Birthday Symposium for Bruce Donald

Duke Gardens, Durham NC

July 17, 2018

Morning Session

Introduction & Welcome
Chittu Tripathy, Mark Hallen, Ryan Lilien.

What I've Learned from Bruce and How it Has Changed My Science,
Terrence Oas, *Professor of Biochemistry, Duke university.*

I will briefly summarize the impact of Bruce on both my administrative as well as scientific life. I first encountered Bruce when I was leading the Bioinformatics and Genomic Technology PhD Program. I will describe his willingness to help us apply for an NIH training grant. I will also describe the impact that our interactions have had on my appreciation for continuous distributions and the role that they can play in interpretation of structural NMR data. I will also mention the potential utility of this approach in the interpretation of structural data from other methods.

From Robotics and Computational Geometry to Molecular Biology,
Jack Snoeyink, *Professor of Computer Science, UNC Chapel Hill.*

Bruce Donald's PhD thesis, "A Theory of Error Detection and Recovery: Robot Motion Planning with Uncertainty in the Geometry of the Environment," did not prefigure his excessive number of affiliations: "Professor of Computer Science at Duke University, Professor of Chemistry, and Professor of Biochemistry in the Duke University Medical Center. He is also Professor of Electrical and Computer Engineering in the Duke Pratt School of Engineering." I'll briefly explain why the foundational work that Bruce did in computational geometry plays a role in so many areas, but will leave it to him to explain why he would want to make it official that he is supposed to contribute to so many areas. Nevertheless, these areas are glad that he has done so.

Computational Structural Biology, Robotics and the Algorithms In Between,
Ramgopal Mettu, *Associate Professor, Department of Computer Science, Tulane University.*

In this talk I'll give a brief overview of my work in the areas of algorithms, computational biology and most recently, robotics. Algorithms for discrete (and continuous) optimization are of course the common thread, but just as important is how a problem is chosen, modeled, and solved. While working with Bruce as a postdoc we devised novel algorithms for NMR data interpretation that had both nice theoretical properties, and also worked on actual experimental data. This experience provided a template for future work, and I'll describe how my subsequent research has at least attempted to be both theoretically interesting and practically relevant.

Development of a Novel 3D-Topography Scanning and Analysis System for Firearm Forensics,
Ryan Lilien, *Founder, Cadre Forensics; Adjunct Professor of Computer Science, University of Toronto*.

I will describe my group's work developing an accurate, fast, and low-cost 3D imaging and analysis system for firearm forensics. Which is, of course, the next logical field to transition to after computational biology. As portrayed in the movies, ammunition cycled through a firearm picks up small microscopic imperfections (*i.e.*, toolmarks) unique to that firearm. Microscopic examination of these marks allows firearm examiners to assess the likelihood of common origin (*e.g.*, the linking of a cartridge case found at a crime scene to a test fire from a suspect's firearm). At Cadre, we are developing a novel 3D scanning and analysis system for cartridge cases. Our system has been validated and is now in use

at the FBI's primary firearm and toolmark laboratory in Quantico, VA. Experience supports the hypothesis that 3D imaging methods such as ours has the potential to greatly impact the criminal justice system.

Cell-Shape Oscillations and Their Biological Relevance,
Glenn Edwards, *Professor of Physics, Duke University.*

Bruce and I met while serving on Duke's Academic Promotions and Tenure Committee. I'll start with a necessarily cryptic comment about our shared experiences during those three years. We also talked in brief about our research interests on those many walks to and from APT. Having Bruce's attention during this Symposium, I'll review some recent biophysics research on the mechanism for cell oscillations and cell ingression. This research uses mathematical and physical techniques that I did not learn as an undergraduate math major nor as a physics graduate student. Consequently the target biological audience became rather narrow. In retrospect, this project is bit like a drum solo. My wife thinks drum solos are self indulgent and detract from the musical experience. Generally I like drum solos, but are there drum solos in blue grass? In any event, I anticipate Bruce really will understand this mechanism, so this is an opportunity not to be missed. Now for the science: Cell-shape oscillations have generated considerable interest.These oscillations occur at low-Reynolds number and from this perspective it was surprising that a band of oscillation frequencies (5.7 ± 0.9 mHz) has been resolved experimentally in amnioserosa cells during Drosophila morphogenesis. This talk will review a molecular mechanism that: 1) identifies the key-attributes of a molecular contractile unit that leads to a unified mechanism for producing oscillations and the steady forces that promote ingression; 2) characterizes a regulatory process that switches contractile units from oscillatory to steady-force production; and 3) incorporates the role of actin cross-links in scaling piconewton actomyosin forces to nanonewton intercellular forces.

