# Improving virtual screening through physics-based models

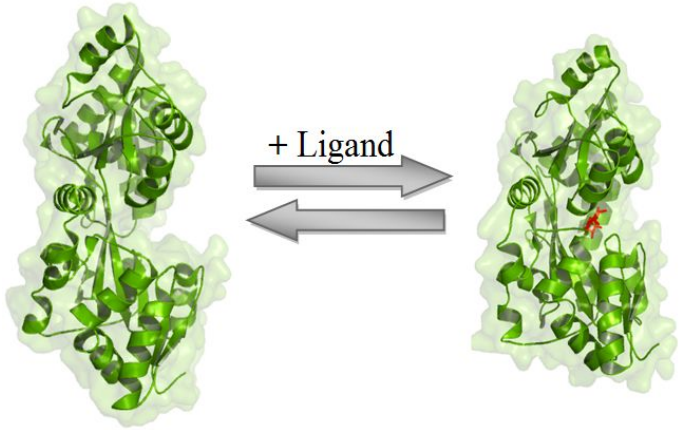
# Introduction

# Virtual Screening

- Can we computationally predict the **binding affinity** and **pose** of protein-ligand interactions?

# Predicting Binding Affinity

- What? Binding strength between ligand and target (protein)
- How? Estimate **free energy** differences (ΔF)




***State A: Open*** ***State B: Closed***

# Free Energy (review)

- Amount of work the system can perform
- Gibbs free energy ($G = H - TS$)
- Helmholtz free energy ($A = U - TS$)
- $A \leftrightarrow B$
  - $K_{eq} = [A]/[B]$
  - $K_{eq} = \exp(-\Delta G_0/RT)$
  - $[A]/[B] = \exp(-\Delta G_0/RT)$

# Predicting Binding Affinity

- Conceptually: *if we mix protein and ligand, what fraction of time will ligand bind protein?*

# Predicting Binding Affinity

- Conceptually: *if we mix protein and ligand, what fraction of time will ligand bind protein?*
- Ergodic hypothesis:

“over long periods of time, the time spent by a system in some region of the phase space of microstates with the same energy is proportional to the volume of this region”

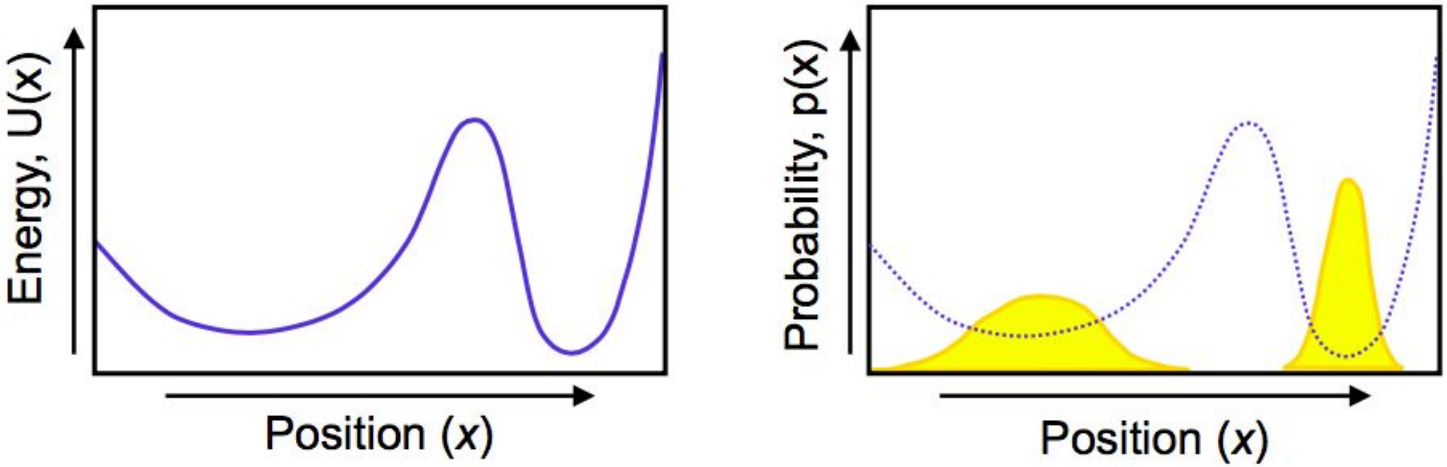
*-Wikipedia*

# Boltzmann Distribution

- Higher energy ~ lower probability

$$p(\boldsymbol{x}) \propto \exp\left(-U(\boldsymbol{x}) / k_B T\right)$$

Via Dror CS279 Ligand Docking Slides

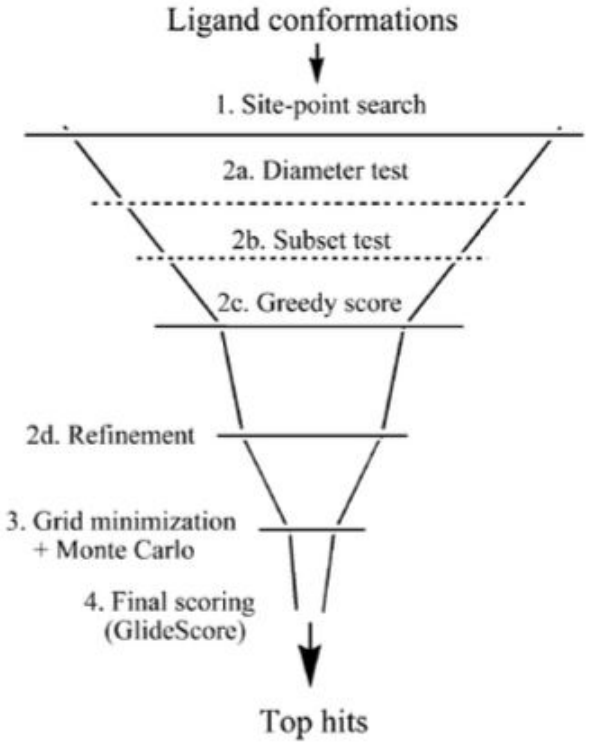


# First Approach

- Approximate free energy difference by directly observing system over time
  - $\Delta G = -RT \ln(P_1/P_0)$
  - Easy!

# First Approach

- Simulations too slow!
  - Must search across many ligand positions and orientations




Friesner et al., *J Med Chem* 47:1739, 2004

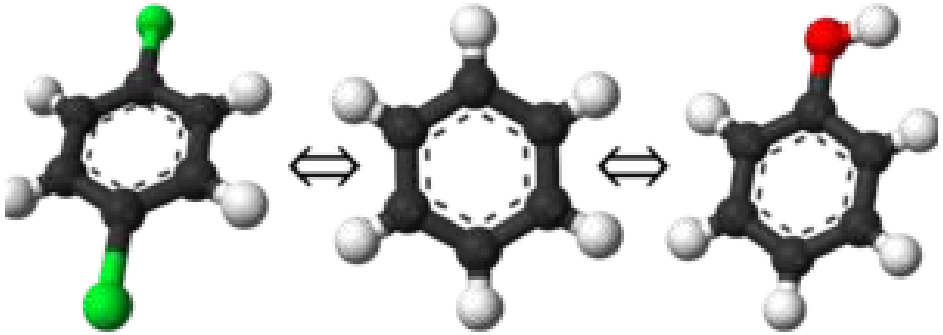
# Alternative Free Energy Calculations

- Non-alchemical approaches
  - Monte Carlo methods
  - Potential of mean force
- Alchemical approaches
  - Free energy perturbation
  - Bennett acceptance ratio
  - Thermodynamic integration
  - Free energy perturbation
  - Umbrella sampling

