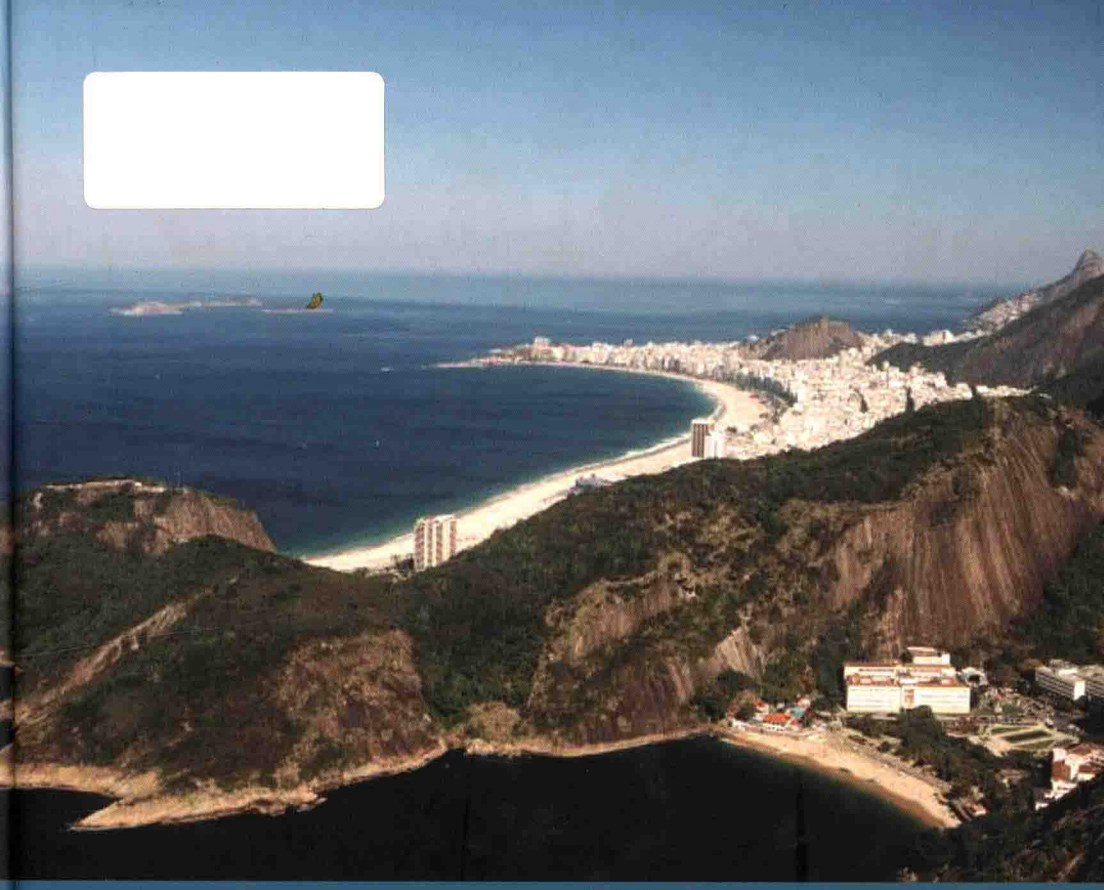

BIOMAT 2010

International Symposium on
Mathematical and Computational Biology

edited by Rubem P Mondaini

BIOMAT 2010

International Symposium on
Mathematical and Computational Biology

Rio de Janeiro, Brazil
24 – 29 July 2010

edited by

Rubem P Mondaini
Federal University of Rio de Janeiro, Brazil

World Scientific

NEW JERSEY · LONDON · SINGAPORE · BEIJING · SHANGHAI · HONG KONG · TAIPEI · CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

BIOMAT 2010
International Symposium on Mathematical and Computational Biology

ISBN-13 978-981-4343-42-8
ISBN-10 981-4343-42-0

Printed in Singapore by World Scientific Printers.

