Computational Structural Methods for Ligand (Drug) Discovery

Trent E Balius
Frederick National Laboratory for Cancer Research
NCI RAS Initiative

Guest lecture for MSB 530 Bioinformatics

2019/08/13

The hook: Why you should care

- Drugs impact human heath and quality of life!
- Drug discovery is a long and expensive process.
- Almost all recent drugs have been touch by computational methods: molecular graphics, molecular docking, free energy calculations …

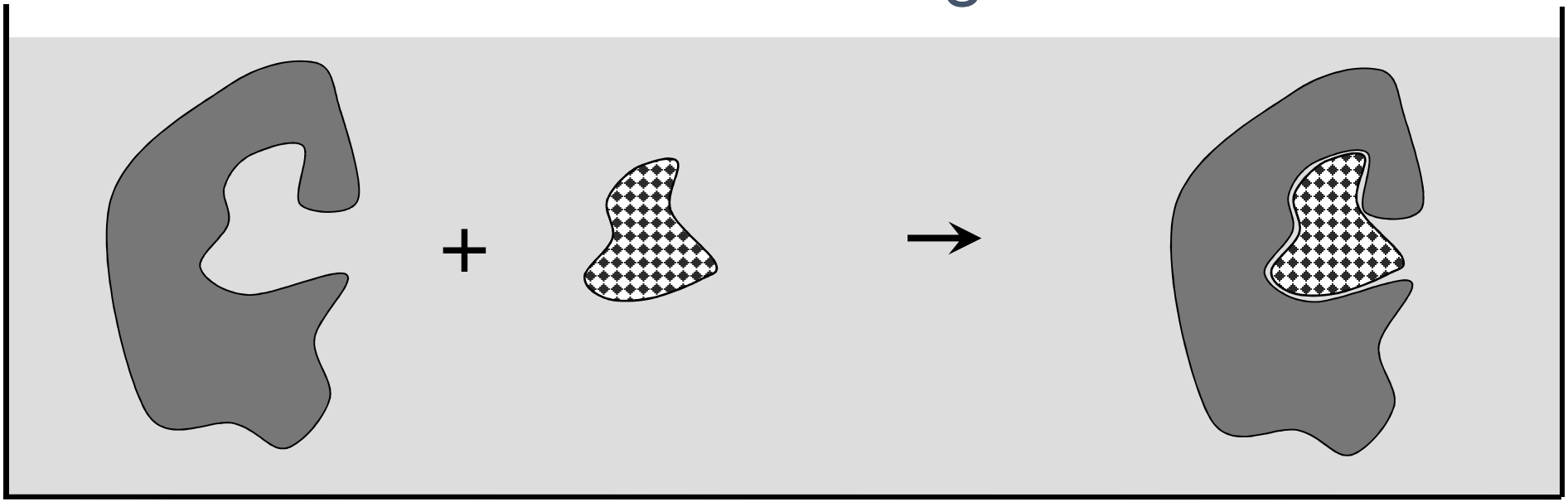
Outline

- The binding event using a thermodynamic cycle
- Introduction to molecular mechanics
- Introduction to molecular dynamics
- Introduction to molecular docking
- Applications from my work
    - Large-scale docking
    - Receptor flexibility
    - Receptor desolvation

Outline

- The binding event using a thermodynamic cycle
- Introduction to molecular mechanics
- Introduction to molecular dynamics
- Introduction to molecular docking
- Applications from my work
    - Large-scale docking
    - Receptor flexibility
    - Receptor desolvation

Molecular Recognition



- Understanding the binding event -- important for drug discovery
- Structure-based, Targeted drug discovery
- Computational methods

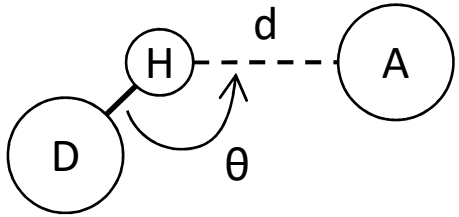
Important molecular forces for ligand binding

- Electrostatics
- Van der Waals
- Hydrogen bond
- Dipolar interactions
- Quadrupole interactions

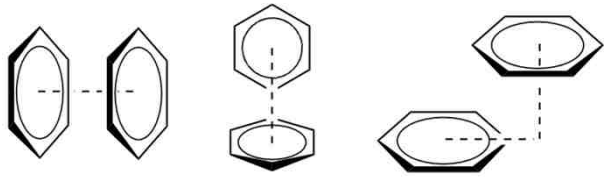
- Interactions with water (solvent)
  - Hydrophobic effect

- Entropy

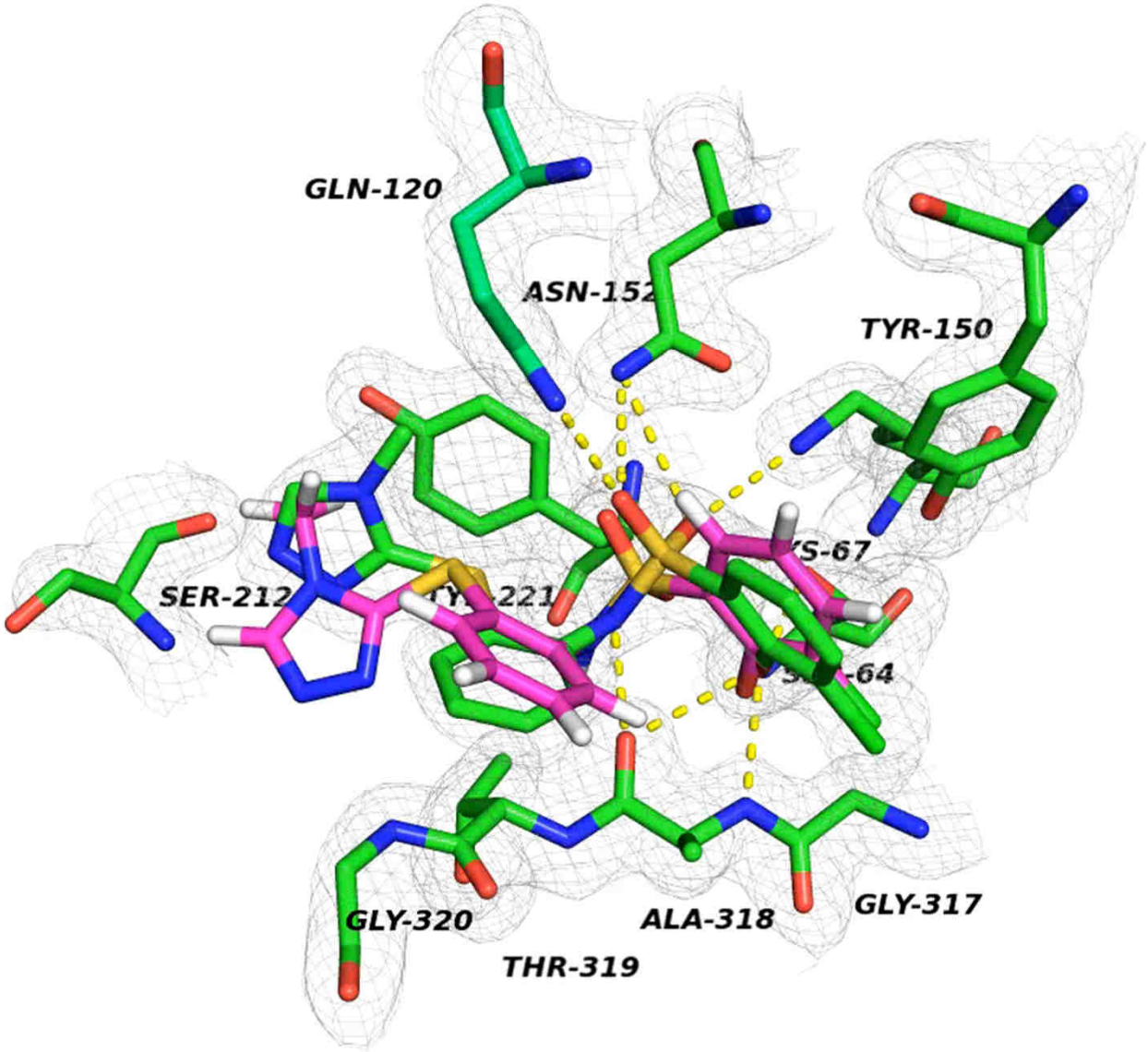
Hydrogen bond



Pi-stacking

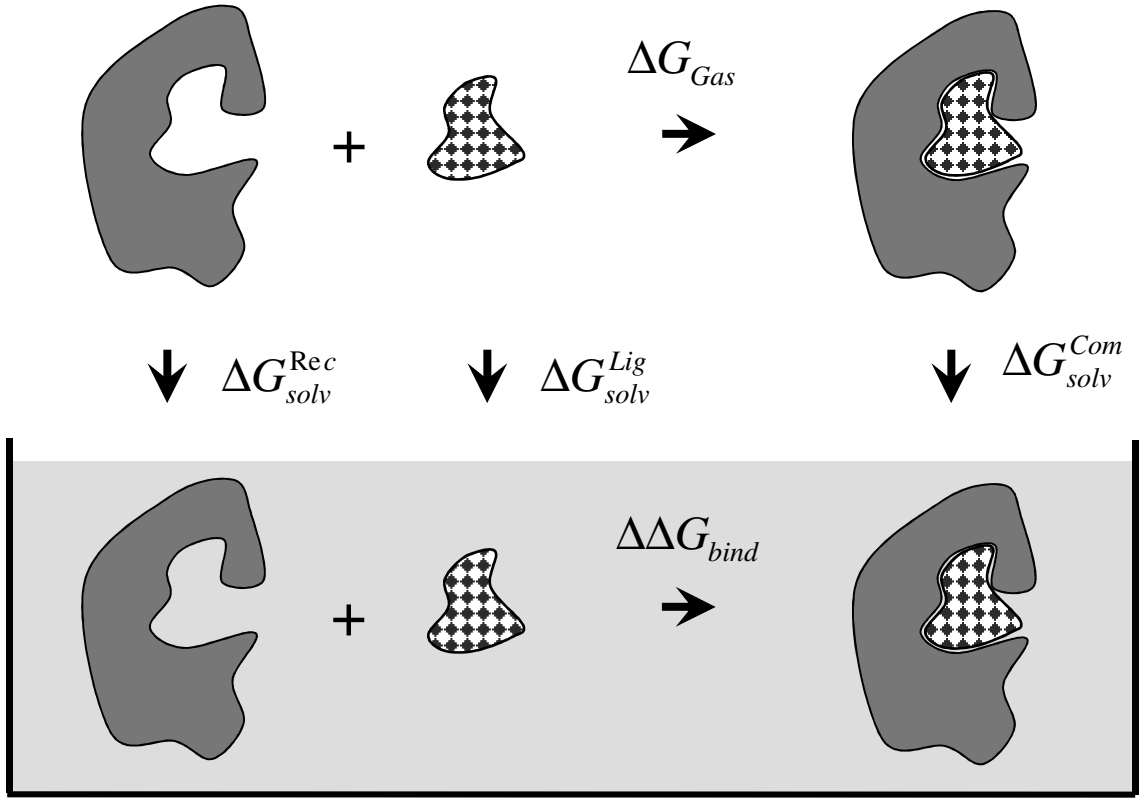


Molecular Recognition



Critical role of water in receptor-ligand binding

Thermodynamic cycle of the binding event



$$\Delta G_{gas} = \Delta G_{VDW} + \Delta G_{ES}$$

$$\Delta G_{desolv} = \Delta G_{solv}^{Com} - \left(\Delta G_{solv}^{Lig} + \Delta G_{solv}^{Rec}\right)$$

