Improving virtual screening through physics-based methods

Nikhil Goel, Jimmy Yu
January 30, 2018

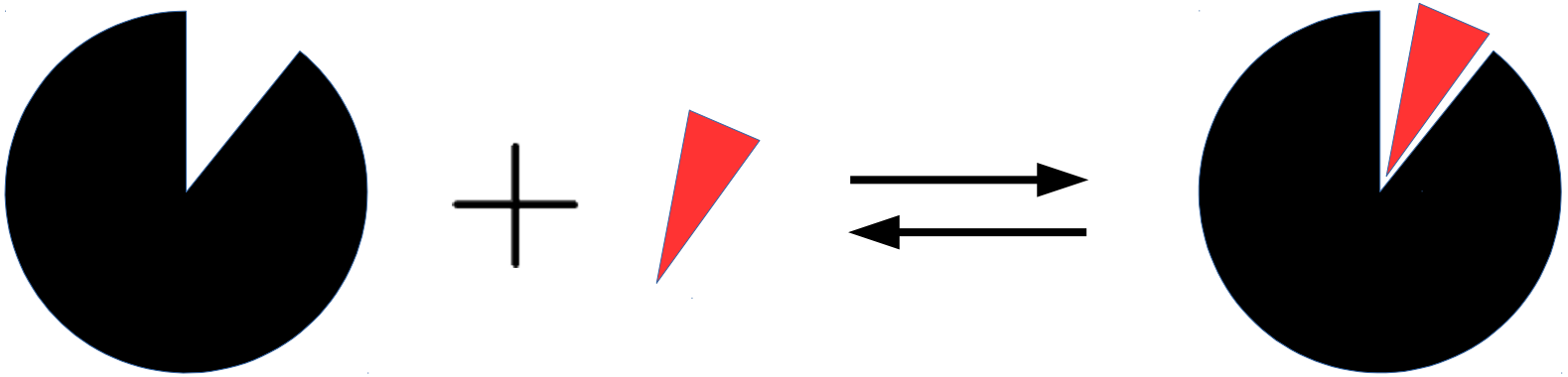
Motivation: drug discovery

- Problem: R&D is expensive – the space of possible small molecule drugs to synthesize and test is prohibitively large
- Accurate prediction of protein–ligand binding affinities desired from computational drug design
- Until recently, the value of computational screening in drug discovery has been limited

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

L. Wang, *et al. J. Am. Chem. Soc.* 2015.

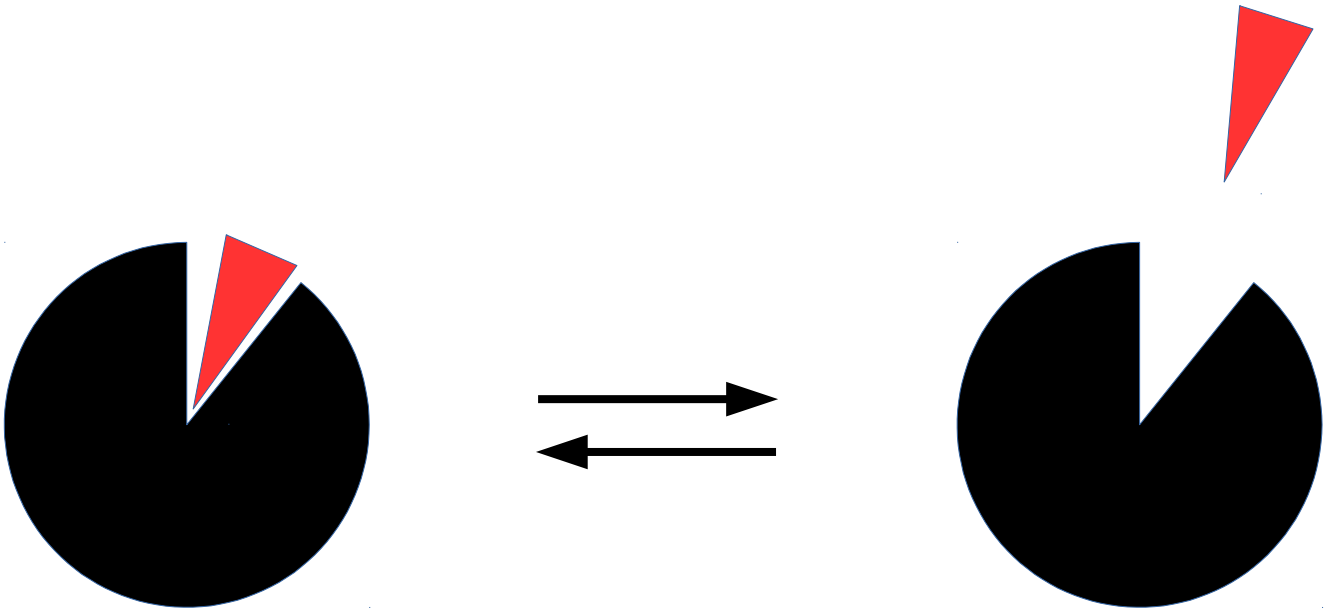
Definition: binding affinity



$$K_d \approx \frac{[\text{ligand}][\text{target}]}{[\text{complex}]}$$

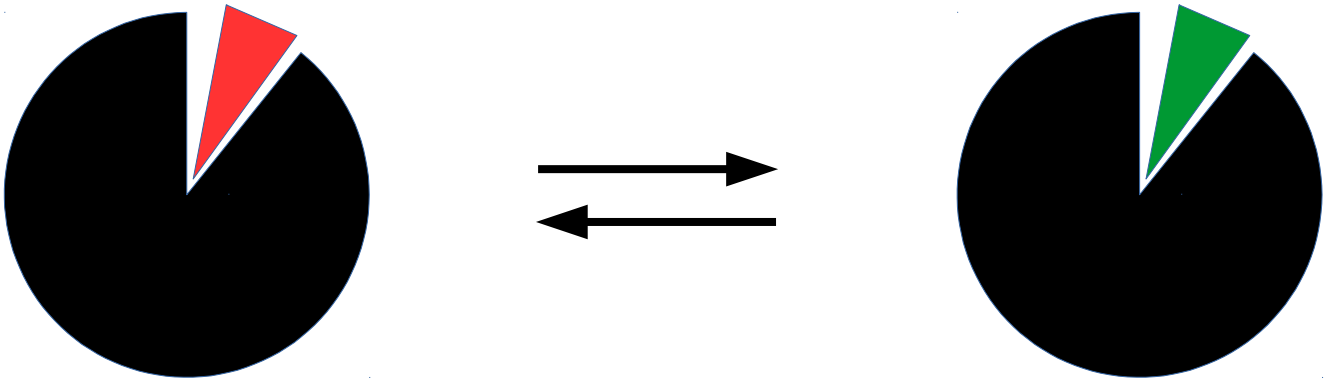
$$\Delta G^0_{\text{bind}} = RT \ln(K_d)$$

