OMB Number: 4040-0001
Expiration Date: 06/30/2016

APPLICATION FOR FEDERAL ASSISTANCE
SF 424 (R&R)

3. DATE RECEIVED BY STATE	**State Application Identifier**

1. TYPE OF SUBMISSION*

❍ Pre-application ● Application ❍ Changed/Corrected Application

4.a. Federal Identifier

b. Agency Routing Number

c. Previous Grants.gov Tracking Number

2. DATE SUBMITTED 2016-02-04

Application Identifier

5. APPLICANT INFORMATION

Organizational DUNS*: 064931884

Legal Name*: Sloan Kettering Institute for Cancer Research
Department:
Division:
Street1*: 1275 York Avenue
Street2:
City*: New York
County: New York
State*: NY: New York
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 10065-6007

Person to be contacted on matters involving this application
Prefix: First Name*: Annmarie Middle Name: L Last Name*: Pacchia Suffix: PhD
Position/Title: Vice President
Street1*: 1275 York Ave.
Street2:
City*: New York
County: New York
State*: NY: New York
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 10065-6007
Phone Number*: 646-227-3273 Fax Number: 212-557-0760 Email: sponsorp@mskcc.org

6. EMPLOYER IDENTIFICATION NUMBER *(EIN) or (TIN)** 1-3162418-2

7. TYPE OF APPLICANT* M: Nonprofit with 501C3 IRS Status (Other than Institution of Higher Education)

Other (Specify):
Small Business Organization Type ❍ Women Owned ❍ Socially and Economically Disadvantaged

8. TYPE OF APPLICATION*

● New ❍ Resubmission
❍ Renewal ❍ Continuation ❍ Revision

If Revision, mark appropriate box(es).

❍ A. Increase Award ❍ B. Decrease Award ❍ C. Increase Duration
❍ D. Decrease Duration ❍ E. Other *(specify)* :

Is this application being submitted to other agencies?* ❍Yes ●No What other Agencies?

9. NAME OF FEDERAL AGENCY*
National Institutes of Health

10. CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER
TITLE:

11. DESCRIPTIVE TITLE OF APPLICANT'S PROJECT*
The role of reorganization energy in achieving selective kinase inhibition

12. PROPOSED PROJECT
Start Date* 09/01/2016
Ending Date* 08/31/2021

13. CONGRESSIONAL DISTRICTS OF APPLICANT
NY-012

SF 424 (R&R) APPLICATION FOR FEDERAL ASSISTANCE

14. PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR CONTACT INFORMATION

Prefix: First Name*: John Middle Name: D Last Name*: Chodera Suffix:
Position/Title: Assistant Member
Organization Name*: Sloan Kettering Institute for Cancer Research
Department: Computational Biology
Division:
Street1*: 1275 York Ave.
Street2:
City*: New York
County: New York
State*: NY: New York
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 10065-6007
Phone Number*: 646-888-3400 Fax Number: Email*: john.chodera@choderalab.org

15. ESTIMATED PROJECT FUNDING

a. Total Federal Funds Requested*	$2,142,500.00
b. Total Non-Federal Funds*	$0.00
c. Total Federal & Non-Federal Funds*	$2,142,500.00
d. Estimated Program Income*	$0.00

16.IS APPLICATION SUBJECT TO REVIEW BY STATE EXECUTIVE ORDER 12372 PROCESS?*

a. YES ○ THIS PREAPPLICATION/APPLICATION WAS MADE AVAILABLE TO THE STATE EXECUTIVE ORDER 12372 PROCESS FOR REVIEW ON:

DATE:

b. NO ● PROGRAM IS NOT COVERED BY E.O. 12372; OR

○ PROGRAM HAS NOT BEEN SELECTED BY STATE FOR REVIEW

17. By signing this application, I certify (1) to the statements contained in the list of certifications* and (2) that the statements herein are true, complete and accurate to the best of my knowledge. I also provide the required assurances * and agree to comply with any resulting terms if I accept an award. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. (U.S. Code, Title 18, Section 1001)

● I agree*

** The list of certifications and assurances, or an Internet site where you may obtain this list, is contained in the announcement or agency specific instructions.*

18. SFLLL or OTHER EXPLANATORY DOCUMENTATION File Name:

19. AUTHORIZED REPRESENTATIVE

Prefix: First Name*: Annmarie Middle Name: L Last Name*: Pacchia Suffix: PhD
Position/Title*: Vice President
Organization Name*: Sloan Kettering Institute for Cancer Research
Department: Research & Technology Mgmt
Division: Office of Rsch & Project Admin
Street1*: 1275 York Ave.
Street2:
City*: New York
County: New York
State*: NY: New York
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 10065-6007
Phone Number*: 646-227-3273 Fax Number: 212-557-0760 Email*: sponsorp@mskcc.org

Signature of Authorized Representative*

Annmarie L Pacchia

Date Signed*

02/04/2016

20. PRE-APPLICATION File Name:

21. COVER LETTER ATTACHMENT File Name:cover_letter1011909030.pdf

424 R&R and PHS-398 Specific Table Of Contents

	Page Numbers
SF 424 R&R Cover Page	1
Table of Contents	3
Performance Sites	4
Research & Related Other Project Information	5
Project Summary/Abstract(Description)	6
Project Narrative	7
Facilities & Other Resources	8
Equipment	9
Other Attachments	10
Authentication_of_Key_Resources_Plan1011909182	10
Research & Related Senior/Key Person	11
PHS398 Cover Page Supplement	17
PHS 398 Modular Budget	19
Personnel Justification	25
PHS 398 Research Plan	26
Specific Aims	27
Research Strategy	28
Bibliography & References Cited	40
Letters Of Support	44
Resource Sharing Plans	46

OMB Number: 4040-0010
Expiration Date: 06/30/2016

Project/Performance Site Location(s)

Project/Performance Site Primary Location

❍ I am submitting an application as an individual, and not on behalf of a company, state, local or tribal government, academia, or other type of organization.

Organization Name:	Sloan Kettering Institute for Cancer Research
Duns Number:	064931884
Street1*:	1275 York Ave.
Street2:	
City*:	New York
County:	New York
State*:	NY: New York
Province:	
Country*:	USA: UNITED STATES
Zip / Postal Code*:	10065-6007
Project/Performance Site Congressional District*:	NY-012

File Name

Additional Location(s)

Funding Opportunity Number: PA-13-302. Received Date: 2016-02-04T15:46:46.000-05:00

OMB Number: 4040-0001
Expiration Date: 06/30/2016

RESEARCH & RELATED Other Project Information

1. Are Human Subjects Involved?* ○ **Yes** ● **No**

1.a. If YES to Human Subjects

Is the Project Exempt from Federal regulations? ○ Yes ○ No

If YES, check appropriate exemption number: _ 1 _ 2 _ 3 _ 4 _ 5 _ 6

If NO, is the IRB review Pending? ○ **Yes** ○ **No**

IRB Approval Date:

Human Subject Assurance Number

2. Are Vertebrate Animals Used?* ○ **Yes** ● **No**

2.a. If YES to Vertebrate Animals

Is the IACUC review Pending? ○ Yes ○ No

IACUC Approval Date:

Animal Welfare Assurance Number

3. Is proprietary/privileged information included in the application?* ○ **Yes** ● **No**

4.a. Does this project have an actual or potential impact - positive or negative - on the environment?* ○ **Yes** ● **No**

4.b. If yes, please explain:

4.c. If this project has an actual or potential impact on the environment, has an exemption been authorized or an environmental assessment (EA) or environmental impact statement (EIS) been performed? ○ Yes ○ No

4.d. If yes, please explain:

5. Is the research performance site designated, or eligible to be designated, as a historic place?* ○ **Yes** ● **No**

5.a. If yes, please explain:

6. Does this project involve activities outside the United States or partnership with international collaborators?* ○ **Yes** ● **No**

6.a. If yes, identify countries:

6.b. Optional Explanation:

	Filename
7. Project Summary/Abstract*	Abstract1011714733.pdf
8. Project Narrative*	Project_Narrative1011909031.pdf
9. Bibliography & References Cited	References1011973129.pdf
10.Facilities & Other Resources	Facilities1011909183.pdf
11.Equipment	Equipment1011909185.pdf
12. Other Attachments	Authentication_of_Key_Resources_Plan1011909182.pdf

Funding Opportunity Number: PA-13-302. Received Date: 2016-02-04T15:46:46.000-05:00

PROJECT SUMMARY / ABSTRACT

The ability to rationally design small molecules that bind with high affinity and specificity to one or more biomolecular targets would radically transform drug discovery. Current approaches require many rounds of screening, modeling, and synthesis in a trial-and-error approach that is costly, time-consuming, and ineffective. After decades of work on the study of biomolecular interactions, there remains an enormous gulf between what we claim to understand about biomolecular association and our ability to put this knowledge into practice. This gulf is especially wide for the design of selective kinase inhibitors, which aim to target one or more specific kinases in order to effectively treat a disease—often cancer—and minimize unwanted toxic side effects.

While the discovery of imatinib was hailed as a breakthrough for its ability to selectively inhibit Abl despite the existence of closely related kinases like Src, it came as a great surprise when the crystal structure of imatinib bound to Src revealed that the Src:imatinib complex was nearly identical to Abl:imatinib. Recent evidence from both experiments and modeling has suggested that a previously underappreciated contribution—the energetic cost of populating the inhibitor-bound conformation—plays a critical role in the selectivity of imatinib for Abl over Src. While this effect has only been studied in the well-studied case of Abl/Src binding to imatinib, it has the potential to be much more general. **We hypothesize that exploiting differences in the energetic cost of confining related kinases to inhibitor binding-competent conformations may be a route to selectivity in targeted kinase inhibition.** Here, we ask how much conformational reorganization energy contributes to the affinity of current FDA-approved noncovalent kinase inhibitors to determine whether existing inhibitors exploit differences in these reorganization energies (perhaps inadvertently) to achieve selectivity, and whether there is a clear route to exploiting this difference in rationally engineering new selective molecules.

We use a combined experimental and computational approach to decompose inhibitor binding affinity and selectivity into contributions from kinase reorganization and binding to individual kinase conformations. We first computationally map the conformations accessible to a diverse panel of human kinase catalytic domains, along with their associated energetics. By using an automated fluorescence assay to measure the affinities of FDA-approved noncovalent inhibitors to this panel and alchemical free energy calculations to determine the inhibitor binding affinities to individual conformations, we can combine these data to quantify the relative contribution of reorganization energy to the affinity and selectivity of kinase inhibition. We then use the introduction of point mutants intended to modulate selectivity via reorganization energies to validate our model, and examine opportunities for exploiting differences in reorganization energy between related kinases or wild-type and mutationally activated kinases as a route to selectivity.

PROJECT NARRATIVE

A new generation of therapies for cancer called *selective kinase inhibitors* target the specific signaling pathways that are disrupted by disease. While the first few drugs in this class suggest these molecules can have tremendous promise as therapeutics, the development of new selective inhibitors tailored toward aberrant signaling pathways in other diseases is still incredibly challenging. We use a combined experimental and computational approach to investigate a previously underappreciated phenomenon that may be responsible for much of the difficulty in developing new selective inhibitors, and develop an approach to harness this effect to make development of targeted inhibitors easier.

Facilities and Other Resources

Computer: All lab members are equipped with laptop computers with integrated graphics processors (GPUs), and have access to high-performance development machines containing a range of modern GPU accelerators. The group has priority access to a high-performance computing cluster with 1920 total hyperthreads and 120 NVIDIA GTX-680 or GTX-Titan GPUs. Project storage is provided by a high-performance shared 1.5PB GPFS storage system. Dedicated servers provide access to Folding@Home, which currently provides $\sim$19 PFLOP/s aggregate computational power in over 350,000 actively computing cores—**equivalent computing facilities would cost tens of millions of dollars**. Network connections are at least 1 Gbit/s throughout MSKCC, with HPC systems connected at 10 Gbit/s.

Laboratory: The Chodera wetlab occupies $\sim$340 square feet of space. The central feature of the wetlab is an integrated platform for fully automated biophysical experiments instrumented for remote monitoring and operation. This system includes a Thermo BenchTrak Orbitor, a Tecan EVO200 with three dispensing technologies (including an HP D300), four Inheco incubators, a BioNex HiG4 centrifuge, Tecan Infinite M1000PRO plate reader (capable of absorbance, fluorescence and FP, luminescence, and AlphaScreen measurements, with injectors installed for kinetics measurements), Caliper GXII microfluidic electrophoresis platform, Roche LC480 qPCR machine, Agilent VCode barcode printer and PlateLoc plate sealer, Thermo MultiDrop Combi reagent dispenser, and Thermo automated Cytomat Hotel. **This platform automates cloning, site-directed mutagenesis, recombinant bacterial protein expression and purification, cell-free transcription and translation, microfluidic gel electrophoresis, Thermofluor protein stability assays, and fluorescence measurements of binding affinities. It can also automate preparation of ITC and SPR experiments that can be conducted at the Rockefeller HTSRC across the street.** There is bench space for one group member to work manually using standard molecular biology tools. A Mettler-Toledo Quantos automated gravimetric solution preparation system ensures compound concentrations are always accurately and traceably prepared. An electronic lab notebook tracks all materials and measurements in the laboratory using barcodes. Shared equipment space, standard laboratory refrigerators and freezers, and common shared equipment (centrifuges, incubators, etc.) is also provided. Both experimental and computational spaces are located in Memorial Sloan-Ketterings new Zuckerman Research Center (ZRC).

Animal: N/A

Office: All lab members have desks in a modern open-plan computational biology working space where the Chodera lab currently occupies $\sim$400 ft^2. Group members are equipped with monitors, backup storage, and other standard workstation accessories. Additional office space includes Dr. Chodera's office, office space for a shared administrative assistant, shared conference rooms, and meeting and library space.

Clinical: N/A

Other Resources:

The Rockefeller high-throughput screening resource center (HTSRC) is located across the street at the Rockefeller University. The HTSRC provides a number of high-throughput binding and biophysical measurement facilities at a minimal cost to us, most notably (1) a GE/MicroCal Auto-iTC200 automated isothermal titration calorimeter capable of processing up to 384 samples unattended, and (2) a Proteon XPR36 SPR instrument (capable of processing 96 samples), among others.

The MSKCC Organic Synthesis Core under the direction of Dr. Ouathek Ouerfelli is a fully-staffed 12-person facility providing organic synthesis and consultation services to MSKCC laboratories.

The MSKCC Analytical NMR Core under the direction of Dr. George Sukenik allows for the unattended 1H-NMR characterization of compounds using an automated sample workflow.

Numerous additional MSKCC core facilities are available to MSKCC researchers, including proteomics and mass spectroscopy, NMR, X-ray crystallography, high-throughput screening, analytical chemistry, DNA sequencing, and bioinformatics consulting. Many of these core facilities are highly automated. Over 30 core facilities are currently available, all directed by Ph.D.-level experts available for consultation.

The QB3 MacroLab core facility is located at the University of California, Berkeley (Dr. Chodera's previous institution), under the technical management of Dr. Scott Gradia. This fully-staffed core facility, which functions as a contract resource, is equipped with robotics capable of high-throughput automated cloning, mutagenesis, bacterial expression, purification, and hanging-drop crystallography. The MacroLab also offers scale-up bacterial expression and purification services to produce mg quantities of protein.

Equipment

Experimental.

Electronic laboratory notebook: An electronic laboratory notebook (ELN) system manages all samples and data, and printed barcodes are used to track all materials received or generated.

