kinetoplast's symmetry axis is preserved, even as the kinetoplast increases in surface area and sometimes exhibits non-two-dimensional extensions (see Figure 1.5).

1.2. Mathematical Models for kDNA Topology

In this section we describe our mathematical and computational models, which we use to describe the minicircle network's topology The topology of a minicircle network is determined by two pieces of information: (1) the type of link connecting any two minicircles and (2) the minicircle mean valence, *i.e*, the average number of minicircles that are linked to any given circle . As discussed in the previous section, Rauch and her colleagues determined that pairs of minicircles are linked only via Hopf links, and Chen and his colleagues found that the mean valence of *C. fasciculata* minicircles is three. In this analysis, and in the studies discussed in the following pages, one of our goals is to test the *confinement hypothesis*, which postulates that the kDNA network's topology results from packing the minicircles into a limited volume. This hypothesis is consistent with theory and experimental results for other macromolecules whose topological complexity increases upon confinement [**ATV**$^+$**02**].

We made the following modeling assumptions, which are supported by the information presented in the previous section:

(1) **Biological assumption** Maxicircles do not play a significant role in determining kDNA topology.

(2) **Biological assumption** The minicircles forms a monolayer, *i.e*, the system is two-dimensional.

(3) **Biological assumption** Minicircle valence is spatially uniform (homogeneous).

(4) **Mathematical assumption** The minicircle centers are located at the vertices of a regular square lattice.

(5) **Mathematical assumption** The minicircles are treated as geometric circles of unit radius.

The second mathematical assumption imposes a distribution on the positions of minicircles. Since the true distribution of minicircle centers is unknown, we choose to start with a simplified model. This simplification is extremely helpful in developing a theoretical analysis framework and, as we discuss in section 3.2, it does not significantly affect our results.

In the case of the third mathematical assumption, there are two possible interpretations. First, minicircles might have a relatively simple, circular shape—this is consistent with two experimental observations that were discussed in the previous section: (1) kDNA minicircles are relaxed rather than highly supercoiled; and (2) the only observed type of link is the Hopf link, which is the only possible link that can be formed with geometrical minicircles. In the second interpretation, we model flexible minicircles generated via a polymer model (*e.g,* wormlike chain or freely jointed chain) by rigid minicircles whose radii are the same as the radii of gyration of the flexible chains. The third mathematical assumption is primarily a simplifying choice postulated *a priori*, but it is consistent with biological evidence that few holes (regions with missing minicircles) are observed within isolated kinetoplasts [**CRW⁺95**].

1.2.1. The *MC-kDNA* Model for Minicircle Topology. In 2011 members of our lab and our collaborators presented a Monte-Carlo-based random knotting/linking model called *Monte-Carlo kDNA* (MC-kDNA) [**DHK⁺12, DAH12**]. *MC-kDNA* was designed to extend the model proposed by Chen and colleagues in 1995 [**CRW⁺95**]. To generate an *MC-kDNA* conformation, we map a finite set of minicircles to the vertices of a finite square region of a square lattice in a one-to-one and onto fashion. Each minicircle (a geometrical circle of unit radius) is centered at a site in a 2-D lattice, which can be square, triangular, hexagonal, or random. We show a sample conformation is in Fig. 1.6. The minicircle density—defined as the number of minicircles per unit area—is given by a^{-2} where a is the lattice constant.

A minicircle has two orientational degrees of freedom—the *tilting angle* and the *azimuthal angle*. These angles describe the circle's orientation in space and specify the vector normal to the plane containg the circle (see Fig. 1.7). The tilting angle, θ, is the angle between the
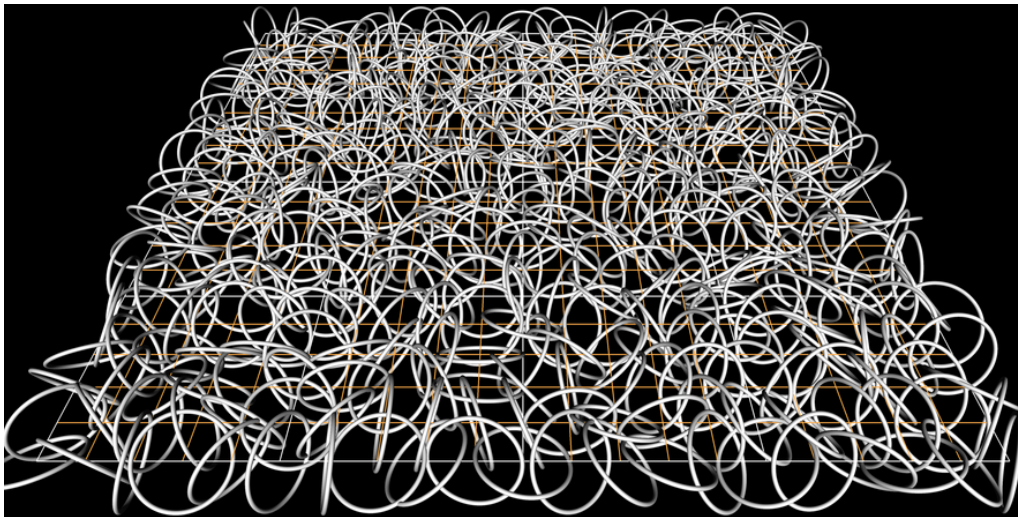
FIGURE 1.6. A grid generated by MC-kDNA

normal vector and the positive z-axis, and the azimuthal angle, ϕ, is the angle between the x-axis and the normal vector's projection onto the xy-plane. Figure 1.7 illustrates how each minicircle is defined. In the standard *MC-kDNA* model, minicircles are oriented randomly, *i.e*, the normal vectors are assumed to be uniformly distributed over the unit sphere. As such, θ is chosen from $[0, \pi]$ (using the well-known result that $\cos\theta$ is uniformly distributed in $[-1, 1]$), and ϕ is taken uniformly from $[0, 2\pi]$.

FIGURE 1.7. Orientation of a minicircle (adapted from [**LPDA20**])

1.2.2. Extensions of the basic random knotting model for kDNA.

Angle restrictions limit the random orientations of minicircle normals to a biologically relevant range.

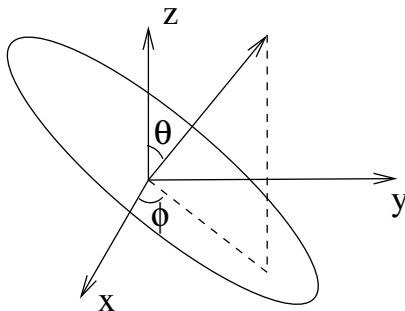In [**ADH12**] Arsuaga and colleagues proposed restrictions on minicircle orientations: $\theta \in [\theta_0, \pi - \theta_0]$ and $\phi \in [\phi_0, \pi - \phi_0] \cup [\pi + \phi_0, 2\pi - \phi_0]$ for positive constants $\theta_0 \geq 0, \phi_0 < \pi$. θ_0 is called the tilting-angle restriction and ϕ_0 is called the azimuthal-angle restriction. In the work discussed here, we will consider only the tilting-angle restriction.

Volume exclusion models electrostatic repulsion.

In [**DHK$^+$12**] Diao and his colleagues modeled excluded-volume effects by placing a hard cylinder-potential around each kDNA minicircle, which themselves are modeled as rigid geometrical circles. This approach has been shown to adequately approximate repulsive electrostatic interactions [**RCV93**]. The authors observed that most of the conformations generated using standard random sampling were invalid and proposed to address this problem using an imputation approach. The idea of this imputation approach is to estimate the various conditional linking probabilities using a simulated annealing algorithm on small (10×10) minicircle grids. These empirically obtained linking probabilities were then used to generate larger grids.

Hard-shell repulsion can be used to model the effects of a positively charged kDNA environment, which could contain concentrations of cations or of histone-like proteins. Positive ions and proteins decrease the effective excluded volume by shielding the repulsion between negative charges. There are four known histone H1-like proteins associated with the kinetoplast in *C. fasciculata* [**Lea02**]. Each of these is small and basic. **KAP1** is implicated in the structure of the kinetoplast *in vivo*, but is not required for viability of the organism. The substantial structural rearrangements induced by **KAP1** gene disruption are not compensated for by the activity of **KAP2**, **3**, or **4**. KAP1 is the most basic of the four proteins

and 50% of its amino-acid residues are either lysine or alanine. KAP2, 3, and 4 also contain a large number of lysine residues, which account for the large positive charges of all four of these proteins [**XHE⁺02**]. KAP1 binds nonspecifically to kinetoplast minicircles, whereas KAP2, 3, and 4 all bind specifically to a single region of minicircle DNA [**XHE⁺02**]. When the KAP1 gene is disrupted, kinetoplasts exhibit a visible change in internal structure under thin- section TEM, although the shape and dimensions of the kinetoplast remain approximately the same [**Lea02**]. KAP2, 3, and 4 also condense kDNA *in vitro* [**XHE⁺02**]. Xu and his colleagues showed that each of these proteins exhibited this condensation phenotype independently of the other KAP proteins at a concentration of 1 μg / μL. These authors also showed that all three proteins are present throughout the kinetoplast via immunofluorescence.

Polymer flexibility can alter the linking behavior of minicircles.

Since kDNA minicircles are not rigid circles and their shapes are affected by factors like polymer energetics [**Vol06**], nucleotide sequence [**MLCE83**], and by possible condensing (or DNA binding) factors [**AEKS99**,**HR98**,**ATS83**] Arsuaga and his colleagues studied the case in which rigid *MC-kDNA* minicircles are replaced by closed equilateral polygons called *freely jointed chains* (FJCs), which have no bending penalty between edges [**ADR14**]. This entails discretizing the continuous DNA backbone, representing it as a set of N equal-length edges connecting pairs of vertices, which are contained in $\mathbb{R}^3$. FJCs capture entropic contributions to minicircle linking (both between pairs of circles and throughout the grid), but do not include energetic effects.

Arsuaga and his colleagues generated minicircle conformations using the generalized hedgehog algorithm [**VHAD09**] as implemented in Knotplot [**Sch**]. The topological state of each simulated freely jointed chain was computed using the Alexander polynomial evaluated at -1, $\Delta(-1)$. Minicircles with $\Delta(-1) \neq 1$ were rejected. The minicircle centers were assigned

to lattice points and their normal vectors were chosen randomly. Once all of the $d \times d$ polygons are positioned, the authors determined the conformation's topology by calculating the Gauss linking number (via Klenin and Langowski's Gaussian integral formulation [**KL00**]).

Distribution of Minicircle Centers

Since the minicircle' topology could be sensitive to the distribution of the circles' centers, members of our lab and our collaborators studied triangular, hexagonal, and random planar lattices [**DAH12**, **RDA13**]. Sample conformations of each type are shown in Fig. 1.8.
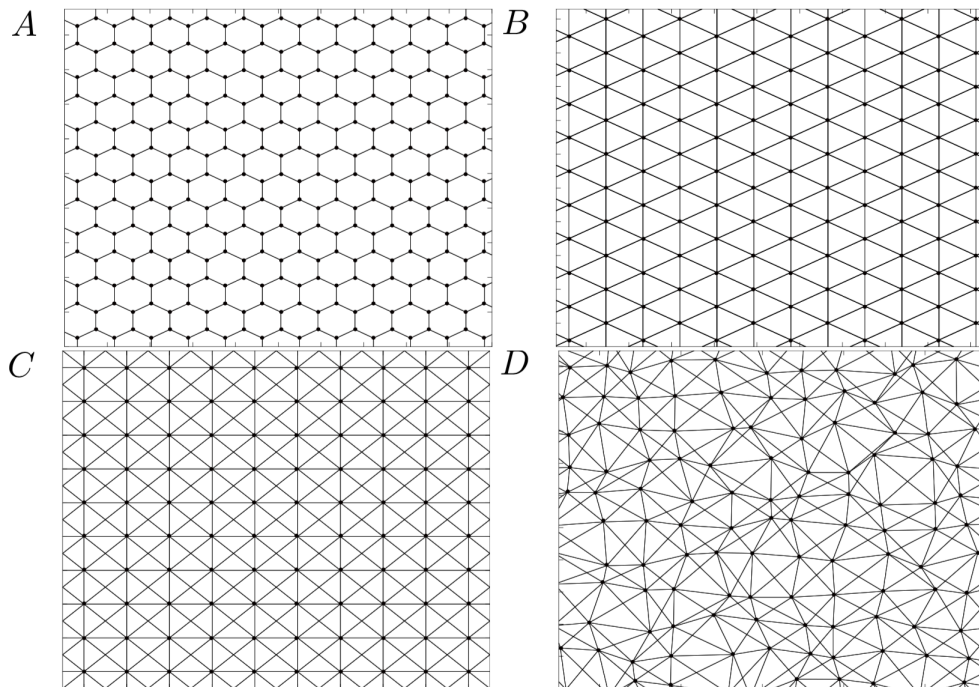
FIGURE 1.8. The hexagonal, triangular, square and random lattices [**DAH12**]

The minicircle density is given by $(\sqrt{3}/6)r^{-2}$ for a triangular lattice and $(\sqrt{3}/9)r^{-2}$ for a hexagonal lattice. Rodriguez and his colleagues generated random lattices by identifying the edges of the lattice (*i.e*, generating a torus) and randomly displacing the vertices of the square lattice uniformly around the position of the minicircle.

1.2.3. Quantitative Descriptions of Minicircle Topology and Definitions.

Definition: Minicircle Linking in the MC-kDNA Model

We say that two minicircles are topologically *linked* if the circles cannot be separated without breaking one of them. Two or more circles can create a *linked cluster* if each circle is linked to at least one other circle in the cluster, and together they form a single object. In other words, no single circle can be separated from the group without a breakage.

Let C_1 and C_2 be two disjoint simple closed and oriented curves in $\mathbb{R}^3$. We say that C_1 and C_2 form an unsplittable link if no topological 2-sphere separates them. Let $\mathbf{r}_1(t_1)$ and $\mathbf{r}_2(t_2)$ be parameterizations of C_1 and C_2, respectively, with $t_1, t_2 \in [0, 2\pi]$. The Gauss linking number $l(C_1, C_2)$ of C_1 and C_2 is defined by:

$$(1.1) \qquad l(C_1, C_2) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{2\pi} \frac{\langle \mathbf{r}_2 - \mathbf{r}_1, \mathbf{r}_1', \mathbf{r}_2' \rangle}{|\mathbf{r}_2 - \mathbf{r}_1|^3} dt_1 dt_2,$$

where the fraction's numerator contains a triple product [**RN11**]. It is well known that the linking number is a topological invariant. In particular, if $l(C_1, C_2) \neq 0$, then C_1 and C_2 are linked. In the case that C1 and C2 are both circles of the same radius r, the linking between them can be determined geometrically—this algorithm is explained in section 3.1.

Definition: The Mean Valence Parameter

The mean minicircle valence was first identified by Chen and colleagues and is a parameter that can be measured experimentally. A minicircle's *valence* is the number of minicircles that are topologically linked to it directly. For a given minicircle conformation (with fixed orientations and lattice constant), the mean valence is taken over all minicircles in the conformation—and is not necessarily an integer. We are interested in the relationship

between minicircle density and mean valence, since Chen and colleagues showed experimentally that the valence of minicircle kDNA doubles after replication (i.e. the density doubles) [**CRW**$^+$**95**].

1.2.4. MC-kDNA and its Extensions—Results and Shortcomings. Chen and his colleagues developed the first quantitative model for the topology of the *C. fasciculata* network [**CRW**$^+$**95**]. As we noted earlier, these authors determined experimentally that each minicircle is linked on average to three of its neighbors through a Hopf link. In 2012, Diao et al. introduced a more sophisticated model, but this model was unable to fully capture the topological properties of the kinetoplast system [**DHK**$^+$**12**]. Their minicircle model demonstrated a linear relationship between mean valence and density. However, at a physiologically relevant density, the model predicts a mean valence of approximately 400, which is far higher than Chen's experimental value of three.

Arsuaga and Diao and their colleagues extended this earlier model by restricting the range of angles from which orientations were sampled [**ADH12**]. This extension was motivated by the idea that, under high confinement *in vivo*, minicircles likely have less orientational freedom. This study produced a better mean-valence prediction than the original work, although the value still remained too high at physiological densities—approximately 11.

We will address this shortcoming—building on the work of Diao and his colleagues—by proposing an energetics model intended to capture the relationship between orientational and topological order. While Arsuaga and his colleagues established the connection between minicircle orientation and the topological properties of the kinetoplast system, their sampling method remained mathematical and uniform (within a restricted range of values). The energetics model that we propose in the following section permits a better sampling method, and it is designed to directly reflect the observed orientational order of minicircles under physiological confinement.

1.3. The Lebwohl-Lasher Model and Selections from Liquid-Crystal Theory

The minicircle normals determine the topological properties of each conformation, but are also of intrinsic physical interest. In particular, their orientations tells us how order emerges in the minicircle structure and how geometrical information is transmitted across space. In this work we treat topological order as an emergent property that arises from local (phenomenological/effective) interactions. More precisely, we propose an energy that couples the principal symmetry axes of ringlike molecules to one another via a modified version of a canonical liquid-crystal Hamiltonian. In the following section, we discuss the historical development of the model class that our Hamiltonian fits into. We present the precise form of this Hamiltonian in Chapter 2.

1.3.1. Introduction to the Lebwohl-Lasher Model. P. Lebwohl and G. Lasher proposed their eponymous model [**LL72**] in 1972 to permit lattice-based computational studies of Maier-Saupe's continuum theory [**MS59**]. The earlier Maier-Saupe theory used a mean-field analysis to predict a first-order nematic-isotropic phase transition in a system of rod-like liquid crystals interacting via a long-range attractive potential.

