THE IMA VOLUMES IN MATHEMATICS
AND ITS APPLICATIONS

EDITORS Craig John Benham
Stephen Harvey
Wilma K. Olson
De Witt L. Sumners
David Swigon

Mathematics of DNA Structure, Function and Interactions

Springer

The IMA Volumes in Mathematics and its Applications

Volume 150

Series Editors
Fadil Santosa Markus Keel

For other titles published in this series, go to
www.springer.com/series/811

Institute for Mathematics and its Applications (IMA)

The **Institute for Mathematics and its Applications** was established by a grant from the National Science Foundation to the University of Minnesota in 1982. The primary mission of the IMA is to foster research of a truly interdisciplinary nature, establishing links between mathematics of the highest caliber and important scientific and technological problems from other disciplines and industries. To this end, the IMA organizes a wide variety of programs, ranging from short intense workshops in areas of exceptional interest and opportunity to extensive thematic programs lasting a year. IMA Volumes are used to communicate results of these programs that we believe are of particular value to the broader scientific community.

The full list of IMA books can be found at the Web site of the Institute for Mathematics and its Applications:

http://www.ima.umn.edu/springer/volumes.html.

Presentation materials from the IMA talks are available at

http://www.ima.umn.edu/talks/.

Video library is at http://www.ima.umn.edu/videos/.

Fadil Santosa, Director of the IMA

* * * * * * * * * *

IMA ANNUAL PROGRAMS

1982–1983	Statistical and Continuum Approaches to Phase Transition
1983–1984	Mathematical Models for the Economics of Decentralized Resource Allocation
1984–1985	Continuum Physics and Partial Differential Equations
1985–1986	Stochastic Differential Equations and Their Applications
1986–1987	Scientific Computation
1987–1988	Applied Combinatorics
1988–1989	Nonlinear Waves
1989–1990	Dynamical Systems and Their Applications
1990–1991	Phase Transitions and Free Boundaries
1991–1992	Applied Linear Algebra
1992–1993	Control Theory and its Applications
1993–1994	Emerging Applications of Probability
1994–1995	Waves and Scattering
1995–1996	Mathematical Methods in Material Science
1996–1997	Mathematics of High Performance Computing

Continued at the back

Craig John Benham · Stephen Harvey · Wilma K.
Olson · De Witt L. Sumners · David Swigon
Editors

Mathematics of DNA Structure, Function and Interactions

Springer

Editors

Craig John Benham
Department of Mathematics
University of California
451 E. Health Sciences Drive
Davis CA 95616
GBSF
USA
http://genomics.ucdavis.edu/benham/benham.php

Stephen Harvey
Department of Chemistry and Biochemistry
Georgia Institute of Technology
Atlanta GA 30332-0400
USA
http://www.chemistry.gatech.edu/faculty/Harvey/

Wilma K. Olson
Department of Chemistry
and Chemical Biology
Rutgers University
610 Taylor Rd.
Piscataway NJ 08854-8087
USA
http://rutchem.rutgers.edu/content_dynamic/faculty/wilma_k_olson.shtml

De Witt L. Sumners
Department of Mathematics
Florida State University
Tallahassee FL 32306-4510
208 Love Bldg.
USA
http://www.math.fsu.edu/∼sumners/

David Swigon
Department of Mathematics
University of Pittsburgh
301 Thackeray Hall
Pittsburgh PA 15260
USA
http://www.math.pitt.edu/∼swigon/

ISSN 0940-6573
ISBN 978-1-4419-0669-4 e-ISBN 978-1-4419-0670-0
DOI 10.1007/978-1-4419-0670-0
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009931719

Mathematics Subject Classification (2000): 92-06, 92B99, 92C40, 92C05

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

FOREWORD

This IMA Volume in Mathematics and its Applications

MATHEMATICS OF DNA STRUCTURE, FUNCTION, AND INTERACTIONS

contains papers presented at a highly successful one-week workshop held on September 16-21, 2007 on the same title. The event was an integral part of the 2007–2008 IMA Thematic Year on “Mathematics of Molecular and Cellular Biology.” We are grateful to all the participants for making this workshop a very productive and stimulating event.

We owe special thanks to Craig John Benham (Davis Genome Center, University of California, Davis), Stephen Harvey (Department of Chemistry and Biochemistry, Georgia Institute of Technology), Wilma K. Olson (Department of Chemistry and Chemical Biology, Rutgers University), De Witt L. Sumners (Department of Mathematics, Florida State University), and David Swigon (Department of Mathematics University of Pittsburgh) for their superb role as workshop organizers and editors of these proceedings.

We take this opportunity to thank the National Science Foundation for its support of the IMA.

Series Editors

Fadil Santosa, Director of the IMA

Markus Keel, Deputy Director of the IMA

PREFACE

Propelled by the success of the sequencing of the human and many related genomes, molecular and cellular biology has delivered significant scientific breakthroughs. Mathematics (broadly defined) continues to play a major role in this effort, helping to discover the secrets of life by working collaboratively with bench biologists, chemists and physicists. The critical need, which has already begun, is the development of a quantitative body of theory for biology. This development of theory is expected to have the same impact on biology as it did on the sciences of physics, chemistry and engineering in the 20th century. People with strong backgrounds in both biology and the mathematical sciences are creating this quantitative body of theory. Because of its outstanding record of interdisciplinary research and training, the IMA was an ideal venue for the 2007-2008 IMA thematic year on Mathematics of Molecular and Cellular Biology. This volume is dedicated to the memory of Nicholas Cozzarelli, a dynamic leader who fostered research and training at the interface between mathematics and molecular biology. Nick was the founding director of the Program in Mathematics and Molecular Biology (PMMB), a national research and training consortium in existence from 1987-2007. Two of the editors of this volume (Olson and Sumners) were members of PMMB, and one of the editors (Swigon) was a PMMB Fellow. Seven of the thirty-one authors of papers in this volume were PMMB Fellows, an indication of the influence of Nick Cozzarelli on research at the mathematics/molecular biology interface. The kickoff event for the IMA thematic year was the IMA tutorial on Mathematics of Nucleic Acids, and the following 6-day IMA workshop Mathematics of DNA Structure, Function and Interactions, held during September 15-21, 2007 in Minneapolis. The workshop consisted of 32 talks and 17 posters, and enjoyed participation by 120 interdisciplinary scientists, a mix of mathematicians, biologists, chemists, physicists and engineers. This volume consists of a remembrance of Nick Cozzarelli by two past members of his Berkeley molecular biology laboratory, and 15 papers contributed by speakers at the tutorial and workshop. It contains of some of the state-of-the-art in mathematical approaches to DNA as of September 2007. A short description of the articles in the volume follows. For a more complete idea of the content of each article, please see the introductions to each article.

1. Nick Cozzarelli: A personal remembrance by Stephen D. Levene and Lynn Zechiedrich. Steve and Lynn were postdocs in the Cozzarelli lab during the period 1989-1997. This remembrance is very perceptive in the description of Cozzarelli as a blast-ahead interdisciplinary scientist, and recounts a hilarious incident at the lab in which Nick accepts an unexpected NIH merit award over the phone.