Quantifying ligand binding



- Bimolecular molecular dynamics simulations
- Sampling problem: need simulations long enough to quantify fraction of time bound and unbound

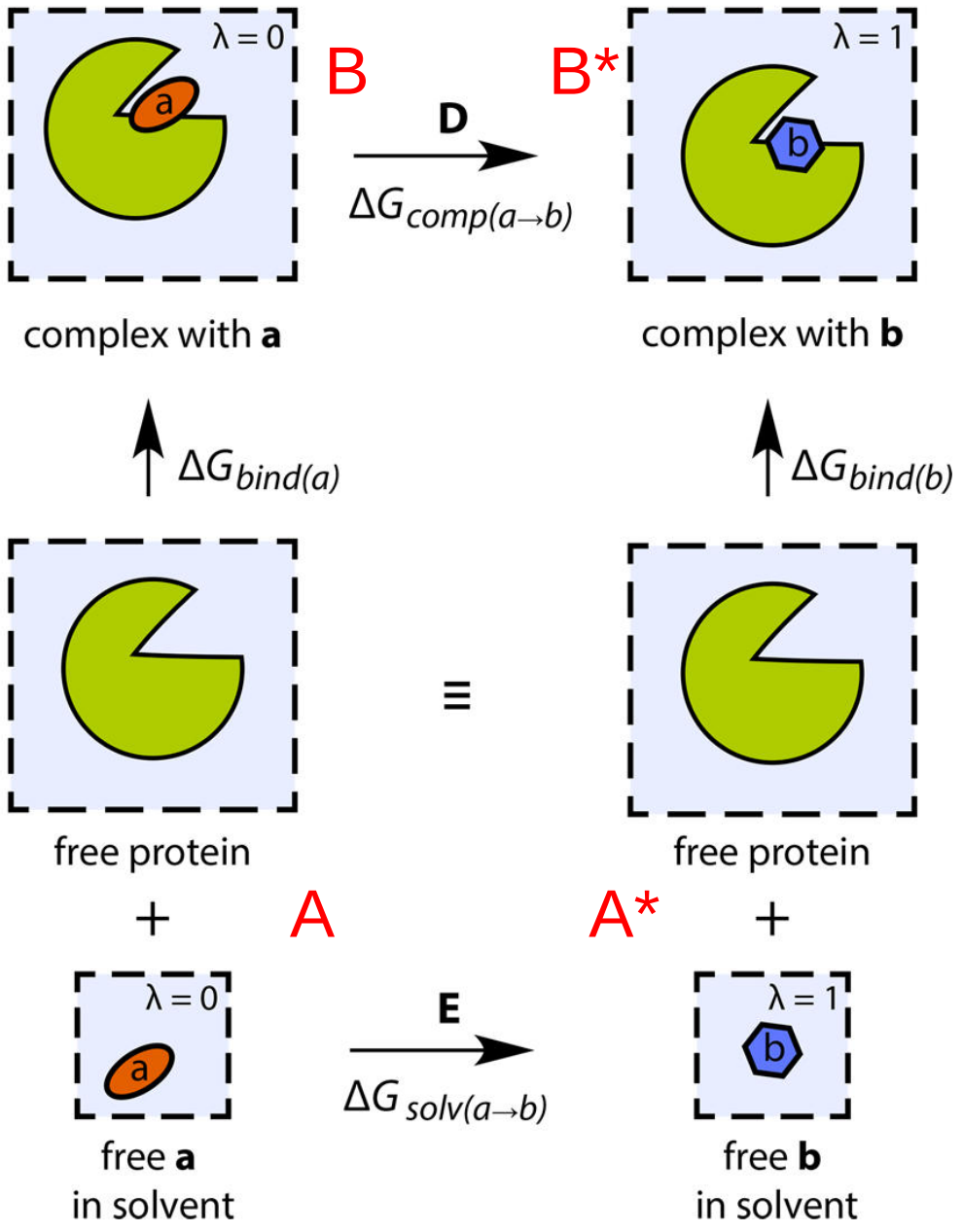
Quantifying ligand binding



- Free energy perturbation (FEP) methods are more efficient

Free energy perturbation

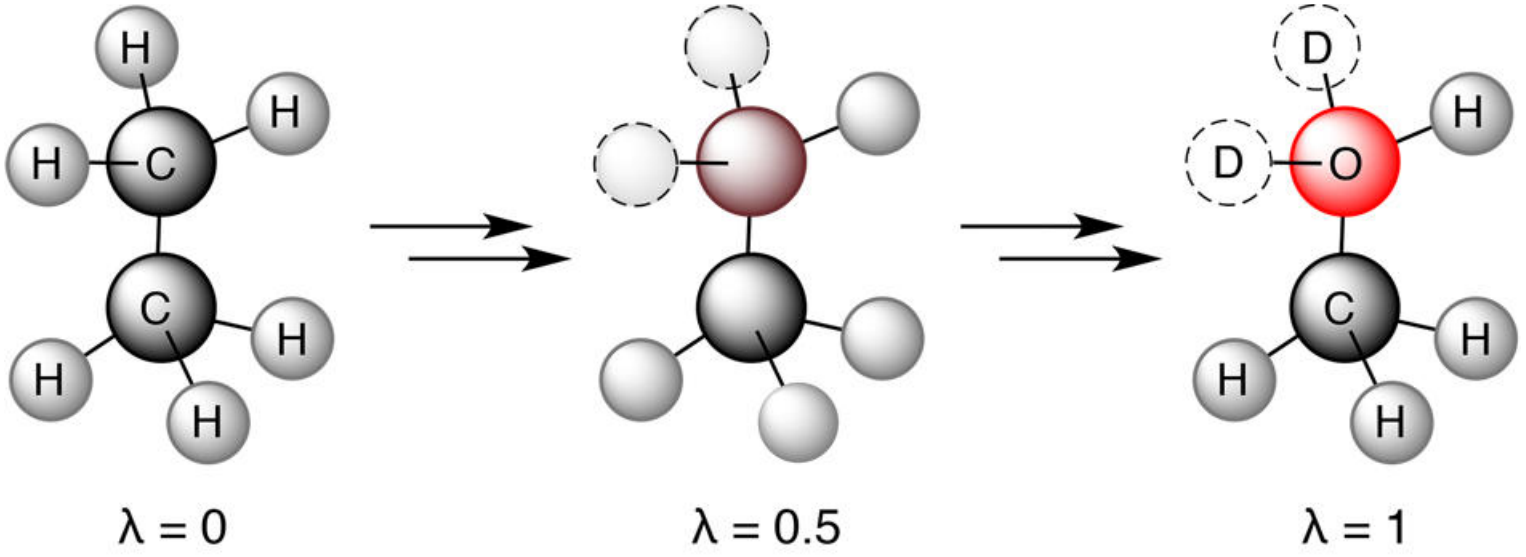
- Morph the ligand from state A to state A* (or B to B*) instead of simulating the binding event
- Simulation to allow systems A and B to sample different conformations