In the Lebwohl-Lasher model, nearest neighbor particles i and j interact via an exchange interaction

(1.2)
$$\frac{-H_{LL}^{(ij)}}{K} = 3\cos^2\theta_{ij} - \frac{1}{2},$$

where K is a positive-definite parameter with units of energy, and the angle θ_{ij} is defined (for unit vectors) by

(1.3)
$$\theta_{ij} \doteq \cos^{-1}(\mathbf{S}_i \cdot \mathbf{S}_j).$$

These authors performed Monte-Carlo simulations of cubic-lattice systems of sizes $10{\times}10{\times}10$ and $20 \times 20 \times 20$. They found the predicted first-order phase transition, and determined
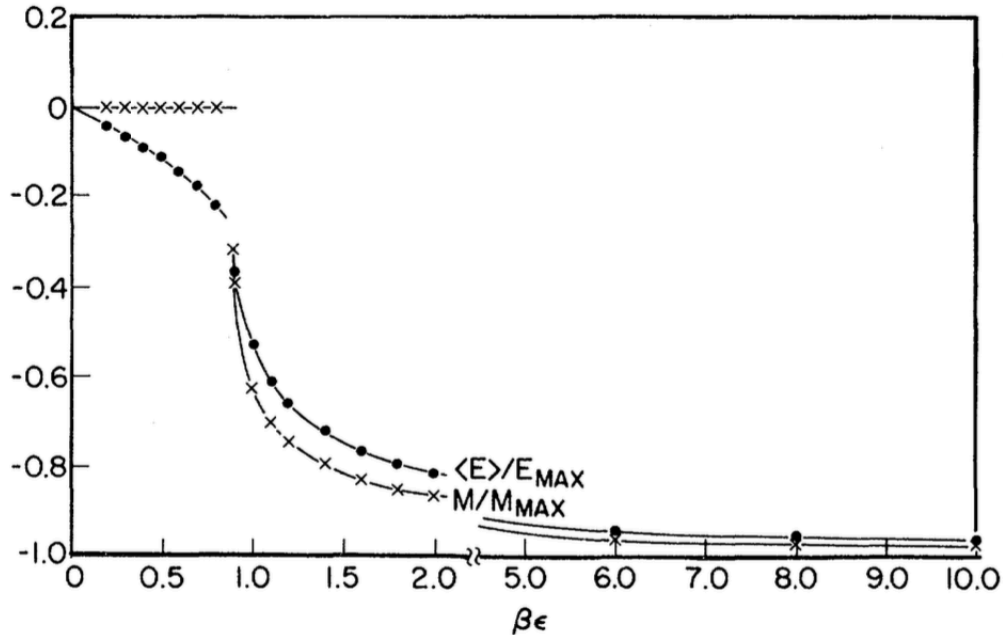
19

FIGURE 1.9. Lebwohl and Lasher plot average energy (normalized to $-1 \leq <E> \leq 0$ and the P_2 order parameter ("M") versus *inverse* temperature [**LL72**]. Both plots exhibit a discontinuity at $T^* = 1.12$.

that it occurs at dimensionless temperature $T^* \doteq k_B T/K = 1.12$ (see Fig. 1.9), which differs from Maier and Saupe's mean-field based estimate of $T^* = 1.30$.

1.3.2. Phase-Transitions and Pseudo-Transitions in Systems with Continuous Symmetries: the Mermin-Wagner Theorem and Kosterlitz-Thouless-like Behavior.

In the 1960s Pierre Hohenberg [**Hoh67**], David Mermin, and Herbert Wagner (the latter two in a separate work [**MW66**]) built a rigorous and general proof for the absence of long-range order (LRO) in a large group of systems with spatial dimension $d \leq 2$. The *Hohenberg-Mermin-Wagner* (HMW) theorem tells us that there are no spontaneous symmetry-breaking phase transitions at finite temperatures for systems with continuous symmetries in dimension $d \leq 2$ [**MW66**].

Hohenberg studied continuum models of superfluids and superconductors, and found that no LRO can exist at finite temperature for $d \leq 2$. Mermin and Wagner established a similar

result for classical *lattice* models of ferromagnets and anti-ferromagnets (Heisenberg models) with SU(2) symmetry [**Hal18**]. While Hohenberg's result permits long-range interactions, Mermin and Wagner's result requires that all interactions be finite, *i.e*, that

$$(1.4) \qquad \sum_{\mathbf{r}} r_\alpha r_\beta J(\mathbf{r}) < \infty,$$

where $\alpha, \beta \in \{x, y\}$ and J is the coupling constant between spins separated by the vector $\mathbf{r}$.

In 1973 Kosterlitz and Thouless published a study [**KT72**] of the classical 2-D xy-model. Here, spins, $\mathbf{S}_i \in \mathbb{R}^2$, occupy sites on a 2-D square lattice and nearest neighbors interact with the potential:

$$(1.5) \qquad \frac{H_{xy}}{J} = \cos(\phi_i - \phi_j),$$

where $J > 0$. The authors found that a sharp transition exists, above which the correlation function decays exponentially, while below which the correlation follows a power-law decay, which is now commonly referred to as "quasi-long-range order" [**Hal18**] Kosterlitz and Thouless study the system by focusing on two harmonic (quadratic) terms in the Hamiltonian's Taylor expansion about a local minimum, E_0 (a "vortex" term and a spin-wave stiffness term):

$$(1.6) \qquad H - E_0 = \frac{J}{2} \sum_{\langle i,j \rangle} (\phi_i - \phi_j)^2 + O((\phi_i - \phi_j)^4).$$

The authors treat the vortex and spin-wave terms independently. They calculated a spin-spin correlation function, $\langle \mathbf{S}_i \cdot \mathbf{S}_j \rangle = \langle \exp[(\phi_i - \phi_j)] \rangle$, and they showed that it decays exponentially in the case of spin-wave elasticity alone. Kosterlitz and Thouless also studied the case of Heisenberg spins ($\mathbf{S}_i \in \mathbb{R}^3$) on the same 2-D square lattice, wherein they found that correlations decay exponentially for all finite temperatures.

1.3.3. Review of Previous Studies on the LL Model. Several versions of the Lebwohl-Lasher model in two and three dimensions have been studied, mostly via Monte-Carlo simulations. We present the key results below.

Luckhurst and colleagues determined the field-dependence of the order parameter $< P_2 >$ in an anisotropic 3-D model.

The authors studied a system of spin-directors located on the points of a cubic lattice [**LSC81**]. The spins interact with their nearest neighbors via the Lebwohl-Lasher potential defined earlier. Spins are also coupled to an external field via a term

$$(1.7) \qquad\qquad\qquad H^{(i)}_{field} = -D \cos^2 \theta_i,$$

where θ_i is the azimuthal angle for spin i (i.e, the angle spin i makes with the z-axis). Note that this choice of field favors spin-directors parallel to the z-axis. The authors performed Monte-Carlo importance sampling on systems of size $10 \times 10 \times 10$.

The authors characterized the field dependence of the order parameter $\langle P_2 \rangle = \langle \cos(\theta) \rangle$, which measures the system's order in the z-direction (see Fig. 1.10). They find a reasonably good agreement between mean-field predictions and their MC results when they simulate the system at $T^* = 1.182$.

Chiccoli and colleagues find power-law behavior in the correlation function for a planar isotropic model.

This was the first study of a 2-D Lebwohl-Lasher model. The authors considered nearest neighbor spins interacting via the potential defined in section 1.3.1 [**CPC88**]. They performed Monte-Carlo simulations on square lattices of sizes $5 \times 5, 10 \times 10, 20 \times 20, 60 \times 60$, and 80×80. The authors observed an orientational pseudo-transition in the system at di-
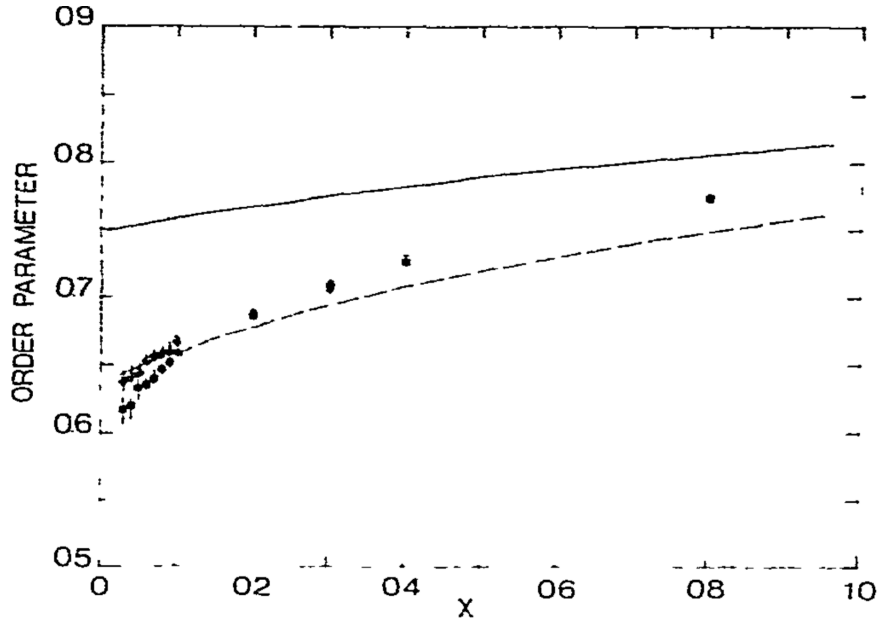
FIGURE 1.10. Luckhurst and colleagues determined the field dependence of $\langle P_2 \rangle$ in an anisotropic cubic-lattice Lebwohl-Lasher model [**LSC81**]. The two lines show their mean-field predictions for $T^* = 1.0$ (solid) and $T^* = 1.182$ (dotted).

mensionless temperature $T^* = 0.6$ (see Fig. 1.11), which they confirmed with heat-capacity measurements. The authors saw a sharp—but continuous and analytic—increase in the order parameter starting from a low but nonzero value. This is consistent with the Mermin-Wagner theorem.

The authors also observed a power-law decay in the orientational correlation function below the pseudo-transition temperature (Kosterlitz-Thouless-like behavior), and an exponential decay in the function above this temperature (see Fig.1.12).

1.4. SARS-CoV-2 and the Spike Protein

In this work we use structural modeling tools and molecular-dynamics (MD) simulations to characterize two viral strains, S477N and S477N/D614G. We compare these results with the empirical population dynamics of the strains, which we calculate using a pHMM. Our
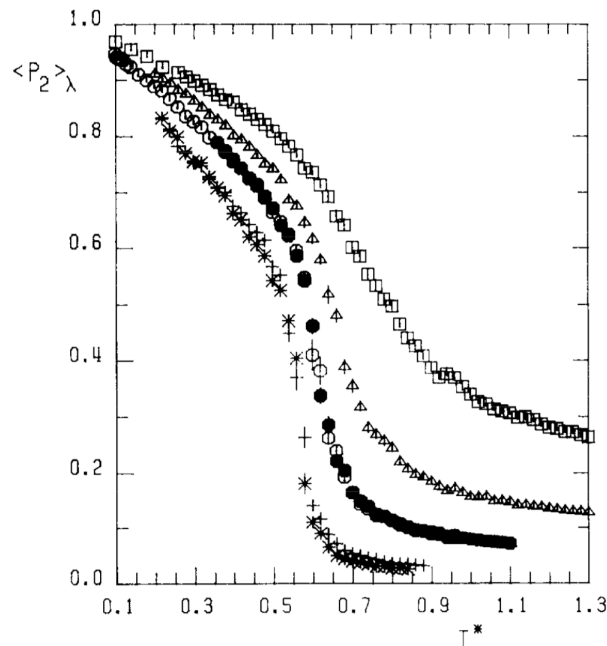
FIGURE 1.11. Chiccoli and colleagues calculate the dependence of the P_2 order parameter on dimensionless temperature, $T^* = k_B T/K$ [**CPC88**]. The symbols correspond to different lattice sizes: 5×5 (squares), 10×10 (triangles), 20×20 (circles—empty denote heating and filled denote cooling), 60×60 (pluses), and 80×80 (stars).

dynamics study includes nearly 150,000 viral sequences, which were collected between early January and the end of October in 2020. We identify several residues located in the spike RBD that contribute substantially to changes in S-ACE2 binding affinities, some of which are hydrogen-bonding residues. Finally, we perform MD binding simulations on five serum-isolated antibodies from recovered individuals: 7CJF, 6XE1, 7KFX, 7KFV, and 7KFW. The latter three proteins (the 7KF family) were isolated from a single individual. The 7KF family binds directly to the RBD's active site and its members differ by minor somatic mutations [**CCP$^+$20**]. 7CJF [**GHZ$^+$21**] and 6XE1 [**HSW$^+$20**] also bind to the RBD active site and their their genomic compositions differ from the 7KF family.

1.4.1. SARS-CoV-2 and the Covid-19 Pandemic. SARS-CoV-2, the betacoronavirus that causes the COVID-19 illness, is an enveloped virus that contains a unique 29-kb
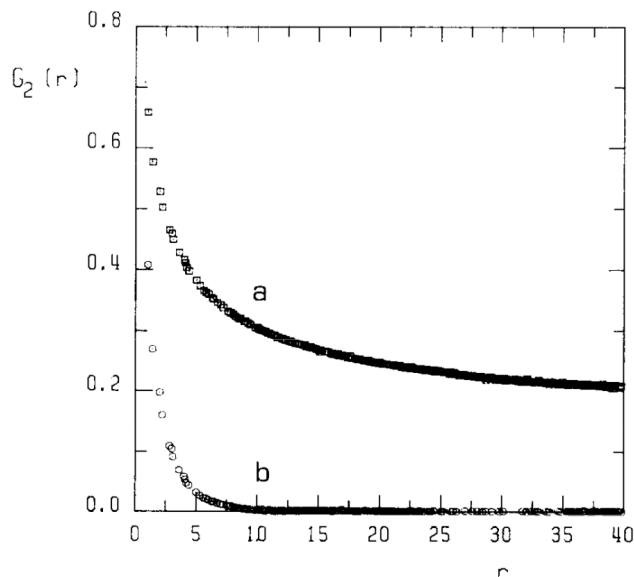
FIGURE 1.12. Chiccoli and colleagues plot the correlation function for the P_2 order parameter as a function of spin separation, r, for the 80×80 lattice [**CPC88**]. The authors find a power-law decay below the pseudo-transition temperature ($T_a^* = 0.54$) and an exponential decay above this critical temperature ($T_b^* = 0.72$).

single stranded (+)RNA genome [**WZY$^+$20**]. The disease was first identified in December of 2019 as caused by a novel coronavirus, with an initial outbreak reported in Wuhan, China. On March 11, 2020 the WHO declared the outbreak a pandemic The SARS-CoV-2 genome encodes for twenty proteins—the envelope (E), membrane (M), nucleocapsid (N), and spike (S) proteins, as well as sixteen *non-structural proteins* (NSPs). Proteins E, M, and S are expressed on the surface of the viral envelope, while the N-protein protects the viral genome. The spike protein mediates cell infection by binding to angiotensin converting enzyme-2 (ACE2), on the surfaces of human bronchial and lung epithelial cells [**ZJM$^+$**].

1.4.2. The Spike Protein and the ACE2 Receptor. The S-protein is a trimeric glycoprotein. The membrane-distal portion of each 1,208-residue protomer is called the receptor-binding domain (RBD; residues 319-526), which contacts the "helix-turn-helix" (HTH) motif of ACE2. Each protomer has two subunits, S1 and S2, with an "RRAR"

furin cleavage site at their intersection. This is one of the few significant differences between SARS-CoV-2 and RaTG13. The high-resolution cryo-EM structure determined by Wrapp and his colleagues shows that typically only one of these RBDs is in the "up" or active conformation in the pre-fusion (pre-binding) trimer; the other two are "down" and less solvent-accessible [**WWC**$^+$]. When the S-protein binds ACE-2, it undergoes significant changes. The non-binding RBDs are flipped to the up position, which destabilizes the structure, and then the S1 domains are shed and the protein rearranges itself.

Recent genomics studies have reported numerous amino-acids substitutions in the spike protein [**SGCR20**, **LWN**$^+$], and we report the population dynamics of two of those mutations, S477N and D614G, here. S477N substitutes ASN for SER at locus 477 and D614G substitutes GLY for ASP at site 614. We used a probabilistic method called a *profile Hidden-Markov Model* (pHMM) to construct a time series for the population occupancies of individual amino-acid substitutions in a viral-genome data set.

1.4.3. There are several reported mutations in the S-Protein. A subset of the known mutations demonstrates fitness by attaining relatively large population fractions, and some of these mutations have reported binding improvements. For example, molecular dynamics simulations predict that the RBD substitutions V367F, W436R, and N354D/D364Y all improve ACE2 binding [**OZD**$^+$**20**]. Strains characterized by the D614G substitution quickly dominated the viral population [**KFG**$^+$**20**], which we verfiy in our pHMM results. *In-vitro* studies by Yurkovetskiy and others [**YWP**$^+$**20**, **ZJM**$^+$] indicate that S trimers containing the D614G substitution alone bind 5.3x worse than wild-type (WT) trimers. Residue 614 is located far from the RBD and Yurkovetskiy and colleagues report that the substitution disrupts an inter-protomer hydrogen bond between D614 and T859 in WT. These authors propose that residue G614 acts as an open "latch," increasing the probability that each RBD will be in its binding-ready "up" conformation. However, this is inconsistent with

their surface-plasmon resonance measurements at 37 degrees C, which show that the equilibrium association constant, K_a, decreases relative to the WT value. Instead, it is possible that G614$\tilde{\text{O}}$'s disruption of the inter-protomer interaction destabilizes the trimer, increasing thermal fluctuations that cause the complexed proteins to disassociate.

There is not yet a consensus as to whether the S477N mutation produces a binding-driven fitness advantage. Two molecular dynamics (MD) studies disagree on this point: Singh and colleagues report that S477N improves ACE2 binding [**SSK**$^+$], while Ou and colleagues find that S477N does not increase affinity [**OZD**$^+$**20**]. More stable binding between S and ACE2 can lead to higher cell infection rates [**PLL**$^+$], which could correlate with worse epidemiological outcomes, e.g: higher viral load, increased transmission efficiency, and more severe clinical profiles. In a similar fashion, there is evidence that a small number of RBD mutations (e.g, E484K in the "Brazilian strain," P.1, and in the "South African strain," B.1.351, and L452R in the "CA strains," B.1.427/429) decrease neutralizing antibody activity in individuals who have recovered from COVID-19 or who have received several of the vaccines [**?**]. This means that individuals exposed to these strains are at somewhat increased risk of contracting COVID-19, in spite of their status as either recovered or vaccinated individuals.

