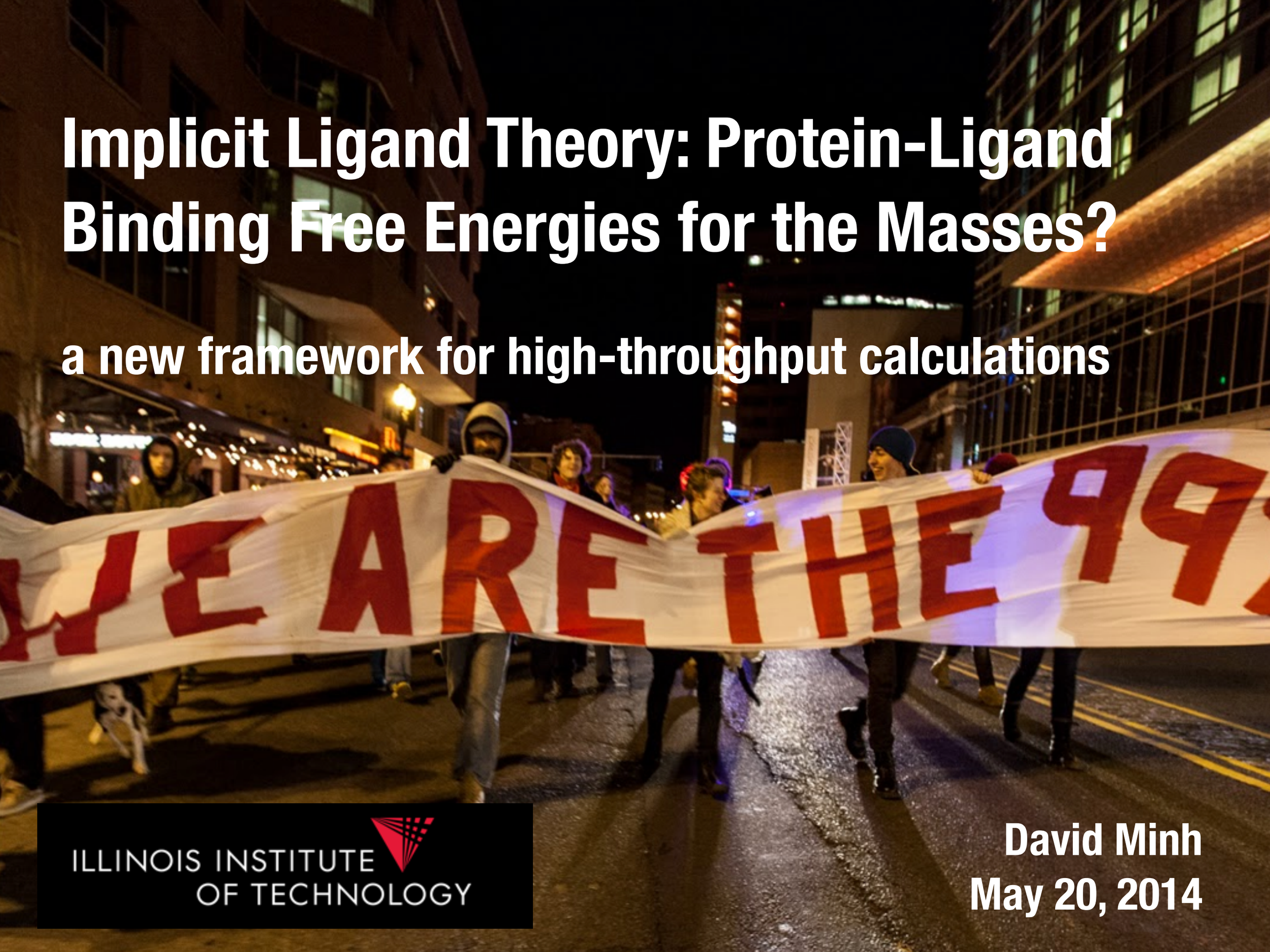
Implicit Ligand Theory: Protein-Ligand Binding Free Energies for the Masses?

a new framework for high-throughput calculations



ILLINOIS INSTITUTE OF TECHNOLOGY

David Minh

May 20, 2014

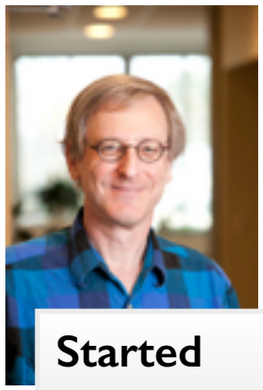
“Gunsaulus said that with a million dollars he could build a school where *students of all backgrounds* could prepare for meaningful roles in a changing industrial society”

Philip Danforth Armour, Sr.

Frank Wakely Gunsaulus

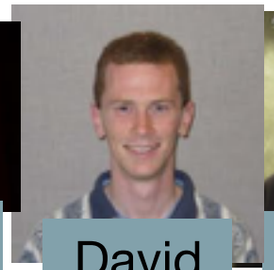
ILLINOIS INSTITUTE OF TECHNOLOGY


ILLINOIS INSTITUTE OF TECHNOLOGY

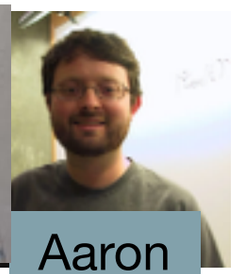


Started postdoc with David Beratan

John Chodera



David Mobley



Aaron Virshup

Vijay Pande

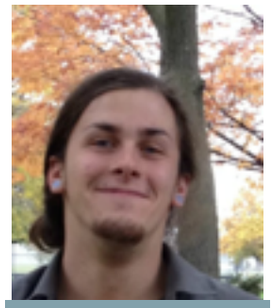
Peter Eastman

Submitted theory paper to JCP

OpenMM visiting scholar at Stanford

Michael Sherman

Laurentiu Spiridon starts postdoc



John Clark

2012 Jun 2013 Jun 2014 Jun

Chen Li

Yu-Ting Yu

Derived theory

Mar Sep Mar Sep Mar

Submitted revision with host-guest model system

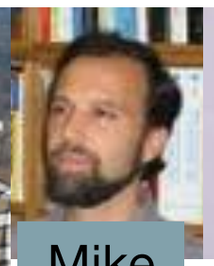
ACS Thomas Kuhn paradigm shift award symposium

Started assistant professor job at IIT

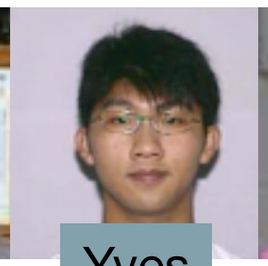
DHFR binding free energies

Bing Xie

Emilio Gallichio

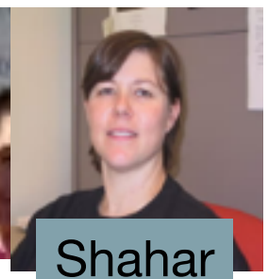


Mike Gilson



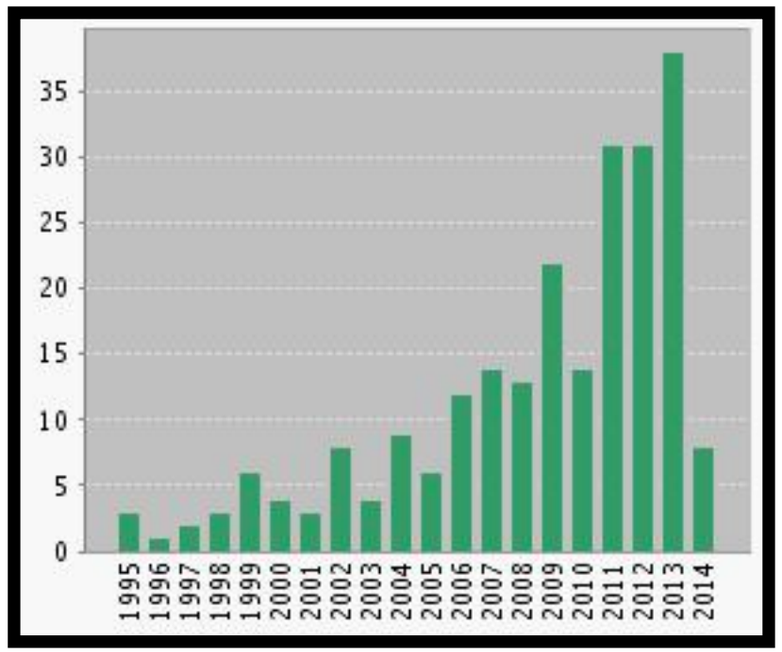
Yves Wang

Clayton Jarrat

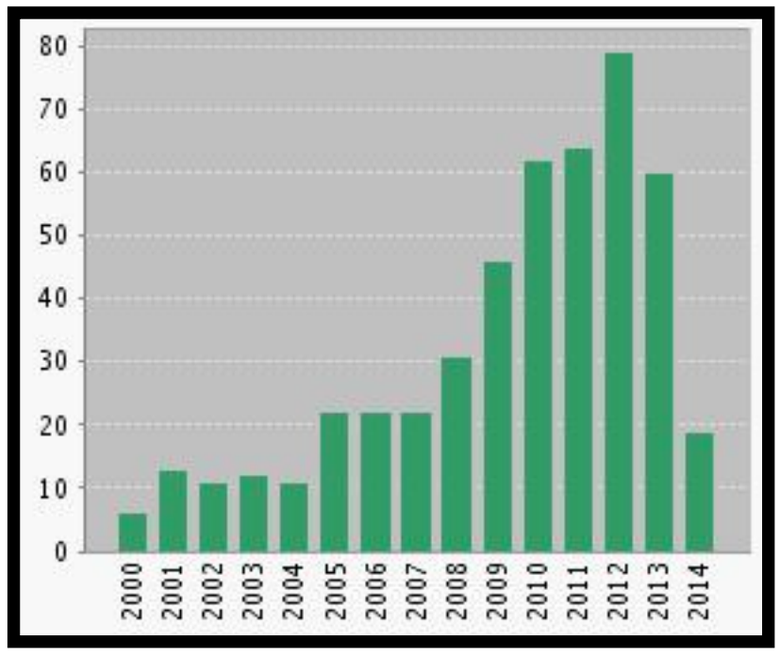


Shahar Keinan

Web of Science, May 2014

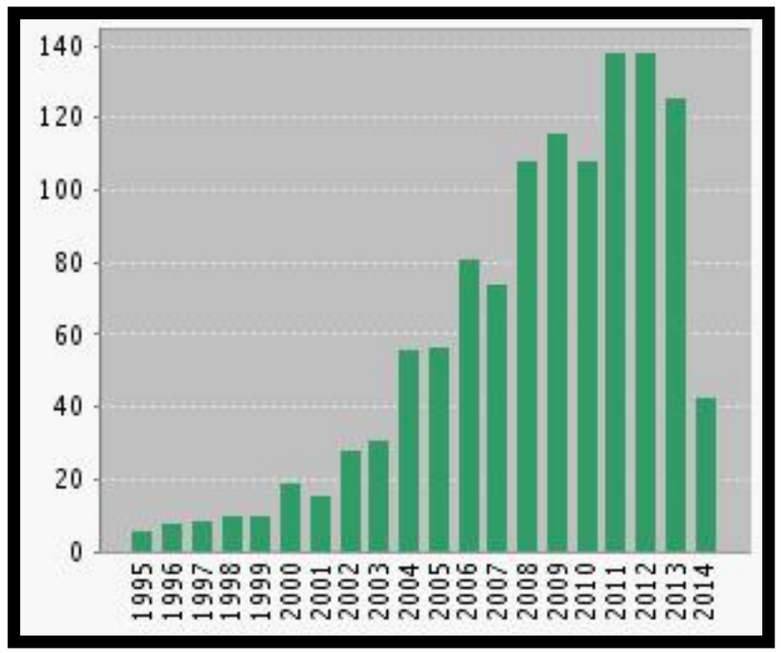


alchemical, 246

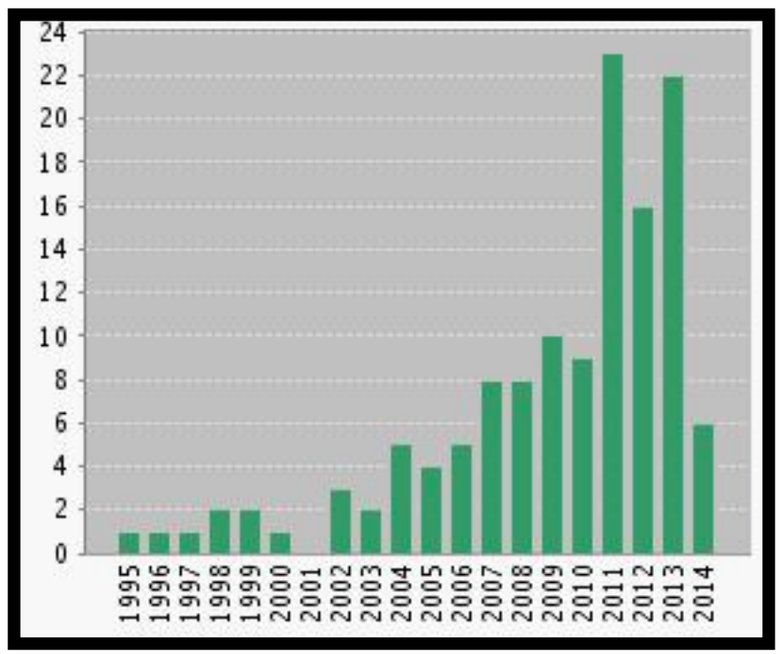


MM/PBSA, 480

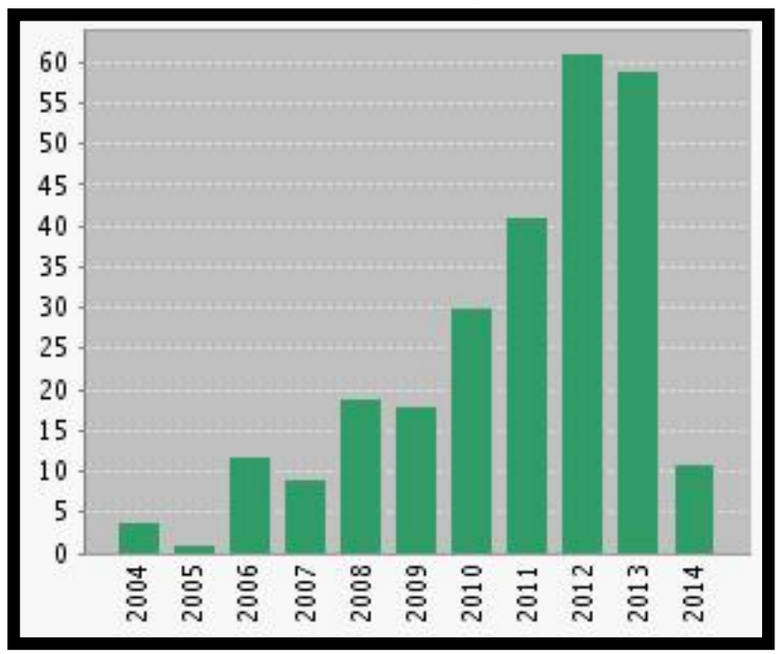
molecular docking,
15,121



protein-ligand
molecular docking,
1186

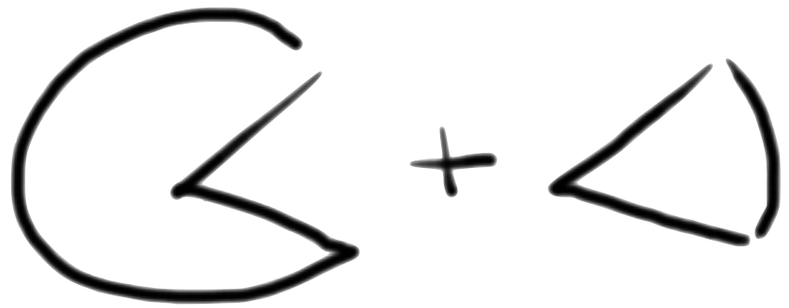


alchemical free energy, 131



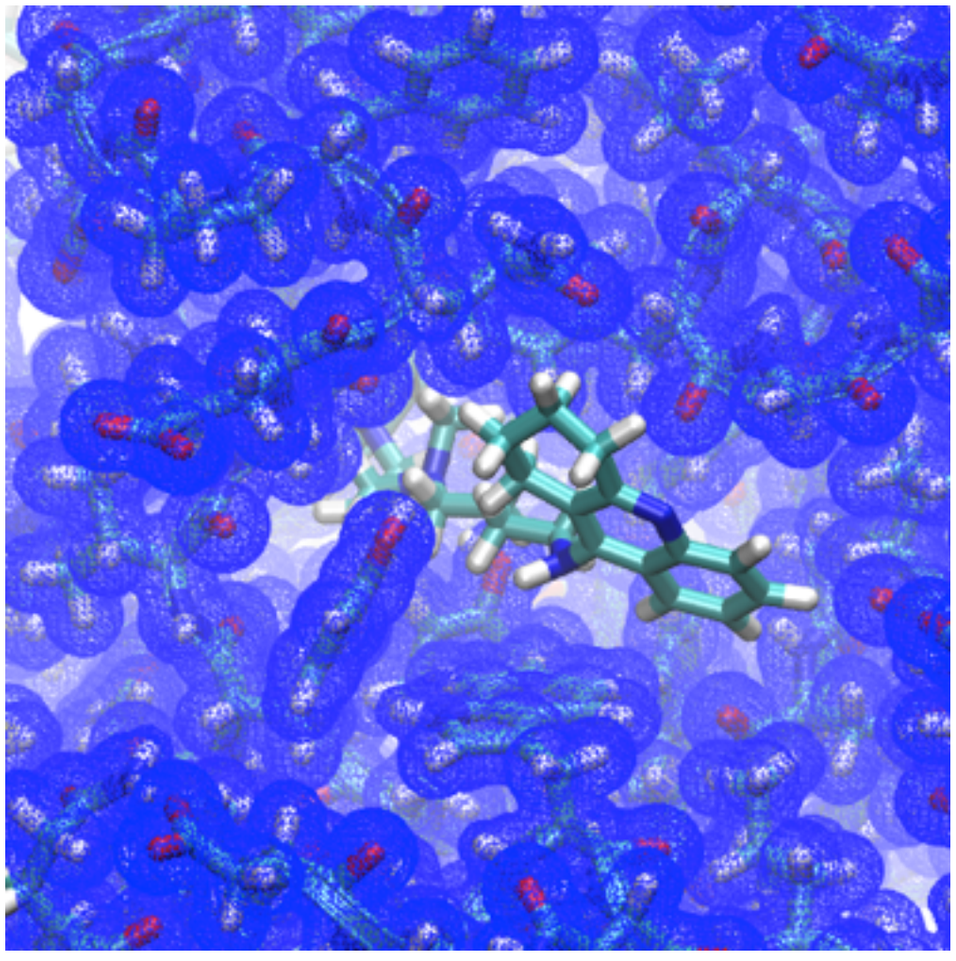
MM/GBSA, 265

Molecular Docking



- Why is docking popular?
  - It is easy to use
  - It is ***fast***
- Why is docking fast?
  - Rigid receptors
    - no internal degrees of freedom
    - pre-calculated interaction grids
  - It is focused on minimization, not statistical sampling
- Can free energy calculations apply some ideas from docking?
  - Yes
  - With implicit ligand theory, free energy calculations can use rigid receptors.

