

08w5104 Rigidity, Flexibility, and Motion: Theory, Computation and Applications to Biomolecules

Walter Whiteley (York University)
Maria Kurnikova (Carnegie Mellon University),
David Richardson (Duke University)
Jack Snoeyink (University of North Carolina)
Michael Thorpe (Arizona State University)

July 6–July 11, 2008

1 Overview of the Field

Following from work on the genome, the focus is shifting to protein and RNA structure and function. Much of the function of a biomolecule is determined by its 3-D structure (shape) and motions (often in combination). While static structures of many new proteins is being determined by x-ray crystallography and by nuclear magnetic resonance techniques (work represented by several invitees), understanding the function of biomolecules requires understanding of conformational changes of these structures in time, e.g. dynamics. Therefore much interest has been focused recently in experimental and computational approaches to study dynamics of biomolecules on different timescales ranging from femtoseconds (local vibrations) to motions on the micro-to millisecond timescale (large-scale motions). In other words, the emphasis now is to describe a biomolecule in an ensemble of appropriate conformations rather than just as a single, static structure. In this context the impact of other contacts such as ligands (drugs), or binding of other proteins, RNA, DNA etc. to the multidimensional energy landscape need to be addressed. It is apparent that experimental approaches yield important but limited information. Computational approaches have to play a crucial role in this goal. Computer modeling of possible motions; the conformational accessible space, unfolding pathways; multiple configurations with different biological functions and paths between these are just some of the objectives of current computational geometry and applied mathematics.

The mathematical theory of rigidity, and related techniques from geometric constraint theory (CAD, robotics), are tools for such computer modeling, and the development of fast algorithms. Applications of such techniques to protein flexibility have been expanding over the last ten years [8, 22, 21, 24, 25]. The core mathematical theory has also been evolving in multiple ways, through the work of mathematicians and computer scientists [7, 10, 26]. A short summary of the state of the art for combinatorial analysis of rigidity includes three problems:

- (a) the general problem of predicting whether a graph, build in 3-space as a bar and joint framework, will be rigid or flexible, for almost all realizations, is a long-standing problem, going back at least to James Clerk Maxwell.
- (b) the general problem of predicting whether a graph built with vertices as rigid bodies, and edges as hinges, in 3-space, will be rigid or flexible for almost all choices of lines for the hinges, has a simple

combinatorial solution and an efficient algorithm (a result of Tay and Whiteley, extending the basic theorem of Tay) [20, 26];

- (c) the general problem of frameworks extracted from covalent bonds of molecular structures (with fixed angles at the bonds) is covered by the algorithms of (b) (the Molecular Framework Conjectures) although it is a special class of frameworks under (a) [21, 25]. Recent experimental and theoretical work provides strong evidence for these conjectures, and the proof has recently been announced [16]. The implementations of this approach provides fast combinatorial algorithms for modeling and predictions (see flexweb.asu.edu).

Work in computational geometry (represented by other participants) has investigated the computational complexity of a variety of algorithms and questions around folding and unfolding chains, polygons and other simplified models that would relate to proteins [12, 18]. This includes the recent solution of the Carpenters' Rule Problem (can a plane linear linkage, laid out with out crossings, always be straightened in the plane without any crossing during the motion?), which combined computational geometry with results in rigidity theory [3, 19, 18]. Work on linkages in 3-space confirms that the 3-D problem is significantly harder, but also indicates that some results can be obtained. The computational modeling of molecular motions involves problems such as ring closure and inverse kinematics, which are central to robotics. Work in robotics has studied the kinematics of larger scale structures subject to geometric constraints, using probabilistic road map algorithms and other sampling techniques. These have recently been extended to combine rigidity decomposition and probabilistic road maps (see the material at parasol.tamu.edu/foldingserver/), and key people from this work are among the invitees.

