# Mathematics and physics of polymer entanglement: Emerging concepts and biomedical applications

## BIRS Workshop 10w5100

## January 10 – January 15, 2010

## Final Report

# 1 Introduction

This workshop focused on the mathematics associated with an array of cutting edge problems at the interface between the mathematical, physical, and biological sciences. In particular, the researchers targeted questions from work arising from molecular biology studies of DNA and other biopolymers for which an interdisciplinary approach could yield unique insights.

In the last decade or so, tremendous advances in the understanding of DNA behavior, including the effects of (i) storage (in viral capsids, eukaryotic nuclei, or bacterial cells), (ii) entanglement (knots and links), (iii) replication, (iv) transcription into RNA, and (v) repair and recombination (including site-specific and general), have been made at the hands of researchers working at the interface of mathematics, biology, and physics. Not only has the understanding of DNA as a biopolymer advanced rapidly, but emerging concepts have reached beyond the scope of DNA to a general understanding of the previously little-explored basic relationship between the local geometry of chain juxtaposition and global topology in polymer chains. Numerical simulations of lattice models as well as continuum freely-jointed and wormlike chain models demonstrated convincingly that the degree of 'hookedness' of an observed local juxtaposition correlates well with global topological complexity and the likelihood that a topoisomerase-like segment passage at the given juxtaposition would disentangle. This is a new paradigm opening up many avenues of computational and experimental research.

Moreover, these novel numerical results also serve to suggest a wealth of questions and conjectures that may be fruitfully addressed by field theory arguments from physics and by rigorous mathematics. Indeed, during the same time, we have noted a drastic increase in the precision in the language of biologists, with their incorporation of such important concepts as 'conjecture', 'hypothesis', and 'theory' following the traditional mathematical usage. An improved understanding of the languages of each of the three disciplines improves the communication, and as such, the understanding of each other. Increasing the awareness of mathematicians to (i) the complexity of the biological problems, as well as (ii) the cutting edge research results, even before they are published, facilitates an increased understanding of biopolymers, a primary goal of this workshop.

In the past, there have been efforts (such as the 2007 BIRS workshop 07w5095, The Mathematics of Knotting and Linking in Polymer Physics and Molecular Biology) to bring together researchers in these areas. Our goal was to include more Biologists/Experimentalists than before. As a result, this was the first opportunity for many of the invitees from different disciplines to meet each other.

Thus the workshop enabled the advancement of existing collaborations at the interface between Biology, Mathematics and Physics and encouraged the development of new ones. The conference was timely for an additional important reason. Research funding for the pure mathematical and physical sciences has decreased recently. However, together with this troubling trend, there is an increase in funding opportunities for mathematicians and physicists working at the interface of the biological sciences, perhaps particularly in regard to medically relevant research.

Rich in important problems only answerable with an interdisciplinary approach, the study of DNA polymer science has had extraordinary successes quite recently, with the vast majority of these occurring at the interface of disciplines. Bringing a cadre of researchers working at the interface of polymer science to the Banff International Research Station for Mathematical Innovation and Discovery provided the opportunity to bridge these fields.

# 2 Presentations

A total of 35 researchers from Biology (12), Chemistry (1), Physics (7), Computer Science (1), and Mathematics (14) attended the meeting. This group was quite diverse. The participants had a mixture of experience, from grad students through senior professors; came from a wide-variety of institutions, from teaching colleges to research-intensive universities, medical schools, and government research institutes; and represented eight different countries.

The workshop had 28 presentations representing a wide body of interdisciplinary research on the DNA polymer. These can be categorized into roughly five areas: 1) linking number and supercoiling in DNA; 2) modeling polymers and entanglement; 3) confinement effects; 4) length effects; and 5) DNA sequence-specific effects and DNA replication factories. We introduce each of these topics and then discuss the presentations

## 2.1 Linking number and supercoiling in DNA

Recent emerging results have been made in all atom simulations of DNA. Whereas coarse-grained models have been extremely useful for understanding polymer behavior (for example, knotting, linking, and in general how they are packed into small spaces), the next series of questions must begin to include the surprising way that the change in linking number, $Lk$, is manifested in DNA. The observed bimodal response of DNA to $Lk$ shows complete collapse of the DNA helix in sequence-dependent localized regions of the biopolymer with a concomitant relaxation back to B-form DNA in the rest of the biopolymer. At the same time, for the overwound helix, elastic polymer rod models work perfectly well. Mathematically and physically, this means, at least in the helix unwinding direction, that the assumptions of elastic rod theory are wrong and suggests that perhaps an asymmetric torsional potential would be physically more appropriate.

The workshop contained five talks on this subject.