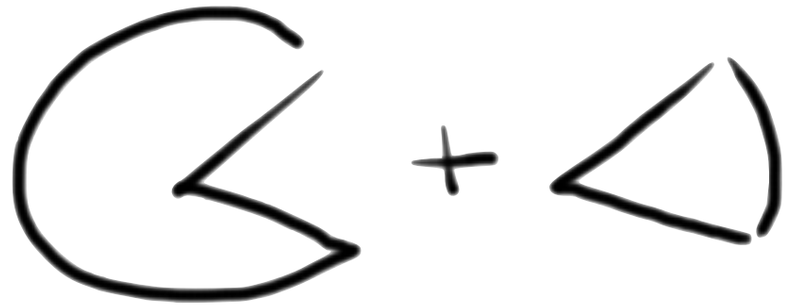
with 500 orientations,
1 ligand every 5 seconds
Coleman et al., PloS One 2013



Statistical Mechanics of Noncovalent Association

$$\Delta G^\circ = -\beta^{-1} \ln \left(\frac{C^\circ C_{RL}}{C_R C_L} \right)$$

$$\Delta G^\circ = -\beta^{-1} \ln \left(\frac{Z_{RL,N} Z_N}{Z_{R,N} Z_{L,N}} \frac{C^\circ}{8\pi^2} \right)$$



R L

$$Z_{RL,N} = \int I_\xi e^{-\beta U(r_{RL}, r_S)} dr_{RL} dr_S$$

$$Z_{Y,N} = \int e^{-\beta U(r_Y, r_S)} dr_Y dr_S$$

$$Z_N = \int e^{-\beta U(r_S)} dr_S$$

$$R + L \rightleftarrows RL$$

C_R	free receptor concentration
C_L	free ligand concentration
C_{RL}	complex concentration
C°	standard state concentration (1 M)

Gilson et al, Biophys J 1997

Implicit Solvent Theory

$$Z_X \equiv \frac{Z_{X,N}}{Z_N} = \frac{\int e^{-\beta U(r_X, r_S)} dr_X dr_S}{\int e^{-\beta U(r_S)} dr_S}$$

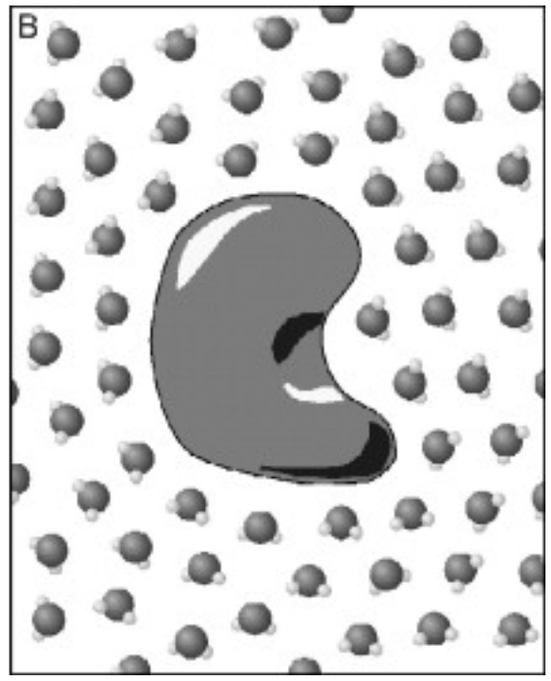
$$= \frac{\int e^{-\beta [\psi(r_X, r_S) + U(r_X) + U(r_S)]} dr_X dr_S}{\int e^{-\beta U(r_S)} dr_S}$$

$$= \int e^{-\beta [U(r_X) + W(r_X)]} dr_X$$

$$W(r_X) = -\beta^{-1} \ln \left(\frac{\int e^{-\beta \psi(r_X, r_S)} e^{-\beta U(r_S)} \, dr_S}{\int e^{-\beta U(r_S)} \, dr_S} \right)$$

Interaction Energy

$$\psi(r_X, r_S) = U(r_X, r_S) - U(r_X) - U(r_S)$$



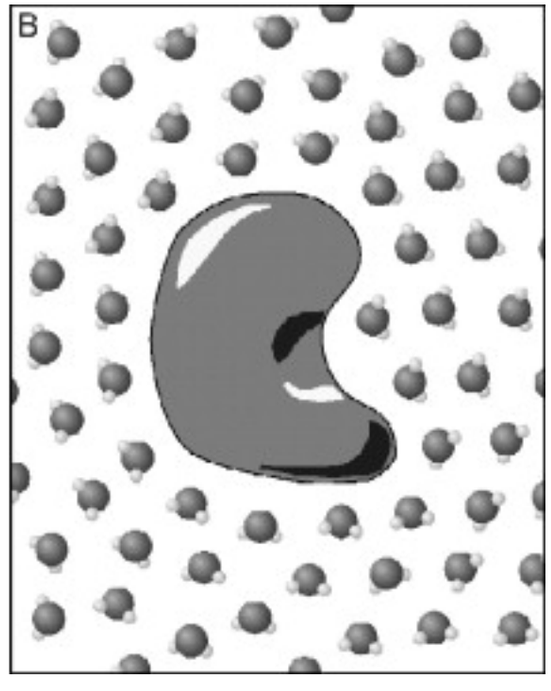
Dong et. al.,
Methods in Cell Biology 2008

Implicit Solvent Theory

$$\Delta G^\circ = -\beta^{-1} \ln \left(\frac{Z_{RL,N} Z_N}{Z_{R,N} Z_{L,N}} \frac{C^\circ}{8\pi^2} \right)$$

$$= -\beta^{-1} \ln \left(\frac{Z_{RL}}{Z_R Z_L} \frac{C^\circ}{8\pi^2} \right)$$

$$Z_X = \int e^{-\beta[U(r_X)+W(r_X)]} dr_X$$



Dong et. al.,
Methods in Cell Biology 2008

Implicit Ligand Theory

$$Z_X = \int e^{-\beta[U(r_X)+W(r_X)]} dr_X$$

Effective Potential Energy

$$\mathcal{U}(r_X) = U(r_X) + W(r_X)$$

$$\Delta G^\circ = -\beta^{-1} \ln \left(\frac{\int I_\xi e^{-\beta \mathcal{U}(r_{RL})} dr_{RL}}{\int e^{-\beta \mathcal{U}(r_R)} dr_R \int e^{-\beta \mathcal{U}(r_L)} dr_L} \frac{C^\circ}{8\pi^2} \right)$$

$$= -\beta^{-1} \ln \left(\frac{\int I_\xi e^{-\beta[\mathcal{U}(r_R)+\Psi(r_{RL})+\mathcal{U}(r_L)]} dr_{RL}}{\int e^{-\beta \mathcal{U}(r_R)} dr_R \int e^{-\beta \mathcal{U}(r_L)} dr_L} \frac{C^\circ}{8\pi^2} \right)$$

$$= -\beta^{-1} \ln \left(\frac{\int e^{-\beta[B(r_R)+\mathcal{U}(r_R)]} dr_R}{\int e^{-\beta \mathcal{U}(r_R)} dr_R} \frac{\Omega C^\circ}{8\pi^2} \right) \quad \Omega = \int I_\xi d\xi_L$$

Effective Interaction Energy

$$\Psi(r_{RL}) = \mathcal{U}(r_{RL}) - \mathcal{U}(r_R) - \mathcal{U}(r_L)$$

$$B(r_R) = -\beta^{-1} \ln \left(\frac{\int I_\xi e^{-\beta \Psi(r_{RL})} e^{-\beta \mathcal{U}(r_L)} \, dr_L d\xi_L}{\int I_\xi e^{-\beta \mathcal{U}(r_L)} \, dr_L d\xi_L} \right)$$

Implicit Ligand Theory

- Rigorous binding free energies
- Rigid receptor

Binding Free Energy

$$\Delta G^\circ = \beta^{-1} \ln \left\langle e^{-\beta B} \right\rangle_R^{r_R} + \Delta G_\epsilon$$

Binding PMF

$$B(r_R) = \beta^{-1} \ln \left\langle e^{-\beta \Psi} \right\rangle_{L,I}^{r_L, \epsilon_L}$$

Effective Interaction Energy

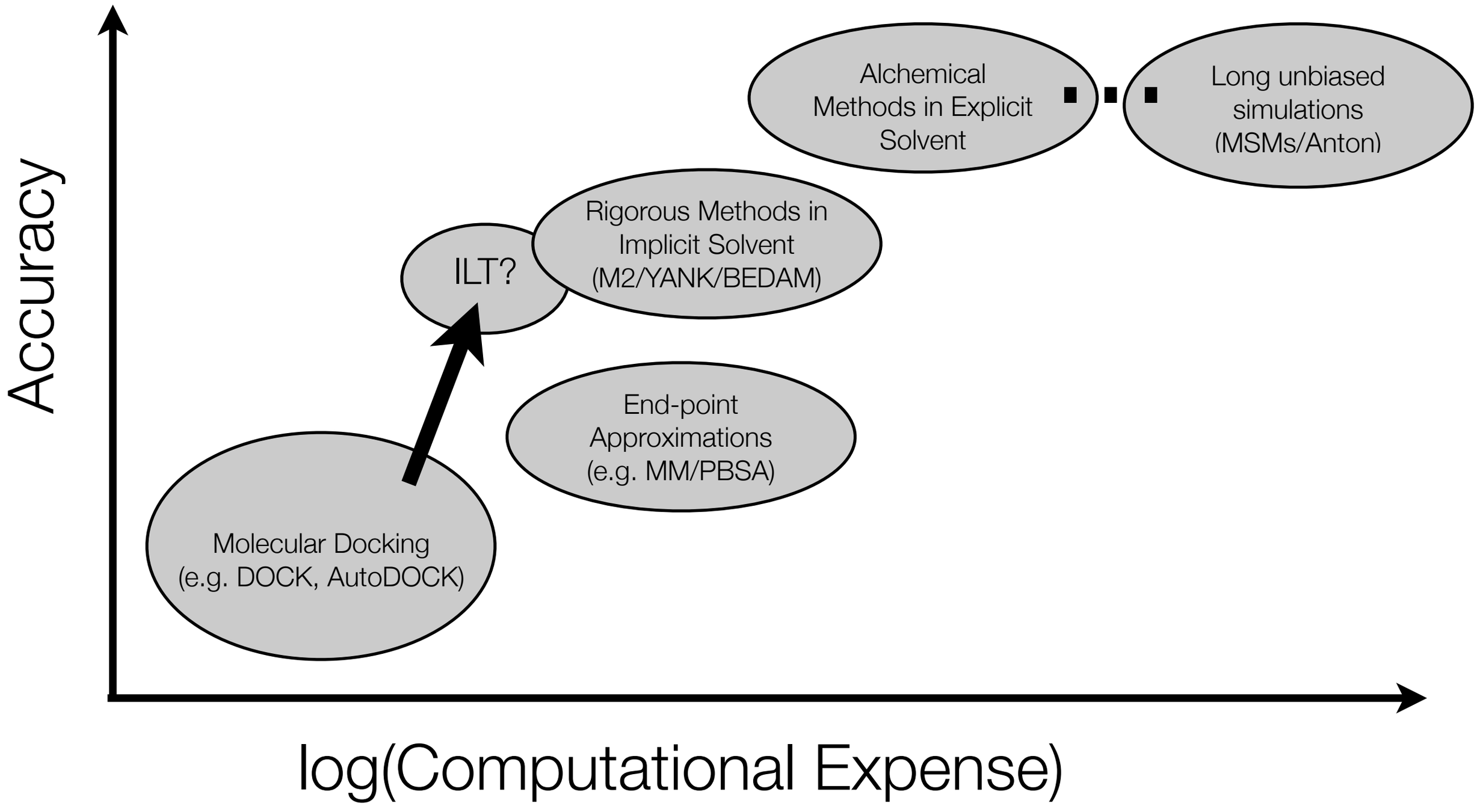
$$\Psi(r_{RL}) = \mathcal{U}(r_{RL}) - \mathcal{U}(r_R) - \mathcal{U}(r_L)$$

Effective Potential Energy

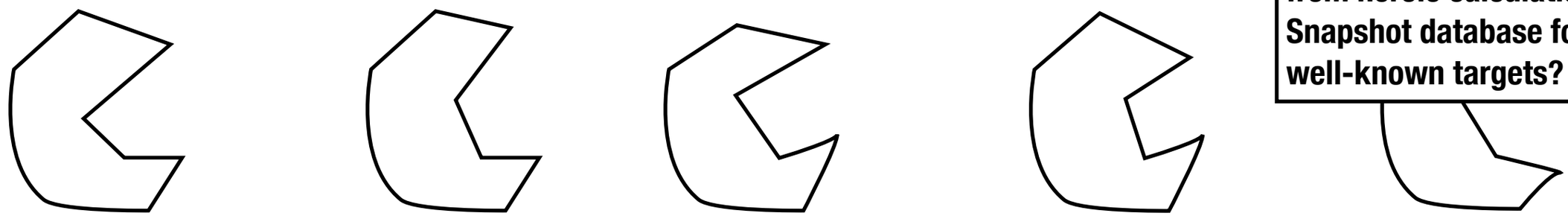
$$\mathcal{U}(r_X) = U(r_X) + W(r_X)$$

Minh, Journal of Chemical Physics 137:104106, 2012

Structure-Based Free Energy Methods



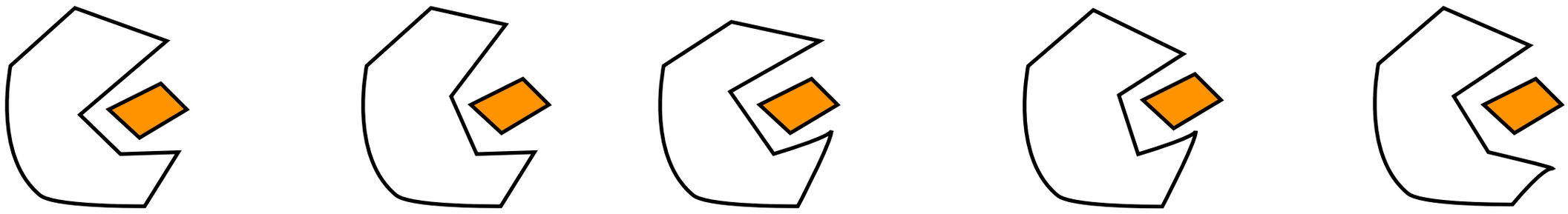
I. Sample configurations of the receptor



Only needs to be done once!
Unbiased MD simulation/
Umbrella Sampling/
Markov State Model
from heroic calculations.
Snapshot database for
well-known targets?

New type of free energy
Lessons more broadly
applicable?