BIOMAT 2010

International Symposium on
Mathematical and Computational Biology

Preface

The present book contains the papers selected for publication by the Referee Board of the BIOMAT Consortium (http://www.biomat.org). The papers have been selected among all submitted full papers of the BIOMAT 2010 International Symposium on Mathematical and Computational Biology. As a rule, the Editorial Board kept the level of acceptance on 20% of the submitted papers in order to preserve the rigour of analysis and critical evaluation of new scientific results and the soundness of reviews. The Editorial Board has also decided to proceed with a double-blind style for the work of the Referee Board on the stages of abstract and full paper. Due to the increasing number of submitted contributions every year, a third evaluation has been necessary for a final decision. This is done by invited Keynote Speakers during the presentation of the paper in the conference.

The BIOMAT Consortium has been working since 2001 on its fundamental mission of enhancing the number of practitioners on the Mathematical Modelling of Biosystems in Latin America and Developing Countries of five continents. The Consortium has been also working during these ten years of realization of international conferences for motivating the future scientific career of young graduate studies on the research areas of Mathematical and Computational Biology and Biological Physics.

The scientific seriousness of the Consortium and its leadership in the representation of Latin American countries in these research areas, was recognized by scientific societies and institutions worldwide. The BIOMAT Consortium has been invited to organize the 2010 Annual Meeting of the Society for Mathematical Biology (http://www.smb.org). This was the first meeting of the SMB Society in South America and it was held in Rio de Janeiro from 24^{th} to 29^{th} July 2010 in a joint conference with the BIOMAT 2010 International Symposium. Sixteen senior scientists have been invited by the BIOMAT Consortium, as Keynote and Plenary Speakers of the joint conference. Parallel sessions of the SMB 2010 Meeting have been mixed with sessions of contributed papers of the BIOMAT 2010 and SMB Mini-Symposium sessions. This joint conference has ended a cycle of organization of international conferences which was created on April 2001 in Rio de Janeiro, with the First BIOMAT Symposium. A new cycle is now starting in the year 2011, with the decision of organizing a number of BIOMAT symposia in other countries. Latin American countries are the

candidates for the first symposia of this new cycle.

We acknowledge the Board of Trustees of three Brazilian sponsoring agencies - Coordination for the Improvement of Higher Education Personell - CAPES, The National Research Council for Scientific and Technological Development - CNPq, Foundation for Research Support of Rio de Janeiro State - FAPERJ. On behalf of the BIOMAT Consortium/ Institute for Advanced Studies of Biosystems, we thank their Directors and Authorized Representatives. Special thanks are due to Prof. José Oswaldo de Siqueira from CNPq for his expertise as a scientific administrator and his senior understanding of the scientific importance of the BIOMAT series of International Symposia.

Thanks are also due to the Commander-in-chief of the School of Naval War- Brazilian Navy, in Rio de Janeiro, to the Rector and the Dean of the Federal University of Rio de Janeiro State - UNIRIO, Prof. Malvina Tuttman and Prof. Luiz Amancio de Sousa Jr., respectively, and the Rector of the Federal University of Rio de Janeiro - UFRJ, Prof. Aloisio Teixeira. Thank you very much indeed for the available facilities on these institutions as well as for some financial help which was provided by UFRJ.

Last but not least, we thank our collaborators which have integrated the BIOMAT Consortium Administrative Staff during the Joint Conference SMB 2010 - BIOMAT 2010, Leonardo Mondaini, Felipe Mondaini and Sandro Pereira Vilela. The latter has also provided some help with the LaTeX version of the accepted papers. Other members of the Staff like Wanderson da Rocha, José Ricardo da Rocha, Janice Justo, Renata Figueiredo and Aline Coutinho should be also remembered for their good organizational skills during the conference.

Rubem P. Mondaini
President of the BIOMAT Consortium
Chairman of the SMB 2010 - BIOMAT 2010 Joint Conference

Rio de Janeiro, December 2010

Editorial Board of the BIOMAT Consortium

Mariano Ricard	Havana University, Cuba
Michael Meyer-Hermann	Frankfurt Inst. for Adv. Studies, Germany
Nicholas Britton	University of Bath, UK
Panos Pardalos	University of Florida, Gainesville, USA
Peter Stadler	University of Leipzig, Germany
Philip Maini	University of Oxford, UK
Pierre Baldi	University of California, Irvine, USA
Ramit Mehr	Bar-Ilan University, Ramat-Gan, Israel
Raymond Mejía	National Institutes of Health, USA
Reidun Twarock	University of York, UK
Richard Kerner	Université Pierre et Marie Curie, Paris, France
Robijn Bruinsma	University of California, Los Angeles, USA
Rui Dilão	Instituto Superior Técnico, Lisbon, Portugal
Ruy Ribeiro	Los Alamos National Laboratory, New Mexico, USA
Timoteo Carletti	Facultés Universitaires Notre Dame de la Paix, Belgium
Vitaly Volpert	Université de Lyon 1, France
William Taylor	National Institute for Medical Research, UK
Zhijun Wu	Iowa State University, USA