Jonathan Fogg (Baylor College of Medicine, USA) spoke on *Supercoiling in DNA minicircles: To get the big picture, think small.* DNA supercoiling has a dramatic effect on its function. Indeed, for many biological processes a distinct threshold of supercoiling must be reached before the reaction can occur. Although the global conformational changes that occur as a result of supercoiling are reasonably well understood, relatively little is known about the consequences of DNA supercoiling on the local level. These sequence-specific conformational changes must surely dictate how proteins recognize and metabolize DNA. Even the largest DNA binding proteins are very small relative to chromosomal or plasmid DNA and are, therefore, unable to sense global DNA topology. Fogg, with his research group, developed and utilized a protocol to produce milligram quantities of supercoiled minicircle DNA, as small as 250 base pairs (bp). Individual topoisomers were isolated ranging from $\sigma = +0.08$ to $-0.19$. Their supercoiled minicircle substrates provide a unique insight into the local DNA structure of supercoiled DNA and how this is recognized and manipulated by enzymes. Several unexpected aspects of supercoiled DNA were revealed from their studies of DNA minicircles. They discovered that the addition or subtraction of three base pairs has a profound effect on the gel electrophoretic mobility of small DNA circles. The topoisomers of a 336 bp minicircle display a very regular pattern of electrophoretic mobility. When they generated topoisomers of 333 bp and 339 bp minicircles, however, several of these minicircles show unexpected electrophoretic behavior. They also discovered a topoisomer that appears to flip between open and writhed conformations, akin to the "frustrated" minicircles detected in their computational simulations. Notably, they found that positively supercoiled minicircles have a much higher propensity to writhe than negatively supercoiled minicircles, even in the absence of added divalent metal ions. In contrast, limited writhe was observed for negatively supercoiled minicircles in the absence of added divalent metal ions, demonstrating the importance of electrostatic effects on DNA structure. Many models of DNA elasticity incorrectly assume that positively supercoiled DNA is equal and opposite to negatively supercoiled DNA. Their findings prove there is a distinct asymmetry.

Sarah Harris (University of Leeds, England) spoke on *Computer Simulations of DNA Supercoiling at the Atomic Level.* The discovery of the structure of duplex DNA revealed how cells store genetic information. However, researchers are far from understanding the more complex biological question of how this information is regulated and processed by the cell. DNA topology and supercoiling is known to affect DNA transcription as changes in topology affect DNA conformation, and can thereby modify the interaction between regulatory DNA-binding proteins and their target sites. Small DNA circles offer a controllable model system for the systematic exploration of the dependence of DNA structure on supercoiling. Harris's research group uses computer simulation to explore the supercoiling-dependent conformation of small DNA circles, in particular their writhe, and how this is affected by supercoiling, salt concentration, DNA sequence and the size of the circles. The calculations use atomistic molecular dynamics simulation, and employ both implicit and explicit solvent models. They have been systematically testing their computational models against experimental data for small circles of between 65 and 214 bp. They have also been investigating the supercoiling-dependent binding of a 3rd DNA strand (triplex formation) to a target site within a writhed DNA circle for comparison with experimental data. These preliminary calculations are designed to explore the thermodynamics of supercoiling-dependent binding, and use triplex formation as a model system for exploring the importance of supercoiling in DNA recognition in general.

Steve Levene (University of Texas at Dallas, USA) spoke on *Loop-mediated regulation by lac repressor: does DNA supercoiling play a role?* Interactions of *E. coli* lac repressor (LacI) with a

pair of operator sites on the same DNA molecule can lead to the formation of looped nucleoprotein complexes both *in vitro* and *in vivo*. The lac system is a major paradigm for loop-mediated gene regulation in prokaryotic cells; however, the complex interplay between DNA topology, modulation of chromosome topology by architectural-DNA binding proteins, and loop-mediated regulation remain poorly understood. Levene discussed the effects of DNA supercoiling on LacI mediated looping *in vitro* investigated by a combination of fluorescence resonance energy transfer studies, semi-analytical DNA elasticity calculations, and Monte Carlo simulation.

David Levens (Center for Cancer Research NIH, USA) spoke on *Genome-wide functional correlation between transcription, DNA conformation and topology*. His group is investigating the role of dynamic supercoiling in the regulation of gene expression and DNA structure *in vivo* and *in vitro*. They have developed a method to map unpaired bases across the genome using potassium permanganate. Besides the expected signature of transcription bubbles at promoters, other sites of non-B-DNA occurring outside of genes were often sensitive to transcription inhibition, suggesting a long-distance coupling between transcription and DNA conformation via transmission of mechanical stress (dynamic supercoils). Such stress is generated as DNA is threaded through the RNA polymerase active site and propagated to remote sequences. Supercoil sensitive unusual DNA structures may contribute to the real-time self-regulation of many genes. Previously his group has demonstrated the existence and measured the magnitude of such dynamical supercoils *in vivo*. Now, they have developed an approach to build a genome-scale map of DNA supercoiling using psoralen intercalation as a probe. The map shows that negative supercoiling often propagates to or beyond 2 kilobase (kb) upstream of active promoters. This supercoiling contributes to the formation of a variety of non-B DNAs, including quadruplex and Z-DNA. These non-B DNA structures may be recognized by proteins and contribute to a variety of control mechanisms. Overlaying the maps of DNA supercoiling and conformation with the *in vivo* binding sites of structure-sensitive transcription factors as well as sites of topoisomerase I and II action may reveal new modes of transcriptional regulation on a global scale.