2. Mathematical methods in DNA topology: Applications to chromosome organization and site-specific recombination, by Javier Arsuaga, Yuanan Diao, and Mariel Vazquez. This paper explores some of the uses of knot theory and 3-dimensional manifold topology to model chromosome organization and the binding and mechanism of site-specific DNA recombination enzymes. The paper reviews both theoretical and computational topological methods.

3. Conformational statistics of DNA and diffusion equations on the Euclidean group by Gregory S. Chirikjian. Using wormlike chain models for DNA, this paper studies the problem of determining the probability density of end-to-end chain position and orientation. Solutions are obtained by solving the Fokker-Planck equation that describes a diffusion process on the Euclidean motion group.

4. Perspectives on DNA looping, by Laura Finzi. This paper presents a survey of the field of DNA looping, with emphasis on three repressor systems lac, gal and phage lambda. The paper concentrates on the insight gained on transcriptionally-relevant DNA looping mechanisms by single-molecule approaches.

5. Differences between positively and negatively supercoiled DNA that topoisomerases may distinguish, by Jonathan M. Fogg, Daniel J. Catanese, Jr. Graham Randall, Michelle C. Swick, and Lynn Zechiedrich. This article presents a new biological perspective on DNA supercoiling, including a review of the functional importance and practical issues encountered in laboratory work. It provides hints of the features of DNA structure and energetics that topoisomerases may utilize in controlling the supercoiled state of DNA.

6. Calibration of tethered particle motion experiments, by Lin Han, Bertrand Lui, Seth Blumberg, John F. Beausang, Philip C. Nelson, and Rob Phillips. The Tethered Particle Motion (TPM) method has been used to observe and characterize a variety of protein-DNA interactions including DNA looping and transcription. This paper describes a detailed calibration of TPM magnitude as a function of DNA length and particle size, exploring how experimental parameters such as acquisition time and exposure time affect the apparent motion of the tethered particle

7. Difference topology: Analysis of high-order DNA-protein assemblies, by Makkuni Jayaram and Rasika Harshey. This paper studies Difference topology, a method for deciphering the DNA topology within DNA-protein complexes that are not readily amenable to standard structural analyses. The logic is to trap the crossings formed by distinct DNA segments by tying them into knots or links by site-specific DNA inversion and deletion, respectively, carried out by a recombinase. The number of such crossings can then be counted by analytical methods such as gel electrophoresis or electron microscopy.

8. Useful intrusions of DNA topology into experiments on protein-DNA geometry, by Jason D. Kahn, James R. Jenssen, and Vasavi Vittal.

This paper studies the use of small DNA minicircles to characterize protein-induced DNA bending and twisting. In every case studied, topological characterization of minicircle synthesis or properties has led to unexpected geometric or mechanistic conclusions.

9. Topological analysis of DNA-protein complexes, by Soojeong Kim and Isabel K. Darcy. Tangles have been used to model protein-bound DNA. The protein is represented by a 3D ball and the protein-bound DNA is represented by the strings embedded in the 3D ball. This paper reviews tangle analysis of protein-DNA complexes involving three or four segments of DNA.

10. Closing the loop on protein-DNA interactions: Interplay between shape and flexibility in nucleoprotein assemblies having implications for biological regulation, by Stephen D. Levene and Yongli Zhang. The formation of DNA loops by proteins bound at distant sites along a single molecule is an essential mechanistic aspect of many biological processes including gene regulation, DNA replication, and recombination. This paper describes a rigorous theory for DNA loop formation that connects the global mechanical and geometric properties of both DNA and protein, with applications to the problem of loop-mediated gene repression in vivo by lac repressor.

11. Four-way helical junctions in DNA molecules, by David M.J. Lilley. Four-way (Holliday) junctions are branch points in DNA where four helices are interconnected by the mutual exchange of strands. This paper presents a short review focusing on recent developments in understanding the structure and dynamics of DNA four-way junctions.

12. Micromechanics of single supercoiled DNA molecules, by John F. Marko. This paper reviews the theory of the mechanical response of single DNA molecules under stretching and twisting stresses. Using established results for the semiflexible polymer including the effect of torsional stress, and for the free energy of plectonemic supercoils, a theory of coexisting plectonemic and extended DNA is constructed and shown to produce phenomena observed experimentally.

13. Flexibility of nucleosomes on topologically constrained DNA, by Andrei Sivolob, Christophe Lavelle and Ariel Prunell. This paper reviews results on nucleosome conformational flexibility, its molecular mechanism and its functional relevance. The initial approach combined both empirical measurement and theoretical simulation of the topological properties of single particles reconstituted on DNA minicircles.

14. The mathematics of DNA structure, mechanics, and dynamics, by David Swigon. A brief review is given of the main concepts, ideas, and results in the fields of DNA topology, elasticity, mechanics and statistical mechanics. Discussion includes the notions of the linking number, writhe, and twist of closed DNA, elastic rod models, sequence-dependent base-pair level models, statistical models such as helical worm-like chain and freely jointed chain, and dynamical simulation procedures.

15. Paradox regained: A topological coupling of nucleosomal DNA wrapping and chromatin fibre coiling, by Andrew Travers. The folding and unfolding of the chromatin fibre is a fundamental control point for the regulation of eukaryotic transcription. This paper presents a novel solution to the so-called linking number paradox problem and shows that this solution implies that the chromatin fibre acts a tunable coil.

16. Statistical-mechanical analysis of enzymatic topological transformations in DNA molecule, by Alexander Vologodskii. This paper reviews computational approaches to the analysis of action of two classes of DNA enzymes: topoisomerase and recombinase. Comparing the simulated distribution with corresponding experimental data serves as a model test. The major principles and assumptions of the approach, which is based on the simulation of an equilibrium set of DNA conformations, are discussed.

On behalf of the editors, I would like to thank the authors of papers for contributing to this volume, and for their cooperation in the editorial process. Special thanks go to Patricia V. Brick and Dzung N. Nguyen for their excellent assistance in preparing papers for the volume publisher.

Craig John Benham
Department of Mathematics
University of California, Davis
http://genomics.ucdavis.edu/benham/benham.php

Stephen Harvey
Department of Chemistry and Biochemistry
Georgia Institute of Technology
http://www.chemistry.gatech.edu/faculty/Harvey/

Wilma K. Olson
Department of Chemistry and Chemical Biology
Rutgers University
http://rutchem.rutgers.edu/content_dynamic/faculty/wilma_k_olson.shtml

De Witt L. Sumners
Department of Mathematics
Florida State University
http://www.math.fsu.edu/~sumners/

David Swigon
Department of Mathematics
University of Pittsburgh
http://www.math.pitt.edu/~swigon/

CONTENTS

Foreword v

Preface vii

Nick Cozzarelli: A personal remembrance 1
Stephen D. Levene and Lynn Zechiedrich

Mathematical methods in DNA topology: Applications to chromosome organization and site-specific recombination 7
Javier Arsuaga, Yuanan Diao, and Mariel Vazquez

Conformational statistics of DNA and diffusion equations on the Euclidean group 37
Gregory S. Chirikjian

Perspectives on DNA looping 53
Laura Finzi

Differences between positively and negatively supercoiled DNA that topoisomerases may distinguish 73
Jonathan M. Fogg, Daniel J. Catanese, Jr., Graham L. Randall, Michelle C. Swick, and Lynn Zechiedrich