# Alchemical free energy methods

# Alchemical Free Energy Calculations

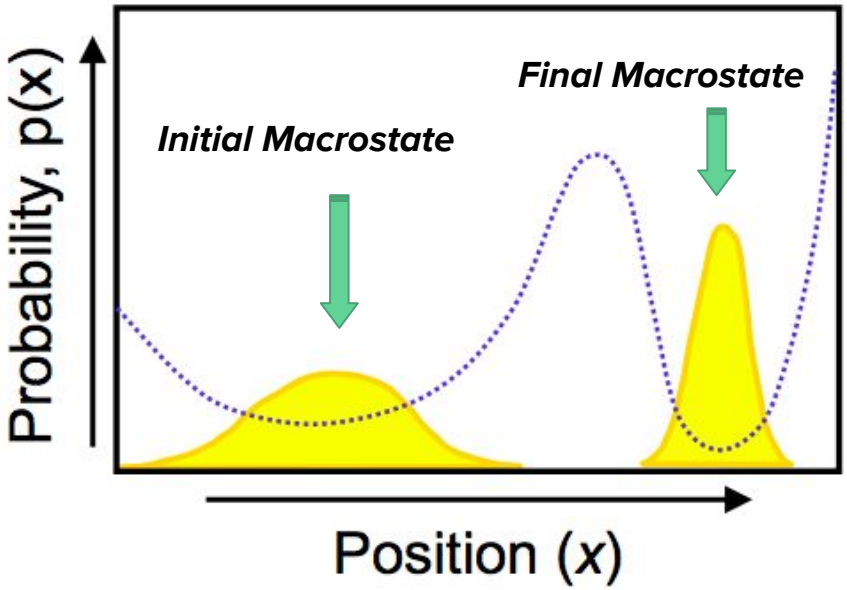
- Possibly non-physical intermediate states
- New enthusiasm due to methodological advances



***Initial State*** ***Chosen Intermediate*** ***Final State***

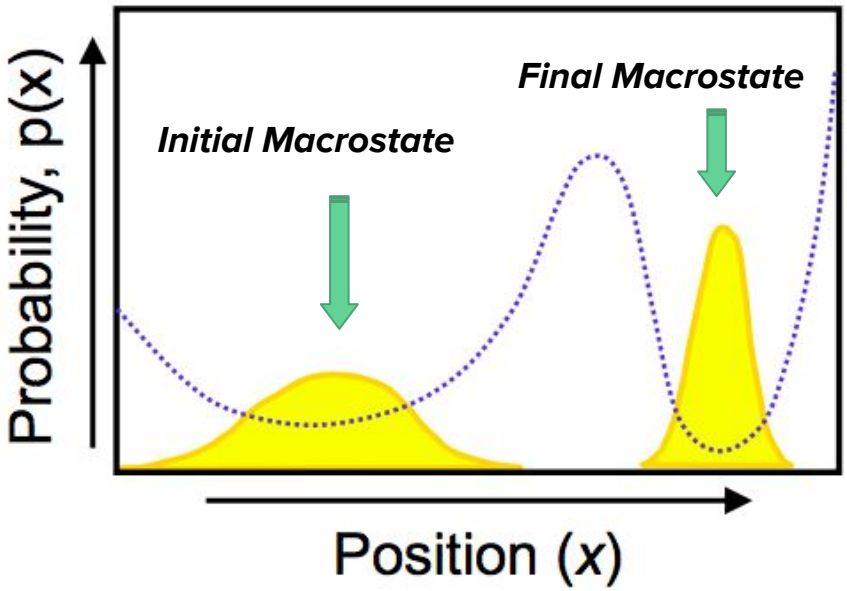
# Alchemical Free Energy Calculations

- Why alchemical intermediates?



# Alchemical Free Energy Calculations

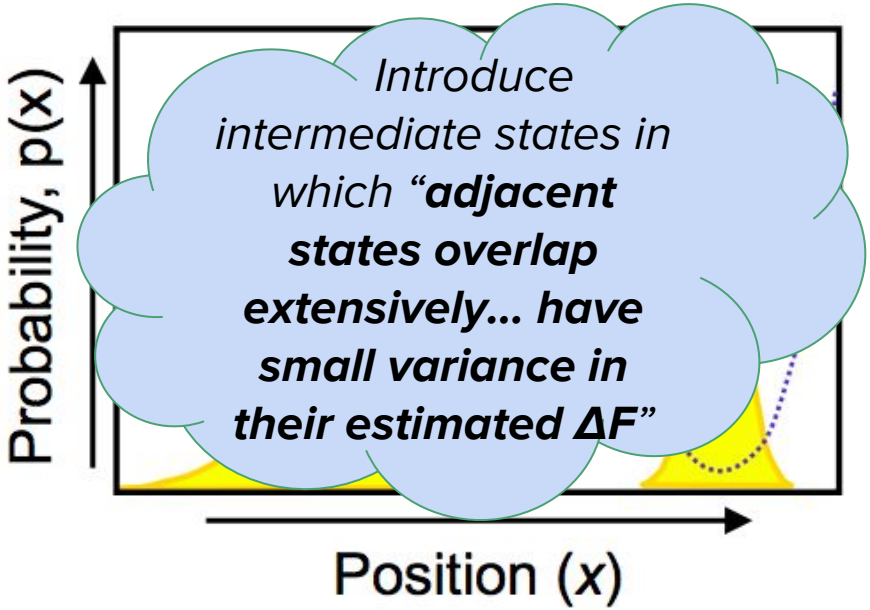
- Why alchemical intermediates?

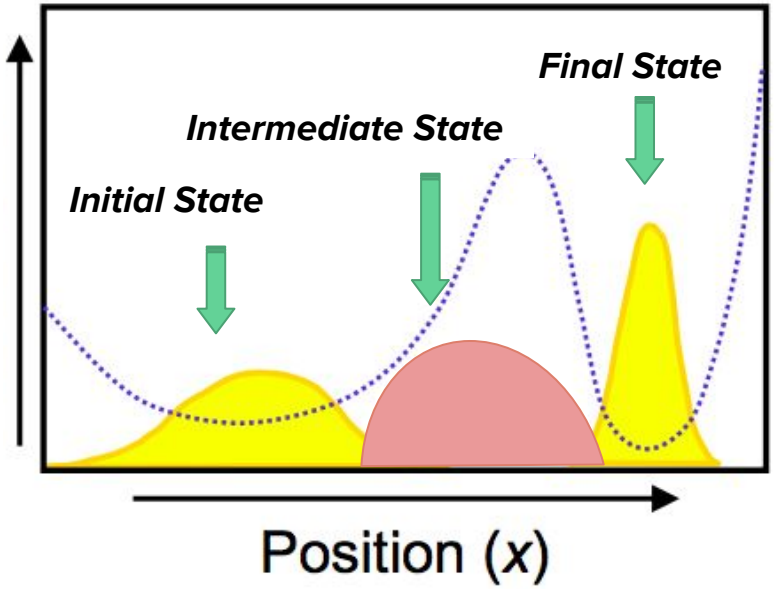



*Modern free energy difference estimation methods* ***"converge slowly unless two states overlap significantly in phase space"***