Lynn Zechiedrich (Baylor College of Medicine, USA) spoke on *Supercoiled minicircles as gene therapy vectors*. To study DNA supercoiling and DNA topoisomerases, Zechiedrich's group created a way to make milligram quantities of minicircle DNAs of a few hundred base pairs. These DNAs have been extremely useful for this purpose and Fogg also discussed this work. Zechiedrich presented data showing that supercoiled minicircles are superior vectors for delivering DNA into human cell types that no other DNA vector has been previously able to penetrate. In cells, DNA sequence is transcribed from these minicircles into small RNAs that regulate gene expression. Even small genes can be expressed from supercoiled minicircles. Supercoiled minicircles resist sheer forces associated with gene therapy delivery and are significantly less susceptible to the nucleases in human serum than normal plasmid DNA vectors of a few thousand base pairs. These data show that supercoiled minicircles are a promising new tool for gene delivery.

## 2.2 Modeling polymers and entanglement

Developing and analyzing models of polymer and biopolymer entanglements is a multistage and interdisciplinary process. In order to be able to make direct quantitative comparisons with experi-

ment, often polymer models must be highly complex and studied primarily by computer simulation. Such models are less likely to be mathematically tractable, however, and hence it is also often useful to investigate simplified polymer models with the goal of making qualitative comparisons with experiment. At the same time, defining and characterizing the nature of "entanglements" can raise questions of a purely mathematical nature. Thus to understand polymers such as DNA, a combination of efforts is necessary. As examples, researchers study individual polymers moving seemingly at random, such as wormlike rods and freely jointed chains, and collections of polymers, as in the case of chromosome territories (seen in the confinement section). These models provide a convenient framework for studying problems like the effect of local strand passages and the clasp conjecture for topoisomerase (i.e. that topoisomerase II acts preferentially at clasps). For example, one might use lattice polygons to study changes in knotting resulting from strand passage in certain configurations or pass to more topological methods using a tangle model. While researchers agree on what a knot is, there are subtleties concerning knot types such as chirality and orientation reversability, which become more problematic when one studies compositions of knots. The knot tables only tell part of the story, disregarding many of the properties of the actual configurations which become quite important in the physical world. These configurations hold other secrets as well, properties shared by all knotted configurations, such as quadrisecants, which can be studied using a combination of geometric and topological methods.

The workshop contained six talks on this subject.

Yitzhak Rabin (Bar-Ilan University, Israel) spoke on *Coupling of Twist and Writhe in Short DNA Rings*. While bending and twist can be treated as independent degrees of freedom for linear DNA molecules, the loop closure constraint introduces a coupling between these variables in circular DNA. Rabin performed Monte Carlo simulations of worm-like rods with both bending and twist rigidity, in order to study the coupling between the writhe and twist distributions for various DNA lengths. He found that for sufficiently short DNA, the writhe distribution differs significantly from that of a model with bending energy only and showed that the factorization approximation introduced by previous researchers coincides, within numerical accuracy, with his simulation results. Rabin concluded that the closure constraint is fully accounted for by the White-Fuller relation.

Hue Sun Chan (University of Toronto, Canada) spoke on *Selective Segment Passages at Hooked and Twisted Juxtapositions Consistently Rationalize the Decatenating, Unknotting and Supercoil-Tightening Actions of Type-2 Topoisomerases*. The mathematical basis of the hypothesis that type-2 topoisomerases recognize and act at specific DNA juxtapositions has been investigated by coarse-grained lattice polymer models, showing that selective segment passages at "hooked" juxtapositions can result in dramatic reductions in catenane and knot populations. The lattice modeling approach is now extended to account for the hallmark narrowing of variance of linking number ($Lk$) of DNA circles by type-2 topoisomerases. In general, the steady-state variance of $Lk$ resulting from selective segment passages at a specific juxtaposition geometry $j$ is inversely proportional to the average linking number, $\langle Lk \rangle_j$, of circles with the given juxtaposition. Based on this formulation, Chan demonstrated that selective segment passages at 'hooked' and 'twisted' juxtapositions reduces the variance of $Lk$. The dependence of this effect on model DNA circle size is remarkably similar to that observed experimentally for type-2 topoisomerases, which appear to be less capable in narrowing $Lk$ variance for small DNA circles than for larger DNA circles. This behavior is rationalized by a substantial cancellation of writhe in small circles with hook-like juxtapositions. For an extended

set of juxtapositions in their model, Chan's research group detects a significant correlation between the juxtapositions' supercoil simplification potential and their logarithmic decatenating potential as well as their logarithmic unknotting potential, a trend reminiscent of scaling relations between corresponding experimental measurements on type-2 topoisomerases from a variety of organisms. The consistent agreements between theory and experiment their group achieved argue strongly for type-2 topoisomerase action at hook- or twist-like DNA juxtapositions.

