

Imran Haque

ihaque@cs.stanford.edu

<http://www.cs.stanford.edu/people/ihaque>

- EDUCATION
- ◇ **Stanford University**, Stanford, CA. (2006-2011)
M.S. and Ph.D. in Computer Science
(thesis: *Accelerating Chemical Similarity Search Using GPUs and Metric Embeddings*)
Advisors: Vijay Pande (Chemistry) and Daphne Koller (Computer Science)
 - ◇ **University of California, Berkeley**, Berkeley, CA. (2002-2006)
B.S. (Hons) *summa cum laude* in Electrical Engineering and Computer Science, May 2006
- RESEARCH INTERESTS
- Drug design, computational structural biology, protein design.
 - Democratization of information in the public interest
- AWARDS
- ◇ **Fellowships and Scholarships**
 - National Science Foundation Graduate Research Fellowship (2006 - 2009)
 - UC Berkeley Alumni Scholarship (2005 - 2006)
 - SRCEA Undergraduate Research Asst. Program Scholarship (2004-2006)
 - UC Berkeley Chancellor's Scholarship (2002-2006)
 - Robert Byrd Honors Scholarship (2002-2006)
- PUBLICATIONS
- ◇ Beauchamp KA, ... , **Haque IS**, and Pande VS. "MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale." *Journal of Chemical Theory and Computation* 2011. Published online ahead of print 6 Sep 2011.
 - ◇ Pronk S, ... , **Haque I**, ... , and Lindahl E. "Copernicus: A New Paradigm for Parallel Adaptive Molecular Dynamics." Accepted to *Supercomputing* 2011.
 - ◇ **Haque IS** and Pande VS. "Error Bounds on the SCISSORS Approximation Method". *Journal of Chemical Information and Modeling*, published online ahead of print 18 Aug 2011.
 - ◇ **Haque IS**, Pande VS, and Walters WP. "Anatomy of High-Performance 2D Similarity Calculations". *Journal of Chemical Information and Modeling*, published online ahead of print 20 Aug 2011.
 - ◇ **Haque IS** and Pande VS. "Large-Scale Chemical Informatics on GPUs". In *GPU Computing Gems, Emerald Edition*; Wen-Mei Hwu ed. Burlington, MA: Morgan Kaufmann. 2011.
 - ◇ **Haque IS** and Pande VS. "SCISSORS — A Linear-Algebraical Technique to Rapidly Approximate Chemical Similarities". *Journal of Chemical Information and Modeling*: 2010, 50, 1075-1088.
 - ◇ **Haque IS**, Pande VS, and Walters WP. "SIML: A Fast SIMD Algorithm for Calculating LINGO Chemical Similarities on CPUs and GPUs". *Journal of Chemical Information and Modeling*: 2010, 50(4), 560-564.
 - ◇ **Haque IS** and Pande VS. "Hard Data on Soft Errors: A Large-Scale Assessment of Real-World Error Rates in GPGPU". To appear in *Resilience 2010: 3rd Workshop on Resiliency in High Performance Computing* (held in conjunction with CCGrid 2010).
 - ◇ Ponder JW, ... , **Haque IS**, ... , Head-Gordon TL. "Current Status of the AMOEBA Polarizable Force Field". *Journal of Physical Chemistry B* 114(8), 2549-2564 (2010).
 - ◇ **Haque IS** and Pande VS. "PAPER — Accelerating Parallel Evaluations of ROCS". *Journal of Computational Chemistry* 31, 117-132 (2010)