Calibration of tethered particle motion experiments 123
Lin Han, Bertrand H. Lui, Seth Blumberg, John F. Beausang, Philip C. Nelson, and Rob Phillips

Difference topology: Analysis of high-order DNA-protein assemblies 139
Makkuni Jayaram and Rasika Harshey

Useful intrusions of DNA topology into experiments on protein-DNA geometry 159
Jason D. Kahn, James R. Jenssen, and Vasavi Vittal

Topological analysis of DNA-protein complexes 177
Soojeong Kim and Isabel K. Darcy

Closing the loop on protein-DNA interactions:
Interplay between shape and flexibility in
nucleoprotein assemblies having implications
for biological regulation 195
Stephen D. Levene and Yongli Zhang

Four-way helical junctions in DNA molecules 213
David M.J. Lilley

Micromechanics of single supercoiled DNA molecules 225
John F. Marko

Flexibility of nucleosomes on topologically constrained
DNA 251
Andrei Sivolob, Christophe Lavelle,
and Ariel Prunell

The mathematics of DNA structure, mechanics,
and dynamics 293
David Swigon

Paradox regained: a topological coupling of nucleosomal
DNA wrapping and chromatin fibre coiling 321
Andrew Travers

Statistical-mechanical analysis of enzymatic topological
transformations in DNA molecules 331
Alexander Vologodskii

Correction to: The Mathematics of DNA Structure, Mechanics, ... C1
and Dynamics

List of workshop participants 347

NICK COZZARELLI: A PERSONAL REMEMBRANCE

STEPHEN D. LEVENE* AND LYNN ZECHIEDRICH†

In the weeks following Nick Cozzarelli's untimely passing two years ago, much was written about his fundamental contributions to molecular biology [1–6]. It is not our intent here to recapitulate an account of his groundbreaking scientific contributions or his outstanding service to the scientific community, which were covered well previously. Instead, we offer

---

*Departments of Molecular & Cell Biology and Physics, University of Texas at Dallas, Richardson, TX 75080.

†Departments of Molecular Virology & Microbiology and Biochemistry & Molecular Biology, Baylor College of Medicine, Houston, TX 77030.

C.J. Benham et al. (eds.), *Mathematics of DNA Structure, Function and Interactions*,
The IMA Volumes in Mathematics and its Applications 150,
DOI 10.1007/978-1-4419-0670-0_1,

a personal portrayal of Nick as seen through the eyes of those who worked in his laboratory. As postdocs in the Cozzarelli laboratory spanning the years between 1989 and 1997 (S.D.L. 1989–92; L.Z. 1990–97), we were privileged to witness firsthand a period when the scope of Nick's interests and the range of techniques he would apply to problems underwent a dramatic expansion.

Nick called himself a biochemist, but one of the many things that made him unique was his ability to readily grasp and apply mathematical and physical concepts to problems involving DNA. Unlike many classically trained biochemists of his generation, Nick was as much at home discussing science with mathematicians and physicists as with colleagues who came from backgrounds similar to his. This was a major attraction of Nick's program for students and postdocs, who came from as diverse a collection of disciplines as one could imagine in a biochemistry laboratory. It is with this perspective that we dedicate this volume to the memory of Nick Cozzarelli.

Steve recalls that when he first arrived in Berkeley Nick's laboratory space was located in Stanley Hall, also known historically as the "virus lab." The building was a vestige of the 1950s, but has since been replaced by a state-of-the-art bioengineering and biophysics building of the same name. The Cozzarelli laboratory at that time was spacious, but remarkable for its lack of any recent renovation. It became clear shortly after one's arrival that the low-tech/high-tech dichotomy reflected in the ambience of the laboratory space mirrored Nick's approach to scientific problems. He was fearless in making use of new technology, sometimes long before the underlying principles became understood (such as the analysis of DNA topology by gel electrophoresis) or before equipment was readily available (such as postdoctoral colleague Junghuei Chen's improvised "PCR" set up, which consisted of a series of beakers over Bunsen burners, a timer and a pair of hands. At 2-minute intervals Junghuei would alternately plunge microcentrifuge tubes into hot- and cold-water baths – it worked beautifully). At the same time, Nick recognized the power of computer simulation and modeling before many other biochemists did and made extensive use of computation to verify or predict experimental outcomes.

Casual acquaintances would characterize Nick as an extrovert; he was wonderful to meet and had something interesting to say on almost any topic to everyone. When Lynn interviewed with him, he took her to lunch at the Berkeley Art Museum on campus. A famous abstract painter was giving a lecture on his art, which was installed then at the museum. When the artist finished his lecture, he asked for any questions and Nick was the first to raise his hand and they ended up having a long, animated conversation about the value of art in scientific publications. Indeed, Nick considered the art of illustration extremely important, which often led to the exchange of multiple figure drafts in the course of preparing manuscripts. This is just one example of many things that Nick passed along to his trainees.

To those who knew him well, “extrovert” was only part of the complicated equation that was Nick. “Intense” was another part. In conversation, Nick listened intently and always identified and grasped the most important points. With near-brutal execution, he would expose any weakness in the argument or the data, usually during laboratory meetings. There are many stories of how, over the years, his students and postdocs would respond to his direct approach. Tears were one response. A quick exit and slam of the door were another. But for the most part, his trainees prepared better and thought harder about their results; learning in the process how to identify weaknesses in an approach. The intensity of Nick’s criticism was never personal, even if sometimes it could feel that way. The focus was always on asking the best questions and answering them conclusively.

Nick approached everything he did in exactly the same way – as a quest for perfection. He recognized and appreciated excellence in all realms-sports, food, music, art. He was a bibliophile. Nick said once that the day he had to stop reading literature for science would be the day he quit science. That explained how he kept up on a wide range of topics and could discuss any of them intelligently and passionately. One of Nick’s most cherished possessions was his and Linda’s Japanese garden, which was stocked with many rare plant specimens. His extraordinary attention to detail was apparent in this beautiful garden, and this trait, too, infused his approach to science.

He was acutely aware of his own faults and limitations; without hesitation, he sought feedback from those around him and could accept criticism as well as he could deliver it. Because of his directness, these interactions seemed natural. Lynn recalls an incident when Nick emerged from his office and blurted out, “Lynn, am I a sexist pig?” After a long pause, his color changed to pale. He hung his head, “Oh no, it’s true.” Her response: “You don’t mean to be, but there are some things that you do and say that could be perceived as sexist by some people.” After a long pause and with a dejected expression, Nick replied: “Please tell me.” It takes a very strong person to look directly at his potential weaknesses. He listened carefully and thoughtfully to her comments, thanked her for her honesty, and committed himself to change. He truly had not realized how some of what he did or said might look or sound. The ability to adapt, and accept criticism without judgment or rancor was an important ingredient of Nick’s success and accounts in large measure for the stunningly successful collaborations that he was involved in throughout his career.

With his directness and intensity Nick could be labeled, perhaps appropriately, as mercurial. However, Nick would always express his feelings and then move on. He was the exact opposite of the passive-aggressive stereotype, who maintains a pleasant facade, but is privately critical. If he was unhappy about something he would confront you (sometimes raising his voice), but this was usually followed up with praise behind your back. He was remarkably supportive when it counted.

