BIOGRAPHICAL SKETCH

NAME: **John D. Chodera**

eRA COMMONS USER NAME: **JCHODERA**

POSITION TITLE: **Assistant Member,** Computational Biology Program

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	Completion Date	FIELD OF STUDY
California Institute of Technology	BS	06/1999	Biology
University of California, San Francisco	PhD	12/2006	Biophysics
Stanford University	Postdoc	2007-2008	Chemistry
QB3 Fellow, University of California, Berkeley	Postdoc	2008-2012	Quantitative Biosciences

A. PERSONAL STATEMENT

My research focuses on the use of rigorous statistical mechanics and physical modeling to develop predictive, quantitative computational models to enable rapid rational engineering of small molecule ligands for use as tool compounds for computational biology or potential therapeutics. This project focuses on developing model systems of complexity intermediate between trivial physical modeling exercises and complex pharmaceutical drug targets that focus on specific challenges to predictive physical modeling; generating datasets for these model systems with particular attention to experimental error and concordance between experimental and modeled systems; running blind challenges that focus the community in ways that significantly advance predictive modeling; and curating and disseminating this data and the results of these competitions.

My laboratory is particularly well suited to developing model protein:ligand systems of intermediate complexity. We have significant expertise in the structural and chemical bioinformatics necessary to mine public databases for potential model systems; we have developed an automated cloning and bacterial expression pipeline to screen for bacterial expression; we develop and employ high-quality automated biophysical assays and Bayesian experimental analysis techniques to measure binding affinities; and we both extensively utilize alchemical free energy calculations and co-organize the primary meetings for this field (Alchemical Free Energy Calculations in Drug Discovery) so are intimately aware of which model systems will provide maximal value to the community for blind challenges and high-quality datasets. In addition, we have experience in coordinating the measurement of physical property datasets by sending students to work with pharma collaborators, analyzing and curating this data, and releasing his data for blind challenges. **I have co-organized the SAMPL challenge series with PI Mobley for the past several years, and my laboratory supplied data for the previous round of SAMPL challenges.**

At the Sloan Kettering Institute, my laboratory consists of twelve theorists and experimentalists that combine theory, advanced simulation algorithms, high performance computing, and automated biophysical measurements to develop quantitative models for predicting and understanding how small molecules (such as drugs) modulate cellular pathways, how mutations lead to drug resistance, and how this resistance can be circumvented or suppressed. My laboratory has extensive experience in the use of alchemical free energy calculations for the computation of protein-small molecule binding affinities, and is actively engaged in efforts to scale our methodologies to aid in the design of high-affinity ligands that bind selectively to desired members of protein families. My laboratory makes heavy use of large-scale computational resources, including the Folding@Home distributed computing platform, national supercomputing resources, and high-performance GPU computing resources at MSKCC. I am also actively involved in developing new high-throughput protocols for high-quality, high dynamic range binding affinity and physical property measurements; laboratory automation techniques; experimental design guided by Bayesian inference and information theoretic principles; and the use of Bayesian inference and bootstrap simulation for accurate assessment of measurement error.

The four publications highlighting our specific expertise for this proposal demonstrate *(a)* our grasp of the challenges facing predictive modeling using modern free energy methods in drug discovery, *(b)* our early recognition of the value in iterative blinded challenges to drive improvements in quantitative predictive modeling, *(c)* our ability to collect useful blind challenge experimental datasets in collaboration with pharma, and *(d)* our focus on the thorough characterization of all sources of experimental error in affinity measurement.

a. **Chodera, J.D**., Mobley, D.L., Shirts, M.R., Dixon, R.W., Branson, K.M., and Pande, V.S. Free energy methods in drug discovery and design: Progress and challenges. *Current Opinion in Structural Biology*, 21:150-160, 2011. PMCID: PMC3085996
b. Mobley, D.L., Graves, A.P., **Chodera, J.D.,** McReynolds, A.C., Shoichet, B.K., and Dill, K.A. Predicting absolute ligand binding free energies to a simple model site. *Journal of Molecular Biology*, 371:1118-1134, 2007. PMCID: PMC2104542
c. Rustenburg, A.S, Dancer, J., Lin, B., Feng, J.A., Ortwine, D.F., Mobley, D.L., and **Chodera, J.D.** Measuring experimental cyclohexane-water distribution coefficients for the SAMPL5 challenge. *Journal of Computer Aided Molecular Design* 30:945, 2016. PMCID: PMC5209288
d. Hanson, S.M., Ekins, S., and **Chodera, J.D.** Modeling error in experimental assays using the bootstrap principle: understanding discrepancies between assays using different dispensing technologies. *Journal of Computer Aided Molecular Design* 29:1073, 2015. PMCID: PMC4696763