II. Estimate the binding PMF for each ligand

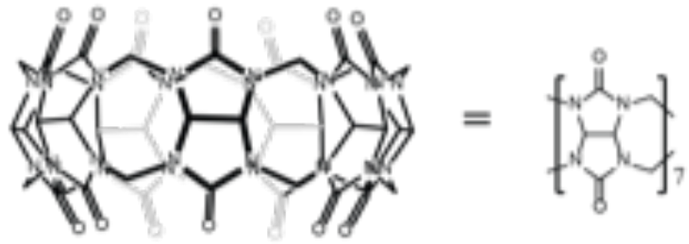


III. Estimate the binding free energy for each ligand

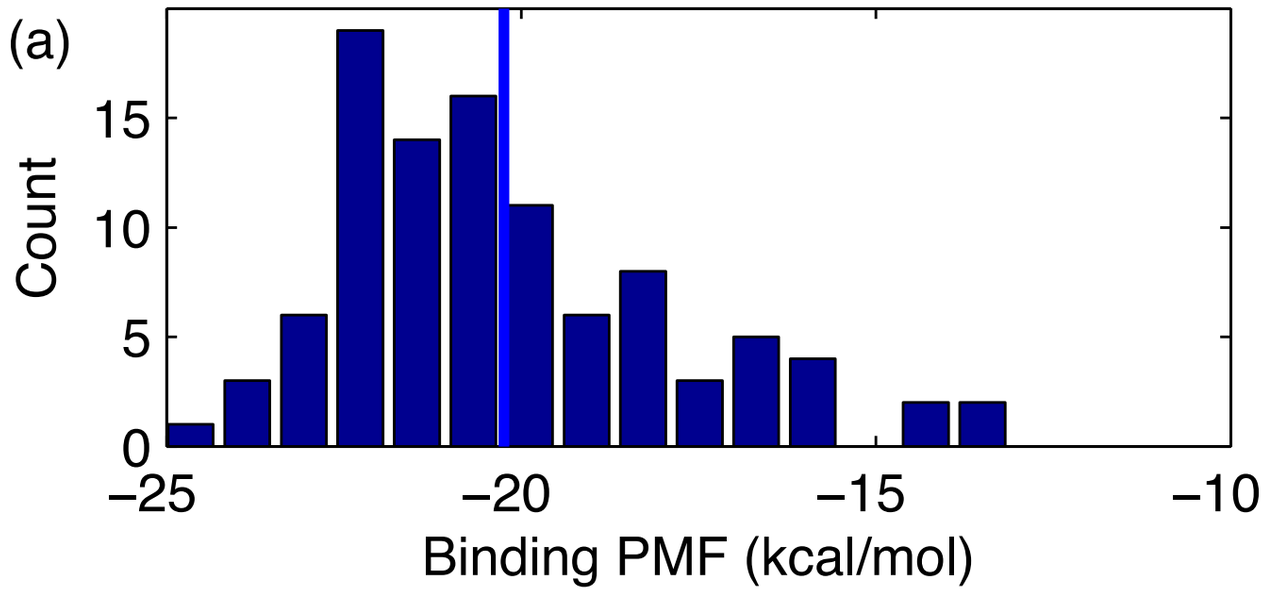
Sample mean of exponential average

$$\hat{B}(r_R) = -\beta^{-1} \ln \frac{1}{N} \sum_{n=1}^{N} e^{-\beta \Psi(r_{RL,n})}$$

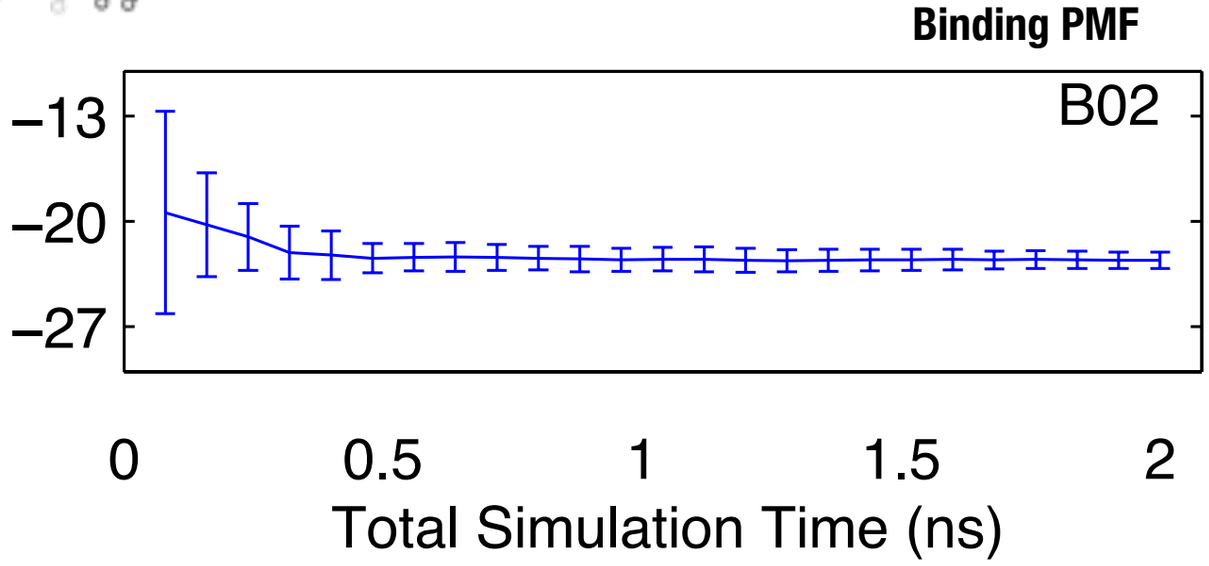
Demonstration on Cucurbit[7]uril



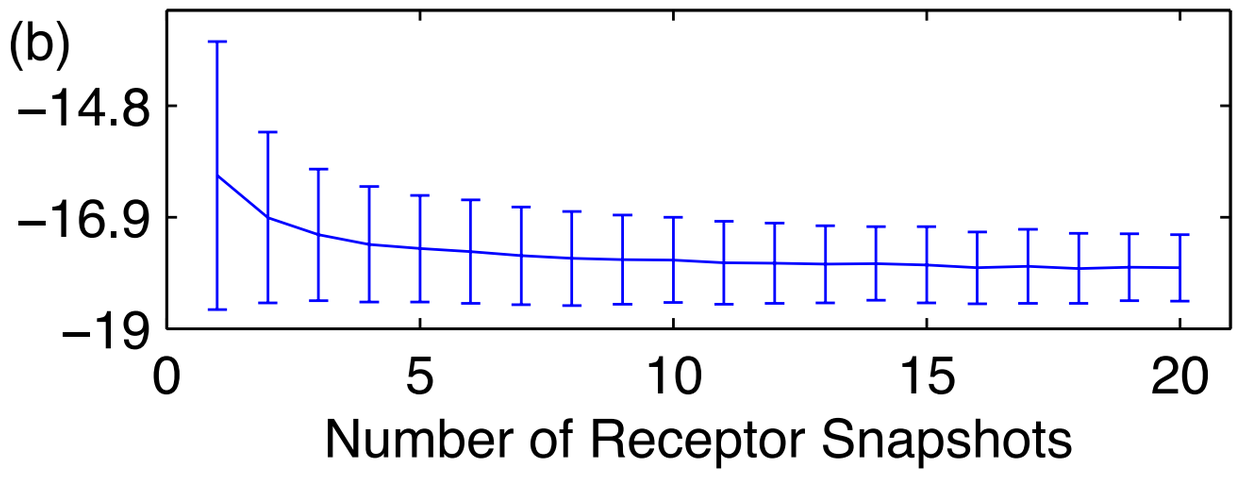
Binding PMFs using Hamiltonian replica exchange in NAMD

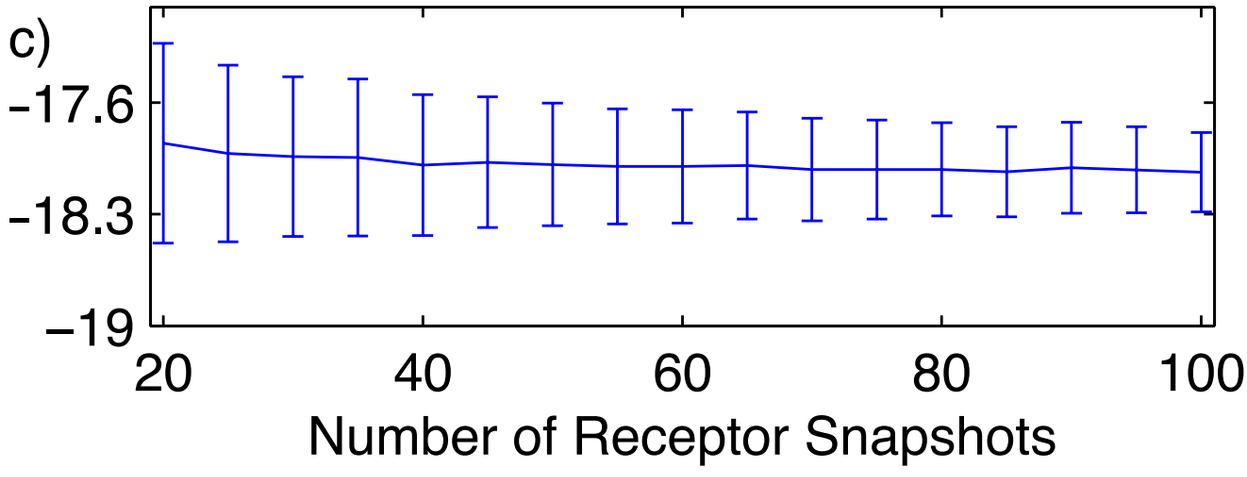


Binding Free Energy



100 receptor snapshots from standard molecular dynamics





Minh, Journal of Chemical Physics 137:104106, 2012.

Ligand $\hat{B}(r_R)$ $\Delta\hat{G}^\circ$	$\min\{\Psi(r_R)\}$ $\min\{\hat{B}(r_R)\}$	$\min\{\Psi(r_R)\}$ EXP	HREX $\min\{\hat{B}(r_R)\}$	HREX EXP
AD1	– 28.6	– 27.2	– 22.0	– 20.1
AD2	– 36.4	– 34.6	– 27.6	– 25.4
AD3	– 38.1	– 36.8	– 27.6	– 26.2
AD4	– 43.1	– 40.4	– 29.8	– 27.1
AD5	– 35.8	– 33.6	– 26.8	– 24.4
B02	– 29.8	– 27.9	– 21.0	– 18.1
B05	– 37.9	– 35.6	– 23.7	– 21.4
B11	– 48.5	– 45.7	– 23.1	– 20.5
F01	– 22.7	– 21.3	– 10.2	– 7.6
F02	– 30.9	– 28.8	– 17.0	– 14.6
F03	– 28.7	– 27.0	– 14.5	– 13.2
F06	– 35.6	– 33.8	– 21.3	– 19.7
R^2_{ITC}	0.849	0.855	0.684	0.704
$RMSE_{ITC}$	17.3	15.3	5.8	4.5
R^2_{Gilson}	0.787	0.795	0.926	0.925
$RMSE_{Gilson}$	15.8	13.9	3.5	2.4
R^2_{Exp}	0.723	0.736	0.996	
$RMSE_{Exp}$	15.5	13.6	2.3	

Minh, Journal of Chemical Physics 137:104106, 2012.

Protein-ligand binding PMF estimation: the method

- Pre-calculated interaction energy grids
  - Not often used with MD
  - Linear scaling, not soft-core potential
    - easier potential energies
    - grids have no singularities
- Thermodynamic cycle includes high temperatures
- Hamiltonian replica exchange
  - Adaptive protocol based on constant thermodynamic length
  - No U-Turn sampler
  - MBAR for analysis

$$\mathcal{L}(\gamma) \equiv \int_0^1 \|\dot{\gamma}\|_\gamma dt = \int_0^1 \sqrt{\sum_{i,j} \dot{\gamma}^i g(\gamma)_{ij} \dot{\gamma}^j} dt,$$

$$g(\lambda)_{ij} \equiv \mathrm{cov}_\lambda(\partial_i \ell_\lambda, \partial_j \ell_\lambda) = \langle \partial_i \ell_\lambda(x) \cdot \partial_j \ell_\lambda(x) \rangle_\lambda,$$

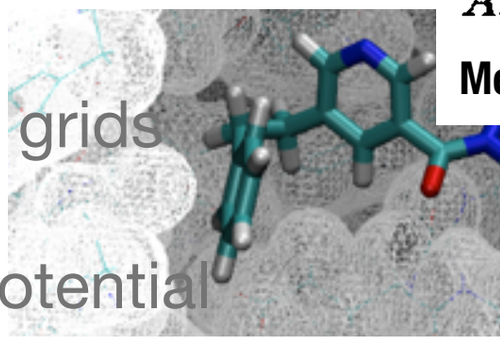
Shenfeld, Xu, Eastwood, Dror, Shaw. Physical Review E 2009

AMBER interaction energies

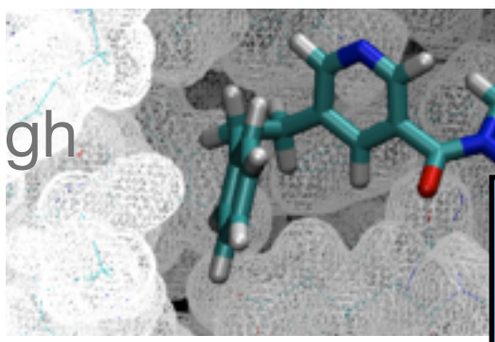
$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332.0 \frac{q_i q_j}{D r_{ij}} \right],$$

$$A_{ij} = \sqrt{A_{ii}} \sqrt{A_{jj}} \quad \text{and} \quad B_{ij} = \sqrt{B_{ii}} \sqrt{B_{jj}},$$

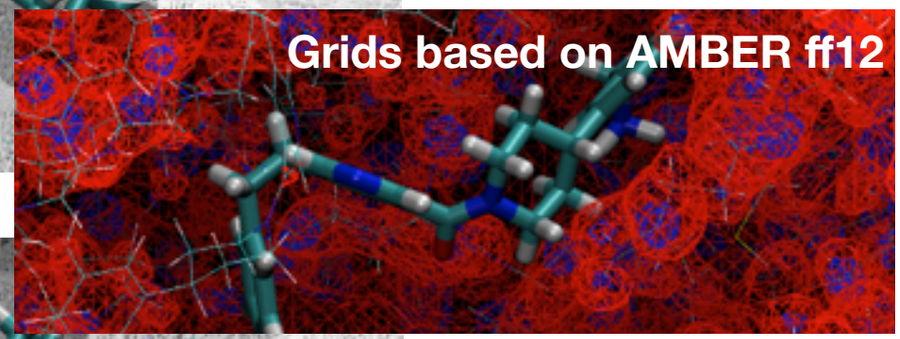
Meng, Shoichet, and Kuntz, J. Comput. Chem.1994



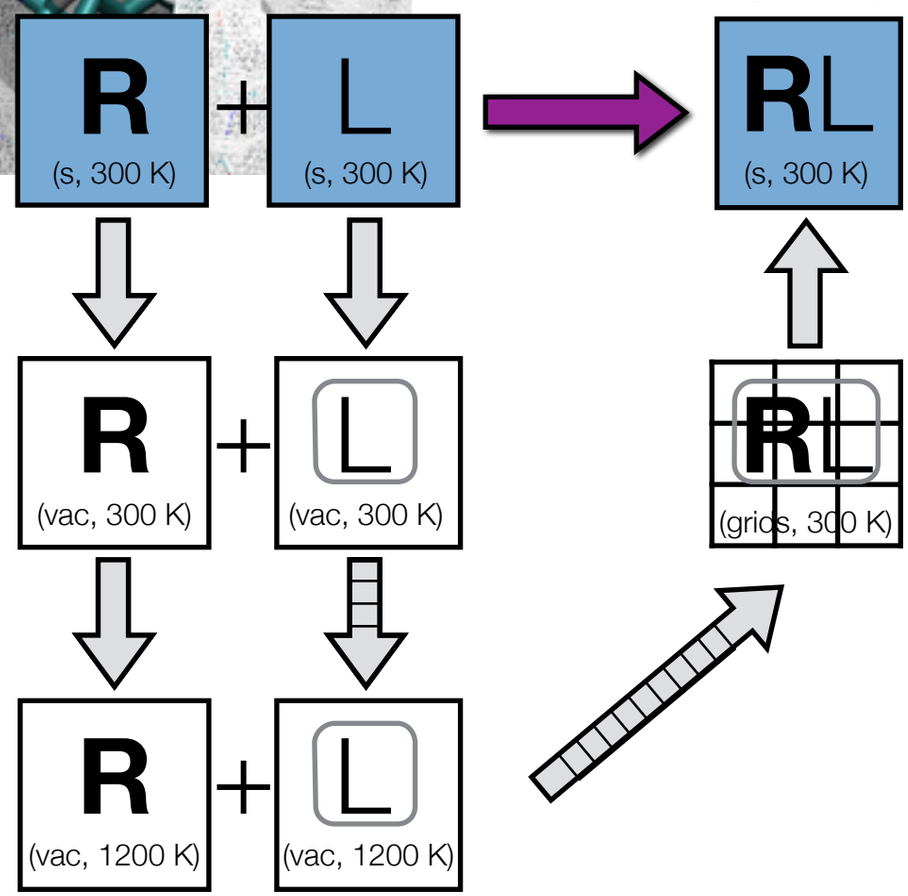
LJ repulsive



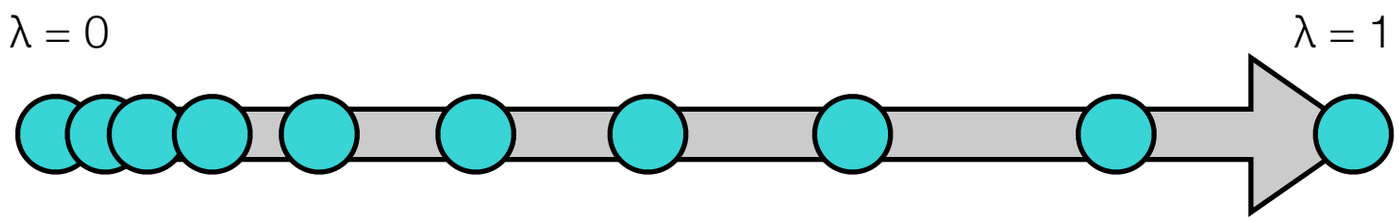
LJ attractive



Electrostatic (PB/SA using APBS)