Nick always appreciated and showed his appreciation to the members of his laboratory for the work that they did. He was generous in crediting others for ideas and data. This was a brilliant training move, one that empowered trainees and helped them to take ownership of their projects, thereby instilling confidence. However, it also speaks to the fact that Nick was never one to get caught up in the scramble for credit, who came first or did what. He put results and scientific goals above all else. Not that if you crossed the line he wouldn't let you know it. He once reviewed a manuscript submitted for publication and the colleague "forgot" to cite one of the ideas that Nick was most proud of: the "poison" hypothesis for the mechanism of quinolone gyrase inhibitors. He immediately picked up the phone and asked the scientist, "What the hell do you think you're doing ignoring our manuscript?" Of course, when the paper was published, Nick and Ken Kreuzer's work was properly credited.

When Nick was talking on the phone or working on a manuscript or grant in the office with his door closed, he was grumpy if you interrupted him. So, everyone in the laboratory was loathe to take a phone call on the laboratory extension for Nick because that meant you would be the one to have to interrupt him. One day Lynn picked up. "I need to speak with Dr. Nicholas Cozzarelli please." She said, "This is not his office phone number. Please call him there." The man responded, "I've been trying to call that number for hours; I need to speak with him- is he there?" Sighing, she said, "Yes, but he must be on the phone or busy if he is not answering his phone (trying hard to communicate that interrupting him was not a good idea)." The man was insistent so she grudgingly knocked on Nick's door. "TELL HIM TO CALL ME ON THIS LINE!" was Nick's response. "Nick, I told him that, but he is very insistent that you come to the phone." The door whipped open and she could see the frustration roil in his eyes as he stomped to the laboratory phone in his house slippers.

There was no missing Nick's half of the ensuing conversation:

> "HELLO?" he shouted into the mouthpiece.
> "YES, I KNOW ABOUT THAT PROGRAM AND IT STINKS. I THINK IT IS A TERRIBLE IDEA AND I HAVE TOLD EVERYONE I COULD..."
> "What?" "Oh, thank you very much... yes, I gratefully accept."

He hung up the phone and turned to Lynn and then-postdoc Roland Kanaar. "I just won one of those NIH merit awards." They took him out for a beer to celebrate a great honor - 10 years of NIH support without having to write a competing renewal.

There are many qualities involved in good mentorship, one of which is to know when to coach and when to let go. Nick had extraordinary intuition in this regard; a gift for knowing how to deliver encouragement at the right moment and in the right way, but giving students and postdocs the freedom to follow their own paths. He was sensitive to individual styles and recognized that science is an endeavor where "one size" does not fit all. The importance of having confidence in one's own abilities and the conviction of one's ideas were stressed; ever the optimist, Nick always looked at the positive first while maintaining a healthy level of skepticism.

One thing that his trainees and close friends would probably all agree on is that Nick was humble and considered himself lucky. "Me? Really" was his response to being told he had been elected into the National Academy of Sciences. Luck was something he always talked about in scientific contexts and we always had the sense of Nick's appreciation for all he was able to accomplish in a career that seems much too short. But it's our view that Nick created his own luck through passion, dedication, integrity, and a candid view of the world that sadly seems to belong to a different era. We miss the values that Nick espoused just as we miss him as a friend, colleague and mentor.

REFERENCES

[1] B. Alberts, *The Nick Cozzarelli I Knew.* Proc. Natl. Acad. Sci. **103**: 6077, 2006.
[2] J.C. Wang, *Nicholas Cozzarelli 1938–2006.* Nat. Struct. Mol. Biol. **13**: 469, 2006.
[3] R. Kanaar and D. Sherratt, *Nicholas (Nick) R. Cozzarelli 1938–2006.* Cell **125**: 415–417, 2006.
[4] C. Bustamante, *Nicholas R. Cozzarelli (1938–2006).* ACS Chem. Biol. **4**: 123–125, 2006.
[5] D. Kennedy, *Nick Cozzarelli.* Science **312**: 159, 2006.
[6] Preface and Dedication to Nicholas Cozzarelli, 1938–2006. Ann. Rev. Biochem. **76**, 2007.

MATHEMATICAL METHODS IN DNA TOPOLOGY: APPLICATIONS TO CHROMOSOME ORGANIZATION AND SITE-SPECIFIC RECOMBINATION

JAVIER ARSUAGA*, YUANAN DIAO†, AND MARIEL VAZQUEZ‡

Abstract. In recent years, knot theory and low-dimensional topology have been effectively used to study the topology and geometry of DNA under different spatial constraints, and to solve the topological mechanisms of enzymes such as site-specific recombinases and topoisomerases. Through continuous collaboration and close interaction with experimental biologists, many problems approached and the solutions proposed remain relevant to the biological community, while being mathematically and computationally interesting. In this paper, we illustrate the use of mathematical and computational methods in a variety of DNA topology problems. This is by no means an exhaustive description of techniques and applications, but is rather intended to introduce the reader to the exciting applications of topology to the study of DNA. Many more examples will be found throughout this book.

Key words. DNA knots, bacteriophage P4, DNA packing, random knots, site-specific recombination, Xer, tangles.

AMS(MOS) subject classifications. Primary 57M25, secondary 92B99.

Motivation. DNA presents high levels of condensation in all organisms. Volume reduction, defined as the ratio between the volume occupied by a given genome and the volume occupied by a random walk of the same length as the genome, ranges from 10^2 in *Escherichia coli* to 10^4 in humans[50].

These large condensation values lead to questions such as how the DNA is packed inside the eukaryotic cell nucleus, the prokaryotic cell, as well as inside other organisms such as DNA viruses. The complexity of the packing problem is magnified when one considers that the DNA molecule needs to be readily available to multiple biological processes essential to the proper functioning of the organism, such as DNA replication, transcription, recombination and repair. The cell has evolved tools to remove unwanted DNA entanglement and solve other topological problems, such as DNA un- (or over-)winding, knotting or linking, and formation of multimers, that may interfere with its functions. DNA topology, the study of geometrical

---

*Department of Mathematics, San Francisco State University, San Francisco, CA 94132, USA (`jarsuaga@sfsu.edu`). This research was supported in part by the Institute for Mathematics and its Applications (IMA). J. Arsuaga is partially supported by NIH grant NIH-2S06GM52588-12.

†Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA (`ydiao@uncc.edu`). Y. Diao is partially supported by NSF grant DMS-0712958.

‡Department of Mathematics, San Francisco State University, San Francisco, CA 94132, USA (`mariel@math.sfsu.edu`). M. Vazquez is partially supported by the Institute for Mathematics and its Applications (IMA), and by NIH grants 2S06GM052588 and 20123550 U56 Mentoring core.

C.J. Benham et al. (eds.), *Mathematics of DNA Structure, Function and Interactions*,
The IMA Volumes in Mathematics and its Applications 150,
DOI 10.1007/978-1-4419-0670-0_2,

(supercoiling) and topological (knotting) properties of circular DNA, provides the necessary experimental and computational techniques to describe and quantify these problems and their solutions.

The paper is divided into two parts. In each part we present an important problem in DNA topology, and the mathematical and computational tools used to address it.