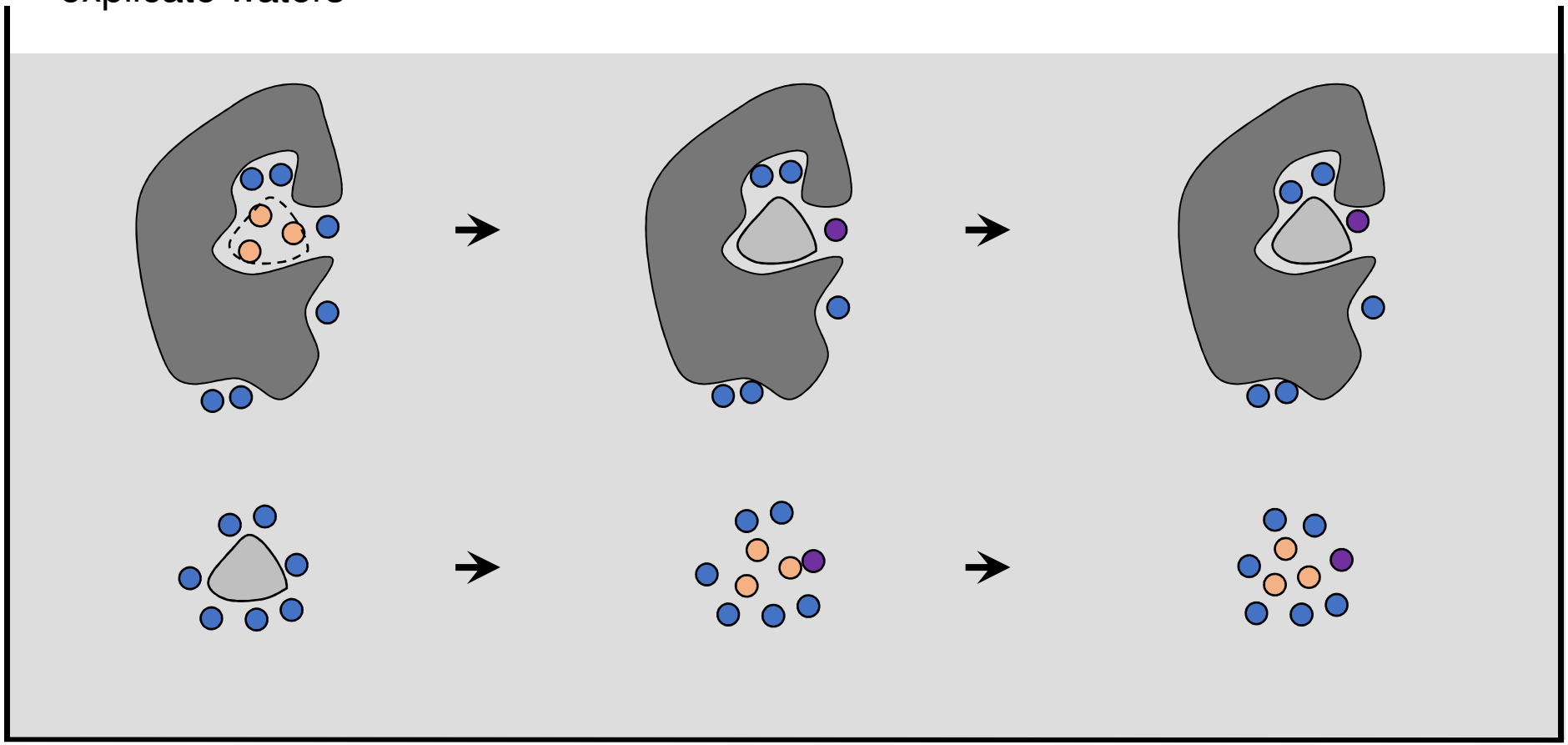
$$\Delta\Delta G_{bind} = \Delta G_{Gas} + \Delta G_{desolv}$$

Water displacement and medicated interactions

Receptor and ligand hydrated with explicate waters

Ligands and receptor Displaces waters.

Waters adjust to new environment



Outline

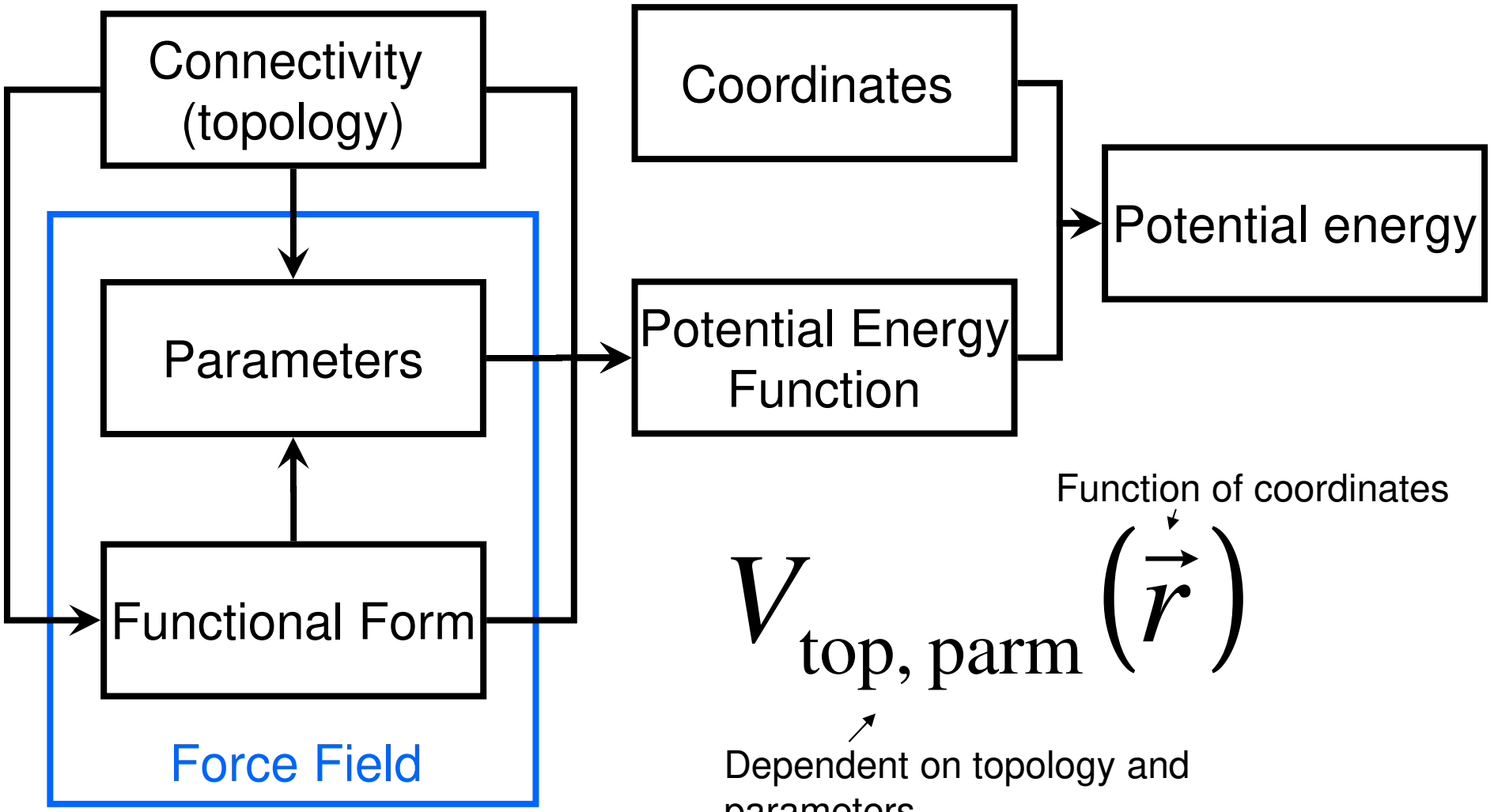
- The binding event using a thermodynamic cycle
- Introduction to molecular mechanics
- Introduction to molecular dynamics
- Introduction to molecular docking
- Applications from my work
    - Large-scale docking
    - Receptor flexibility
    - Receptor desolvation

Molecular Models

- Quantum Mechanics
  - physical, but expensive
  - Schrödinger equation: $\mathrm{H}\,\Psi = \mathrm{E}\,\Psi$
  - wave functions defines electron density

- Molecular Mechanics
  - less physical --> empirical parameterization
  - cheap and accurate

Molecular Mechanics Force Field



Function of coordinates

$$V_{\text{top, parm}}\left(\vec{r}\right)$$

Dependent on topology and parameters

Molecular Mechanics

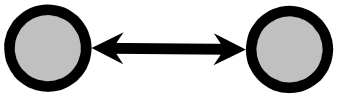
- Every atom is a sphere with a radius (Lennard Jones)
- Point charge is located at each atomic center

  (Coulomb's law)
- Bonds and angles are held by springs to ideal lengths
  - e.g. $V_{bond} = k_b (r - r_0)^2$
  - Hooke's Law, K_b: spring constant, r_0: ideal length
- Dihedrals are represented by sigmoidal function which has energy wells at favorable angles.
- Improper torsions force atoms to be a defined angle to plane.

The "Tinker-toy Model"

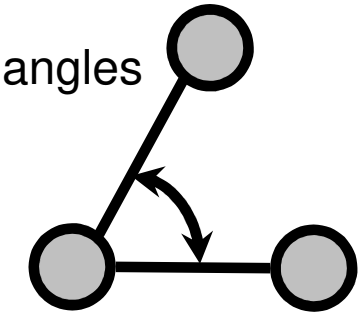
Bonded terms

bonds



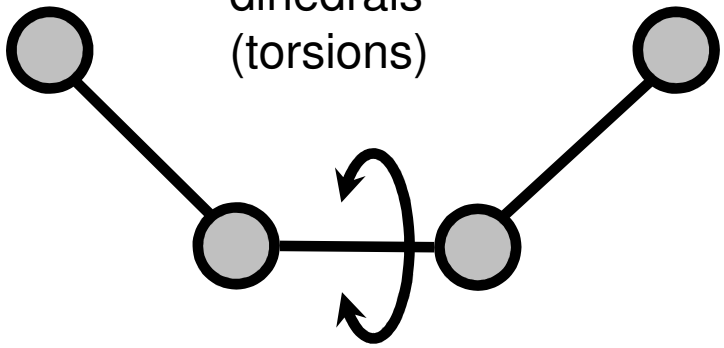
$$V_{bond} = k_b (r - r_0)^2$$

angles



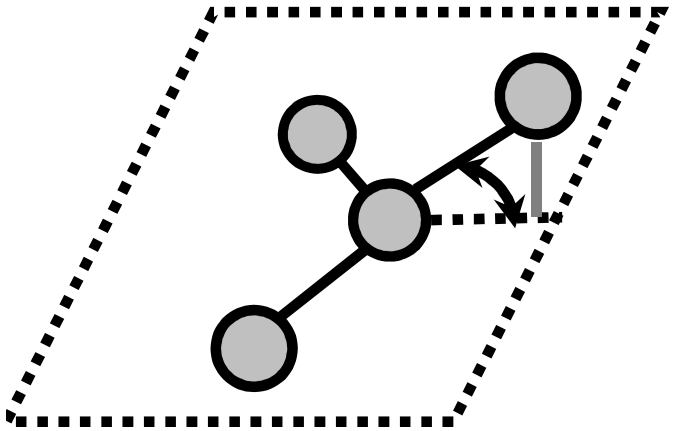
$$V_{angle} = k_\theta (\theta - \theta_0)^2$$

dihedrals
(torsions)