Williams-Noonan, B. J. *et al.* 2017.

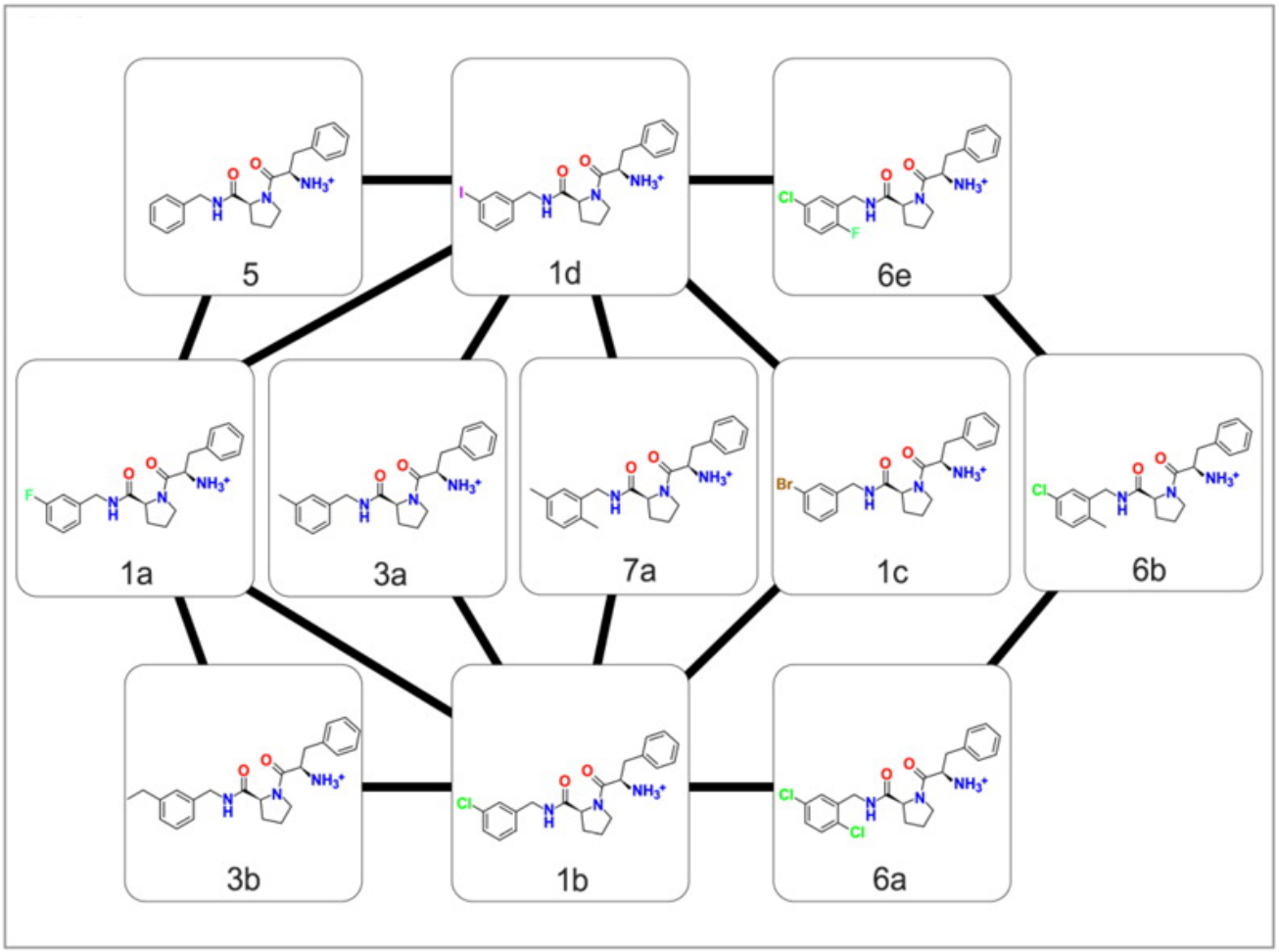
Free energy perturbation

- Transformation from A to A* occurs over a number of discrete steps



Williams-Noonan, B. J. *et al.* 2017.

Exploring the space of perturbations



Wang, L. *et al.* 2015.

