Mariel Vazquez

Research Statement

I am a mathematical biologist specialized in the applications of topology to the study of DNA. In my research I collaborate closely with experimental biologists to ensure that the problems approached and the solutions proposed remain biologically relevant. I use my background in pure mathematics to study DNA topology and DNA rearrangements using analytical (topology, knots, tangles and graphs) and computational (Monte Carlo simulations, computer visualization) methods. During my PhD, I used modern tools in knot theory and low-dimensional topology to study two site-specific recombination systems: Gin of bacteriophage Mu and Xer of *Escherichia coli* [4,10,13][1]. In August 2000 I became a postdoc in Rainer Sachs' group and a Visiting Assistant Professor at UC Berkeley (Math Dept). In Professor Sachs' group I used graph theory, statistical methods and Monte Carlo computer simulations to study genomic rearrangements found in radiation data [5-8,11,12,14]. I also extended the DNA topology work from my dissertation to a more general study of site-specific recombination and started a novel study of type II topoisomerases, which led to my mentoring several undergraduates in research. At the same time, jointly with J. Arsuaga I studied DNA packing in bacteriophage capsids [2,3,9,15]. Since 2000, I have mentored 16 UC Berkeley undergraduates on the study of DNA knots, site-specific recombinases and type II topoisomerases, and have co-authored research papers with some of them [4,17,18]. Jointly with J. Arsuaga and R. Sachs, I applied for an NSF grant in Spring 2002, which fully funded my last three years at Berkeley. In August 2005, I joined the math faculty at San Francisco State University (SFSU). Since 2005 I have tirelessly applied for both internal and external funding. In collaboration with my co-investigators I have brought over $1.7 million to the university in federal research grants, and I have received seed funding from a variety of sources. In 2011, I received an NSF CAREER Award, and in 2012 a Presidential Early Career Award for Scientists and Engineers (PECASE). The CAREER award provides funding to study the process of DNA unlinking by Xer recombination.

My current research interests are in DNA topology and in chromosomal aberrations as described in the following two sections.

I. DNA topology: DNA unknotting, DNA unlinking and chromosome architecture

DNA topology refers to supercoiling, knotting and linking of circular DNA molecules (*e.g.* bacterial chromosomes and naturally occurring plasmids; chloroplast DNA; human mitochondrial DNA).

[1] All references cited in this document correspond those in the list of publications and the CV.

Changes in DNA topology are effected by DNA packing and condensation, as well as by a variety of cellular processes such as DNA replication, topoisomerase activity and recombination. Understanding the mechanisms by which these processes change the topology of DNA is biologically important. Mathematical and computational methods have proven to be invaluable for addressing these problems. In particular, 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 mechanism of enzymes that change the topology of DNA, such as site-specific recombinases and type II topoisomerases. I have extensively studied the action of such enzymes [4,10,13,16-18,20,21-22,25,27], as well as the packing of DNA in confined volumes [2,3,9,15-16,21,28]. From these studies have also resulted a series of theoretical papers on properties of knotted and linked polygons [19,23,26,29]. I use state-of-the-art techniques in mathematical and computational knot theory, low-dimensional topology, and Monte Carlo computer simulations.

a. Topological consequences of DNA replication

Replication of a circular chromosome requires unwinding of the DNA and results in the formation of DNA links where two newly replicated sister chromosomes are interlinked and cannot be separated without double-stranded chain cleavage. Error-free unlinking is required to minimize mutagenesis and to ensure proper segregation at cell division. Characterizing the topological mechanism of DNA unlinking is key to understanding the processes of circular chromosome replication, recombination and segregation at cell division. In my group we study the following aspects of this problem.

DNA unlinking by XerCD-FtsK (NSF CAREER award, May 2011-April 2016)

Approximately one in every eight generations, two newly replicated E. coli chromosomes form a single chromosome dimer. The cell is able to resolve this problem (*i.e.* dimer resolution) prior to cell division by means of two site-specific recombinase enzymes called XerC and XerD and the powerful translocase FtsK, which act at the division septum. My collaborator D. Sherratt (Oxford University) has shown experimentally that when the Xer enzymes act on a plasmid containing two Xer binding sites (*e.g. psi*-sites), they produce a 4-crossing torus link. In my previous work I used the tangle method and a theorem on Dehn surgeries on strongly invertible knots (Hirasawa and Shimokawa, 2000) to prove that there are only three possible mechanisms of action for XerCD at *psi*, and to characterize them [13]. Jointly with the Sherratt lab we recently showed that when acting at the chromosomal *dif* sites, XerCD and the translocase Ftsk (XerCD-Ftsk) are able to mediate sister chromosome unlinking in TopoIV deficient cells [20]. TopoIV is one of the type II topoisomerases in E. coli and is largely responsible for unlinking replication

links. A clear picture of the *in vivo* mechanism of DNA unlinking by XerCD in *E.coli* is not yet available. It is known that FtsK activates XerCD recombination by co-localizing with these enzymes at the *dif* sites. XerCD-FtsK unlinking experiments suggest a stepwise mechanism of action (Fig. 1).