Contents

Population Dynamics

Population Biology

Theoretical Immunology

Computational Biology

USING DNA KNOTS TO ASSAY VIRAL GENOME PACKING

DE WITT SUMNERS

Department of Mathematics, Florida State University, Tallahassee, FL 32306 USA

E-mail: sumners@math.fsu.edu

Bacteriophages pack their double-stranded DNA genomes to near-crystalline density in viral capsids and achieve one of the highest levels of DNA condensation found in nature. Despite numerous studies some essential properties of the packaging geometry of the DNA inside the phage capsid are still unknown. Although viral DNA is linear double-stranded with sticky ends, the linear viral DNA quickly becomes cyclic when removed from the capsid, and for some viral DNA the observed knot probability is an astounding 95%. These observed DNA knots carry information about capsid packing geometry. This talk will discuss comparison of the observed viral knot spectrum with the simulated knot spectrum, concluding that the viral knot spectrum is non-random, writhe directed, and generated by local cholesteric interaction (juxtaposition at a small twist angle) between DNA strands

1. Introduction

DNA knots and links (catenanes) are of biological interest because they can detect and preserve topological information, notably information about the binding and mechanism of enzymes that act on DNA, and information about the geometry of viral genome packing in bacteriophage capsids. In the topological approach to enzymology, DNA substrate plasmids of known supercoiling density and knot/link type are incubated with purified enzyme *in vitro*, and changes from substrate to product in DNA topology/geometry are observed and quantified by gel electrophoresis and electron microscopy. Knotted and linked DNA molecules are very robust, and can survive experimental manipulation and visualization intact. Characterization of knotted products formed by random cyclization of linear molecules has been used to quantify important biochemical properties of DNA such as effective diameter in solution as a function of salt concentration [1,2]. DNA knots and catenanes obtained as products of site-specific recombination on negatively

supercoiled substrate have been a key to understanding enzymatic binding and mechanism [3,4]. In all cases, the development of mathematical and computational tools has greatly enhanced analysis of the experimental results [5,6].

Significant numbers of endogenous DNA knots are found also in biological systems: in *Escherichia coli* cells harboring mutations in the GyrB or GyrA genes [7], bacteriophages P2 and P4 [8,9], and cauliflower mosaic viruses [10]. However, very little biological information about these systems has been inferred from the observed knots. In particular, interpretation of the experimental results for bacteriophages has been limited by the experimental difficulty in quantifying the complex spectrum of knotted products. These difficulties have paralleled those encountered in developing a theory for random knotting of ideal polymeric chains in cases where interactions with other macromolecules and/or confinement in small volumes have a significant function [11-16].

2. Bacteriophage Genome Packing

Bacteriophages are viruses that infect bacteria. They pack their double-stranded DNA genomes to near-crystalline density in viral capsids and achieve one of the highest levels of DNA condensation found in nature. When I was on sabbatical in Berkeley in 1989, Jim Wang described to me the problem of DNA packing in icosahedral viral capsids [8,9], and the high degree of knotting produced when the viral DNA is released from the capsids. Despite numerous studies some essential properties of the packaging geometry of the DNA inside the phage capsid are still unknown. Although viral DNA is linear double-stranded with sticky (cohesive) ends, the linear viral DNA quickly becomes cyclic when removed from the capsid, and for some viral DNA the observed knot probability is an astounding 95%. In the summer of 1998, my PhD students Javier Arsuaga and Mariel Vazquez spent 2 months in the laboratory of Joaquim Roca in Barcelona, supported by the Burroughs Wellcome Interfaces grant to the Program in Mathematics and Molecular Biology. In the Roca laboratory, they infected bacterial stock, harvested viral capsids and extracted and analyzed the viral DNA. They quantified the DNA knot spectrum produced in the experiment, and used Monte Carlo generation of knots in confined volumes to produce a random knot spectrum to be compared to the observed viral DNA knot spectrum. A subsequent experiment was performed in the Roca laboratory on a shorter viral genome [22]. A series of papers were produced as a re-

sult of these experiments [17-21, 23, 24], and I will describe some of the results from these papers, focusing on (and reproducing here) most of the discussion and analysis from [20], and a short discussion of the results from [24].

