The Geometry and Topology of Knotting and Entanglement in Proteins (17w5032)

Mario Eudave-Munoz (Universidad Nacional Autonoma de Mexico),
Dorothy Buck (University of Bath),
Kenneth C Millett (University of California, Santa Barbara),
Eric Rawdon (University of St. Thomas),
Joanna I. Sulkowska (University of Warsaw)

5 November – 10 November 2017

1 Overview of the Workshop

Researchers from theoretical biology, experimental biology, biophysics, geometry, and topology focused their interactions on understanding the consequences of the spatial structure of proteins and nucleic acids, such as DNA, and on how living organisms work. For example, we know that some proteins contain knots while most others do not. How and why did this happen? Was it an accidental mutation that persists (when it need not) or is there a reason why this knotting or other forms of entanglement are necessary? How do enzymes that change the spatial structure of nucleic acids work? How might one devise new enzymes to slow and stop various forms of cancer or alter their toxicity? These grand questions and many others benefited from the cross-disciplinary perspectives, theories, methods, and experience present in this workshop.

Mathematical researchers reported on new insights from symplectic geometry, 3-dimensional geometric topology, as well as theories of knotting, linking, and other forms of entanglement. Biologists and biophysicists shared new biological insights reflected in protein and enzyme experiments, and on the identification, classification, and connection of entanglement for biological structures. New tools such as the databases KnotProt, LinkProt, and LassoProt were presented to the group as well as the results of molecular biophysical experiments. Biomedical presentations brought still another perspective to the discussions. Together, these researchers brought a powerful array of knowledge and experience with which they tackled challenging questions and initiated explorations of potential research collaborations. These new collaborations will lead to important progress on these most critical questions and give direction to new areas of future research.

The organization of the workshop was guided by three goals: first, the exploration of new mathematical methods being used within the context of new experimental methods and data concerning knotting, linking, and other types of entanglement found in proteins or occurring in DNA; second, inspiring research into new mathematical and biological questions based on this recent progress; and third, insuring that all advanced graduate students and post-docs had an opportunity to showcase their research. As a consequence, the workshop provided an ideal environment for these graduate students and post-docs to meet and, together, interact with more senior researchers across a wide variety of fields. Together, they established a common vocabulary for discussing outstanding problems through these multidisciplinary research interactions.

2 Recent Developments and Open Problems

The workshop presentations reported on recent progress in, roughly, seven distinct interacting domains organized around the workshop themes of knotting, linking, and other manifestations of structural entanglement.

From geometry and topology, there are several new contributions that are especially fruitful. First, the recent exploitation of the symplectic geometric structure of the spaces of open and closed polygons have provided methods that enable one to randomly sample "random knots," and to devise rigorous proofs of theorems describing the averages of fundamental measures of open and closed polygons important in their use as course grained models of macromolecules. These random knots are used by researchers to explore the structure of entangled macromolecular chains *in silico*. Acquiring large random samples requires the development of highly efficient algorithms leading to some very attractive new approaches. For example, the concept of orthoschemes accelerates the data generation process when applied to the action coordinates of a symplectic structure. There has also been much interest, recently, in the use of special knot diagrams as a vehicle for studying another notion of "random knots," in particular by means of random selecting from among these classes of diagrams. One example, the classical structure of a braid, has been yielding recent new results. From another perspective, knot and link diagrams encorporate powerful structural properties that allow them to be used in the study of the action of topoisomerases on DNA or the consequences of cystine bridges on the structure and properties of proteins. Slipknots, i.e. knotted arc lying within unknotted arcs, have been detected in some protein structures. They are now also foci of investigations in knot spaces and in knot diagrams. Furthermore, because knots and slipknots can occur in a biological context in which severe spatial constrains occur, recent research continues to explore course grained models subject to such spatial constraints.

Catalyzing the theoretical and methodological developments are the recent experimental discoveries shedding light on the knotted and slipknotted structure of proteins, the role of chaperones, and folding pathways. While there remain facets about which some researchers have differing views, there is a rough agreement sufficient to inform biomedical applications such as the creation of new drugs or explorations to inform the design of new treatment strategies.

Similarly, the study and understanding of chromosome regions, topological and geometric aspects of their spatial structure, and their folding has seen important recent progress.

With recent theoretical mathematical progress, one sees new developments in the understanding of biological and physical polymers that include the synthesis of more complex graphical structures as well as simulation studies of their static and dynamic properties. These simulations have provided new insight to the viseoelastic properties of collections of polymers, highlighted the role of composite knots, and begun the exploration of spatial properties of θ-graphs, and the sedimentation of elastic knots. With the demonstration of the first deep knot in a protein in 2000, the richness and significant presence of entangled structures has enriched the understanding of entangled macromolecular structures with knots, slipknots, lassos, and links (See Figure 1) inspiring and inspired by mathematics.

3 Presentation Highlights

A consensus view of the workshop participants is that the workshop had assembled a diverse array of speakers who, collectively, covered fundamental current research questions, reported on recent advances, described promising new methods or perspectives, and discussed discoveries in contiguous fields of research that might inspire future directions of research.