CHAPTER 2

The 2-D Anisotropic LL Model: A Theoretical System Inspired by Thin Films and Biological Nanostructures

2.1. Methods and Basic Definitions

2.1.1. The 2-D Anisotropic Lebwohl-Lasher Model. The kinetoplast minicircles are known to exhibit an ordered geometry, in which the minicircles align parallel to one another and perpendicular to the plane of the kinetoplast disk. We propose to model this behavior via a Hamiltonian whose low-energy states mirror this biological geometry. Consider a 2-D square lattice of spins $\mathbf{S}_i \in \mathbb{R}^3$, which are unit vectors centered at a lattice site i. We define our Hamiltonian by

$$(2.1) \qquad H = -K \sum_{<i,j>} (\mathbf{S}_i \cdot \mathbf{S}_j)^2 + D \sum_i S_{iz}^2 \ ,$$

where K is the ferroquadrupolar coupling strength, S_{iz} is the z-component of spin i, and D is a field parameter, which breaks the symmetry in the xy-plane and penalizes larger z-components ($K, D > 0$). The first term in the Hamiltonian is a sum over nearest neighbors in the lattice (i.e, short-range interactions).

Quadrupole moments capture director symmetry and provide physically motivated order parameters.

Neither the topological nor the geometric properties that we are considering in this study depend upon the chirality of the minicircles. Although rings of DNA have a well-defined chirality, this property does not seem to play an important role in determining the structural

28

characteristics nor the interesting topological behaviors of the minicircles. Chen and his colleagues took the same approach when they chose to ignore the signs of topological crossings in their network model [**CRW$^+$95**].

Therefore, in this model we use spin directors—rather than spin vectors—which define the axis normal to the disk bounded by the rigid minicircle. Spin directors are most expressed naturally as quadrupole moments—principally, Q_0, which is defined by

$$(2.2) \qquad\qquad\qquad\qquad Q_0 = 3S_z^2 - 1$$

as well as the other four quadrupole moments:

$$(2.3) \qquad\qquad Q_2 = S_x^2 - S_y^2, \;\; Q_{xy} = 2S_xS_y, \;\; Q_{yz} = 2S_yS_z, \;\; Q_{xz} = 2S_xS_z,$$

These moments are the five independent elements of the *quadrupole tensor*, which is a 3×3 traceless, symmetric matrix. Since we choose all of our spin-directors to have unit length, the tensor is also subject to an additional, simple constraint:

$$(2.4) \qquad\qquad\qquad\qquad S^2 = S_x^2 + S_y^2 + S_z^2 = 1$$

Since these are classical spins, the normalization for Q_0 given in the previous definition means that, for randomly oriented directors (in the high-T regime), we expect:

$$(2.5) \qquad\qquad\qquad\qquad \langle Q_0 \rangle = 0.$$

Therefore, Q_0 is a good choice for an orientational order parameter.

We now rewrite the Hamiltonian terms of these five quadrupole moments with the help of the following identities:

$$(2.6) \qquad S_z^2 = \frac{Q_0 + 2}{3}$$

$$(2.7) \qquad S_x^2 = \frac{1}{2}[S^2 + Q_2 - \frac{1}{3}(Q_0 + 1)]$$

$$(2.8) \qquad S_y^2 = \frac{1}{2}[Q_2 - S^2 + \frac{1}{3}(Q_0 + 2)]$$

$$(2.9) \qquad 2S_{ix}S_{iy}S_{jx}S_{iy} = \frac{1}{2}Q_{xy}^{(i)}Q_{xy}^{(j)}$$

We begin by expanding the first term in the Hamiltonian in the spin components:

$$(2.10) \qquad -K\sum_{<i,j>}(\mathbf{S}_i \cdot \mathbf{S}_j)^2 = -K\sum_{<i,j>} \quad [\quad S_{ix}^2 S_{jx}^2 + S_{iy}^2 S_{jy}^2$$

$$(2.11) \qquad + \quad S_{iz}^2 S_{jz}^2$$

$$(2.12) \qquad + \quad 2S_{ix}S_{iy}S_{jx}S_{jy} + 2S_{iy}S_{iz}S_{jy}S_{jz}$$

$$(2.13) \qquad + \quad 2S_{ix}S_{iz}S_{jx}S_{jz}].$$

Using the previous identities, we substitute in the quadrupole moments:

$$(2.14) \qquad -K\sum_{<i,j>}(\mathbf{S}_i \cdot \mathbf{S}_j)^2 = -K \; \sum_{<i,j>} \; [\frac{1}{2}Q_2^{(i)}Q_2^{(j)} + \frac{1}{6}Q_0^{(i)}Q_0^{(j)}$$

$$(2.15) \qquad + \quad \frac{1}{2}Q_{xy}^{(i)}Q_{xy}^{(j)} + \frac{1}{2}Q_{yz}^{(i)}Q_{yz}^{(j)}$$

$$(2.16) \qquad + \quad \frac{1}{2}Q_{xz}^{(i)}Q_{xz}^{(j)}],$$

Here, the superscripted indices in parentheses indicate to which spin the moment belongs, and we have neglected the constant terms.

Similarly, we re-express the Hamiltonian's field term:

$$(2.17) \qquad D\sum_i S_{iz}^2 = \frac{D}{3}\sum_i Q_0^{(i)},$$

again, neglecting the constant term. Together, these two terms comprise the quadrupolar formulation of the Hamiltonian.

2.1.2. We augment our MF theory with Monte-Carlo simulations, which allow us to study low-temperature phase behavior. We generated orientational conformations and took samples using the Metropolis-Hastings algorithm [**PB11**]. The underlying Markov chain was defined by the following move: at each iteration we randomly select one spin to "flip" by changing its orientation to a new, random direction. We then calculate the change in energy due to this perturbation. The energy, E, is determined by the quadrupolar Hamiltonian. In a square lattice all interior sites have coordination number four, but, in order to take into account the finite-size effects of the real kinetoplast, we allow the sites along the edges to have a coordination number of either two or three (when we are not considering periodic boundary conditions).

Sampling: Convergence and Independence

After a warm-up period of N_{warmup}, we sample the order parameters, q_0 and q_2 at intervals of N_{int} steps. For each random variable, we choose N_{warmup} by determining the number of steps, N, at which its mean converges to a stable value. For the order parameter q_0, we determined that the warm-up period should be 400,000 steps. As shown in Fig. 2.1, q_0 converges to a stable value after approximately 400,000 steps. In order to ensure independent sampling, we test the time series' auto-correlation using a method called batch-mean analysis, or BMA, (explained in [**Fis13**]) to detect and correct for dependence in the sample variables.

Given a set of n samples for a parameter θ, we use BMA to determine the number of independent samples. For a sampling interval of N_{int} time steps and a batch size b, the correlation time τ is bounded according to:
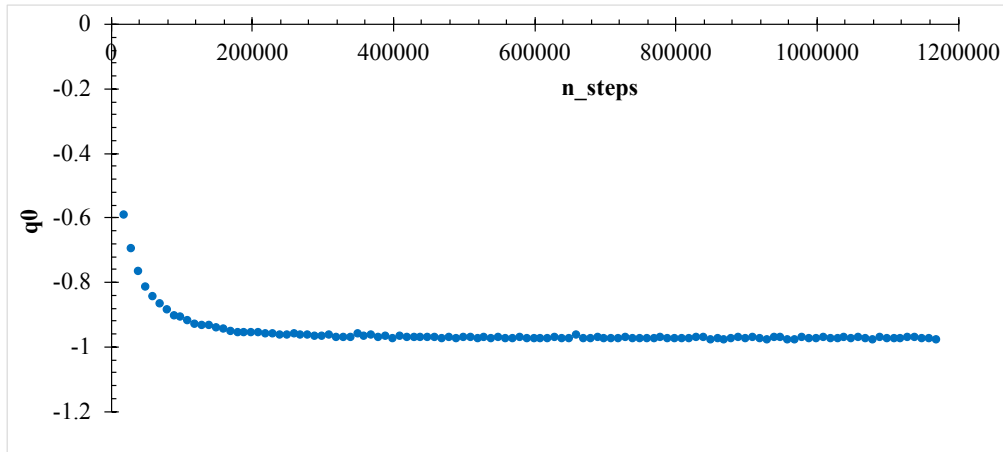
$$\tau < N_{int}b. \tag{2.18}$$

31

FIGURE 2.1. We generated many MCMC t-series to determine the convergence period for the Q_0 order parameter. Each data point gives the sample mean from a t-series of the specified length.

Using the number of independent samples, m, for θ (from BMA), we can calculate its standard error:

$$(2.19) \qquad \sigma_m = \frac{\sigma_n}{\sqrt{m}},$$

where σ_n is the standard deviation in the original set of samples.

2.2. A Mean-Field Theory for the Anisotropic Lebwohl-Lasher Model

In the mean-field approximation, we couple each spin moment to a corresponding molecular-mean field for that moment. For example, the first term in equation 2.14 becomes

$$(2.20) \qquad \sum_{<i,j>} Q_2^{(i)} Q_2^{(j)} \approx \sum_i <Q_2> Q_2^{(i)} = \sum_i q_2 Q_2^{(i)},$$

where we introduce the notation of Chen and Levy [?] in the last equality. In a mean-field approximation, we consider the spins to be noninteracting. Therefore, we can express the

32

partition function for the system as a product of one-particle partition functions:

$$(2.21) \qquad\qquad Z_{\mathrm{MF}} = Z_1^N.$$

The one-particle partition function depends on the single-particle Hamiltonian, which is given by:

$$(2.22) \qquad H_1 = -\frac{K}{2}[q_2 Q_2 + \frac{1}{3}q_0 Q_0 + q_{xy}Q_{xy} + q_{yz}Q_{yz} + q_{xz}Q_{xz}] + \frac{D}{3}Q_0,$$

where we have divided by two to account for double counting. Since the Hamiltonian is invariant under rotations about the z-axis, as well as under reflections across the xy-plane,

$$(2.23) \qquad\qquad q_{xy} = 0 = q_{yz} = q_{xz} = q_2.$$

We show this by expressing each of these moments in terms of spherical coordinates:

$$(2.24) \qquad\qquad Q_{xy} = S_x S_y = \sin^2\theta \sin\phi \cos\phi$$

$$(2.25) \qquad\qquad Q_{yz} = S_y S_z = \sin\theta \cos\theta \sin\phi$$

$$(2.26) \qquad\qquad Q_{xz} = S_x S_z = \sin\theta \cos\theta \cos\phi$$

$$(2.27) \qquad\qquad Q_2 = S_x^2 - S_y^2 = \sin^2\theta \cos^2\phi - \sin^2\theta \sin^2\phi,$$

where we use:

$$(2.28) \qquad\qquad S_x = \sin\theta \cos\phi$$

$$(2.29) \qquad\qquad S_y = \sin\theta \sin\phi$$

$$(2.30) \qquad\qquad S_z = \cos\theta.$$

The expectation value of a variable α is calculated using the expression

$$(2.31) \qquad\qquad <\alpha> = \frac{\int d\Omega \, \alpha e^{-\beta H[\theta,\phi]}}{Z},$$

where Z is the system's partition function; where the Hamiltonian is written as a functional of the two angles; and where the integral is carried out over the surface of the unit sphere.

The expectation value of the xy-moment is therefore:

$$(2.32) \qquad q_{xy} = \frac{\int_0^{2\pi} d\phi \ \sin\phi \cos\phi \int_0^{\pi} d\theta \ \sin^2\theta \ e^{-\beta H[\theta,\phi]}}{Z}.$$

In saying—as we did earlier in this section—that the Hamiltonian is invariant under rotations about the z-axis, we were saying that, for a rotation of this sort taking $\phi \to \phi'$,

$$(2.33) \qquad H[\theta,\phi] = H[\theta,\phi'].$$

This means that the Boltzmann factor is a constant with respect to the first integral (over ϕ) in the numerator of the equation for q_{xy}. If this integral over the azimuthal angle is zero for each of these four moments, than all of their expectation values are zero.

For the first three moments, we verify this as follows:

$$(2.34) \qquad \int_0^{2\pi} d\phi \ \sin\phi \cos\phi = 0 \qquad \Rightarrow \qquad q_{xy} = 0$$

$$(2.35) \qquad \int_0^{2\pi} d\phi \ \sin\phi = 0 \quad \Rightarrow \qquad q_{yz} = 0$$

$$(2.36) \qquad \int_0^{2\pi} d\phi \ \cos\phi = 0 \quad \Rightarrow \qquad q_{xz} = 0.$$

For Q_2, we write:

$$(2.37) \ q_2 \ = \ \frac{\int_0^{2\pi} d\phi \ \cos^2\phi \int_0^{\pi} d\theta \ \sin^3\theta \ e^{-\beta H[\theta,\phi]} - \int_0^{2\pi} d\phi \ \sin^2\phi \int_0^{\pi} d\theta \ \sin^3\theta \ e^{-\beta H[\theta,\phi]}}{Z}$$

$$(2.38) \qquad = \ \frac{\int_0^{\pi} d\theta \ \sin^3\theta (\int_0^{2\pi} d\phi \ \cos^2\phi \ - \int_0^{2\pi} d\phi \ \sin^2\phi) e^{-\beta H[\theta,\phi]}}{Z}.$$

Therefore (again, taking the Boltzmann factor outside of the integral over ϕ):

(2.39)
$$\int_0^{2\pi} d\phi \ \cos^2 \phi - \int_0^{2\pi} d\phi \ \sin^2 \phi = 0$$

(2.40)
$$\Rightarrow \quad q_2 = 0.$$

We now show that the first line of the previous statement is true.

(2.41)
$$\int_0^{2\pi} d\phi \ \cos^2 \phi - \int_0^{2\pi} d\phi \ \sin^2 \phi = \int_0^{2\pi} d\phi \ (\frac{1 + \cos 2\phi}{2}) - \int_0^{2\pi} d\phi \ (\frac{1 - \cos 2\phi}{2})$$

(2.42)
$$= \pi + \frac{1}{4}[\sin 2\phi]_0^{2\pi} - \pi + \frac{1}{4}[\sin 2\phi]_0^{2\pi}$$

(2.43)
$$= 0$$

(2.44)
$$\therefore q_2 = 0.$$

Therefore, H_1 can be written in terms of Q_0 only:

(2.45)
$$H_1 = -\frac{K}{6}q_0 Q_0 + \frac{D}{3}Q_0.$$

Even as q_0 is the only nonzero expectation value (in general) in our theory, it is also the appropriate choice for quantifying the observed orientations of *in-vivo* kDNA strands parallel to the kinetoplast's symmetry axis. We can write this expectation value via its formal definition, and we call this equation the *mean-field equation* (MFE):

(2.46)
$$q_0 = \frac{\int d\Omega \ Q_0(\theta) \ e^{\frac{-H_1(\theta,q_0)}{T}}}{\int d\Omega \ e^{-\frac{H_1(\theta,q_0)}{T}}},$$

where the measure, $d\Omega$, represents the solid angle. This is a self-consistency equation since q_0 appears on both of its sides. This equation allows us to study the orientational phase behavior of the system with respect to the symmetry axis.

In order to write the MFE in full detail, we substitute H_1 into both integrands. We also use the fact that neither integrand depends on ϕ to cancel the solid-angle contribution

35

from ϕ:

$$(2.47) \qquad q_0 = \frac{\int_0^\pi d\theta \ \sin\theta \ (3\cos^2\theta - 1) \exp[-\frac{(3\cos^2\theta - 1)}{3T}(D - \frac{K}{2}q_0)]}{\int_0^\pi d\theta \ \sin\theta \ \exp[-\frac{(3\cos^2\theta - 1)}{3T}(D - \frac{K}{2}q_0)]}.$$

2.2.1. Entropy maximization leads to disorder in the high-temperature limit.

In the limit as $T \to \infty$, the exponentials in the integrands of both integrals in the MF equation go to one, because H_1 remains finite. To evaluate the remaining terms in these integrals, we perform a change of variables:

$$(2.48) \qquad x \ = \ \cos\theta$$

$$(2.49) \qquad dx \ = \ -\sin\theta.$$

Observe that under this transformation, the upper and lower limits of the integrals become, respectively, -1 and 1. With this change of variables, we find that the MF equation evaluates to zero:

$$(2.50) \qquad q_0 \ = \ \frac{\int_{-1}^1 dx \ (3x^2 - 1)}{\int_{-1}^1 dx}$$

$$(2.51) \qquad q_0 \ = \ 0.$$

This outcome agrees with our expectations: for large temperatures the free energy, $F = U - TS$, will maximize the second, entropic by adopting a disordered state characterized by $q_0 = 0$.

2.2.2. Analytical Approaches to Studying the Low-Temperature Limit.

In the absence of the high-temperature simplification made in the previous section—but using the same change of variables—the MF equation can be written in the expanded form:

$$(2.52) \qquad q_0 = \frac{\int_{-1}^1 dx \ (3x^2 - 1) \exp[-\frac{(3x^2 - 1)}{3T}(D - \frac{K}{2}q_0)]}{\int_{-1}^1 dx \ \exp[-\frac{(3x^2 - 1)}{3T}(D - \frac{K}{2}q_0)]}$$

36

We observe that the numerator and the denominator of the right-hand side of this equation contain, respectively, a pseudo-Gaussian and a Gaussian integral. However, both have finite limits.

Since Gaussian integrals over the real line evaluate to closed-form values,

$$(2.53) \qquad \int_{-\infty}^{+\infty} dx\ e^{-Cx^2} = (\frac{\pi}{C})^{\frac{1}{2}}$$

(for $C \in \mathbb{R}^+$), we consider first the case where the integrand constants in the MFE are sufficiently large. Specifically, we require that the standard deviation of the Gaussian function is small with respect to 1, i.e:

$$(2.54) \qquad \frac{1}{T}(D - \frac{K}{2}q_0) > 2.$$

The larger the left-hand side of this inequality, the more accurate the approximation. By applying this approximation several times—along with integrations by parts—we find:

$$(2.55) \qquad q_0 = -1 + \frac{3}{\sqrt{\pi}} \left(\frac{T}{D - \frac{K}{2}q_0} \right)^{\frac{1}{2}} \left\{ \frac{1}{2} \left[\pi T (D - \frac{K}{2}q_0) \right]^{\frac{1}{2}} - \exp\left[-\frac{1}{T} \left(D - \frac{K}{2}q_0 \right) \right] \right\}$$

This equation does not have an analytical solution.