Gravimetric solution preparation: Biophysical measurements of protein-ligand binding affinities are fundamentally limited by the accuracy with which compound concentrations are known. Accurate affinity measurements are absolutely essential to validating and improving computational methodologies for probing and predicting binding affinities, so it is essential that compound concentrations be known precisely and accurately. Methods to *measure* concentrations are generally costly, inaccurate, time-consuming, and often not universally applicable. Precise preparation of initial compound solutions remains the best way to ensure accuracy. Our laboratory is therefore equipped with a high-precision Mettler-Toledo Quantos balance for automated gravimetric solution preparation. Powder dosing heads dose compound directly into solubilization vials on the analytical balance, while liquid dosing heads dispense solvent under argon, ensuring accurate concentrations of compound solutions. Provenance, masses, concentrations, and uncertainties are tracked via barcodes and electronically within our ELN.

Integrated liquid handler and automation platform: By the start of this award, the experimental laboratory will be equipped with a high-throughput automation platform for 96- and 384-well biophysical assays. The automated system will integrate the following instruments: a high-precision liquid-handling platform equipped for vacuum filtration, density measurement, and thermal cooling/heating/shaking; a high-end multimode plate reader with injectors; heating/shaking deep well plate incubators for bacterial culture; a plate centrifuge; a qPCR machine used for PCR and ThermoFluor; a microplate sealer; barcode-based tracking capabilities; an automated plate carousel; and a microfluidic gel electrophoresis system. A LabMinds EasySolution automated buffer preparation system will ensure that all buffers required in large quantities are prepared accurately, reproducibly, and traceably.

Additional automated biophysical characterization: Through the adjacent Rockefeller HTSRC facility, our laboratory also has access a GE/MicroCal Auto-iTC200 automated isothermal titration calorimeter (ITC) and a BioRad Proteon XPR36 surface plasmon resonance (SPR) instrument. Both instruments accommodate 96-well plates for fully automated runs, and are available for our use at low cost. Our laboratory automation platform will allow these experiments to be set up automatically.

Standard molecular biology equipment: The wet laboratory is also equipped for manual molecular biology.

Computational.

Local GPU cluster: The Chodera laboratory has priority access to a new high-performance computing cluster with 480 Intel Xeon E5-2665 hyperthreaded cores (960 effective hyperthreads) and 120 NVIDIA GTX-680 or Titan graphics processor-based accelerators (GPUs). Project storage is provided by a high-performance shared 1.4PB GPFS storage system connected to the computing cluster. Network connections are at least 1 Gbit/s throughout MSKCC facilities, and cluster, GPU, and storage systems are connected with 10 Gbit/s links.

GPU development resources: All Chodera laboratory members are equipped with laptop computers with GPUs capable of GPU-accelerated software development. All members also have access to five development GPU boxes contain an assortment of most available GPUs for development and automated software testing.

Folding@home: The Chodera laboratory is a participating laboratory in the Folding@home Consortium [`http://folding.stanford.edu`]. Folding@home is a distributed computing infrastructure run by Vijay Pande at Stanford University with over 350,000 actively computing cores, making it the most powerful distributed computing project in the world in terms of aggregate performance—19 PFLOP/s of aggregate computational power. The free availability of large quantities of computer time through this network—**which would otherwise cost tens of millions of dollars in hardware and power**—greatly leverages funding provided for this proposal. Access to the Folding@home network is provided via two dedicated servers at MSKCC connected to 180TB of usable storage in a high-availability datacenter.

AUTHENTICATION OF KEY RESOURCES PLAN

Small molecule kinase inhibitors. The identity of vendor-supplied small molecule kinase inhibitors will be validated by 1H-NMR spectroscopy using the MSKCC Analytical NMR Core.

Human kinase domain plasmid constructs. The sequence of engineered human kinase domain plasmid constructs received or generated will be validated against canonical sequences in UniProt by pyrosequencing.

Recombinantly expressed human kinase domain proteins. The molecular weight, concentration, and purity of purified His-tagged recombinantly expressed human kinase domain proteins will be verified using a Caliper GXII microfluidic gel electrophoresis instrument. ThermoFluor melts (thermal denaturation scans in the presence of Cypro Orange, a dye that changes fluorescence upon binding to unfolded proteins) performed using a Roche LC480 qPCR machine will be used to verify protein stability in our buffer systems. Phosphorylation state of the purified kinase domains (which are coexpressed with phosphatase) will be confirmed by mass spectrometry.

OMB Number: 4040-0001
Expiration Date: 06/30/2016

RESEARCH & RELATED Senior/Key Person Profile (Expanded)

PROFILE - Project Director/Principal Investigator

Prefix: First Name*: John Middle Name D Last Name*: Chodera Suffix:

Position/Title*: Assistant Member
Organization Name*: Sloan Kettering Institute for Cancer Research
Department: Computational Biology
Division:
Street1*: 1275 York Ave.
Street2:
City*: New York
County: New York
State*: NY: New York
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 10065-6007

Phone Number*: 646-888-3400 Fax Number: E-Mail*: john.chodera@choderalab.org

Credential, e.g., agency login: JCHODERA

Project Role*: PD/PI Other Project Role Category:

Degree Type: Degree Year:

	File Name
Attach Biographical Sketch*:	Biosketch_Chodera1011909186.pdf
Attach Current & Pending Support:	

Tracking Number: GRANT12081788
Funding Opportunity Number: PA-13-302 . Received Date: 2016-02-04T15:46:46.000-05:00

OMB No. 0925-0001 and 0925-0002 (Rev. 10/15 Approved Through 10/31/2018)

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 characterizing an underappreciated contribution to the binding affinity of small molecule selective kinase inhibitors—the energetic penalty the kinase must pay to adopt a conformation capable of binding the inhibitor. Recent evidence in the well-studied case of imatinib's preference for Abl over Src suggests that this contribution may be key to understanding imatinib's selectivity for Abl despite the almost-identical binding contacts to Src. Our laboratory is uniquely positioned to broadly survey how widespread this phenomenon is in achieving selectivity (perhaps inadvertently) in existing selective kinase inhibitors through the use of a combined experimental and computational approach that allows us to dissect this contribution to the overall affinity and selectivity of these inhibitors. We will utilize techniques I developed for the study of conformational changes in protein folding (Markov state models) with the computational resources available to us through the Folding@home worldwide distributed computing platform to map conformations and associated energetics, and together with GPU-accelerated absolute alchemical binding free energy calculation techniques I helped pioneer, quantitatively assess the potential for exploiting this phenomenon to achieve selectivity. In parallel, we will use our unique automated biophysical platform to assay highly accurate direct binding affinities for a variety of kinase inhibitors bound to a variety of recombinantly-expressed kinases, and validate our computational model through engineering mutants that manipulate this reorganization energy penalty in defined ways.

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 our ability to automate biomolecular simulations at the superfamily scale (a), identify kinetically distinct biomolecular conformational states from these simulations (b), employ alchemical free energy calculations to determine quantitatively accurate binding affinities to individual conformations without the need for docked inhibitor structures (c), and use modeling techniques to quantify the uncertainty in experimental small molecule binding assays (d).

a. Parton, D.L., Grinaway, P.B., Hanson, S.M., Beauchamp, K.A., and **Chodera, J.D.** (2016). Ensembler: Enabling high-throughput molecular simulations at the superfamily scale. *PLoS Computational* Biology, in press. Submitted for PMCID.
b. **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
c. 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
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)***

2005 IBM Almaden Research summer internship, Blue Gene project, under William C. Swope
2007-2008 Postdoctoral Fellow, Department of Chemistry, Stanford University
2008-2012 QB3 Distinguished Postdoctoral Fellow, University of California, Berkeley, Berkeley, CA
2012-present **Assistant Member** and Laboratory Head, Computational 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

Honors and Awards

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

Other Experience and Professional Memberships

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

C. Contributions to Science

1. 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 of the long-time dynamics of biomolecules from many short atomistically-detailed molecular simulations. This technique allows for the characterization of thermally accessible metastable conformational states, along with their associated interconversion kinetics and equilibrium free energies, and is now utilized by many laboratories around he world.

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. Pitera, J.W. and **Chodera**, J.D. On the use of experimental observations to bias simulated observables. *Journal of Chemical Theory and Computation* 8:3445, 2012.
c. 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
d. Prinz, J.H., Wu, H., Sarich, M., Keller, B., Fischbach, M., Held, M., **Chodera, J.D.**, Schütte, C., and Noé, F. (2011). Markov models of molecular kinetics: Generation and validation. *Journal of Chemical Physics* 134:174105. PMID: 21548671

2. 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 algorthms, 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

3. Single-molecule experiments and 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 developed new estimators for the analysis of equilibrium single-molecule data or molecular simulations, demonstrating the ability to use this machinery to reconstruct incredibly accurate potentials of mean force of biomolecules from single-molecule force spectroscopy experiments *(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. Shirts, M.R. and **Chodera, J.D.** Statistically optimal analysis of samples from multiple equilibrium states. (2008) *Journal of Chemical Physics* 129, 124105. PMCID: PMC2671659

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 - 12/31/2016
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

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

Completed Research Support

None

PHS 398 Cover Page Supplement

OMB Number: 0925-0001

1. Project Director / Principal Investigator (PD/PI)

Prefix:
First Name*: John
Middle Name: D
Last Name*: Chodera
Suffix:

2. Human Subjects

Clinical Trial? ○ No ○ Yes
Agency-Defined Phase III Clinical Trial?* ○ No ○ Yes

3. Permission Statement*

If this application does not result in an award, is the Government permitted to disclose the title of your proposed project, and the name, address, telephone number and e-mail address of the official signing for the applicant organization, to organizations that may be interested in contacting you for further information (e.g., possible collaborations, investment)?

● Yes ○ No

4. Program Income*

Is program income anticipated during the periods for which the grant support is requested? ○ Yes ● No

If you checked "yes" above (indicating that program income is anticipated), then use the format below to reflect the amount and source(s). Otherwise, leave this section blank.

Budget Period*	Anticipated Amount ($)*	Source(s)*

PHS 398 Cover Page Supplement

5. Human Embryonic Stem Cells

Does the proposed project involve human embryonic stem cells?* ● No ○ Yes

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s) from the following list: http://grants.nih.gov/stem_cells/registry/current.htm. Or, if a specific stem cell line cannot be referenced at this time, please check the box indicating that one from the registry will be used:

Cell Line(s): Specific stem cell line cannot be referenced at this time. One from the registry will be used.

6. Inventions and Patents (For renewal applications only)

Inventions and Patents*: ○ Yes ○ No

If the answer is "Yes" then please answer the following:

Previously Reported*: ○ Yes ○ No

7. Change of Investigator / Change of Institution Questions

❑ Change of principal investigator / program director

Name of former principal investigator / program director:

Prefix:

First Name*:

Middle Name:

Last Name*:

Suffix:

❑ Change of Grantee Institution

Name of former institution*:

PHS 398 Modular Budget

OMB Number: 0925-0001
Expiration Date: 10/31/2018

Budget Period: 1

Start Date: 09/01/2016 End Date: 08/31/2017

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium Indirect (F&A)*	250,000.00
Consortium Indirect (F&A)	0.00
Total Direct Costs*	250,000.00

B. Indirect (F&A) Costs

	Indirect (F&A) Type	Indirect (F&A) Rate (%)	Indirect (F&A) Base ($)	Funds Requested ($)
1.	MTDC	71.40	250,000.00	178,500.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number): DHHS, Darryl W. Mayes 212-264-2069

Indirect (F&A) Rate Agreement Date: 10/14/2015

Total Indirect (F&A) Costs: 178,500.00

C. Total Direct and Indirect (F&A) Costs (A + B)

Funds Requested ($): 428,500.00

Funding Opportunity Number: PA-13-302. Received Date: 2016-02-04T15:46:46.000-05:00

PHS 398 Modular Budget

Budget Period: 2

Start Date: 09/01/2017 End Date: 08/31/2018

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium Indirect (F&A)*	250,000.00
Consortium Indirect (F&A)	0.00
Total Direct Costs*	250,000.00

B. Indirect (F&A) Costs

	Indirect (F&A) Type	Indirect (F&A) Rate (%)	Indirect (F&A) Base ($)	Funds Requested ($)
1.	MTDC	71.40	250,000.00	178,500.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number): DHHS, Darryl W. Mayes 212-264-2069

Indirect (F&A) Rate Agreement Date: 10/14/2015

Total Indirect (F&A) Costs: 178,500.00

C. Total Direct and Indirect (F&A) Costs (A + B)

Funds Requested ($): 428,500.00

Funding Opportunity Number: PA-13-302. Received Date: 2016-02-04T15:46:46.000-05:00

PHS 398 Modular Budget

Budget Period: 3

Start Date: 09/01/2018 End Date: 08/31/2019

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium Indirect (F&A)*	250,000.00
Consortium Indirect (F&A)	0.00
Total Direct Costs*	250,000.00

B. Indirect (F&A) Costs

	Indirect (F&A) Type	Indirect (F&A) Rate (%)	Indirect (F&A) Base ($)	Funds Requested ($)
1.	MTDC	71.40	250,000.00	178,500.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number): DHHS, Darryl W. Mayes 212-264-2069

Indirect (F&A) Rate Agreement Date: 10/14/2015

Total Indirect (F&A) Costs: 178,500.00

C. Total Direct and Indirect (F&A) Costs (A + B)

Funds Requested ($): 428,500.00

Funding Opportunity Number: PA-13-302. Received Date: 2016-02-04T15:46:46.000-05:00

PHS 398 Modular Budget

Budget Period: 4			
Start Date: 09/01/2019 End Date: 08/31/2020			

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium Indirect (F&A)*	250,000.00
Consortium Indirect (F&A)	0.00
Total Direct Costs*	250,000.00

B. Indirect (F&A) Costs

	Indirect (F&A) Type	Indirect (F&A) Rate (%)	Indirect (F&A) Base ($)	Funds Requested ($)
1.	MTDC	71.40	250,000.00	178,500.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number): DHHS, Darryl W. Mayes 212-264-2069

Indirect (F&A) Rate Agreement Date: 10/14/2015

Total Indirect (F&A) Costs: 178,500.00

C. Total Direct and Indirect (F&A) Costs (A + B)

Funds Requested ($): 428,500.00

Funding Opportunity Number: PA-13-302. Received Date: 2016-02-04T15:46:46.000-05:00

PHS 398 Modular Budget

Budget Period: 5			
Start Date: 09/01/2020 End Date: 08/31/2021			

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium Indirect (F&A)*	250,000.00
Consortium Indirect (F&A)	0.00
Total Direct Costs*	250,000.00

B. Indirect (F&A) Costs

	Indirect (F&A) Type	Indirect (F&A) Rate (%)	Indirect (F&A) Base ($)	Funds Requested ($)
1.	MTDC	71.40	250,000.00	178,500.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number): DHHS, Darryl W. Mayes 212-264-2069

Indirect (F&A) Rate Agreement Date: 10/14/2015

Total Indirect (F&A) Costs: 178,500.00

C. Total Direct and Indirect (F&A) Costs (A + B)

Funds Requested ($): 428,500.00

Funding Opportunity Number: PA-13-302. Received Date: 2016-02-04T15:46:46.000-05:00

PHS 398 Modular Budget

Cumulative Budget Information

1. Total Costs, Entire Project Period

Section A, Total Direct Cost less Consortium Indirect (F&A) for Entire Project Period ($)	1,250,000.00
Section A, Total Consortium Indirect (F&A) for Entire Project Period ($)	0.00
Section A, Total Direct Costs for Entire Project Period ($)	1,250,000.00
Section B, Total Indirect (F&A) Costs for Entire Project Period ($)	892,500.00
Section C, Total Direct and Indirect (F&A) Costs (A+B) for Entire Project Period ($)	2,142,500.00

2. Budget Justifications

Personnel Justification	Budget_Justification1011780826.pdf
Consortium Justification	
Additional Narrative Justification	

Funding Opportunity Number: PA-13-302. Received Date: 2016-02-04T15:46:46.000-05:00

BUDGET JUSTIFICATION

Senior/Key Personnel

John D. Chodera, Ph.D., Principal Investigator (3.0 calendar months effort) will serve as PI and Project Director on this project. He is an Assistant Member (Assistant Professor equivalent rank) at the Sloan Kettering Institute—the basic science arm of the Memorial Sloan Kettering Cancer Center—with extensive experience in using large-scale biomolecular simulations to map conformational dynamics, alchemical free energy calculations for ligand binding, and the use and interpretation of biophysical experiments. He has a publication track record spanning over 15 years of highly regarded work in these fields. He has a decade of experience with the Folding@home worldwide distributed computing project that will be utilized to map kinase domain conformations and energetics, wrote the GPU-accelerated alchemical free energy calculation code that will be used to compute small molecule binding affinities, designed the automated biophysical wetlab that will be used for experimentally measuring binding affinities, and developed an early version of the high-throughput fluorescence assay that will be used in these measurements. He also has extensive experience with computing biophysical observables—including NMR data—from biomolecular simulations. He will manage the project and actively supervise the work being performed in this proposal.