DNA structure was an important central feature of several presentations (e.g. from presentations by Dorothy Buck, De Witt Sumners, and Lynn Zechiedrich). These talks connected historical roots of knotting to current day research. Notably, the study of "minicircles" and their shapes has led to potential biomedical applications for drug delivery, in particular for some forms of cancer. That work harmonized with the biomedical drug design issues discussed by Claudia Benitez-Cardoza and Ellinor Haglund. In these latter presentations, we encountered "lassos," which were also featured in the work presented by Dabrowski-Tumanski, Niemyska, and Sulkowska. This new fundamental research was echoed in a number of other talks providing a new research stimulus. Recent efforts to undersand chromosones was the subject of the lectures of Arsuaga and of Onuchic, in which they discussed new efforts to analyze the large collections of data characteristic

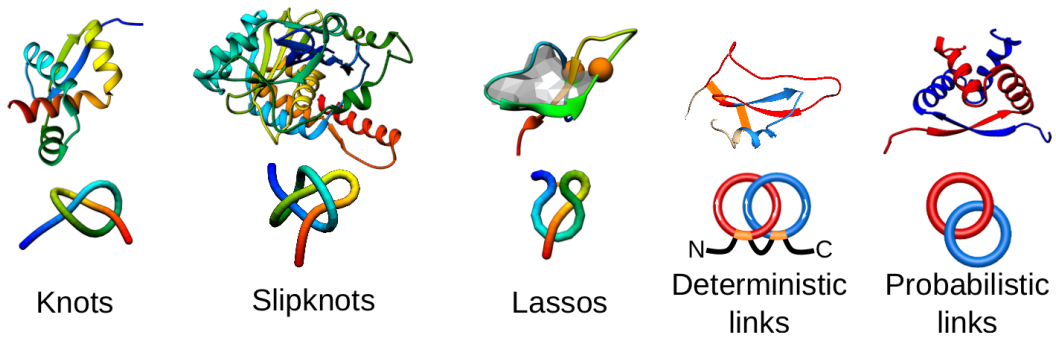

Figure 1: Knots, slipknots, lassos, cystene bridge links, and random closure links that are formed in proteins (from Sulkowska abstract)

of this research objective. Virnau reviewed his computational research on knotted protein structures and described his new efforts to generate samples to extremely long macromolecules, a matter of still wide-spread interest even with the new methods made available in light of the symplectic geometry of spaces of open and closed polygons. Jackson's report on experimental studies of knotted proteins provided an important perspective on the applications of numerical simulations, and highlighted several important open questions. Plunkett described her ergodic method of sampling spaces of polygonal arcs subject to a constraint on their thickness. This thickness dimension is of great significance since other methods are unable to control the thickness of the sampled structures and, yet, most biological chains behave as though the chains have some inherent thickness. Her algorithm, and that of Chapman for polygonal rings, are the only ones rigourously proved to be ergodic and able to control for thickness. Deguchi reported on sampling of structures and exposed structural properties of composite knots as well as giving new insights into the scaling exponent for entangled random polygons.

4 Outcome of the Workshop

During the workshop, clearly younger participants took the initiative to establish working relationships with others across disciplines and nationalities, as well as interacting with senior researchers they had never encountered beyond reading their publications. In fact, given the open physical environment of workshop, one could easily observe research discussions at one of the several blackboards at the venue as well as small group meetings of longtime collaborators from distant institutions. In some cases participants encountered differences about experimental or computational outcomes. As a number of participants had been together in earlier meetings, they had established a culture of respectful but frank discussion that was adopted by all participants. Both the openness and the energetic engagement in exploring different perspectives made this workshop an exceptionally productive one. One expects that the seeds planted during this short time together will ultimately result in new research directions and publications for the participants, and those they mentor, in the future.

References

[1] M. Di Pierro, R. R. Cheng, E. Lieberman Aiden, P. G. Wolynes, and J. N. Onuchic, De novo prediction of human chromosome structures: Epigenetic marking patternsn encode genome architecture, *PNAS* **46** 2017, 12126-12131.

[2] H. Zhang and S.E. Jackson, Characterization of the Folding of a 52-Knotted Protein Using Engineered Single-Tryptophan Variants. *Biophys J* **111** 2016, 2587-2599.

[3] S.E. Jackson, A. Suma, and C. Micheletti, How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. *Curr Opin Struct Biol* **42** 2016, 6-14.

[4] J .I. Sulkowska, E Rawdon, K.C. Millett, J. Onuchic, and A .Stasiak, Conservation of complex knotting and slipknotting patterns in proteins, *PNAS* **109** 2012, E1715-23.

[5] P. Dabrowski-Tumanski and J. I. Sulkowska, Topological Knots and Links in Proteins, *PNAS* **114** 2017, 34153420

[6] P. Dabrowski-Tumanski and J. I. Sulkowska, To Tie or Not to Tie? That is the question, *Polymers*, **9** 2017, 454