# Alchemical Free Energy Calculations

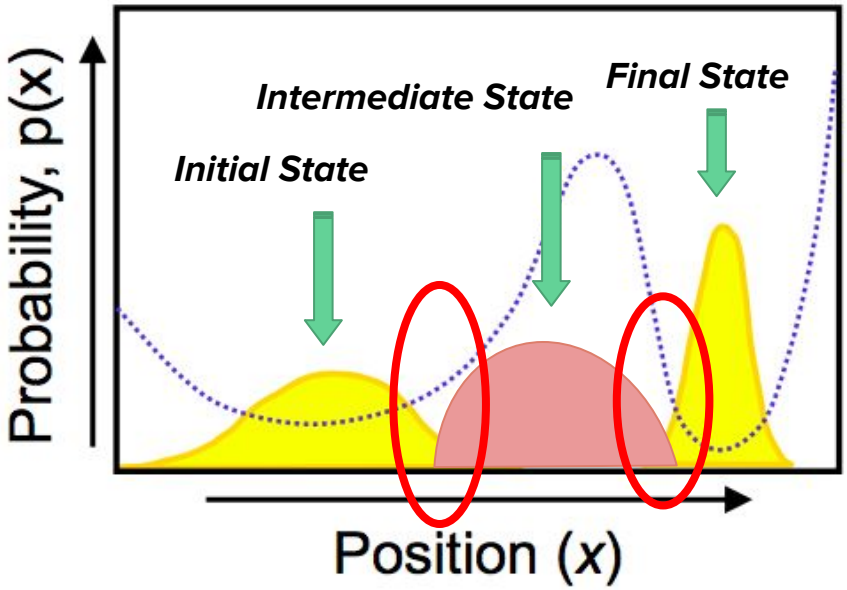
- Why alchemical intermediates?



# Alchemical Free Energy Calculations

- Take advantage of overlapping phase space (in red) to compute relative free energy difference



# Alternative Free Energy Calculations

- Non-alchemical approaches
  - Monte Carlo methods
  - Potential of mean force
- Alchemical approaches
  - Free energy perturbation
  - Bennett acceptance ratio
  - Thermodynamic integration
  - Free energy perturbation
  - Umbrella sampling

# Statistical Mechanics (review)

- Describe system according to Hamiltonian
  - $H(p_1, \dots, p_n, r_1, \dots, r_n)$
- Canonical partition function
  - $Z = \int \exp(-\beta H(\Gamma)) \, d\Gamma$
- Energy difference as ratio of partitions (Zwanzig, 1954)
  - $\Delta F = -\beta^{-1} \ln \frac{Z_1}{Z_0}$

# Free Energy Perturbation

- $\Delta F = -\beta^{-1} \ln \langle e^{-\beta \Delta E_{0\to 1}} \rangle_0$
- Two states characterized by $H_0$, $H_1$ in which

  $\Delta E_{0\to 1}$ equals the energy difference $H_0(\Gamma)$ - $H_1(\Gamma)$ at a point in phase space $\Gamma$

- $\Delta F = -\beta^{-1} \ln \frac{Z_N}{Z_0} = -\beta^{-1} \ln \frac{Z_1}{Z_0} \cdots \frac{Z_2}{Z_1} \cdots \frac{Z_N}{Z_{N-1}} = \sum_{n=1}^{N-1} \Delta F_{n\to n+1}$

# Alternative Methods

- Bennett acceptance ratio (Bennett, 1976)
  - $$\Delta F = -\beta^{-1} \ln \frac{\langle f(\Delta E_{0\to 1}\rangle_0}{\langle f(\Delta E_{1\to 0})\, \exp[-\beta \Delta E_{0\to 1}]\rangle_1}$$
- Weighted histogram analysis method
  - $$\Delta F_k = -\beta^{-1} \ln \sum_{k=1}^{K} \left\langle \left[ \sum_{k'=1}^{K} (n_{k'}/n_k) \exp[\beta(\Delta F_{k'} - (\Delta E_{k\to k'}))] \right]^{-1} \right\rangle_k$$
- Thermodynamic Integration
  - $$\Delta F = \int_0^1 d\lambda \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda$$

# Challenges and future outlook

# Alchemical Free Energy Calculations

- **Modeling** and simulation set-up
- **Sampling** of relevant configurations with appropriate probability
- **Analysis** of results to obtain estimates

# Modeling

- Require full atomistic model of system
- Selection of alchemical intermediates
  - Minimizing variance between adjacent states
- Selection of force field
  - Polarizable force fields
  - Explicit vs implicit solvent representation

# Sampling

- Ideally sample from equilibrium distribution such that all relevant states are sampled to reach convergence
- Markov State Models
  - Numerous short simulations to identify metastable states
  - Restrict conformations

# Analysis

- Lack of literature
  - E.g. which parameters have significant impact for specific systems
- Lack of error assessment
  - E.g. error incurred from omissions in free energy approximations

# Moving Forward

- Automate preparation of systems
- Facilitate high-throughput use & evaluation
- Establish standardized benchmarks
- Organize periodic prediction challenges

# Questions?

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

# Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field

Lingle Wang,[†] Yujie Wu,[†] Yuqing Deng,[†] Byungchan Kim,[†] Levi Pierce,[†] Goran Krilov,[†] Dmitry Lupyan,[†] Shaughnessy Robinson,[†] Markus K. Dahlgren,[†] Jeremy Greenwood,[†] Donna L. Romero,[‡] Craig Masse,[‡] Jennifer L. Knight,[†] Thomas Steinbrecher,[†] Thijs Beuming,[†] Wolfgang Damm,[†] Ed Harder,[†] Woody Sherman,[†] Mark Brewer,[†] Ron Wester,[‡] Mark Murcko,[†] Leah Frye,[†] Ramy Farid,[†] Teng Lin,[†] David L. Mobley,[⊥] William L. Jorgensen,[∥] Bruce J. Berne,[§] Richard A. Friesner,[§] and Robert Abel*,[†]

[†]Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States

[‡]Nimbus Discovery, 25 First Street, Suite 404, Cambridge, Massachusetts 02141, United States

[§]Department of Chemistry, Columbia University, 3000 Broadway, New York, New York 10027, United States

[∥]Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States

[⊥]Departments of Pharmaceutical Sciences and Chemistry, University of California—Irvine, Irvine, California 92697, United States

Karen Lee

CS371 Presentation

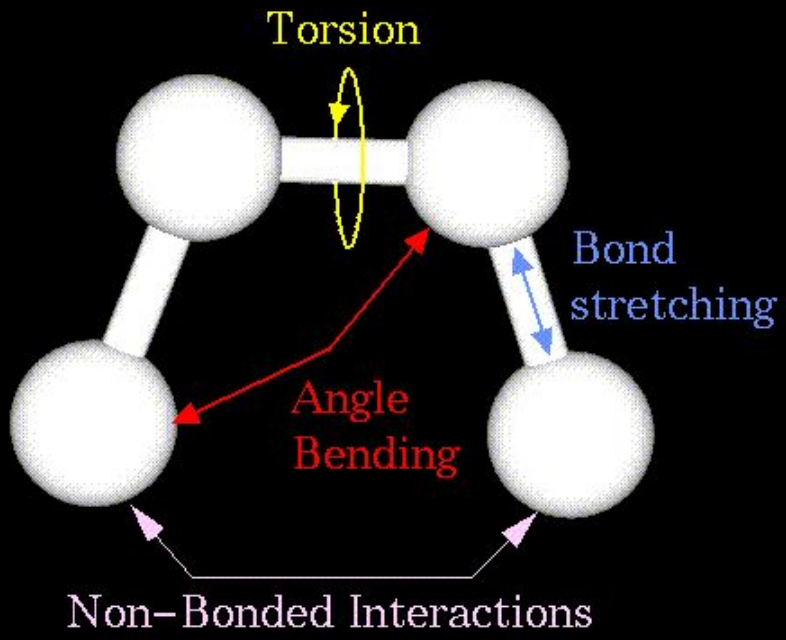
# RESEARCH BACKGROUND: PROTEIN-LIGAND INTERACTION

- Binding affinity is important for maximizing therapeutic effect
- Computational chemistry and computer-aided drug discovery (CADD)
- The approach: free-energy simulation
  - Free energy perturbation (FEP) technology
- Current challenges: lack of large-scale validation and technical challenges