Michael Szafron (University of Saskatchewan, Canada) spoke on *Knotting Probabilities Resulting from a Local Strand Passage in a Knot-type K SAP*. Also motivated by understanding the action of type-2 topoisomerases on DNA, Szafron and Soteros have developed a self-avoiding polygon (SAP) lattice model to investigate the effect of random local strand passages on the knot-type of a ring polymer. For increasing SAP sizes, the limiting knot transition probability estimates obtained from Monte Carlo data for this model were presented. Evidence was provided that these limiting knot transition probabilities depend on the local juxtaposition at the strand passage site. This evidence provides further support for the hypothesis (mentioned above in the work of Chan's group) that selective segment passages according to the local juxtaposition geometry can reduce knot populations.

Dorothy Buck (Imperial College, England) spoke on *Topological Analysis of DNA Knotting and Unknotting*, joint work with Ken Baker and Andrew Lobb. Many protein-DNA interactions, such as site-specific recombination and (type II) topoisomerase-mediated unknotting and unlinking, act by cutting and resealing (double-stranded) DNA segments in a localized way. These enzymatic reactions can be modelled in terms of tangles, 3-dimensional balls with two properly embedded arcs, each representing a segment of DNA. The action of the protein can be thought of as removing one tangle and replacing it with another, e.g. a topoisomerase-initiated crossing change as replacing a (+1) tangle with a (-1) tangle, leaving the rest of the DNA unchanged. This replacement can be straightforward (as in the topoisomerase example above) or quite complex. Because of the plectonemic supercoiling of DNA, 'rational tangles' (formed by an alternating series of horizontal and vertical twists) are the most biologically relevant. Buck classified all possible rational tangles that can replace, in any prescribed manner, a given rational tangle, thus elucidating all possible protein mediated localized changes of DNA.

Jason Cantarella (University of Georgia, USA) spoke on *Intrinsic Symmetries of Knots and Links*. Given a link composed of several circular strands of DNA, each component is oriented and uniquely labeled by its sequence of base pairs. Can these components be reoriented? Can they switch places? The group of transformations of this type which can be realized by an isotopy of the link is called the "intrinsic" symmetry group of the link. Cantarella presented the first computations of the intrinsic symmetry groups of links with 8 and fewer crossings. The traditional definition of the symmetry group of a link is the mapping class group $MCG(S^3, L)$ of the pair $S^3$, $L$. The symmetry groups are the images of the traditional symmetry groups of links under the natural homomorphism from $MCG(S^3, L)$ onto $MCG(S^3) \times MCG(L)$.

Teresita Ramirez-Rosas (Grand Valley State University, USA) spoke on *Looking for a lower bound for the number of quadrisecants*. Ramirez-Rosas has been interested in finding a lower bound for the number of quadrisecants for a polygonal knot in general position in terms of its crossing number. Her immediate goal is to show the following:

*Conjecture: A knot $K$ with crossing number, $cr(K)$, has at least $\frac{1}{2}\left(\frac{2cr(K)+1}{3}\right)^2$ quadrisecants.*
Ramirez-Rosas discussed some ideas that might lead us to find a lower bound for the number of quadrisecants. In particular, she talked about one of her results that can help us to solve this conjecture: given $x \in K$ the number of trisecants with starting or ending point at $x$ is at least $\frac{2cr(K)+1}{3}$.

### 2.3 Confinement effects

In many practical situations of interest, macromolecules do not have full configurational freedom due to the constraints of geometric confinement, for example, when polymers are confined between two parallel planes as in models of steric stabilization of dispersions or DNA molecules contained in a capsid. Macromolecules so confined exhibit significantly different average and individual structure in comparison with those in free environments. Also, effective confining arises in the case of macromolecules that have specific hydrophobic and hydrophilic regions or when regions have restricted flexibility or torsion. While, in general, one might believe that great progress has occurred in understanding the storing, knotting, and winding of polymers, in fact rather little is known rigorously and many fundamental questions seem just beyond our grasp, both theoretically or via numerical studies. Further effort is clearly needed and promising steps are being taken in these areas.

The workshop contained five talks on this subject.

Javier Arsuaga (San Francisco State University, USA) spoke on *Modeling of Chromosome Intermingling by Partially Overlapping Uniform Random Polygons*, joint work with Yuanan Diao and Rob Scharein. During the early phase of the cell cycle the eukaryotic genome is organized into chromosome territories. The geometry of the interface between any two chromosomes remains a matter of debate and may have important functional consequences. The Interchromosomal Network model (introduced by Branco and Pombo) proposes that territories intermingle along their periphery. In order to partially quantify this concept Arsuaga's group investigated the probability that two chromosomes form an unsplittable link. They used the uniform random polygon (URP) as a crude model for chromosome territories and modeled the interchromosomal network as the common spatial region of two overlapping uniform random polygons. This simple model allows one to derive some rigorous mathematical results as well as to perform computer simulations easily. They found that the probability that a uniform random polygon of length $n$ partially overlaps a fixed polygon is bounded below by $1 - O(1/\sqrt{n})$. Arsuaga's group used numerical simulations to estimate the dependence of the linking probability of two uniform random polygons on the amount of overlapping. They propose that this dependence relation may be modeled as $\frac{1-ae+b(1-e)}{e\sqrt{mn}+b(1-e)}$ where $e > 0$. They used these results to model the data published by Branco and Pombo and observed that for the amount of overlapping observed experimentally the URPs have a non-zero probability of forming an unsplittable link.