In Part I we discuss the formation of knots in bacteriophages and its implications for phage packing geometry. Bacteriophages are viruses that propagate in bacteria. Most dsDNA bacteriophages pack their genome in a similar way inside the capsid, a proteinic enclosure with icosahedral symmetry. In the 1980s Liu and colleagues found that DNA extracted from bacteriophages P4 and P2 capsids was mostly knotted [63, 64]. The origin of these knots and whether they contained any information about the organization of the DNA inside the capsid remained unexplored. Here we will describe our current knowledge on how these DNA knots are formed, in particular we will focus on different mathematical models that have been proposed to explain their formation. We will also emphasize how this problem has been amenable to interdisciplinary studies and has generated new mathematics [2, 7, 66].

Part II deals with the resolution of topological obstructions arising during replication of the E. coli chromosome. The bacterial chromosome, a $4.6 Mbp$ double-stranded DNA circle, is condensed 10^3 times inside the nucleoid. The two DNA strands are wrapped around each other an average of $420,000$ times in the supercoiled bacterial chromosome and therefore the DNA double-helix must be unwound in order to be copied. Interwinding of newly replicated sister chromosomes in a partially replicated chromosome forms precatenanes, which become catenanes (links) upon completion of replication. Without careful management by cellular machinery, replication of the bacterial chromosome would lead to two sister molecules highly linked together. The cell must solve the topological problem of separating the two linked sister chromosomes to ensure proper segregation at cell division. Unlinking of replication catenanes is mainly achieved by the type II topoisomerase Topo IV (reviewed in [41, 81]).

Furthermore, stalled or broken replication forks are repaired by homologous recombination. Occasionally crossing-over by homologous recombination generates DNA dimers, which may be knotted [84]. The dimers are resolved by Xer recombination. The Xer system consists of enzymes XerC and XerD, which act cooperatively and co-localize at the septum with the protein FtsK. FtsK plays an essential role in dimer resolution, coordinates chromosome segregation and cell division (reviewed in [9]). Recent experimental evidence shows that XerCD-FtsK recombination can unlink catenanes formed by site-specific recombination *in vitro* [52], as well as catenanes formed by replication *in vivo* [47]. Here we will review the tangle method for site-specific recombination. We will illustrate the method with applications to Xer recombination. The analysis will lead to several possible topological pathways followed by the enzymes. The question is posed as to whether the different pathways are simple planar projections of the same 3-dimensional topological mechanism.

Part I. DNA Knotting in Bacteriophages.

In this part of the paper we present the problem of DNA knotting in bacteriophage P4 as well as the various tools from the theory of random knotting used to approach this problem. Bacteriophage P4 knots are formed by random cyclization. In section 1 we introduce the problem of random cyclization of DNA in free solution. We discuss several computational methods currently used to simulate this process, as well as the corresponding analytical results to estimate the knotting probability of a random polygonal curve in $\Re^3$. This work is used as a framework to study the problem of DNA knotting in bacteriophages. Section 2 deals with cyclization of DNA in confined volumes. First, we review some of the experimental results on DNA knots found in the bacteriophage P4 system. This is followed by the description of three computational models and how these models have been used to address the biological problem. In section 3 we discuss the limitations of these approaches and future directions.

1. Cyclization of DNA molecules in free solution.

1.1. Experimental studies on random cyclization of DNA molecules. Random cyclization of long linear DNA molecules with sticky (i.e. complementary) ends produces knots with non-trivial probability. This knotting probability was independently estimated by Rybenkov et al. [78] and by Shaw and Wang [83]. Both groups showed that the formation of these knots depends on the length of the DNA molecule and on the ionic conditions of the solution (i.e. the effective diameter of the DNA molecule). In [78] it was found that the knotting probability for P4 DNA molecules circularized in solution was 3% and that the trefoil was the prevalent knot population followed by smaller amounts of the four crossing knot and even smaller amounts of the five crossing knots. Monte-Carlo simulations of idealized polymer chains (e.g [58, 66]) and analytical results [34] support these experimental results as explained below.

1.2. Simulations of Gaussian and equilateral random polygons without confinement. The wormlike chain is the most accurate polymer model for simulating DNA in solution. However other models such as the equilateral random polygon (ERP) or the Gaussian random polygon are good for estimating properties of long DNA molecules in the bulk and at the same time are more amenable to the development of rigorous analytical results. Several algorithms have been proposed for generating samples of equilateral random polygons. These include the crankshaft algorithm [58, 68], the hedgehog algorithm [58] and the pairwise rotation

Table 1

Summary results concerning various random polygons. Results in parenthesis are numerical results and – is the case when the numerical results vary because of the choices of parameters. GP_n *stands for a Gaussian random polygon of n edges,* EP_n *stands for an equilateral random polygon of n edges,* CEP_n *stands for an equilateral random polygon of n edges within a confined space (usually a sphere with a predetermined radius),* R_n *stands for a uniform random polygon of n edges and* SP_n *for spooling random model.*

	mean ACN	leading coeff.	knotting prob
GP_n	$O(n \ln n)$	$1/2\pi$	$\geq 1 - e^{-n^{\epsilon}} \to 1$
EP_n	$O(n \ln n)$	$3/16$	$\geq 1 - e^{-n^{\epsilon}} \to 1$
CEP_n	$O(n^2)$	$--$	$(\geq 1 - e^{-n^{\epsilon}} \to 1)$
R_n	$O(n^2)$	$(.115)$	$(\geq 1 - e^{-n^{\epsilon}} \to 1)$
SP_n	$O(n^2)$	$--$	$(\geq 1 - e^{-n^{\epsilon}} \to 1)$

algorithm [73]. The crankshaft algorithm is fairly popular among some researchers. In this algorithm two vertices of the polygon are selected at random, dividing the polygon into two subchains. One of the two subchains is selected at random (with equal probabilities for each subchain), and the selected subchain is rotated through a random angle around the axis connecting the two end vertices of the subchain. In the hedgehog algorithm an ERP is first generated and at each step two vectors are selected at random, rotated with respect to their sum and placed back in the polygon. The crankshaft algorithm generates an ergodic Markov chain in the space of all ERPs of fixed length [68]. However the main drawback with this algorithm is that the correlation between samples is very high and therefore many configurations need to be generated in between any two samples in the Markov chain. The hedgehog algorithm on the other hand generates independent samples of polygons however it is unknown whether it is ergodic or not. These algorithms have helped to estimate some of the biologically relevant properties, such as the knotting probability and the mean of the Average Crossing Number distribution (ACN), for equilateral polygons as shown in Table 1. The leading coefficient 3/16 in the equilateral random polygon case means the mean ACN of equilateral random polygons of length n is of the form $(3/16)n \ln n + O(n)$. This number is $1/2\pi$ for the case of Gaussian random polygons.