In computational biophysics and biochemistry, there are web implementations of several algorithms [e.g. FIRST available at flexweb.asu.edu] for modeling the rigidity and initial flexibility of biomolecules as frameworks to which this mathematical theory applies [21, 22]. These models incorporate rigidity features, including the Molecular Framework Conjecture. These algorithms are fast enough to be used as preliminary screening in areas such as ligand docking in drugs [12] as well as some simulations of unfolding pathways. Motions simulated over larger time-scales are also being developed, using programs such as ROCK (also on flexweb). A single static image, plus a rigidity simulation generated an ensemble of conformations similar to the ensemble of conformations generated from NMR data. Key people in these simulations and comparisons participated, along with others who are exploring related computational methods. Other methods, such as Gaussian Network Models (also fast and simple) and Molecular Dynamics Simulations (slow but with more detail) also offer predictions. It is important to compare these methods and search for ways to refine and combine them; some of the organizers and invitees are working on these problems. Significant and suggestive initial results have been obtained, but much work remains.

The mathematical and computational models have become more sophisticated, offering qualitative and quantitative predictions for the behaviour of biological complexes. Meanwhile, dynamics measured using X-ray crystallography, NMR [15, 23, 27], fluorescence and other experimental techniques are expanding. We are now entering a period of comparison of predictions with experimental observations, which is challenging at all levels. What motions does the experimental data measure (and on what time scale)? How do different measurements compare? What properties of the mathematical models correlate with measurements? What role to other mathematical models and assumptions play in generating the experimental evidence deposited in sites such as the Protein Data Bank? What is the reliability of data and the scope of the models. Among the organizers and the invitees, we have some leading experts in this area, who will assist with the vital task of grounding the models in the best evidence from experiment, both as a caution and a stimulation to the computational and mathematical modelers.

Of course, such intensive work on modeling stimulates new mathematical problems, which has generated progress on existing unsolved problems. Work in rigidity has always engaged collaborations with other disciplines, including material science, physics, engineering, and now biology and chemistry. The organizers and invitees include representatives from many of these fields who are already collaborating with mathematicians. Each of these fields is in rapid evolution, due both to new theoretical results and to new experimental techniques that modify our assumptions and raise new questions. The work is increasingly interdisciplinary and the workshop proposal reflects that reality. Any major mathematical progress will have potential impact within mathematics and with work in these other fields.

The July 20004 BIRS Workshop 04w5017: Modeling Protein Flexibility and Motions, offered an unusually interdisciplinary gathering of people from these diverse communities. This workshop was a continuation of the first workshop, though as a half sized group - with restricted numbers and a sharper focus.

2 Recent Developments and Open Problems

At the time the workshop was held, the Molecular conjecture was an important unsolved problem. Since the workshop, a manuscript has circulated, and the consensus is that this conjecture has been confirmed [16]. This completes the basis for the FIRST algorithm (flexweb.asu.edu) and supports the ongoing application of these 'generic' techniques to determine the first-order rigidity of molecular models extracted from the protein structures.

In the period prior to the workshop, there was significant progress on global rigidity in the plane - and there was a major unsolved general problem of further characterizing generic structures (including molecules) which are globally rigid in 3-space. During the workshop, some significant conjectures were generated, and important new results are presented below.

3 Presentation Highlights

We held several joint sessions with The Biology-Combinatorics Interface: Addressing New Challenges in Computational Biology 08w5069. In the first two days we held four joint lectures, giving an overview of the problems each group was addressing - and laying the groundwork for further informal conversations during the five days.

One of these survey talks, by Michael Thorpe, covered a range of methods from rigidity analysis and finite motion simulation, and a range of applications including zeolite (a crystalline like material), modeling protein dynamics and virus capsids and symmetry. The lecture is available at the BIRS web site for viewing and downloading.

Maria Kurnikova presented a biochemistry tutorial, including a focus on stability vs instability (a recurring theme after this talk) and insights from Molecular Dynamics Simulation (MDS).

Walter Whiteley led off a discussion of current problems - potential areas for work during the workshop. In the remaining periods, we mixed talks with working sessions, to enable people to engage in more detailed conversations to share work and to develop more refined questions and conjectures. One special feature of our workshop was the number of hands on models, and materials for constructing additional models. These were used during talks, within working groups and in the evenings in the common room. These models were essential to clarifying points and exploring possibilities - as people gathered around tables and explored situations such as the possible 'flips' for structures which were rigid, but not 'stable' (see below).

