

# Some Geometric and Computational Challenges Arising in Structural Molecular Biology

Bruce R. Donald

Department of Computer Science, Department of Chemistry, and Department of Biochemistry  
Duke University and and Duke University Medical Center, Durham, NC, USA

[www.cs.duke.edu/brd/](http://www.cs.duke.edu/brd/)

[brd+socg19@cs.duke.edu](mailto:brd+socg19@cs.duke.edu)

---

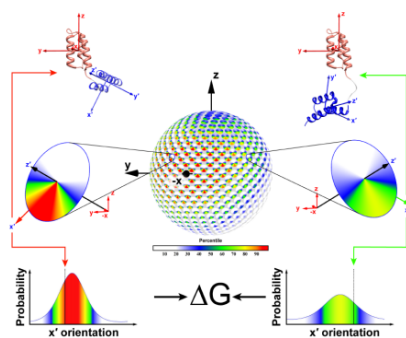
## Abstract

---

Computational protein design is a transformative field with exciting prospects for advancing both basic science and translational medical research. New algorithms blend discrete and continuous geometry to address the challenges of creating designer proteins. I will discuss recent progress in this area and some interesting open problems.

I will motivate this talk by discussing how, by using continuous geometric representations within a discrete optimization framework, broadly-neutralizing anti-HIV-1 antibodies were computationally designed that are now being tested in humans – the designed antibodies are currently in eight clinical trials, one of which is Phase 2a (NCT03721510). These continuous representations model the flexibility and dynamics of biological macromolecules, which are an important structural determinant of function.

However, reconstruction of biomolecular dynamics from experimental observables requires the determination of a conformational probability distribution. These distributions are not fully constrained by the limited geometric information from experiments, making the problem ill-posed in the sense of Hadamard. The ill-posed nature of the problem comes from the fact that it has no unique solution. Multiple or even an infinite number of solutions may exist. To avoid the ill-posed nature, the problem must be regularized by making (hopefully reasonable) assumptions.



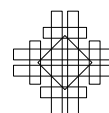
I will present new ways to both represent and visualize correlated inter-domain protein motions (see the figure above). We use Bingham distributions, based on a quaternion fit to circular moments of a physics-based quadratic form. To find the optimal solution for the distribution, we designed an efficient, provable branch-and-bound algorithm that exploits the structure of analytical solutions to the trigonometric moment problem. Hence, continuous conformational PDFs can be determined directly from NMR measurements. The representation works especially well for multi-domain systems with broad conformational distributions. For more information please see Y. Qi et al. [1].

Ultimately, this method has parallels to other branches of geometric computing that balance discrete and continuous representations, including physical geometric algorithms, robotics, computational geometry, and robust optimization. I will advocate for using continuous distributions for protein modeling, and describe future work and open problems.



© Bruce R. Donald;  
licensed under Creative Commons License CC-BY  
35th International Symposium on Computational Geometry (SoCG 2019).  
Editors: Gill Barequet and Yusu Wang; Article No. 2; pp. 2:1–2:2

Leibniz International Proceedings in Informatics  
LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



## 2:2 Geometric and Computational Challenges

**2012 ACM Subject Classification** Theory of computation → Design and analysis of algorithms

**Keywords and phrases** Geometric computing, computational biology, Continuous Interdomain Orientation Distributions of Protein Conformations

**Digital Object Identifier** 10.4230/LIPIcs.SoCG.2019.2

**Category** Invited Talk

**Related Version** Y. Qi et al. [1]

---

### References

- 1 Yang Qi, Jeffrey W. Martin, Adam W. Barb, François Th  lot, Anthony K. Yan, Bruce R. Donald\*, and Terrence G. Oas\*. Continuous interdomain orientation distributions reveal components of binding thermodynamics. *Journal of Molecular Biology*, 430(18, Part B):3412 – 3426, 2018. doi:10.1016/j.jmb.2018.06.022.

---

\*Corresponding Authors