1.3. Analytical results for knotting of polygons without confinement: Gaussian and equilateral random polygon models. A Gaussian random vector $X = (x, y, z)$ is a random point whose coordinates x, y and z are independent standard normal random variables (with mean $= 0$ and variance $= 1$). The pdf (probability density function) of X is the joint pdf of x, y and z, which is

$$f(X) = \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{x^2+y^2+z^2}{2}} = \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{|X|^2}{2}}.$$

A Gaussian random walk of n steps (denoted by GW_n) consists of $n+1$ consecutive points $X_0 = (0,0,0) = O$, X_1, X_2, ... , X_n such that $Y_{k+1} = X_{k+1} - X_k$ ($k = 0, 1, ..., n-1$) are independent Gaussian random vectors. It follows that the joint pdf for all the vertices is

$$\begin{aligned} f(X_1, X_2, ..., X_n) &= \left(\frac{1}{\sqrt{2\pi}}\right)^{3n} e^{-\frac{1}{2}(|Y_1|^2+|Y_2|^2+\cdots+|Y_n|^2)} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^{3n} e^{-\frac{1}{2}(|X_1|^2+|X_2-X_1|^2+\cdots+|X_n-X_{n-1}|^2)}. \end{aligned}$$

A Gaussian random polygon GP_n is a conditioned GW_n of n steps such that the last vertex X_n coincides with the starting point $X_0 = O$. Thus, if we let $g_n(X_n)$ be the pdf of X_n for a GW_n,

then the joint pdf of X_1, X_2, ..., X_{n-1} of a GP_n is

$$h(X_1, X_2, ..., X_n) = f(X_1, X_2, ..., X_n)/g_n(O).$$

The one advantage of the Gaussian random polygons (over other random polygon models) is that the joint probability density function of its vertices is of an explicitly nice form. This enabled the derivation of the following result concerning the knotting probability of a GP_n [34].

THEOREM 1.1. *[34] Let $\mathcal{K}$ be any knot type, then there exists a positive constant ϵ such that GP_n contains $\mathcal{K}$ as a connected sum component with a probability at least* $1 - \exp(-n^\epsilon)$ *provided that n is large enough.*

One can obtain a similar result for equilateral random polygons.

Suppose Y_1, Y_2, ... , Y_n are n independent random vectors uniformly distributed on S^2. An equilateral random walk of n steps, denoted by EW_n, is defined as the sequence of points in the three dimensional space $\mathbf{R}^3$: $X_0 = O$, $X_k = Y_1 + Y_2 + \cdots + Y_k$, $k = 1, 2, ..., n$. Each X_k is called a vertex of the EW_n and the line segment joining X_k and X_{k+1} is called an edge of EW_n (which is of unit length). Notice that the coordinates of each point are not independent from each other due to the fact that the distance between consecutive points in the polymer needs to be one. If the last vertex X_n of EW_n is fixed, then we have a conditioned random walk $EW_n|X_n$. In particular, EW_n becomes a polygon if $X_n = O$. In this case, it is called an equilateral random polygon and is denoted by EP_n. The joint probability density function $f(X_1, X_2, ..., X_n)$ of the vertices of an EW_n is $f(X_1, X_2, ..., X_n) = \varphi(U_1)\varphi(U_2)\cdots\varphi(U_n) = \varphi(X_1)\varphi(X_2 - X_1)\cdots\varphi(X_n - X_{n-1})$. Where $\varphi(U_i)$ is the density function of selecting a random point over the surface of the sphere.

Let X_k be the k-th vertex of an EW_n ($n \geq k > 1$), its density function is defined by

$$\begin{aligned} f_k(X_k) = \int\int\cdots\int \varphi(X_1)\varphi(X_2 - X_1)\cdots \\ \varphi(X_k - X_{k-1})dX_1dX_2\cdots dX_{k-1} \end{aligned} \tag{1.1}$$

and it has the closed form $f_k(X_k) = \frac{1}{2\pi^2 r}\int_0^\infty x\sin rx\left(\frac{\sin x}{x}\right)^k dx$ [74]. In the case of EP_n, the density function of the vertex X_k can be approximated by a Gaussian distribution, as given in the following lemma [30, 33, 34].

Lemma 1.1. *Let X_k be the k-th vertex of an EP_n and let h_k be its density function, then*

$$h_k(X_k) = \left(\sqrt{\frac{3}{2\pi\sigma_{nk}^2}}\right)^3 \exp\left(-\frac{3|X_k|^2}{2\sigma_{nk}^2}\right) + O\left(\frac{1}{k^{5/2}} + \frac{1}{(n-k)^{5/2}}\right), \quad (1.2)$$

where $\sigma_{nk}^2 = \frac{k(n-k)}{n}$.

In other words the density of the k step of an EP_n can be approximated by a Gaussian distribution with mean 0 and a standard deviation that depends on The vertex number k and on the distance from the vertex to the origin (or first vertex in the polygon).

This lemma provided the key link to apply the methods used in [34] for the Gaussian random polygons to the equilateral random polygons, which leads to the following theorem.

Theorem 1.2. *[30] Let $\mathcal{K}$ be any knot type, then there exists a positive constant ϵ such that EP_n contains $\mathcal{K}$ as a connected sum component with a probability at least $1 - \exp(-n^\epsilon)$, provided that n is large enough.*

Numerical studies on EP_n suggest a scaling law of $1 - \exp(-n/a)$ with $a = 244 \pm 5$ (see [66] and references therein).

The above two theorems imply that a long GP_n or EP_n contains many connected sum components (with a high probability), which makes it highly unlikely for the polygon to be achiral. This is stated in the following corollary. However, this only provides reason for the long GP_n and EP_n to favor chiral knots than achiral ones. For relatively short polygons, this is not clear.

Corollary 1.1. *[30, 34] There exists some constant $\theta > 0$ such that the probability that a GP_n or an EP_n is a chiral knot is at least $1 - \frac{1}{n^\theta}$.*

The determination of the knot type of a circular molecule can tell us its topological (minimum) crossing number, i.e., the minimum number of crossings one will see no matter how this molecule is artificially stretched, twisted, or bent. However, the average crossing number (ACN), defined as the average of crossing numbers over all orthogonal projections of the molecule, is a more natural geometric measure of the molecule entanglement as it refers to the actual number of crossings that can be perceived while observing a non-perturbed trajectory of a given molecule [55]. Furthermore, it is believed that DNA knots migrate in gel eletrophoresis accordingly with their ACN [99].

The following theorems are presented in [31, 32] and establish the $O(n \ln n)$ behavior of the mean ACN for the Gaussian and equilateral random polygons as illustrated in Table 1.

Theorem 1.3. *Let χ_n be the ACN of an equilateral random walk of n steps; then*

$$E(\chi_n) = \frac{3}{16} n \ln n + O(n).$$

On the other hand, if χ'_n is the ACN of a Gaussian random polygon of n steps, then

$$E(\chi'_n) = \frac{1}{2\pi} n \ln n + O(n).$$

2. Cyclization of DNA molecules in confined volumes: DNA knotting in bacteriophage P4 capsids.

2.1. Experimental studies on DNA knots in bacteriophages. In dsDNA bacteriophages the volume of the bacteriophage genome is reduced 100 times inside the capsid [53]. This volume reduction imposes severe physical constraints on the DNA molecule. For instance the DNA molecule is under (at least) 50 atmospheres of pressure [42, 93] and at a concentration of 800mg/ml [56]. Despite these conditions the dsDNA molecule is believed to preserve its double helical structure [8] and not to have sequence-specific associations with the protein capsid. A number of models have been proposed to describe the organization of the viral chromosome under such extreme conditions of condensation. These include coaxial and concentric spooling models [4, 20, 35, 76, 82], coaxial models [10], toroidal models [51, 72], and liquid-crystaline models [61].