B. Positions and Honors

POSITIONS AND EMPLOYMENT (*current positions in* ***bold)***

2012-present **Assistant Member** and Laboratory Head, Computational and Systems Biology Program, Sloan Kettering Institute for Cancer Research, MSKCC (primary appointment)
2013-present **Assistant Professor**, Program in Physiology, Biophysics, and Systems Biology, Weill Cornell Graduate School of Medical Sciences
2013-present **Faculty Member**, Tri-Institutional PhD Program in Chemical Biology
2013-present **Faculty Member**, Tri-Institutional PhD Program in Computational Biology and Medicine
2015-present **Faculty Member**, Gerstner Sloan Kettering Graduate School of Medical Sciences, MSKCC
2008-2012 QB3 Distinguished Postdoctoral Fellow, University of California, Berkeley, Berkeley, CA
2007-2008 Postdoctoral Fellow, Department of Chemistry, Stanford University
2005 IBM Almaden Research summer internship, Blue Gene project, under William C. Swope

HONORS AND AWARDS

2013-2106 Louis V. Gerstner Young Investigator Award
2008-2012 QB3-Berkeley Distinguished Postdoctoral Fellowship
2005-2006 IBM Predoctoral Fellowship
2005 Frank M. Goyan Award for outstanding work in Physical Chemistry, UCSF
2000-2005 Howard Hughes Medical Institute Predoctoral Fellowship

OTHER EXPERIENCE AND PROFESSIONAL MEMBERSHIPS

2000-present Member, American Chemical Society
2014-present Scientific Advisory Board, Schrödinger

C. CONTRIBUTIONS TO SCIENCE

To date, I have published more than 55 articles in peer-reviewed journals, which have collectively received over 5932 citations in the literature. My current h-index is 33, and my i10-index is 49.

1. Accurate alchemical free energy calculations of ligand binding affinities. With the aim of enabling true computer-guided design of small molecules as potential therapeutics and chemical probes, I have spent the better part of a decade developing alchemical free energy methodologies into a quantitative, predictive tool for accurate computation of small molecule binding affinities to biomolecular targets. Work I have led or contributed to has benchmarked and improved the accuracies of free energy calculations, fixed deficiencies in methodologies, helped establish best practices, developed new efficient simulation algorithms, and exploited high-performance graphics computing hardware (GPUs) to greatly advance our progress toward this goal. We have made effective use of model systems and blind tests as a means of identifying systematic improvements in methodologies. Key papers demonstrate the capability of GPU-based free energy calculations to discover and compute affinities to new binding sites, review challenges facing the deployment of these techniques in drug discovery, address the problem of multiple kinetically-trapped conformational states contributing to binding, and demonstrate the power of cycles of experiment and computation to drive improvements.

a. Wang, K., **Chodera, J.D.,** Yang, Y., and Shirts, M.R. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *Journal of Computer-Aided Molecular Design* 27:989, 2013. PMCID: PMC4154199
b. **Chodera, J.D**., Mobley, D.L., Shirts, M.R., Dixon, R.W., Branson, K.M., and Pande, V.S. Free energy methods in drug discovery and design: Progress and challenges*. Current Opinion in Structural Biology*, 21:150-160, 2011. PMCID: PMC3085996
c. Mobley, D.L., **Chodera, J.D.,** and Dill, K.A. Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. Journal of Chemical Theory and Computation, 3:1231-1235, 2007. PMCID: PMC2562444
d. Mobley, D.L., Graves, A.P., **Chodera, J.D.,** McReynolds, A.C., Shoichet, B.K., and Dill, K.A. Predicting absolute ligand binding free energies to a simple model site. *Journal of Molecular Biology*, 371:1118-1134, 2007. PMCID: PMC2104542