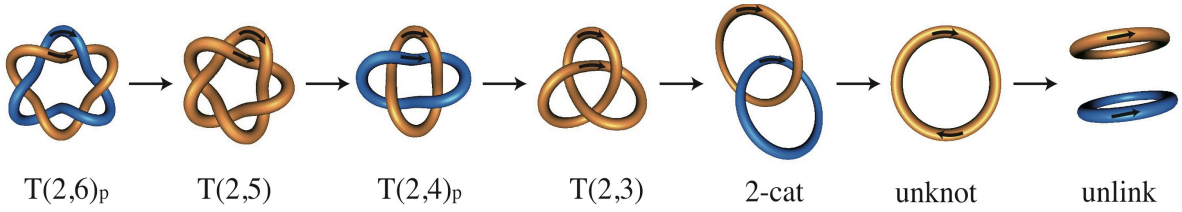

Figure 1: Proposed shortest pathway of DNA unlinking by Xer-FtsK-*dif*.

My long-term research goal in this project is to determine the topological mechanism of action of the XerCD site-specific recombination system in *E.coli.* My first objective is to characterize the topological mechanism of DNA topology simplification by the XerCD-FtsK system using knot theory, low-dimensional topology, computer simulations and visualization. In pursuit of the research objective, the central hypothesis is that, *after activation by FtsK at the division septum, XerCD unlinks DNA links in a stepwise manner*. I will test this hypothesis by first using the tangle method to find possible topological pathways of DNA unknotting and unlinking by site-specific recombination on small DNA knots and links, and second by performing computer simulations of DNA unlinking by site-specific recombination.

The work outlined here is done in close collaboration with experimental biologists David J. Sherratt (Iveagh Professor of Microbiology, Oxford University, UK) and Ian Grainge (U. of Newcastle, Australia), with polymer physics experts Christine Soteros and Michael Szafron (U of Saskatchewan, Canada), and with low-dimensional topologists Koya Shimokawa (Saitama University, Japan) and Kai Ishihara (Yamaguchi University, Japan). This collaborative work combines experimental, analytical and computational results to discriminate between pathways for Xer recombination, identify probable pathways, and determine deviations from randomness. For example, using recent results on band surgery and polynomial invariants (Kawauchi 2009, Kanenobu 2009) we show that there is a unique shortest pathway between any T(2,2m) link and the unlink. To study the mechanisms of action at each step of the pathway, we use tangle calculus and the characterization of band surgeries from the unknot to the unlink (Scharlemann 1989), from the trefoil to 2-cat (Darcy *et al,* 2011), and from 2-cat to the unknot (Bleiler-Litherland, 1989; Hirasawa-Shimokawa 2000). The uniqueness relies on the assumption that recombination reduces the complexity of its substrates at each step. If this assumption is relaxed, other pathways arise. We have developed a Monte Carlo method (Recombo) to simulate site-specific recombination *in silico*, and are using it to assign weights to individual recombination steps, and thus give

further evidence that DNA unlinking by XerCD-Ftsk acts by a stepwise pathway as indicated in Figure 1. We expect to submit two papers for publication before the end of 2012 reporting on this work.

Mechanism of topology simplification by type II topoisomerases (NIH SCORE; Jan 2007-Dec 2011; M. Vazquez, PI.).

Type II topoisomerases are essential to every living organism and are targets of numerous anti-bacterial and anti-cancer drugs. Their main cellular role is to modulate DNA supercoiling and to eliminate undesired DNA entanglement, such as knots produced by random strand-exchange and links produced by replication of circular chromosomes. The local action of Type II topoisomerases has been described: they mediate a passage of two double-stranded DNA segments through each other by creating a transient double-strand break in one of the segments. However random occurences of this simple mechanism cannot explain a number of experimental observations. In 1997 the Cozzarelli lab reported that type II topoisomerases can simplify DNA topology below thermal equilibrium values. A number of models have been proposed to explain this result, but no model has provided a detailed answer.