In the context of our visual and kinesthetic exploration of examples, we participated as subjects in an ongoing research project of Natasha Meyers, an Anthropologist of Science at York University. With our informed consent, Natasha observed and collected field notes on how people in such an interdisciplinary environment communicated and interacted. In addition, Natasha gave a one hour talk about her prior work on interactions between supervisors and graduate students who were constructing 3-D models of proteins, as well as some larger questions for her ongoing research. Overall, these interactions provided an additional awareness of how we communicated, and what situations lead to confusion or to clarity for people coming from diverse intellectual communities.

With a number of senior graduate students participating, we ensured that each of them had an opportunity to present and to receive feedback about fruitful directions and additional methods which could be applied to their work. Areas for such follow-on discussions included the interactions of symmetry and rigidity, and the exploration of periodic structures (real and mathematical zeolites), as well as flatness of substructures as an additional feature which altered the rigidity of structures. These themes engaged a substantial number of the participants - and the further exchanges among those participants have continued since this workshop.

Derek Wilson presented particular protein modeling challenge - Acylphosphatase from *Sulfolobus Solfataricus* (Sso AcP) [?]. We were given HD Exchange data (colored by a convention) on two conformations of a molecule. The challenge was to consider mathematical techniques, including rigidity analysis for individual

conformations or for an ensemble of conformations from the Protein Data Bank, and generate a ‘comparable’ coloring / prediction of which parts were ‘rigid’ and which were ‘flexible’ (variable). This discussion integrated with our growing awareness of communication issues / variations, so that on the last day we held a discussion on ways of color coding measures of ‘flexibility’ both as detected in diverse samples of biological data and in mathematical simulations.

In Thursday, we had another pair of talks shared by the two workshops: Mary Condon on RNA Secondary Structure and David Richardson and Jack Snoeyink on RNA tertiary structure.

Before the workshop and throughout the week, a wiki site was used to post a number of conjectures and open problems. There were regular updates, including posting of presentations and revised conjectures. This was quite fruitful and continued to be available for some months following the workshop.

4 Scientific Progress Made

As anticipated, there was substantial interaction around the use of the words *rigid*, *flexible* and *stable* in the various disciplines. Over the week, through examples and explorations, there was a refined consensus that:

1. flexible as used in generic rigidity, and in molecular discussions, described a configuration with a continuous path of variation in shape.
2. stable, as used in molecular descriptions, coincided with *global rigidity* in the discussions of mathematicians and computer scientists - particularly for generic configurations (or configurations where small variations did not change the global rigidity). This generally coincides with redundant rigidity, where removing one constraint still leaves rigidity (see the next item);
3. minimal infinitesimal matched up with what biochemists and biophysicists would call fluctuating, or having small range floppy nodes.

One of the key examples, explored with models and with mathematical theories was the five-fold ring - which is generically globally rigid (and redundantly rigid) and occurs in the basic structure of proline - the unusual amino acid which is rigid rather than flexible across the carbon bonds between amide plates.

From this analysis, there were two threads of further discussion. (A) A recognition that in ‘coloring’ flexibility, it was at least important to use three colors: red, for flexible; grey, for minimally infinitesimally rigid; and blue for redundantly rigid / stable. Of course the flexible and the redundantly rigid could be further colored to show ‘degrees’ but comparisons might well only be possible across fields with these three categories. (B) It was important to develop a more complete mathematical and computational theory for stability (global rigidity). This was evident in the extensive entries on the wiki for conjectures and problems around global rigidity, and in the progress made to generate new conjectures and begin work on verifying these conjectures and extending methods for further work.

In the second thread, a conjecture was developed, by Tibor Jordan Meera Sitharam, and Walter Whiteley, that a generic body bar framework is globally rigid if and only if it is redundantly rigid. As the section below indicates, this conjecture has been proven, using methods explored during the workshop and extended afterwards. This conjecture extends to the claim that a generic molecular framework is globally rigid if and only if it is redundantly rigid. This further conjecture remains unsolved, but is now more accessible because of the results for body-bar frameworks.