Other Personnel

Sonya M. Hanson, D.Phil., Postdoctoral Fellow (9.0 calendar months effort) has extensive experience in both experimental structural biology and biochemistry as well as biomolecular simulation. Dr. Hanson received her D.Phil. from Oxford University, working with Mark Sansom, as part of the prestigious National Institutes of Health Oxford-Cambridge Scholars Program. She has been extensively involved in developing the automated fluorescence assay that will be used to accurately measure direct binding affinities of kinase inhibitors to recombinantly expressed kinases for this project, and is fully trained in all aspects of protocol design and operation of the robotic platform. She was also involved in the design and implementation of the computational pipeline for automated modeling of kinase conformations and simulation preparation that will be used to conduct kinase simulations on the Folding@home distributed computing platform. She will play a key role in the design and execution of robotic experiments, setting up and analyzing the kinase domain simulations on Folding@Home, and designing and analyzing the binding free energy calculations. Dr. Hanson will also assist Dr. Chodera in supervising Ms. Rodriguez and Mr. Albanese in this project. Effort outside the 9 calendar months of effort budgeted for this project will be spent on career development activities.

Lucelenie Rodriguez, Laboratory Technician (6.0 calendar months effort) has a B.S. in Cellular and Molecular Biology, participated in the Amgen Scholars and UCSF Training Program, and has two years of experience as a laboratory technician. She will assist in the development and execution of all wetlab aspects of the project, including the automated biophysical experiments. Dr. Hanson has trained her on various aspects of the automation system necessary to aid in the design and execution of the experiments described.

Steven A. Albanese, Graduate Student in the Gerstner Sloan Kettering Graduate School of Biomedical Sciences (12.0 calendar months effort) received B.A. degrees in Biology (Cell and Molecular) and Chemistry from Cornell University. Mr. Albanese will be involved in carrying out all aspects of the project, with both experimental and computational components being integral aspects of his graduate training.

PHS 398 Research Plan

Please attach applicable sections of the research plan, below.

OMB Number: 0925-0001

1. Introduction to Application (for RESUBMISSION or REVISION only)	
2. Specific Aims	Specific_Aims1011909177.pdf
3. Research Strategy*	Research_Strategy1011973130.pdf
4. Progress Report Publication List	
Human Subjects Sections	
5. Protection of Human Subjects	
6. Inclusion of Women and Minorities	
7. Inclusion of Children	
Other Research Plan Sections	
8. Vertebrate Animals	
9. Select Agent Research	
10. Multiple PD/PI Leadership Plan	
11. Consortium/Contractual Arrangements	
12. Letters of Support	LetterOfSupport_Pande1011855911.pdf
13. Resource Sharing Plan(s)	Resource_Sharing1011855847.pdf
Appendix (if applicable)	
14. Appendix	

SPECIFIC AIMS

Cancer is the second leading cause of death in the United States, accounting for nearly 25% of all deaths; in 2015, over 1.7 million new cases were diagnosed, with over 580,000 deaths. Many of these cancers involve the dysregulation of kinases, which play a central role in cellular signaling pathways. Mutations, translocations, or upregulation events can cause one or more kinases to become highly active and cease responding normally to regulatory signals. As a result, much of the effort in developing treatments for these diseases (and perhaps 30% of current drug development) has focused on shutting down aberrant kinases with targeted inhibitors.

Tyrosine kinase inhibitors (TKIs) in particular have proven themselves powerful therapeutics in the treatment of human cancers. The high selectivity of some TKIs such as imatinib—which potently inhibits just a small fraction of the human kinome to treat chronic myelogenous leukemia (CML)—is believed to be responsible for their effectiveness and low toxicity. Unfortunately, even when selective TKIs are available, the inexorable emergence of resistance mutations limits the duration over which the patient will derive therapeutic benefit, requiring a switch to second- and third-line selective TKIs—if they exist—as resistance develops. Ultimately, drug resistance is thought to be the reason for treatment failure in over 90% of patients with metastatic cancer.

The development of *new* selective kinase inhibitors remains incredibly challenging due to the fact that these inhibitors are almost universally targeted toward the ATP binding site shared by all kinases, but must bind with high affinity to only one (or a few) out of more than 500 human kinases to minimize unintended effects. While the discovery of imatinib was hailed as a breakthrough for its ability to selectively inhibit Abl over closely related kinases like Src, it came as a great surprise when the crystal structure of imatinib bound to Src was nearly identical to the Abl-bound structure. Recent evidence from experiments and simulation has suggested that a previously underappreciated contribution—the energetic cost of populating the inhibitor-bound conformation—plays a critical role in imatinib's selectivity. While this effect has only been examined in the well-studied case of Abl/Src binding to imatinib, it has the potential to be much more general. **We hypothesize that exploiting differences in the energetic cost of confining the kinase to the binding-competent conformation is a route to selectivity in targeted kinase inhibition.** Here, we ask how much conformational reorganization energy contributes to the selectivity and affinity of current noncovalent clinical kinase inhibitors to determine whether existing inhibitors (perhaps inadvertently) exploit differences in these reorganization energies to achieve selectivity, and whether this difference can be exploited to engineer new selective molecules.

We use a combined experimental and computational approach to decompose inhibitor binding affinity and selectivity into contributions from kinase reorganization and binding to individual kinase conformations:

Aim 1. Create an energetic atlas of the conformations accessible to human kinase catalytic domains. Using the Folding@home worldwide distributed computing platform, we will use massively parallel molecular simulations to map the conformational dynamics of kinase domains, generating an atlas of thermally accessible conformations and associated energetics using the Markov state model approach we originally developed to study conformational states transiently populated during protein folding. We will validate this map through the use of acrylodan labeling at locations predicted to be sensitive to ligand-induced conformational changes.

Aim 2. Quantify the contribution of kinase reorganization energy to inhibitor selectivity and affinity. We will use a novel automated platform to express a diverse panel of recombinant kinase domains and a newly developed fluorescence assay to directly measure the affinities of noncovalent FDA-approved kinase inhibitors to the entire panel. Combined with alchemical free energy calculations to individual kinase conformations, we will dissect the contribution of kinase reorganization energies and direct binding affinities using models that integrate experimental and computational data.

Aim 3. Identify opportunities to exploit differences in reorganization energies to achieve selectivity. We will validate our model by engineering mutations computationally identified to modulate inhibitor selectivities via manipulating differences in reorganization energies rather than the affinity for the inhibitor-bound conformation. In parallel, we will identify new opportunities to exploit differences in reorganization free energies between closely related kinases and between wild-type kinases and variants with clinically-identified oncogenic activating mutations.

This project will have a number of important implications for human health and our understanding of the biophysical determinants of selectivity. Structure-based drug design efforts will immediately benefit from a detailed understanding of the importance of reorganization energy and conformational energetics in determining affinity and selectivity. The release of an atlas of kinase conformations and energetics will provide new opportunities for the rational design of both ATP-competitive and allosteric kinase inhibitors. In addition, opportunities to exploit differences in reorganization energies of closely related kinases or wild-type and oncogenic forms of the kinase can present new paths to achieving higher efficacy without incurring additional off- or on-pathway toxicity.

SIGNIFICANCE

Kinases play a critical role in cellular signaling pathways. Perturbations to these pathways due to mutation, translocation, or upregulation events can cause one or more kinases to become highly active and cease responding normally to regulatory signals, often with disastrous consequences. Kinase dysregulation has been linked to a number of diseases, such as cancer, diabetes, and inflammation. **Cancer is the second leading cause of death in the United States, accounting for nearly 25% of all deaths; in 2015, over 1.7 million new cases were diagnosed, with over 580,000 deaths** [1]. 50–70% of current cancer drug discovery [2] (and perhaps 30% of *all* current drug development effort [3, 4]) has focused on shutting down aberrant kinases with targeted inhibitors.

Selective kinase inhibitors (**Figure 2B**) have already proven their potential as both therapeutics for the treatment of kinase dysregulation diseases and powerful chemical tools for probing pathway function. The discovery of imatinib (**Figure 1A**), which specifically targets the Abl kinase dysregulated in chronic myelogenous leukemia (CML) patients to abate disease progression, was transformative in revealing the enormous therapeutic potential of selective kinase inhibitors, kindling hope that this remarkable success could be recapitulated for other cancers and diseases [5]. While there are now 31 FDA-approved selective kinase inhibitors, these molecules were approved for targeting only 13 out of ∼500 human kinases, with the vast majority targeting just a handful of kinases; **the discovery of therapeutically effective inhibitors for other kinases has proven remarkably challenging.**

A primary difficulty in the design of new selective kinase inhibitors is the high similarity of active sites among the ∼500 members of the human kinome (**Figure 2A**). Because small-molecule kinase inhibitors generally target the 'druggable' ATP binding site, the functional constraints of ATP hydrolysis make the exploitation of minor differences in the ATP binding sites of active kinases difficult for current drug discovery strategies. Imperfect selectivity can lead to off-target toxicity, while on-target inhibition of non-dysregulated pathways can limit maximum tolerated doses, reducing the effectiveness of therapy. Worse yet, in cancer, even when selective kinase inhibitor therapy is available, the inexorable emergence of resistance limits the duration over which the patient will derive therapeutic benefit, requiring clinicians turn to second-line inhibitors. Unfortunately, few second-line drugs are available, and none are immune to resistance. Ultimately, over 90% of metastatic cancer patients die of the disease due to a lack of effective therapeutics [6]. **There is dire need for new selective inhibitors: for a broader variety of kinases implicated in other cancers, for oncogenic forms to minimize on-pathway toxicity, and for drug-resistant mutants that arise during therapy.**

Recently, evidence has emerged that the selectivity of imatinib for Abl over the closely related Src kinase is due in large part to a previously unappreciated contribution—the *reorganization energy* of the kinase, or energetic penalty that must be paid for the kinase to adopt a binding-compentent conformation (**Figure 1**). While imatinib achieves an enormous 4.6 kcal/mol selectivity for Abl over Src, it makes nearly identical contacts in the ATP-binding site in

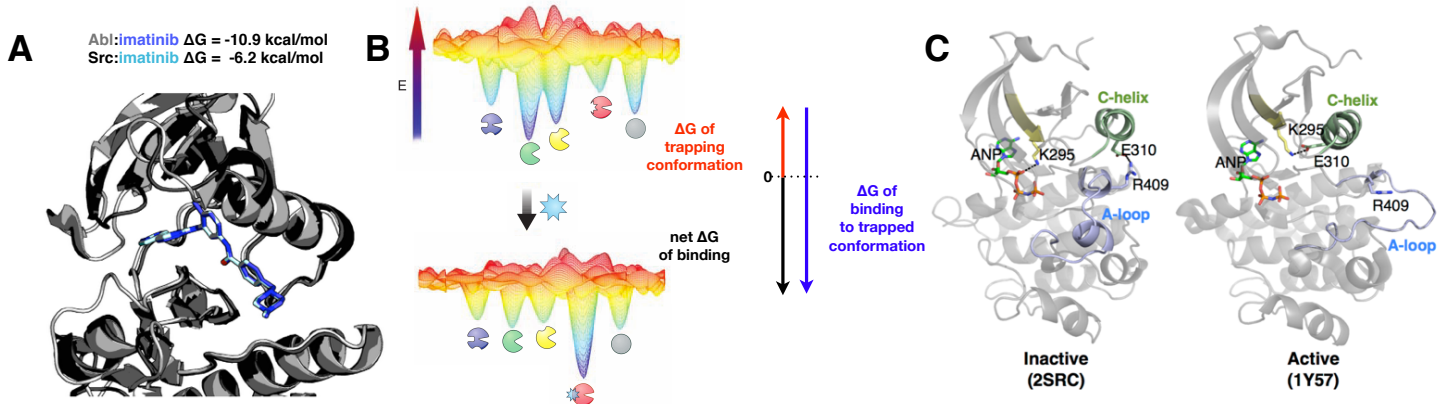


Fig. 1. **Contribution of conformational reorganization energy to selectivity of imatinib among closely related Abl and Src kinases.** *(A)* Crystal structures of imatinib bound to Abl kinase [7] (grey) and Src kinase [8] (black) show essentially identical binding contacts, suggesting the 4.6 kcal/mol selectivity difference is due to the reorganization free energy required to adopt a binding-competent conformation [8–13]. *(B)* Cartoon depiction of multiple conformational states with differing inhibitor binding competence [14]. The arrow (E) denotes the conformational *reorganization free energy* the kinase must pay to adopt a particular conformational state in relation to the lowest free energy conformation. For inhibitor (blue) to bind the the binding-competent conformation (red), an energetic penalty must be paid to confine the kinase to the binding-competent state; favorable binding free energy can then cause this conformation to become the new lowest free energy "ground state". The total binding free energy of a selective inhibitor is the sum of an unfavorable reorganization free energy for the kinase to adopt a binding-competent configuration and a favorable free energy for inhibitor binding to this conformation; inhibitor binding to multiple conformations yields a more complex expression, but is still computable given the reorganization free energies and binding affinities for individual conformations. *(C)* Crystal structures have revealed some of the distinct conformations that can be populated by Src kinase, but are unable to quantify the reorganization free energy cost for adopting that conformation.

both Abl and Src with both kinases locked in essentially identical conformations [7, 8] (**Figure 1A**) from which little selectivity can be derived; instead, a great deal of selectivity is attained from the much larger energetic cost that must be paid to put Src into an imatinib-binding conformation compared to the smaller cost Abl must pay [8–13]. The net inhibitor binding free energy is the sum of the reorganization energy and the binding free energy to that conformation (**Figure 1B**)[1]. While indirect, the aggregate of available data on the imatinib:Abl/Src system is now sufficiently compelling to suggest that this mechanism of attaining selectivity has significant merit.