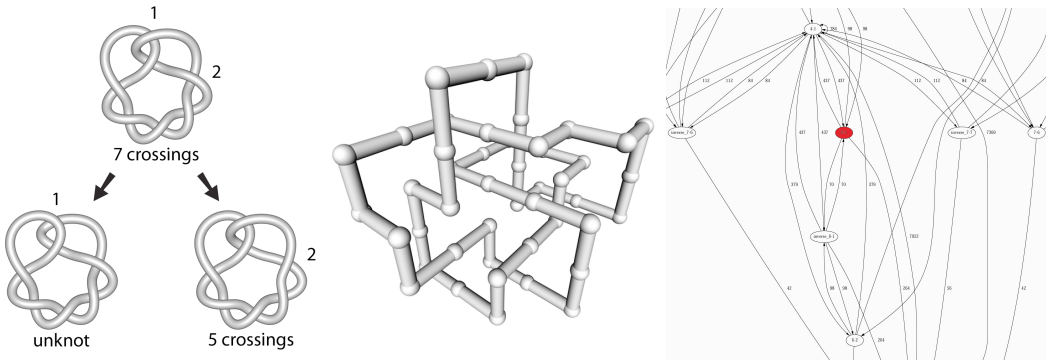

Figure 2. The image on the left illustrates how strand-passage on crossing 1 unknots the 7-crossing twist knot (7_2) in one step, while selection of crossing 2 yields a 5-crossing knot (5_2). The center image shows and embedding of the 7_2 knot in the simple cubic lattice Z^3. The figure on the right shows a portion of a network of interactions between knot types as determined by single strand-passages. The red node is the unknot.

We have focused on the process of DNA unknotting by type II topoisomerases. Our hypothesis is that *local geometrical features of knotted DNA can guide strand-passage by type-2 topoisomerases*. To test this hypothesis we have first designed Monte Carlo methods that sample the space of all configurations of polygonal chains of given length and knot type using different continuous and lattice models. Second we have generated knot distributions using an unbiased (random) strand-passage simulation assuming that any crossing has equal probability of being acted-upon [17]. Our results agree with those presented by others and show that random unknotting produces many more knots than observed experimentally by unknotting by type II topoisomerases. How do type II topoisomerases achieve the observed reduction in

knotting? This question has led us to propose various mechanisms of binding where a topological bias guided by bending properties of DNA or by local chirality driving the knotting below the expected equilibrium values (Fig. 2) [27; 3 papers in preparation]. From this study stemmed a few theoretical studies of random knots in the simple cubic lattice [23,26-27,29].

Mainly undergraduates have conducted this study. Most of them are Berkeley students enrolled in the URAP program. Until now, most of our effort has gone to developing robust computational tools to sample ensembles of polygons and to target different topologies for strand-passage. Our future plans include the implementation of a model developed by our collaborator J. Roca (CID-CSIC Barcelona) that considers the interplay of three different strands at the time of unknotting. I am currently working with URAP students on a 3D strand-passage model, which discriminates between supercoiled and clasped regions of a knot (Fig. 2).

b. Difference topology and 3-string tangles

In [22] we developed new topological methods for analyzing difference topology experiments involving 3-string tangles. This work is a perfect example on how a novel biological question inspires new mathematics, which then feed back to the biology.

Difference topology is an experimental technique used to study the degree of entanglement of the DNA in a protein-DNA complex. In this technique, circular DNA is incubated with the protein(s) under study. A site-specific recombinase of known mechanism (*e.g.* Cre) is then added to the reaction. DNA substrates are designed to contain sites specific to the chosen recombinase (*e.g. loxP*) at carefully selected locations along the DNA. Site-specific recombination on each of these substrates results in changes in their DNA topology. The DNA conformation in the original protein-DNA complex has a direct impact on the topology of each recombination product. Determining the knot/link type of the products gives information on the pre-recombinant DNA-protein conformation. In their paper, Pathania, Jayaram and Harshey (2002) used difference topology experiments to propose a shape for the DNA bound within the Mu transposase protein complex. They studied the recombination products under the assumption that the structure was branched supercoiled, and proposed a unique structure for the complex (Fig. 3 left and center).

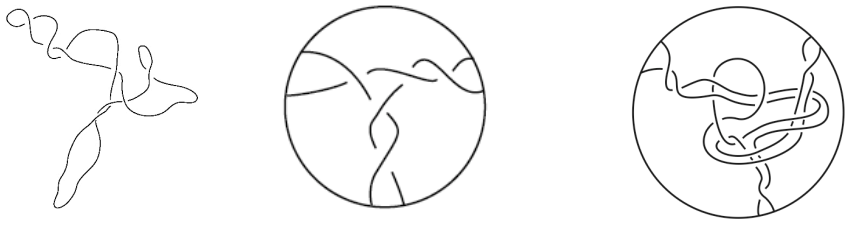

Figure 3. Branched supercoiled circle (left); PJH solution (center); another solution tangle (right).