Afternoon Session 1

NMR.Needs.Bruce (Pump up the Volume),
Jeffrey Hoch, *Joseph Meyerhoff Visiting Professor of Chemical Physics, Weizmann Institute of Science; Professor of Molecular Biology and Biophysics, UConn Health; Director, Gregory P. Mullen NMR Structural Biology Facility; Director, NMRbox.org: National Center for Biomolecular NMR Data Processing and Analysis.*

Bruce Donald made his mark in a number of areas before turning his attention to computational aspects of biomolecular NMR spectroscopy - but his influence is most sorely needed in NMR as nowhere else. A curse of modern computational hardware is that it is possible to take dumb ideas and with enough brute force emit something...OK. However not everything that can be computed should. Bruce brings a sense and sensibility to computation in bioNMR that we as a field need to emulate in order to truly succeed. Through an example (or two) I'll show how bedrock principles embodied by Bruce's work could earn a new generation of assistant professors tenure.

The Ubiquitous Sparsity: From Modeling Protein Structures Using Sparse NMR Data to Understanding Customer Behavior,
Chittu Tripathy, *Staff Data Scientist, Walmart Labs.*

My work in the D-Lab was mainly focused on using Sparse NMR data, specifically, the residual dipolar couplings and residual chemical shift anisotropy data to determine the 3D protein structural folds at atomic resolution. After moving to an industry job, I only began to realize why modeling with sparse data is even more important. In past 5 years, I have worked on real-world problems involving millions of customers of two retail giants of our time, Amazon and Walmart, primarily focusing on modeling customer's behavior and their interactions with the products, ranging in domains such as last-mile logistics, personalization, customer-segmentation, targeting with an aim to understand the 'WHY' behind a customer-action. Sitting at the core of these problems, that are perceived to derive inferences from the enormous amount of data, are a number of sparse-data problems. I will talk about some of the models I have worked on in the recent past, and how my training in the D-lab naturally extends to this area.

Protein Backbone and Protein Surface: New Solutions to Old Problems,
Lincong Wang, *College of Computer Science and Technology, Jilin University, China.*

The first part of my talk focuses on a novel representation of protein backbone as a polyline linking consecutive peptide plane centers. This representation is the basis of

a highly-accurate, distance-based algorithm for protein secondary structure element assignment that uses none of backbone hydrogen bonds. Next I present the characterization of protein-ligand interactions using solvent-excluded surface (SES) and a series of SES-defined physical and geometrical properties. For example we have proposed a new model for protein-protein interaction interfaces specified in terms of the large differences in SES-defined properties between completely-buried and partially-buried surface atoms.

On the Topic of Velvet Paintings and Rotamer Hippopotami (with Some Science),
Jonathan Jou (JJ), *Postdoctoral Research Associate at Duke.*

What follows will be 20ish minutes chronicling my time in the lab, highlighting personally memorable anecdotes, hopefully exciting science, and more than a few good lab memories. Topics include, but are not limited to: protein design, algorithms, lab photos, graduation gifts, thesis cover artwork, personal gaffes, and maybe even clip from a music video, if I can get my hands on it.

Afternoon Session 2

PLUG (Pruning of Local Unrealistic Geometries) Removes Restrictions on Biophysical Modeling for Protein Design,
Mark A. Hallen, *Research Assistant Professor, Toyota Technological Institute at Chicago.*

Protein design algorithms must search an enormous conformational space to identify favorable conformations. As a result, those that perform this search with guarantees of accuracy generally start with a conformational pruning step, such as dead-end elimination (DEE). However, the mathematical assumptions of DEE-based pruning algorithms have up to now severely restricted the biophysical model that can feasibly be used in protein design. To lift these restrictions, I propose to prune local unrealistic geometries (PLUG) using a linear programming-based method. PLUG's biophysical model consists only of well-known lower bounds on interatomic distances. Based on 96 test cases, PLUG is at least as effective at pruning as DEE for larger protein designs?the type that most require pruning. When combined with the LUTE protein design algorithm, PLUG greatly facilitates designs that account for continuous entropy, large multi-state designs with continuous flexibility, and designs with extensive continuous backbone flexibility and advanced non-pairwise energy functions.