2.2.3. Numerical Evaluation of the Mean-Field Equation. Since the MF equation does not have an analytical solution in general, we employed a numerical solution method, which is valid for all temperatures. We use a *grid search* algorithm to systematically find MFE solutions in the plane $K = 1$ within the T-K-D parameter space. Grid search has an advantage over standard root-finding algorithms ("numerical solvers"), which require an initial guess based on physical intuition as to where the answer lies.

For a given small choice of $\epsilon > 0$, our algorithm identifies an approximate solution to the MF equation if the following inequality is satisfied for a given triplet (T, D, q_0):

$$|\langle Q_0 \rangle - q_0| < \epsilon.$$

Since equation 2.52 contains a dependence on T^{-1}, the equation is undefined for $T = 0$. To avoid this issue, we limit our search to $T \geq 0.1$. Evaluating a large number of integrals (i.e, two per grid point) for the MFE introduces a computational challenge. We deal with this by observing that we can express each of these integrals—which are Gaussian or pseudo-Gaussian—in terms of the error function, which can be efficiently evaluated computationally with standard libraries.

Our first result (see Fig. 2.2) is a phase portrait for the plane $K = 1$ with $\epsilon = 0.01$. We study values of D and T that are within one order of magnitude of K in either direction, and we sample the plane in step sizes $dT = 0.1 = dD$. The colormap gives the order parameter expectation value q_0 that satisfies the MFE at each point. As expected, we observe a *crossover-type transition*, where the order parameter decreases continuously and monotonically from a disordered state as temperature decreases (see Fig. **??**). As discussed in section 2.2, a state with $q_0 = 0$ is disordered with respect to the z-component of its spins.

At high temperature we see that states are expected to be disordered in $\mathbf{S}_{iz}$. As the strength of our effective field decreases, states become more disordered. However, $q_0 > 0$ for all values of T, because we only study nonzero field conditions. Finally, at low temperatures
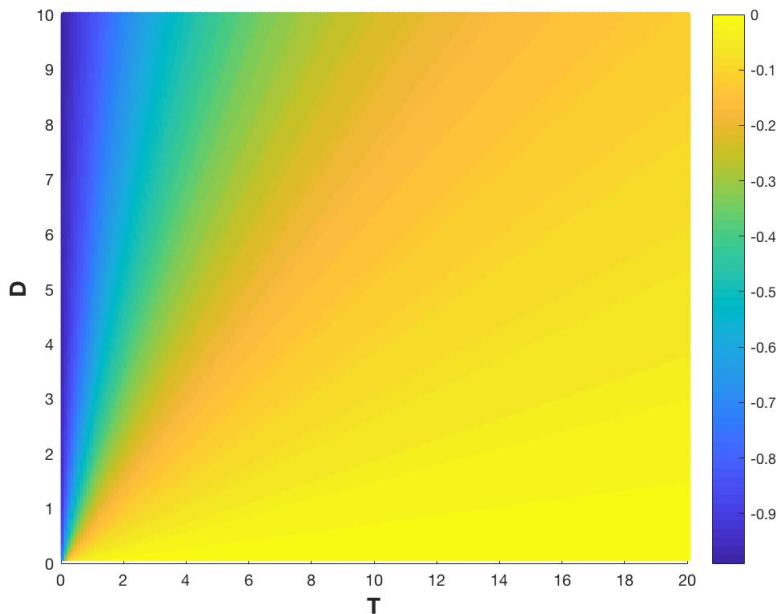
FIGURE 2.2. We used the grid-search algorithm to construct a phase portrait for q_0, which consists of approximate solutions to the MF equation (2.52). For fixed $K = 1$, our analysis predicts a crossover transition to an ordered phase (blue), which is sharp (but continuous) when D and T are small.

the system is expected to enter a highly ordered state with $q_0 \approx -1$. This represents conformation whose spin-directors are nearly parallel to the z-axis, and we note that this order persists at $T = 0.1$ even as D becomes very small.

2.3. Numerical Results

2.3.1. The 2-D anisotropic Lebwohl-Lasher system exhibits a pseudo-transition to field-parallel order in the MFT and in MC Simulations.

Our Monte-Carlo simulations demonstrate a pseudo-transition to field-parallel order at $T_c^* = 0.55$, which appears as a sudden, continuous decrease in the order parameter q_0 (see Fig. 2.3). At $T_{c,0}^*$, $q_0(T)$ exhibits a change in concavity, matching the location of a finite peak in this function's first derivative (which plays the role of heat capacity; see Fig. 2.4).
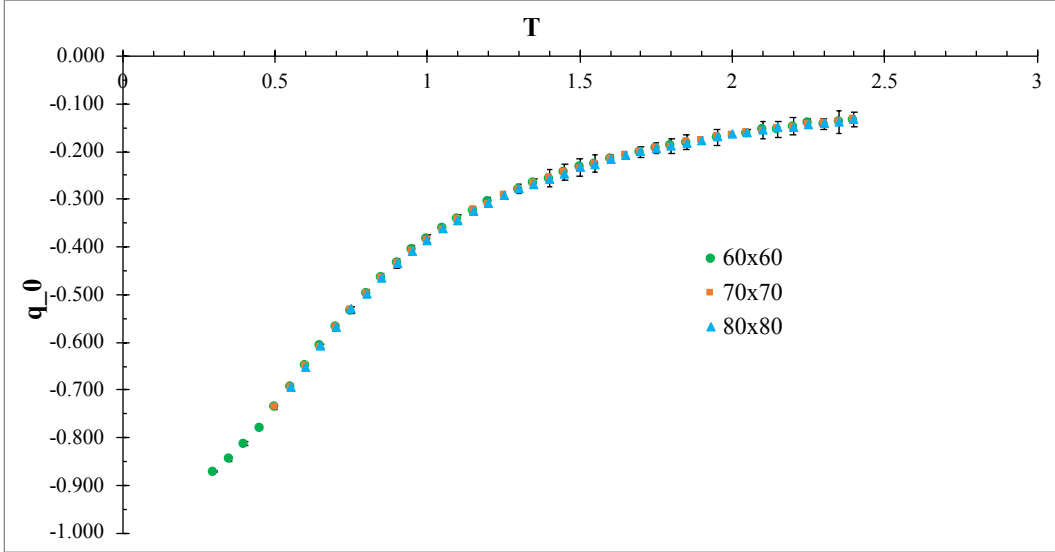
FIGURE 2.3. **Field-parallel pseudo-transition in** q_0: Our MC results indicate a pseudo-transition to field-parallel order. This behavior persists in several studied lattice sizes, with transition temperatures of $T_{c,0}^*(60 \times 60) = 0.55; T_{c,0}^*(70 \times 70) = 0.60;$ and $T_{c,0}^*(80 \times 80) = 0.65$. Standard error is included for the 60×60 data.
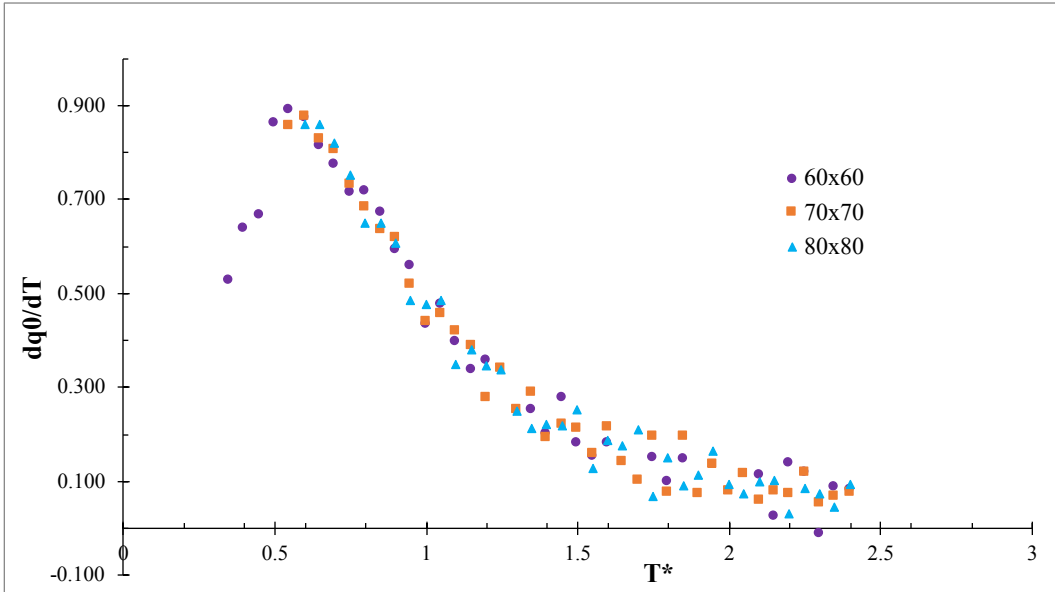

FIGURE 2.4. **Location of** q_0 **Pseudo-Transition** The location of the peak in the first derivative of q_2 indicates the location of the pseudo-transition.

At large temperatures ($T^* \gtrsim 2.5$), q_0 reflects a largely disordered phase, even as the nonzero field maintains $\hat{\mathbf{z}}$ as a preferred direction ($q_0 < 0$).

Note that Fig. 2.3 shows the intersection of the plane $D = 1$ with the 2-D surface represented by the phase portrait shown in Fig. 2.2. This intersection is a curve. Our Monte-Carlo samples for q_0 match our theoretical predictions from the mean-field equation within error for $T > 0.6$. At low temperatures, the MF approach slightly over-estimates order, as we expect. We simulated lattices of various sizes: $60 \times 60, 70 \times 70$, and 80×80, and we plot these results together.

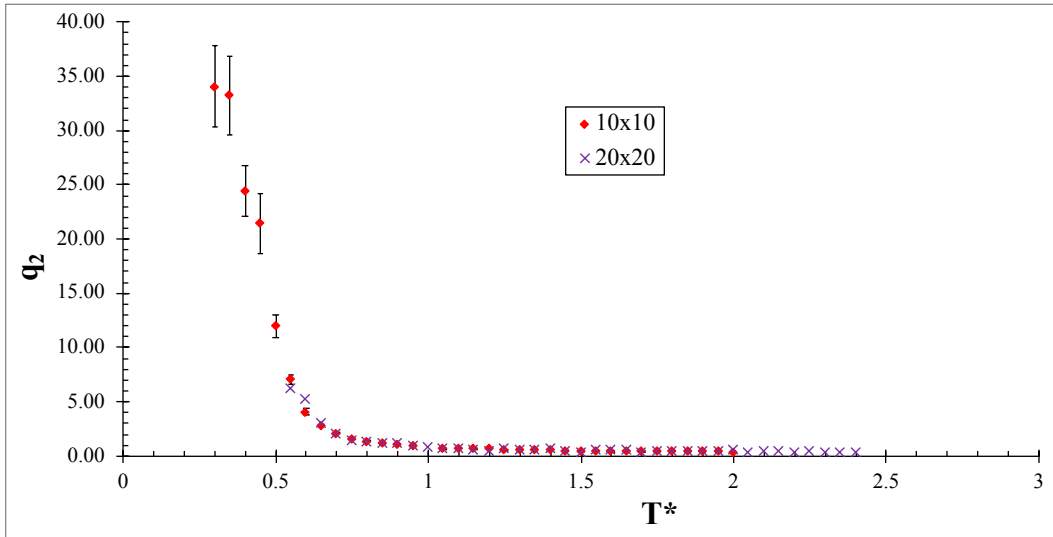

FIGURE 2.5. **Symmetry-breaking pseudo-transition in** q_2: We calculated the q_2 order parameter for several lattice sizes. The error bars (10×10) are standard error, as defined in the Methods.

2.3.2. A spontaneous symmetry-breaking pseudo-transition occurs in the xy-plane. Our simulations for the 10×10 lattice show that the system undergoes a spontaneous symmetry-breaking pseudo-transition at $T^*_{pc,2} = 0.43$ (see Fig. 2.5). This transition breaks the system's rotational symmetry about the z-axis, creating a preferred direction in the xy-plane. This MC result diverges from the MFT's prediction; as discussed in section 2.2, the mean-field expectation value for Q_2 is 0.

2.3.3. The q_2 phase transition coincides with an increase in planar correlation.

We calculated the second moment, ξ_2^2, of the q_2 correlation function, $G_2^{(j)}$ according to:

$$(2.56) \qquad \xi_2^2 \doteq \frac{\sum_{j=1}^{N} r_{jr}^2 \langle G_2^{(jr)} \rangle}{\sum_{j=1}^{N} \langle G_2^{(jr)} \rangle},$$

where the correlation function's expectation value is defined by:

$$(2.57) \qquad \langle G_2^{(j)} \rangle = \langle G_2^{(jr)} \rangle = \langle Q_2^{(j)} Q_2^{(r)} \rangle.$$

In the definitions above, we select a reference spin r near the lattice's center (without loss of generalization due to the Hamiltonian's translational invariance and our use of PBCs in the simulations).

ξ_2^2 increases sharply to a plateau at at $T^* = 0.55$ (see Fig. 2.6), which is close to the location of the q_2 pseudo-transition.
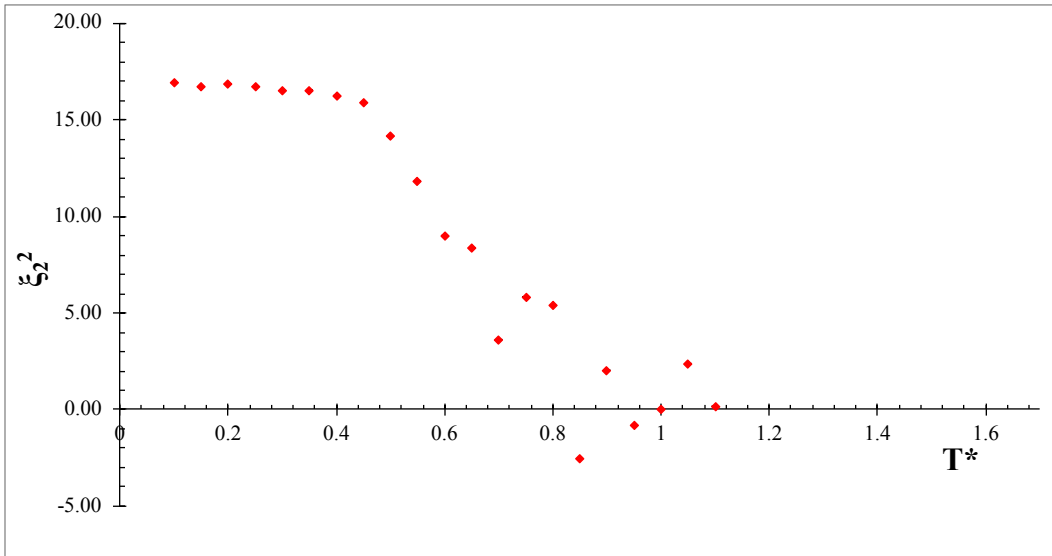

FIGURE 2.6. **Correlation function second moment for q_2:** We calculated the second moment of the correlation in Q_2, ξ_2^2.

Topological Entanglement in Trypanosome kDNA: The Relationship Between Orientational and Topological Order

3.1. Models and Methods

3.1.1. Determining if Two Rigid Circles Are Linked. When we model the minicircles as rigid geometrical circles, we can determine the linking between a pair of minicircles using a straightforward geometrical condition. We prefer this method to the more computationally intensive task of calculating the Gauss integral for the pair of circles [**DHK$^+$12**] In this algorithm we define three planes—one plane to contain each of the two minicircles and a third plane to be discussed shortly—and determine if the point at the intersection of these three planes is within the radii of both circles. As we introduced in section 1.2, a minicircle, C_1, is fully described by the set $C_1 = (r, \mathbf{S}_1, P_1)$, where we have fixed the radius at $r = 1$ and where $\mathbf{S}$ is the minicircle's normal vector and $P_1 = (x_1, y_1, z_1)$ is the minicircle's center.

Let plane 1 be the plane containing C_1 and plane 2 be that containing C_2. These two planes are defined by the familiar equations:

$$\text{plane 1: } 0 = S_{1x}(x - x_1) + S_{1y}(y - y_1) + S_{1z}(z - z_1)$$

$$\text{plane 2: } 0 = S_{2x}(x - x_2) + S_{2y}(y - y_2) + S_{2z}(z - z_2)$$

Now, let's define a third plane, plane 3, which contains the midpoint of the line segment connecting the centers of the two circles and which is perpendicular to this line segment.

The line segment $\overline{P_1 P_2}$ is parallel to the vector

$$
\begin{pmatrix}
x_2 - x_1 \\
y_2 - y_1 \\
z_2 - z_1
\end{pmatrix},
$$

so these are the three components of a normal vector to plane 3. This line segment's midpoint is:

$$
O = \left(\tfrac{x_2 - x_1}{2}, \tfrac{y_2 - y_1}{2}, \tfrac{z_2 - z_1}{2} \right).
$$

Therefore, we can define plane 3 by the equation:

$$
0 = (x_2 - x_1)\left(x - \frac{x_2 - x_1}{2}\right) + (y_2 - y_1)\left(y - \frac{y_2 - y_1}{2}\right) + (z_2 - z_1)\left(z - \frac{z_2 - z_1}{2}\right).
$$

Our three planes are specified by three equations in three variables, so we know that they intersect at a single point, Q. C_1 and C_2 form a Hopf link if and only if the distance separating the points $|P_1 Q| < r$.