# Research Goal

Develop and apply an FEP protocol that enables highly accurate binding affinity predictions across over 200 ligands and 10 targets

# WHAT IS A FORCE FIELD?




http://c125.chem.ucla.edu/NIH/MolMecha

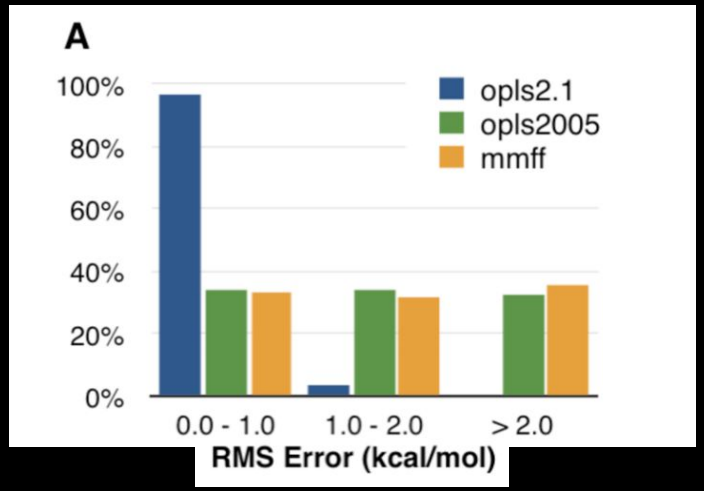
# WHAT IS A FORCE FIELD?

- Set of equations and empirical parameters to describe the potential energy of a protein as a function of its atomic coordinates
- Potential energy function divided into 2 classes:
  - Bonded interactions
  - Nonbonded interactions

$$E_{MM} = E_{bond} + E_{ang} + E_{tors} + E_{vdw} + E_{ele} + E_{cross}$$

# IMPROVED FORCE FIELD OPLS2.1

- Incorporates a robust model for non-bonded interactions
- 1200 new torsion profiles and 10,000 new torsional parameters added
- 7000 new bend types



# FEP/REST ALGORITHM TO IMPROVE MOLECULAR DYNAMIC PREDICTIONS

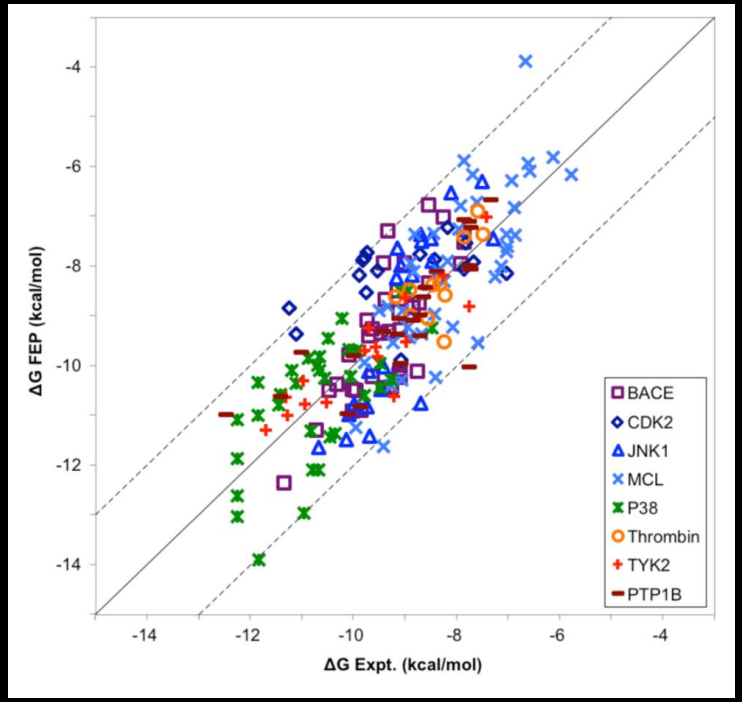
- A newly developed FEP/REST algorithm enables simulations of selected subset with higher effective temperature regime
- Selection of REST region
  - Protein residues close to binding pocket
  - Uniform set of key protein residues from crystallography
  - Ligand involved directly in perturbation
  - 25 atom cutoff

# RESULTS FROM FEP/REST COMPUTATIONS

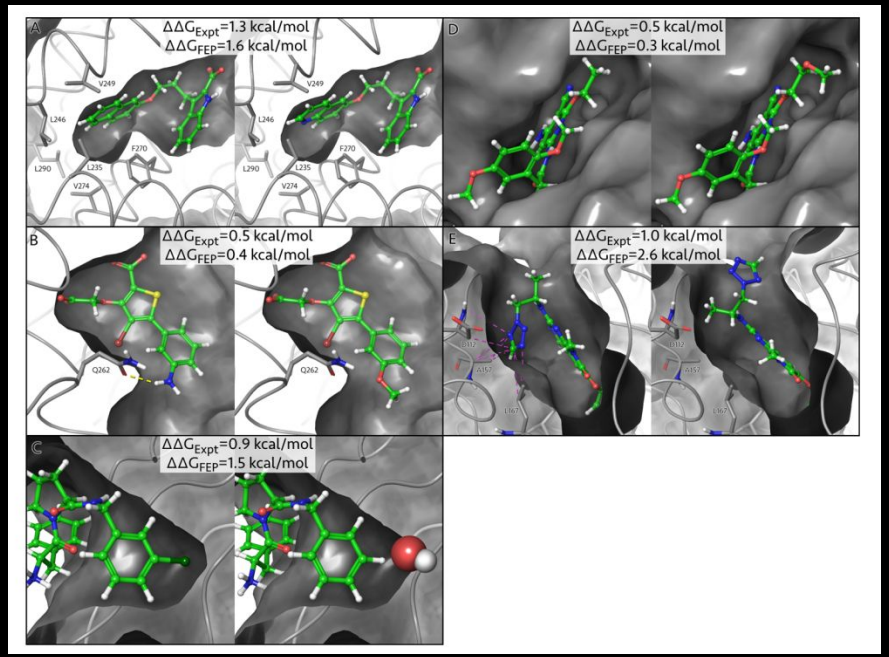
**Table 2. Relative Binding Free-Energy Calculation Results**[a]