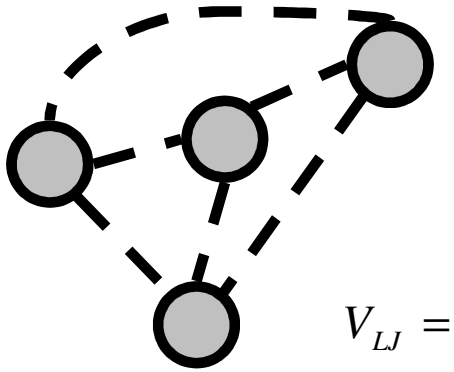


$$V_{dihidral} = K_\chi (1 + \cos(n\chi - \delta))$$

improper



$$V_{improper} = k_\varphi (\varphi - \varphi_0)^2$$



Through space
interactions

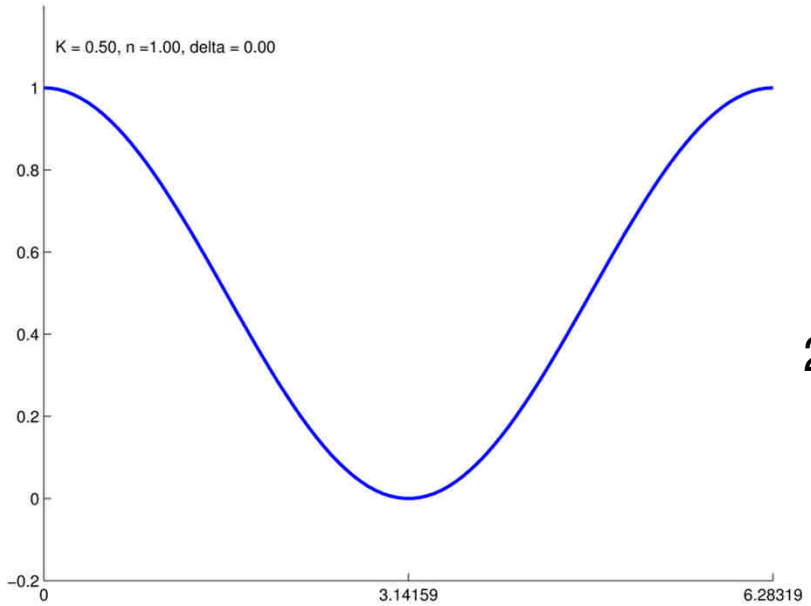
$$V_{coul} = \frac{q_i \; q_j}{\varepsilon \; r_{ij}}$$

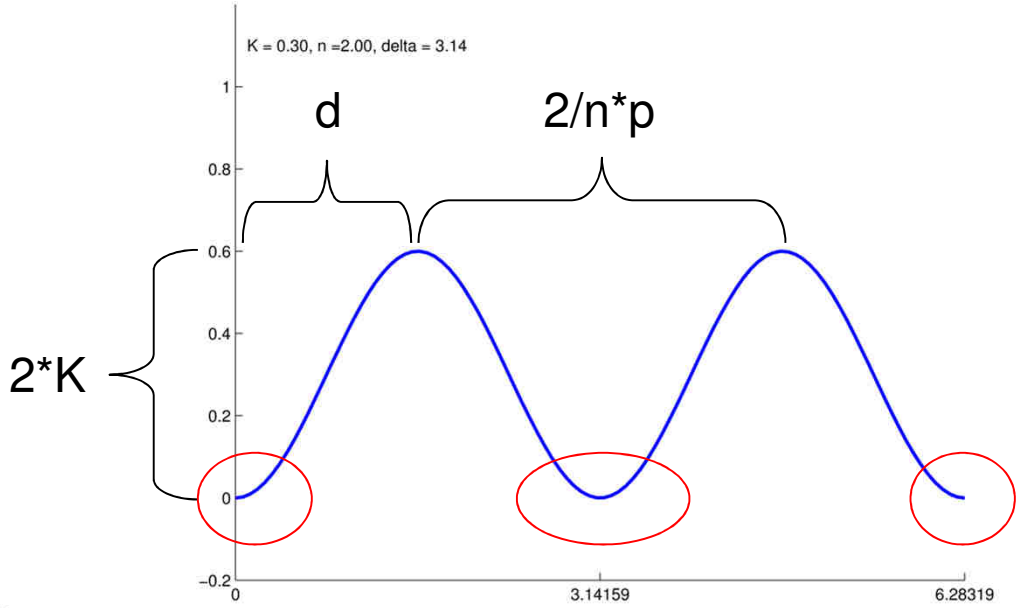
$$V_{LJ} = \varepsilon_{i,j} \left[\left(\frac{R\min_{i,j}}{r_{i,j}} \right)^{12} - 2 \left(\frac{R\min_{i,j}}{r_{i,j}} \right)^{6} \right]$$

$$R\min_{i,j} = \frac{R\min_i + R\min_j}{2} \qquad \varepsilon_{i,j} = \sqrt{\varepsilon_i * \varepsilon_j}$$

Dihedral Term is a Sigmodal Function

$$V_{dihidral} = K_{\chi}(1+\cos(n\chi-\delta))$$

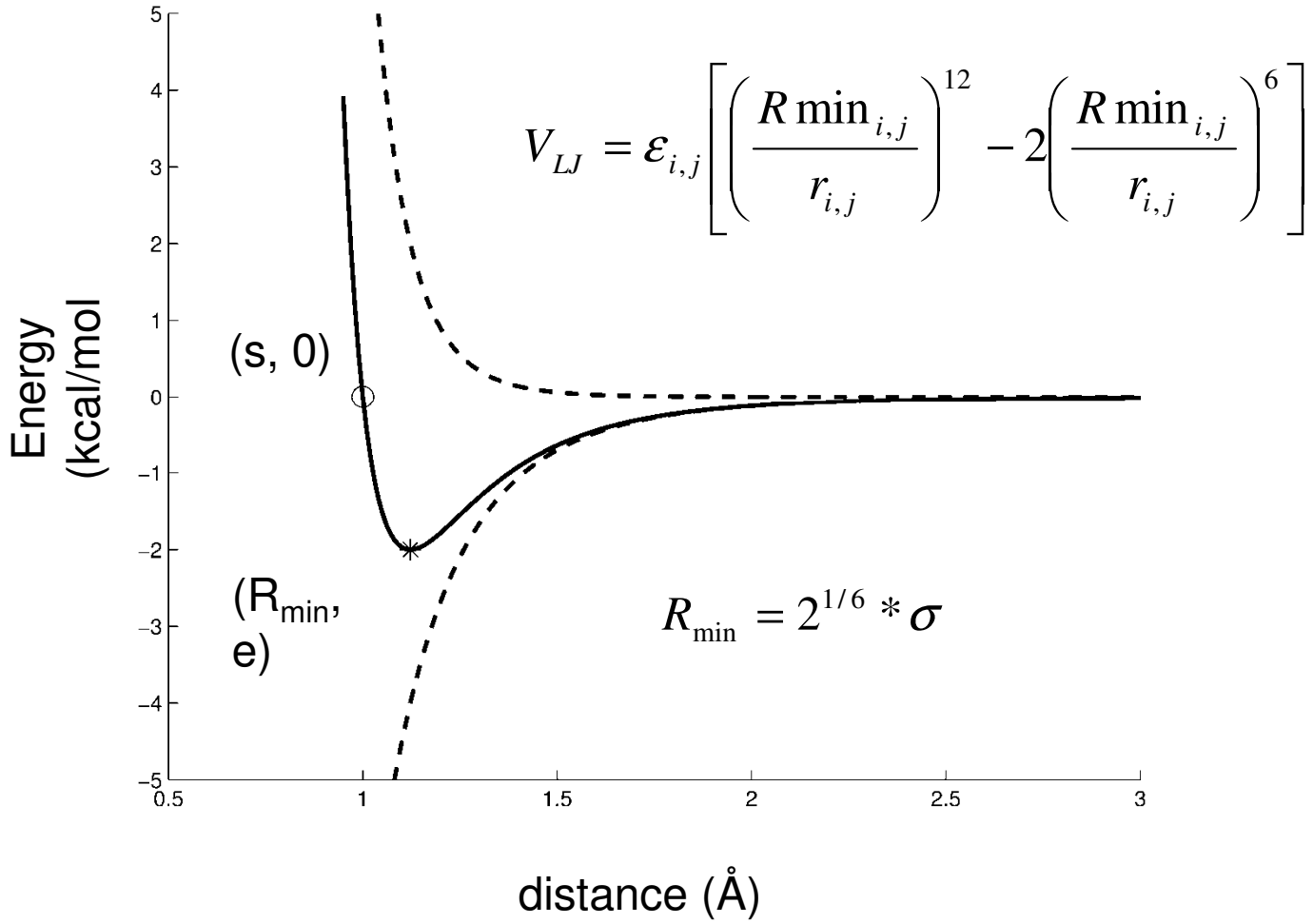




p (3.14) radians = 180 degrees

Molecular Modelling Principles and applications,
Leach Pearson Prentice hall second edition (chapter 4)

Lennard-Jones Equation



Molecular Modelling Principles and applications,
Leach Pearson Prentice hall second edition (chapter 4)

Potential Energy Function

$$V = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{impropers} k_\varphi (\varphi - \varphi_0)^2$$

$$+ \sum_{dihedrals} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \varepsilon_{i,j} \left[\left(\frac{R\min_{i,j}}{r_{i,j}} \right)^{12} - 2 \left(\frac{R\min_{i,j}}{r_{i,j}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon r_{ij}}$$

Different Force-Field

- CHARMM
- AMBER
- GROMOS
- OPLS

Parameterization

- Experimental observables
- Quantum Mechanical calculations

Interdependences among parameters

Molecular Modelling Principles and applications,
Leach Pearson Prentice hall second edition (chapter 4)

Mackerell, Vol. 25, No. 13, Journal of Computational Chemistry

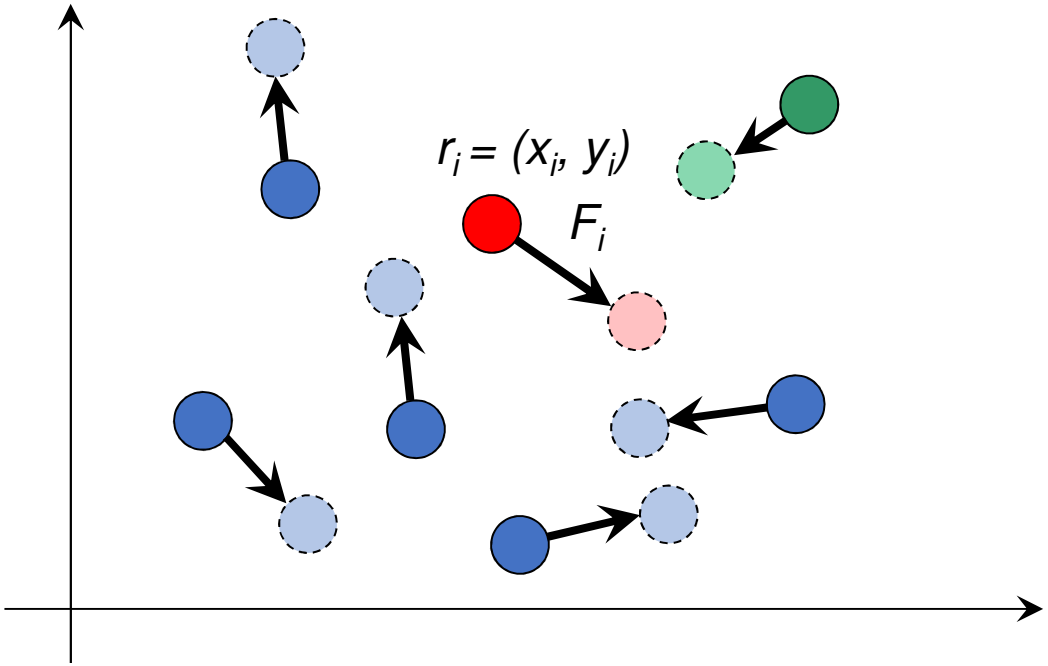
Outline

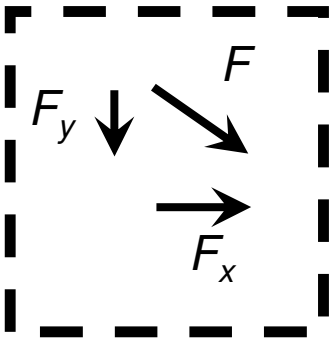
- The binding event using a thermodynamic cycle
- Introduction to molecular mechanics
- Introduction to molecular dynamics
- Introduction to molecular docking
- Applications from my work
    - Large-scale docking
    - Receptor flexibility
    - Receptor desolvation

Molecular Dynamics

$$E(\vec{r})$$
$$\vec{F}_i = -\nabla_i E$$
$$\vec{a}_i = \vec{F}_i / m_i$$
$$\vec{r}_{n,i} = \vec{r}_i + \iint \vec{a}_i dt$$

$$\vec{F}_i = -\nabla_i E_{top,parm}(\vec{r}) = -\begin{bmatrix} \frac{\partial}{\partial x_i} E_{top,parm}(\vec{r}) \\ \frac{\partial}{\partial y_i} E_{top,parm}(\vec{r}) \end{bmatrix}$$





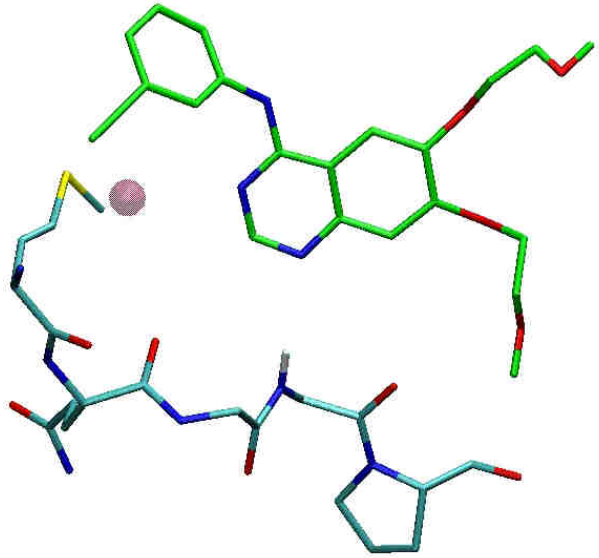
Molecular Dynamics

- Newton Equations

$$E\left(\mathrm{X}_{\text{position}}\right)$$

$$F = -\nabla E$$

$$\mathrm{X}_{\text{position}}^{\text{new}} = \iint \frac{F}{m} \partial t^2$$



- Differential Eq.