- ◇ Jed W. Pitera, **Imran Haque**, and William Swope. "Absence of reptation in the high-temperature folding of the trpzip2 beta-hairpin peptide". *Journal of Chemical Physics* 124, 141102 (2006)
- INVITED TALKS
 - ◇ Hyperience: 5th National Dutch Informatics Congress. Urk, Netherlands, 24 Nov 10. "Folding@Everywhere: Computational Biochemistry in the New Era of HPC". (**Keynote Presentation**)
 - ◇ Resilience Summit at Los Alamos Computer Science Symposium. Santa Fe, NM, 13 Oct 2010. "Hard Data on Soft Errors: A Global-Scale Assessment of GPGPU Memory Soft Error Rates (Updated)".
 - ◇ Workshop on Bio-Molecular Simulations on Future Computing Architectures (Oak Ridge National Laboratory). Oak Ridge, TN. 17 Sep 2010. "Hybrid Vigor: Using Heterogeneous HPC to Accelerate Chemical Biology".
 - ◇ 3rd Workshop on Resiliency in High Performance Computing in Clusters, Clouds, and Grids (Resilience 2010) @ CCGrid 2010. Melbourne, Australia, 17 May 2010. "Hard Data on Soft Errors: A Global-Scale Assessment of GPGPU Memory Soft Error Rates".
 - ◇ CSIRO Molecular Health and Technologies. Melbourne, Australia, 13 May 2010. "Bigger, Longer, and Uncut: Chemical Informatics at Scale".
 - ◇ CUP XI Conference, Santa Fe, NM, 10 Mar 2009. "LINGOs and GPUs".
 - ◇ NVIDIA Corporation, Santa Clara, CA, 2 Dec 2009. "Do GPUs Really Need ECC? A Global-Scale Assessment of GPU Memory Soft Error Rates".
 - ◇ NSF-NAIS Workshop on Intelligent Software. Edinburgh, UK, 19-21 Oct 2009. "Of Jacquard looms and Jaccard coefficients: multithreading biomolecular simulations in a GPU world".
 - ◇ Vertex Pharmaceuticals, Cambridge, MA, 27 Apr 2009. "GPUs: Fast. Too Furious?".

RESEARCH EXPERIENCE

Computational biology / biochemistry

- ◇ **Development of a computational model to predict toxicological (ADMET) properties of small-molecule drug leads.**
Using molecular shape and chemical similarity scoring over large screening databases (tens of thousands to millions of compounds) and machine learning techniques based on known ADMET data to predict pharmacological properties of unassayed compounds. *With Kim Branson. Advisor: Vijay Pande.*
- ◇ **Fast similarity search algorithms over combinatorial chemical libraries**
Improved on PQS, an existing algorithm at Vertex Pharmaceuticals to do chemical similarity search over combinatorial chemical libraries, by developing clustering-based techniques to prioritize libraries to search. Achieved order-of-magnitude runtime speedups with no loss of result quality. *With Brian Goldman, Pat Walters (Vertex Pharmaceuticals). (June - September 2009)*
- ◇ **Invention and GPU implementation of a new algorithm for the LINGO chemical similarity measure**
Developed a new algorithm to evaluate the LINGO chemical similarity metric, with 2x speedup on a CPU and over 80x speedup on a GPU relative to commercial code. *Advisor: Pat Walters (Vertex Pharmaceuticals). (June - September 2009)*
- ◇ **Development of PAPER, a GPU-accelerated molecular alignment code**
Wrote PAPER, a GPU (graphics processing unit)-accelerated implementation of the Gaussian overlap optimization algorithm to find the optimal overlay of small molecules. Demonstrated multiple order of magnitude speedups versus a reference CPU-based code, and accuracy comparable to a commercial implementation of the algorithm. *Advisor: Vijay Pande. (April - October 2008)*
- ◇ **Characterizing ribose binding protein and a designed indole acetic acid binding protein**
Performed site-directed mutation on natural and computationally-redesigned ribose binding proteins to introduce attachment points for FRET fluorophores to measure protein domain motion. Expressed modified protein in *E. coli* and purified protein using FPLC. Did organic synthesis, HPLC purification, and LC/MS characterization of a

modified fluorescent dye from literature.

Advisor: Pehr Harbury. (April - June 2007)

- ◇ **Exploring enzyme conformational dynamics using molecular dynamics.**
Set up molecular dynamics simulations of the cyclophilin A peptide prolyl isomerase to investigate its conformational dynamics during catalysis.
With John Chodera. Advisor: Vijay Pande (January - March 2007)
- ◇ **Investigation of the links between genomic structural rearrangements and cancer phenotypes**
Applied a genetic “module network” machine learning approach to investigate effects of DNA copy-number variation in the prostate cancer genome on cancer phenotype.
With Brad Gulko, Sharareh Noorbaloochi, Keyan Salari. Advisor: Daphne Koller. (October 2006 - October 2007)
- ◇ **Summer Research Internship, IBM Almaden Research Center.**
Performed data-mining of a large dataset from molecular dynamics simulation of the trpzip2 mini-protein to find evidence of a novel mode of protein folding. Collaborated with Dan Bonachea at UC Berkeley to port the Titanium language to the BlueGene/L supercomputer. Developed a scalable k-medians clustering code in Titanium.
With Jed Pitera, IBM Almaden Research Center (Jun - Aug 2005)