Modern approach to FEP

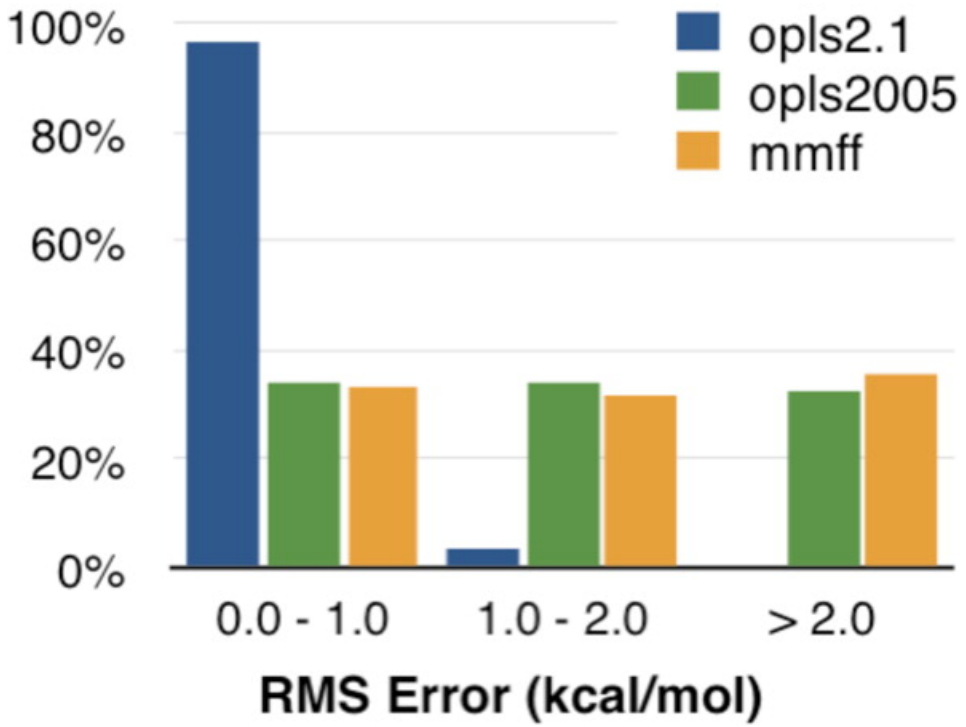
- FEP methods have been around for over 30 years, but previously limited by high computational cost
- High throughput screening possible thanks to the parallelization afforded by GPUs
- Accuracy suitable for guiding lead optimization due to force field and sampling algorithm improvements

OPLS2.1: Force field for drug-like molecules

- Optimized Potential for Liquid Simulations
- Force fields are usually not fit to data for drug-like molecules
- OPLS2.1 force field is more robust in the treatment of drug-like ligands.

OPLS2.1 in FEP

- If a missing torsion is identified, augment the force field with new fitted parameters
- 89/199 ligands had at least one missing torsion



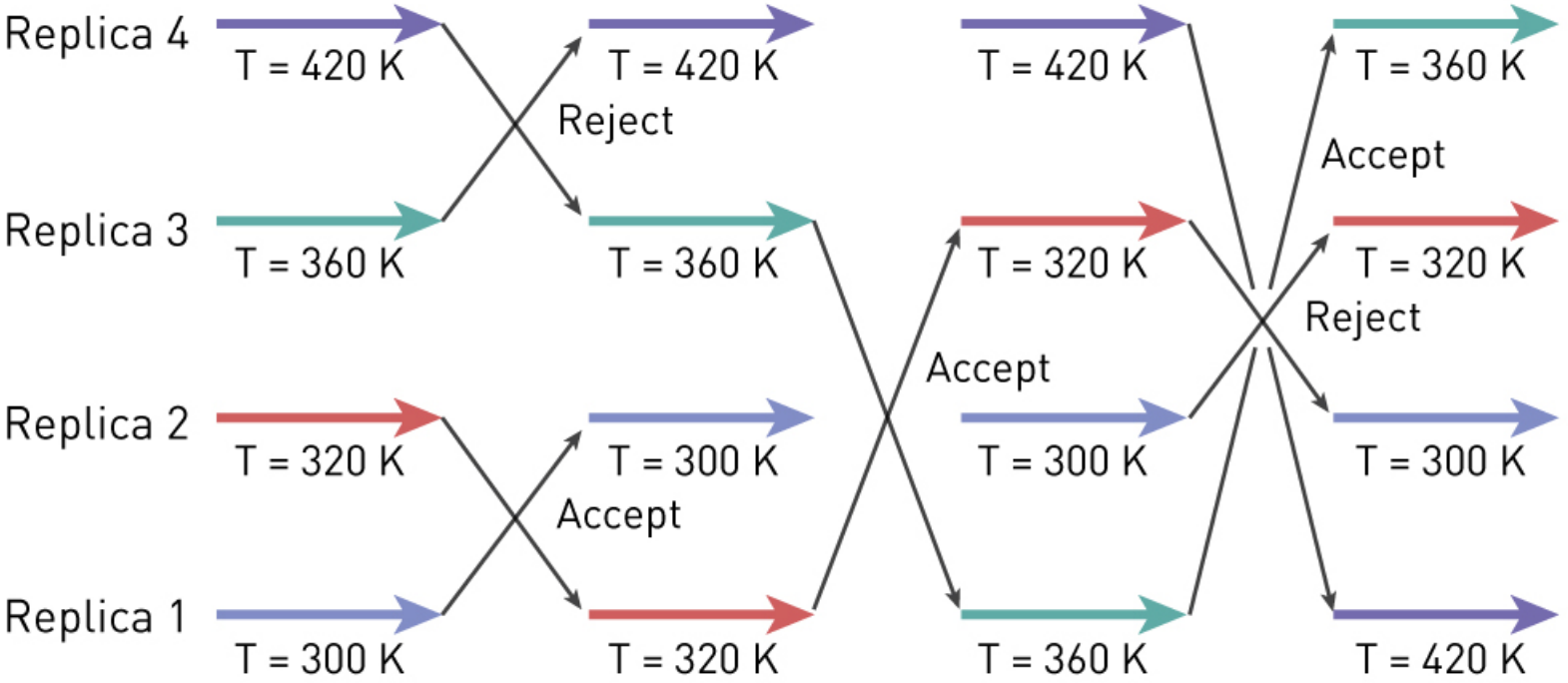
Wang, L. *et al.* 2015.