3.2. Survey of Results from the *MC-kDNA* Model and its Extensions

3.2.1. Pair-Linking Probability. Initial studies on the *MC-kDNA* model focused on how the linking probability of minicircle pairs changes with their separations and how the different assumptions of the model affect these probabilities. Diao and van Rensburg determined that the linking probability between two minicircles decreases linearly with separation (proof in [**DvR98**]):

THEOREM 3.2.1. *For two geometrical minicircles of unit radius whose centers are at a distance 2r (r > 0) apart and their orientations are uniformly distributed on the unit sphere,*

the probability, p_r, for them to form an unsplittable link is given by:

$$
(3.1) \qquad p_r = \begin{cases} 1 - r, & 0 < r < 1, \\ 0, & r \geq 1. \end{cases}
$$

MC-kDNA with Angle Restrictions

A titling-angle restriction $\theta_0 > 0$ forces the minicircle system to adopt a more "vertical" conformation (*i.e*, their normal vectors are close to the xy-plane). This restriction decreases the linking probability between two minicircles as shown in Fig. 3.1. The relationship
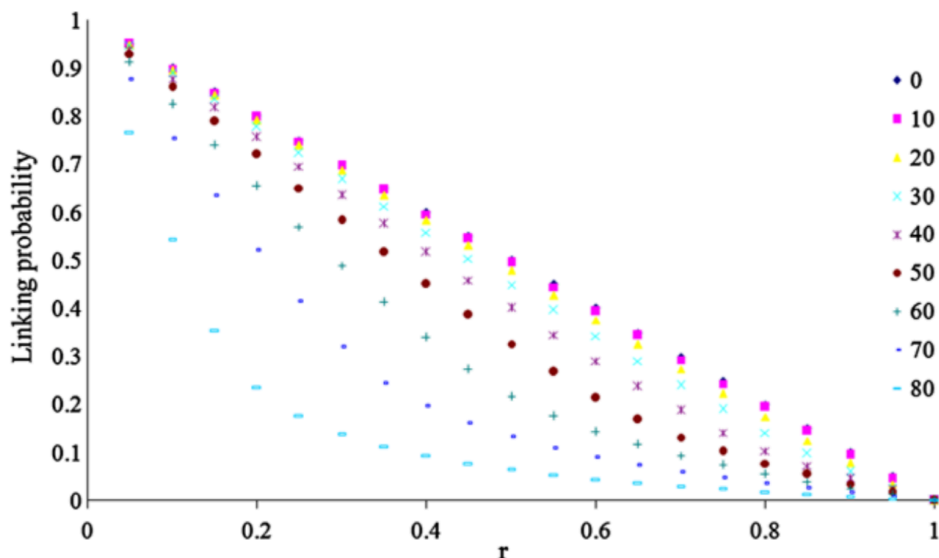

FIGURE 3.1. The linking probability between two minicircles with centers separated by a distance r for various values of θ_0 (in degrees as indicated in the legend). The data is from [**ADH12**].

between tilting-angle restriction and linking probability between two minicircles at fixed separation becoming closer to $P(r) = 1 - r$ when minicircles are very close or very far and having a more pronounced difference when the distance is between 0.2 and 0.6.

MC-kDNA with Steric Effects

Diao and his colleagues used the hard-cylinder model discussed in section 1.2 to estimate the linking probability as a function of minicircle separation for various levels of volume exclusion [**DRKA15**]. Interestingly, we observe that although the probability decreases
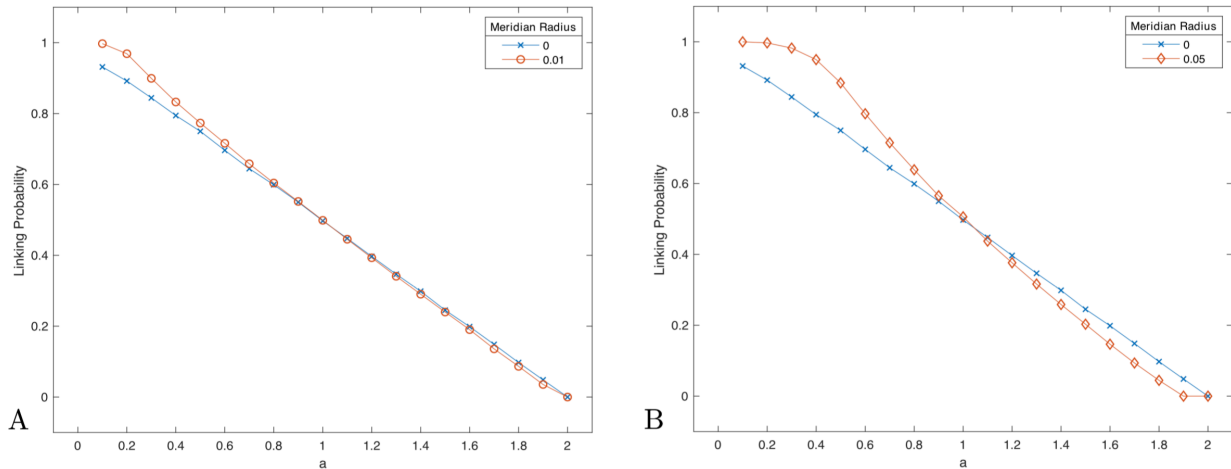

FIGURE 3.2. The linking probability (y-axis) is plotted against the separation (x-axis) for pairs of minicircles with volume exclusion. Each panel shows one curve for thin circles (no volume exclusion) and one curve for the indicated value of cylinder thickness ("meridian radius"). The slopes of the linking-probability functions increase as the degree of volume exclusion increases. Each point corresponds to 10,000 samples. The data is from [**DRKA15**]

monotonically in separation, r, the slopes of the linear portions of the curves increase as the thicknesses of the minicircles increase (see Fig. 3.2). In other words, as the extent of volume exclusion increases, the linking probability becomes more sensitive to minicircle separation. At very small separations, the function becomes nonlinear (and less steeply sloped) as fewer volume-excluding conformations result in a topological link.

MC-kDNA with Flexibile Minicircles

Klenin and colleagues computationally studied the linking probability of two closed, freely jointed chains as a function of r, the separation between their respective centers of mass [**Kea88**]:

# of segments	estimated linking probability
20 segments	$P(r) = 0.682 \exp(-0.214 r^{2.61})$
40 segments	$P(r) = 0.767 \exp(-0.063 r^{2.75})$
80 segments	$P(r) = 0.871 \exp(-0.048 r^{2.24})$

In a similar study, Arsuaga and his colleagues studied chain lengths in the range 16–20 [**ADR14**]. The results are summarized in Figure 3.3, where an exponential decay phe-
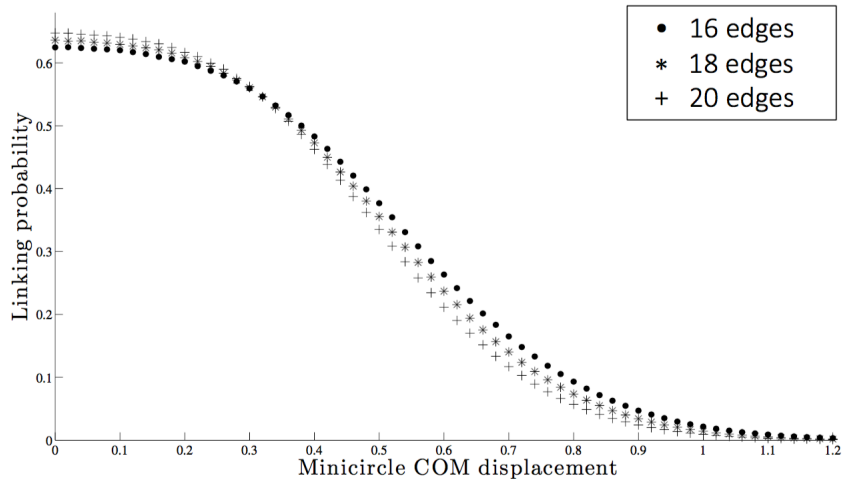

FIGURE 3.3. Linking probability between two polygonal minicircles modeled by closed freely jointed chains as a function of the distance between their centers of mass (COM displacement). The data is from [**ADR14**].

nomenon is also apparent. In the case of flexible minicircles, it's possible that topological links more complicated than Hopf-type might form. However, given the relatively short length minicircle arc length (with respect to the Kuhn length), it might be unlikely for links other than the Hopf-type to form and they might only occur rarely.

Arsuaga and his colleagues addressed this question by studying the distribution of linking numbers among pairs of FJCs [**ADR14**]. Their results indicate that for polygons with $n = 16, 18$, and 20 segments, the majority of linked pairs have absolute linking numbers of 1. For example, in a grid dimension 100×100 with $n = 16$ edges, these authors found that less than 15% of linked pairs have absolute linking number larger than 1 with densities near the mean percolation density. These percentages only increased slightly when the density increased to values near the mean saturation density, indicating that most of the linked pairs we encounter are indeed Hopf links.

3.2.2. Minicircle Mean Valence. Next, we describe how the mean valence changes as we change the different parameters of the simulation. As before we first describe the behavior of the mean valence for the initially proposed model and then present our results on how this behavior changes. Unless otherwise noted, the numerical results presented in this chapter were obtained using sample grids of dimensions ranging from 100×100 to 1000×1000 with a sample size of 1000.

MC-kDNA with Angle Restrictions

In [**ADH12**] the authors considered values for the tilting angle restriction of $\theta_0 = 0, 30, 60, 87$ degrees with no azimuthal-angle restriction. The authors' numerical results demonstrate that there is a strong linear relationship between the mean valence and the minicircle density as shown in Figure 3.4. The simulations show that the mean valence decreases as tilting-angle restriction increases. However, for a fixed tilting-angle restriction, the relationship between mean valence and density remains linear, with regression equations given below (all with $R^2 > 0.9$):
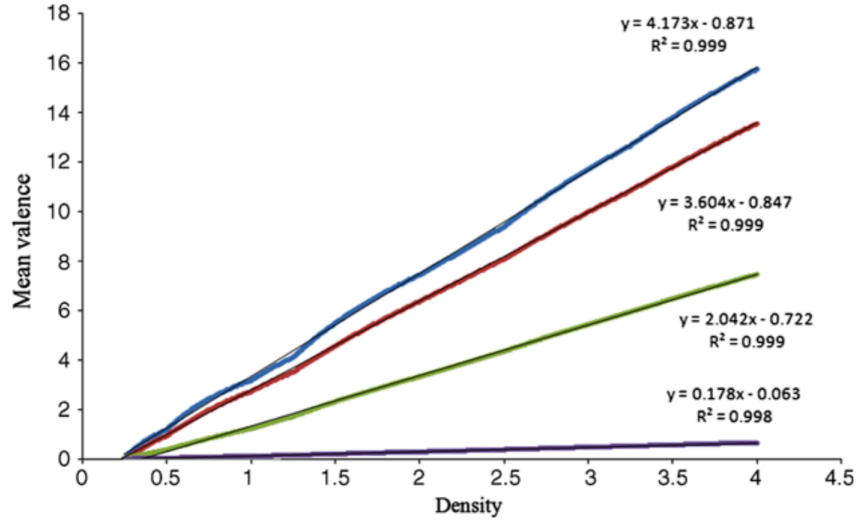
FIGURE 3.4. Minicircle mean valence is linear in density for each value of the tilting-angle restriction. These values are—from top to bottom—$\theta_0 = 0, 30, 60, 87$ degrees [**ADH12**].

angle restriction	linear predictor
$\theta_0 = 0$ (std. *MC-kDNA*)	$y = 4.1732x - 0.8716$
$\theta_0 = 30$	$y = 3.604x - 0.847$
$\theta_0 = 60$	$y = 2.042x - 0.063$

These results indicate that when minicircle orientations are unrestricted (*i.e*, sampled uniformly from points on the unit sphere), the *MC-kDNA* model predicts a mean valence much larger than the experimentally determined value from Chen et al. 1995 at physiological densities [**ADH12**]. As noted previously, electron microscopy studies indicate that kDNA minicircles adopt average orientations that are parallel to the plane of the kinetoplast disk [**DRKA15**].

MC-kDNA with Flexible Minicircles

In their 2014 work, Arsuaga and his colleagues studied the flexible-minicircle *MC-kDNA* model [**ADR14**]. As shown in Fig. 3.5, the authors found a linear relationship between mean
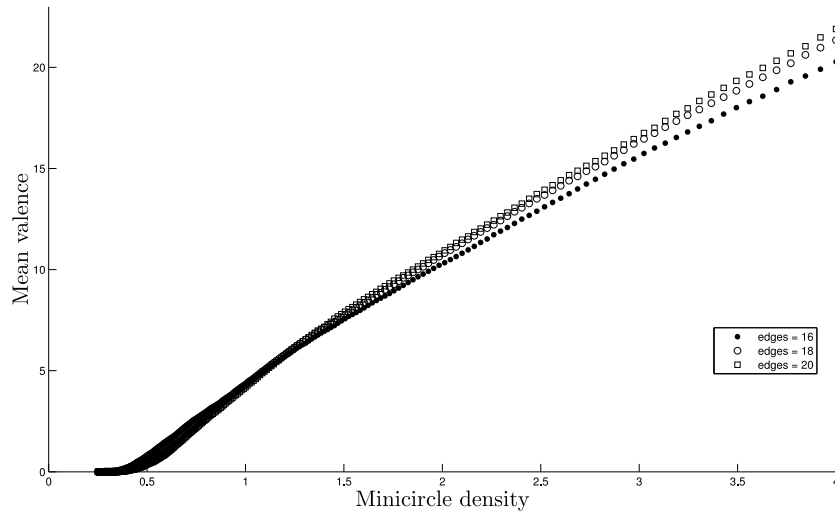

FIGURE 3.5. Minicircle mean valence is linear in density for each value of the tilting-angle restriction. These values are—from top to bottom—$\theta_0 = 0, 30, 60, 87$ degrees [**ADR14**].

valence and circle density. There are a few features present in Arsuaga's results, however, that are not present in the rigid-circle *MC-kDNA* model. First, there is a nonzero linking probability for $\rho < 0.5$, while rigid minicircles cannot link above this density. Second, at high densities the mean valence is higher that for rigid circles, which suggests that flexible minicircles have a higher chance of linking with non-neighboring minicircles. Consistent with this observation, longer chains have a higher mean valence than shorter ones.

Minicircle mean valence increases linearly with density.

Diao and his colleagues showed analytically that the mean-valence of the square-lattice *MC-kDNA* model increases linearly with increasing density [**DHK$^+$12**]:

THEOREM 3.2.2. *For large values of the density, the mean valence, $\langle V \rangle$, of any given minicircle in the square-lattice MC-kDNA model is $O(\rho)$. More specifically, we have:*

$$(3.2) \qquad\qquad 0.9\rho < \langle V \rangle < \frac{16}{3}\rho.$$

This result is consistent with the experimental observation that when the number of minicircles doubles during kDNA replication, the mean valence doubles from three to six [**DHK$^+$12**]. We discuss the theoremÕs proof in the following paragraphs.

We begin by observing that, since minicircles have radius 1, they must be separated by a distance $d \in (0, 2)$ for linking to be possible. For a square lattice, a particular site
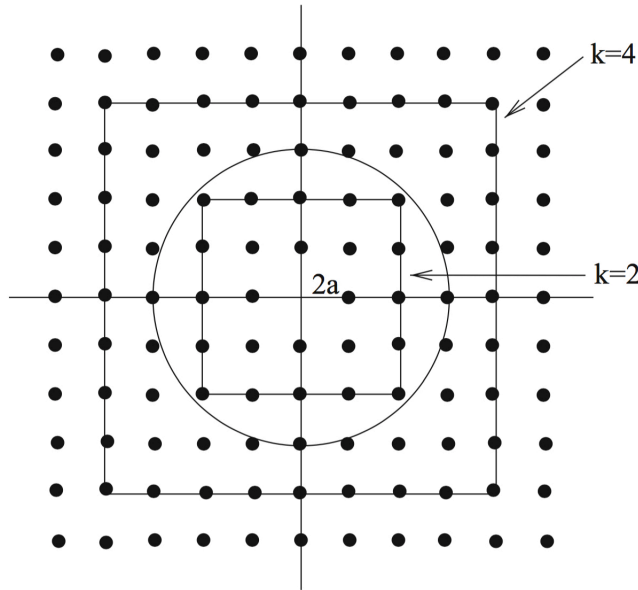

FIGURE 3.6. A central minicircle is surrounded by a series of concentric squares labelled by k (adapted from [**DHK$^+$12**]).

is surrounded by nested squares, each identified by an integer $k \in \mathbb{Z}^+$ (where the square

$k = 1$ contains the nearest and next-nearest neighbors of the center lattice site) (see Fig. 3.6). The sites contained in a particular square are separated from the center by a distance $d \in [2ak, 2\sqrt{2}ak]$, where a closest site is at the center of one of the squareÕs sides and a farthest site is at one of the square's corners. Note that the authors choose to make the lattice constant $2a$ in this proof, rather than a.

The remainder of this proof consists of counting the number of lattice sites a distance $d < 2$ from the center site, and then summing the linking probabilities between the center minicircle and each of these others. For a pair of minicircles separated by $d \in (0, 2)$, Diao and his colleagues show in Corollary 1 that

$$P_{\text{link}} = 1 - d.$$

Now, the reader can quickly verify by counting (again, see Fig. 3.6) that the square k has $8k$ lattice sites (*e.g*, square $k = 2$ has 16 lattice sites). Therefore, we expect that the square k contributes a quantity $8k(1 - d)$ to the mean valence of the center circle. For a given lattice constant $2a$, $k < 1/a$. In other words, lattice sites outside this square correspond to circles that are too far away to link with the center circle.

Given the maximum and minimum distances established in the second paragraph of this proof for square k, we can sum over all of the squares surrounding the center lattice site to find the bounds on mean valence:

$$(3.3) \qquad \sum_{k=1}^{1/a} 8k(1 - \sqrt{2}ak) \le \langle V \rangle \le \sum_{k=1}^{1/a} 8k(1 - ak).$$

We evaluate these sums as follows:

$$(3.4) \qquad 8\sum_{k=1}^{1/a} k - 8\sqrt{2}a\sum_{k=1}^{1/a} k^2 \le \langle V \rangle \le 8\sum_{k=1}^{1/a} k - 8a\sum_{k=1}^{1/a} k^2.$$

Now, we assume that a is small (ρ is large), which allows us to treat all values of k as integers. Under this assumption, both of these series have well known closed-form values:

(3.5)
$$\frac{4}{a}\left(\frac{1}{a}+1\right) - \frac{4\sqrt{2}a}{3}\frac{a}{a}\left(\frac{1}{a}+1\right)\left(\frac{2}{a}+1\right) \leq \langle V \rangle \leq \frac{4}{a}\left(\frac{1}{a}+1\right) - \frac{4}{3}\frac{a}{a}\left(\frac{1}{a}+1\right)\left(\frac{2}{a}+1\right),$$

which reduces to:

(3.6)
$$0.92\frac{1}{a^2} + O\left(\frac{1}{a}\right) \leq \langle V \rangle \leq \frac{4}{3}\frac{1}{a^2} - \frac{4}{3}.$$

For the left-hand side of this inequality, we observe that for small a the inverse-square term dominates the $O(1/a)$ terms. Therefore, we substitute $\rho = 1/(4a^2)$ and we find:

(3.7)
$$0.9\rho \leq \langle V \rangle \leq \frac{16}{3}\rho.$$

Importantly, this result tells us that mean valence is linear in density. Diao and his colleagues computationally confirmed this behavior in the same work.