Computer Science/Electrical Engineering

- ◇ **Investigation of memory soft error rates in GPGPU**
Developed MemtestG80, a custom test code to assess memory error rates on NVIDIA G80/GT200-based graphics cards. Ran MemtestG80 on over 50,000 hosts on the Folding@home distributed computing network to find that, in their installed environments, two-thirds of tested GPUs exhibited a detectable, pattern-sensitive rate of memory soft errors. *Advisor: Vijay Pande. (January - November 2009)*
- ◇ **Development of a next-generation distributed computing platform**
Designed and demonstrated a prototype architecture for the next generation of the Folding@Home distributed computing platform, based on VMWare virtualization to present a common environment across heterogeneous client machines.
Advisor: Vijay Pande. (January - March 2007)
- ◇ **Development of gCensus, a web-based GIS program**
Developed gCensus, a web-based geographic information system (GIS) offering visualization and mapping of US Census data using Google Earth.
Available online at <http://gecensus.stanford.edu>
Individual project (Jun 2006 - present)
- ◇ **Design and Analysis of LDPC communication code decoders.**
Helped design and implement, on arrays of FPGAs, partially-parallel architectures for decoder units for LDPC communications codes. Empirically studied the error-correcting capabilities of various LDPC codes at high signal-to-noise ratios.
With Zhengya Zhang, Lara Dolecek. Advisor: Bora Nikolic, UC Berkeley. (Aug 2004 - May 2006)
- ◇ **Library Optimization in the Titanium Programming Language.**
Optimized, debugged, and benchmarked the Titanium Domain library, which is critical to array operations in Titanium, a Java dialect for parallel computing.
With Dan Bonachea. Advisor: Kathy Yelick, UC Berkeley. (May 2004 - May 2005)

TEACHING EXPERIENCE

- ◇ Teaching Assistant, Stanford BMI 214/CS 274, Representations and Algorithms for Computational Molecular Biology, Russ Altman. (Spring 2010)
- ◇ Teaching Assistant, Stanford CS 148, Introduction to Computer Graphics and Imaging, Pat Hanrahan. (Fall 2010)
- ◇ Teaching Assistant, Stanford CS 109, Introduction to Probability for Computer Scientists. (Winter 2011)

MEMBERSHIPS ◇ Eta Kappa Nu (National EE honor society)

- PRESENTED RESEARCH
- ◇ **Imran Haque** and Vijay Pande. “Accelerating Biochemical Data Mining 1,000,000×”. Research poster at BCATS conference, Stanford, CA, 6 Nov 2010.
 - ◇ **Imran Haque** and Vijay Pande. “Cheminformatics at Scale — Bigger, Longer, and Uncut”. Research poster at EuroCUP IV conference, Bergisch-Gladbach, Germany, 28 Apr 2010.
 - ◇ **Imran Haque** and Vijay Pande. “Lies, Damned Lies, and AUC Confidence Intervals”. Research poster at EuroCUP IV conference, Bergisch-Gladbach, Germany, 28 Apr 2010.
 - ◇ **Imran S. Haque** and Vijay S. Pande. “GPUs — TeraFLOPS or TeraFLAWED?”. Research poster at SC09 (Supercomputing 2009) conference, Portland, OR, 17 Nov 2009.
 - ◇ **Imran Haque** and Vijay Pande. “GPUs — TeraFLOPS, or TeraFLAWED?”. Research poster at EuroCUP III conference, Toledo, Spain, 22 Apr 2009.
 - ◇ Kim Branson, **Imran Haque**, and Vijay Pande. “I See a Color Force Field and I Want It Painted Black”. Research poster at CUP X conference, Santa Fe, NM, 9 Mar 2009.
 - ◇ **Imran Haque** and Vijay Pande. “Rochambeau — Playing Games with ROCS”. Research poster at CUP X conference, Santa Fe, NM, 9 Mar 2009.
 - ◇ **Imran Haque**, John Chodera, Michael Shirts, David Mobley, Vijay Pande. “Toward Quantitative Prediction of Binding Affinities to JNK3 by Alchemical Free Energy Methods”. Research poster at CUP IX conference, Santa Fe, NM, 17 Mar 2008.
 - ◇ Brad Gulko, **Imran Haque**, Sharareh Noorbaloochi, and Keyan Salari. “Role of DNA Copy Number Alterations in the *trans*-Regulatory Network of Cancer Cells”. Research poster at National Cancer Institute Integrative Cancer Biology Program meeting at Stanford University, 13 Feb 2007.