Sampling the conformation space

- Improving sampling in FEP by using replica exchange
- Required for accurate statistics and proper representation of the conformations sampled by the system

Replica exchange

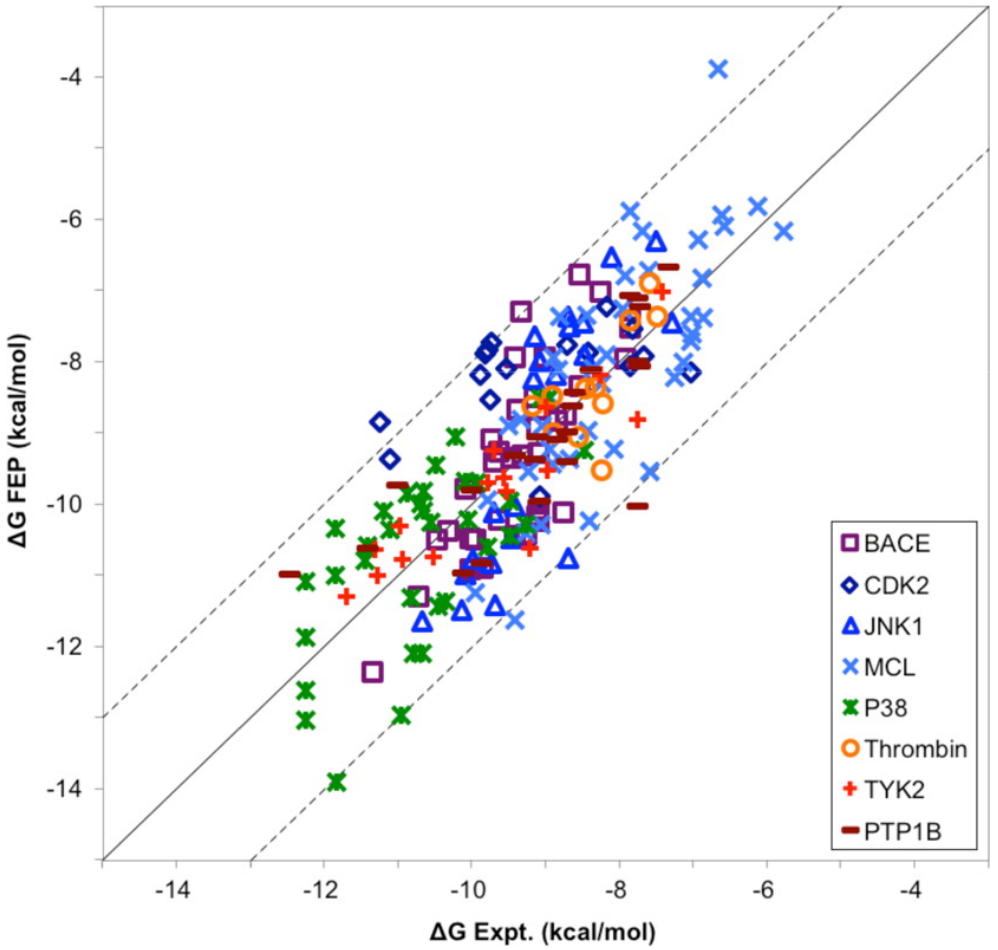
- Replica systems run in parallel at different temperatures can exchange coordinates to allow for enhanced sampling of the phase space



Sugita, Y., RIKEN Theoretical Molecular Science Laboratory

Results: method validation

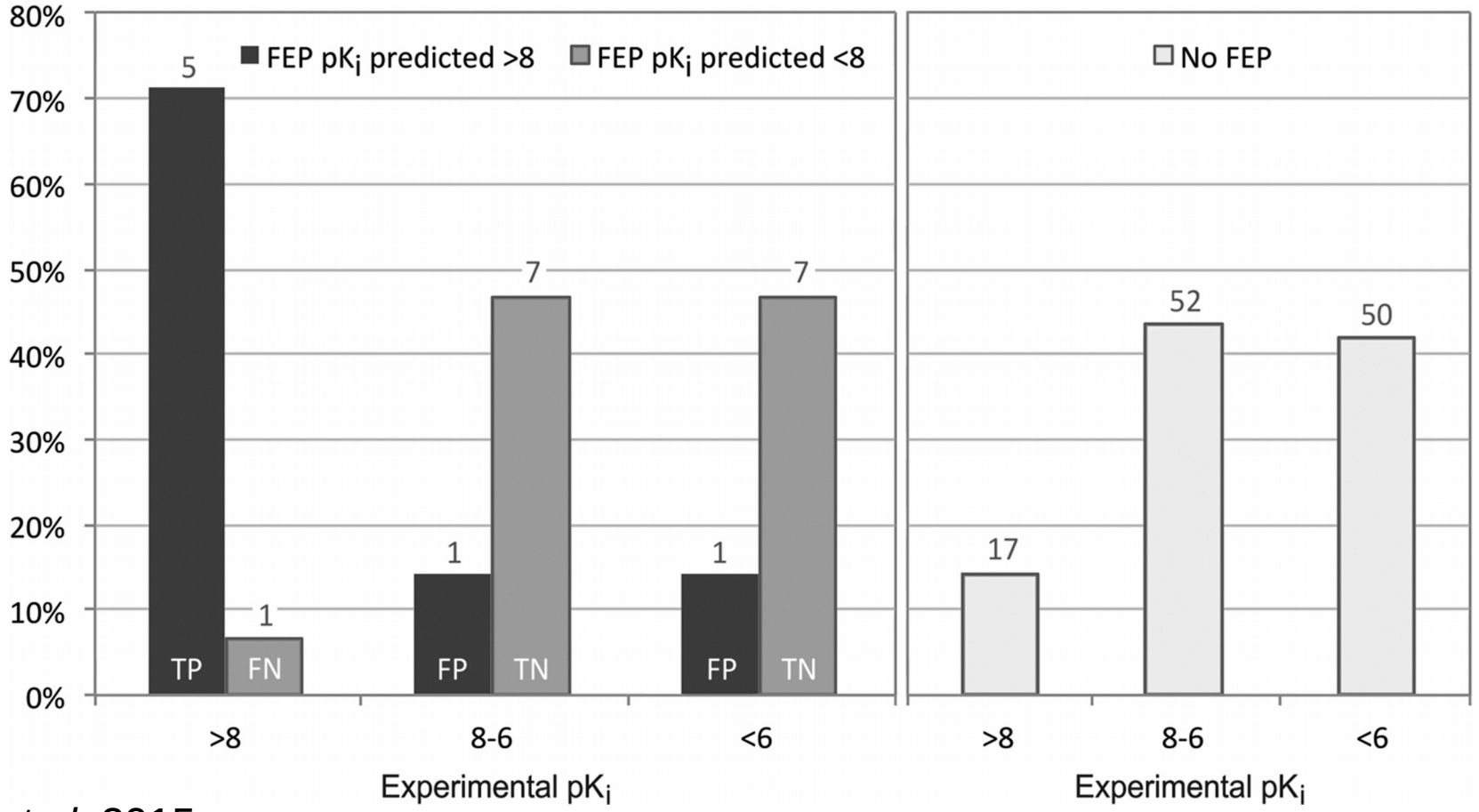
- For eight different receptors, perturbed ligands are predicted to bind with affinities within 2 kcal/mol of experimental results.