| | system | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | BACE | CDK2 | JNK1 | MCL1 | p38 | PTP1B | thrombin | Tyk2 |
| no. of compds | 36 | 16 | 21 | 42 | 34 | 23 | 11 | 16 |
| binding affinity range (kcal/mol) | 3.5 | 4.2 | 3.4 | 4.2 | 3.8 | 5.1 | 1.7 | 4.3 |
| crystal structure | 4DJW | 1H1Q | 2GMX | 4HW3 | 3FLY | 2QBS | 2ZFF | 4GIH |
| series ref | 46 | 47 | 48 | 49 | 50 | 51 | 45 | 52,53 |
| no. of perturbations | 58 | 25 | 31 | 71 | 56 | 49 | 16 | 24 |
| MUE FEP | 0.84 ± 0.08 | 0.91 ± 0.12 | 0.78 ± 0.12 | 1.16 ± 0.10 | 0.80 ± 0.08 | 0.89 ± 0.12 | 0.76 ± 0.13 | 0.75 ± 0.11 |
| RMSE FEP | 1.03 ± 0.08 | 1.11 ± 0.12 | 1.00 ± 0.15 | 1.41 ± 0.12 | 1.03 ± 0.09 | 1.22 ± 0.17 | 0.93 ± 0.15 | 0.93 ± 0.12 |
| avg $\sigma_{cc}$ | 0.65 | 0.57 | 0.30 | 0.91 | 0.76 | 0.94 | 0.93 | 0.46 |
| obs *R*-value FEP | 0.78 ± 0.07 | 0.48 ± 0.19 | 0.85 ± 0.07 | 0.77 ± 0.05 | 0.65 ± 0.09 | 0.80 ± 0.08 | 0.71 ± 0.24 | 0.89 ± 0.07 |
| *P*-value FEP | $3.9 \times 10^{-5}$ | $1.2 \times 10^{-2}$ | $7.0 \times 10^{-8}$ | $2.2 \times 10^{-7}$ | $1.6 \times 10^{-7}$ | $7.8 \times 10^{-6}$ | $1.1 \times 10^{-2}$ | $2.3 \times 10^{-7}$ |
| obs *R*-value, MW | 0.14 | −0.48 | −0.39 | −0.55 | −0.46 | −0.84 | −0.48 | 0.00 |
| obs *R*-value, MM-GB/SA | −0.40 | −0.53 | 0.65 | 0.42 | 0.66 | 0.67 | 0.93 | 0.79 |
| obs *R*-value, Glide SP | 0.00 | −0.56 | 0.24 | 0.59 | 0.14 | 0.55 | 0.53 | 0.79 |
| anticip FEP *R*-value | 0.64 ± 0.09 | 0.73 ± 0.11 | 0.64 ± 0.12 | 0.71 ± 0.07 | 0.67 ± 0.08 | 0.79 ± 0.07 | 0.37 ± 0.26 | 0.74 ± 0.10 |
| anticip exptl *R*-value | 0.88 ± 0.03 | 0.92 ± 0.03 | 0.88 ± 0.04 | 0.91 ± 0.02 | 0.89 ± 0.03 | 0.94 ± 0.02 | 0.68 ± 0.15 | 0.92 ± 0.03 |

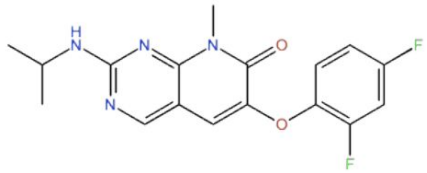
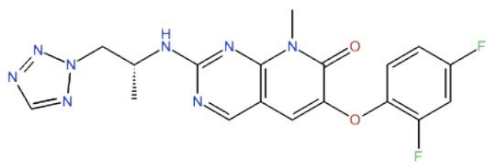
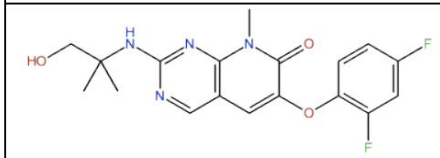
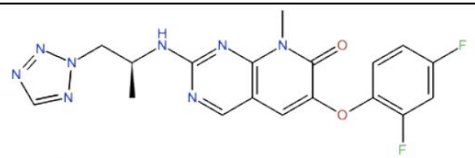
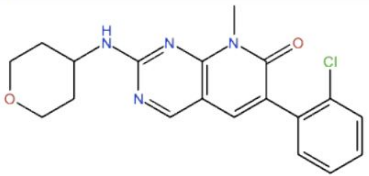
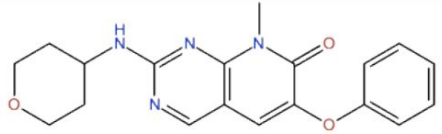
# VALIDATION OF FEP/REST ALGORITHM



# DIFFERENT INTERACTION TYPES CAPTURED BY FEP

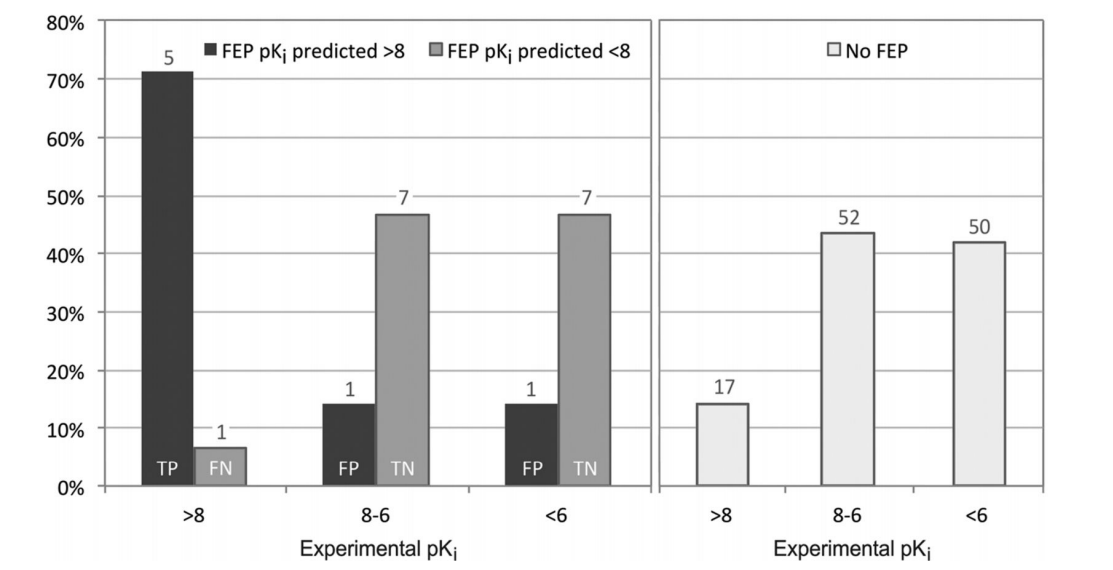


# LIGAND PERTURBATIONS

| | |
|---|---|
|  3fly |  3fmh |
|  2z |  3fmk |
|  2c |  3flz |

# APPLICATION TO DRUG-DESIGN PROJECTS

- Project 1: Developing selective inhibitors of IRAK4
- Project 2: Developing inhibitors for TYK2




Maintain affinity at $pK_i$>8

# SUMMARY: MAIN FINDINGS

1. Improved force field that incorporates additional torsional parameters and nonbonded interactions
2. A newly developed FEP/REST algorithm
3. Shows the viability of applying FEP to drug-discovery projects

# QUESTIONS?

nature chemistry

ARTICLES

PUBLISHED ONLINE: 25 MAY 2014 | DOI: 10.1038/NCHEM.1954

# Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery

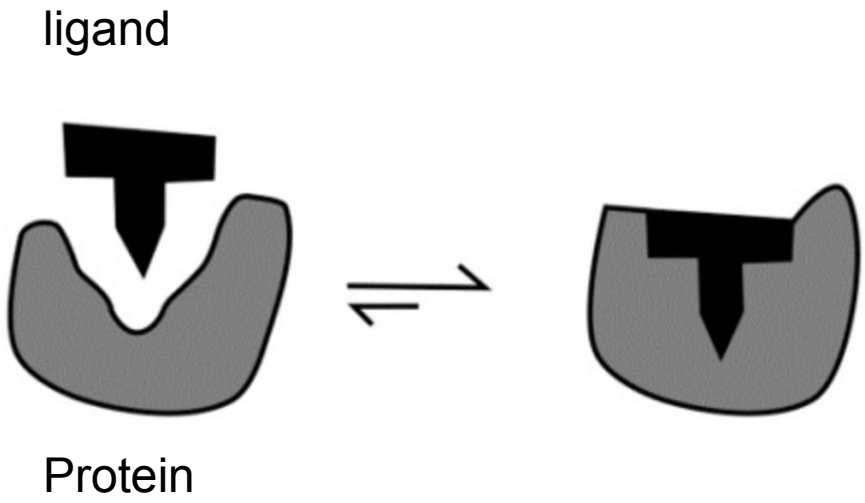
Marcus Fischer[1,2†], Ryan G. Coleman[1†], James S. Fraser[3*] and Brian K. Shoichet[1,2*]