  (velocity verlet algorithm)
  - propagate to get motion

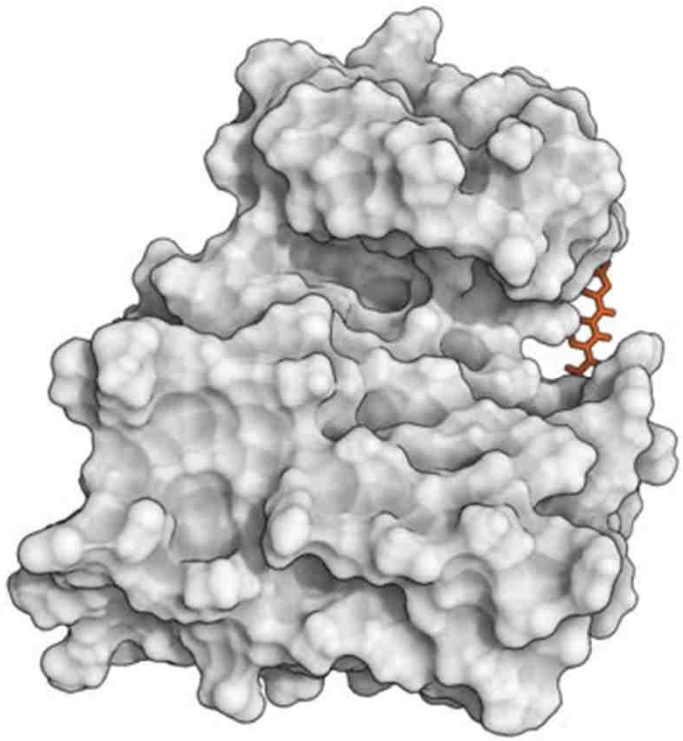
2 fs time step

- Energy functions:

$$E = E_{bonded} + E_{steric} + E_{elect}$$

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral}$$

Simulate Binding of Ligand to Protein



4-μs simulation of dasatinib binding to Src kinase;

Binding occurs 2.5 μs into the simulation (in the 7th second of the movie)

2 fs = 1 time step

1s = 1,000,000 μs

4 μs = 2 billon time steps

Shan, Y; et al. *J. Am. Chem. Soc.*2011, 133, 24, 9181-9183

Some Applications for Modular Dynamics

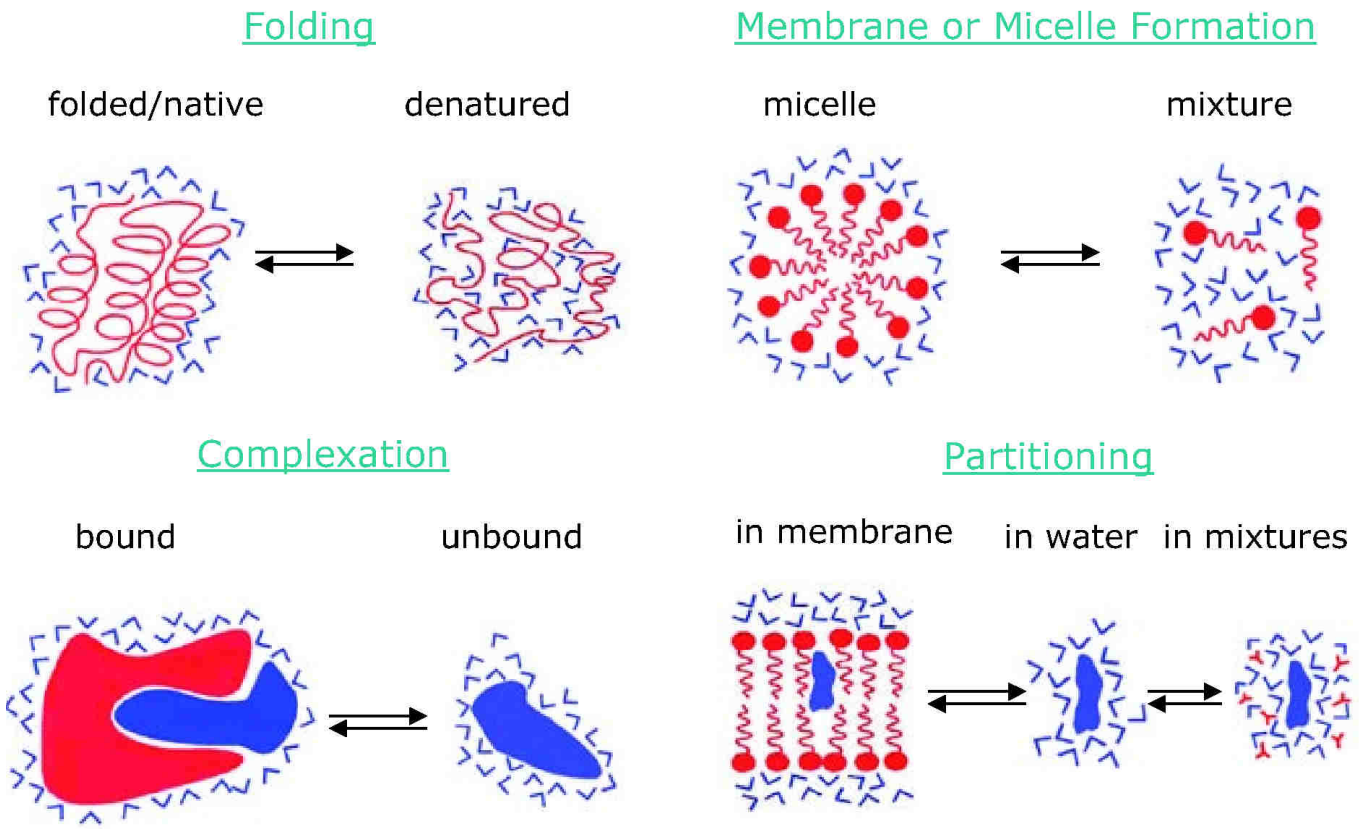


Figure 1. Four biomolecular processes that are governed by thermodynamic equilibria.

Angew. Chem. Int. Ed. 2006, 45, 4064 – 4092

Things to Consider for Molecular Simulations

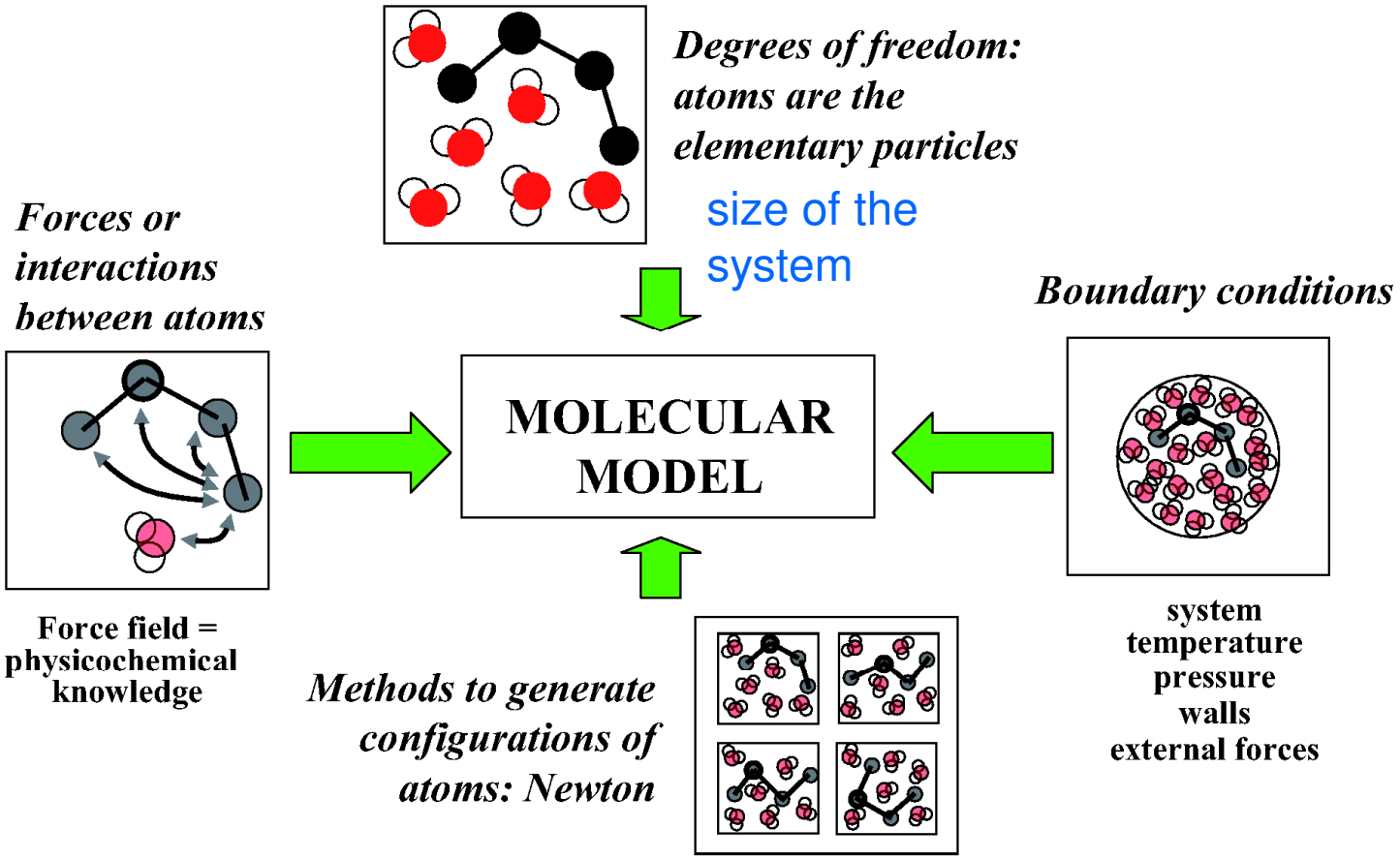


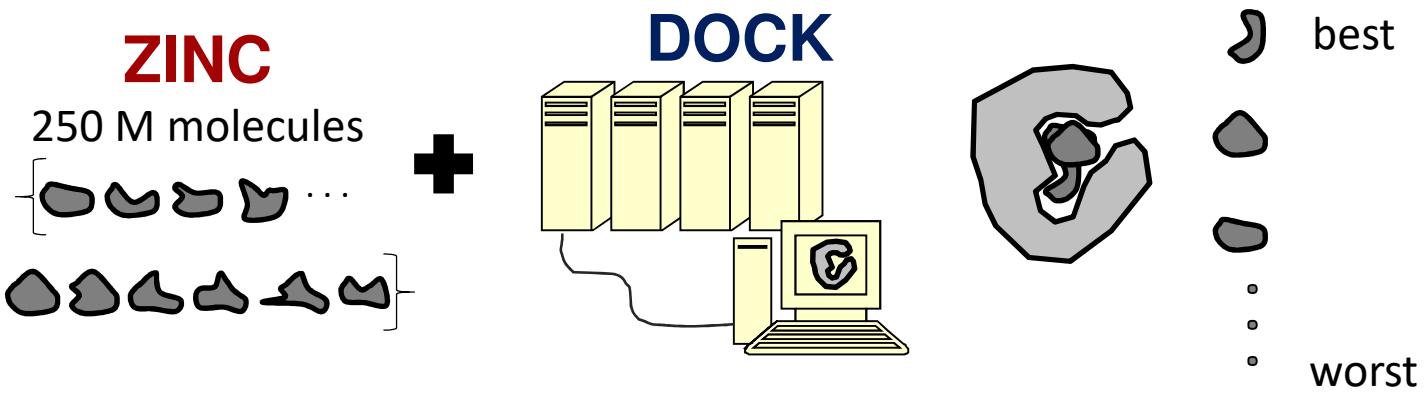
Figure 2. Four basic choices in the definition of a model for molecular simulation.