All icosahedral bacteriophages with double-stranded DNA genomes are believed to pack their chromosomes in a similar manner [25]. During phage morphogenesis, a procapsid is first assembled, and a linear DNA molecule is actively introduced inside it by the connector complex [26,27]. At the end of this process, the DNA and its associated water molecules fill the entire capsid volume, where DNA reaches concentrations of 800 mg/ml [28]. Some animal viruses [29] and lipo-DNA complexes used in gene therapy [30] are postulated to hold similar DNA arrangements as those found in bacteriophages. Although numerous studies have investigated the DNA packing geometry inside phage capsids, some of its properties remain unknown. Biochemical and structural analyses have revealed that DNA is kept in its B form [31-33] and that there are no specific DNA-protein interactions [34,35] or correlation between DNA sequences and their spatial location inside the capsid, with the exception of the cos ends in some viruses. Many studies have found that regions of the packed DNA form domains of parallel fibers, which in some cases have different orientations, suggesting a certain degree of randomness [32,33]. The above observations have led to the proposal of several long-range organization models for DNA inside phage capsids: the ball of string model [36], the coaxial spooling model [32,36,37], the spiral-fold model [38], and the folded toroidal model [39]. Liquid crystalline models, which take into account properties of DNA at high concentrations and imply less global organization, have also been proposed [33]. Cryo EM and spatial symmetry averaging has recently been used to investigate the surface layers of DNA packing [40]. In [20], the viral DNA knot spectrum was used to investigate the packing geometry of DNA inside phage capsids.

3. Bacteriophage P4

Bacteriophage P4 has a linear, double-stranded DNA genome that is 10-11.5 kb in length and flanked by 16-bp cohesive cos ends [41]. It has long been known that extraction of DNA from P4 phage heads results in a large proportion of highly knotted, nicked DNA circles [8,9]. DNA knotting probability is enhanced in P4 derivatives containing genome deletions [42] and in tailless mutants [43]. Most DNA molecules extracted from P4 phages are circles that result from the cohesive-end joining of the viral genome.

Previous studies have shown that such circles have a knotting probability of about 20% when DNA is extracted from mature P4 phages [8]. This high value is increased more than 4-fold when DNA is extracted from incomplete P4 phage particles (which we refer to as "capsids") or from non-infective P4 mutants that lack the phage tail (which we refer to as "tailless mutants" [8]). Knotting of DNA in P4 deletion mutants is even greater. The larger the P4 genome deletion, the higher the knotting probability [42]. For P4 vir1 del22, containing P4's largest known deletion (1.6 kb deleted [44]), knotting probability is more than 80% [43]. These values contrast with the knotting probability of 3% (all trefoil knots) observed when identical P4 DNA molecules undergo cyclization in dilute free solution [1,45]. These differences are still more striking when the variance in distribution of knot complexity is included. Although knots formed by random cyclization of 10-kb linear DNA in free solution have an average crossing number of three [1,45], knots from the tailless mutants have a knotting probability of 95% and appear to have very large crossing numbers, averaging about 26. [8,42,43].

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

In [20] it is shown that the DNA knots provide information about the global arrangement of the viral DNA inside the capsid. The distribution of the viral DNA knots is analyzed by high-resolution gel electrophoresis. Monte Carlo computer simulations of random knotting for freely jointed polygons confined to spherical volumes is performed. The knot distribution produced by simulation is compared to the observed experimental DNA knot spectrum. The simulations indicate that the experimentally observed scarcity of the achiral knot 4_1 and the predominance of the torus knot 5_1 over the twist knot 5_2 are not caused by confinement alone but must include writhe bias in the packing geometry. Our results indicate that the packaging geometry of the DNA inside the viral capsid is non-random and

writhe-directed.