How widespread is this phenomenon? Do other known selective kinase inhibitors, perhaps inadvertently, achieve their selectivity by exploiting differences in reorganization energies between closely related kinases? While unexplored outside of the well-studied imatinib:Abl/Src system, understanding the potential for exploiting differential reorganization energies to attain selectivity in other kinases could have significant implications for engineering clinically useful effective kinase inhibitors. Both off-target (binding to other kinases or biomolecules) and on-target (binding to wild-type, non-dysregulated kinase) toxicity can limit maximum tolerated dose, ultimately limiting the effectiveness of therapy and potentially resulting in an inability to achieve clear efficacy, and similarities between closely-related kinases may make it difficult to engineer selective molecules by exploiting only differences in their binding pockets; a new route to achieving selectivity among closely related kinases or between wild-type and mutant kinases that exploits characterized differences in reorganization energies could offer new paths to the development of useful selective inhibitors that minimize toxicity.

Assessing the role for reorganization energy in selectivity and affinity more broadly is complicated by lack of methodologies readily able to measure the equilibrium populations of conformations that kinase domains can adopt. Kinase inhibitors can reach picomolar affinities, so paying an energetic price of even $6k_BT$ is feasible while still attaining nanomolar affinity; observing these conformations in the *apo* kinase experimentally would require a technique capable of characterizing populations less than 0.2% populated, which is far below the level achievable by X-ray crystallography. While crystallography can observe these conformations when crystal structures of the conformations stabilized by allosteric inhibitors can be obtained, the structure alone is insufficient to deconvolute the contribution of reorganization energy inhibitor affinity. X-ray crystallographic data cataloging accessible conformations is also woefully incomplete, with a multitude of distinct structures only available for a small number of kinases [15]. While NMR techniques have shown promise—in particular, relaxation-dispersion experiments capable of quantifying minor populations provided certain conditions of conformational exchange rates and chemical shift separations are met [16]—these methods don't yet provide the atomistic detail needed to be useful for exploiting these structures for rational drug design.

We propose a combined computational and experimental approach that will allow us to decompose inhibitor binding selectivity and affinity into contributions from kinase reorganization and binding to individual kinase conformations. Collectively, the experiments proposed here will allow us to address the motivating questions above while providing new opportunities to achieve selectivity in kinase inhibition.

[1] When the inhibitor can bind to multiple conformations, the net binding free energy is $\Delta G_{bind} = -k_B T \ln \sum_n e^{-(\Delta G_n^{\mathrm{reorg}} + \Delta G_n^{\mathrm{bind}})/k_B T}$.

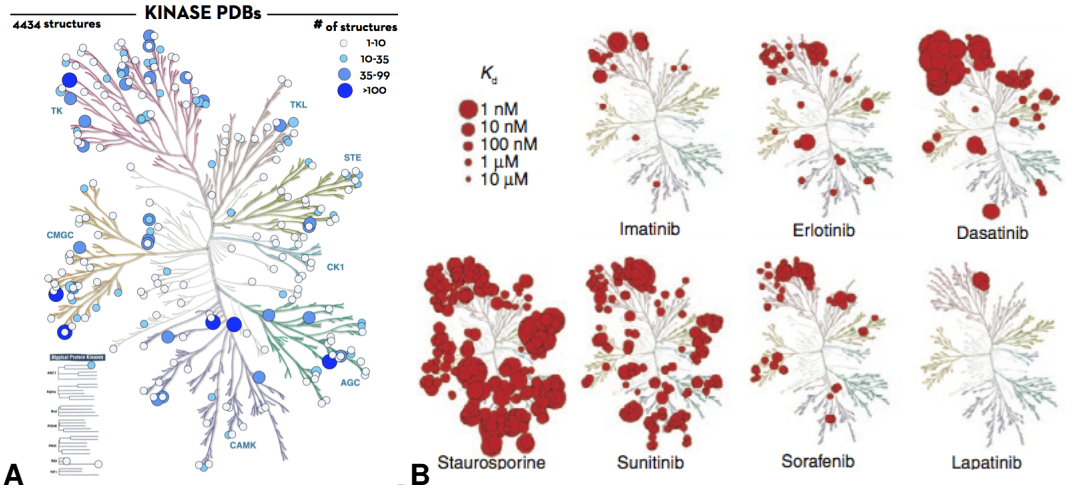


Fig. 2. **Incomplete structural coverage of kinases and diversity of selectivity profiles for small molecule kinase inhibitors. (A)** Human kinases structural data is incomplete, with few or no structures available for many kinases [15, 17]; **(B)** Kinase inhibitors may have wildly diverse selectivity profiles, and even selective inhibitors like imatinib are not exquisitely selective [18, 19].

INNOVATION

The role of conformational reorganization energy in kinase inhibitor selectivity and affinity is poorly understood. We hypothesize that the conformational reorganization energy that must be paid to confine kinases to inhibitor binding-competent conformations can be a significant contribution to the selectivity and affinity of selective kinase inhibitors (**Figure 1).** If so, an accurate accounting of reorganization energies would be essential for increasing success rates in the rational design of selective inhibitors, and exploiting differences in reorganization energies may present a novel route to attain selectivity in closely related kinases, or between oncogenic and wild-type forms.

To address this hypothesis, we propose an approach that combines advanced computational and experimental tools to allow us to decompose inhibitor binding selectivity and affinity into contributions from kinase reorganization and binding to individual conformations. A combined computational/experimental approach is currently the *only* way to address this issue given that existing biophysical techniques cannot characterize reorganization energies, inhibitor binding affinities to distinct conformations, and structures. Our broad survey will indicate whether reorganization energy currently plays—or has the opportunity to play—a significant role in attaining selectivity, with the potential to trigger a paradigm shift in how we engineer kinase inhibitors to attain desired selectivity profiles.

Elucidating human kinase domain conformations and energetics using automated superfamily-scale modeling, massively distributed molecular simulations on Folding@home, and Markov state models. Mapping the accessible conformations and energetics of a large number of kinase domains in order to determine their inhibitor-bound reorganization energies requires us to bring together a unique set of technologies. Because the vast majority of kinase domains have few available experimental structures of distinct conformations and are generally missing coordinates for mobile regions, we have built a new tool that automates the preparation of molecular simulations on the superfamily scale by modeling human kinase domain sequences onto all available distinct kinase domain structures [15]. To accumulate sufficient statistics to map the relevant transitions among highly populated *apo* and inhibitor-bound conformations, we utilize the Folding@home worldwide distributed computing network, which provides access to over 350,000 actively computing cores providing over 19 PFLOP/s of aggregate computing power—most in the form of graphics processing units (GPUs)—-making it the largest computer available for biological research [20, 21]. Finally, to distill the terabytes of trajectory data for each kinase domain into distinct conformations and associated energetics, we make use of the powerful *Markov state model* (MSM) methodology we co-developed for the study of low-population conformations in protein folding to here identify conformations and reorganization energies relevant for inhibitor binding [22–28]. The alternative kinase conformations revealed in the process—which may not have yet been uncovered by structural biology—will also present new opportunities for allosteric inhibition [14, 29].

Automated fluorescence measurements of inhibitor binding affinities to recombinantly-expressed kinases. Deconvoluting the contribution of reorganization energy from overall inhibitor binding affinity requires a method for accurately quantifying the net absolute inhibitor binding affinity for a variety of inhibitors, and rationally manipulating this contribution via mutations requires a system in which mutants can easily be engineered, expressed, and purified. To overcome limitations of standard approaches that involve mammalian or insect cell expression and unreliable inhibitor binding measurements that depend on incomplete knowledge of catalytic mechanism to extract absolute affinities, we introduce two new approaches. The first is a panel of recombinantly expressable human kinase domains engineered through a structural bioinformatics effort [30], made possible through coexpression with phosphatase [31]. This allows the rapid and facile engineering of mutants to rationally test our models by directly manipulating the reorganization energy via mutations distal to the binding site. The second is a fluorescence binding assay that reports *directly* on the binding of quinazoline- and quinoline-scaffold kinase inhibitors, developed from a a cuvette-based assay [32] to a high-throughput assay in our laboratory and expanded to also access the binding affinities of non-fluorescent inhibitors via competition assays. In order to accurately account for experimental uncertainty in measured affinities—which is poorly characterized in literature reports—we use a novel Bayesian inference analysis scheme, which allows for meaningful reconciliation with other data.

GPU-accelerated alchemical free energy calculations of inhibitor binding to distinct kinase conformations. We utilize alchemical free energy calculations to compute absolute binding free energies of kinase inhibitors to individual kinase configurations and kinase conformation population changes in response to point mutations. While alchemical free energy methods have been around since the mid-1980s, only recently have they reached useful accuracy in blind tests [33]. These methods use molecular mechanics forcefields to rigorously compute binding affinities that include all relevant statistical mechanical effects, including entropic and enthalpic contributions. To make these techniques fast enough for practical use, we have built a new free and open-source GPU-accelerated code YANK [http://github.com/choderalab/yank] [34], which combines numerous algorithmic improvements we have previously developed to achieve quantitative accuracy [23, 24, 34–40] into a single code.

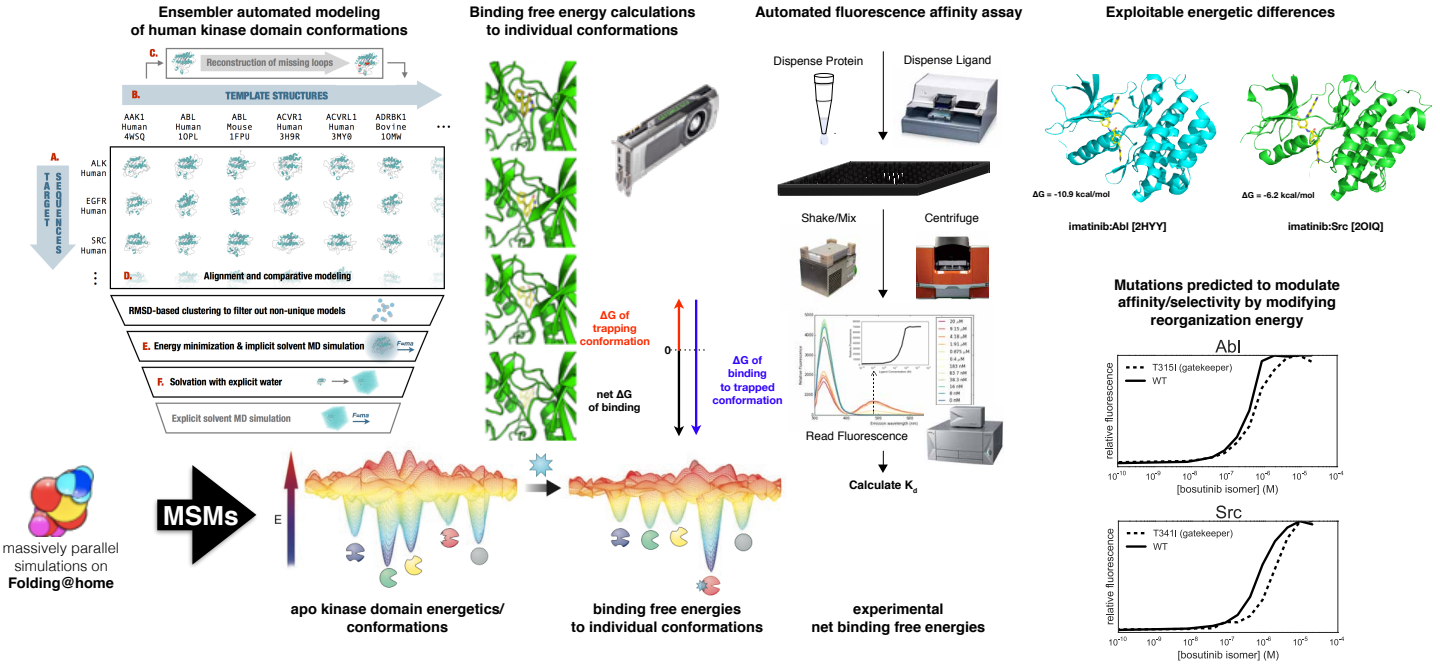
APPROACH

The clinical kinase inhibitor imatinib and related second-generation tyrosine kinase inhibitors have provided crucial insight into the origin of kinase inhibitor selectivity, in addition to clues as to why the design of selective inhibitors has proven so challenging. Imatinib binds tightly to Abl kinase—the kinase that becomes constitutively activated in CML patients due to a gene translocation event [41]—but binds 4.6 kcal/mol more weakly to the closely related Src kinase [8]. Surprisingly, imatinib has been cocrystallized with both Abl and Src in essentially identical conformations in which the drug makes essentially identical contacts in each kinase (**Figure 1A**) [8, 41]. Evidence from crystal structures, thermodynamic measurements, and computer simulations has hinted that this high selectivity is due to the difference in energetic costs for Src or Abl to adopt a binding-competent conformation [8–13], but the scope of this effect in selective kinase inhibition in general is unknown.

We hypothesize that conformational selection represents a general mechanism for achieving selectivity. This would mean that accurate accounting for the reorganization energy of confining a kinase to an inhibitor binding-competent state (**Figure 1B**) is critical for the design of selective inhibitors. This hypothesis immediately raises several quantitative questions: To what degree is inhibitor selectivity due to differences in reorganization energies for kinases to adopt binding-competent conformations, rather than differences in direct interactions with the kinase once this conformation is adopted? How much does this vary among inhibitors? Could quantitative models incorporating reorganization free energies allow us to design new inhibitors with desired selectivity profiles?

To address these questions, we require (1) a way to measure inhibitor binding affinities to a diverse set of well-behaved, manipulable kinase systems; (2) an approach to identify distinct kinase conformations and their equilibrium populations (or conformational free energies) in the absence of inhibitor; and (3) a way to quantify the inhibitor binding affinity to each distinct conformation. To further validate quantitative hypotheses about how much of the free energy of binding is due to these conformational reorganization energies, we will design and test mutations predicted to modulate the affinity via the reorganization energy.

Our approach combines computation and experiment to develop a quantitative, predictive understanding of small molecule affinity and selectivity for wild type and mutant forms of human kinases. While this proposal does not focus on the *development* of new kinase inhibitors, we construct a framework that will enable the rational *design* of new inhibitors with desired selectivity profiles.



AIM 1: Create energetic atlas of human kinase domain conformations

AIM 2: Quantify contribution of reorganization energy to inhibitor selectivity/affinity

AIM 3: Identify opportunities to exploit reorganization energies / mutational validation

Aim 1. Create an energetic atlas of the conformations accessible to human kinase catalytic domains.

Overview. To quantify the role of kinase conformational reorganization energy in the selectivity and affinity of kinase inhibitors in the broader context of the ∼500-member human kinome (**Figure 2A**), we adopt a computational strategy to map thermally accessible conformations and their associated energetics with large-scale distributed molecular simulations. We focus on the kinase catalytic domains—the molecular targets for selective kinase inhibitors—starting with a set of 51 for which we have demonstrated bacterial expression [30], and expanding this set as more constructs are engineered as part of **Aim 2**. We emphasize the importance of using the *same* kinase domain sequence constructs for both simulation and experiment to minimize discrepancies. After extracting kinetically metastable conformations and their energetics, we validate the the computational models by comparing to available NMR data. In addition to producing useful data on its own, this Aim will provide data for Aim 2 to dissect the contribution from reorganization energy to binding and for Aim 3 to identify opportunities to exploit differences in reorganization energy as well as validate our models via point mutations that modulate the reorganization energy.