Angew. Chem. Int. Ed. 2006, 45, 4064 – 4092

Outline

- The binding event using a thermodynamic cycle
- Introduction to molecular mechanics
- Introduction to molecular dynamics
- Introduction to molecular docking
- Applications from my work
    - Large-scale docking
    - Receptor flexibility
    - Receptor desolvation

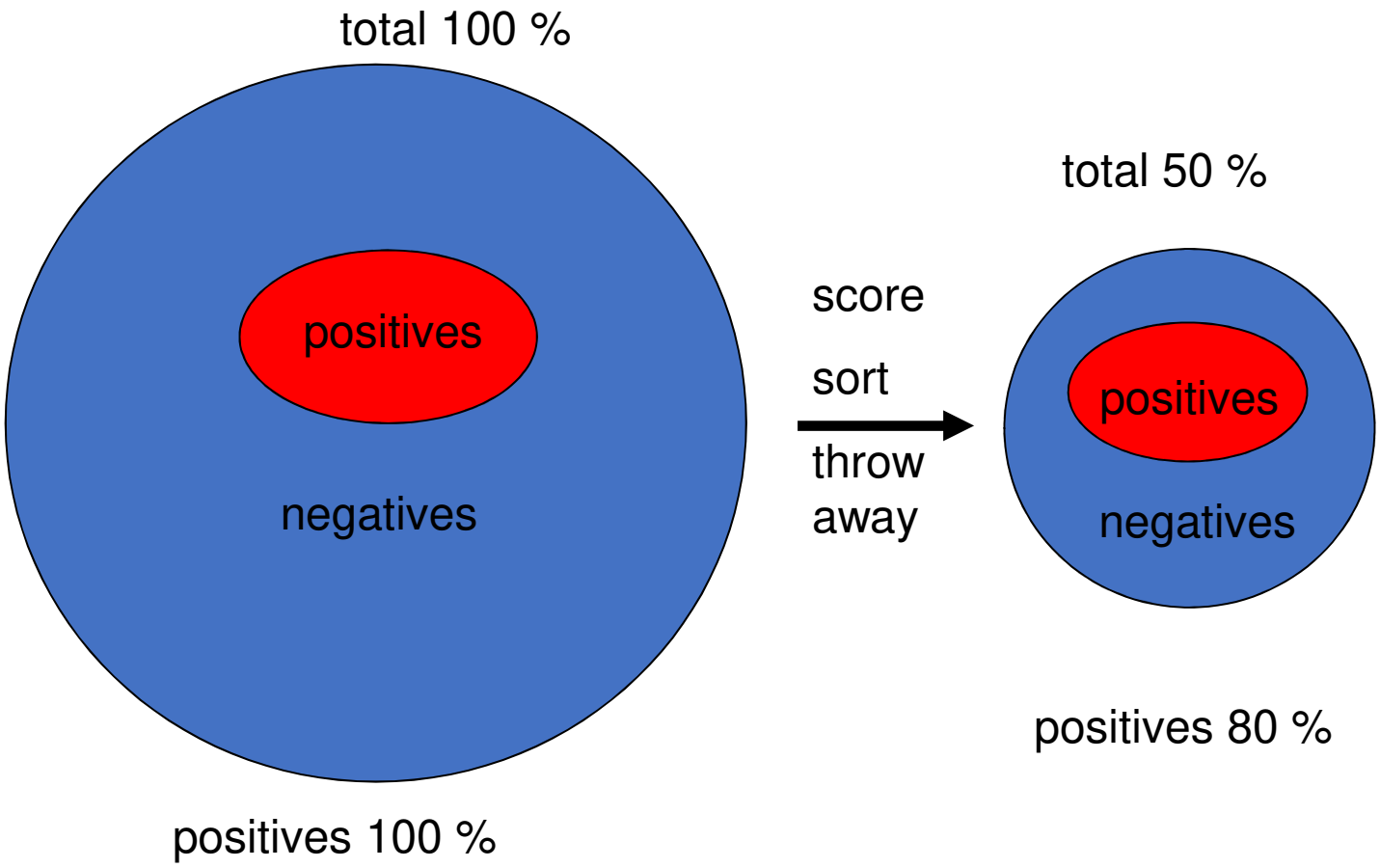
Docking is ligand discovery tool



Docking Tasks

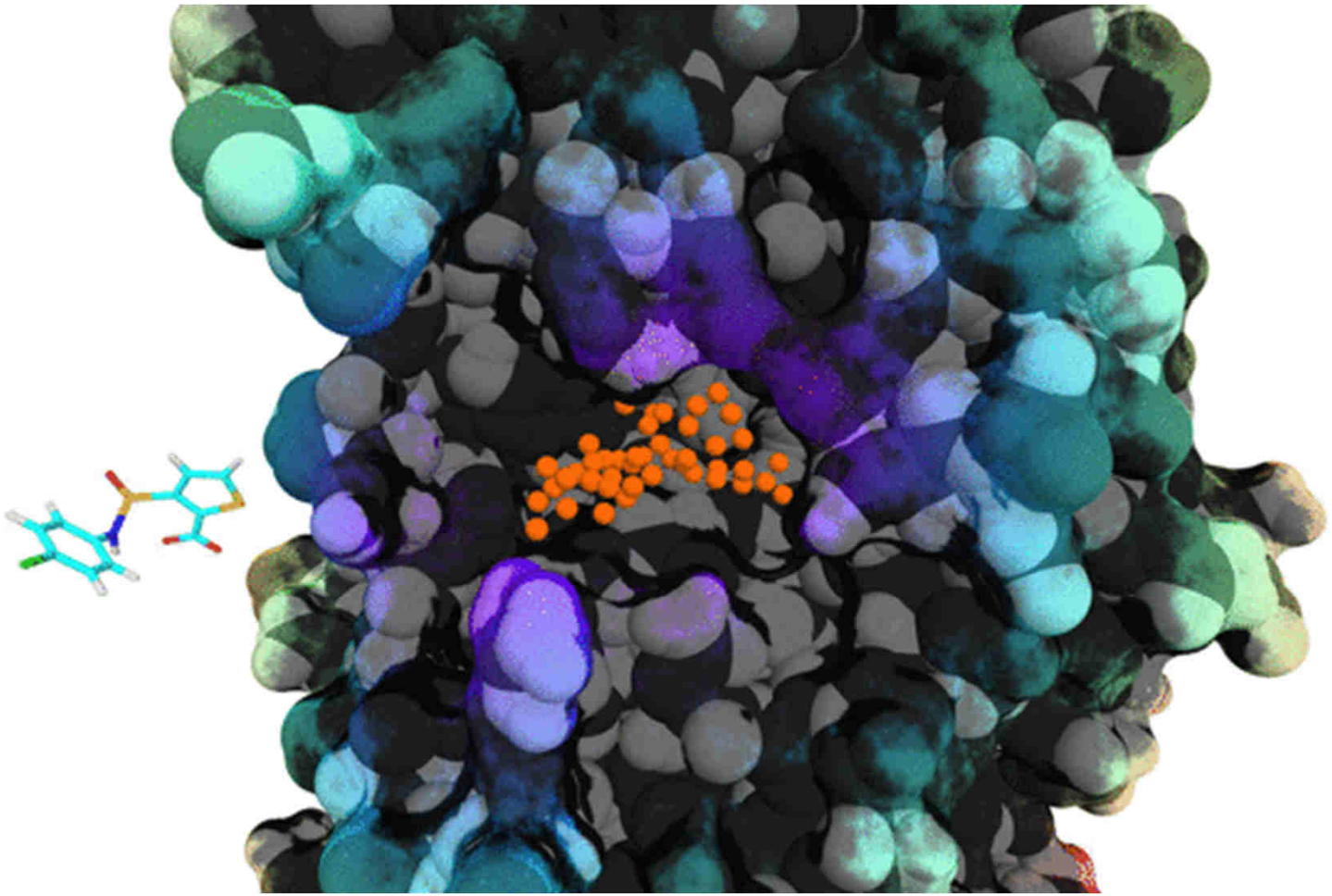
- Sampling
- Scoring
- Balance of speed and accuracy