2. Quantitative experimental biophysics. I have been involved in the development of new techniques for the analysis of a variety of biophysical measurements. In the field of single-molecule force spectroscopy, I developed new techniques for the analysis of both nonequilibrium and equilibrium experiments. Working with force spectroscopists at UC Berkeley, I developed data analysis techniques crucial to demonstrating that nascent polypeptide chains translated by the ribosome have their folding properties modulated by electrostatic interactions with the ribosome *(a)*, mechanical characterization of the molten globule state of a protein *(b),* and limitations of constant-force-feedback experiments *(c)*. I have also worked extensively with biophysical techniques for the measurement of protein-ligand binding affinities, focusing on highly accurate quantitative measurements using techniques such as isothermal titration calorimetry *(d)*.

a. Kaiser, C., Goldman, D.H., **Chodera, J.D.**, Tinoco, I. Jr., and Bustamante, C. (2011) The ribosome modulates nascent protein folding. *Science* 334:1723. PMCID: PMC4172366
b. Elms, P.J., **Chodera, J.D.**, Bustamante, C., Marqusee, S. (2012) The molten globule state is unusually deformable under mechanical force. *Proceedings of the National Academy of Sciences USA* 109:3796. PMCID: PMC3309780.
c. Elms, P.J., **Chodera, J.D.**, Bustamante, C.J., Marqusee, S. (2012) Limitations of constant-force-feedback experiments. Biophysical Journal, 103, 1490, 2012. PMCID: PMC3471466
d. Tellinghuisen, J., and **Chodera, J.D.** Systematic errors in isothermal titration calorimetry: Concentrations and baselines. *Analytical Biochemistry* 414:297, 2011.

3. Biomolecular conformational dynamics and structural biology. Biological macromolecules are not static entities, but populate a variety of kinetically metastable conformational states critical to binding and function. The long lifetimes of these metastable states present a challenge for molecular simulation, which are generally limited in length to a few microseconds. Together with collaborators at Stanford, the IBM Almaden Research Center, and the Freie Universität Berlin, I developed an approach to use *Markov state models* (MSMs) to build stochastic models of the long-time dynamics of biomolecules from many short atomistically-detailed molecular simulations *(a)*. This technique allows for the characterization of thermally accessible metastable conformational states and their kinetics, permits direct connection between computation and experiment *(b)*,

and is now utilized by many laboratories around he world. More recently, we have also demonstated how biomolecular simulation and modeling can be used in conjunction with biophysical and biochemical experiments to understand important aspects of enzyme function *(c,d)*.

a. **Chodera, J.D.**, Signhal, N., Pande, V.S., Dill, K.A., and Swope, W.C. (2007). Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *Journal of Chemical Physics* 126, 155101. PMID: 174616665
b. Noé, F., Doose, S., Daidone, I., Löllmann, M., Sauer, M., **Chodera, J.D.**, and Smith, J.C. (2011). Dynamical fingerprints: A theoretical framework for understanding biomolecular processes by combination of simulation and kinetic experiments. *Proceedings of the National Academy of Sciences USA* 108:4822, 2011. PMCID: PMC3064371
c. Cypers, S., Ruff, E., Behr, J.M., **Chodera, J.D.,** and Levinson, N.M. A conserved water-mediated hydrogen bond network governs allosteric activation in Aurora kinase A. *Nature Chemical Biology* 13:402, 2017. PMCID: Submitted
d. Intlekofer, A., Wang, B., Liu, H., Shah, H., Carmona-Fontaine, C., Rustenburg, A., Salah, S., Gunner, M.R., **Chodera, J.D.,** Cross, J.R., and Thompson, C.B. Acidification enhances production of L-2-hydroxyglutarate through alternative substrate use by dehydrogenase enzymes. *Nature Chemical Biology* 13:494, 2017. PMCID: Submitted.

4. Advances in molecular simulation algorithms and methodologies. Throughout my career, I have been active in the development of new algorithms to increase the efficiency of molecular simulations, establish best practices, benchmark and improve molecular mechanics forcefields, and exploit novel computing paradigms. Key advances include recognizing replica exchange simulations can be considered a form of Gibbs sampling *(a)*, new estimators for combing simulation data from a variety of temperatures *(b)*, the development of a new GPU-accelerated molecular simulation framework (*c)*, and a simple solution to the longstanding problem of detecting when a simulation has sufficiently equilibrated *(d)*.