The mean valence has a low sensitivity to the distribution of minicircle centers.

Arsuaga and colleagues studied the relationship between the mean valence and the underlying (regular) lattice [**ADH12**]. The following table summarizes their findings:

lattice type	linear predictor
hexagonal	$y = 4.157x - 0.825$
triangular	$y = 4.1451x - 0.8329$
square	$y = 4.1732x - 0.8716$

These results shows that, among the studied regular lattices, the spatial distribution of minicircle centers has a rather small impact on this generally linear relationship as stated in Theorem 2. In all cases, $R^2 > 0.9$.

3.3. Results

3.4. *T. brucei* Minicircle Topology

3.4.1. Experimental Determination of Minicircle Topology. Chen and his colleagues determined the mean valence of the *Crithidia fasciculata* minicircle network to be three [**CRW**⁺**95**]. We propose to reproduce these results for *C. fasciculata*, and then to extend them to determine the topology of the *T. brucei* kinetoplast. In Chen's assay, the authors partially digested isolated (purified) *C. fasciculata* kDNA networks with the restriction enzyme XhoI, which cuts each minicircle at a single location. The authors carried out a time-course assay, preparing mixtures containing a ratio of 1 μg of kDNA to 10 U of XhoI. These reactions occurred at 37°C for 10-90 minutes, before XhoI was heat-deactivated for 10 minutes at 70°C.

Because of the sequence heterogeneity in *T. brucei*, we will use *E. coli* topo-IV, a type-II topoisomerase, instead of XhoI (topo-IV is not site-specific). Topo-IV is commercially available from TopoGEN. Type-II topoisomerases like topo-IV can perform strand passage to unlink a pair of linked minicircles. Under this type of action, the valences of both minicircles decrease by one. An important difference between using XhoI and a type-II topoisomerase is that, in the former case, the enzyme linearizes a given minicircle, necessarily removing it from the network. However, in the latter case, the enzyme will instead change the valence by one—the minicircle may remain linked to other circles in the network.

In the case of XhoI, partial digestion means that only a subset of the minicircles are cut by this restriction enzyme. As we just discussed, when XhoI acts on a *C. fasciculata* minicircle, the minicircle is linearized—it is released from the network and the valences of all of the circles to which it was previously linked are decreased by one. On the other hand, topo-IV acts on a pair of linked minicircles, and only these two circles see a decrease in their valences.

Chen and colleagues proposed using linking graphs to mathematically model the topology of the minicircle system. Linking graphs are objects containing a set of nodes, which each represent a minicircle, and edges, which can connect two nodes, thereby signifying that these two minicircles are linked. Therefore, XhoI partial digestion means that a subset of the nodes in the linking graph are removed; topo-IV partial digestion, on the other hand, means that a subset of the edges in the graph are removed, leaving the nodes alone.

3.4.2. Simulated *T. brucei* Unlinking Assay. We expect the distribution of minicircle monomers, dimers, and trimers released from the network to be different in the case of topo-IV treatment than in the case of XhoI treatment. In the previous section, we explained how the action of topo-IV can be modelled mathematically by the elimination of an edge in the linking graph.

Simulation Methods

We carried out a preliminary study of the expected distribution of connected components (i.e: monomers, dimers, etc.), extending software used in [**ILK$^+$18**]. In these simulations, we generate 70×70 square lattices of randomly oriented minicircles using the method proposed in [**DHK$^+$12**], which is explained in section 1.2. We then simulate topo-IV unlinking action by randomly unlinking each pair of linked minicircles with a specified probability. For each value of the unlinking probability, we generated10 minicircle lattices.

Predicted Oligomer Distribuions

When we compare our distributions (see right panel of Figure 3.7) from simulated topo-IV digestion to Ibrahim and colleagues' results for XhoI digestion (left panel), we note first that the fraction of minicircles that are released from the networks in oligomer form is much higher for the simulated topo-IV results. For example, only 10% of the minicircles from the XhoI simulation are released as intact monomers at $p = 0.7$, whereas more than 50% of those
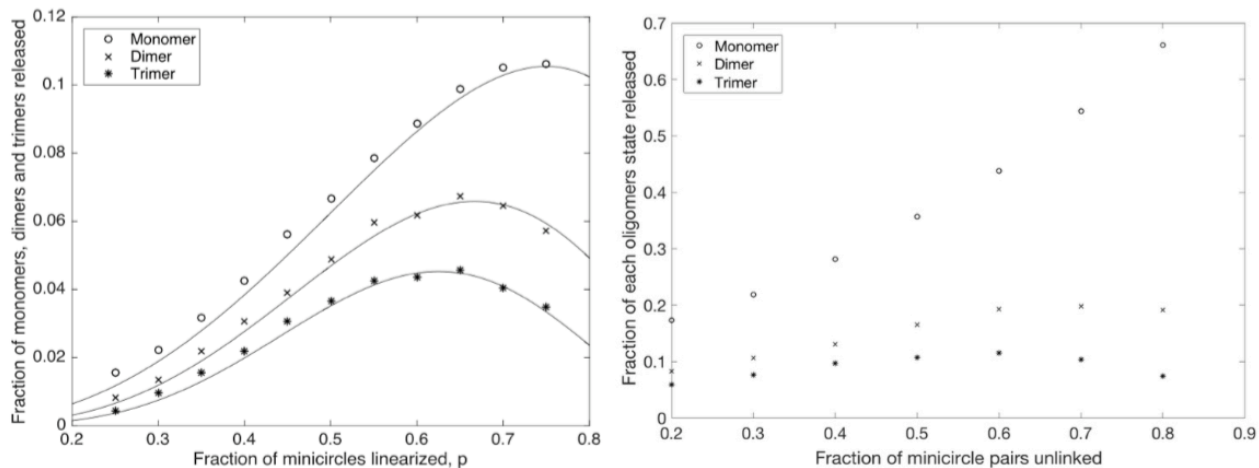
FIGURE 3.7. Left: Ibrahim and colleagues simulated minicircle linearization via XhoI on a square lattice [**ILK$^+$18**]. Right: Our simulations of minicircle unlinking via topo-IV on a square lattice

in our topo-IV study are released as monomers. This is the result we expected. Because XhoI is a restriction enzyme, it linearizes each minicircle that it acts on. Topo-IV, on the other hand, simply unlinks pairs of linked minicircles. Since no minicircles are linearized, we expect the fractions of each oligomer state to be higher than its counterpart in the XhoI simulations.

Moreover, in the XhoI simulations, all three oligomer states begin to decrease as p exceeds a certain value. Biologically, this also makes sense, because all three oligomers states become less likely when XhoI linearizes a large fraction of the minicircles. On the other hand, the fraction of monomers in the topo-IV simulations is monotonically increasing in p. This simply means that, as more pairs of circles are unlinked, individual intact minicircles are increasingly likely to be released from the network. Finally, we note that the vertical ordering of the three states' distributions is the same in both simulations, i.e, monomers are more common than dimers, which in turn are more common than trimers.

56

Determing Mean Valence from Oligomer Distributions

Given the distributions of minicircle oligomers that Chen and colleagues observed in their partial digestion assay, these authors used a formula that they derive in their study to calculate the mean valence:

$$V = \frac{\log m - \log N - \log p}{\log q}.$$

In this formula, m is the average number of intact monomers released from the network, and p is the fraction of monomers linearized. The authors work under the assumptions that every minicircle in the network has an equal probability of encountering the enzyme Xhol (being cut) and that the minicircles have homogeneous valence. The authors observe that the ratio of released circles to total circles—m/N—is given by pqV. In other words, this ratio is equal to the product of the probability for a circle to remain intact, p, and probability that the V circles linked to it are all cut by the enzyme, qV.

In the case of *T. brucei* experiments, this formula (which was derived for the XhoI assay) does not apply. Instead, we propose a simulation-based approach. We would generate conformations of randomly oriented minicircles according to the method described in [**DHK$^+$12**]. For a density greater than the critical percolation density, $\rho_{\mathrm{perc}} \approx 0.637$, (which ensures that a relatively large number of minicircles will be linked) [**DHK$^+$12**], we will calculate the linking between each pair of minicircles using the algorithm proposed in [**DHK$^+$12**], and then determine the average valence for each conformation.

After this, we will group together conformations whose mean valences are close (within a small range) to integer values—for simplicity—between one and twenty. This range is motivated by the mean-valence predictions discussed in section 3.2. For each set of conformations (associated with a particular integer mean valence), we will perform the simulated topo-IV unlinking assay, and determine the corresponding distributions of monomers, dimers, and trimers, similar to those shown in Figure 3.7. We will compare the oligomer distributions

from our topo-IV experiment with these simulated distributions to determine the mean valence that most closely matches our experimental results.

CHAPTER 4

The S477N mutation confers a potential binding fitness advantage consistent with global SARS-CoV-2 population dynamics.

4.1. Methods

Genomic Data Set and Sequence Pre-Processing. We obtained SARS-CoV-2 sequences for this study from the GISAID database on Nov 11, 2020 [**SM17**]. Our data set contains FASTA files for every complete human SARS-CoV-2 nucleotide sequence (from all geographical locations) available in GISAID. The sequences were then aligned using `ClustalOmega` with the default parameters [**SH18**]. We found that `ClustalOmega` ran faster on our data set than common alternatives like `ClustalW` [**THG94**] and `MUSCLE` [**Edg04**].

After aligning the sequences, we extracted the spike protein by comparing the aligned sequence with the NCBI's SARS-CoV-2 reference sequence (NC_045512.2; "WT") [**WZY$^+$20**]. We then removed any sequences that contained missing nucleotides. We organized the remaining 149,354 spike-protein sequences chronologically by collection date and partitioned them into 109 sets, with each set containing an average of 1,300 sequences. We chose to prioritize data sets of relatively equal size rather than choosing equal-length time windows. This choice scales the pHMM probabilities similarly for each data set.

A Profile Hidden-Markov Model Identifies Point Mutations. pHMM's are probability models that identify the evolutionary changes of an input set of sequences [**Dur13**]. A particular model is a forward-feeding machine defined by three states (excluding the arbitrary start and end): match, insert, and delete. Match is the model's default state and nucleotides are emitted from this state with a probability specified by the input data. The

insert state represents a sequence insertion and has nucelotide emission probabilities determined by the occurrence of individual nucleotides in the input data set. Unlike match and insert, the delete state does not emit any nucleotides. pHMM outputs are contained in three matrices: match-state emission probabilities, insert-state emission probabilities, and transmission probabilities. We built a pHMM and calculated the associated matrices for each of the 109 sequence partitions. We used the match-state probabilities to determine the genomic locations and frequencies of point mutations.

After building a pHMM for the full global data set, we sorted the data by the reporting country. For the country-level results (see Figs. 4.2 and 4.3), we plot a three-bin moving average, which accounts for the variability in day-to-day reporting for each country. In Fig. 4.3 we show the fraction of each country's reported sequences in a given three-bin period. Gaps in two of the time-series reflect time points with an undefined moving average (i.e, zero denominator). France is the only country with reported double-mutant samples that we omitted due to its small sample size.

Molecular modeling software predicts S-ACE2 binding affinity. We used the molecular-modelling package YASARA [**KV14**] to substitute individual residues and to search for minimum-energy conformations on the resulting modified structures of the S-ACE2 complex. For all of the structures, we carried out an energy-minimizatio (EM) routine, which include steepest descent and simulated annealing (until free energy stabilizes to within 50 J/mol) minimizations to remove clashes. For the WT and S477N complexes, we subsequently performed molecular-dynamics (MD) simulations. Each simulation ran for 10 nanoseconds using the AMBER14 force field [**MMK$^+$15**] for solute, GAFF2 [**WWC$^+$04**] and AM1BCC [**JJB02**] for ligands, and TIP3P for water. The cutoff was 8 Å for Van der Waals forces (AMBER's default value [**HAO$^+$06**]) and no cutoff was applied for electrostatic forces (using the Particle Mesh Ewald algorithm [**EPB$^+$95**]). The equations of motions were integrated with a multiple timestep of 1.25 fs for bonded interactions and 2.5

fs for non-bonded interactions at $T = 298$ K and $P = 1$ atm (NPT ensemble) via algorithms described in [**KV15**].

We used the 7A94 post-fusion S-ACE2 complex [**BWX$^+$20**] for our WT and S477N calculations. 7A94 is a 3.90-Å cryo-EM structure that consists of a glycosylated S homotrimer bound to a single ACE2 protein (ACE2 residues 19-615). Yurkovetskiy and colleagues report that the D614G substitution induces substantial structural changes, mediated by attenuated inter-protomer interactions. For the two strains containing the D614G mutation, we consturcted a hybrid PDB template. Specifically, we used the 7KDL structure—a 2.96-Å cryo-EM glycosylated S homotrimer containing G614—to model the majority of the S trimer [**?**]. In order to get a biologically plausible ACE2 binding mode (since there are no published post-fusion complexes with D614G), we used YASARA's superpose tool [**KV14**] to structurally align, then replace, the single 7KDL "up" RBD with 7A94's bound RBD. This preserves the positioning of 7A94's ACE2 chain with respect to the latter RBD, which also contains a more complete set of atomic coordinates than the 7KDL RBD. We used the full S-trimer in all of our simulations to account for the inter-protomer interactions and to be consistent between experiments.

Prior to calculating the free energy, we carry out several pre-processing steps on the structure including an optimization of the hydrogen-bonding network [**KDHK12**] to increase the solute stability and a pK_a prediction to fine-tune the protonation states of protein residues at the chosen pH of 7.4 [**KV15**]. We added NaCl ions with a physiological concentration of 0.9 We calculated binding free energy for the energy-minimized structure using the molecular mechanics/generalized Born surface area (MM/GBSA) method [**HWLW01**, **SLT$^+$14**, **CLS$^+$16**], which is implemented by the HawkDock server [**WWW$^+$19**]. We observed in our simulations that HawkDock consistently overestimates the magnitude of binding free energy relative to *in-vitro* methods. To an analyze the residue-level changes in each mutant, we used HawkDock to identify those S-protein residues that contribute most strongly to binding.

Additionally, we used `YASARA` to predict those residues participating in hydrogen bonds for each variant and to calculate the strengths and separations of each bond.

4.2. Results

4.3. Time stratification of sequence data shows frequent emergence of S477N single- and double-mutant strains in multiple countries.
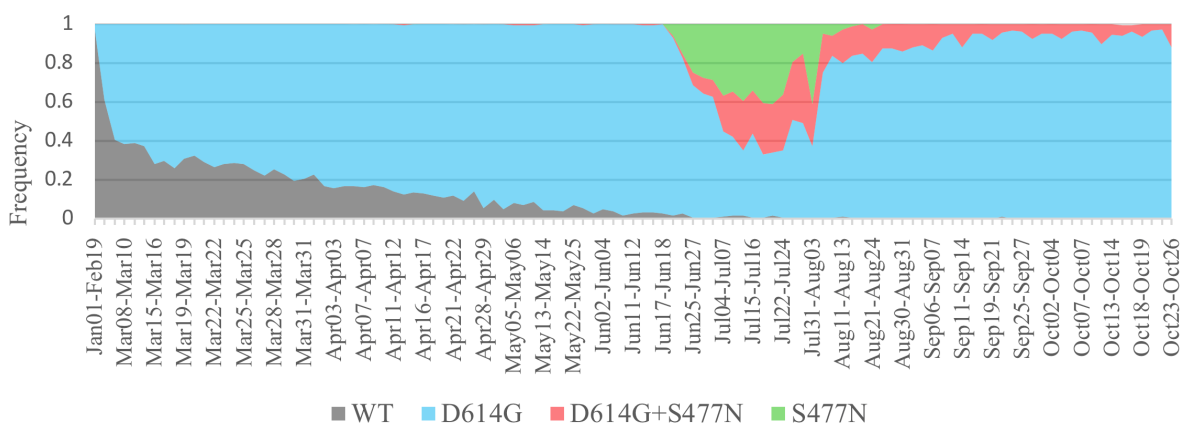

FIGURE 4.1. **Global Population Fractions of 477, 614 Mutants** We show the fraction of sequences in each time window that contains the specified mutation(s) in our full global data set. (Figure kindly provided by Sofia Jakovcevic [**Pea**])

The D614G substitution (a non-synonymous single-nucleotide substitution: A23364G) appears in our first data set, with collection dates between 1/1 and 2/19 (see Fig. 4.1; courtesy of Sofia Jakovcevic). The frequency of this strain increases roughly monotonically in subsequent data sets until the 3/6 - 3/9 interval, at which point GLY becomes the consensus residue in locus 614 (ie, GLY appears in > 50% of sequences). D614G rises to a stable plateau above 90% until mid-June, when we detect a second substitution, S477N (G22992A), which occurs as both a single-mutant (in Australia only—see Fig. 4.2) and a double-mutant strain, S477N/D614G (in Australia and Europe–see Fig. 4.3).