Bacteriophage P4 is an icosahedral phage of radius $r = 180$Å and a linear dsDNA genome of 11.5 kb ($l = 120 \times 10^3$Å). The genome is flanked by two 16*bp* long single stranded complementary sequences of DNA called *cos* sites. During phage morphogenesis a protein enclosure called capsid is assembled first. This is followed by the packing of a single linear DNA molecule into the capsid through the portal vertex. Infective viruses keep at least one of their *cos* sites attached near the portal [21]. This attachment prevents the two *cos* ends from meeting within the capsid and circularizing the chromosome. However in the experiments performed by Liu and colleagues it was found that most of the DNA molecules phenol extracted from bacteriophage P4 are circular and non-trivially knotted [63, 64].

Recent work [5, 6, 91, 92] reproduced and extended the results of Liu et al. Figure 1 shows a two dimensional gel of DNA knots from bacteriophage P4 in which different conditions are used in each dimension [91]. In this figure the top spot corresponds to the unknotted molecule followed, along a bell-shaped curve, by the trefoil knot, the figure eight (four-crossing) knot, and so on. The spot ahead of the bell is the linear chain. The most remarkable fact about this distribution is that about 95% of the DNA molecules are knotted and only about 2% are knots between 3 and 10 crossing knots.

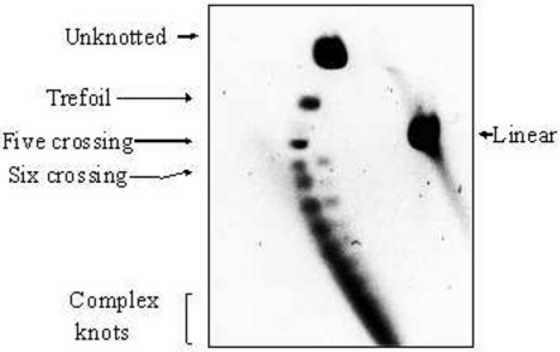

FIG. 1. *Two dimensional gel of knots extracted from bacteriophage P4.*

Furthermore the large majority of the population consist of knots with 30 crossings or more [5]. These results are in high contrast with those obtained by random cyclization of P4 DNA molecules in free solution (see Section 1.1 [78]) and suggested that knots extracted from bacteriophage P4 are formed inside the phage capsid and therefore may be used as reporters for chromosome organization in P4. Despite the small percentage of knots with less than 8 crossings (i.e. those that can be separated by gel electrophoresis) two important properties of the knot distribution were revealed. First the four crossing knot is mostly absent and second the torus knots 5_1 and the 7_1 are more probable than the twist knots 5_2 and 7_2 (contrary to what is expected in free solution). The theoretical work described next aims at explaining these experimental results. In this review we will focus only on the problems of knotting probability and complexity.

2.2. Random knotting within a confined space. A simple approach to study the knotting probability and complexity of P4 knots is by generating ensembles of random polygons inside different convex volumes. Next we describe three models: the Confined Equilateral Random Polygon (CEP_n), the Uniform Random Polygon (URP_n) and the Random Spooling (SP_n).

The Confined Equilateral Random Polygon.

In this model we consider ERPs confined to spheres of certain radius r and use CEP_n to denote such a polygon of length n. Figure 2 shows an example of such polygon.

Unfortunately, the extra condition that confines the polygon to a sphere of radius r completely invalidates the approximation formula given in Lemma 1.2 for the vertex X_k of CEP_n. Intuitively, a CEP_n would be more likely to be knotted than an EP_n. Indeed, this is confirmed by numerical studies [5, 65, 66] which were pioneered by Michels and Wiegel [65].

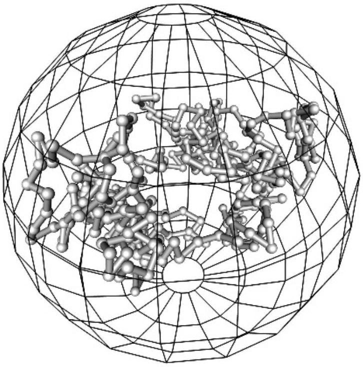

FIG. 2. *An equilateral random polygon inside a sphere.*

In their studies molecular dynamics algorithms were used to sample closed polygons and the knotting probability was computed. Michels and Wiegel found that the knotting probability of a polygon inside a sphere increases with respect to that in $\Re^3$ following a $\exp(N^{\alpha}/r^3)$ law with $\alpha = 2.28$. In more recent work [5, 66] large ensembles of CEP_n were generated by the crankshaft algorithm. In [66] the scaling law proposed by Michels and Wiegel was confirmed and the coefficient $\alpha = 2.15 \pm 0.04$ improved.

In [5], the combination of experimental and theoretical results led to propose that the effect of the confinement during the random cyclization process of the DNA molecule is one of the key drivers in the formation of knots in the P4 system. This argument has been extended by D. Smith′ group to explain the knotting of chains in confined volumes [75]

The Uniform Random Polygon.

Developing analytical results for CEP_n is a very difficult problem. An alternative model was proposed in [68] as a way to study the random behavior of circular DNA molecules packed in phage capsids that may provide clues about showing some of these analytical results. For $i = 1, 2, ..., n$, let $U_i = (u_{i1}, u_{i2}, u_{i3})$ be a three-dimensional random point that is uniformly distributed in the unit cube C^3 (or in a unit ball) such that U_1, U_2, ..., U_n are independent. Let e_i (called the i-th edge) be the line segment joining U_i and U_{i+1}, then the edges e_1, e_2, ..., e_n define a *uniform random polygon* R_n in the confined space (either the cube or the sphere), where e_n is the line segment joining U_n and U_1. A polygon of length n is denoted by URP_n.

While the knotting probability of an R_n has not been analytically determined (even in the case of $n \to \infty$), a numerical study carried out in [3] provided convincing data that the knotting probability of an R_n quickly approaches 1 as n approaches infinity. Figure 4 is the plot of the percentage of URPs with non-trivial determinant (i.e. those whose Alexander poly-

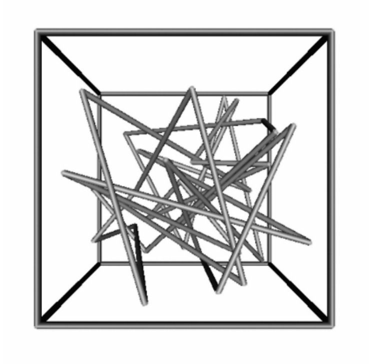

FIG. 3. *A uniform random polygon confined in the unit cube.*

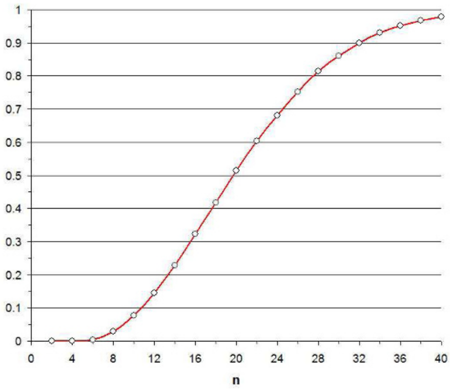

FIG. 4. *The lower bound of knotting probability for URPs up to 40 segments.*

nomial evaluated at $t = -1$ is non-trivial). Since the trivial knot has a trivial determinant, the results give a lower bound of the knotting probability. Figure 4 below is the plot of the data. The fitting curve used here is $1 - \exp(-0.000082n^3)$, although this is not to be expected as a general rule since the trivial knot probability of an R_n is at least of order $\exp(-n \ln n)$ as shown in [3].