On traversing thermodynamic state space



$$\mathcal{L}(\gamma) \equiv \int_0^1 \|\dot{\gamma}\|_\gamma dt = \int_0^1 \sqrt{\sum_{i,j} \dot{\gamma}^i g(\gamma)_{ij} \dot{\gamma}^j} dt,$$

$$g(\lambda)_{ij} \equiv \mathrm{cov}_\lambda(\partial_i \ell_\lambda, \partial_j \ell_\lambda) = \langle \partial_i \ell_\lambda(x) \cdot \partial_j \ell_\lambda(x) \rangle_\lambda,$$

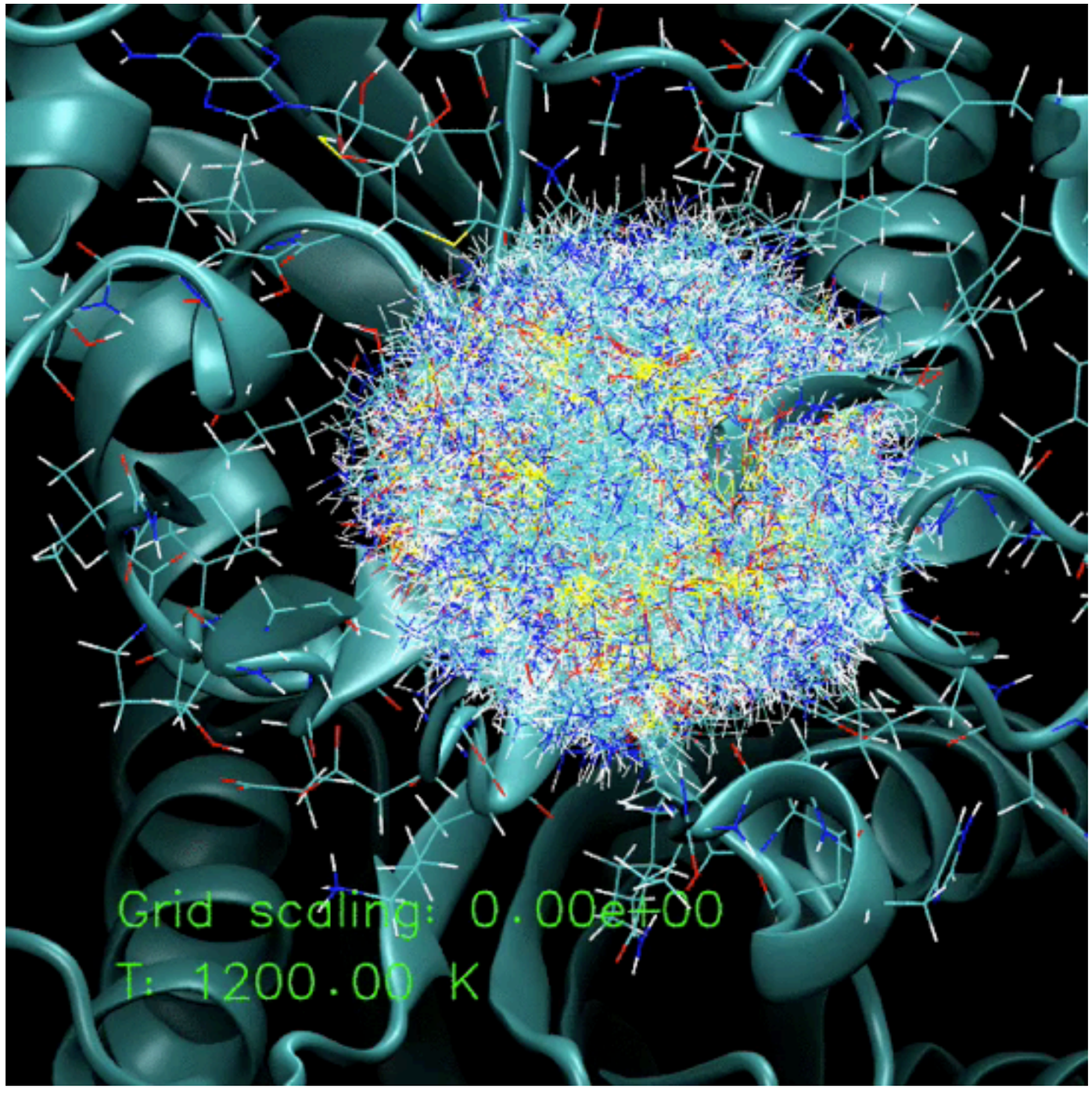
Shenfeld, Xu, Eastwood, Dror, Shaw. Physical Review E 2009

Initialization Strategy:
1. start with n random seeds
2. sample state K
3. **determine state K+1**
4. resample (obtain n seeds for K+1 from samples for state K)

$$h_\lambda = \beta[U_{MM}(x) + \lambda\Psi(x)]$$
$$g(\lambda) = \beta^2\lambda^2\sigma^2[\Psi(x)] + C$$
$$\frac{d\mathcal{L}(\lambda)}{dt} = \beta\lambda\sigma[\Psi(x)]\frac{d\lambda}{dt}$$

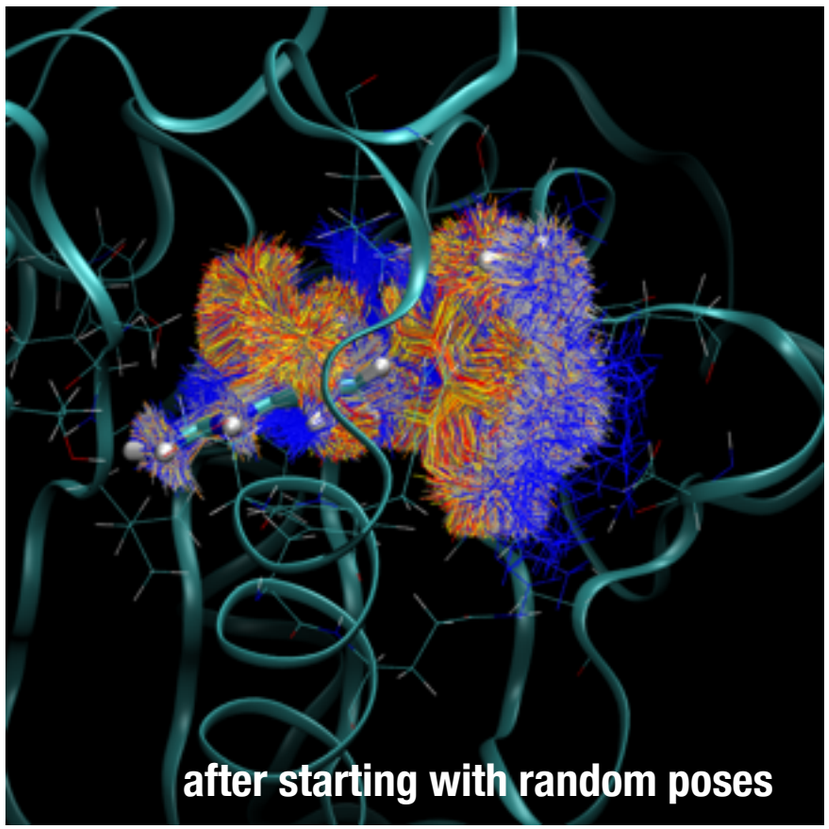
$$\Delta\lambda^i = \frac{s}{\sigma_0\left[\partial^i l_\lambda\right]}$$

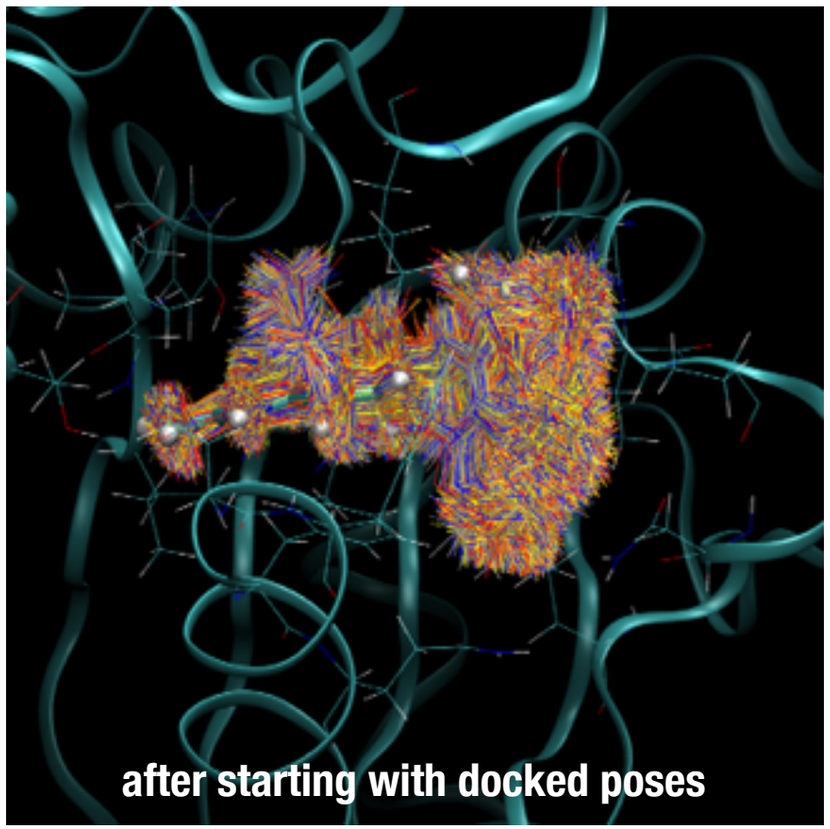
Example ensembles: 1hnn (adrenaline synthesis)

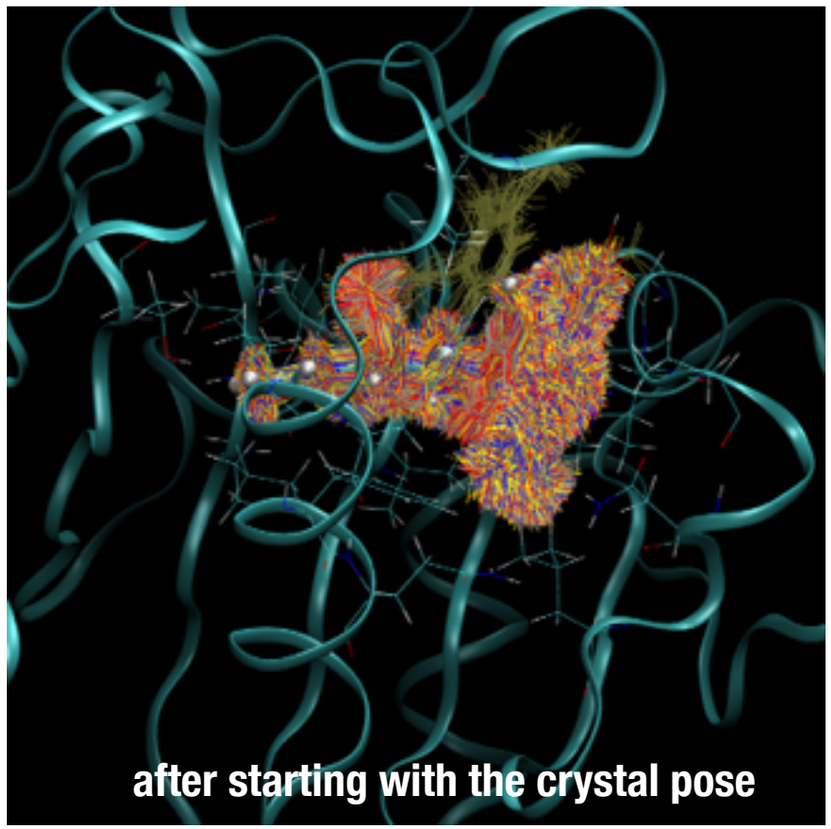


Protein-ligand binding PMF estimation: lessons I. it is best to start from docked configurations

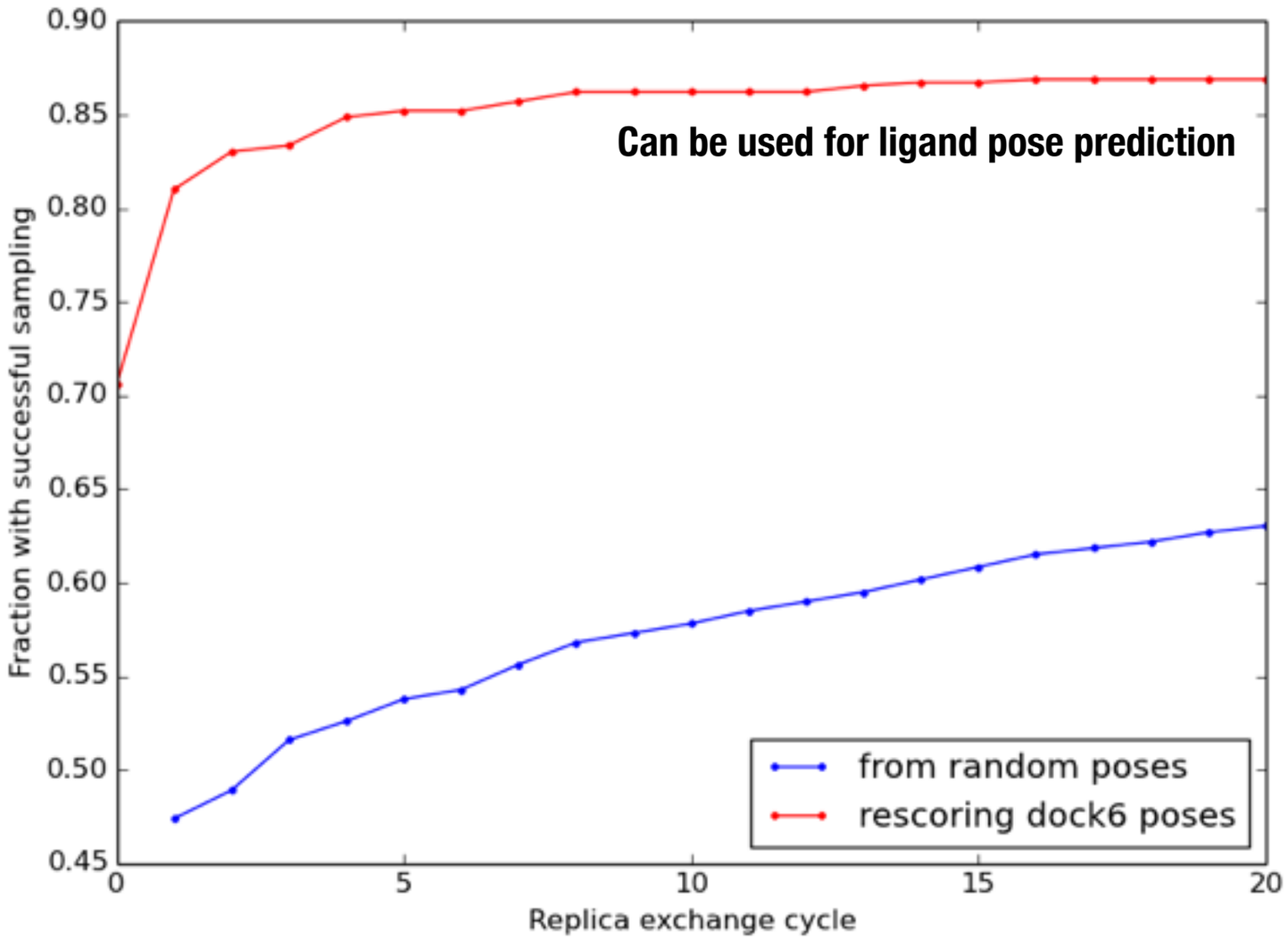
Redocking to 1s3v (dihydrofolate reductase)
Seven independent sampled ensembles of ligands
fully interacting with the grid at 300 K