Enrichment for positives through docking

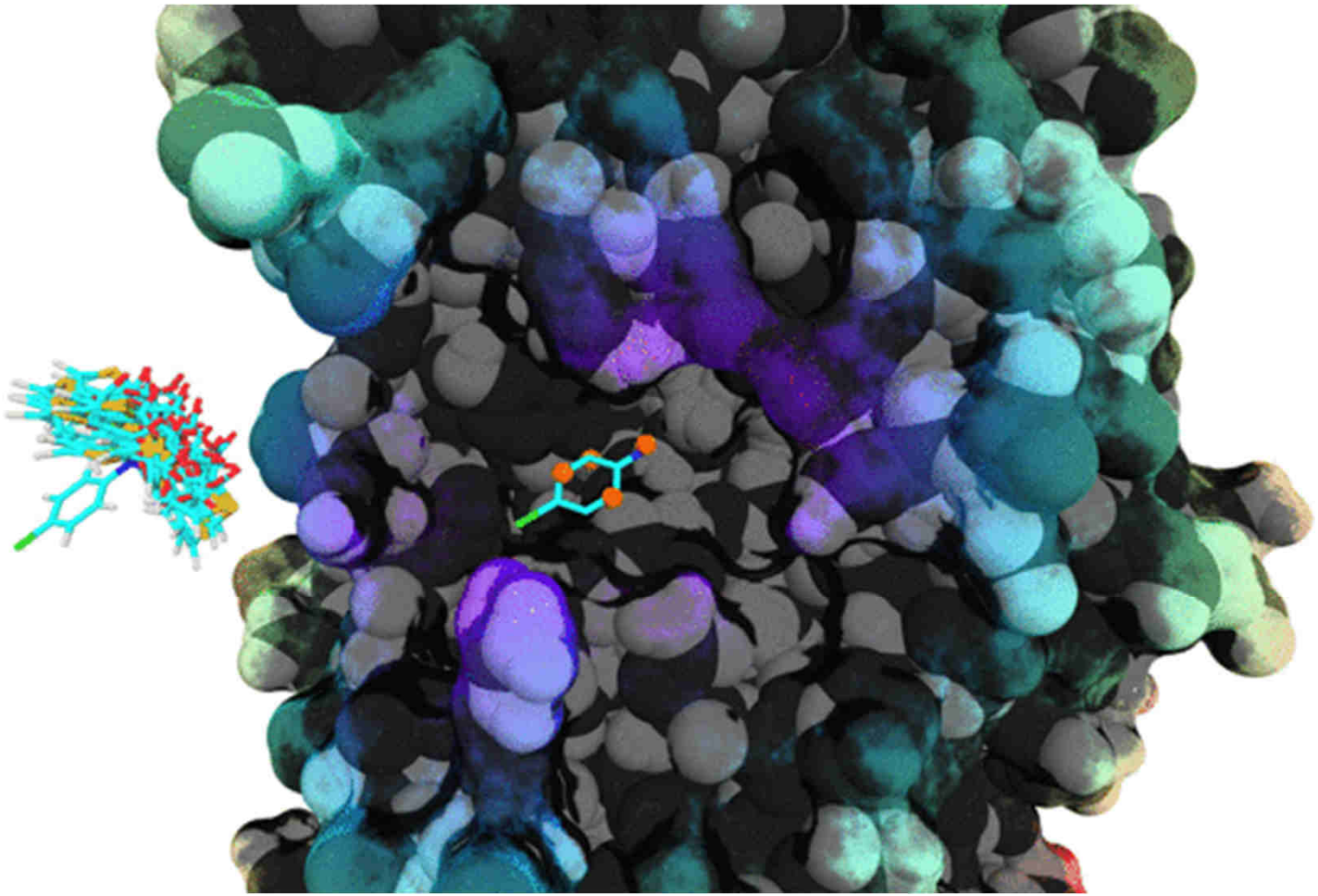


J. Med. Chem., 2006, 49 (23), pp 6789–6801

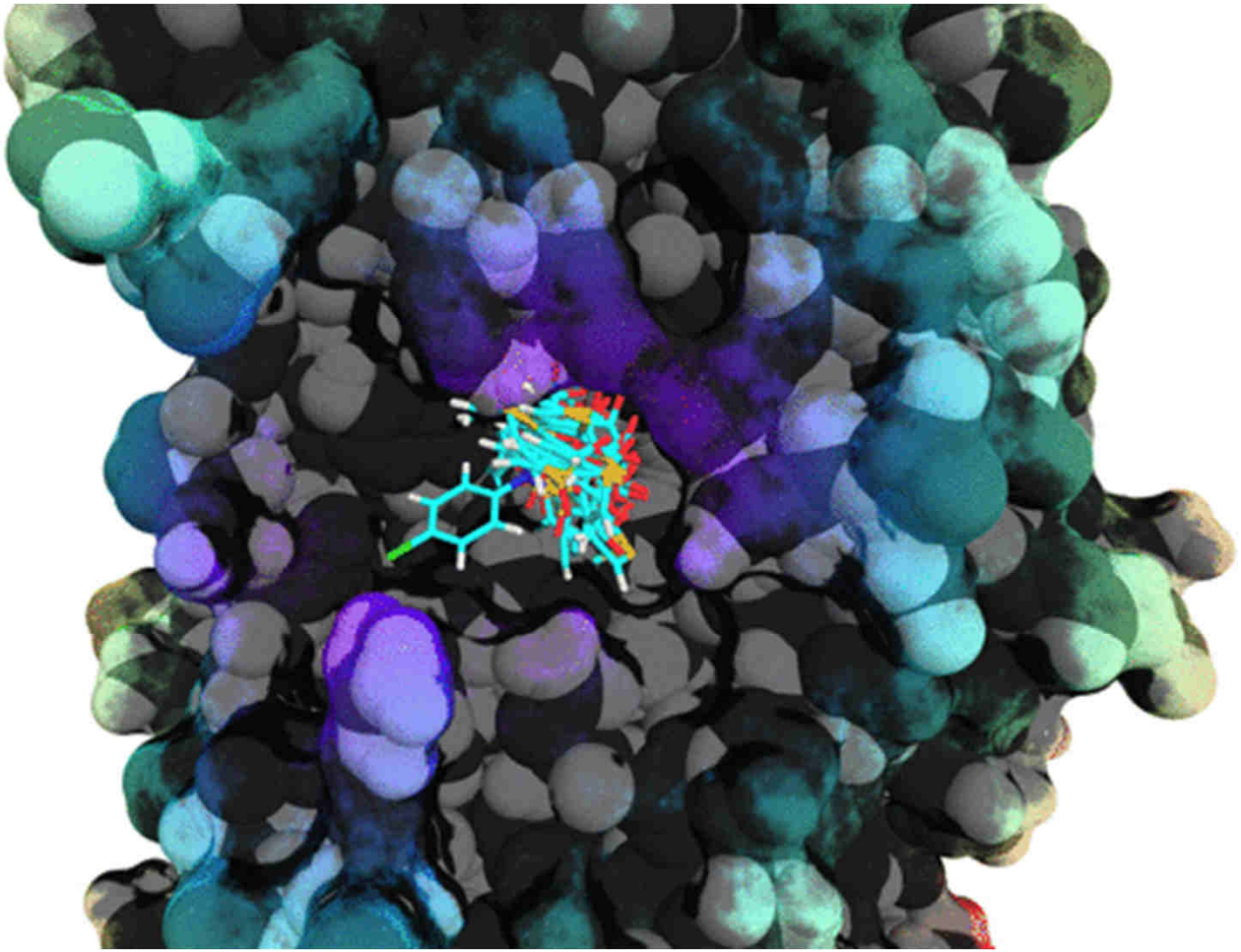
How it works in DOCK3.7—the movie



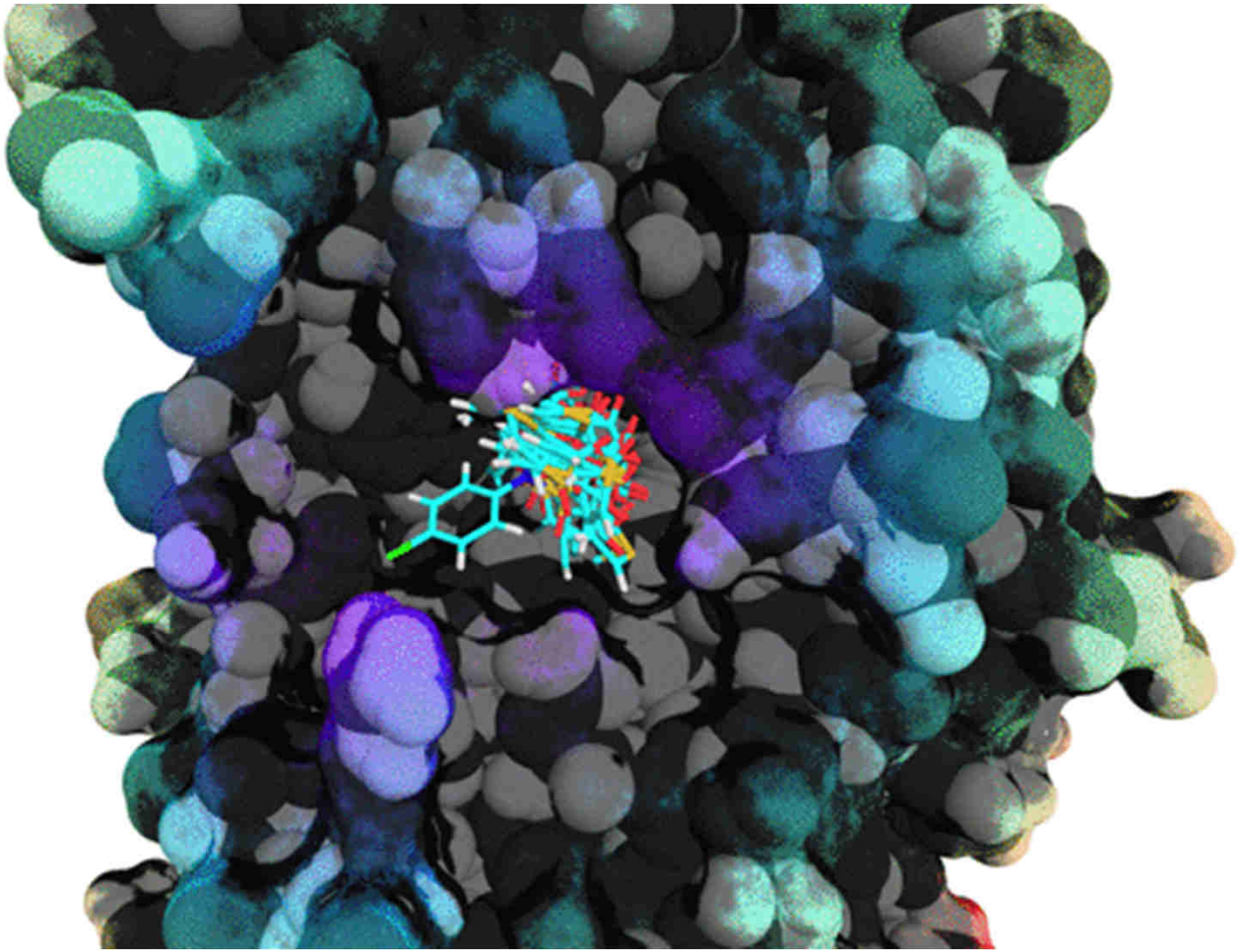
How it works in DOCK3.7—the movie



How it works in DOCK3.7—the movie

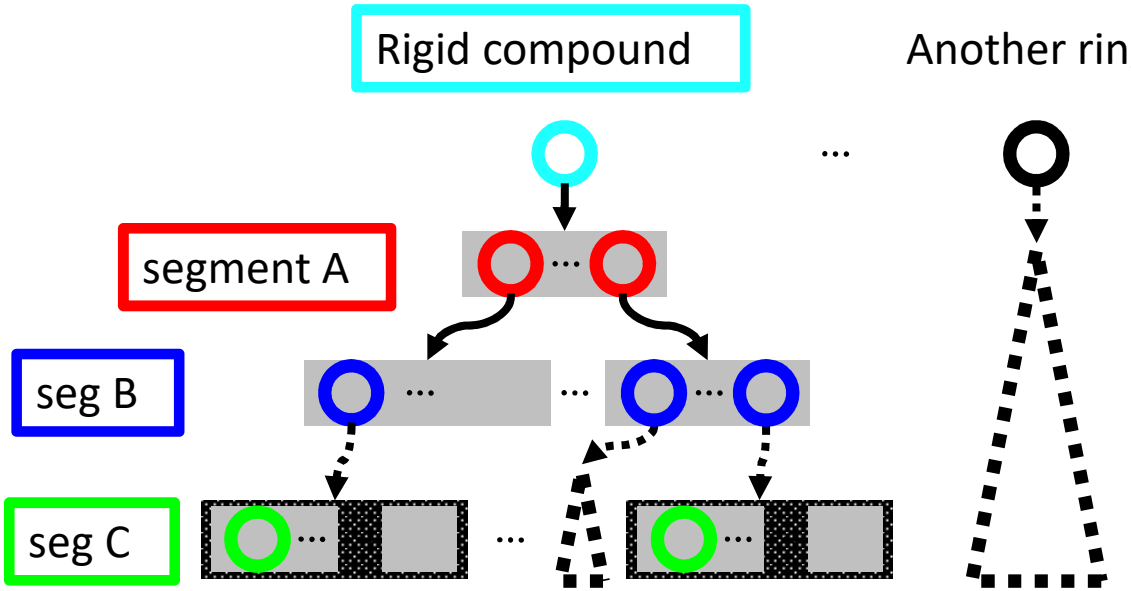


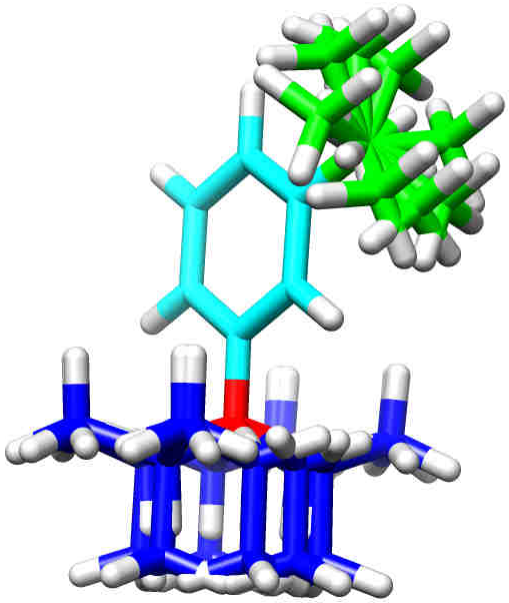
How it works in DOCK3.7—the movie

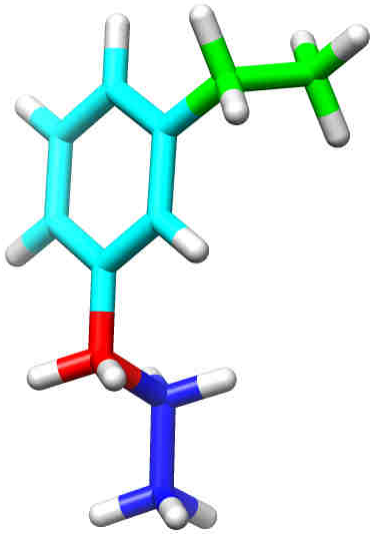


Ligand Sampling, Degrees of Freedom

Database Construction





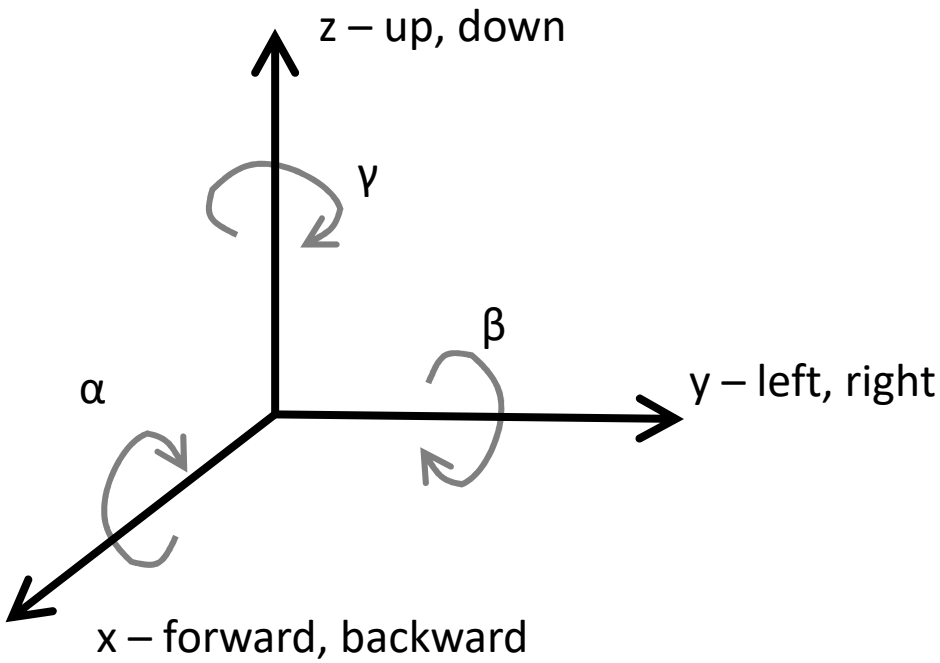