Work continued on how symmetry (a common feature of families of proteins, such as dimers) impacts the rigidity of the corresponding framework model and the underlying molecule. This was a topic first explored during the prior 2004 workshop, and has become an important area of mathematical work, combining mathematical rigidity and the representation theory of groups. Some further collaborative work was developed (see below for a resulting paper) and topics for continuing exploration were listed. These topics played a valuable part in the recently completed thesis of Bernd Schulze [17]. Further recent specific connections to protein modeling are mentioned below.

5 Outcome of the Meeting

There was substantial overlap with with two day workshop immediately following this five day workshop, both in terms of participants and in terms of topics / conjectures / collaborations. With that in mind, several of the resulting collaborations and papers will reasonably be reported in both reports.

As mentioned above, one theme which rose in importance through the preparations for the workshop, the conjectures posted on the wiki site, and discussions of what ‘stability’ in biochemistry translated to in mathematics, and the mathematical discussions was Global Rigidity. One outcome was the conjectures for body bar frameworks linking redundant rigidity with global rigidity, generically. This core conjecture is confirmed in the paper of Connely, Jordan and Whiteley [5]. From this result, there are natural questions of extensions to molecular structures - so there is an extended Global Rigidity conjecture for molecular structures: a generic molecular model is globally rigid if and only if it is redundantly rigid. We note that there are fast algorithms for redundant rigidity, encoded in programs such as FIRST at flexweb.asu.edu.

The paper above was based on some extensions of the prior stress matrix methods for confirming global rigidity in all dimensions. An other paper flowing from the workshop, by Connelly and Whiteley [6], confirmed projective transformations and coning as a valuable methods for global rigidity, also giving a mathematical basis for transferring results on global rigidity to other metrics, such as spheres, hyperbolic geometry, and even broader metrics built on the shared projective geometry of the structures.

On the impact of symmetry on rigidity / flexibility, continuing discussions in several informal sessions during the workshop generated the core results which are in the paper [11], authored by Simon Guest, Bernd Schulze and Walter Whiteley. This extension of prior results for bar and joint frameworks of [4] to more general frameworks suggests that versions for molecular structures are within reach. Since that time, the thesis of Bernd Schulze [17] has clarified a number of issues, and generated new results which have potential applications back to molecules. This work is also ongoing.

The discussion generated by an example brought by the biochemist Derek Wilson provided some ideas of mathematical / computational models which could potentially address the data given. As a result, a new collaboration of Adnan Sljoka (mathematics graduate student) and Derek Wilson (biochemistry) on explicit ways of incorporating ensemble information in the mathematical theory to give a plausible account / prediction of the observed HD data for This collaboration has continued, forming part of Ph.D. thesis work of Adnan Sljoka, and a draft paper is being polished for submission.

The presentation of Adam Watson on his Ph.D. work on Flatness and Rigidity demonstrated situations where some simple geometry not detected in the usual counts gave additional flexibility to a structure. The ensuing discussion laid the seeds for further correspondence on when symmetry was sufficient to induce the corresponding flatness, again an extra situation not detected even in the symmetry adapted counts of [4, 11]. There are plans to present these connections in a joint paper of Adam Watson, Bernd Schulze, and Walter Whiteley.

It is clear that, overall, the workshop supported new collaborations, and supported the developed of all the graduate students who participated. In the summer of 2009, a workshop in Budapest provided follow up for a number of the mathematical topics which arose during this five day workshop and the follow-on two day workshop.

References

- [1] Bemporad, F.; Gsponer, J.; Hopearuoho, H. I.; Plakoutsi, G.; Stati, G.; Stefani, M.; Taddei, N.; Vendruscolo, M.; Chiti, F. Biological function in a non-native partially folded state of a protein *Embo J.* **27** 2008, 1525-1535.
- [2] G. Crippen and T. Havel, *Distance Geometry and Molecular Conformation*, John Wiley & Sons, 1988
- [3] R. Connelly, E. Demaine and G. Rote, Straightening polygonal arcs and convexifying polygons. *Discrete & Computational Geometry*, **30** 2003, 205239.
- [4] R. Connelly, P. Fowler, S. Guest, B. Schulze and W. Whiteley, When is an symmetric framework iso-static? *International Journal of Solids and Structures* **Vol. 46**, 2009, 762-773.