Wisam Reid

BIOPHYS 371

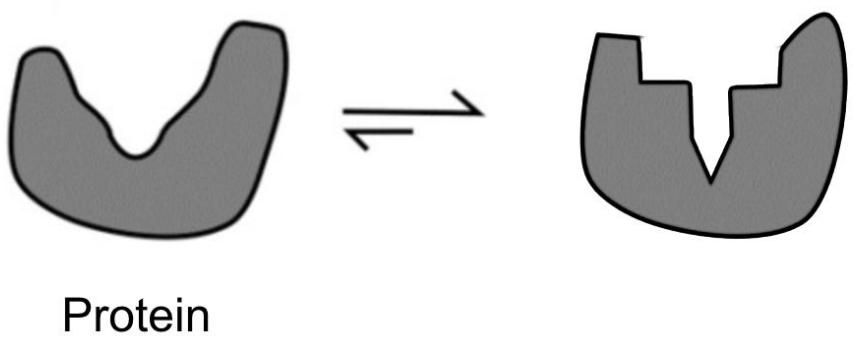
Paper Presentation

# Apo proteins (with no ligand bound) may transiently transform into the conformations adopted in ligand complexes




Du, Xing et al. “Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods.” Ed. Tatyana Karabencheva-Christova. International Journal of Molecular Sciences 17.2 (2016): 144. PMC. Web. 30 Jan. 2017.

# Apo proteins (with no ligand bound) may transiently transform into the conformations adopted in ligand complexes

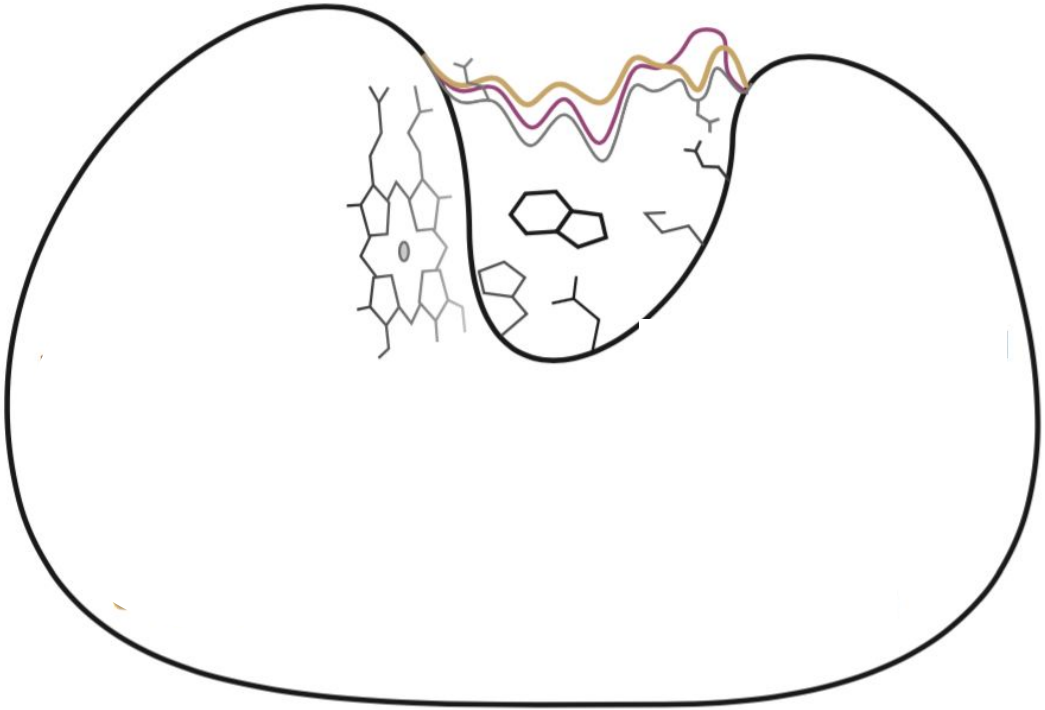



Du, Xing et al. “Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods.” Ed. Tatyana Karabencheva-Christova. International Journal of Molecular Sciences 17.2 (2016): 144. PMC. Web. 30 Jan. 2017.

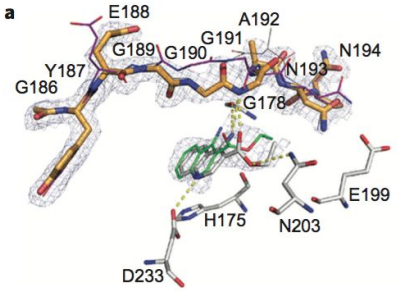
# Can these conformations may be used to address two long-standing problems?

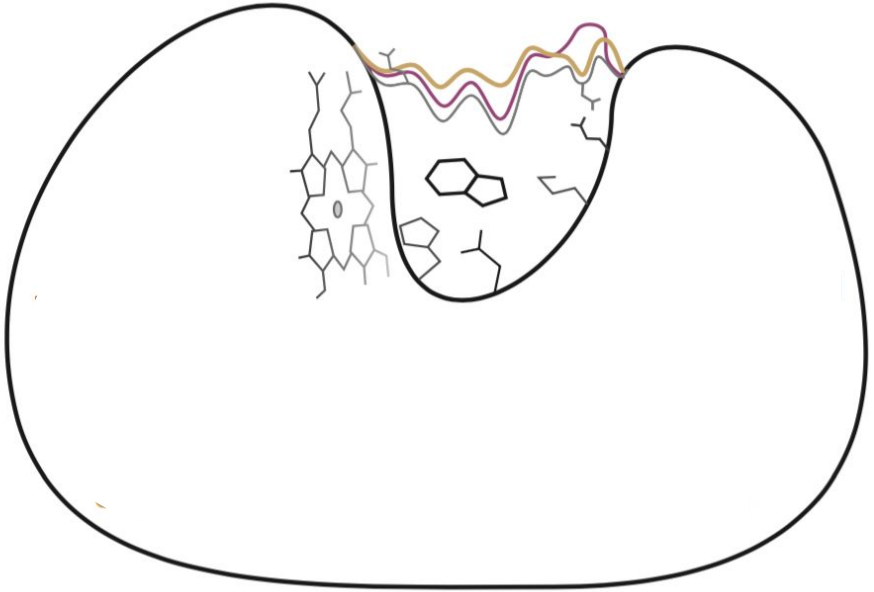
- Ligand discovery
- Sampling protein states to determine which states they are most likely to be in