Subaim 1.1 *Simulate a panel of human kinase domains on Folding@home starting from multiple conformations*

Approach: In order to compute the reorganization energies (free energy cost of confining the kinase to various inhibitor-bound conformations), we will computationally map thermodynamically relevant conformations of a panel of *apo* kinases. Specifically, we will construct *Markov state models* (MSMs) of conformational dynamics from large-scale molecular simulation data [22–28]. These models aggregate molecular dynamics simulation trajectory data to identify *kinetically metastable* regions of conformation space that correspond to distinct conformations with long kinetic lifetimes (∼100 ns), as well as interconversion rate constants and corresponding equilibrium populations (or free energies). This approach has been successful in identifying conformations and intermediates relevant to protein folding [24], connecting simulations to biophysical experiments [27], and more recently, monitoring structural transitions in kinases [29], making it a natural choice to for this approach. Our panel will initially consist of 51 kinases we have expressed in bacteria [30], and will expand as **Aim 2** adds more bacterially expressible constructs to our panel. Kinase domains will initially be unphosphorylated to match expression conditions, but subsequent work will examine the influence of phosphorylation on reorganization energies and binding affinities.

To simulate the entire human kinome, we will use the Folding@home distributed computing platform. Folding@home (FAH) [http://folding.stanford.edu] is a worldwide distributed computing project where donors worldwide run a client program to contribute their computing power for biomolecular simulation. With a current aggregate sustained throughput of 19 PFLOP/s from over 350,000 active computing cores, FAH is the most powerful distributed computing platform in the world. Our laboratory—a member of the FAH consortium—will utilize the OpenMM GPU-accelerated Folding@home simulation core [42] to generate simulation data starting from Ensembler-generated initial simulation conformations for the kinases in our panel. Typical consumer GPUs on FAH produce ∼40 ns/day of simulation data using current best practice protocols (Langevin dynamics with 2 fs timestep, weak 5/ps collision rate, particle-mesh Ewald [PME] electrostatics, long-range dispersion correction, and Monte Carlo barostat), with typical aggregate throughput on the order of ∼50 μs/day/kinase. Simulations will initially focus on Src and Abl to optimize use of Ensembler models, followed by kinase expression panel members within the tyrosine kinase family, followed by all kinases in the expression panel. Initial conformation generation, simulations, and analysis will proceed individually by related groups of kinases in a staggered fashion.

Many of the large number of kinase constructs we must simulate lack structural information, or have structures for only one conformation. For this reason, we have developed an automated simulation preparation pipeline termed to produce initial models for simulation and characterization (**Figure 3**) called Ensembler [15]. To ensure we do not miss relevant ligand binding-competent conformations that may have been observed only in structures of related kinases, Ensembler uses a robust distributed computing framework to model all human kinase domain sequences of interest onto *every available kinase structure from any organism*, sorted by sequence identity. Poor models are culled by subsequent rounds of more detailed implicit and explicit solvent simulations. The remaining models are equilibrated and prepared for distributed simulation on FAH automatically. Poor initial models that represent irrelevant conformations will subsequently be automatically terminated by simulation instability or culled by ergodic subspace trimming [43] during MSM construction with MSMBuilder [44, 45] and pyEMMA [46]. All models will be made available online—tyrosine kinase models are already available via Dryad.

Preliminary data: As a proof of concept, we have applied this automated modeling and simulation pipeline to the catalytic domain of human Src and Abl kinases using the AMBER99SB-ILDN forcefield with TIP3P water (**Figure 3**) [15]. The procedure generated a total of 4003 distinct conformations for Src and 4280 conformations for Abl that survived all stages of the pipeline. **In a short pilot project of approximately one month, we collected over 1.5 *milliseconds* of aggregate simulation data for Src on Folding@Home**; we expect much less data per kinase will be needed after further optimization.

Potential Problems and Alternative Strategies: As a proof of concept, we modeled all human tyrosine kinases (TKs) using Ensembler (**Figure 3**) [15]. While it is possible we may encounter unexpected difficulties in extending this procedure other kinase families, the worst-case scenario is that we confine our study to tyrosine kinase family members for which we have expressing constructs (currently 10 with expression >2 μg/mL, though this number is expected to increase). We do not yet know what fraction of Ensembler-generated models will end up contributing to the maximum connected subspace following ergodic subspace trimming [43] during subsequent MSM construction; should pilot experiments find conformations from remote-homology templates are predominantly culled, we will focus simulation effort on models derived from templates above a sequence identity cutoff. We plan to utilize recent AMBER forcefields for these simulations, though initial control experiments on kinsaes with available NMR data will compare several AMBER and CHARMM forcefields to ensure there is sufficient concordance with experiment.

Subaim 1.2 *Extract long-lived conformations and reorganization free energies using Markov state models*

Approach: In order to extract kinetically metastable (long-lived) conformations with lifetimes τ $>$100 ns and their associated reorganization free energies (**Figure 2**), we will construct *Markov state models* (MSMs) [22–28] from the accumulated simulation data generated by Folding@home in **Aim 1.1**. We will make use of the MSM construction tools MSMBuilder [44, 45] and pyEMMA [46]. We will generalize an approach that has been recently found to be successful for Src kinase [29] which focused on residue contacts known to be associated with kinase activation by automatically identifying interresidue distances associated with residue contacts that change during the course of the the aggregate simulation data. Ergodic subspace trimming [43] will prune away disconnected regions of conformation space arising from simulations initiated from poor initial models that failed to interconvert with the

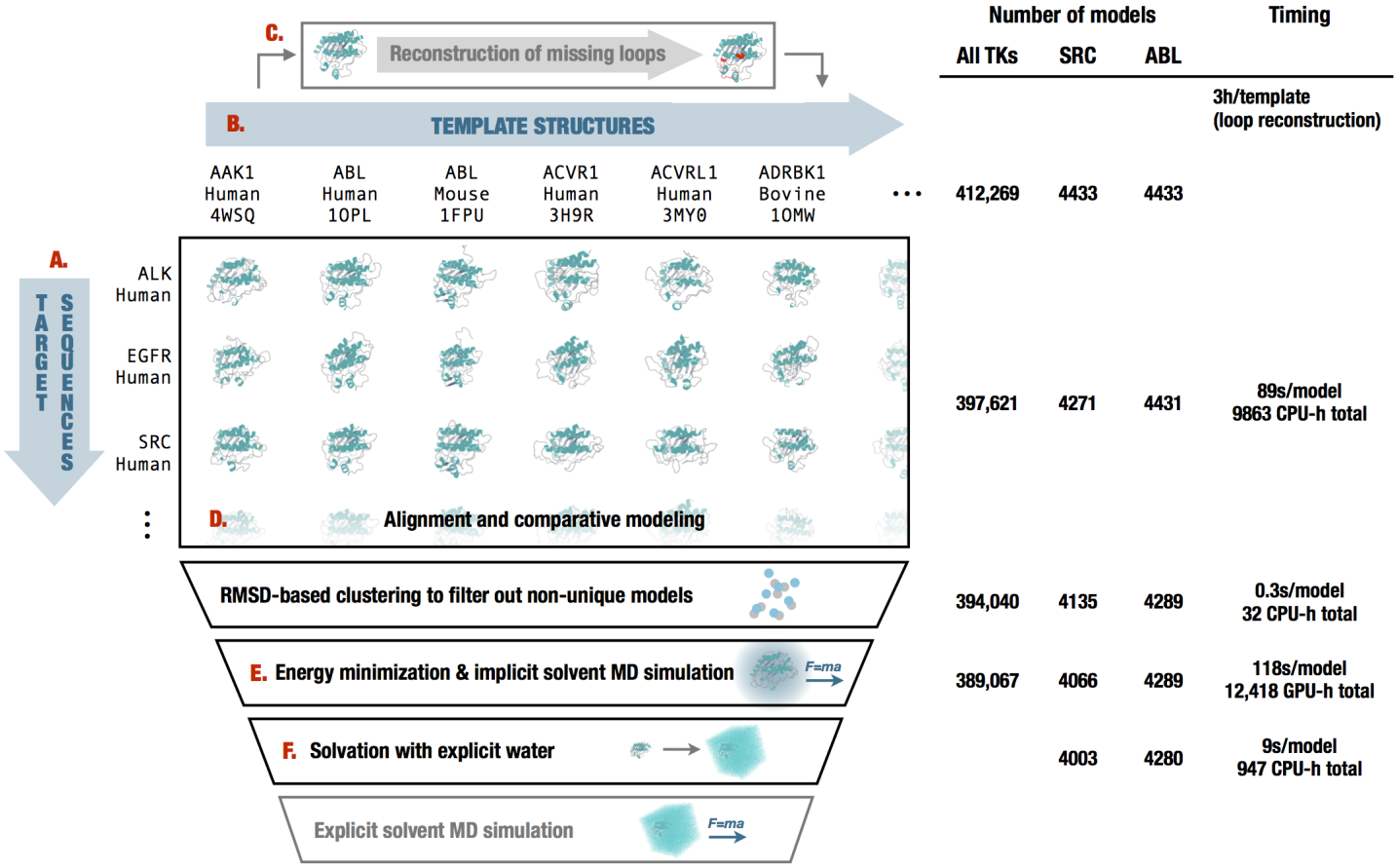


Fig. 3. **Automated modeling workflow from *Ensembler* illustrating how kinase domain simulations are automatically prepared.** In order to map the variety of accessible kinase conformations for a diverse panel of human kinase domains, we must automate the preparation of kinase structures for subsequent simulation in a variety of plausible conformations. Comparative models are generated for target sequences (human kinase domain constructs) using kinase domains from any organism (prioritized by sequence identity) as template structures. Since many kinases are missing large loops, template structures first have missing loops regenerated in plausible conformations using ROSETTA loopmodel [47, 48]. Distinct conformations are selected via RMSD-based clustering, energy minimized, solvated using an algorithm to select optimal box size, and equilibrated prior to deployment on Folding@home for distributed molecular simulations prior to the generation of Markov state models from which conformations and free energies are extracted. Statistics and computational effort for modeling Src, Abl, and the entire tyrosine kinase (TK) family are shown at right.

rest of the conformation space. Kinetic lumping will utilize whichever form of metastability maximization [23], PCCA+ [49], Bayesian hidden Markov models [50], or Bayesian agglomerative clustering [51] is found to produce the most robust models, as judged by bootstrapping over subsets of trajectories.

This procedure will yield what we will refer to as distinct *accessible conformations*, which will consist of clustered configurations that correspond to kinetically metastable conformational states of each kinase. Because downstream use of these conformations will involve sampling within these kinetically metastable conformations to compute inhibitor binding affinities to individual conformations (**Aim 2.3**), we require these conformations to have internal kinetic barriers that can be overcome within a timescale of $\tau \ll 100$ ns, so we utilize a cutoff lifetime of $\tau \sim 100$ ns for determining the number of kinetically distinct kinetically conformations. We will use RMSD to identify which conformations existing cocrystal kinase:ligand structures correspond to in order to identify conformations likely competent for binding kinase inhibitors; in **Aim 2**, all conformations less than $6k_BT$ in reorganization energy penalty will be subjected to binding free energy calculations[2]. Each conformational state will also have an associated reorganization free energy that reflects the free energy cost that must be paid to confine the kinase to this conformational state (with corresponding statistical uncertainties estimated using Bayesian methods [25, 52]).

Several questions can immediately be addressed: How different are the accessible ($<6k_BT$) conformations of related and unrelated kinases? Are available crystal structures representative of conformations we expect could be stabilized by ligand binding, or is more conformational diversity present than previously appreciated? This kinome structural and energetic atlas will provide the critical tools for probing the origin of inhibitor selectivity in **Aim 2** and opportunities for allosteric modulation in **Aim 3** as simulations and Markov state models are produced.

Potential Problems and Alternative Strategies: It is unlikely (but possible) that no transitions will be seen between the modeled inactive and active conformational states even in 10,000 distributed Folding@Home simulations running for one month of wall clock time. If necessary, more complex enhanced sampling techniques such as adaptive seeding [44] or temperature reweighting [53, 54] can be used should convergence prove difficult.

Subaim 1.3 *Validate Markov state models against existing NMR data*

Approach: While **Aim 2** will utilize a combination of experimental inhibitor binding affinity measurements (which probes the aggregate binding to *all* binding-competent conformations) and alchemical binding free energies to individual conformations to dissect the contribution of reorganization energy to inhibitor selectivity and affinity, it is useful to independently validate the conformations and associated reorganization energies extracted from the Markov state models (MSMs) to ensure dominant conformational states are accurately described. Fortunately, solution NMR data represents an excellent way to reconcile MSMs with biophysical experiment, with a long history of comparisons between these techniques [55–57]. While many of these observables (such as chemical shifts and ^{3}J-coupling constants) are primarily population-weighted averages of properties from individual conformations (and hence sensitive to high-population conformations), certain NMR observables (such as NOE peak volumes, which are exquisitely sensitive to even small populations for short interproton differences) are highly nonlinear and provide useful validation that low-population conformations are faithfully captured.

A wide variety of tools now exist to easily compute various NMR observables, including chemical shifts, ^{3}J-coupling constants, and NOE peak volumes. A diverse battery of these observables were recently used to assess the quality of a variety of biomolecular forcefields on a diverse set of 524 NMR measurements using an automated toolkit that is now being developed by our laboratory [https://github.com/choderalab/TrustButVerify] [58]. A wealth of solution NMR data for kinases exists both in the Biological Magnetic Resonance Data Bank (BMRB) [59] and in a number of publications (e.g. [13], where data is available over a range of imatinib concentrations, and [60], where a number of inhibitor-bound spectra are reported).

We will use our automated NMR comparison suite to carry out an exhaustive benchmark of computed NMR observables against available NMR data for the kinases for which we generate MSMs in **Aim 1.2**. We will focus first on the small subset of kinases for which we compare multiple forcefields, since this will allow us to use experimental data to aid selection of the best forcefield for subsequent kinase MSM generation. When NMR data in the presence of various concentrations of inhibitor are available for these kinases (such as [13]), we will also combine this with the ligand-bound conformational structures and binding affinities computed in **Aim 2.3** to determine the concentration-dependent shift in computed NMR properties. The result will be an extensive benchmark between molecular simulation data and available NMR data of kinases with facile bacterial expression, which will serve to select optimal forcefields for kinase modeling, gauge the trustworthiness of MSM-derived conformations on a per-kinase basis, and identify opportunities for additional NMR investigations of kinases.

[2] A $6k_BT$ reorganization free energy penalty corresponds to a 500-fold loss in binding affinity. Because kinase inhibitors can have up to picomolar affinity, a 500-fold loss could in principle still yield sub-nanomolar affinity.

Potential Problems and Alternative Strategies: While chemical shift prediction has improved greatly over the last two decades, there are still limitations in its ability to quantitatively predict experimental data [58]. We will mitigate this by looking for concordance between chemical shift prediction tools as an indicator of expected accuracy. Exact protein constructs used in simulations may differ from NMR experiments; we will prioritize constructs with minimal deviations. Some NMR data may be for phosphorylated forms of kinases, in which case we will utilize relative free energy calculations to account for phosphorylation-induced changes in conformational state populations and dynamics. If NOEs prove inaccurate for quantitative comparison because of dynamical effects, we will include relaxation effects from conformational exchange [57].

Aim 2. Quantify the contribution of kinase reorganization energy to inhibitor selectivity and affinity.

Overview. To quantitatively assess the contribution of conformational reorganization energy (**Figure 1B**) to kinase inhibitor selectivity and affinity across a broad range of inhibitors, we need two of three components for each kinase:inhibitor pair: (1) the conformational reorganization energies for binding-competent conformations, (2) the net overall binding affinity of the inhibitor to the kinase, and (3) the binding affinities of the inhibitor to individual conformations. We use a combination of computation and experiment to quantify *all three* components, overdetermining the problem to ensure that our investigation will not be dependent on the failure of one of these techniques. Having quantified the reorganization energies (1) independently in **Aim 1**, in parallel, we measure net binding affinities (2) of a panel of noncovalent kinase inhibitors to a diverse set of bacterially-expressed kinases using a new highly accurate fluorescence assay with high dynamic range, as well as compute the binding affinities to individual conformations (3) using state-of-the-art GPU-accelerated alchemical free energy calculations. The result—in addition to a wealth of new experimental binding data—is the first large-scale survey quantifying the degree that conformational reorganization energies are responsible for inhibitor selectivity, as well as an assessment of reorganization energy costs that must be paid for inhibitors to bind their intended targets.