Fully Grown Conformers

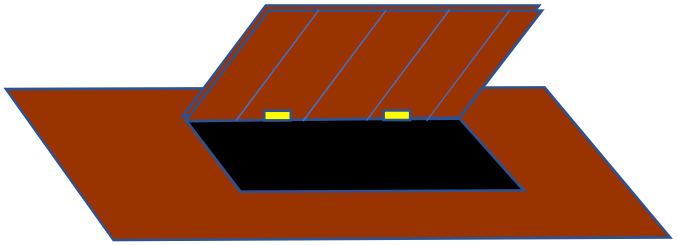
Branch
(conformer)

5 internal Degrees of Freedom

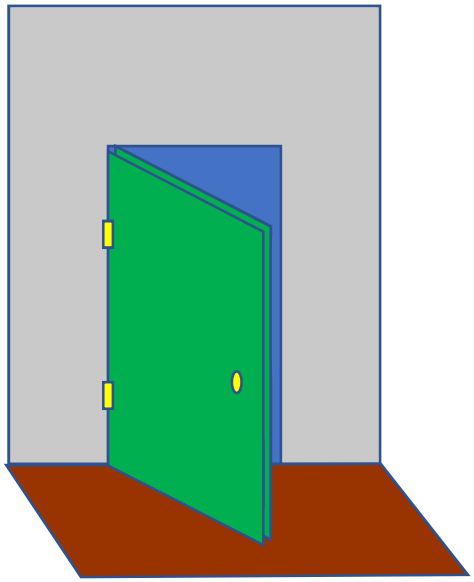
Six degrees for freedom



Trap door (hinge on y axis)



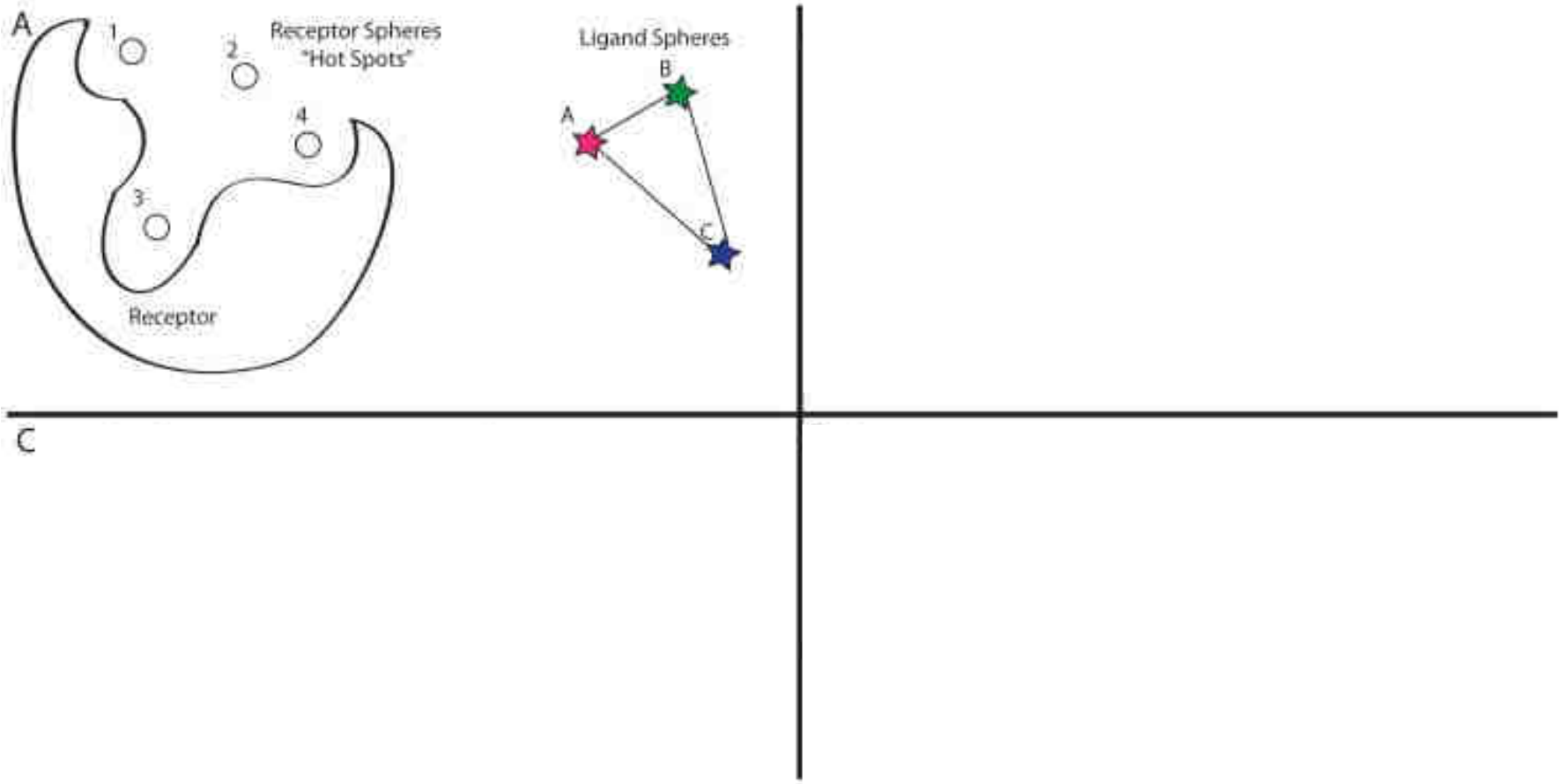
front door (hinge on z axis)



doorknob (x axis)

How Sampling works (orientational)

A toy example illustrating the matching sphere orientational matching algorithm

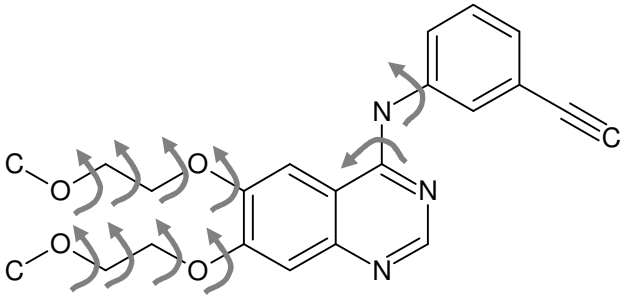


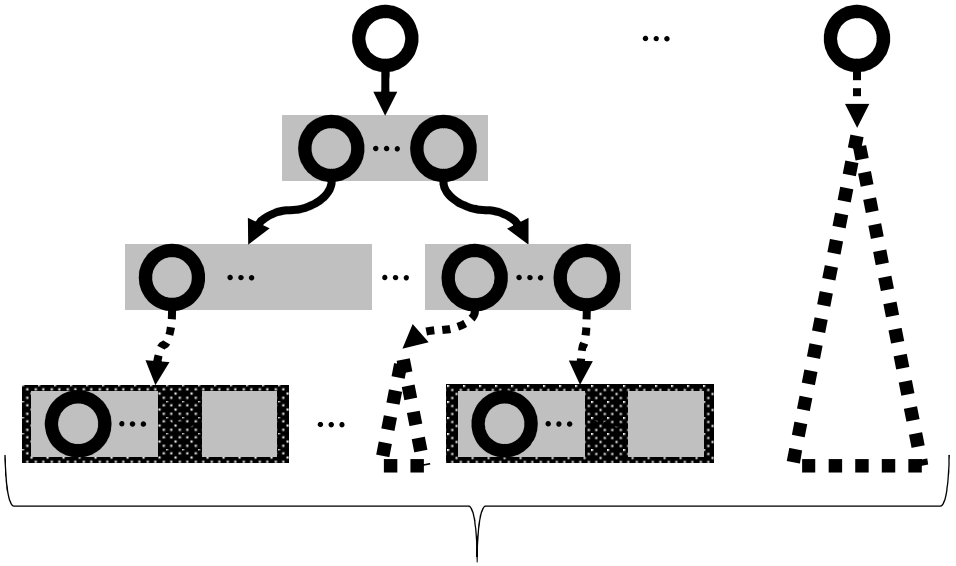
Coleman, RG et al. PLoS One. 2013; 8(10): e75992.