# Soft docking: Assessing three loop conformations of cytochrome c peroxidase (CcP)

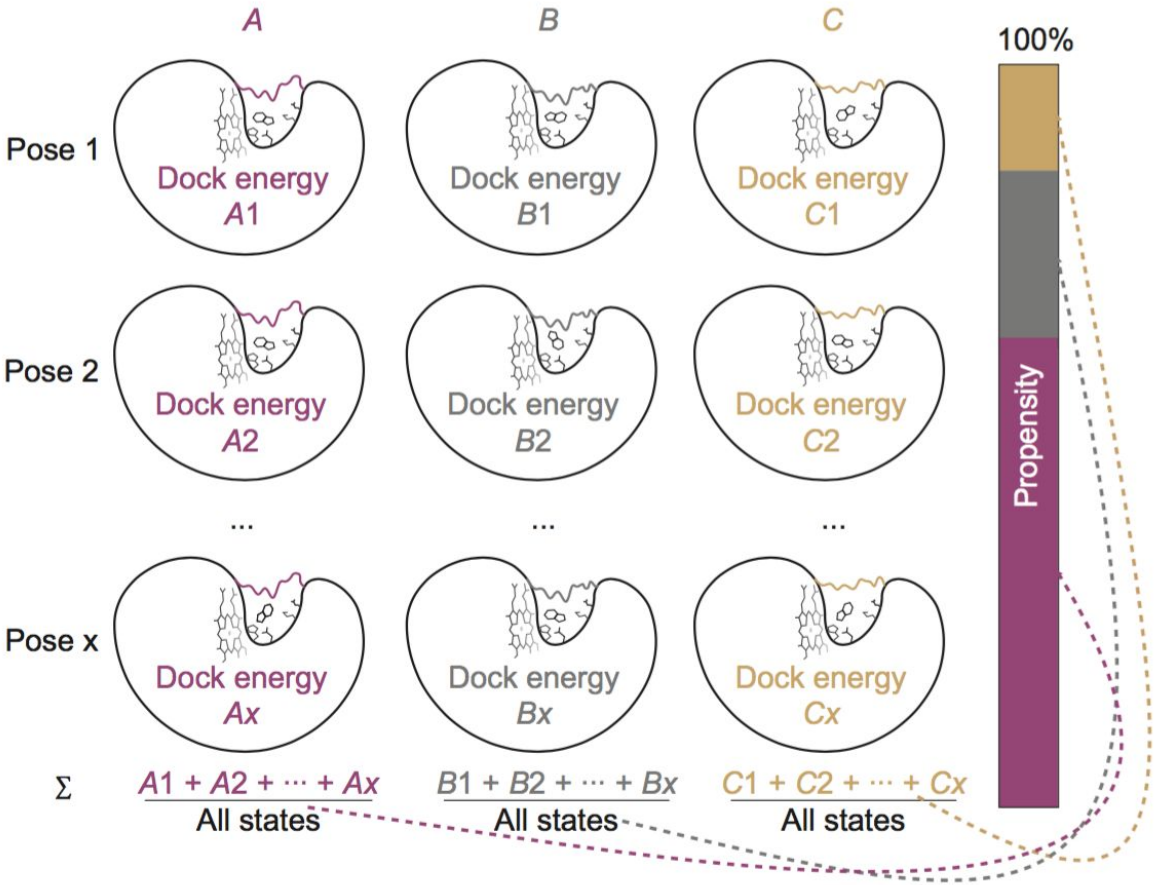


# Prospective docking predictions

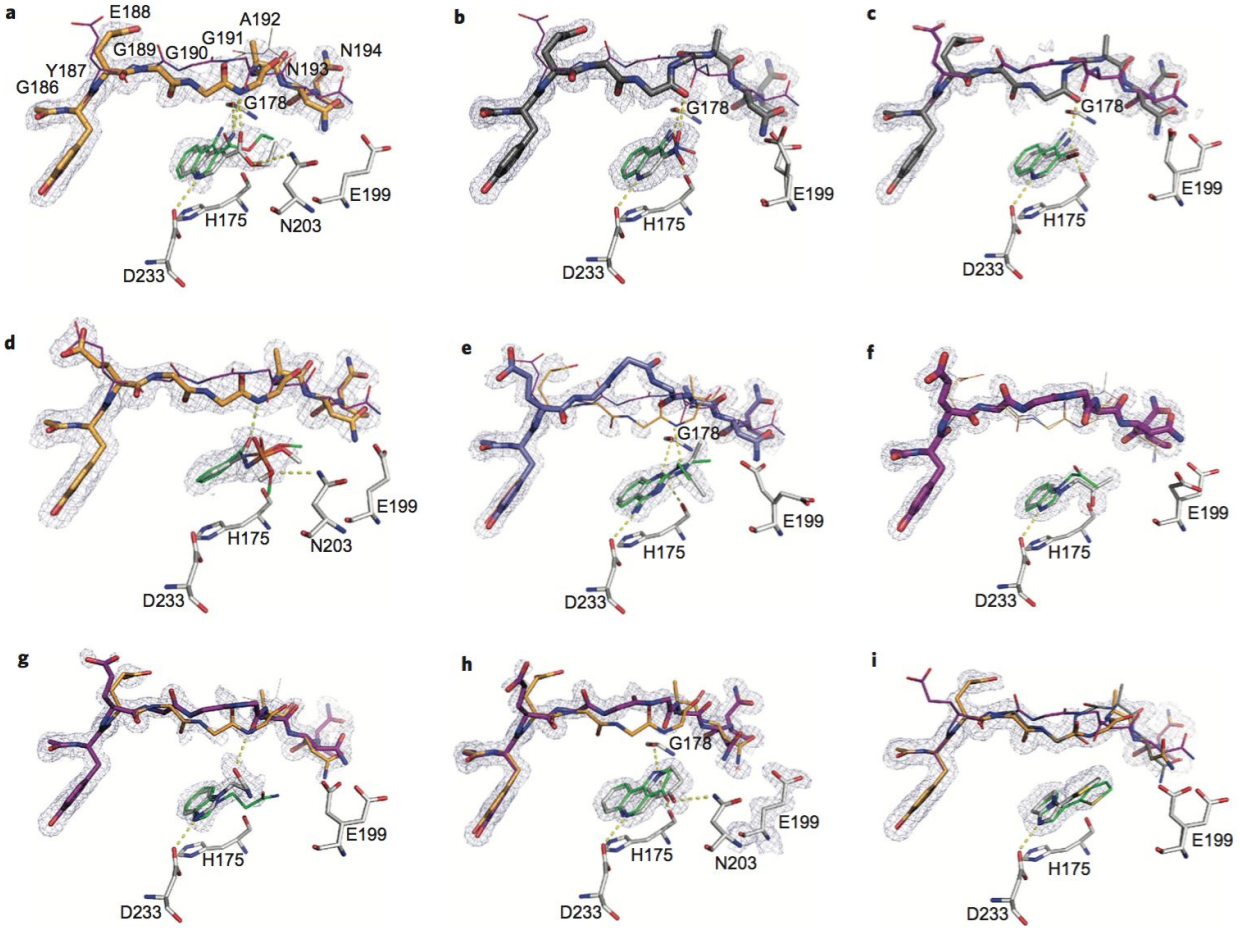


# Prospective docking predictions



# Prospective docking predictions



# Two strategies for modeling protein flexibility in docking screens for new ligands.

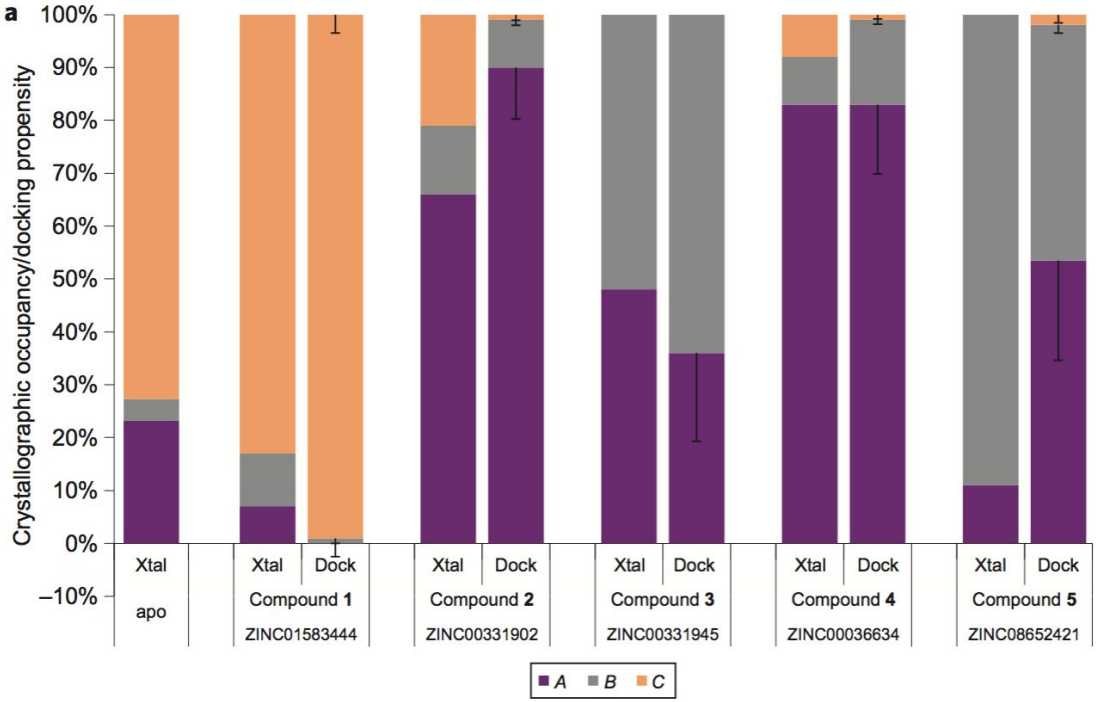
- ‘Soft Docking’

- Explicitly represent, and dock into, multiple receptor conformations

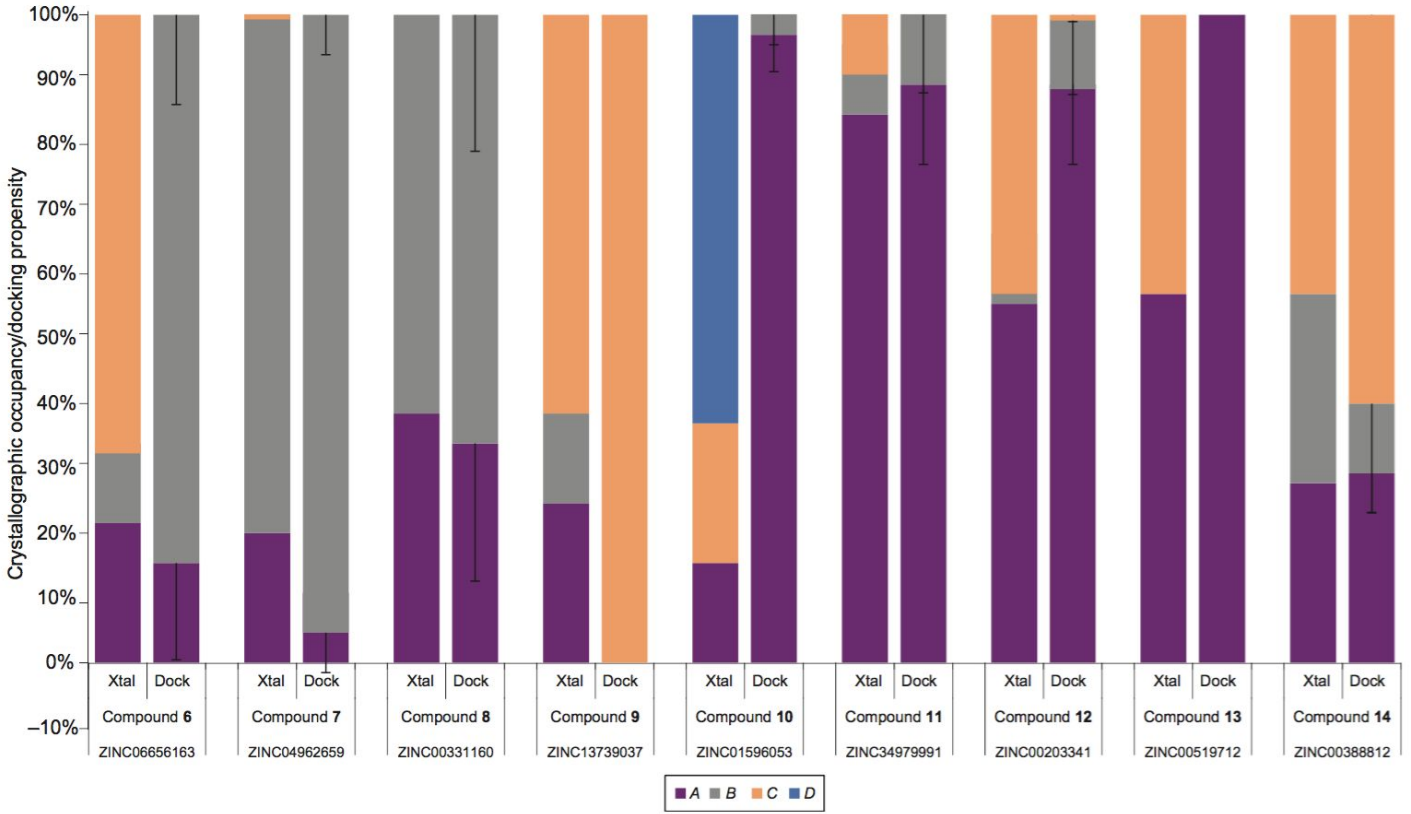
# Hypothesis

Given recent advances in crystallographic refinement, we can incorporate protein flexibility and conformational energy penalties in docking
screens to improve ligand discovery.

# Results


a
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
–10%
Crystallographic occupancy/docking propensity
Xtal
apo
Xtal
Dock
Compound 1
ZINC01583444
Xtal
Dock
Compound 2
ZINC00331902
Xtal
Dock
Compound 3
ZINC00331945
Xtal
Dock
Compound 4
ZINC00036634
Xtal
Dock
Compound 5
ZINC08652421
A
B
C

# Results


Crystallographic occupancy/docking propensity
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
–10%
Xtal
Dock
Compound 6
ZINC06656163
Compound 7
ZINC04962659
Compound 8
ZINC00331160
Compound 9
ZINC13739037
Compound 10
ZINC01596053
Compound 11
ZINC34979991
Compound 12
ZINC00203341
Compound 13
ZINC00519712
Compound 14
ZINC00388812
A
B
C
D

# Summary

- Partial occupancy modelling enabled the prospective prediction of ligands with new chemotypes and new physical properties
- The described methods yielded high correspondence between the loop propensities and ligand geometries
    - The observed loops and residue conformations matched well those predicted

# Caveats

- Only a narrow range of conformations above the ground state can be observed reliably in this method.
    - i.e. The D conformation observed in the complexes of ligand 10, was unanticipated because it was not observed in the apo structure.
- Docking scores, even when physics-based, leave out important terms and make substantial approximations

# Thank You