Subaim 2.1 *Express a set of kinase domains in a high-throughput robotic expression protocol*

It is first necessary to establish a diverse panel of well-behaved human kinases that express well in a high-throughput system that allows for the subsequent introduction of specific mutations to perturb conformational energetics (**Aim 3.3**) and assay of changes in binding affinity against a panel of small molecule kinase inhibitors (**Aim 2.2**). While human kinase expression in baculovirus-infected insect cells have demonstrated good success rates [61, 62], this cannot compete in cost or convenience with bacterial expression, and a survey of 62 full-length non-receptor human kinases found that over 50% express well in *E. coli* [61]. Truncation to the soluble kinase domain alone simplifies biochemical manipulation and computational modeling, retains the inhibitor-targeted domain, and facilitates the study of receptor-type kinases that would otherwise be membrane-associated.

The advent of robot-assisted protocols for automated cloning, site-directed mutagenesis, and expression in 96-well formats has transformed the productivity of many academic research laboratories, including our own. We have built a fully automated biophysical workstation [http://choderalab.org/resources] that is capable of automated cloning, site-directed mutagenesis, bacterial expression in 96-well 1 mL culture formats, purification, and subsequent quantitative fluorescence binding assays to accurately measure inhibitor binding affinity (**Aim 2.2**).

We will to produce a set of recombinant kinase domain constructs for automated bacterial expression. We will clone into vectors appropriate for high-throughput autoinducible expression, attaching N-terminal His-tags with TEV cleavage sites. Kinases will be prioritized by relevance to human disease, inhibitor binding, availability of structural data, and known bacterial expression. Construct boundaries will be identified from existing constructs from the PDB, previous literature efforts, transferred from homologous kinases, or determined algorithmically based on sequence. Coexpression with phosphatase (YopH for Tyr kinases, lambda for Ser/Thr) will enhance expression and ensure kinases remain unphosphorylated [31], which will be confirmed by mass spectrometry. If necessary, N-terminal fusions with solubility-promoting domains will be made, or genes resynthesized to optimize codon usage. Yield and purity will be assayed by microfluidic gel electrophoresis, analytical gel filtration will ensure constructs are monomeric, and thermostability in working buffers will be assessed by thermal melts in the presence of a dye that binds unfolded protein (ThermoFluor) [63].

Preliminary Data: As a proof of concept that an automated robotic expression pipeline can produce sufficient quantities of kinase for our assays, we cloned 96 human kinase domain constructs into a standard bacterial expression vector using domain boundaries known to have been expressed in bacteria through any protocol from header entries in the protein databank. Expression was carried out in a 96-well 1 mL deep-block plate with 900 μL Rosetta2 cells from Novagen. A total of 52 of these kinase domains (including many kinases implicated in cancer) were found to have expression yields $>$2 μg/mL culture, 35 of which were $>$10 μg/mL, which is more than sufficient for our automated fluorescence binding affinity measurements [30]. [Full expression results are available online at http://choderalab.org/kinome-expression.]

Potential Problems and Alternative Strategies: In the worst-case scenario, our cloning and expression efforts will not uncover additional recombinantly expressable kinase domain constructs, leaving us with a diverse panel of 52 kinases [30]. This set can be expanded in cases where we will not need to engineer mutations by using large-scale protein preps of several liters to increase the number of kinases in the set to any of those for which we find detectable, but not high, expression. We will continue to monitor the PDB for deposition of new kinase structures with reported bacterial expression using the Expression Explorer tool we have developed [https://github.com/choderalab/ExpressionExplorer]. Because our fluorescence assay (**Aim 2.2**) is not dependent on enzymatic activity, we may also find that inactivating key catalytic residues might expand our kinase set while minimally perturbing inhibitor affinities for study.

Subaim 2.2 *Measure inhibitor binding affinities to a panel of recombinantly expressed kinases*

Approach: To deconvolve the contribution of reorganization energy to selectivity and affinity, we require a high-throughput assay that can provide accurate assessments of kinase inhibitor binding affinities across a large range of affinities to a diverse set of kinases using only μg-scale quantities of recombinant kinases. While public binding affinity databases contain numerous kinase inhibitor affinities, they are unsuitable for our purposes due to both their unreliability [64] and their incompleteness, as well as and our need to make new measurements on mutated kinases (**Aim 3.3**). ATP-competitive kinase inhibitors based on quinazoline or quinoline scaffolds (which includes a number of drugs and drug candidates) undergo a large increase in fluorescence upon binding [32], and we have used this property to develop a general label-free approach to measuring the binding affinities of small molecule kinase inhibitors to kinases in an automated assay either directly or via competition, requiring only a few μg of kinase per measurement (**Figure 4**). These molecules are are inexpensive, available in gram quantities, bind with a range of affinities to the ATP binding site, and their affinities can be *directly* measured over a wide dynamic range (nM–mM) by titration into kinase. Since other selective inhibitors not based on these scaffolds are also ATP-competitive, their affinities can readily be measured via competition assays (**Figure 5**).

We make use of the fact that all available ATP-competitive kinase inhibitors will bind to *every* kinase—only varying in the magnitude of their affinity [18, 19]. **We propose to use quanzoline- and quinoline-scaffold inhibitors as fluorescent probes in an inexpensive high-throughput assay to directly measure binding affinities.** We will scale up this assay [32] to microplate format (96- or 384-well), using titration of increasing concentrations of fluorescent probe compounds to measure the affinities of either the probe compound or non-fluorescent kinase inhibitors via competition (**Figure 5**). We will perform measurements on the entire set of 28 FDA-approved noncovalent kinase inhibitors, in addition to some compounds of clinical interest that are readily available.

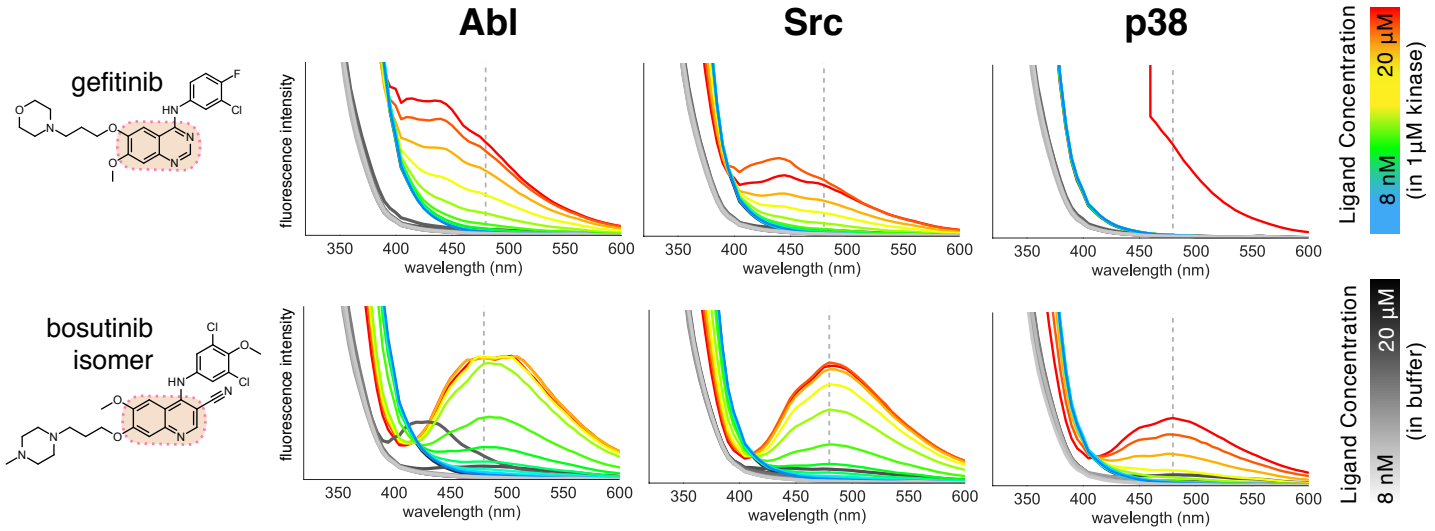


Fig. 4. **Automated high-accuracy fluorescence assay for measuring direct binding of fluorescent kinase inhibitor probes.** By using quinazoline or quinoline ATP-competitive kinase inhibitor probes that greatly increase fluorescence upon binding, we can rapidly and directly measure the affinity of the probes over a very large dynamic range (nM–mM) or measure other inhibitor affinities by competition assays. While this phenomenon was initially described using a cuvette-based assay [32], we have developed it into an automated high-throughput assay. *Left:* Gefitinib (quinazoline scaffold) and an isomer of bosutinib (quinoline scaffold), just two of a large number of fluorescent probes that can be used for affinity measurement in this assay. *Right:* Fluorescence emission spectra (excitation at 280 nm) for different concentrations of probe compound shows large increase in fluorescence upon binding to kinase (1 μM in 100 μL volume in 96-well plates). The dashed vertical line (480 nm) indicates emission wavelength used for high-throughput form of this assay, which allows even weak (mM) binding to be quantified. Numerous quinazoline- and quinoline-scaffold kinase inhibitors are usable as probes.

Preliminary Data: A variety of quinazoline and quinoline scaffold ATP-competitive inhibitors exhibit useful fluorescence changes in a plate format assay (two are shown in **Figure 4**). Excitation near 350 nm gives a large increase of fluorescence emission on binding, while excitation at 280 nm to excite kinase Trp/Tyr yields superior signal-to-noise ratio, presumably due to FRET [Nicholas Levinson, private communication]. Bayesian analysis of single-wavelength fluorescence measurements (excitation 280 nm / emission 480 nm) allows accurate quantification of inhibitor affinity confidence intervals across at least six orders of magnitude of K_d [https://github.com/choderalab/assaytools]. Competition assays can be used to access the binding affinities of non-fluorescent ATP-competitive inhibitors (**Figure 5**). Assay modeling suggests further optimization can ensure robust, accurate results over a dynamic range of nM–mM with only 1–2 μg kinase/meaurement; excellent results were obtained by Levinson in a low-throughput fluorimeter assay using only 200 ng kinase/measurement [32].

Potential Problems and Alternative Strategies: While the dynamic range of the assay is fundamentally limited at the low end by minimal detectable fluorescence and at the high end by the solubility limit of fluorescent probes (generally $\sim$50 μM), competition assays can extend this dynamic range for non-fluorescent inhibitors. Should assay design prove problematic, we can use Bayesian experimental design to automate optimization of assay concentrations [65]. If a single probe may not provide the desired dynamic range of affinities for some kinases, we can utilize a multitude of probes and combine their measurements to enhance dynamic range. Fluorescence interference may occur with some non-probe compounds if the unbound inhibitor absorbs in the emission band of the kinase:probe complex; in this case, we can monitor the emission at multiple wavelengths, use a probe compound with a different emission wavelength, or omit these inhibitors from the set. Analysis models will account for effects such as the inner filter effect at high concentrations. As the phosphorylation state can have a large effect on affinity for some inhibitors, we will also attempt to measure affinity for phosphorylated kinase by allowing kinases to autophosphorylate at high concentration; phosphorylation will be confirmed by mass spectrometry.

Subaim 2.3 *Compute alchemical binding free energies of inhibitors to individual kinase domain conformations*

Approach:

To quantify the degree to which conformational selection and kinase reorganization energies determine inhibitor selectivity, we must be able to dissect the overall inhibitor binding free energy into contributions arising from trapping the kinase in a binding competent conformation and subsequently binding the ligand once trapped in this conformation (**Figure 1**). We note that multiple binding-competent conformations with differing affinities may exist, and have developed a general statistical mechanical framework for this [36]. Together with either reorganization energies from **Aim 1** or net binding affinities from **Aim 2.2**, we can use binding free energies to individual conformations to determine to what degree selectivity is achieved via conformational reorganization energy. **Alchemical free energy calculations provide a rigorous approach to computing binding affinities to trapped conformations at modest computational cost** [33, 34, 66–70]. These approaches use molecular simulations of a series of alchemically perturbed intermediate states that 'decouple' the small molecule from the rest of the system, allowing the total free energy of binding (which includes all entropic and enthalpic effects) to be efficiently computed. Using short GPU-accelerated binding free energy calculations on local resources, we will compute the binding affinity from each inhibitor in our set to each distinct low-lying ($<6k_BT$) kinase conformational state identified in **Aim 1**. Because we have used Markov state models to identify distinct conformations with long lifetimes, these calculations only need to be long enough to overcome internal kinetic barriers within these conformations ($\ll$100 ns). We will use an alchemical Hamiltonian-exchange protocol that will allow the ligand to automatically find appropriate binding modes in the ATP binding site [34].

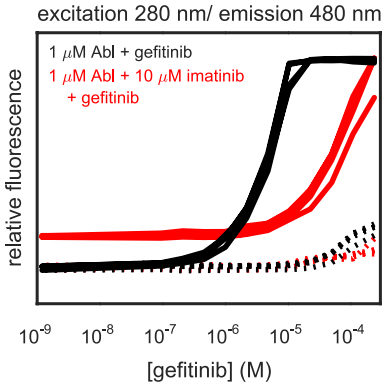


Fig. 5. **Inhibitor affinity can be measured by competition with fluorescent probes.** The fluorescence assay described in **Figure 4** can access affinities of non-fluorescence kinase inhibitors (imatinib) by competition with an already-characterized fluorescent probe compound (gefitinib). Bold lines denote the presence of 1 μM kinase; dashed lines denote the absence of kinase.

Potential Problems and Alternative Strategies: Discrepancies between computed and measured net affinities beyond what can be explained by rigorous accounting for statistical error can be attributed to three sources: (1) incomplete conformational sampling (due to missed conformational states); (2) deficiencies in the treatment of chemical effects such as protonation state changes (in either the protein [71] or the ligand [72]), tautomerization [73], or treatment of the salt environment or essential cofactors; and (3) deficiencies in the forcefield. While a detailed dissection of all of these effects is beyond the scope of this proposal, automated sensitivity analysis can suggest which of these effects may be predominantly responsible for discrepancies in instances where they arise.

Subaim 2.4 *Validate models with acrylodan labeling experiments*

Approach: Recently, site-directed acrylodan labeling [74] has been used in kinases as a sensitive reporter of ligand-induced kinase conformational changes, a technique termed 'FLiK' (Fluorescent Labeling in Kinases) [75–77]. Specific FLiK labeling sites have been used to discriminate between known classes of allosteric inhibitors—generally classified into Type I, II, III, or IV [75–77]. Originally intended to aid in high-throughput screening assays for allosteric inhibitors, we can use these labeling techniques to validate our kinase conformation and energetic atlas. By additionally computing the conformation-dependent acrylodan fluorescence using a simple biophysical fluorescence model (e.g. [78]), we can compute the equilibrium acrylodan fluorescence as a function of ligand concentration. This can be compared with experimental measurements (where acrylodan is used to label site-directed Cys surface mutants in Cys-light backgrounds) to report on specific conformational changes. Computational analysis of conformations identified in **Aim 1** will determine optimal labeling sites to report on binding-induced conformational stabilization.