4. Knot Type Probabilities for Phage P4 DNA in Free Solution

The probability that a DNA knot K of n statistical lengths and diameter d is formed by random closure in free solution is given by $P_K\,(n,\ d) = P_K\,(n,\ 0)\,e^{-rd/n}$, where r depends on the knot type and equals 22 for the trefoil knot K and 31 for the figure 8 knot K^*[33]. The knotting probability of a 10-kb DNA molecule cyclized in free solution is 0.03 (25, 26), which implies an effective DNA diameter near 35$\mathring{A}$. Because $P_K(34,\ 0) = 0.06$ and $P_{K^*}\,(34,\ 0) = 0.009$, then $P_K\,(34,\ 35) = 0.027(1/36$ times that of the unknot) and $P_{K^*}\,(34,\ 35) = 0.003(1/323$ times that of the unknot). These values were used to estimate the fractions of the knot K and the knot K^* generated for P4 DNA in free solution.

5. Monte Carlo Simulation

Knotting probabilities of equilateral polygons confined into spherical volumes were calculated by means of Markov-chain Monte Carlo simulations followed by rejection criteria. Freely jointed closed chains, composed of n equilateral segments, were confined inside spheres of fixed radius, r, and sampled: values of n ranged from 14 to 200 segments; r values, measured as multiples of the polygonal edge length, ranged from 2 to infinity. Excluded volume effects were not taken into account. Markov chains were generated by using the Metropolis algorithm [46]. The temperature, a computational parameter, was held at T = 300 K to improve the efficiency of the sampling algorithm. Other values of T produced similar results, thus indicating that the computation is robust with respect to this parameter. Chains contained inside the sphere were assigned zero energy. Chains lying partly or totally outside the confining sphere were assigned an energy given by the maximum of the distances of the vertices of the chain to the origin. Only chains with zero energy were sampled. A random ensemble of polygons was generated by the crankshaft algorithm as follows: (i) two vertices of the chain were selected at random, dividing the polygon into two subchains, and (ii) one of the two subchains was selected at random (with equal probabilities for each subchain), and the selected subchain rotated through a random angle around the axis connecting the two vertices. This algorithm is known to generate an ergodic Markov chain in the set of polygons of fixed length [47]. Correlation along the subchains was computed by using time-series

analysis methods as described by Madras and Slade [48]. Identification of the knotted polygons was achieved by computing the Alexander polynomial $\Delta(t)$ [49-51] evaluated at t = -1. It is known that $\Delta(-1)$ does not identify all knotted chains; however, for polygonal chains not confined to a spherical volume, nontrivial knots with trivial $\Delta(-1)$ values rarely occur. This circumstance has been observed by using knot invariants, such as the HOMFLY polynomial, that distinguish between knotted and unknotted chains with higher accuracy than the Alexander polynomial. Computer simulations for small polygons (< 55 segments) show that the knotting probabilities obtained by using $\Delta(-1)$ agree with those obtained by using the HOMFLY polynomial. Furthermore, Deguchi and Tsurusaki have reported that the value of $\Delta(-1)$ can almost always determine whether a given Gaussian polygon is unknotted for lengths ranging from 30 to 2,400 segments [51]. Each selected knotted polygon was further identified by evaluating its Alexander polynomial at t = -2 and t = -3. Although the Alexander polynomial is an excellent discriminator among knots of low crossing number and its computation is fast, it does not distinguish completely among some knotted chains [for example, composite knots $3_1\#3_1$ and $3_1\#4_1$ have polynomials identical to those of prime knots 8_{20} and 8_{21}, respectively [RW]. Evaluation of the polynomial at t = -2 and t = -3 is also ambiguous because the Alexander polynomial is defined up to units in $Z[t^{-1}, t]$, and therefore the algorithm returns $(\pm n^{\pm m}\Delta(-n))$, where $n = 2$ or 3 and m is an integer. To deal with this uncertainty, we followed van Rensburg and Whittington [RW] and chose the largest exponent k such that the product $(\pm n^{\pm m}\Delta(-n))n^{\pm k}$ is an odd integer with $n = 2$ or 3. This value was taken as the knot invariant. To compute the writhe, we generated > 300 regular projections and resulting knot diagrams for each selected polygon. To each of the projected crossings a sign was assigned by the standard oriented skew lines convention. The directional writhe for each diagram was computed by summing these values. The writhe was then determined by averaging the directional writhe over a large number of randomly chosen projections. To generate writhe-directed random distributions of polygons, we used a rejection method in which polygons whose writhe was below a positive value were not sampled.