Wang, L. *et al.* 2015.

Application to drug discovery

- Proof of principle on inhibitors for IRAK4
- F = False, T = True, P = Positive, N = Negative



Wang, L. *et al.* 2015.

Limitations

- Force field accuracy is benchmarked against other force fields and QM data (chemical torsions and conformational energies), but not validated against experiment.
- OPLS2.1 force field is possibly over fit.
- There are important constraints on the ligand (other than tightness of binding) that are important that are not addressed and the possible perturbations of candidate ligands is limited by the starting structure.

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*]

Nikhil Goel

Problem: Proteins Fluctuate

Conformation Sampling

Existing Strategies

A battle of biases

Soft Docking

Bias: Flexibility

- Loosen the steric component.
- Which ligands *might* be accommodated by certain protein rearrangements?
- Consequence: Increased docking false positives.

Explicit Docking

Bias: Known Structures

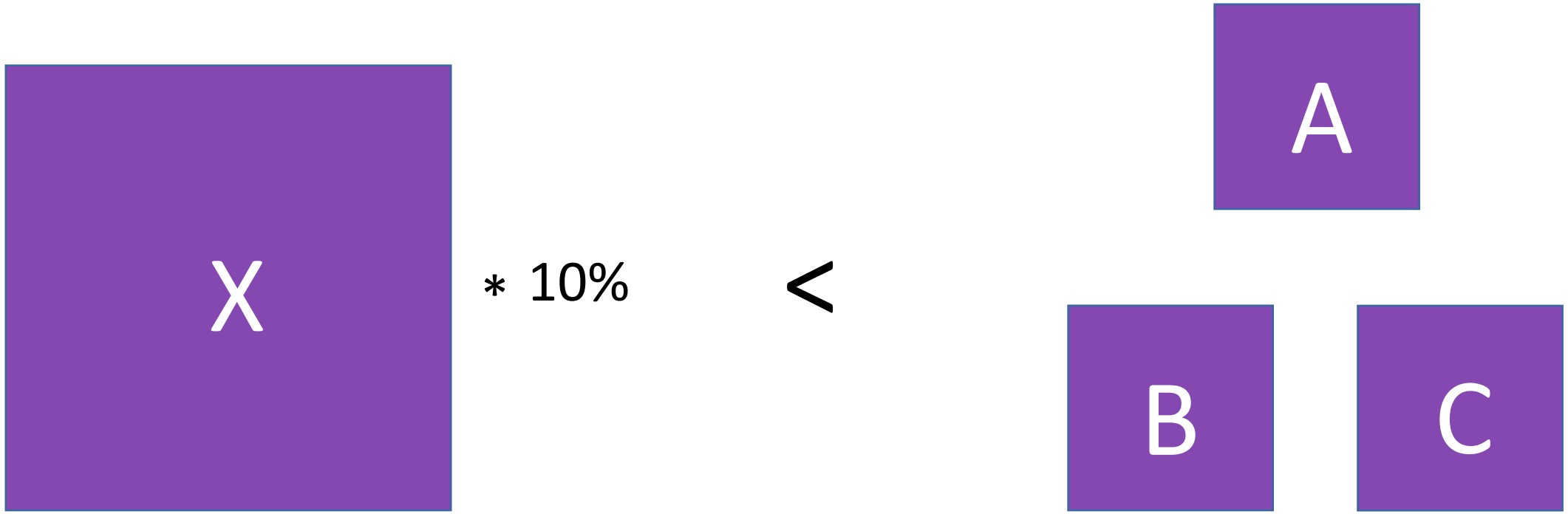
- Ensures accessible states.
- Which ligands *are* accommodated by certain protein rearrangements?
- Consequence: Limited # of states.