Potential Problems and Alternative Strategies: It is possible that we may find it difficult to engineer Cys-light background mutations by eliminating existing Cys residues. Fortunately, kinase domains generally have few or no Cys residues; 66% of our currently-expressing panel has no Cys residues, and the remainder only possessed a single Cys, with the exception of one kinase that had two. We can mitigate the potential for success by attempting to engineer Cys residues into multiple sites and ensuring that the kinases have unperturbed binding affinities as a result of both the introduction of the Cys and labeling with acrylodan.

Aim 3. Identify opportunities to exploit differences in reorganization energies to achieve selectivity.

The data from **Aim 1** and **Aim 2** provides rich opportunities for extracting additional insight. We will use the collective data to identify opportunities to exploit differences in reorganization energy to develop new selective ATP-competitive inhibitors that take advantage of specific conformational states that may present similar binding pockets but require different reorganization energies to access in different kinases. We will also look for additional opportunities to exploit cryptic pockets that are revealed by our conformational and energetic atlas of the kinase panel that have not yet been revealed by structural biology. Finally, we will demonstrate our understanding of the importance of reorganization energies by engineering kinase mutations distal to the binding site designed to modulate affinity by manipulating reorganization energies.

Subaim 3.1 *Categorize ATP binding sites in accessible kinase conformations by their potential for achieving selectivity through reorganization free energy differences*

Approach: We will assess the *potential* for exploiting differences in reorganization energies between kinases as a mechanism to attain selective inhibition. By using a pharmacophore-like similarity metric to compare the similarity of ATP binding sites from all accessible conformations of individual kinases obtained in **Aim 1.2**, we can cluster these to identify binding sites that are relatively unique and sites that appear to be highly similar in multiple kinases. While unique binding sites exposed in particular accessible conformations provide opportunities for achieving selectivity by directly exploiting differences in binding site interactions, we are particularly interested in whether sites that are similar in multiple kinases provide an opportunity to achieve selectivity via differences in the reorganization energy to access the conformation(s) that the binding pocket is available in. Each pocket will be assigned a *reorganization energy selectivity potential* score based on the difference in reorganization energies among kinases that share that pocket: If this score is positive, there is potential to achieve some degree of selectivity due to reorganization energy by designing an inhibitor to bind to that pocket, with the score indicating how much selectivity energy is available. If the score is zero, there is no potential for attaining selectivity via differences in reorganization energies. Large negative scores would indicate this pocket is a poor choice to target for designing selective inhibitors because differences in reorganization energy actually *oppose* selectivity.

Because other considerations (such as expression levels in various tissues) also factor into this consideration, we will build an online tool for exploring these pockets and reorganization energies for specified subsets of kinases. While actively designing, synthesizing, and assaying new small molecules is outside the scope of this proposal, this resource will aid other laboratories in their efforts to achieve selectivity by providing structures against which virtual screens or rational design could be performed.

Potential Problems and Alternative Strategies: Defining a sensible binding site similarity metric is critical to being able to distinguish opportunities for exploiting differences in the ATP binding sites. Fortunately, a number of similarity metrics exist [79–81], and the availability of a wealth of crystallographic data in which the same ligand appears in multiple kinases can help ascertain whether the similarity metric is capable of identifying binding pockets that can accommodate similar ligands.

Subaim 3.2 *Identify opportunities for developing non-ATP-competitive allosteric modulators*

Approach: The conformations and reorganization energies extracted in **Aims 1.2** present a wealth of useful useful information for the subsequent design of new kinase inhibitors. One potential use of these conformations is the design of allosteric inhibitors which may bind in locations outside the ATP binding pocket targeted by most kinase inhibitors. There has been enormous recent interest in the potential these inhibitors have for achieving selectivity by being unconstrained by the requirements of ATP-competitive binding. The recent discovery of GNF-2 [82]—a selective allosteric Abl inhibitor that binds to the myristate binding site—has also generated great excitement over the potential therapeutic route of combining allosteric and ATP-competitive inhibitors to overcome resistance [83].

Following on the success of this strategy for identifying cryptic allosteric sites in β-lactamase [84], we will examine the low-lying (up to $\sim 6 k_B T$ of reorganization energy) conformations identified in **Aim 1.2**, and use a small molecule binding site identification algorithm such as LigSite [85] or SiteMap [86,87] that attempt to locate druggable pockets on individual configurations within these metastable conformations. This will allow us to assess the potential for designing new ligands that could access pockets previously unobserved by structural biology techniques by paying an energetic penalty to populate these conformations.

We will compile the set of putative cryptic binding sites for each kinase in our study, cross-correlating sites across kinases to determine if any of these pockets might be particularly suitable for exploiting *differences* either in binding site contacts or reorganization free energy in order to achieve selectivity. Structures, putative binding sites, associated reorganization energies, and corresponding annotations regarding potential exploitability to achieve selectivity will be rapidly made available online as analysis is completed to aid ongoing drug discovery efforts.

Potential Problems and Alternative Strategies: Different site-finding algorithms are likely to differ in the pockets they identify; to mitigate this, we plan to use multiple site-finding algorithms if possible. We will further mitigate the potential for missing opportunities for additional analyses by making the full conformational datasets (stripped of solvent to make them of accessible size) publicly available to other laboratories to make maximal use of this data.

Subaim 3.3 *Introduce mutations to probe reoganization energy differences and potential for achieving selectivity for oncogenic mutations*

Approach: In order to demonstrate our understanding of the role of reorganization energies in inhibitor selectivity and affinity, we will attempt to manipulate net inhibitor binding affinities in a controlled fashion by introducing mutations that perturb *only* reorganization energies and not ligand binding free energies to specific conformations. To do this, for a small number of kinases that express well in our automated bacterial expression pipeline, we will identify mutations outside the binding site that are predicted by alchemical binding free energy calculations to not perturb inhibitor binding affinities to individual conformations, but where relative protein mutation free energy calculations suggest the mutation will modulate reorganization energies among kinase conformational states. We will use our automated mutagenesis pipeline to engineer a variety of mutations matching these criteria, measuring the changes in binding affinities to our panel of kinase inhibitors for mutants with sufficiently high expression. If the overall affinity to the inhibitor is modulated in the predicted manner, this demonstrates both our ability to understand and control the reorganization energy. If there are instead significant discrepancies for some mutants, we will work with our structural biology collaborator Markus Seeliger (Stony Brook) to see if crystal structures could be obtained for these mutants (apo and inhibitor-bound) to identify whether our computational conformation atlas may have missed important conformations not yet observed by crystallography.

Mutations appearing in tumors that may exert their activating effects by modulating conformational populations—either increasing active populations or reducing inhibitor affinity—are of particular interest due to the potential role of reorganization energy in their effects and the possibility for exploiting reorganization energy differences to achieve selectivity. We will examine clinically-identified mutations from cancer mutation datasets (such as the TCGA [http://cancergenome.nih.gov/] via the cBioPortal for Cancer Genomics [http://cbioportal.org]) to identify mutations outside the ATP-binding site, engineer these mutations, and assay their impact on inhibitor affinities using the fluorescence assay of **Aim 2.2**. In parallel, we will introduce these mutations computationally to assess their impact on both conformational reorganization energies and inhibitor binding affinities using alchemical free energy calculations. Opportunities to exploit reorganization energy differences between wild-type and oncogenic mutants via binding sites exposed in specific conformations will be noted, and the potential role reorganization energy in achieving inhibitor resistance assessed.

Potential Problems and Alternative Strategies: While many of the mutants we attempt to engineer may not express, we consider this low risk due to the abundance of possible mutants that can be easily screened for study. Mutations can cause destabilization, dimerization, or other artifacts, making it vital we apply the same quality control tests described in **Aim 2.1** to ensure our mutants are stable, pure, monomeric, and unphosphorylated.

Bibliography and References Cited

[1] American Cancer Society, . Cancer Facts & Figures 2015, 2015.

[2] Cohen, P. and Alessi, D. R.: Kinase drug discovery—what's next in the field? ACS Chemical Biology. 8: 96–104, 2013.

[3] Cohen, P.: Protein kinases—the major drug targets of the twenty-first century? Nature Rev. Drug Disc. 1: 309–315, 2002.

[4] Cohen, P. and Tcherpakov, M.: Will the ubiquitin system furnish as many drug targets as protein kinases? Cell. 143: 686–693, 2010.

[5] Stegmeier, F., Warmuth, M., Sellers, W. R., and Dorsch, M.: Targeted cancer therapies in the twenty-first century: Lessons from imatinib. Clin. Pharm. & Therap. 87: 543–552, 2010.

[6] Wilson, T. R., Johnston, P. G., and Longley, D. B.: Anti-apoptotic mechanisms of drug resistance in cancer. Curr. Cancer Drug Targets. 9: 307–319, 2016.

[7] Cowan-Jacob, S. W., Fendrich, G., Floersheimer, A., Furet, P., Liebetanz, J., Rummel, G., Rheinberger, P., Centeleghe, M., Fabbro, D., and Manley, P. W.: Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukemia. Acta Cryst. D. 63: 80–93, 2007.

[8] Seeliger, M. A., Nagar, B., Frank, F., Cao, X., Henderson, M. N., and Kuriyan, J.: c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty. Structure. 15: 299–311, 2007.

[9] Levinson, N. M., Kuchment, O., Shen, K., Young, M. A., Koldobskly, M., Karplus, M., Cole, P. A., and Kuriyan, J.: A Src-like inactive conformation in the Abl tyrosine kinase domain. Plos Biology. 4: e144, 2006.

[10] Seeliger, M. A., Ranjitkar, P., Kasap, C., Shan, Y., Shaw, D. E., Shah, N. P., Kuriyan, J., and Maly, D. J.: Equally potent inhibition of c-Src and Abl by compounds that recognize inactive kinase conformations. Cancer Res. 69: 2384–2392, 2006.

[11] Aleksandrov, A. and Simonson, T.: Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases. J. Biol. Chem. 285: 13807–13815, 2010.

[12] Lin, Y.-L., Meng, Y., Jiang, W., and Roux, B.: Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. Proc. Natl. Acad. Sci. USA. 110: 1664–1669, 2013.

[13] Agafonof, R. V., Wilson, C., Otten, R., Buosi, V., and Kern, D.: Energetic dissection of Gleevec's selectivity toward human tryosine kinases. Nature Struct. Mol. Biol. 21: 848–852, 2014.

[14] Lee, G. M. and Craik, C. S.: Trapping moving targets with small molecules. Science. 324: 213, 2009.

[15] Parton, D. L., Grinaway, P. B., Hanson, S. M., Beauchamp, K. A., and Chodera, J. D.: Ensembler: Enabling high-throughput molecular simulations at the superfamily scale. PLoS Computational Biology. in press.

[16] Korzhnev, D. M. and Kay, L. E.: Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. Acc. Chem. Res. 41: 442–451, 2008.

[17] http://www.sgc.ox.ac.uk/research/kinases/.

[18] Fabian, M. A., Biggs, W. H., Treiber, D. K., Atteridge, C. E., Azimioara, M. D., Benedetti, M. G., Carter, T. A., Ciceri, P., Edeen, P. T., Floyd, M., Ford, J. M., Galvin, M., Gerlach, J. L., Grotzfeld, R. M., Herrgard, S., Insko, D. E., Insko, M. A., Lai, A. G., Lélias, J.-M., Mehta, S. A., Milanov, Z. V., Velasco, A. M., Wodicka, L. M., Patel, H. K., Zarrinkar, P. P., and Lockhart, D. J.: A small molecule–kinase interaction map for clinical kinase inhibitors. Nature Biotech. 23: 329–336, 2005.

[19] Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., Chan, K. W., Ciceri, P., Davis, M. I., Edeen, P. T., Faraoni, R., Floyd, M., Hunt, J. P., Lockhart, D. J., Milanov, Z. V., Morrison, M. J., Pallares, G., Patel, H. K., Pritchard, S., Wodicka, L. M., and Zarrinkar, P. P.: A quantitative analysis of kinase inhibitor selectivity. Nature Biotech. 26: 127–132, 2008.

[20] Shirts, M. and Pande, V. S.: Screen savers of the world unite! Science. 290: 1903–1904, 2000.

[21] http://fah-web.stanford.edu/cgi-bin/main.py?qtype=osstats.

[22] Chodera, J. D., Swope, W. C., Pitera, J. W., and Dill, K. A.: Long-time protein folding dynamics from short-time molecular dynamics simulations. Multiscale Model. Simul. 5: 240–251, 2006.

[23] Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A., and Swope, W. C.: Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. J. Chem. Phys. 126: 155101, 2007.

[24] Prinz, J.-H., Wu, H., Sarich, M., Keller, B., Fischbach, M., Held, M., Chodera, J. D., Schütte, C., and Noé, F.: Markov models of molecular kinetics: Generation and validation. J. Chem. Phys. 134: 174105, 2011.

[25] Chodera, J. D. and Noé, F.: Probability distributions of molecular observables computed from Markov models: II. uncertainties in observables and their time-evolution. J. Chem. Phys. 133: 105102, 2011.

[26] Bacallado, S., Chodera, J. D., and Pande, V. S.: Bayesian comparison of Markov models of molecular dynamics with detailed balance constraint. J. Chem. Phys. 131: 045106, 2009.
[27] Noé, F., Doose, S., Daidone, I., Löllmann, M., Sauer, M., Chodera, J. D., and Smith, J. C.: Dynamical fingerprints: A theoretical framework for understanding biomolecular processes by combination of simulation and kinetic experiments. Proc. Natl. Acad. Sci. USA. 108: 4822, 2011.
[28] Chodera, J. D. and Noé, F.: Markov state models of biomolecular conformational dynamics. Curr. Opin. Struct. Biol. 25: 135–144, 2014.
[29] Shukla, D., Meng, Y., Roux, B., and Pande, V. S.: Activation pathway of src kinase reveals intermediate states as targets for drug design. Nature Comm. 5: 3397, 2014.
[30] Parton, D. L., Hanson, S. M., Rodríguez-Laureano, L., Albanese, S. K., Jeans, C., Gradia, S., Seeliger, M. A., and Chodera, J. D.: A panel of recombinant human kinase domain constructs for automated bacterial expression. bioRxiv. http://dx.doi.org/10.1101/038711 pp, 2016.
[31] Seeliger, M. A., Young, M., Henderson, M. N., Pellicena, P., King, D. S., Falick, A. M., and Kuriyan, J.: High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. Protein Sci. 14: 3135–3139, 2005.
[32] Levinson, N. M. and Boxer, S. G.: Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain. PLoS One. 7: e29828, 2012.
[33] 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. Curr. Opin. Struct. Biol. 21: 150, 2011.
[34] Wang, K. K., Chodera, J. D., Yang, Y., and Shirts, M. R.: Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. J. Comput. Aid. Mol. Des. 27: 989, 2013.
[35] Mobley, D. L., Chodera, J. D., and Dill, K. A.: On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. J. Chem. Phys. 125: 084902, 2006.
[36] Mobley, D. L., Chodera, J. D., and Dill, K. A.: Confine and release: Obtaining correct free energies in the presence of protein conformational change. J. Chem. Theor. Comput. 3: 1231–1235, 2007.
[37] Shirts, M. R., Mobley, D. L., Chodera, J. D., and Pande, V. S.: Accurate and efficient corrections for missing dispersion interactions in molecular simulations. J. Phys. Chem. B. 111: 13052–13063, 2007.
[38] Shirts, M. R. and Chodera, J. D.: Statistically optimal analysis of samples from multiple equilibrium states. J. Chem. Phys. 129: 124105, 2008.
[39] Nilmeier, J. P., Crooks, G. E., Minh, D. D. L., and Chodera, J. D.: Nonequilibrium candidate Monte Carlo: A new tool for equilibrium simulation. Proc. Natl. Acad. Sci. USA. 108: E1009, 2011.
[40] Chodera, J. D. and Shirts, M. R.: Replica exchange and expanded ensemble simulations as gibbs sampling: Simple improvements for enhanced mixing. J. Chem. Phys. 135: 194110, 2011.
[41] Deininger, M., Buchdunger, E., and Druker, B. J.: The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood. 105: 2640–2653, 2005.
[42] 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., and Klein, C.: OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. J. Chem. Theor. Comput. 9: 461, 2012.
[43] Scalco, R. and Caflisch, A.: Equilibrium distribution from distributed computing (simulations of protein folding). J. Phys. Chem. B. 115: 6358, 2011.
[44] Bowman, G. R., Huang, X., and Pande, V. S.: Using generalized ensemble simulations and markov state models to identify conformational states. Methods. 49: 197–201, 2009.
[45] Beauchamp, K. A., Bowman, G. R., Lane, T. J., Maibaum, L., Haque, I. S., and Pande, V. S.: MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond timescale. J. Chem. Theor. Comput. 7: 3412—3419, 2011.
[46] Scherer, M. K., Trendelkamp-Schroer, B., Paul, F., Prez-Hernndez, G., Hoffmann, M., Plattner, N., Wehmeyer, C., Prinz, J.-H., and No, F.: PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. Journal of Chemical Theory and Computation. October 2015.
[47] Mandell, D. J., Coutsias, E. A., and Kortemme, T.: Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nature Methods. 6: 551, 2009.
[48] Song, Y., DiMaio, F., Wang, R. Y.-R., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D.: High-resolution comparative modeling with RosettaCM. Structure. 21: 1735–1742, 2013.
[49] Kube, S. and Weber, M.: A coarse graining method for the identification of transition rates between molecular conformations. J. Chem. Phys. 126: 024103, 2007.
[50] Chodera, J. D., Elms, P., Noé, F., Keller, B., Kaiser, C. M., Ewall-Wice, A., Marqusee, S., Bustamante, C. J., and Hinrichs, N. S.: Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty. arXiv:1108.1430. 2011.