It turns out that the mean ACN behavior for an R_n is much easier to determine, both analytically and numerically than for CEPs. Consider a uniform random polygon R_n with n edges e_1, e_2, ..., e_n in that consecutive order. Let $a(e_i, e_j)$ be the average crossing number between e_i and e_j, then the ACN of R_n is

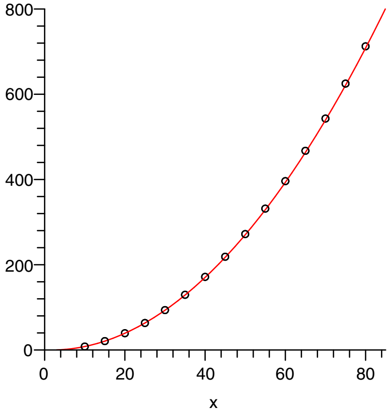

FIG. 5. *The mean ACN of uniform random polygons up to 80 vertices.*

$$\chi_n = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i-1,i,i+1} a(e_i, e_j).$$

It follows that the expected value of the average crossing number of R_n is

$$E(\chi_n) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i-1,i,i+1} E(a(e_i, e_j)) = p(n-3)n.$$

This establishes the $O(n^2)$ behavior of the mean ACN of an R_n as shown in Table 1. Numerical studies in [2] produced the following near perfect fit using $E(\chi_n) \approx 0.115(n-3)n$.

There have not been enough numerical studies on the knot types of the CEP_n's [5, 66] and R_n's [67] to indicate their bias against achiral knots, even though this is generally expected for long random polygons since achiral knots are much rarer than the achiral ones within large knots.

The Random Spooling model.

The last model we discuss is the *random spooling model.* This model incorporates features from the random knotting models (described above) into the spooling and toroidal models [4, 10, 20, 35, 51, 72, 76, 82]. In the standard spooling model DNA fibers spool around an axis forming coaxial spherical layers. In [59] the knot type of molecular dynamics generated spooling conformations was studied and it was found that most of these conformations were unknotted. These results together with the wide distributions of knots that are observed in P4 suggested that current theoretical models of DNA packing disregard the effect of random fluctuations which

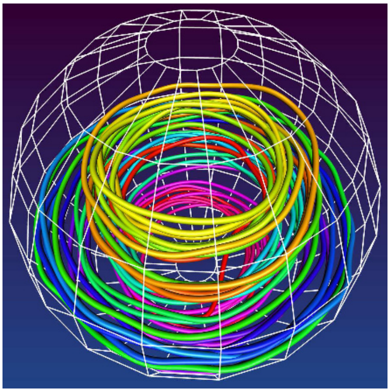

FIG. 6. *The random spooling model.*

in fact may play an important role in the packing of the viral chromosome. We recently proposed [7] that fibers follow spooling trajectories and at the same time they intermingle, as illustrated in the figure 6. This intermingling between fibers of different coaxial layers increases the knotting probability.

Some initial simulation and analytical results have been published [7]. For instance we have estimated the complexity of the average crossing number in the direction of the spooling axis as stated in the next theorem.

THEOREM 2.1. *[7] Let P_n^s be a spooling random polygon, then the average number of crossings in its projection to the xy-plane perpendicular to its center axis is of the order of $O(n^2)$.*

Although the knotting probability has not yet been shown to increase to 1 as suggested by the numerical results shown in Figure 7 a relationship between the writhe of the projection along the spooling axis and the knot type has been proven. The following theorem is a consequence of a theorem due to Morton[69]

THEOREM 2.2. *[7] Let $w(D_n)$ be the writhe of the projection in the direction of spooling axis and $\sigma(P_n^s)$ the number of times the spool goes around its axis. If $|w(D_n)| \geq \sigma(P_n^s)$, then P_n^s is a non-trivial knot. Furthermore, in this case P_n^s cannot be an achiral knot.*

This theorem shows that spooling conformations with high writhe are knotted. This agrees well with some of our results that relate DNA knotting and writhing in bacteriophage P4 [6, 11] and suggests that high writhe may also play an important role in the formation of knots in P4.

3. Conclusions. Here we have discussed the problem of knotting by random cyclization in free solution and in confined volumes. In both cases we have presented experimental, analytical and computational results. By comparing our experimental results with those obtained in free solution we

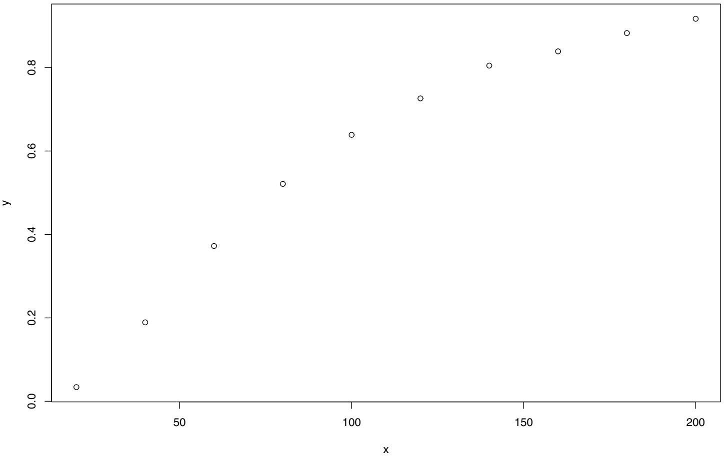

FIG. 7. *Knotting probability as a function of the length of the chain for the random spooling model.*

concluded that knotting in bacteriophage P4 occurs before, or very soon after, the disruption of the capsid and therefore P4 knots can be used as reporters of DNA packing. The large amount of knotting is still a feature that is not fully explained by current mathematical models. In this review we have presented three random knotting models: the confined equilateral polygon, the uniform random polygon and the spooling random polygon. All these models present consistent results however they do not reach the high levels of complexity found in bacteriophages. This is specially true if more accurate representations of the DNA molecule are taken into account. Nevertheless some information about the biological system has been extracted from these theoretical models. For instance our current simulations results suggest that DNA knotting in P4 is mainly driven by the confinement imposed by the capsid during the cyclization reaction, and perhaps also by possible biases introduced by the arrangement of the viral chromosome [6, 11]. Importantly none of the current idealized models proposed in the literature account for the formation of knots inside the capsid and previous simulations results failed to do so [59] thus suggesting that they may not reflect some important properties of the DNA packing. The random spooling model is our first attempt to address this issue. It remains to be seen if such models can reproduce the knot distributions observed experimentally. New experimental results have recently obtained for P4 deletion mutants whose genomes range between 5 and 8 *kb* [92]. These experiments hold a great promise for unveiling the properties that drive knotting in phage capsids as well as some of the essential features of the viral packing.