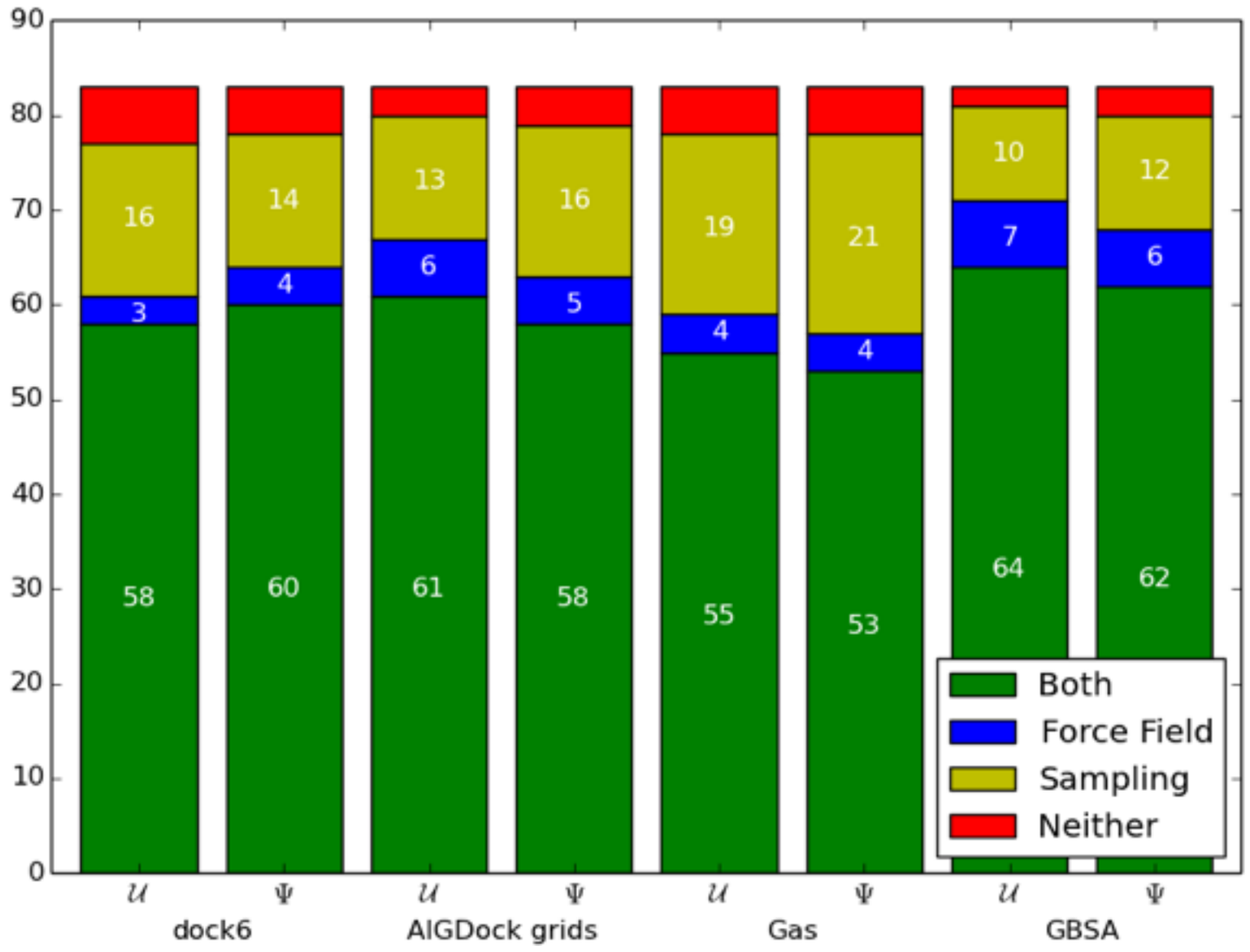


I. it is best to start from docked configurations



Redocking to the Astex Diverse Set (Hartshorn et al, J. Med. Chem. 2007) with UCSF dock6
Sampling success = obtain crystal pose (with 2 A RMSD) in final thermodynamic state
Each cycle is 1 to 1.5 hrs on a single CPU

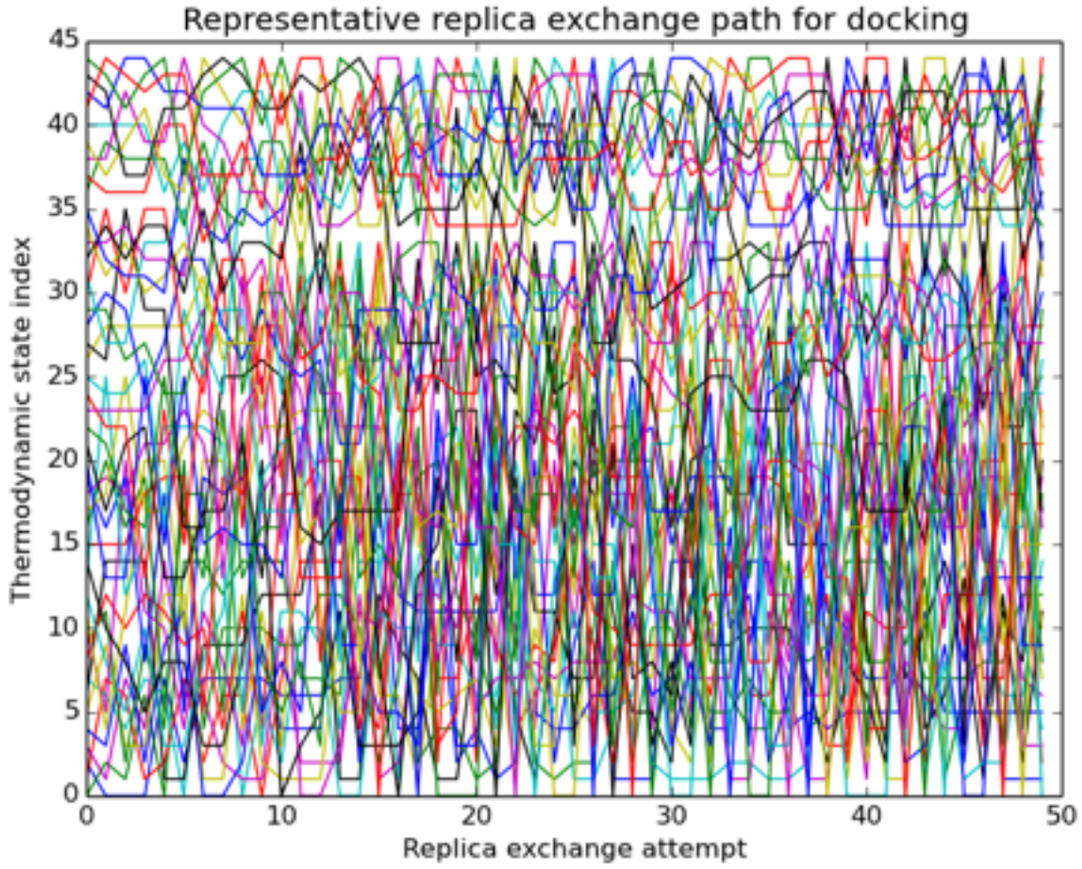
II. MM force fields can improve pose prediction

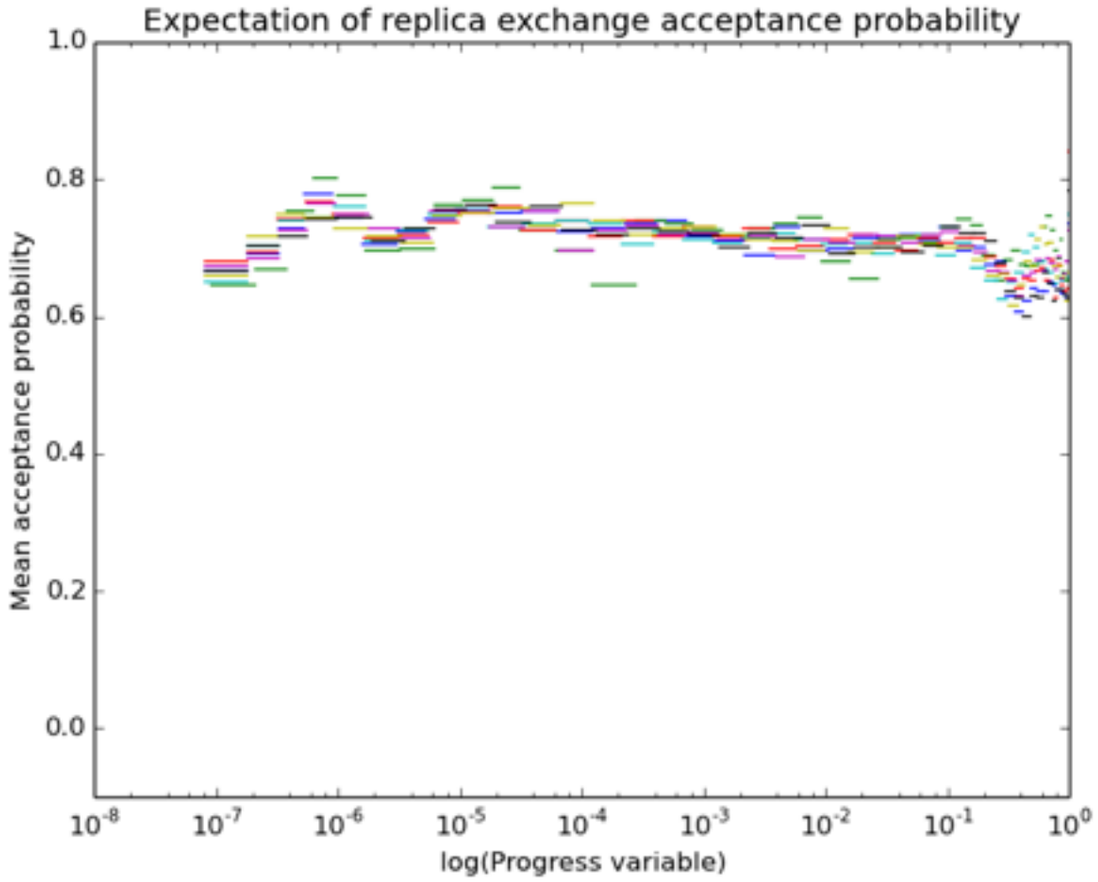


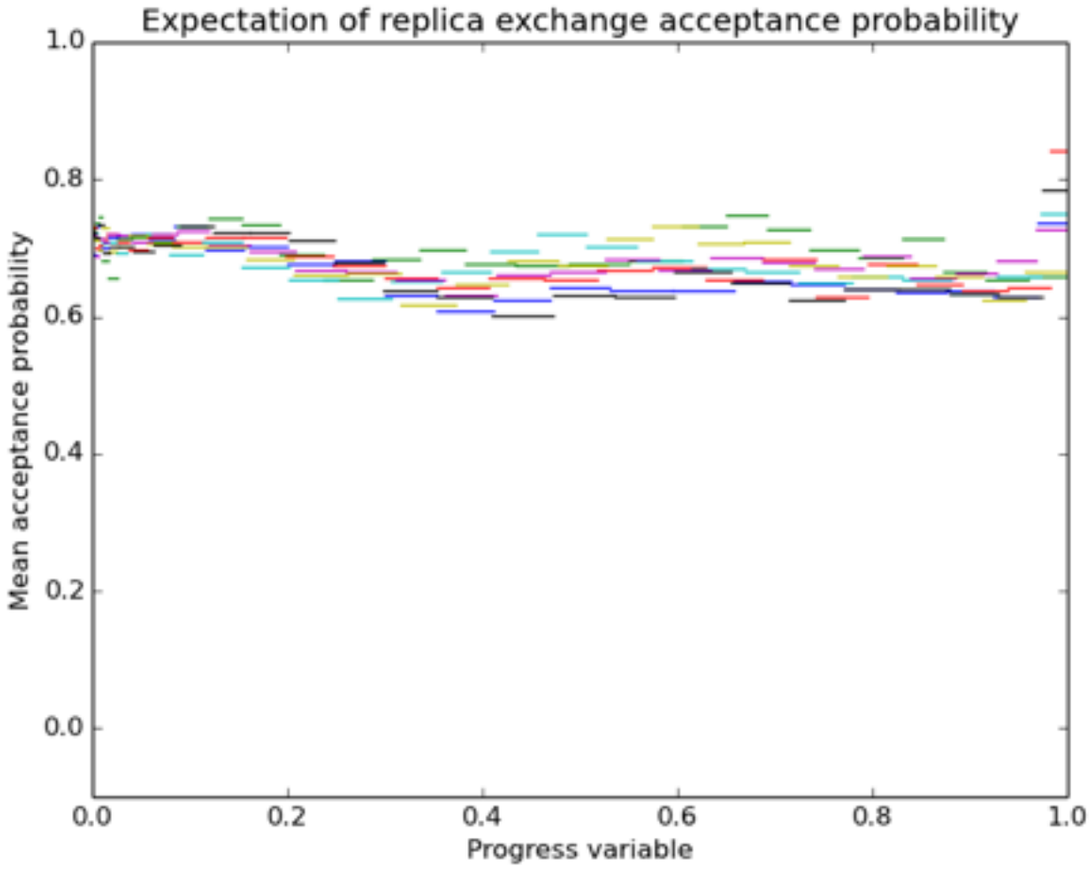
Redocking to the Astex Diverse Set (Hartshorn et al, J. Med. Chem. 2007) with UCSF dock6
Sampling success = obtain crystal pose (with 2 A RMSD), in 74/85 complexes
Force field success = crystal pose is the lowest energy structure

III. the adaptive protocol works

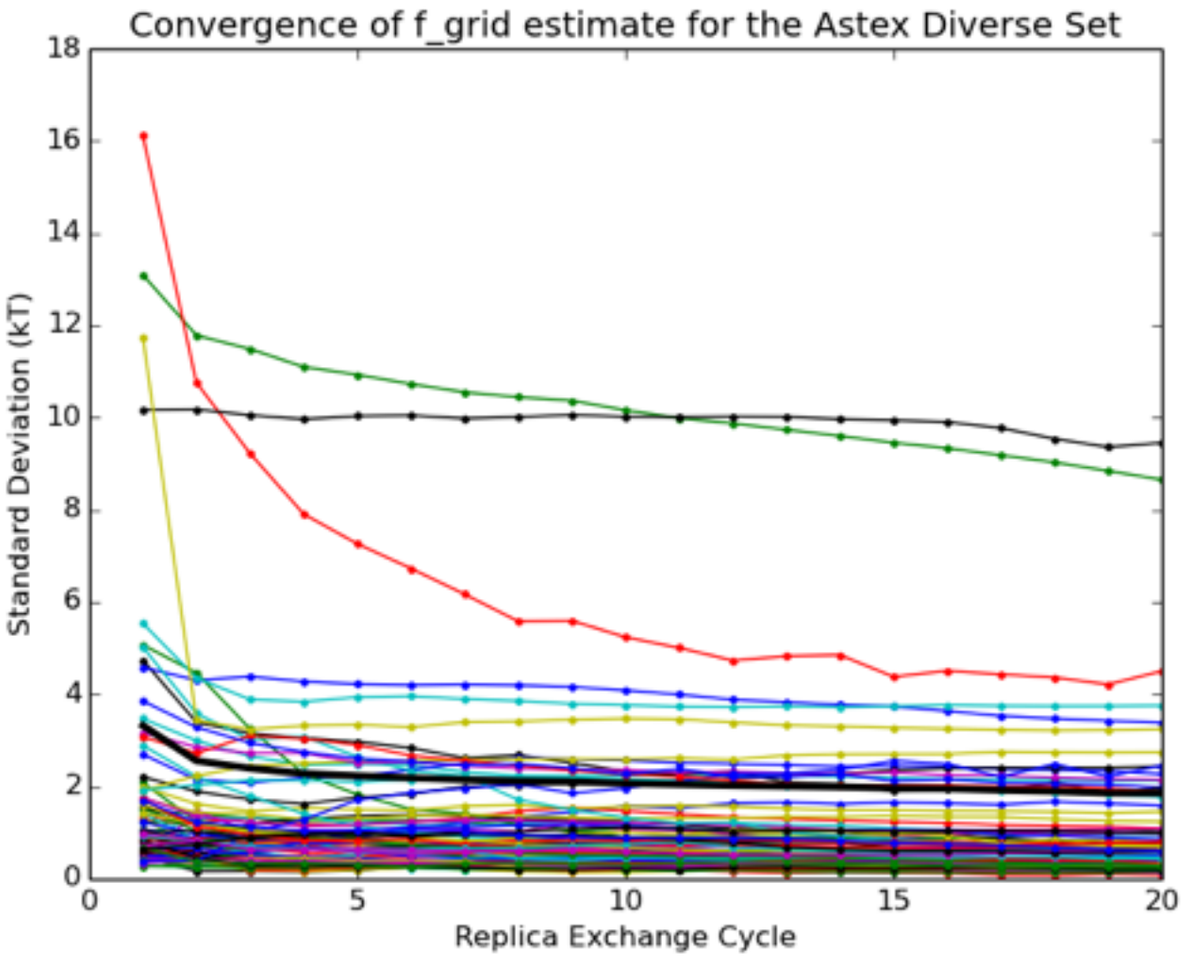
Redocking to 1s3v (dihydrofolate reductase)
Seven independent binding PMF calculations after starting with docked poses