- [5] R. Connelly T. Jordan and Walter Whiteley, Generic global rigidity of body-bar Frameworks, EGRES TR-2009-13, www.cs.elte.hu/egres/.
- [6] R. Connelly and W. Whiteley, Global rigidity: The effect of coning, to appear in *Discrete & Computational Geometry*.
- [7] R. Connelly and W. Whiteley, Second-order rigidity and pre-stress stability for tensegrity framework, *SIAM J. Discrete Math.* **9** (1996), 453-492.
- [8] Philip Duxbury and Michael Thorpe (eds), *Rigidity Theory and Applications*, Academic/Kluwer, 1999.
- [9] P.W. Finn and L.E. Kavraki, Computational Approaches to Drug Design. *Algorithmica*, pp. 347-371, 1999.
- [10] J. Graver, B. Servatius and H. Servatius, *Combinatorial Rigidity*, Graduate Studies in Mathematics, AMS, (1993)
- [11] S. Guest, B. Schulze, and W. Whiteley, When is a bar-body framework isostatic? preprint 2009.
- [12] V. Schneke and L.A. Kuhn, Flexibility screening for molecules interacting with proteins, in *Rigidity Theory and Applications*, Academic/Kluwer, 1999, 385-4000.
- [13] Donald Ingber, The Architecture of Life, *Scientific American*, (1998), 2-11,
- [14] D. Ingber, S. Heidemann, P. Lamoureux and R. Busbaum, Opposing views on tensegrity as a structural framework for understanding cell mechanics: Controversies in Physiology, *J. App. Physiol.* **89** (2000), 1663-1678.
- [15] T. Mamonova, B. Hesperheide, M. Thorpe and M. Kurnikova, Protein flexibility using constraints from molecular dynamics simulations, *Physical Biology* **2** (2005).
- [16] N.Katoh, S.Tanigawa, A proof of the molecular conjecture, preprint 2009, arXiv:0902.0236.
- [17] B. Schulze, Combinatorial and Geometric Rigidity with Symmetry Constraints, Ph.D. Thesis, Mathematics and Statistics, York University, 2009.
- [18] M. Soss and G. Toussaint, Geometric and computational aspects of polymer reconfiguration, *J. Math. Chem.* **27** (2001), 303-318.
- [19] I. Streanu. Combinatorial Roadmaps in Configuration Spaces of Simple Planar Polygons. In S. Basu and L. Gonzalez-Vega editors, *Proceedings of the DIMACS Workshop on Algorithmic and Quantitative Aspects of Real Algebraic Geometry in Mathematics and Computer Science* 2003, 181-206.
- [20] T-S. Tay and W. Whiteley, Generating isostatic frameworks, *Structural Topology* **11** (1985), 20-69.
- [21] M.F. Thorpe, Ming Lei, A.J. Rader, Donald J. Jacobs and Leslie A. Kuhn, Protein Flexibility Predictions using Graph Theory, *Proteins* **44**, (2001), 150-165
- [22] Michael Thorpe and Leslie Kuhn (eds.), *Protein Flexibility and Folding*, Elsevier, 2001
- [23] Wells S., Menor S, Hesperheide B. and Thorpe M. F 2005, Constrained Geometric Simulation of Diffusive Motion in Proteins, *Physical Biology* **2** (2005)127-136 .
- [24] Walter Whiteley, Rigidity of molecular structures: generic and geometric analysis, in *Rigidity Theory and Applications*, Academic/Kluwer, 1999, 21-46.
- [25] Whiteley, W. 2005, Counting out the flexibility of molecules, *Physical Biology* **2** 1-11.
- [26] Walter Whiteley, Matroids from Discrete Geometry, in *Matroid Theory*, J. Bonin, J. Oxley and B. Servatius (eds.), AMS Contemporary Mathematics, 1996, 171-313.
- [27] M. Zhang and L.E. Kavraki, Solving Molecular Inverse Kinematics Problems for Protein Folding and Drug Design. in *Currents in Computational Molecular Biology*, pp. 214-215, ACM Press, April 2002.