[51] Bowman, G. R.: Improved coarse-graining of markov state models via explicit consideration of statistical uncertainty. J. Chem. Phys. 137: 134111, 2012.
[52] Noé, F.: Probability distributions of molecular observables computed from Markov models. J. Chem. Phys. 128: 244103, 2010.
[53] Chodera, J. D., Swope, W. C., Noé, F., Prinz, J.-H., Shirts, M. R., and Pande, V. S.: Dynamical reweighting: Improved estimates for dynamical properties from simulations at multiple temperatures. J. Chem. Phys. in press pp, 2011.
[54] Prinz, J.-H., Chodera, J. D., Pande, V. S., Smith, J. C., and Noé, F.: Optimal use of data in parallel tempering simulations for the construction of kinetic models of peptide dynamics. J. Chem. Phys. in press pp, 2011.
[55] Morcos, F., Chatterjee, S., McClendon, C. L., Brenner, P. R., López-Rendón, R., Zintsmaster, J., Ercsey-Ravasz, M., Sweet, C. R., Jacobson, M. P., Peng, J. W., and Izaguirre, J. A.: Modeling conformational ensembles of slow functional motions in Pin1-WW. PLoS Comput. Biol. 6: e1001015, 2010.
[56] Beauchamp, K. A., McGibbon, R., Lin, Y.-S., and Pande, V. S.: Simple few-state models reveal hidden complexity in protein folding. Proc. Natl. Acad. Sci. USA. 109: 17807–17813, 2012.
[57] Xia, J., jie Deng, N., and Levy, R. M.: NMR relaxation in proteins with fast internal motions and slow conformational exchange: Model-free framework and markov state simulations. J. Phys. Chem. B. 117: 6625–6634, 2013.
[58] Beauchamp, K. A., Lin, Y.-S., Das, R., and Pande, V. S.: Are protein force fields getting better? a systematic benchmark on 524 diverse NMR measurements. J. Chem. Theory Comput. 8: 1409–1414, 2012.
[59] Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Wenger, R. K., Yao, H., and Markley, J. L.: BioMagResBank. Nucl. Acids Res. 36: D402–D408, 2008.
[60] Vajpai, N., Strauss, A., Fendrich, G., Cowan-Jacob, S. W., Manley, P. W., Grzesiek, S., and Jahnke, W.: Solution conformations and dynamics of abl kinase-inhibitor complexes determined by nmr substantiate the different binding modes of imatinib/nilotinib and dasatinib. J. Biol. Chem. 283: 18292–18302, 2008.
[61] Chambers, S. P., Austen, D. A., Fulghum, J. R., and Kim, W. M.: High-throughput screening for soluble recombinant expressed kinases in *Escherichia coli* and insect cells. Protein Expression and Purification. 36: 40–47, 2004.
[62] Wang, L., Foster, M., Zhang, Y., Tschantz, W. R., Yang, L., Worrall, J., Loh, C., and Xu, X.: High yield expression of non-phosphorylated protein tyrosine kinases in insect cells. Protein Express. Pur. 61: 204–211, 2008.
[63] Lavinder, J. L., Hari, S. B., Sullivan, B. J., and Magliery, T. J.: High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. J. Am. Chem. Soc. 131: 3794–3795, 2009.
[64] Kramer, C., Kalliokoski, T., Gedeck, P., and Vulpetti, A.: The experimental uncertainty of heterogeneous public k_i data. J. Med. Chem. 55: 5165–5173, 2012.
[65] Huan, X. and Marzouk, Y. M.: Simulation-based optimal Bayesian experimental design for nonlinear systems. J. Comput. Phys. 232: 288–317, 2013.
[66] Tembe, B. L. and McCammon, J. A.: Ligand-receptor interactions. Computers & Chemistry. 8: 281–283, 1984.
[67] Michael R. Shirts, D. L. M. and Chodera, J. D.: Alchemical free energy calculations: Ready for prime time? Annu. Rep. Comput. Chem. 3: 41–59, 2007.
[68] Michel, J. and Essex, J.: Hit identification and binding mode predictions by rigorous free energy simulations. J. Med. Chem. 51: 6654–6664, Oct 2008.
[69] Michel, J. and Essex, J. W.: Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. J. Comput. Aided Mol. Des. 24: 639–658, 2010.
[70] Gallicchio, E. and Levy, R. M.: Advances in all atom sampling methods for modeling protein–ligand binding affinities. Curr. Opin. Struct. Biol. 21: 161–166, 2011.
[71] Shan, Y., Seeliger, M. A., Eastwood, M. P., Frank, F., Xu, H., Jensen, M. O., Dror, R. O., Kuriyan, J., and Shaw, D. E.: A conserved protonation-dependent switch controls drug binding in Abl kinase. Proc. Natl. Acad. Sci. USA. 106: 139–144, 2009.
[72] Czodrowski, P., Sotriffer, C. A., and Klebe, G.: Protonation changes upon ligand binding to trypsin and thrombin: Structural interpretation based on pKa calculations and ITC experiments. J. Mol. Biol. 367: 1347–1356, 2007.
[73] Martin, Y. C.: Let's not forget tautomers. J. Comput. Aid. Mol. Des. 23: 693–704, 2009.
[74] Prendergast, F. G., Meyer, M., Carlson, G. L., Ilda, S., and Potter, J. D.: Synthesis, spectral properties, and use of 6-acryloyl-2-dimethylaminonaphthalene (acrylodan). a thiol-selective, polarity-sensitive fluorescent probe. J. Biol. Chem. 258: 7541, 1983.

[75] Simard, J. R., Getlik, M., Grütter, C., Pawar, V., Wulfert, S., Rabiller, M., and Rauh, D.: Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. J. Am. Chem. Soc. 131: 13286–13296, 2009.
[76] Simard, J. R., Getlik, M., Grütter, C., Schneider, R., Wulfert, S., and Rauh, D.: Fluorophore labeling of the glycine-rich loop as a method of identifying inhibitors that bind to active and inactive kinase conformations. J. Am. Chem. Soc. 132: 4152–4160, 2010.
[77] Schneider, R., Becker, C., Simard, J. R., Getlik, M., Bohlke, N., Janning, P., and Rauh, D.: Direct binding assay for the detection of Type IV allosteric inhibitors of Abl. J. Am. Chem. Soc. 134: 9138–9141, 2012.
[78] Lane, T. J., Bowman, G. R., Beauchamp, K., Voelz, V. A., and Pande, V. S.: Markov state model reveals folding and functional dynamics in ultra-long MD trajectories. J. Am. Chem. Soc. 133: 18413–18419, 2011.
[79] Hoffman, B., Zaslavskiy, M., Vert, J.-P., and Stoven, V.: A new protein binding pocket similarity measure based on comparison of clouds of atoms in 3D: application to ligand prediction. BMC Bioinformatics. 11: 99, 2010.
[80] Spitzer, R., Cleves, A. E., and Jain, A. N.: Surface-based protein binding pocket similarity. Proteins. 79: 2746–2763, 2011.
[81] Desaphy, J., Azdimousa, K., Kellenberger, E., and Rognan, D.: Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. J. Chem. Inf. Model. 52: 2287–2299, 2012.
[82] Adrián, F. J., Ding, Q., Sim, T., Velentza, A., Sloan, C., Liu, Y., Zhang, G., Hur, W., Ding, S., Manley, P., Mestan, J., Fabbro, D., and Gray, N. S.: Allosteric inhibitors of Bcl-abl-dependent cell proliferation. Nature Chem. Biol. 2: 95–102, 2006.
[83] Zhang, J. and et al., : Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. Nature. 463: 501–507, 2010.
[84] Bowman, G. R. and Geissler, P. L.: Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. Proc. Nat. Acad. Sci. 109: 11681–11686, 2012.
[85] Hendlich, M., Rippmann, F., and Barnickel, G.: LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins. J. Mol. Graph. Model. 15: 359–363, 1997.
[86] Halgren, T.: New method for fast and accurate binding-site identification and analysis. Chem. Biol. Drug Des. 69: 146–148, 2007.
[87] Halgren, T.: Identifying and characterizing binding sites and assessing druggability. J. Chem. Inf. Model. 49: 377–389, 2009.

STANFORD
UNIVERSITY

DEPARTMENT OF CHEMISTRY

Vijay S. Pande, PhD
Professor, Department of Chemistry
Stanford University
James H. Clark Center
318 Campus Drive, S295
Stanford, CA 94305-5444
Tel: 650-723-3660
pande@stanford.edu

January 29, 2016

John D. Chodera
Assistant Member
Computational Biology Program
Memorial Sloan Kettering Cancer Center
email: choderaj@mskcc.org

Dear John,

I am writing to express my enthusiasm and support for your proposal to address the role of conformational reorganization energy in kinase inhibitor recognition using Folding@home. As we have been long-time collaborators, you know that the worldwide Folding@home distributed computing network that has grown out of my laboratory is the *largest* computational resource for biology in the world, with hundreds of thousands of active computing cores providing over 19 PFOP/s of sustained computational power, much of this in the form of graphics processing units (GPUs). We look forward to the continued use of this unique resource by your laboratory, which has also been involved in maintenance of the Folding@home infrastructure and GPU-accelerated OpenMM simulation cores.

We are also enthusiastic that you are continuing to make use of the GPU-accelerated OpenMM simulation library in your open-source code YANK for alchemical free energy calculations in this project. Your laboratory has been involved in OpenMM development for a number of years, and has made contributions to aid the ability of OpenMM to perform these alchemical binding free energy calculations to compute small molecule binding affinities.

I look forward to our continued productive collaboration on Folding@home and OpenMM, and wish you success on your project.

Sincerely,

Vijay Pande
Henry Dreyfus Professor of Chemistry and (by courtesy) of Structural Biology and of Computer Science
Director, Folding@home Distributed Computing Project
General Partner, Andreessen Horowitz

Department of Pharmacological Sciences
School of Medicine

Stony Brook Medicine

January 29, 2016

To:

John Chodera
Assistant Member
Computational Biology Program
Sloan Kettering Institute

Dear John,

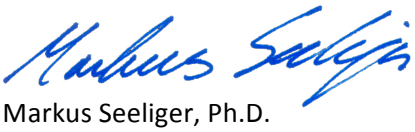
I am delighted to write in support of your proposal "The role of reorganization energy in achieving selective kinase inhibition". My own work has made seminal contributions to our understanding of the role of reorganization energy in the recognition of Abl by imatinib; this project blends computation and experiment to address how widespread this phenomenon might be for selective kinase inhibitors of clinical interest, and whether it might allow differences in reorganization energy to be exploited.

We have been collaborating for some time as you have been developing your initial panel of recombinantly expressed kinase domains using the phosphatase coexpression technique I developed while in the Kuriyan lab at Berkeley. The initial results of this expression experiment---in which 51/96 kinase domains showed >2 ug/mL expression---has generated a number of reagents we hope will be broadly useful to our laboratories and many other laboratories interested in working with these kinases. We look forward to continuing to help advise your laboratory as you engineer more constructs to expand the number of human kinase domains that can be bacterially expressed for subsequent biophysical, biochemical, and structural biology work.

We are also excited to see that you plan to engineer point mutations to directly manipulate kinase energetic landscapes to test your models of the role of reorganization energy in inhibitor selectivity and affinity. While your laboratory will focus on fluorescence measurements for these constructs, we would be delighted to assist with our deep expertise in X-ray crystallography of kinases and their complex should you discover particularly interesting mutations in kinases we are able to crystallize.

Best of luck with your grant application, and we look forward to continuing to work together.

Warm regards,



Markus Seeliger, Ph.D.
Assistant Professor of Pharmacological Sciences
Stony Brook University Medical School
++1 631 444 3558
markus.seeliger@stonybrook.edu

Resource Sharing Plan

Numerous useful and shareable resources will be generated during the course of these project, all of which will be made freely available to the research community at the earliest opportunity.

In addition to the release of information through timely open-access publications (with emphasis on complete inclusion of all primary data), we will make every attempt to also release materials and data as they are generated.

This includes, but is not limited to:

Kinase constructs and mutants. Plasmids containing kinase constructs and mutants with verified expression will be made available through nonprofit plasmid banks such as AddGene [http://www.addgene.org] and DNASU [http://dnasu.org].

Kinase inhibitor affinity and selectivity measurements. Kinase inhibitor binding affinity measurements to wild-type and mutant kinase catalytic domains will be submitted (along with primary data) to a public database such as BindingDB [http://www.bindingdb.org]. Primary data will be made available through our group website [http://choderalab.org].

Software. All computer software developed for this project will be made freely available through free (libre) open source software licenses (such as LGPL) on online collaborative public code repositories such as GitHub [http://github.com], where codes produced by our laboratory are currently hosted [http://github.com/choderalab/].

Simulation datasets. All simulation datasets will be shared, when practical, through the online repositories such as GitHub [http://github.com], Dryad [http://datadryad.org/], FigShare [http://figshare.com], and our group website [http://www.choderalab.org/data/].

Experimental protocols. In addition to publishing detailed accounts in papers, all experimental protocols will be made available online on our group website [http://choderalab.org].

Simulation protocols and best practices. We will continue to actively support and help maintain the online repository of simulation protocols, best practices, and references for alchemical free energy calculations at the community site alchemistry.org

3D printable laboratory parts. Numerous useful 3D printed parts are fabricated in our laboratory to aid in our research projects. Electronic printable versions of these parts are made available on both our group website [http://www.choderalab.org/3dparts/] and the NIH 3D Print Exchange [http://3dprint.nih.gov].