a. **Chodera, J.D.**, and Shirts, M.R. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *Journal of Chemical Physics* 135:194110, 2011. PMID: 22112069
b. Prinz, J.H, **Chodera, J.D.,** Pande, V.S., Swope, W.C., Smith, J.C., Noé, F. (2011) Optimal use of data in parallel tempering simulations for the construction of discrete-state Markov models of biomolecular dynamics. *Journal of Chemical Physics* 134, 244108. PMCID: PMC3139503
c. Eastman, P., Friedrichs, M., **Chodera, J.D.,** Radmer, R., Bruns, C., Ku, J., Beauchamp, K., Lane, T.J., Wang, L.P., Shukla, D., Tye, T., Houston, M., Stitch, T., Klein, C., Shirts, M.R., and Pande, V.S. OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *Journal of Chemical Theory and Computation* 9:461, 2012. PMCID: PMC3539733
d. **Chodera, J.D.** A simple method for automated equilibration detection in molecular simulations. *Journal of Chemical Theory and Computation,* in press. Submitted to PMC.

5. Nonequilibrium statistical mechanics. The discovery of the Jarzynski equality (JE) in 1997 and the Crooks fluctuation theorem (CFT) in 1999 touched off a revolution in the field of statistical mechanics, providing for the first time exact relationships between the behavior of systems driven out of equilibrium and their equilibrium counterparts. I have been heavily involved in efforts to produce robust, reliable, and useful statistical estimators from these theorems, enabling the analysis of both nonequilibrium molecular simulations and real nonequilibrium biophysical experimental data to produce optimal estimates of physical properties like free energies and equilibrium expectations, along with good estimates of error (*a* and *b)*. Together with Gavin Crooks and David Min, I developed a new efficient simulation methodology that exploits nonequilibrium driving---nonequilibrium candidate Monte Carlo (NCMC)---which can increase the acceptance probability of Monte Carlo in complex systems moves by orders of magnitude *(c).* More recently, we have shown how nonequilibrium theorems and estimators can yield new insight into the errors made in simulating physical systems by discretizing dynamical equations of motion for computer simulation (*d*).

a. Minh, D.D.L and **Chodera, J.D.** Optimal estimators and asymptotic variances for nonequilibrium path-ensemble averages. *Journal of Chemical Physics* 131, 134110, 2009. PMCID: PMC2771048

b. Minh, D.D.L. and **Chodera, J.D.** Estimating equilibrium ensemble averages using multiple time slices from driven nonequilibrium processes: Theory and application to free energies, moments, and thermodynamic length in single-molecule pulling experiments. *Journal of Chemical Physics* 134, 024111, 2011. PMID 21241084.
c. Nilmeier, J.P., Crooks, G.E., Minh, D.D.L., and **Chodera, J.D**. Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation. *Proceedings of the National Academy of Sciences USA* 108, E1009, 2011. PMCID: PMC3215031
d. Sivak, D.A., **Chodera, J.D.,** and Crooks, G.E. Using nonequilibrium fluctuation theorems to understand and correct errors in equilibrium and nonequilibrium simulations of discrete Langevin dynamics. Physical Review X, 011007, 2013.

Complete list of published work available at MyNCBI Collections:
http://www.ncbi.nlm.nih.gov/sites/myncbi/john.chodera.1/bibliography/43349161/public

D. RESEARCH SUPPORT

Ongoing Research Support

I8-A8-058 (PI: Luo) 1/1/2015 – 6/30/2017
Starr Cancer Consortium
Designing Sinefungin Scaffolds as Specific Protein Methyltransferase Inhibitors
Our long-term goal is to develop PMT inhibitors for epigenetic cancer therapy, with the current objective to establish a drug-discovery pipeline with sinefungin analogues.
Role: Co-Investigator

SK2016-0731 (PI: Chodera) 8/29/2016 – 8/28/2018
Merck KG
Benchmarking absolute alchemical free energy calculations with YANK
The objective of this project is to perform a large-scale benchmark of alchemical free energy calculations using the GPU-accelerated open source code YANK.
Role: Investigator

Completed Research Support

SK2015-0252 (PI: Chodera) 7/1/2015 – 10/31/2016
AstraZeneca
Evaluating the potential for Markov state models of conformational dynamics
Our goal is to evaluate the potential for Markov state models of conformational dynamics to describe the mechanism of slow off-rate inhibition in the human kinases CK2 and SYK.
Role: Investigator