In [22], in collaboration with I. Darcy (Math, U. Iowa), and J. Luecke (Math, UT Austin), we removed the branched supercoiled assumption and provided a rigorous mathematical analysis of the experimental data. We model the DNA-protein complex determined by the Mu transposase bound by the DNA as a 3-string tangle. We translate each difference topology experiment into a system of 3-string tangle equations. The challenge here is that, while 2-string tangles have been extensively studied in the context of site-specific recombination, a classification of rational 3-string tangles in unavailable, and so is 3-string tangle calculus. Particularly challenging is the problem of determining rationality of a three-string tangle. In [22] we use knotted graphs to characterize the solutions to the equations. We exhibit infinite families of solutions fitting the experimental data and argue that there is a unique 3-string tangle with at most 9 crossings satisfying the Mu tangle equations (call it PJH). The PJH tangle is rational, and it coincides with the solution proposed by the biologists (Fig. 3 center). Further, we show that all solutions to the mathematical problem, other than PJH, are too complex to be biologically reasonable (e.g. Fig. 3 right). We also provide information on the minimum number of experiments needed to reach the same conclusion. This work can easily be extended to other protein-DNA complexes involving 3-string tangles, and is amenable to computer implementation. I. Darcy and SJ Kim extended the work to 4-string tangles. This work also inspired new mathematical results on the planarity of tetrahedral graphs.

c. Chromosome architecture

Topological characterization of DNA organization in bacteriophages (NSF Math Biology, Sept 2009-Aug 2013; co-PIs: J. Arsuaga, Y. Diao).

DNA extracted from bacteriophage P4 is knotted with high probability and the knot distribution is very complex [3]. In 1985, Liu and Wang proposed to use these knots as a general assay for unknotting by type II topoisomerases. However, at that time the knot identification techniques were very limited and the distribution of knots was unknown. We are extending this assay to better characterize the action of type-2 topoisomerases using the new information about P4 knots obtained by our group and others [15], as well as a new cosmid system that allows packing of DNA's (as short as 5Kb). In particular we are using the experimental system developed by our collaborator J. Roca to produce knot distributions of 5Kb cosmids and perform one-step unknotting by type II topoisomerases. We have developed computer simulations that model changes in a population of knots when treated with type II topoisomerase (section 1.a). The experiments will help validate our theoretical approach. Furthermore we postulate that stepwise unknotting of these P4 knots will shed light on the original knot distribution and on the underlying architecture of the packaged DNA. The work outlined above is done in collaboration with J. Arsuaga, Y. Diao (UNC Charlotte), J. Roca (CID-CSIC, Spain). Analytical, numerical and experimental results of this work are reported in [2,3,9,15,16,19,21,28]

Reconstructions of three dimensional genome architecture from chromatin conformation capture data (Joint DMS-NIGMS initiative, pending; co-PIs J. Arsuaga and M. Segal)

The three dimensional (3D) architecture of eukaryotic chromatin is widely acknowledged to play critical roles in nuclear and cellular function. There is growing recognition that gene regulation and cancer-driving gene fusions are influenced by 3D organization of chromosomes. Until recently, our understanding of chromatin structure has been limited by constraints on the direct observation of highly-condensed material at the genomic level. New high-resolution molecular techniques are changing this situation. For example, novel genome-scale assays such as chromatin conformation capture (CCC) now permit the elicitation of data on chromatin contacts, creating unprecedented opportunities for studying chromatin organization and exploring its influence on various biological processes. Most analyses of CCC data to date have focused on the one-dimensional (1D) contacts level. More effort is needed to develop 3D reconstructions, evaluate their accuracy and reproducibility, and apply these reconstructions to the analysis of biological processes. Our hypothesis is that chromatin contact data can be reliably used to determine 3D genome structures and to assess their downstream impact on biological function. To test this hypothesis, we will develop new reconstruction algorithms and refine existing ones. We will undertake a systematic evaluation of their performance and will investigate the reproducibility of the obtained reconstructions under perturbations using computational and statistical tools.