New Strategy

Flexible Docking

Higher-Energy Alternatives

Bias: Local Minima

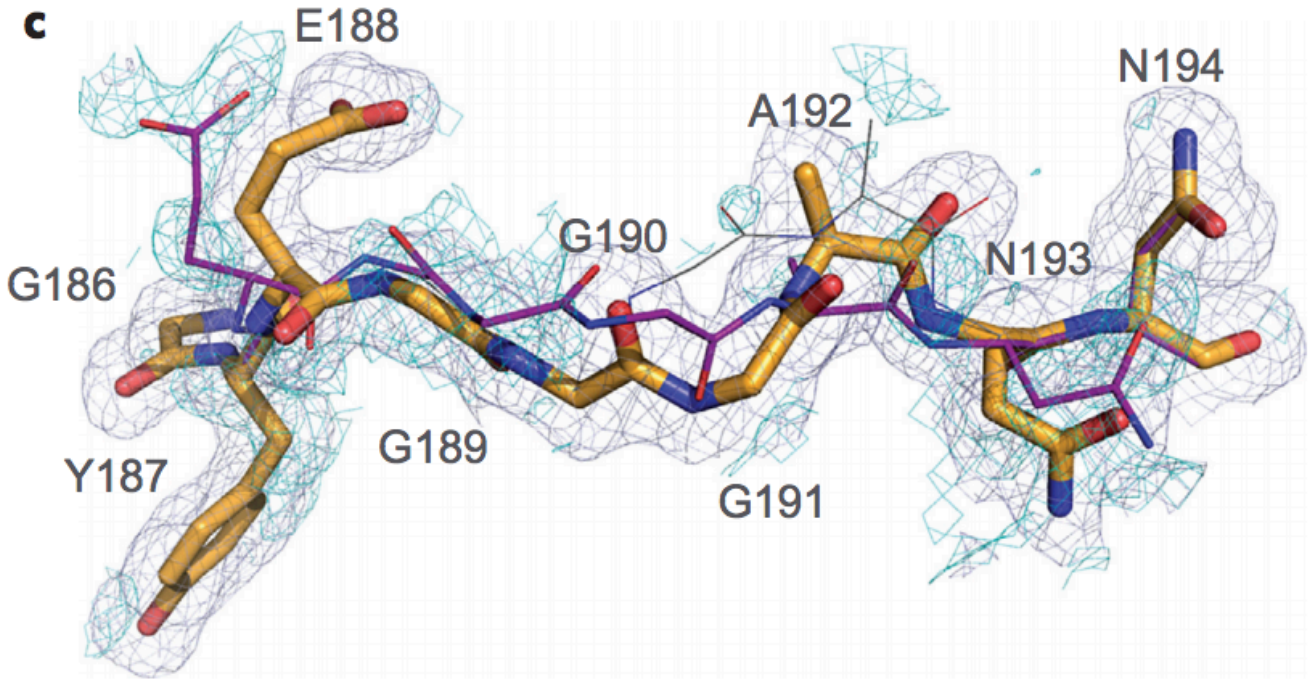


Anionic cavity site in cytochrome c peroxidase

583,363 compounds

x

16 energy-weighted conformations

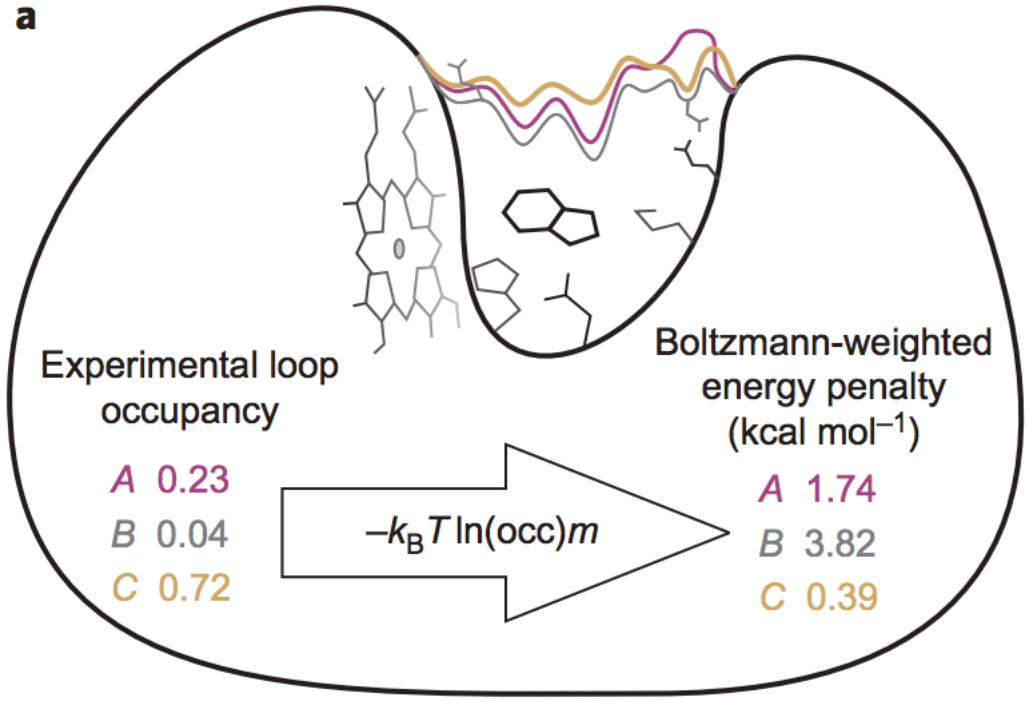


Decomposed interaction energies into additive function

->

16 receptor states with only 2.4x speed cost compared with 1 state

Occupancies -> Energy Penalties

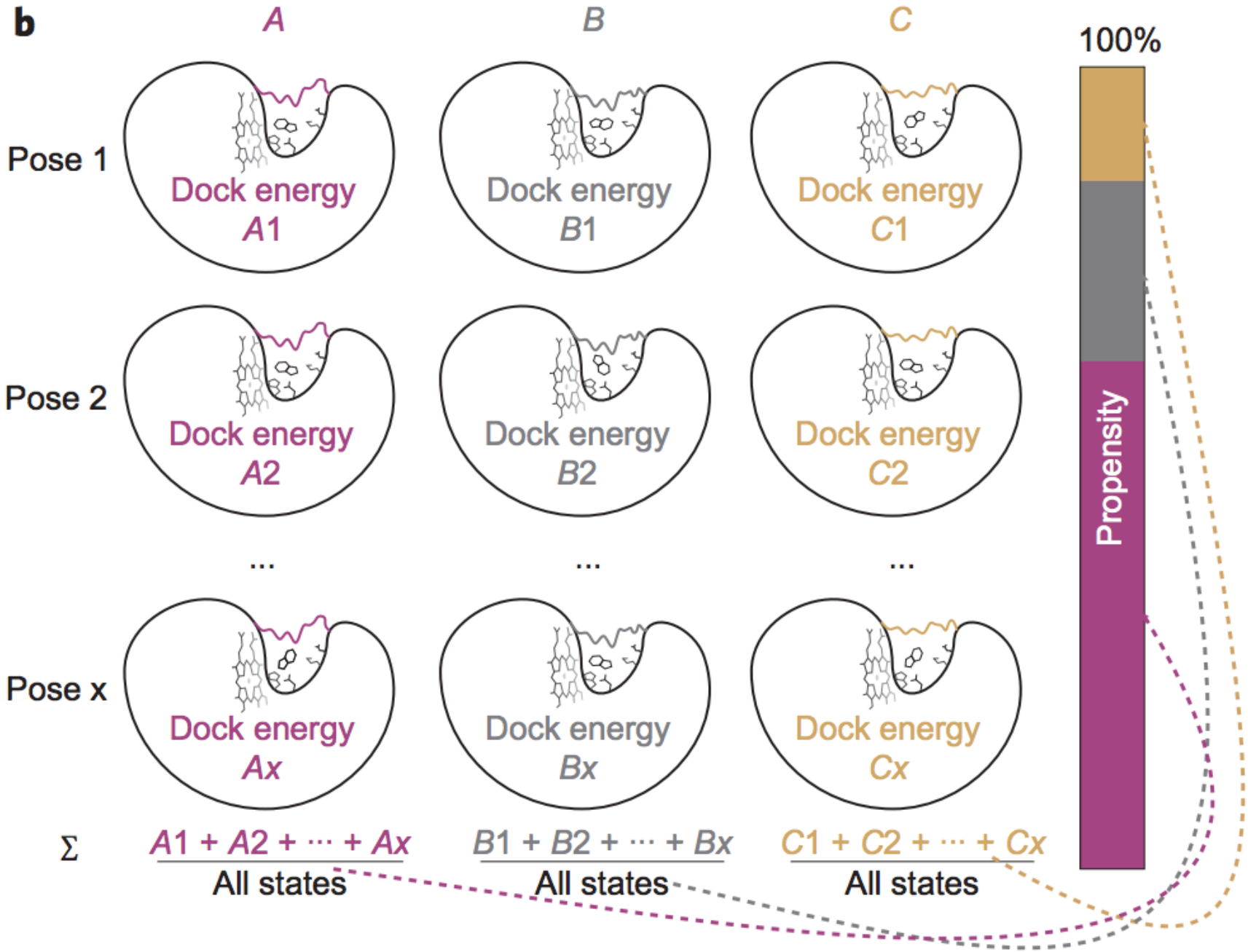


$$\text{energy penalty (conformation } A) = -k_B T \ln(\text{occ}(A))m$$

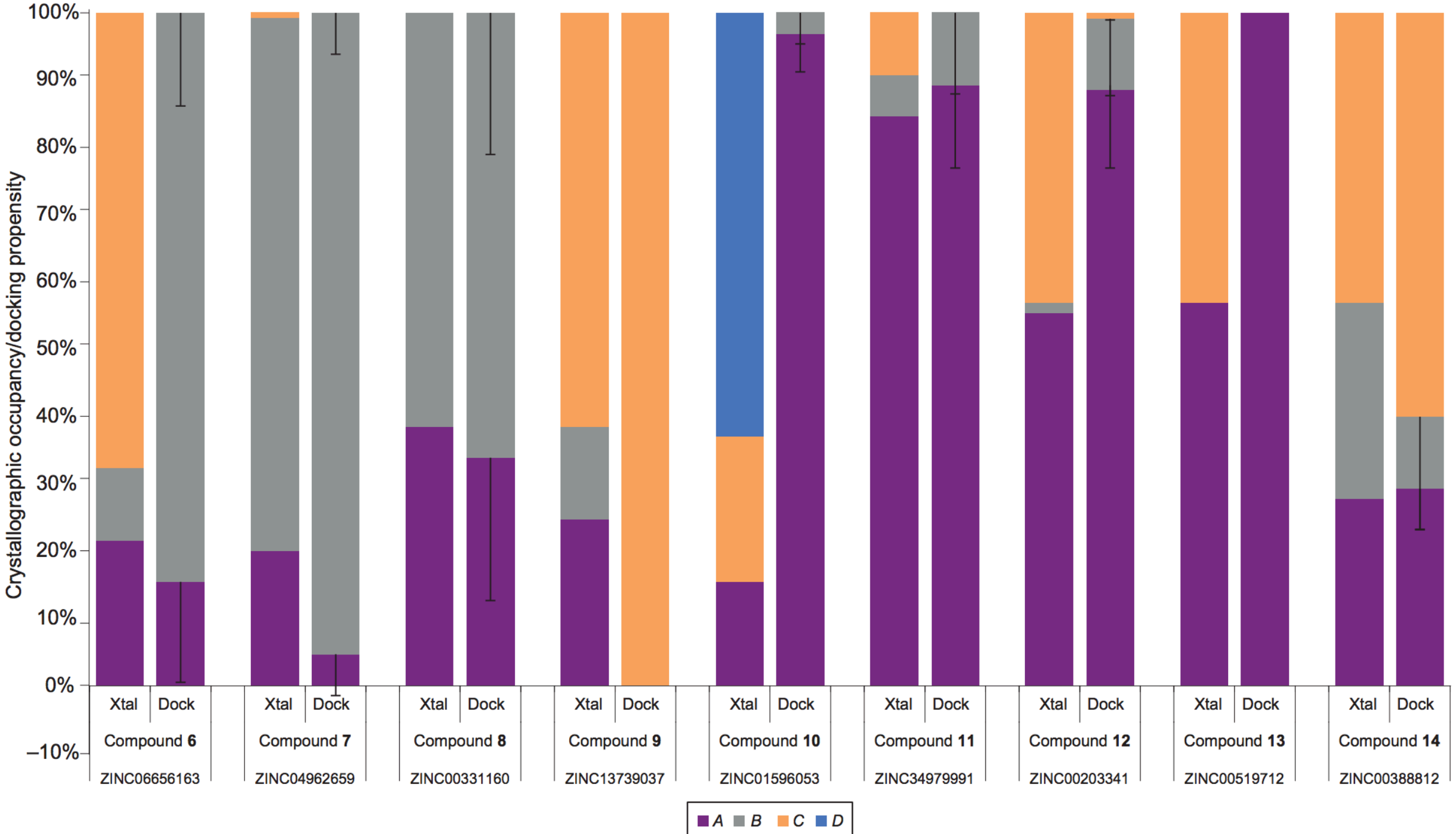
k_B = Boltzmann constant
T = temperature (K)
m = flexible weighting multiplier

$$\text{dock energy (loop } X\text{, ligand } Z) = \text{energy penalty(loop } X) + \Sigma_{\text{atom } z \in \text{ligand } z} \text{Vdw}(z) + \text{elstat}(z) + \text{ligand desol}(z)$$

$$\text{propensity}_{\text{ligand } Z, \text{loop } X} = \frac{\Sigma_{x \in \text{loop } X} e^{\text{dock energy}(x,Z)/kT}}{\Sigma_{y \in \text{all states}} e^{\text{dock engery}(y,Z)/kT}}$$


b
A
B
C
100%
Pose 1
Dock energy A1
Dock energy B1
Dock energy C1
Pose 2
Dock energy A2
Dock energy B2
Dock energy C2
...
...
...
Pose x
Dock energy Ax
Dock energy Bx
Dock energy Cx
Propensity
Σ
A1 + A2 + ··· + Ax
All states
B1 + B2 + ··· + Bx
All states
C1 + C2 + ··· + Cx
All states

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

Three Principle Observations

1) Alternative protein conformations with calculated energies.

2) Prediction of ligands with new properties.

3) We can apply this method to more proteins!

Strengths

- Novel Technique.
- Interesting findings in terms of the relationship between loop occupation and ligand propensity.
- Ligands with new phenotypes and chemical properties!

Limitations

- Data Availability
- Would this work for “messier” systems?
- In selecting a “simpler” system, have they created an ad hoc method?
- Docking scores make assumptions.