Rob Scharein (Hypnagogic Software, Canada) spoke on *Bounds for the minimum step number of knots in the simple cubic lattice*, joint work with K. Ishihara, J. Arsuaga, Y. Diao, K. Shimokawa and M. Vazquez. Knots are found in DNA as well as in proteins, and they have been shown to be

good tools for structural analysis of these molecules. An important parameter to consider in the artificial construction of these molecules is the minimum number of monomers needed to make a knot. Scharein addressed this problem by characterizing, both analytically and numerically, the minimum length (also called minimum step number) needed to form a particular knot in the simple cubic lattice. His group's analytical work is based on improvement of a method introduced by Diao to enumerate conformations of a given knot type for a fixed length. This method allows one to extend the previously known result on the minimum step number of the trefoil knot $3_1$ (which is 24) to the knots $4_1$ and $5_1$ and show that the minimum step numbers for the $4_1$ and $5_1$ knots are 30 and 34, respectively. Using an independent method based on the BFACF algorithm, Scharein provided a complete list of numerical estimates (upper bounds) of the minimum step numbers for prime knots up to ten crossings, which are improvements over current published numerical results. They enumerated all minimum lattice knots of a given type and partitioned them into classes defined by BFACF type-0 moves.

Michael Schmid (Baylor College of Medicine, USA) spoke on *How can DNA get in and out of a virus capsid?* Double stranded DNA phages and viruses encapsidate their genome into a preformed capsid shell through one icosahedral vertex, which contains a portal protein complex. ATP is consumed, and the DNA is inserted, probably involving twisting. Extrusion of the DNA during cell or bacterial infection is accomplished through the same vertex. Schmid's lab (National Center for Macromolecular Imaging, Baylor College of Medicine) has determined the structure of several phages and viruses by cryoelectron microscopy (cryoEM). This technique aligns and averages thousands of individual 2D projection images in random orientations to produce a 3D reconstruction of the virus. Recently his lab has been able to perform this reconstruction without applying icosahedral symmetry, thus are able to see the unique vertex and the other non-icosahedral features. Clues as to the packing of the DNA include: 1) concentric shells of DNA spooled around the axis defined by the unique vertex, 2) a roughly hexagonal packing of the DNA helices against each other, 3) the terminus (last in) of the DNA runs up the axis toward the portal, among others. Many questions remain.

Cristian Micheletti (International School for Advanced Studies, Italy) spoke on *Coarse-grained simulations of DNA in confined geometries*. The packing of DNA inside bacteriophages arguably yields the simplest example of genome organisation in living organisms. An indirect indication of how DNA is packaged is provided by the detected spectrum of knots formed by DNA that is circularised inside the P4 viral capsid. The experimental results on the knot spectrum of the P4 DNA can be compared to results of coarse-grained simulation of DNA knotting in confined volumes. Micheletti started by considering a standard coarse-grained model for DNA which is known to be capable of reproducing the salient physical aspects of free (unconstrained) DNA. Specifically the model accounts for DNA bending rigidity and excluded volume interactions. By subjecting the model DNA molecules to spatial confinement it was found that confinement favours chiral knots over achiral ones, as found in the P4 experiments. However, no significant bias of torus over twist knots was found, contrary to what was found in P4 experiments. A good consistency with experiment can be found, instead, upon introducing an additional interaction potential accounting for the tendency of contacting DNA strands to order as in cholesteric liquid crystals. The degree of localization of the obtained knots was discussed in connection with the process of genome ejection out of the phage.

Alexander Grosberg (New York University, USA) spoke on *Large scale organization of DNA in chromosomes*. Recent experiments confirmed an old theoretical prediction that human genome (and presumably that of other eukaryotes) on the large scale (above the nucleosome size) is organized in the form of a crumpled fractal globule stabilized by the topological effects. Grosberg analyzed the application of the globule structure as a model for chromosome territories.

### 2.4 Length effects

The typical length of DNA in a cell ranges from thousands of base pairs in a virus, $\sim 4$ megabase pairs in bacteria, to $\sim 3$ billion base pairs in mammals or equivalently $\sim 10$ to 10 million Kuhn lengths. How does the length of DNA influence its topological and geometric properties such as knotting, linking and supercoiling? Is an organism's natural length of DNA optimal in terms of minimizing the possibility of topological obstructions to vital cellular processes such as replication and transcription while maximizing the amount of information that can be stored? In order to address this kind of question, theorists investigate the length dependence of the topological and geometric properties of model polymers. For lattice models of polymers, one can obtain mathematical proofs for the limiting behavior of, for example, knotting and linking probabilities as polymer length goes to infinity. Well established statistical mechanics and field theory arguments can also be used to predict the finite length scaling behavior of polymer properties such as the knotting probability or the average squared radius of gyration. Determining the length scale for which this scaling behavior is relevant, however, requires computer simulations and comparison to experiments. In general, much work remains on both the theory and experimental side in order to further bridge the gap. The mathematical facet of this work brings together topologists, geometers, statisticians, and computational scientists.