II. Analysis of chromosome aberrations: Graph theory and Computational Homology (NSF grant, 2002-2004, PI R.K. Sachs and; NIH RIMI, 2008- 2013, PI B. Macher, Project leaders J. Arsuaga, M. Vazquez)

During my postdoc at UC Berkeley and under the direction of Prof. R. Sachs I worked on the analysis or radiation induced chromosome aberrations. When radiation tracks cross the cell nucleus they introduce double-strand breaks (DSBs). If left unrepaired, DSBs may produce undesired chromosomal rearrangements (*aberrations*) that may lead to cancer, or drive cell death. Understanding the mechanisms of DSB repair is a problem of utmost biological importance. We approached this problem by analyzing radiation-induced chromosomal aberrations statistically and by describing their mechanisms of production with biophysical models and Monte Carlo computer simulations [5,8,12]. We also developed a graph-theoretical characterization of chromosome rearrangements [6,12].

Copy number changes (i.e. amplifications and deletions) are another type of chromosome aberrations commonly observed in cancer. In 2008, J. Arsuaga and I were funded as co-leaders of subproject of an NIH RIMI grant (PIs Macher/Corrigan). In this project we proposed to use methods of persistence

homology to detect copy number changes in breast cancer. One reason why copy number changes are important is because they can affect oncogenes and tumor suppressor genes. Their presence along the genome can be detected using high-throughput techniques such as Comparative Genomic Hybridization (CGH) arrays. The algorithm, initially developed by J. Arsuaga, assigns an n-dimensional surface to each CGH profile (*i.e.* patient) and performs an association study between the topological properties of the network and any phenotype of the CGH profile. Such association has allowed us to identify regions of the genome that are commonly found in a given subpopulation of patients (*i.e.* recurrent *vs* non-recurrent) [24] and has led us to propose new aberrations in the recurrent population that we intend to further investigate.

One of the drawbacks of the method is that some characteristics of the aberration are lost when the CGH profile is mapped to a cloud of points. We are currently developing a method to further analyze the complexity of the aberrations. In this method we connect those points in the cloud that are consecutive along the genome. Our preliminary results show that the geometry of these curves is a reflection of the size of the aberration being analyzed (Fig. 4).

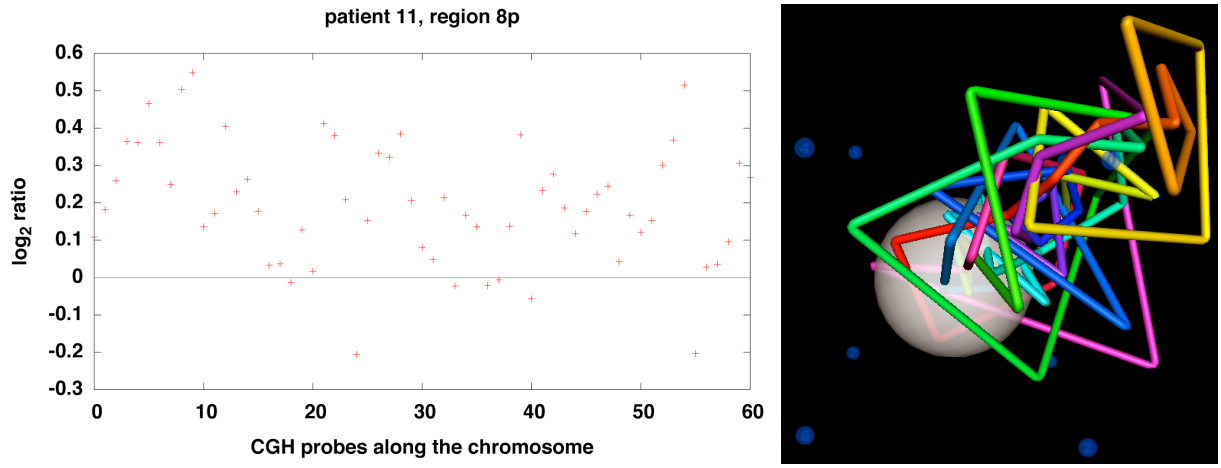

Figure 4: The graph on the left illustrates the raw CGH data for the p arm of the 8th chromosome of patient 11 (Climent et al. Cancer Res. 67 (2007)). The *y*-axis represents the *log*2 ratio of tumor DNA to reference DNA clone copy number. A sliding widow of length 3 generates coordinates for a cloud of points in R^3. The figure on the right illustrates a polygonal chain connecting points in the 3-dimensional cloud of points. The connectivity of the points illustrates the length and position of several amplifications.

This work is done in collaboration with Dr. J Arsuaga (SFSU) and Dr. C. Park (UCSF).