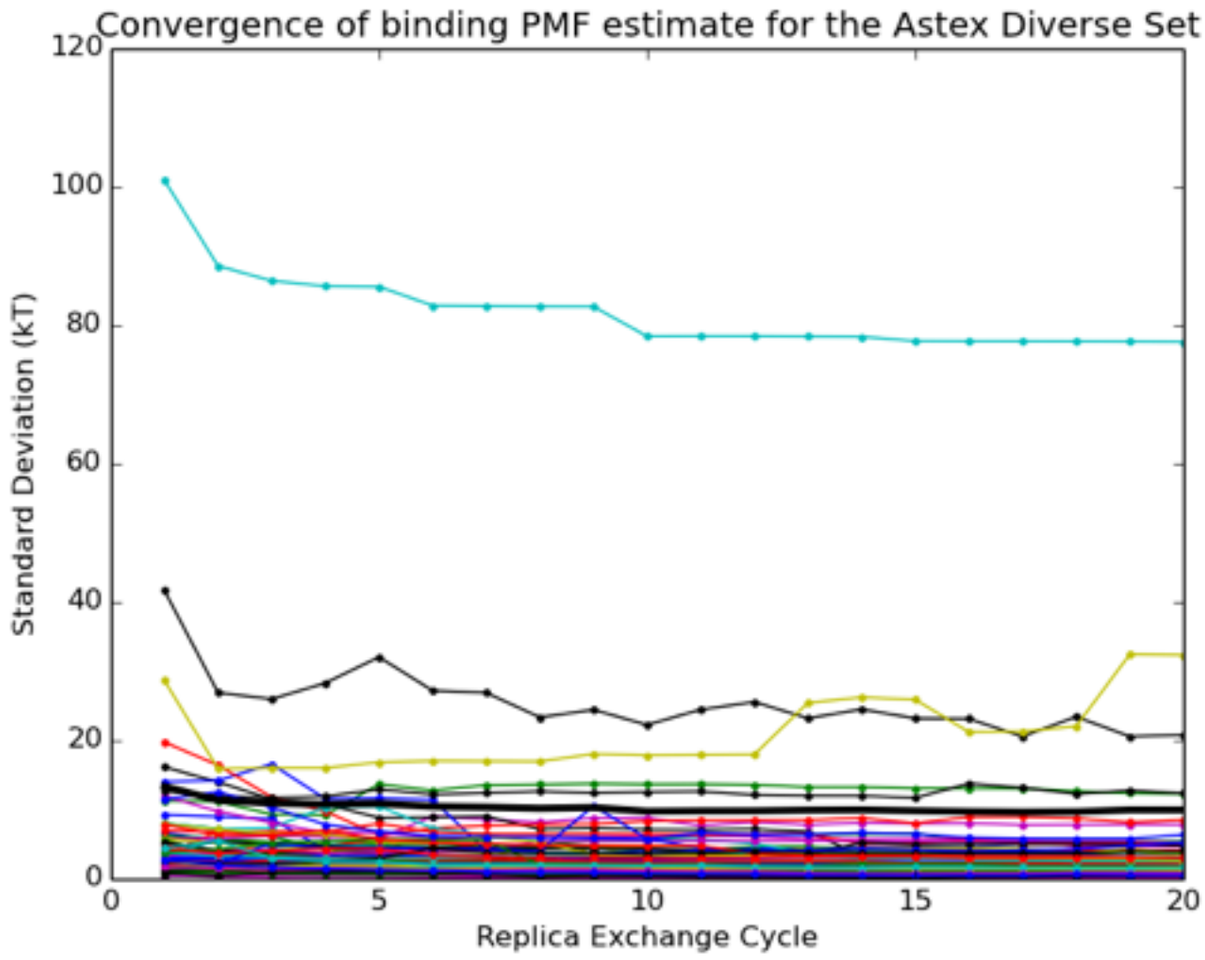




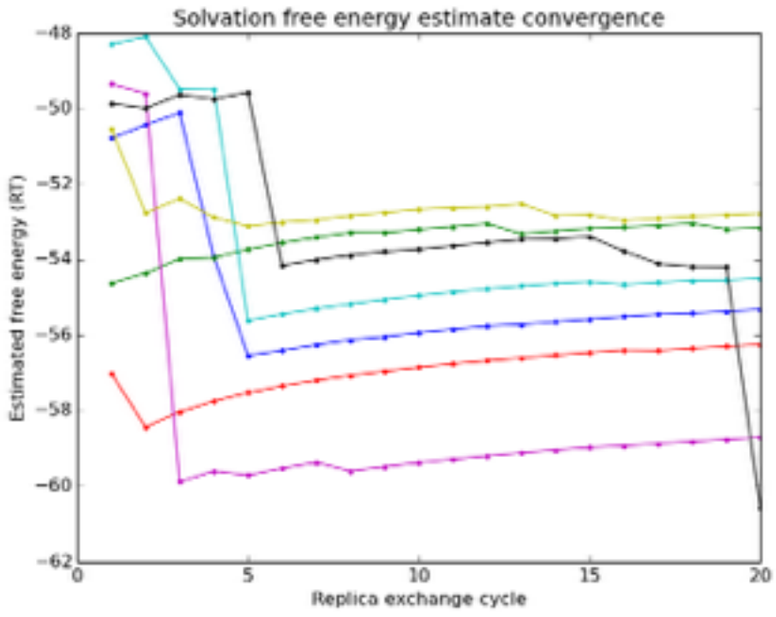


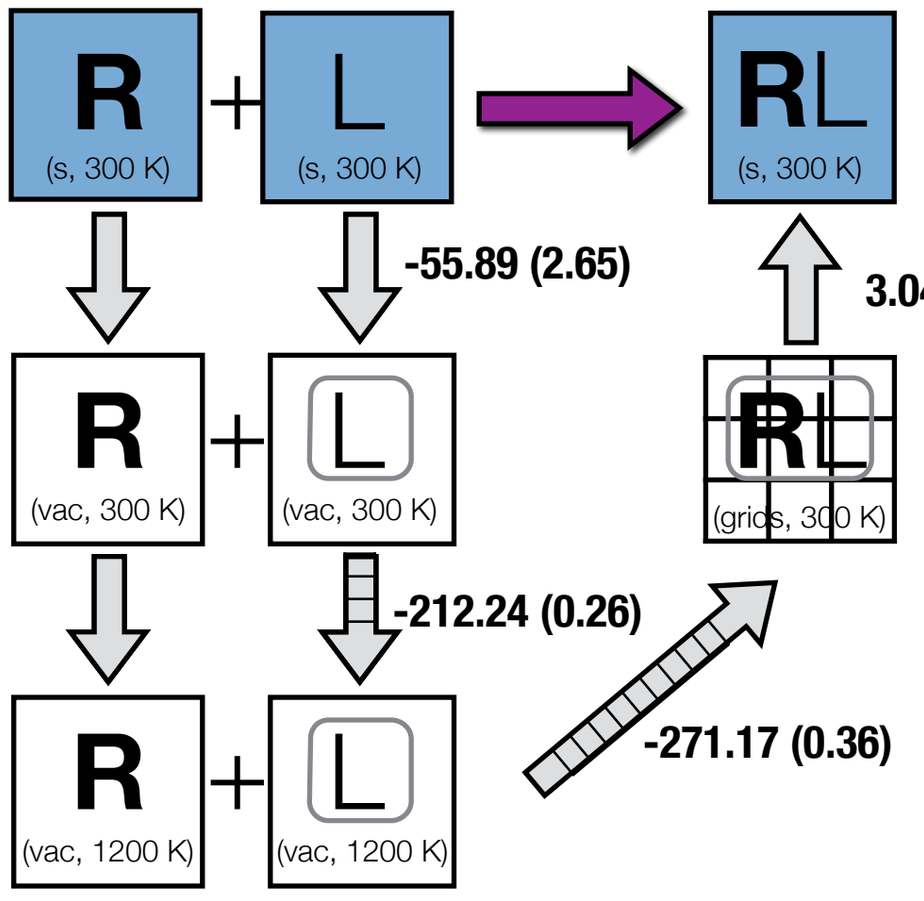
IV. convergence is highly system-depedendent

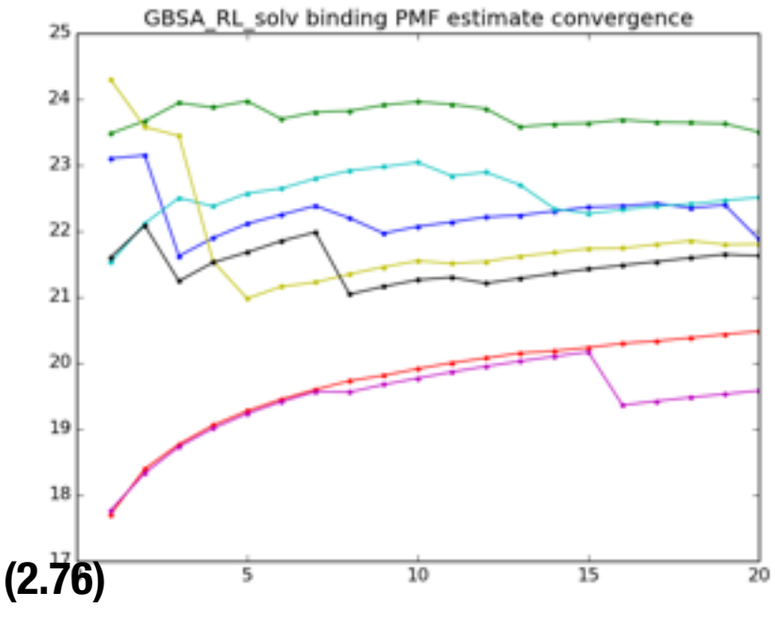


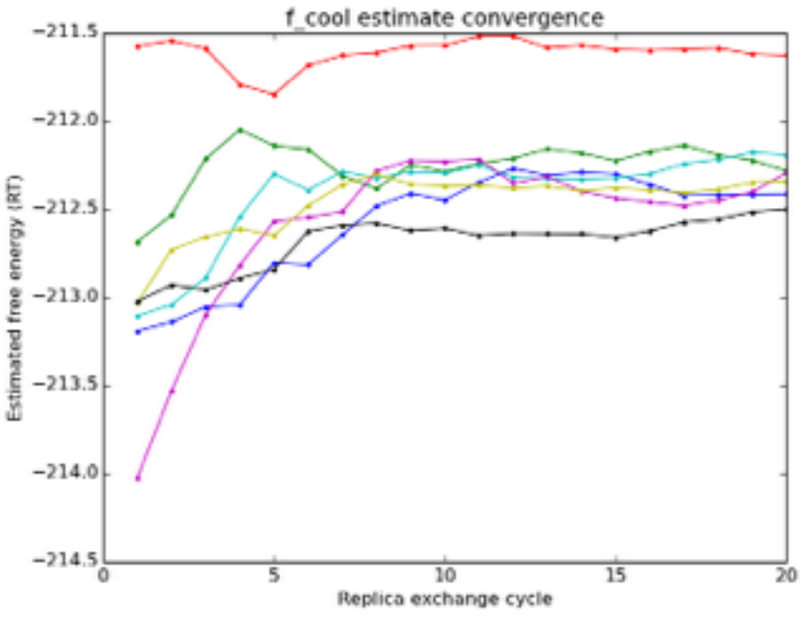


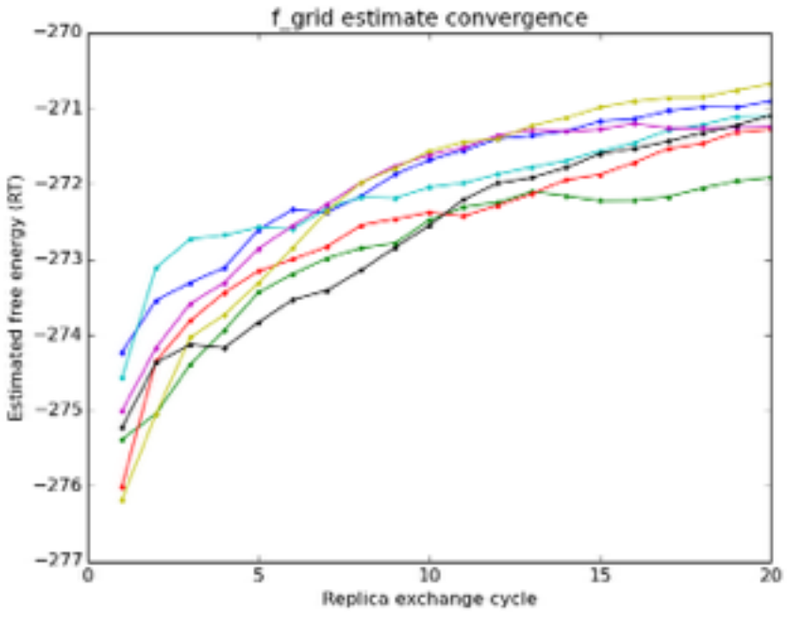
IV. configuration space overlap between GBSA and gas phase limits binding PMF precision