The workshop contained six talks on this subject.

Tetsuo Deguchi (Ochanomizu University, Japan) spoke on *Effective scaling approximations for knotting probability, topological swelling and the distance distribution of random knots*. Deguchi discussed various scaling approximations for the probability of random knotting and the mean square radius of gyration for random knots as functions of the number of segments. He also introduced an effective scaling formula for the distribution of the distances between two segments of polygon. For an illustration, consider knotting probability. For off-lattice models Deguchi numerically evaluated the probability of random knotting as a function of the number of nodes. He then found that two types of fitting formulas are quite effective, one for describing asymptotic behavior and another one for describing finite-size random knotting probability. Although the latter formula should be valid for a limited range of the number of nodes, it has a nice factorization property by which one can predict the probability of composite knots from those of the constituent prime knots. Deguchi's scaling approximations are particularly effective for finite-size random knots and should be fundamental in application to real ring polymers since all ring polymers have some finite number of segments. These results can be compared to experiments in the near future.

Bertrand Duplantier (Centre Energie Automique/Saclay, France) spoke on *Partition Function of a Freely-Jointed Chain in Half-Space*, joint work with Olivier Bernardi and Philippe Nadeau. When

lecturing about the Physics of Biological Polymers in 2007 at EPFL (Lausanne), Duplantier was asked by Andrzej Stasiak about the statistics of a discrete freely-jointed chain anchored at a plane in three space, and under traction by a force. This problem is relevant to the description of DNA under traction and of proteins in translocation across a membrane. Surprisingly, the calculation of the canonical partition function is non-trivial, and must be done via a functional recursion over the number of monomers. The enumeration of configurations also involves specific combinatorial aspects, which bring in cell decompositions, Motzkin paths and bijections to trees, a long way from the original biological question!

Stu Whittington (University of Toronto, Canada) spoke on *Pattern theorems: What we know and what we wish we knew.* Pattern theorems are a way to show that certain events occur with high probability, and were used to show that lattice polygons (a model of ring polymers) are knotted with high probability when the polygon is large. Over the last twenty years new ways to prove pattern theorems have emerged and pattern theorems have been proved for new situations. Whittington's talk reviewed the idea behind pattern theorems and showed how they can be used to prove results about topological and geometrical entanglement complexity. In spite of the progress made recently there are still many areas where a pattern theorem would be useful or where a sharper form of a pattern theorem would give improved results. Some of these open questions were discussed.

Mahshid Atapour (York University, Canada) spoke on *Exponential Growth of the Number of $n$-edge Linked Clusters in $\mathbb{Z}^3$ and the Consequences in Entanglement Percolation.* An animal in the simple cubic lattice is a finite connected subgraph of $\mathbb{Z}^3$. Let $a_n$ be the number (up to translation) of $n$-edge animals in $\mathbb{Z}^3$. In 1967, Klarner proved that $a_n$ grows exponentially. Let $e_n$ be the number (up to translation) of all $n$-edge linked clusters, i.e. subgraphs of $\mathbb{Z}^3$ in which the connected components (animals) are (topologically) non-splittable. Atapour explained how it can be proved that $e_n$ also has a finite exponential growth rate. She also mentioned some of the important consequences of this result in entanglement percolation.

Andrew Rechnitzer (University of British Columbia, Canada) spoke on *Counting knotted polygons (nearly).* The Rosenbluth method of simulating self-avoiding walks has become one of the standard methods for studying polymer statistics. The algorithm was originally developed in the 1950s by Hammersley & Morton and Rosenbluth & Rosenbluth, but suffered from poor convergence. This changed in the mid 90s with Grassberger's development of a pruned and enriched implementation called PERM. It was soon followed by multicanonical and flat-histogram implementations which have become indispensable tools for exploring the critical behaviour of polymer systems. Combinatorially, one can think of the Rosenbluth method as a technique of 'approximate enumeration' which produces precise estimates of the number of conformations of a particular size and energy. This same method can then be applied to many other combinatorial problems provided there is a unique and unambiguous way of constructing the underlying objects. Unfortunately, self-avoiding polygons (SAPs) are not such a system. Rechnitzer discussed this history and described two recent extensions of the original Rosenbluth algorithm which allow the approximate enumeration of two-dimensional SAPs and SAPs of fixed knot type in three dimensions in joint work with Buks van Rensburg.