Both S477N and S477N/D614G begin to gain population share in mid-June and in the following two months, all four strains are present in Australia, with the two strains containing
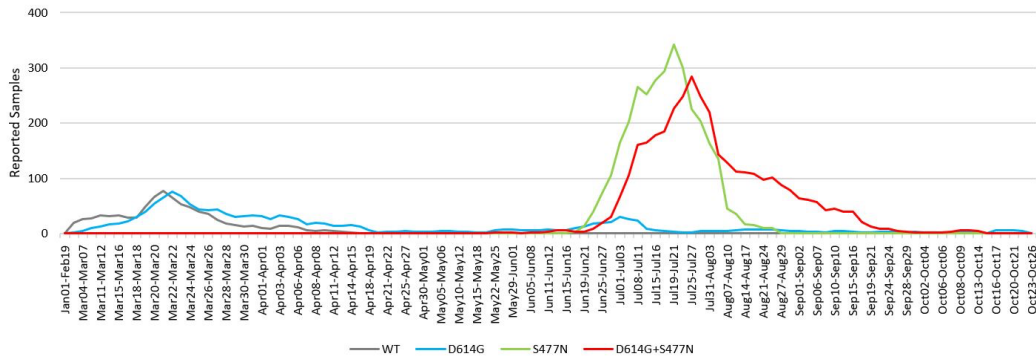
62

FIGURE 4.2. **Population Fractions in Australia** We show the number of reported samples in Australia for each of the four strains. This data is shown as a three-bin moving average. Samples from these four categories account for nearly all reported samples from the country. (Figure kindly provided by Sofia Jakovcevic [**Pea**])
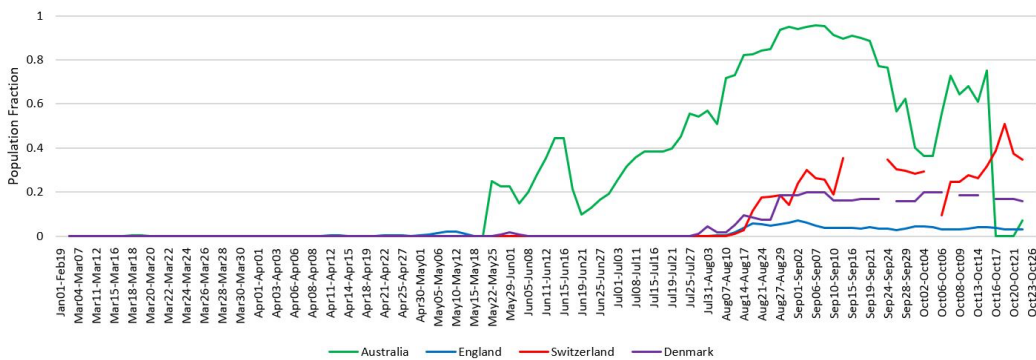

FIGURE 4.3. **Population Fractions in Europe** We calculate the fraction of each country's reported sequences with the S477N/D614G substitutions. Gaps in an individual time series indicate that no samples were reported during the specified period. The data is shown as a three-bin moving average. (Figure kindly provided by Sofia Jakovcevic [**Pea**])

N477 dominating reported samples. Between 7/4 and 7/7, ASN becomes the consensus residue at locus 477 globally. Australian COVID-19 infections fell dramatically in earlier August after the introduction of strict public-health measures, and SER regains majority population share at site 477 globally after the 7/31 - 8/3 period.

We note that the S477N/D614G maintains a large population share in Australia even as overall infections fall. In Europe this strain also maintains a substantial population

share–approximately 20% of the viral population in Denmark and a higher percentage in Switzerland after emerging simultaneously in these countries in early August. From mid-August until the end of October, the viral population returns to a nearly bipartite state, split between S477N/D614G and D614G, which dominates.

MD simulations suggest an advantage for S477N when added to either WT or D614G. We calculated binding free energies for both substitutions individually and for the combined variant. We found that S477N/D614G binds more strongly than does D614G: $\Delta\Delta G = -9.18$ kcal / mol; and that S477N's affinity improves with respect to WT: $\Delta\Delta G = -1.62$ kcal/mol (see Fig. 4.4). While $\Delta\Delta G$ for the former pair is significant outside of error bars, the latter pair is not, due to relatively noisy MD simulations. Pairwise binding improvements for the two strains containing S477N are consistent with their apparent fitness advantages over WT and D614G, respectively, demonstrated by our pHMM results. Specifically, S477N gains population share in mid-June even as WT is nearly eliminated from the viral population, and D614G demonstrates a fitness advantage with respect to the double mutant as it becomes the dominant strain in early August.

As we explain in the Introduction, two recent papers have provided *in-vitro* evidence indicating that D614G's infectivity improvement over WT is not explained by improvements in binding affinity. Since locus 477 is unlikely to affect the structure-modifying effects of the inter-protomer G614 interaction, we believe that binding effects should not substantially diminish the previously noted competitive advantage of a strain containing the D614G substitution alone. Note that we should be cautious in comparing the WT and S477N pair with the two D614G variants, since these two sets of simulations are based on partially different PDB templates.

Our population-dynamics results in Australia (see Fig. 4.2) are consistent with a binding fitness advantage for S477N with respect to WT. These country-level results also show an example of S477N/D614G out-competing D614G, even as the double mutant strain emerges
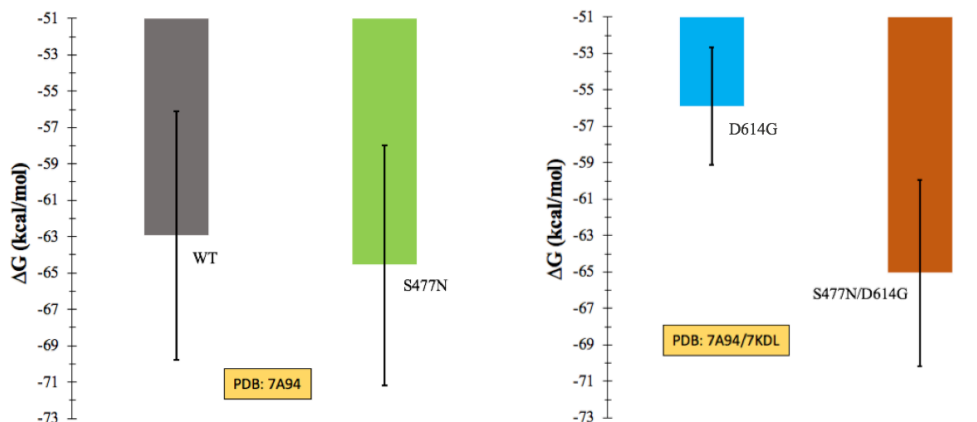
FIGURE 4.4. **MD Binding Affinities** We calculated binding free energies, ΔG, for all four strains using the MM/GBSA method. The two strains containing S477N bind more strongly than their respective counterparts. Calculations for the left and right pairs involved partially different starting PDBs and we are cautious about inter-pair comparisons.

later in several European countries, attaining a stable $\sim 20\%$ population share in Denmark and an increasing share above one-third in Switzerland.

The S477N mutation improves binding by influencing hydrogen bonding for N501. We used YASARA to compare S-ACE2 hydrogen bonds between S and ACE2 in WT and in S477N. We found that the S477N substitution induces more stable hydrogen bonding involving RBD residue N501, which is found at the far end of the S-ACE2 interfacial region (shown in Fig. 4.5) with respect to locus 477. We performed three separate WT MD simulations and four S477N simulations. We observe a more persistent H-bond between acceptor N501 and donor Y41 (in the ACE2 HTH) in the S477N strain (average duration is 2.8 ns) than in WT (average duration is 0.7 ns—see Fig. 4.6). This bond occurs between two oxygen atoms—an N-terminal O connected to the N501 amide group (OD1.N501) and an O connected to the hydroxyl group in Y41 (OE1.Y41). In the WT case, this bond is present for less than 3-ns in trajectories B and C and is completely absent in trajectory A.
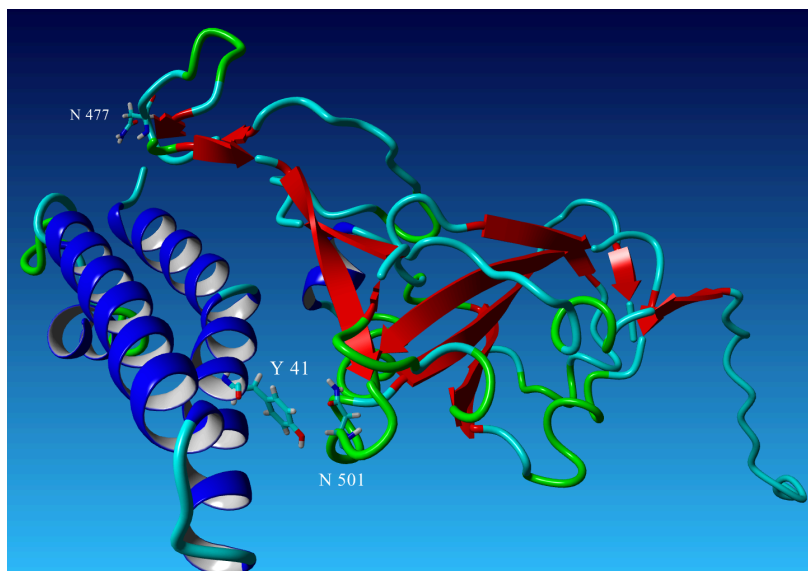
FIGURE 4.5. **S477N's Effects on the Y41-N501 H-Bond** S residue N501 forms hydrogen bonds with ACE2 residue Y41, both of which are found in the interfacial region of the bound complex. This interaction persists longer in the S477N simulations (see next figure) than in WT; an S477N snapshot is shown here. In this image, the RBD is to the right and the HTH is to the left (with the rest of the ACE2 residues removed for a cleaner view).

However, the interaction appears in all four mutant trajectories and is present throughout nearly the entirety of trajectory G.

S477N contributes a possible antibody-escape advantage in a majority of RBD-ab complexes. We performed MD simulations and GBSA binding calculations to measure the effects of the S477N mutation on RBD binding to five neutralizing antibodies: 7CJF, 7KFX, 6XE1, 7KFV, and 7KFW. These antibodies were isolated in sera from recovered individuals, and each one binds to the RBD in the same region that the RBD uses for its primary ACE2 interactions. We found that the S477N substitution weakens RBD-ab binding for a majority of the tested antibodies: 7CJF, 7XE1, and 7KFW (see Fig. 4.8). This
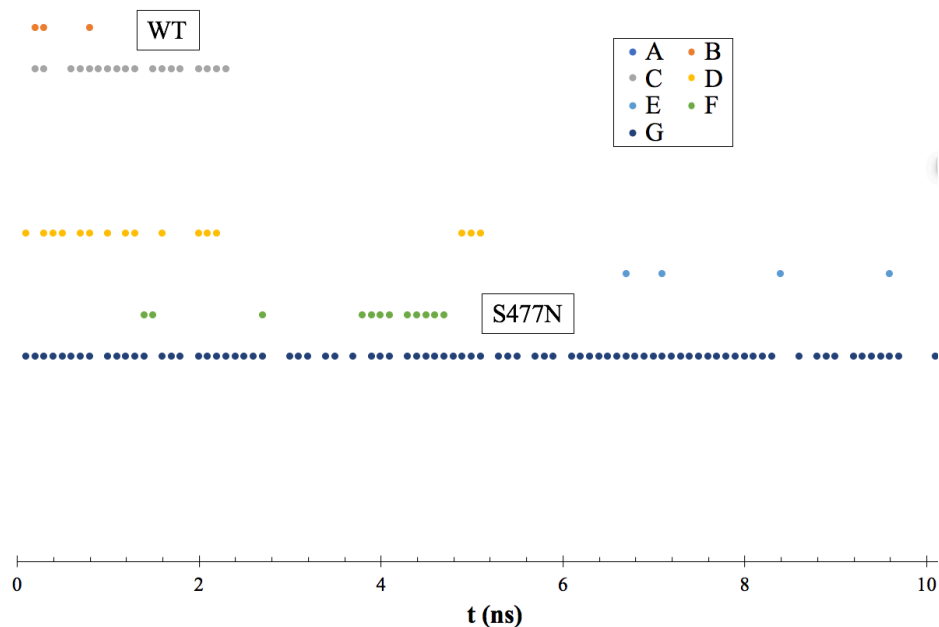
FIGURE 4.6. **Persistent Inter-complex H-Bond in Mutant** RBD residue N501 forms an inter-complex H-bond with ACE2 residue Y41, which persists for longer durations in mutant MD simulations than in WT. Each point indicates that the bond was detected at the specified time step (0.1-ns intervals) in a given simulations. This interaction was never detected in WT trajectory A.

indicates that S477N viral mutants could be less prone to antibody neutralization by antibodies similar to these three molecules. For 7KFX and 7KFV, our simulations predict that similar antibodies will bind more strongly to the mutant, which would decrease antibody escape. In each WT-mutant pair, the magnitudes of the differences, $\Delta\Delta G$, are relatively large, even as the noisy MD simulations lead to large error bars. Increased antibody escape is an additive fitness advantage for the S477N strain, on top of the binding advantage that we report in an earlier section. This advantage is consistent with S477N's empirical fitness advantage in our pHMM analysis
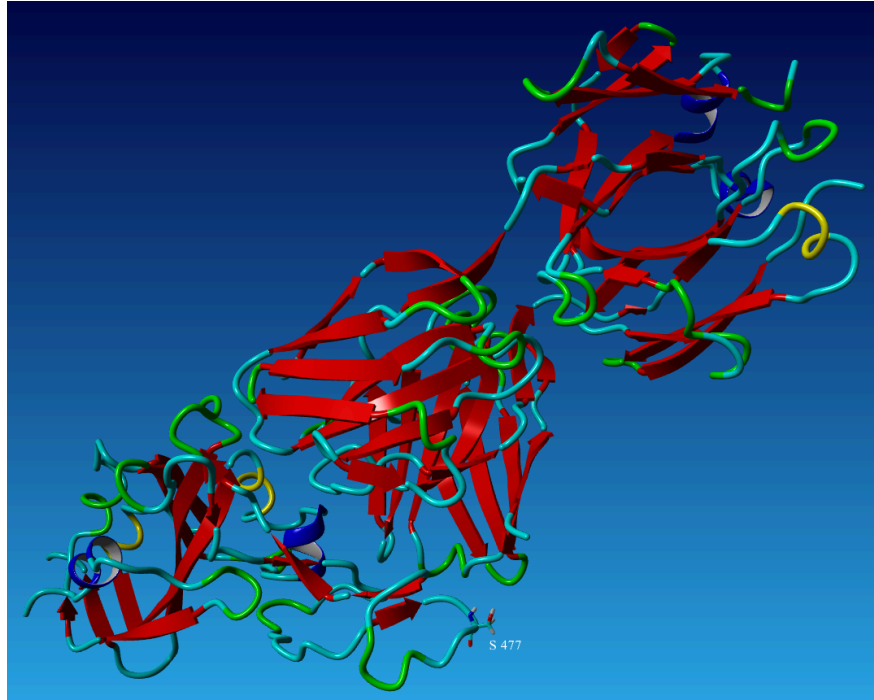
FIGURE 4.7. **WT 7KFW-RBD Complex** We show a simulation snapshot of the 7KFW antibody bound (center, upper right) to a WT RBD (bottom left). We've labeled the RBD substitution site 477 in the images bottom.

FIGURE 4.8. **Predicted ab-RBD Binding Effects** S477N weakens binding in a majority of RBD-ab complexes, which provides a potential fitness advantage by decreasing antibody neutrlization (i.e, in recovered individuals. We report the means and standard deviations of individual 10-ns MD simulations with 5 snapshots spaced by 2 ns.

CHAPTER 5

Discussion

5.1. Models for Minicircle Topology

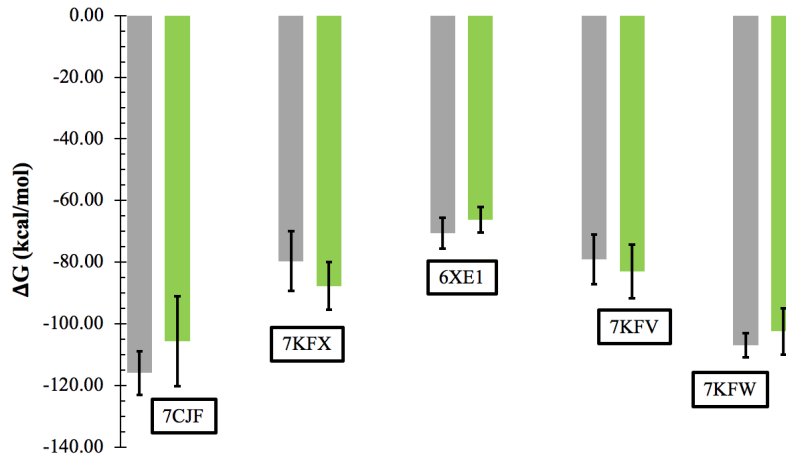The three-dimensional organization of genomes has received significant attention in recent years due to its connections to cell functions. Recent experimental techniques like chromosome conformation capture assays and advances in electron tomography have given insights into chromatin architecture at unprecedented resolutions. Still, determining the topological properties of any genome remains challenging. Topological considerations are important because they enable exploration of the downstream influence of the 3-D genome on a variety of biological processes. Theoretical studies show that DNA molecules in confined volumes are entangled through knots and links. *In vitro* assays, on the other hand, show that these knots and links inhibit basic DNA processing functions such as replication and transcription. This apparent paradox suggests that cells have evolved mechanisms to perform DNA processing functions in topologically complex environments. kDNA is one of the few genomes to be topologically characterized. Although it is known that kDNA's elegant topological structure is essential for the successful completion of the parasite's life cycle, as well as independent replication of minicircles and maxicircles and a coordinated regulation of transcription, important information for potential drug developments (including the origin, function, and maintenance of the network) is still lacking. The studies and results presented in chapter 3 aim, in part, at developing models that can be used for modeling potential drug targets. Specifically, we discussed several key topological quantities that describe minicircle, namely the mean valence and pair-linking probability, and we explored how various bio-physical parameters can affect these characteristics by studying various minicircle models that take

69

into account of these parameters. A key finding among all of these studies is that minicircle density appears to be an important driving force in minicircle network formation, with the minicircle orientation restrictions also an important contributor. On the other hand, factors like the spatial distribution of minicircles, their flexibility, and steric effects play a secondary role. These modeling approaches and results can be refined if more information is available about the specific shapes and orientations of the minicircles in the kinetoplast (even only in the average sense), and can be potential research topics in the future.

5.2. The S477N Mutant in SARS-CoV-2

Our pHMM results show that the S477N substitution emerges in multiple regions at different points in time and maintains a substantial share of the viral population in each such country. In Australia two strains emerge in the summer of 2020, single-mutant S477N and double-mutant S477N/D614G, and both compete favorably there with the two background strains, WT and D614G. Even as aggressive mitigation measures pushed the number of reported samples down dramatically in early August, the double mutant continues to dominate the Australian viral population for several months. In early August the double mutant emerges in several European countries, including England, Denmark, and Switzerland, and it demonstrates a stable or increasing population fraction in those countries through the end of October as it competes with single-mutant D614G. Our results (Fig. 4.1) also confirm that D614G dominates the viral population on long time scales.

we performed MD simulations to compare the binding affinity of S477N against WT and that of S477N/D614G against S477N. Taken together, our two pairs of simulations suggest that S477N introduces an additive binding advantage against either background, which is consistent with the population dynamics that we report. In the case of the latter pair, the double mutant demonstrates a statistically significant binding improvement over D614G. S477N improves it's binding affinity in part due to strengthened hydrogen bonding involving residue N501, which is in a region distal to locus 477. Using several MD trajectories, we

70

report that the hydrogen bond between oxygen atoms in N501 and the Y41 ACE2 residue is more persistent in the mutant that in WT.