A Computational Pipeline for Design of Novel RNA-like Topologies,
Swati Jain, *Postdoctoral Associate, Department of Chemistry, New York University.*

Designing novel RNA topologies is important for therapeutic and industrial applications. I will describe a recent computational pipeline for design of novel RNA topologies we have developed based on our lab's coarse-grained RNA-As-Graphs (RAG) framework, that represents RNA secondary structures as tree graphs. We have previously enumerated possible graph topologies for graphs up to 13 vertices and identified RNA-like topologies among these. We developed a systematic design pipeline, using recently developed tools for graph-partitioning and fragment assembly (F-RAG). Following partitioning of the target graph, corresponding atomic fragments from our database of RNA substructures (RAG-3D) are combined using F-RAG, and the candidate atomic models are scored using a knowledge-based potential. The sequences of the top scoring models are screened further using available tools for 2D structure prediction. Experimental structure probing using SHAPE-MaP for two sequences agree with our predictions and suggest that our combined tools yield excellent candidates for further sequence and experimental screening.

Vector Embedding of Proteins in a Protein-Protein Interaction Network using Deep Learning,
Mehmet Serkan Apaydin, *Researcher at Inria Sophia Antipolis Mediterranée.*

I plan to obtain a vector embedding of proteins using deep learning similar to the vector embedding developed by Google and Facebook for natural language processing. This has potential applications in drug discovery and diseases such as cancer. The data we have is protein structures in the PDB and their interaction networks which is available in public databases such as Matador. The scientific question we would like to ask is whether we can identify the function of a protein from its context, just as in natural language processing we can understand words of similar meaning using word vectors.

Protein Design for Immunoengineering Powered by

Geometric Deep Learning,
Pablo Gainza, *École polytechnique fédérale de Lausanne.*

After I left Bruce's lab, I have tried to follow his footsteps in exploiting tools from computer science to design biomedically-relevant proteins. My work has mainly focused on the computational design of novel protein-protein interactions (PPIs) applied to immunoengineering (cellular therapies, vaccine design and biologics design). My talk will be divided in two parts. First, I will discuss one such application, the design of PPIs responsive to small molecules to control cellular therapies. Using computational design I developed a suite of chemically induced monomers (CIMs), high affinity PPis that monomerize in the presence of a drug. In collaboration with the Swiss Cancer Center, we have used these CIMs as safety OFF-switches to control the activity of chimeric antigen receptor T-cells extracted from patient donors, and have shown that these are constitutively active unless a drug is present. Most successes in PPI design are redesigns of existing protein binding motifs. The de novo design of PPIs, in contrast, remains an unsolved problem. In the second part of my talk I will discuss my ongoing efforts on de novo design of PPIs based on protein surfaces, exploiting state of the art tools from the field of geometric deep learning. Specifically, I will discuss SeedNet, a deep neural network that learns from real PPIs to predict fragments of proteins that can be reused to target other proteins. We are currently using these tools to computationally novel protein binders against selected cancer immunotherapy targets and to validate immunogen candidates for mimicry of pathogen epitopes.

10 Things To Love About Bruce,
Daniela Rus, *Director, CSAIL; Andrew (1956) and Erna Viterbi Professor; EECS, MIT.*

Bruce Donald has a rare ease of seeing into the future and identifying important technical problems. In the time I have known him, he has made fundamental contributions to the fields of mobile robotics, MEMS, and computational biology. His research considers problems at the edge of what is possible and beyond the boundaries of current research. Bruce has had great impact on science and on all the people he mentored and befriended. He delights us with his passion, insight, great humor, friendship, and lasagna. We are lucky to have Bruce in our lives.

Group Photo

Reception, Dinner, Cake, & Good Times