Buks Janse Van Rensburg (York University, Canada) *Statistics of knotted lattice polygons.* Polygons in the cubic lattice are simple closed curves in three space and have well-defined knot types. The

number of lattice polygons of length $n$ and knot type $K$ in the cubic lattice is $p_n(K)$, where we consider two polygons to be equivalent under translations in the lattice. For example, if $K$ is the unknot $\emptyset$, then $p_4(\emptyset) = 3$, $p_6(\emptyset) = 22$, $p_8(\emptyset) = 207$ and so on. Determining $p_n(K)$ for arbitrary $n$ and knot types $K$ is a difficult numerical problem, but the GAS-algorithm can be used for approximate enumeration of $p_n(K)$. Janse van Rensburg presented the results of simulations resulting from collaborations with Rechnitzer to estimate the approximate values of $p_n(K)$ for some knot types $K$. The scaling of $p_n(K)$ was discussed, and evidence presented that $p_n(K) \sim A_K n^{\alpha-2+N_K} \mu_\emptyset^n$; where $N_K$ is the number of prime components in the knot type $K$ and $\mu_\emptyset^n$ is the growth constant for unknotted polygons. The relative frequencies of knot types in lattice polygons were also discussed.

### 2.5 DNA sequence effects and replication factories

The so-called "base-pair step parameters" provide remarkable predictive powers with regards to the conformation of a DNA polymer. Next approaches should start to include not only nearest neighbor effects, but even next nearest neighbor effects. How to model this mathematically and computationally is an enormous yet exciting new challenge. The DNA sequence, of course, dictates both the structural deformations that occur as a consequence of underwinding and overwinding DNA, as well as the electrostatics. In addition, the DNA sequence, as well as $Lk$, counterions, and water, all come into play in the formation of the so-called "alternative secondary structure of DNA". Researchers have made great inroads into the understanding of these structures and how important they are for DNA. Medically, the structures that result can cause human suffering and account for the cause of several important and fairly common human diseases.

Instead of free, unconstrained DNA filling up space in a cell, in fact the proteins that replicate and transcribe DNA are "fixed" in the cell in what biologists have named "factories". During replication, for example, this means that the DNA moves, at a rate of 100-1000 base pairs/second. In front of the factories, the DNA will have to be transiently overwound and this overwinding is unlikely to be allowed to adopt the geometric configuration of writhe. White's adaptation to DNA of $Lk = Tw + Wr$, therefore, must be now modeled to limit writhe and mathematical considerations of variations in twist and writhe should aid in the understanding of this important biological phenomenon. The single double helix train track in front of the factory, during replication has, behind the factory, split into two train tracks with partial gaps on one side and a nick on the other. The topological interplay between linking number and catenation is likely to be governed by mathematical and physical principals. Thus replicating DNA involves extraordinary structures with tremendous topological complications, and this is a mechanism desperately in need for improved mathematical modeling. At the same time, it is the long-established concepts in DNA topology and knot theory that have helped guide the understanding of this remarkable biopolymer. The mathematics involved includes tangle, braid, knot, link, and polymer modeling. The study of the characteristics of both equilibrium as well as kinetic aspects of DNA now include geometric, spatial, and topological facets that may be implicated in these mechanisms as well as the characteristics of polymers under a variety of solvent conditions. While these studies require advances in computational methods to fully illuminate the equilibrium properties, sufficient information appears already to be available to inform an understanding of experimental observations.

The workshop contained six talks on this subject.

Wilma Olson (Rutgers University, USA) spoke on *Protein-mediated DNA looping and gene expression.* Making sense of gene regulation in living systems requires understanding of the looping properties of DNA in crowded, multi-component systems. The presence of non-specific binding proteins that introduce sharp bends, localized untwisting, and/or dislocation of the DNA double-helical axis, stabilizes functional repression loops ranging from as few as 65 base pairs to as many as tens of thousands of base pairs. As a first step in the analysis of such looping, Olson's group investigated the effects of various proteins on the configurational properties of fragments of DNA, treating the DNA at the level of base-pair steps and incorporating the known effects of various proteins on DNA double-helical structure. The presentation highlighted some of the new models and computational techniques that her group has developed to generate the three-dimensional configurations of protein-mediated DNA loops and illustrate new insights gained from their work about the effects of various proteins on DNA topology and the apparent contributions of the non-specific binding proteins to gene expression.

Phoebe Rice (University of Chicago, USA) spoke on *Structural model for how Sin recombinase exploits topology.* Sin is a DNA recombinase belonging to the serine resolvase family. For various biological reasons, these enzymes convert one large DNA circle into two smaller ones. To prevent other recombination products, the system is regulated by a 'topological filter' - it is only enzymatically active when the two partner sites are brought together in a synaptic complex containing three interdomainal supercoiling nodes. Using crystal structures of individual components, Rice's group constructed a 3-dimensional model for this synaptic complex. They also used biochemical assays to address the details of how this catalytic regulation is enforced. Rice presented preliminary data showing that a different serine resolvase, Tn3, uses different protein-DNA interactions to construct a regulatory complex with the same DNA topology.