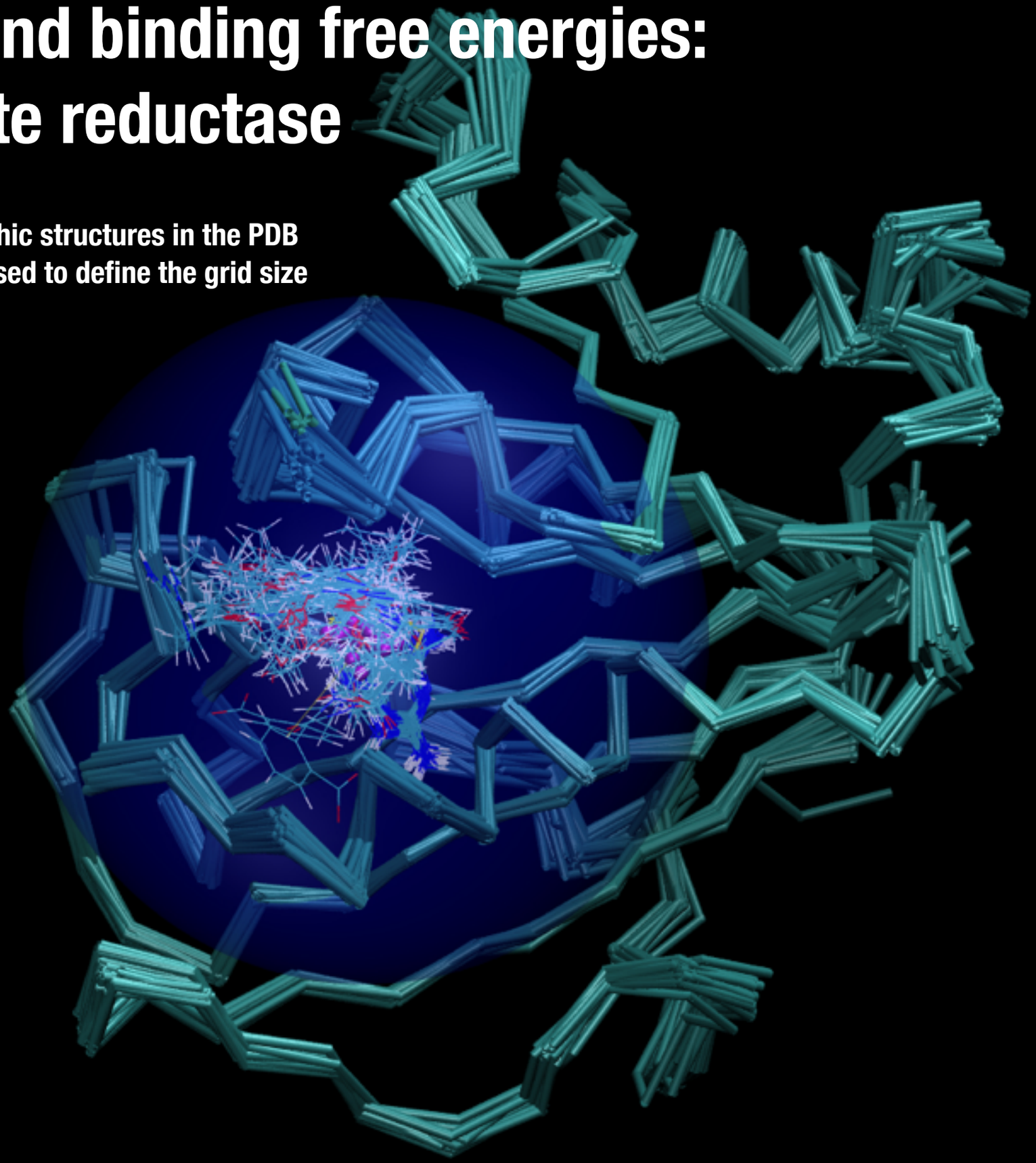






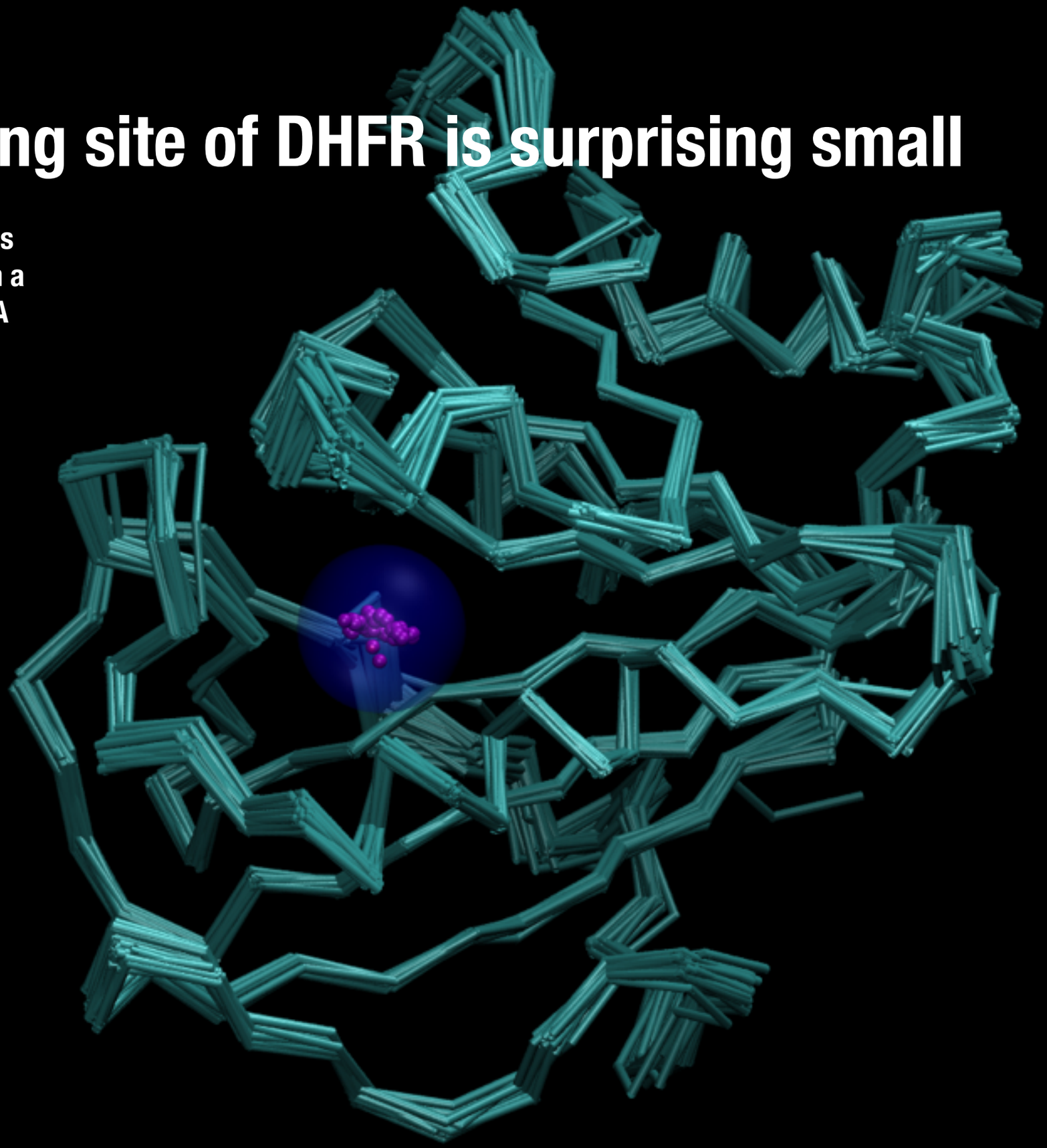
Protein-ligand binding free energies: dihydrofolate reductase

There are 63 crystallographic structures in the PDB
The span of ligands was used to define the grid size

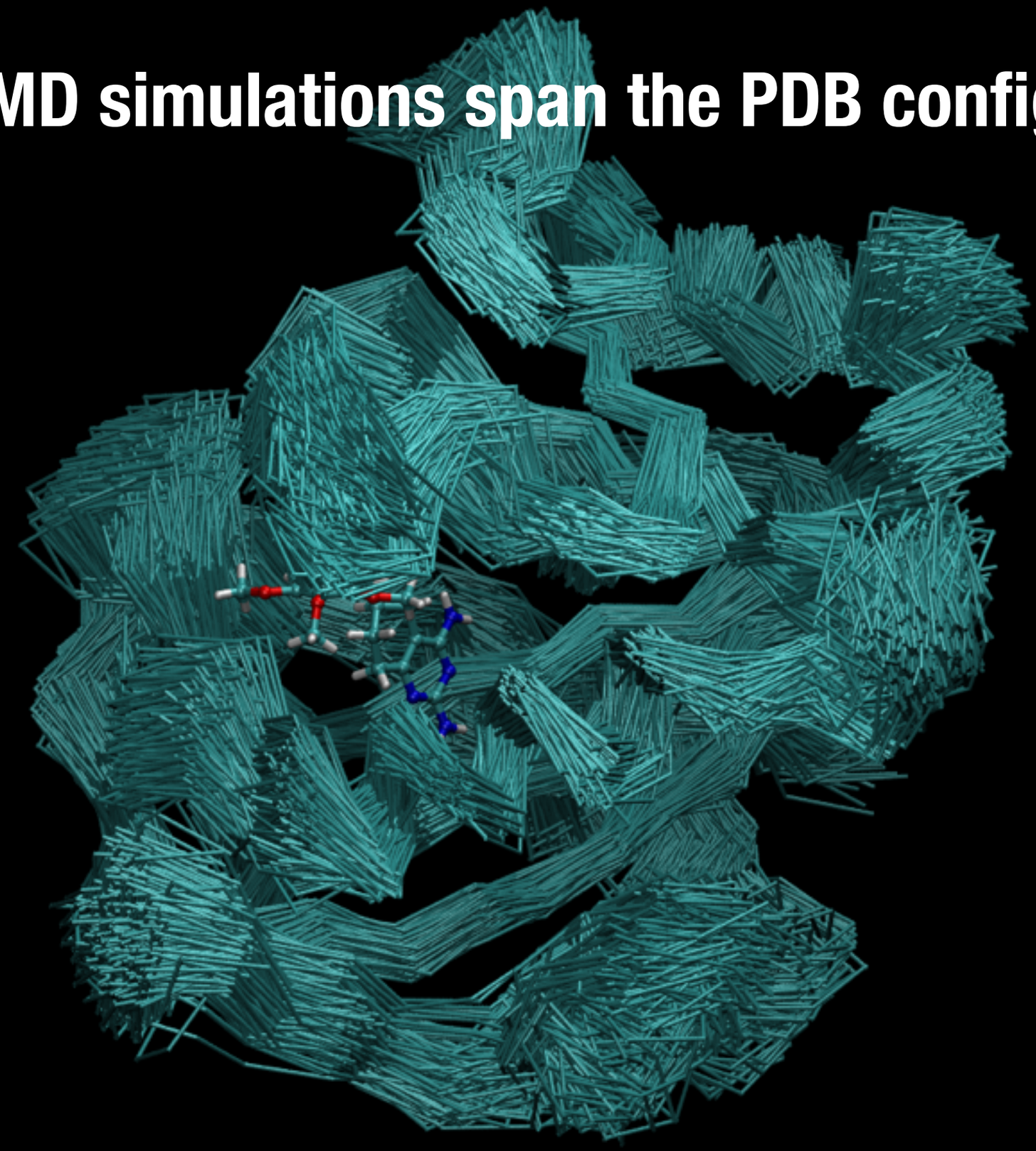


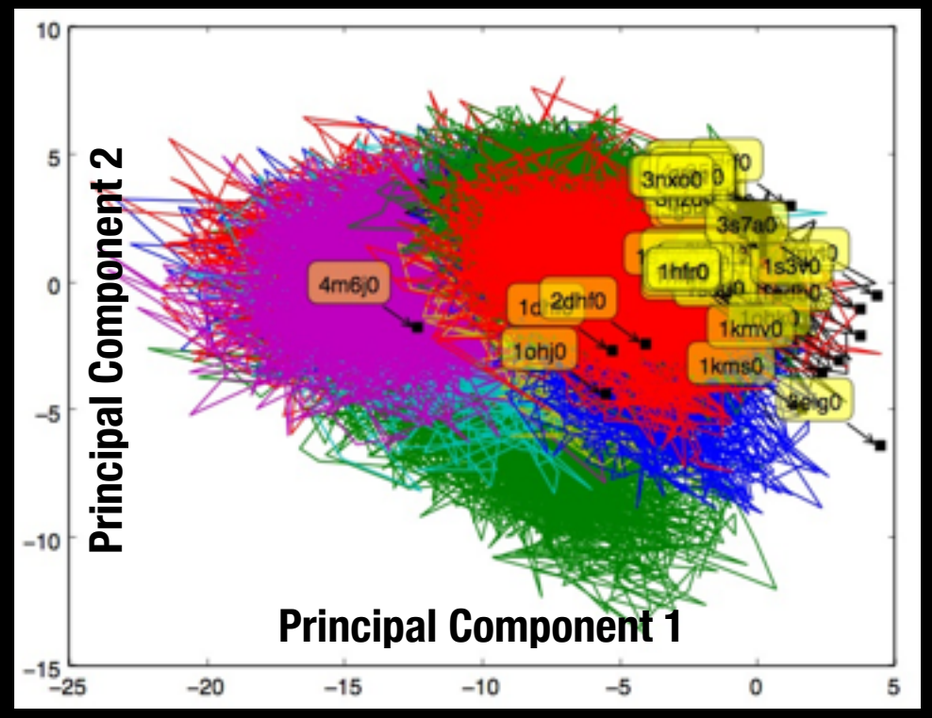
The binding site of DHFR is surprising small

Ligand center of mass coordinates fit within a sphere of radius 3.5 A



MD simulations span the PDB configuration space

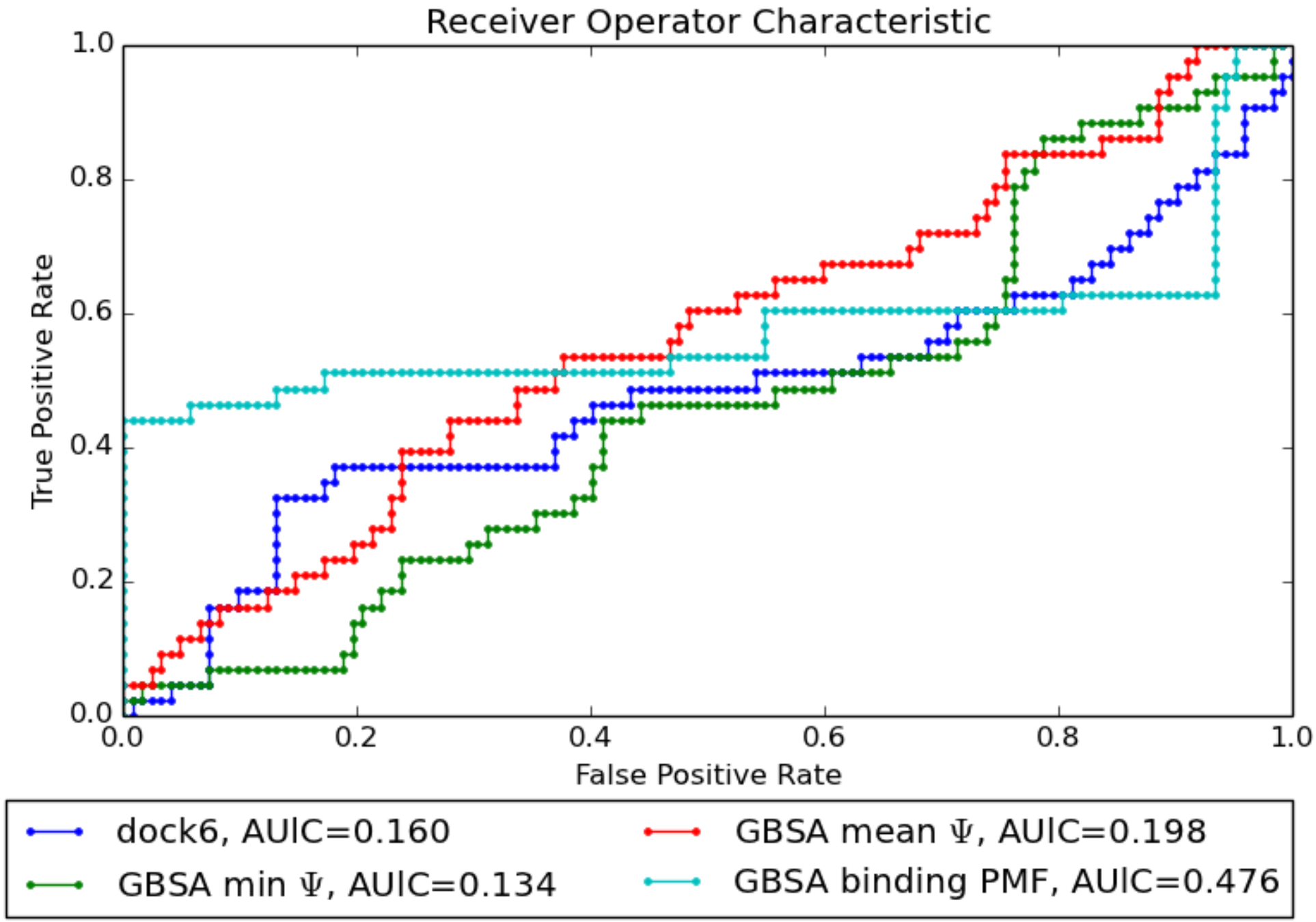




Superimposed snapshots from
5 MD simulations starting from 1pdb (apo)
and 5 MD simulations starting from 1s3v
totaling 2 microseconds

The rmsd between crystal structures
and the snapshots ranges from 0.7 to 3.0 A

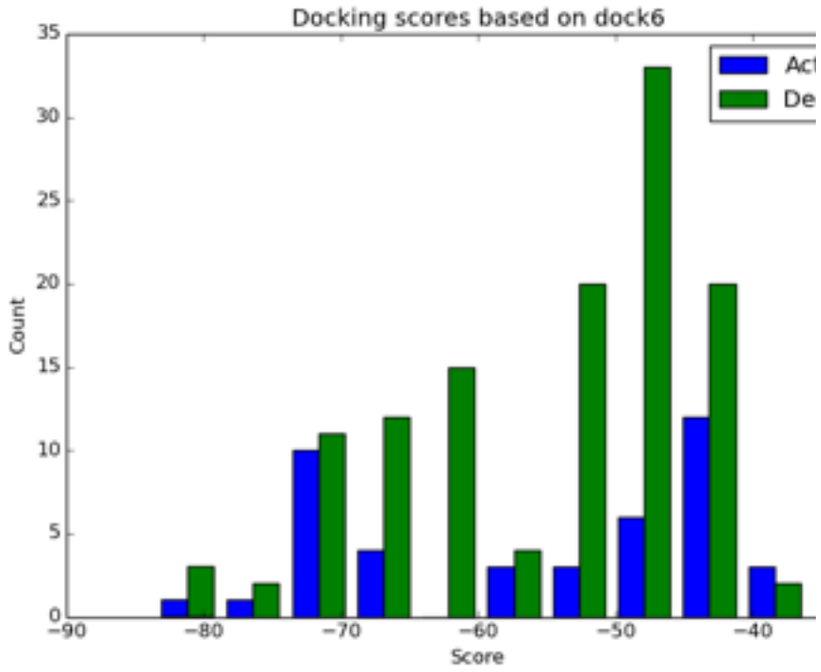
Binding PMFs improve docking performance: ROC