Our simulations demonstrate the important effects of mutations in the two "ends" of the S-RBD region—those neighboring locus 477 or neighboring locus 498—which are the closest points between S and ACE2. In particular, the hydrogen bonding results highlight the importance of the interfacial region that includes RBD residues 501. Residue 501 is believed to play a role in the increased transmissibility of the "UK strain" (B.1.1.7) [?], and our simulations predict that the S477N substitution creates new H-bonding interactions between N501 and partners on ACE2. Finally, our MD simulations of five antibodies isolated in sera from COVID-19-recovered individuals show a possible binding effect for S477N, with a majority of antibody-RBD complexes exhibiting weaker binding to the mutant, suggesting diminished neutralizing activity.

Bibliography

[ADH12]   J. Arsuaga, Y. Diao, and K. Hinson, *The effect of angle restriction on the topological charac-teristics of minicircle networks*, J Stat Phys **146** (2012), no. 2, 434–45.

[ADR14]   J. Arsuaga, Y. Diao, and V. Rodriguez, Mol Math Biol **2** (2014), no. 1, 98–106.

[AEKS99]   K. Abu-Elneel, I. Kapeller, and J. Sholomai, J Biol Chem **274** (1999), 13419–26.

[ATS83]   D. Avrahami, Y. Tzfati, and J. Shlomai, PNAS (1983), 9210511Ð5.

[ATV$^+$02]   Arsuaga, R. K. Tan, M. Vazquez, D. W. Sumners, and S. Harvey, *Investigation of viral dna packaging using molecular mechanics models*, Biophys Chem **101** (2002), 475–84.

[BWX$^+$20]   D. J. Benton, A. G. Wrobel, P. Xu, C. Roustan, S. R. M. P. B. Rosenthal, J. J. Skehel, and S. J. Gamblin, *Receptor binding and priming of the spike protein of sars-cov-2 for membrane fusion*, Nature **588** (2020), 327–330.

[CCP$^+$20]   S. A. Clark, L. E. Clark, J. Pan, A. Coscia, L. G. McKay, S. Shankar, R. I. Johnson, A. Griffiths, and J. Abraham, *Molecular basis for a germline-biased neutralizing antibody response to SARS-CoV-2*, (2020).

[CLS$^+$16]   F. Chen, H. Liu, H. Sun, P. Pan, Y. Li, D. Li, and T. Hou, *Assessing the performance of the mm/pbsa and mm/gbsa methods. 6. capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking*, Physical Chemistry Chemical Physics **18** (2016), 22129–22139.

[CPC88]   C. Chiccoli, P. Pasini, and Z. C, *A monte carlo investigation of the planar lebwohl-lasher model*, Physica **148** (1988), no. 3, 298–311.

[CRW$^+$95]   J. Chen, C. A. Rauch, J. H. White, P. T. Englund, and N. R. Cozzarelli, *The topology of the kinetoplast dna network*, Cell **80** (1995), no. 1, 61–69.

[DAH12]   Y. Diao, J. Arsuaga, and K. Hinson, J Phys A **45** (2012), 035004.

[DG69]   E. Delain and R. G, *Dna ultrastructure of the kinetoplast of trypanosoma cruzi cultivated in-vitro*, C R Acad Sci **268** (1969), 1225–1227.

[DHK⁺12]   Y. Diao, K. Hinson, R. Kaplan, M. Vazquez, and J. Arsuaga, J Math Biol **64** (2012), no. 6, 1087–1108.

[DP67]   E. Delain and C. Pauletti, J Mol Biol **28** (1967), no. 2, 377–82.

[DRKA15]   Y. Diao, V. Rodriguez, M. Klingbeil, and J. Arsuaga, *Orientation of dna minicircles balances density and topological complexity in kinetoplast dna*, PLoS One **10** (2015), no. 6, e0130998.

[Dur13]   R. Durbin, *Biological sequence analysis*, Cambridge University Press, 2013.

[DvR98]   Y. Diao and J. van Rensburg, *Topology and geometry in polymer science*, IMA Vols Math, Appl **103** (1998), 79–88.

[Edg04]   R. C. Edgar, *Muscle: multiple sequence alignment with high accuracy and high throughput*, Nucleic Acids Research **32** (2004), 1792–1797.

[Eng78]   P. T. Englund, Cell **74** (1978), 157–168.

[EPB⁺95]   U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *A smooth particle mesh ewald method*, The journal of chemical physics **103** (1995), no. 19, 8577–8593.

[Fis13]   G. Fishman, *Monte carlo: Concepts, algorithms, and applications, springer series in operations research and financial engineering*, Springer, New York, 2013.

[FTWE92]   M. Ferguson, A. F. Torri, D. C. Ward, and P. T. Englund, Cell **70** (1992), 621Ð629.

[GBZH13]   L. Gheiratmand, A. Brasseur, Q. Zhou, and C. Y. He, J Biol Chem **288** (2013), 3489–3499.

[GHZ⁺21]   Y. Guo, L. Huang, G. Zhang, Y. Yao, H. Zhou, S. Shen, B. Shen, B. Li, X. Li, Q. Zhang, M. Chen, D. Chen, J. Wu, D. Fu, X. Zeng, M. Feng, C. Pi, Y. Wang, X. Zhou, M. Lu, Y. Li, Y. Fang, Y.-Y. Lu, X. Hu, S. Wang, W. Zhang, G. Gao, F. Adrian, Q. Wang, F. Yu, Y. Peng, A. G. Gabibov, J. Min, Y. Wang, H. Huang, A. Stepanov, W. Zhang, Y. Cai, J. Liu, Z. Yuan, C. Zhang, Z. Lou, F. Deng, H. Zhang, C. Shan, L. Schweizer, K. Sun, and Z. Rao, *A SARS-CoV-2 neutralizing antibody with extensive spike binding coverage and modified for optimal therapeutic outcomes*, Nature Communications **12** (2021), no. 1.

[Gor08]   H. U. Goringer, Ann Rev Microbiol **66** (2008), 65–82.

[Hal18]   B. I. Halperin, *On the hohenberg–mermin–wagner theorem and its limitations*, J Stat Phys **175** (2018), 521–529.

[HAO⁺06]   V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Comparison of multiple amber force fields and development of improved protein backbone parameters*, Proteins: Structure, Function, and Bioinformatics **65** (2006), no. 3, 712–725.

[Hoh67]   P. C. Hohenberg, *Existence of long-range order in one and two dimensions*, Phys Rev **158** (1967), 383–386.

[HR98]   J. C. Hines and D. S. Ray, Mol Biochem Parasitol **94** (1998), 41Ð52.

[HSW⁺20]   N. K. Hurlburt, E. Seydoux, Y.-H. Wan, V. V. Edara, A. B. Stuart, J. Feng, M. S. Suthar, A. T. McGuire, L. Stamatatos, and M. Pancera, *Structural basis for potent neutralization of SARS-CoV-2 and role of antibody affinity maturation*, Nature Communications **11** (2020), no. 1.

[HW80]   J. H. J. Hoeijmakers and P. J. Weijers, *The segregation of kinetoplast dna networks in trypanosoma brucei*, Plasmid **4** (1980), no. 1, 97–116.

[HWLW01]   T. Hou, J. Wang, Y. Li, and W. Wang, *Assessing the performance of the mm/pbsa and mm/gbsa methods: I. the accuracy of binding free energy calculations based on molecular dynamics simulations*, journal of Chemical Information Modeling **51** (2001), 69–82.

[ILK⁺18]   L. Ibrahim, P. Liu, M. Kliingbeil, Y. Diao, and J. Arsuaga, *Estimating properties of kinetoplast dna by fragmentation reactions*, J Phys A **52** (2018), 034001.

[IME93]   D. IPerez-Morga and P. T. Englund, *The structure of replicating kinetoplast dna networks*, J Cell Biol **123** (1993), no. 5, 1069–1079.

[JJB02]   A. Jakalian, D. B. Jack, and C. I. Bayly, *Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation*, journal of computational chemistry **23** (2002), no. 16, 1623–1641.

[JSE08]   R. E. Jensen, L. Simpson, and P. T. Englund, Trends Parisitol **24** (2008), no. 10, 428–431.

[KDHK12]   E. Krieger, R. L. Dunbrack, R. W. Hooft, and B. Krieger, *Assignment of protonation states in proteins and ligands: Combining pk a prediction with hydrogen bonding network optimization*, Computational Drug Discovery and Design, Springer, 2012, pp. 405–421.

[Kea88]   K. V. Klenin and et al., J Biomol Struct Dyn **6** (1988), 1173Ð–85.

[Kea16]   S. Kaser and et al., PNAS **113** (2016), E4467–E4475.

[KFG⁺20]   B. Korber, W. M. Fischer, S. Gnanakaran1HyejinYoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E.Giorgi, T. Bhattacharya, B. Foley, K. M.Hastie, M. D.Parker, D. G.Partridge, C. M.Evans, T. M.Freeman, and T. I. de Silva, *Tracking changes in sars-cov-2 spike: evidence that d614g increases infectivity of the covid-19 virus.*, Cell **182** (2020), 812–827.

[KL00]   K. V. Klenin and J. Langowski, Biopolymers **54** (2000), no. 5, 307Ð17.

[KT72]   J. M. Kosterlitz and D. J. THouless, *Ordering, metastability, and phase transitions in two-dimensional systems*, J Phys C **6** (1972), 521–529.

[KV14]   E. Krieger and G. Vriend, *Yasara viewÑmolecular graphics for all devicesÑfrom smartphones to workstations*, Bioinformatics **30** (2014), no. 20, 2981–2982.

[KV15]   ———, *New ways to boost molecular dynamics simulations*, journal of computational chemistry **36** (2015), no. 13, 996–1007.

[Lea02]   J. Lukes and et al., Eukaryot Cell **1** (2002), no. 4, 495–502.

[Lea08]   B. Liu and et al., Trends Parisitol **21** (2008), no. 8, 363–369.

[LL72]   P. A. Lebwohl and G. Lasher, *Nematic-liquid-crystal order—a monte carlo calculation*, Phys Rev A **6** (1972), no. 1, 426–429.

[LPDA20]   P. Liu, R. Polischuk, Y. Diao, and J. Arsuaga, *Characterizing the topology of kinoplast dna using random knotting*, Topology and Geometry of Biopolymers: AMS Special Session **746** (2020).

[LSC81]   G. R. Luckhurst, P. Simpson, and Z. C, *Computer simulation studies of anisotropic systems: The effect of external fields*, Chem Phys Let **78** (1981), no. 3, 429–433.

[LWN$^+$] Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao, Q. Zhang, H. L. Huan, L. Nie, H. Qin, M. Wang, Q. Lu, X. Li, Q. Sun, J. Liu, L. Zhang, X. Li, W. Huang, and Y. Wang, *The impact of mutations in sars-cov-2 spike on viral infectivity and antigenicity*.

[MLCE83]   J. C. Marini, S. D. Levene, D. M. Crothers, and P. T. Englund, Cold Spring Harb Symp Quant Biol **47** (1983), no. 1, 279–83.

[MMK$^+$15] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, *ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb*, journal of chemical theory and computation **11** (2015), no. 8, 3696–3713.

[MS59]   W. Maier and A. Saupe, Z Naturforsch **13** (1959), no. a, 564–570.

[MW66]   N. D. Mermin and H. Wagner, *Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic heisenberg models*, Phys Rev Lett **17** (1966), 1133–1136.

[OZD$^+$20] J. Ou, Z. Zhou, R. Dai, S. Zhao, X. Wu, J. Zhang, W. Lan, L. Cui, J. Wu, D. Seto, J. Chodosh, G. Zhang, and Q. Zhang, *Emergence of sars-cov-2 spike rbd mutants that enhance viral infectivity through increased human ace2 receptor binding affinity*, BioRxiv (2020), 991844.

[PB53]   K. R. Porter and J. Blum, *A study in microtomy for electron microscopy*, Anat Rec **117** (1953), no. 4, 685–710.

[PB11]   R. K. Pathria and P. K. Beale, *Statistical mechanics*, Elsevier, Burlington, MA, 2011.

[Pea] R. Polischuk and et al., *The s477n mutation confers a potential binding fitnessadvantage consistent with global sars-cov-2population dynamics.*, in prep.

[PLL+] J. A. Plante, Y. Liu, J. Liu, H. Xia, B. A. Johnson, K. G. Lokugamage, X. Zhang, A. E. Muruato, J. Zou, C. R. Fontes-Garfias, D. Mirchandani, D. Scharton, J. P. Bilello, Z. Ku, Z. An, B. Kalveram, A. N. Freiberg, V. D. Menachery, X. Xie, K. S. Plante, S. C. Weaver, and P.-Y. Shi, *Spike mutation d614g alters sars-cov-2 fitness.*

[RCV93] V. V. Rybenkov, N. R. Cozzarelli, and A. V. Vologodskii, PNAS **90** (1993), 5307–11.

[RDA13] V. Rodriguez, Y. Diao, and J. Arsuaga, J Phys: Conf Ser **454** (2013), 01207.

[Rea93] C. A. Rauch and et al., EMBO J **12** (1993), 403Ð411.

[RN58] R. Ricca and B. Nipoti, J Knot Theor Rami **20** (1958), 1325–43.

[RN11] _____, J Knot Theor RAMI **20** (2011), 1325–43.

[Sch] R. Scharein, *Knotplot.*

[SE95] T. A. Shapiro and P. T. Englund, Ann Rev Microbiol **49** (1995), 117–143.

[Sea14] F. Schnarwiler and et al., *Trypanosomal tac40 constitutes a novel subclass of mitochondrial -barrel proteins specialized in mitochondrial genome inheritance*, PNAS **111** (2014), 7624–7629.

[SGCR20] J. Singer, R. Gifford, M. Cotton, and D. Robertson, *Cov-glue: A web application for tracking sars-cov-2 genomic variation*, Preprints (2020), 2020060225.

[SH18] F. Sievers and D. G. Higgins, *Clustal omega for making accurate alignments of many protein sequences*, Protein Science **27** (2018), 135–145.

[SHM58] G. Steinert, F. H, and S. M, *Synthesis of deoxyribonucleic acid in the parabasal body of trypanosoma mega.*, Exp Cell Res **15** (1958), 632–635.

[SLT+14] H. Sun, Y. Li, S. Tian, L. Xu, and T. Hou, *Assessing the performance of mm/pbsa and mm/gbsa methods. 4. accuracies of mm/pbsa and mm/gbsa methodologies evaluated by various simulation protocols using pdbbind data set*, Physical Chemistry Chemical Physics **16** (2014), 16719–16729.

[SM17] Y. Shu and J. McCauley, *Gisaid: Global initiative on sharing all influenza data Ð from vision to reality.*, EuroSurveillance **22** (2017).

[SSK+] A. Singh, G. Steinkellner, K. Koch, K. Gruber, and C. C. Gruber, *Serine 477 plays a crucial role in the interaction of the sars-cov-2 spike protein with the human receptor ace2.*

[Tea16] Trikin and et al., PLoS Pathog **12** (2016), no. 5, e1005586.

[THG94]   J. D. Thompson, D. G. Higgins, and T. J. Gibson, *Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*, Nucleic Acids Research **22** (1994), 4673–4680.

[VHAD09]   R. Varela, K. Hinson, J. Arsuaga, and Y. Diao, J Phys A **42** (2009), no. 9, 095204.

[Vol06]   A. Vologodskii, *Computational studies of rna and dna*, Springer, 2006.

[WHF+77]   P. Weislogel, J. Hoeijmakers, A. Fairlamb, C. Kleisen, and P. Borst, Biochimica et Biophysica Acta **478** (1977), 169–179.

[WWC+]   D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan, *Cryo-em structure of the 2019-ncov spike in the prefusion conformation.*

[WWC+04]   J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *Development and testing of a general amber force field*, journal of computational chemistry **25** (2004), no. 9, 1157–1174.

[WWW+19]   G. Q. Weng, E. C. Wang, Z. Wang, H. Liu, D. Li, F. Zhu, and T. J. Hou, *Hawkdock: a web server to predict and analyze the structures of protein-protein complexes based on computational docking and mm/gbsa*, Nucleic Acids Research **47** (2019), W322–W330.

[WZY+20]   F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, M. L. Yuan, Y. L. Zhang, F. H. Dai, Y. Liu, Q. M. Wang, J. J. Z. L. Xu, E. C. Holmes, and Y. Z. Zhang, *A new coronavirus associated with human respiratory disease in china*, Nature **579** (2020), 265–269.

[XHE+02]   C. W. Xu, J. Hines, M. Engel, D. Russel, and R. D, Molec Cell Biol **16** (2002), no. 2, 564–576.

[YWP+20]   L. Yurkovetskiy, X. Wang, K. E. Pascal, C. Tomkins-Tinch, T. P. Nyalile, Y. Wang, A. Baum, W. E. Diehl, A. Dauphin, C. Carbone, K. Veinotte, S. B. Egri, S. F. Schaffner, J. E. Lemieux, J. B. Munro, A. Rafique, A. Barve, P. C. Sabeti, C. A. Kyratsous, N. V. Dudkina, K. Shen, and J. Luban, *Structural and functional analysisof the d614g sars-cov-2 spike protein variat*, Cell **183** (2020), 739Đ751.

[ZJM+]   L. Zhang, C. B. Jackson, H. Mou, A. Ojha, H. Peng, B. D. Quinlan, E. S. Rangarajan, A. Pan, A. Vanderheiden, M. S. Suthar, W. Li, T. Izard, C. Rader, M. Farzan, and H. Choe, *SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity, 10.1038/s41467-020-19808-4, journal = Nature Communications, volume = 11, year = 2020, pages = 6013.*

[ZLRC+08]   Z. Zhao, M. Lindsay, A. Roy Chowdhury, D. Robinson, and P. Englund, EMBO J **27** (2008), 143–154.