Georgi Muskhelishivili (Jacobs University, Germany) spoke on *General organisational principles of the transcriptional regulation system: a tree or a circle?* The fundamental problem in attempting a holistic description of the transcriptional regulation system is of a methodological nature and lies in the necessity of integrating the systemic and structural-molecular views. Recent advances of systemic approaches to gene expression provide unforeseen opportunities for relating extensive datasets describing the transcriptional regulation system as a whole. However, due to the multifaceted nature of the phenomenon, these datasets often contain logically distinct types of information determined by the underlying approach and adopted methodology of data analysis. Consequently, to integrate the datasets comprising information on the states of chromatin structure, transcriptional regulatory network and cellular metabolism, a novel methodology enabling interconversion of logically distinct types of information is required. Muskhelishivili provided a holistic conceptual framework for the analysis of the global transcriptional regulation as a system coordinated by the structural coupling between the transcription machinery and DNA topology, acting as interdependent sensors and determinants of metabolic functions. In this operationally closed system any transition in physiological state represents an emergent property determined by shifts in structural coupling, whereas genetic regulation acts as a genuine device converting one logical type of information into the other.

Jue D. Wang (Baylor College of Medicine, USA) spoke on *Co-orientation of Replication and Tran-*

*scription Preserves Genome Integrity.* In many bacteria, there is a genome-wide bias towards co-orientation of replication and transcription, with essential and/or highly expressed genes further enriched co-directionally. Wang's group previously found that reversing this bias in the bacterium Bacillus subtilis slows replication elongation, and proposed that this effect contributes to the evolutionary pressure selecting the transcription-replication co-orientation bias. This selection might have been based purely on selection for speedy replication; alternatively, the slowed replication might actually represent an average of individual replication-disruption events, each of which is counter-selected independently because genome integrity is selected. To differentiate these possibilities and define the precise forces driving this aspect of genome organization, Wang's group generated new strains with inversions either over 1/4 of the chromosome or at ribosomal RNA (rRNA) operons. Applying mathematical analysis to genomic microarray snapshots, they found that replication rates vary dramatically within the inverted genome. Replication is moderately impeded throughout the inverted region, which results in small but significant competitive disadvantage in minimal medium. Importantly, replication is strongly obstructed at inverted rRNA loci in rich medium. This obstruction results in disruption of DNA replication, activation of DNA damage response, loss of genomic integrity and cell death. Wang's results strongly suggest that preservation of genome integrity drives the evolution of co-orientation of replication and transcription, a conserved feature of genome organization.

Tim Hughes (University of Toronto, Canada) spoke on *High nucleosome occupancy is encoded at human regulatory sequences.* Active eukaryotic regulatory sites are characterized by open chromatin, and yeast promoters and transcription factor binding sites (TFBSs) typically have low intrinsic nucleosome occupancy, i.e. these sequences are disfavored when naked DNA and histone octamers are assembled *in vitro.* Hughes showed that in contrast to yeast, DNA at human promoters, enhancers, and TFBSs generally encodes high intrinsic nucleosome occupancy. In most cases his group examined, these elements also have high experimentally measured nucleosome occupancy *in vivo.* These regions typically have high G+C content, which correlates positively with intrinsic nucleosome occupancy, presumably due to high bend, twist, tip etc. parameters, as well as reduced probability of rigid, nucleosome-excluding polyA-like sequences. Hughes proposed that high nucleosome affinity is directly encoded at regulatory sequences in the human genome to restrict access to regulatory information that will ultimately be utilized in only a subset of differentiated cells. He also proposed that nucleosomes may play direct roles in the function of active enhancers. Their findings also present a functional consequence of variation in base content that is observed at diverse scales in eukaryotic genomes.

Wei Yang (NIH, USA) spoke on *Lessons learnt from a DNA helicase UvrD.* How do molecular motors convert chemical energy to mechanical work? Helicases and nucleic acids offer simple motor systems for extensive biochemical and biophysical analyses. Atomic resolution structures of UvrD-like helicases complexed with DNA in the presence of AMPPNP, ADPPi, and Pi reveal several salient points that aid understanding mechano-chemical coupling. Each ATPase cycle causes two motor-domains to rotationally close and open. At a minimum, two motor-track contact points of alternating tight and loose attachment convert domain rotations to uni-directional movement. A motor is poised for action only when fully in contact with its track and, if applicable, working against a load. The orientation of domain rotation relative to the track determines whether the movement is linear, spiral or circular. Motors powered by ATPases likely deliver each power stroke in two parts, before and after ATP hydrolysis.

# 3 Conclusion

The workshop was designed to intentionally maximize relaxed interactions among the diverse participants. On the last day of the workshop, we held a short meeting to get feedback on the format and the participants were asked what they liked, what they might change, whether or not they learned anything new, and whether they will start new collaborations from the meeting.

Some of the comments about the meeting included:

"I talked to several biologists in a more detailed way than at many meetings."

"I have attended quite a few conferences covering material that lies at the interface between mathematics and other sciences; this conference was far and away the one with the most communication between fields. The talks were very accessible and it was clear that ideas were genuinely flowing between disciplines."

"I think it was very successful in general and inspired me personally to think about applications of maths to genetic regulation."

"The meeting had the right balance between the various communities and, in my view, all speakers were able to communicate effectively their research to an audience with a very mixed background."

In the weeks that have followed since our meeting, we the organizers continue to receive notice of one of our speakers visiting the country or laboratory of another, interactions that might never have taken place without the extraordinary opportunity afforded by BIRS.