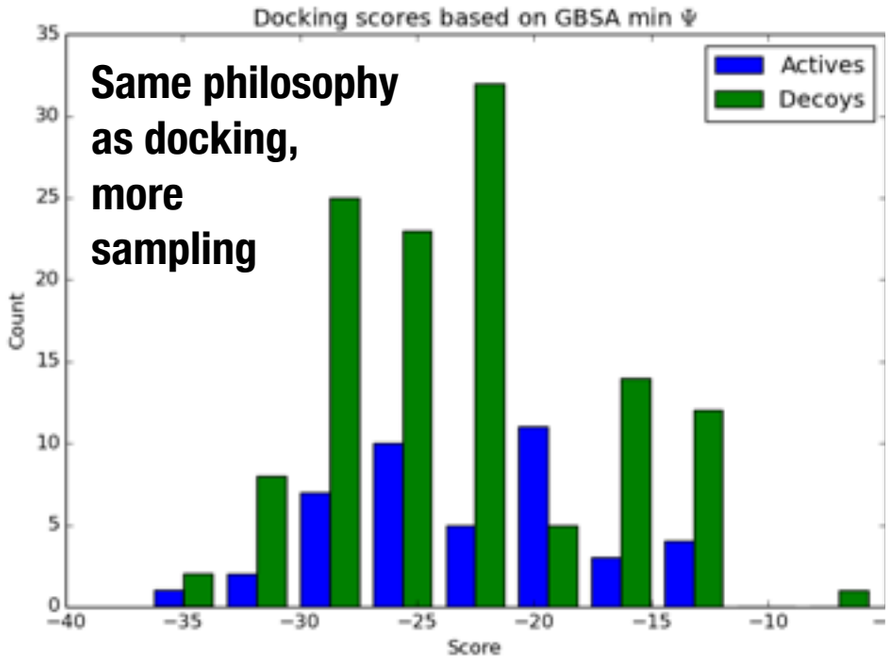


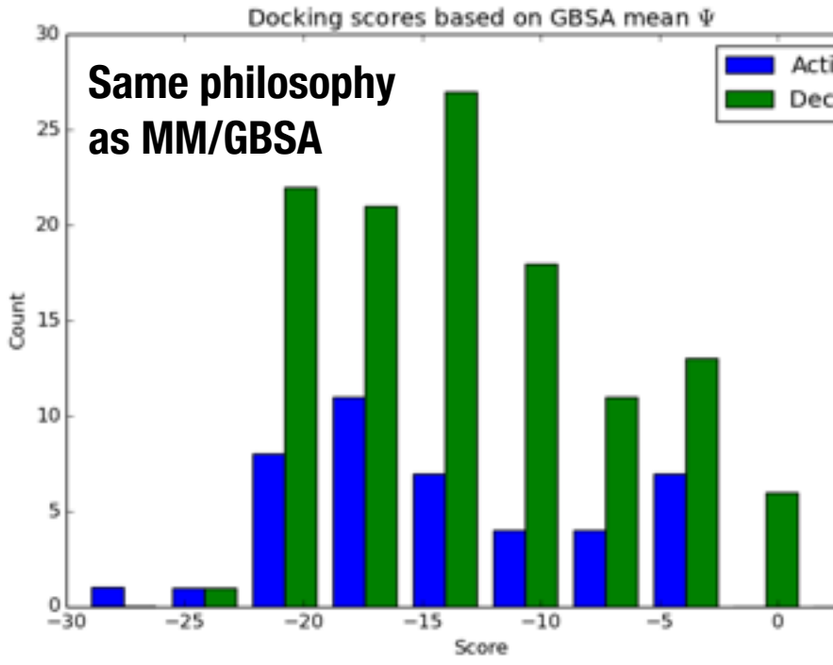
Random AUIC is 0.14462

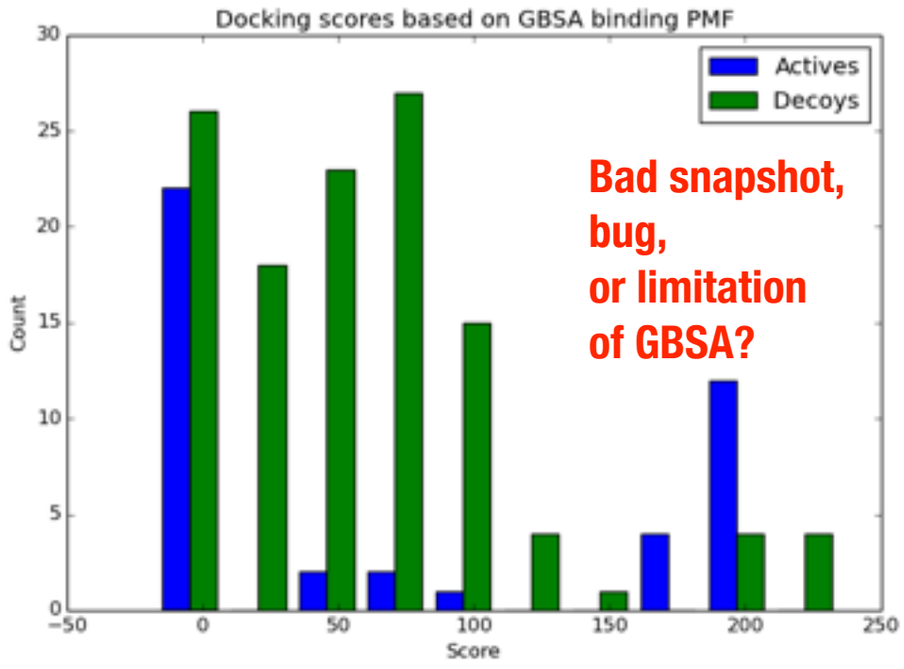
Binding PMFs improve docking performance

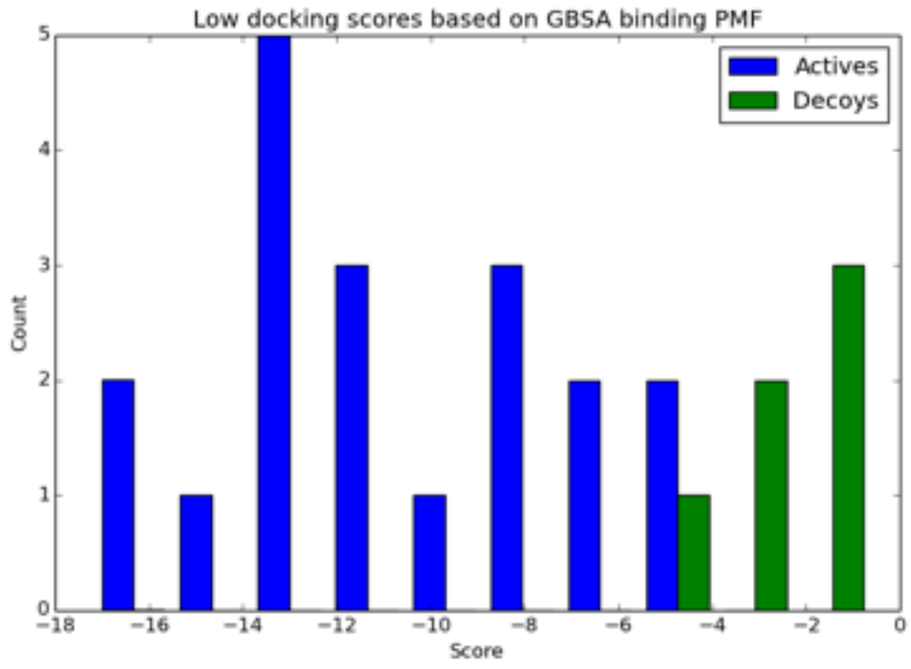
43 actives from DHFR crystal structures
122 decoys using DUD-E server (http://dude.docking.org/)
Docked to first snapshot of simulation starting from 1pdb, an apo structure of DHFR
All scores are in kcal/mol
Site confinement free energy is 1.08 kcal/mol



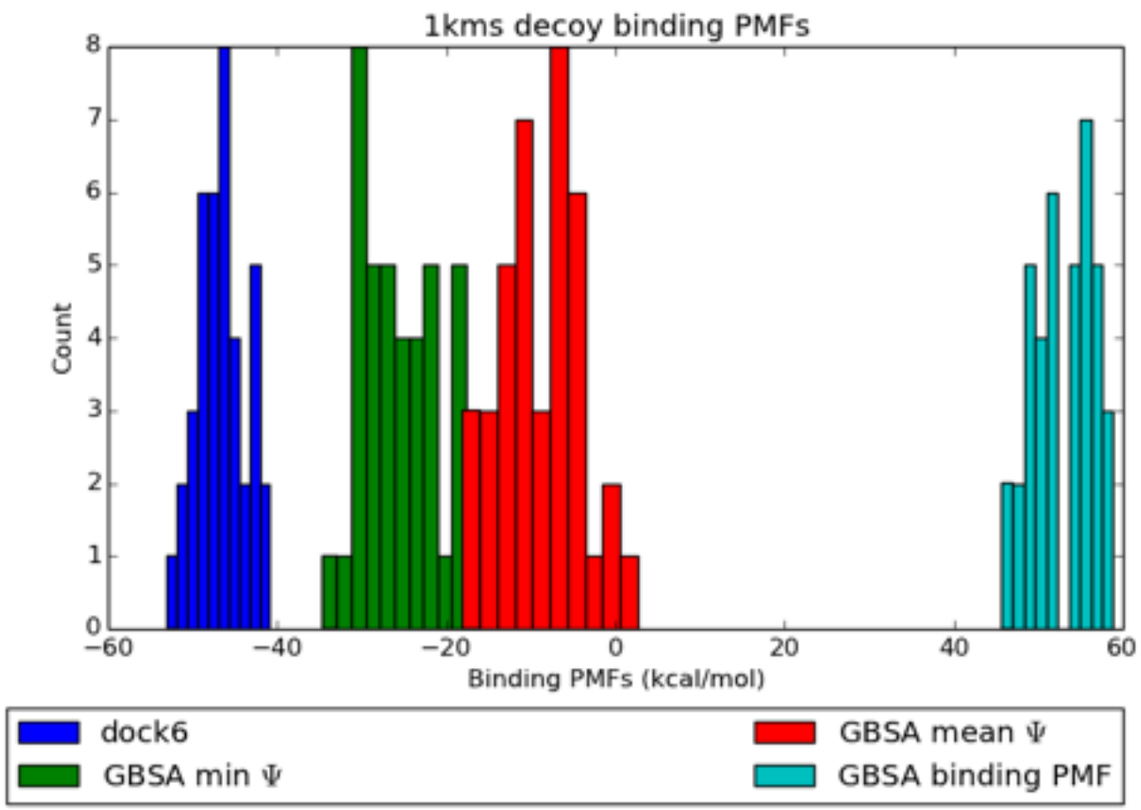


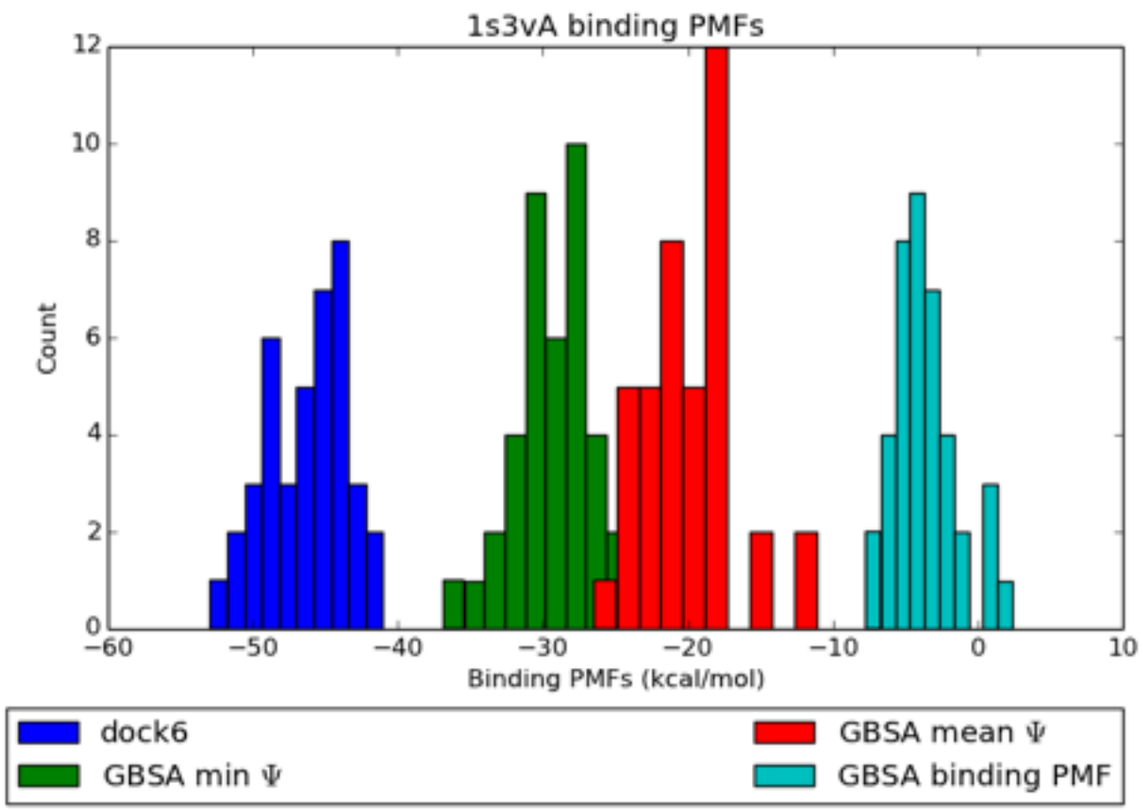


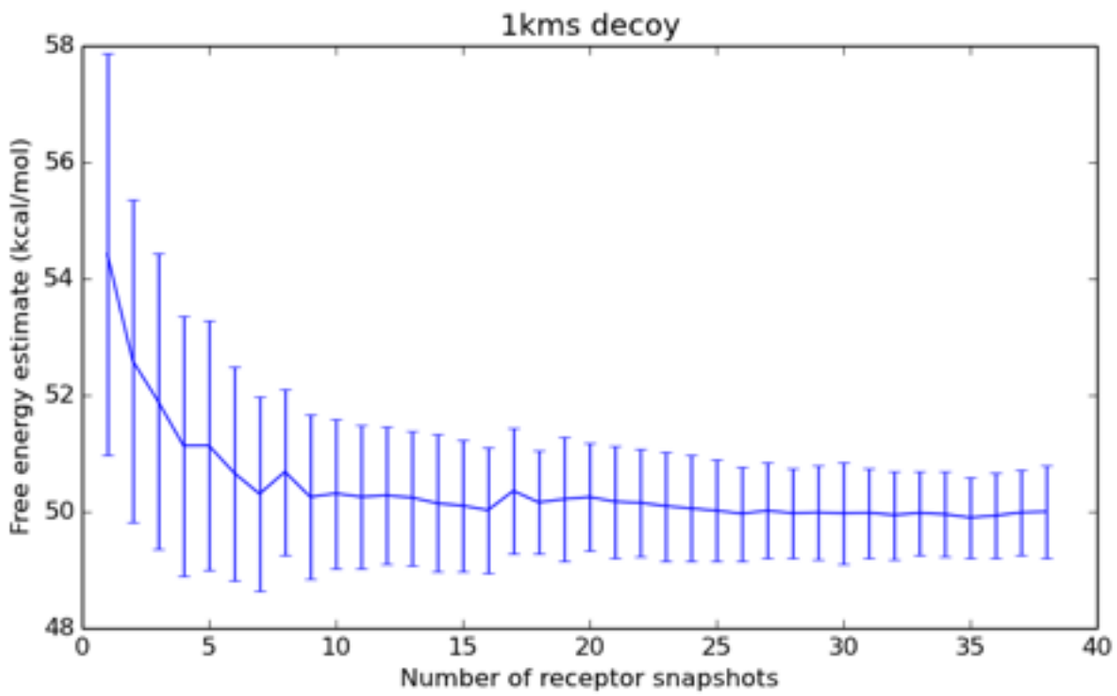


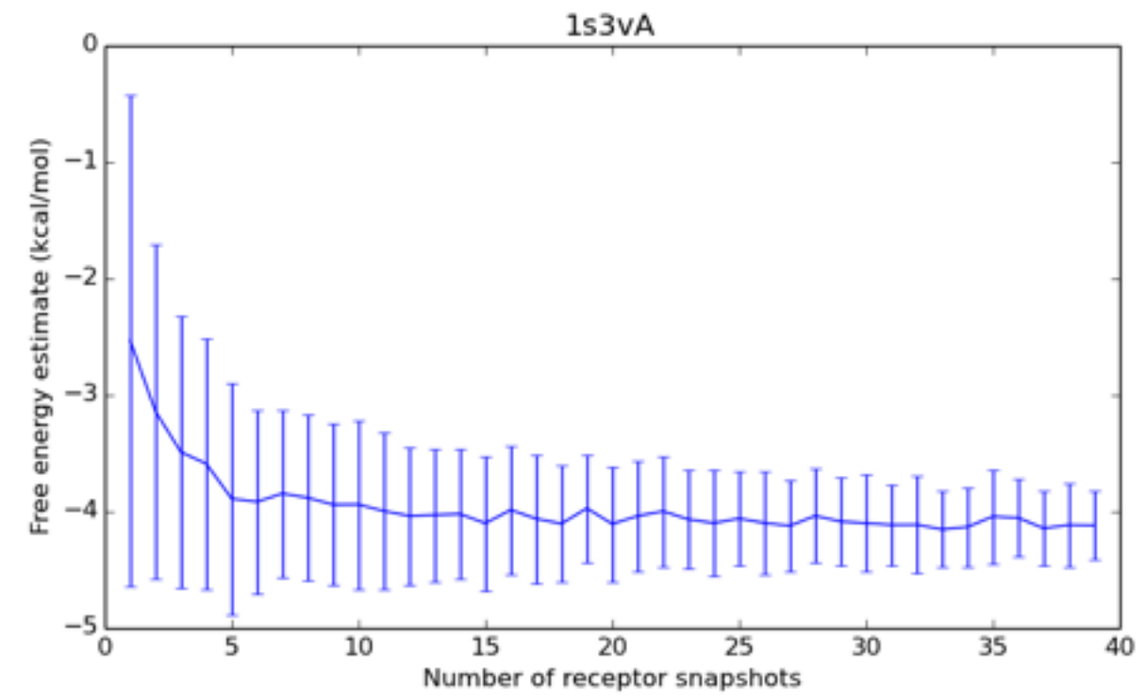


Free energies may require surprisingly few snapshots









Future directions?

- Enhanced sampling methods/M2 for faster binding PMF estimation
- Different strategies for receptor sampling and weighting to account for induced fit: umbrella sampling, Markov State Models.
- Improved solvent models for sampling and postprocessing
- New systems for testing and applications
- Open to suggestions and collaborations