How DOCK 3.7 works

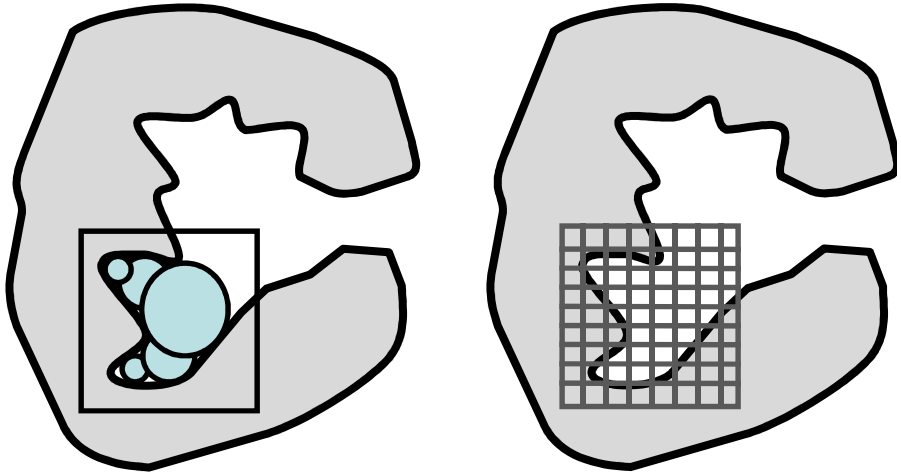
Preparation, Sampling, and Scoring

Ligand Sampling –
database construction





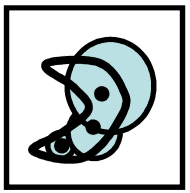
Dockable database file

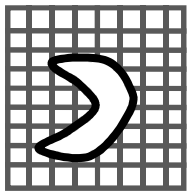


DOCKING

orient

score



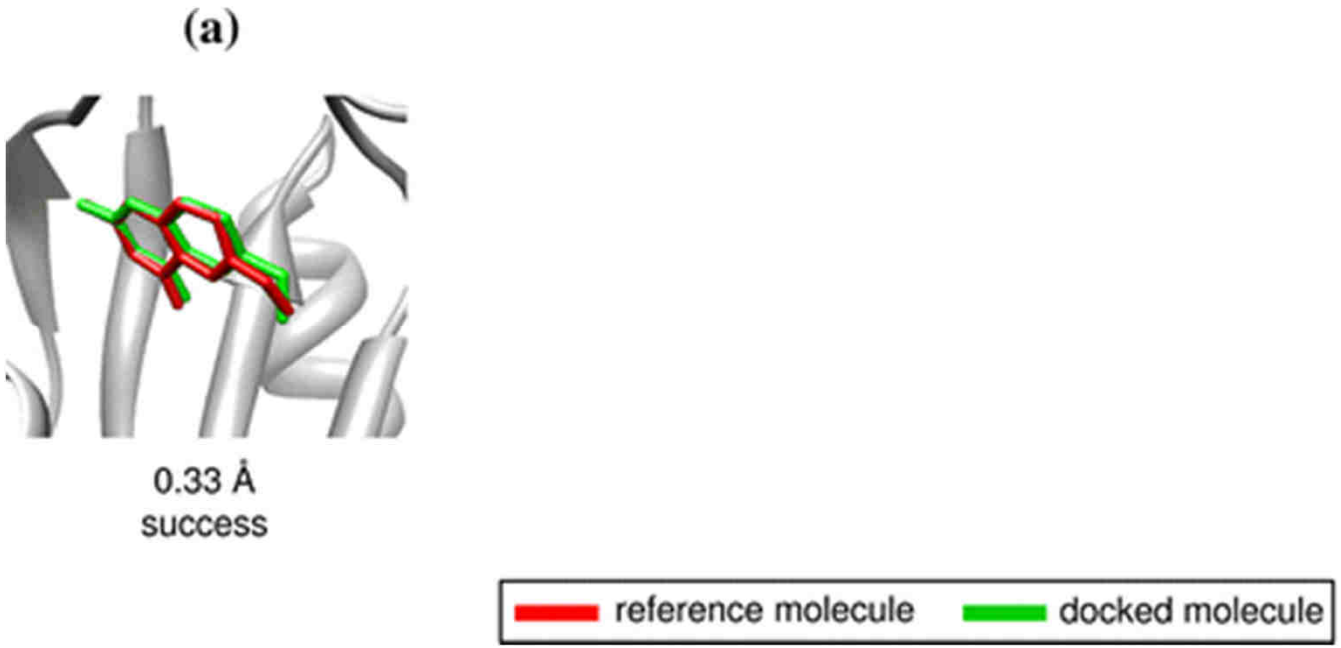


Scoring using a grid to speed up the calculations

How to evaluate docking methods

- Pose reproduction, reproduce the crystallographic poses
- Enrichment calculations, make sure ligand found in the top of the rank orders lists.
- Prospective testing on model cavities, make a predication, and test it!

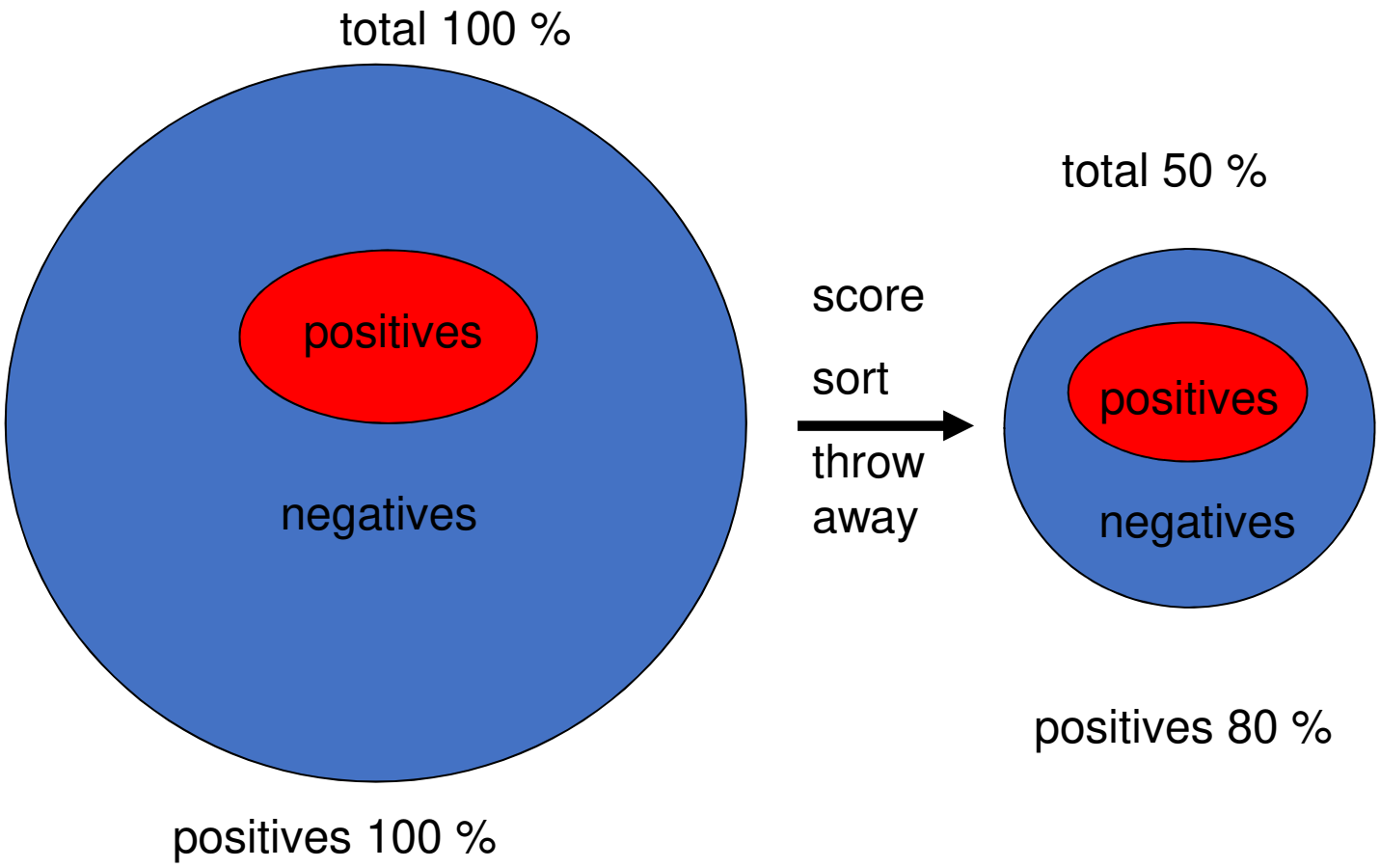
Pose Reproduction: RMSD Calculations



We can correct for molecular symmetry using the Hungarian algorithm

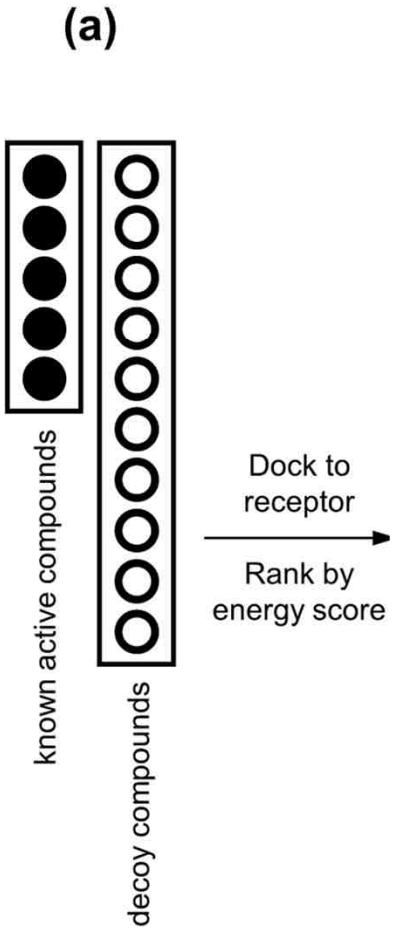
It is important that we are obtaining the correct binding mode.
Right for the right reasons

Enrichment Background

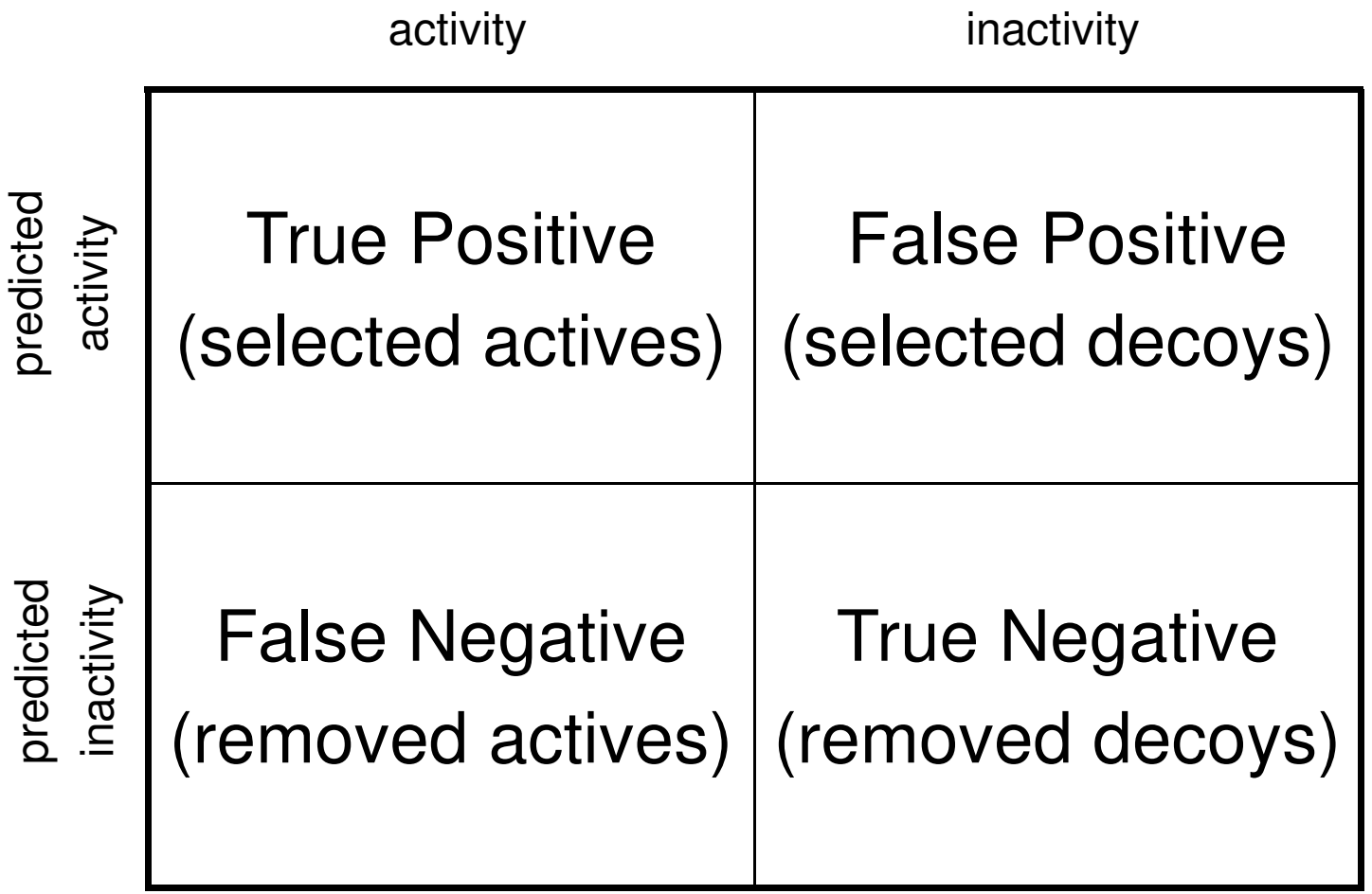


J. Med. Chem., 2006, 49 (23), pp 6789–6801

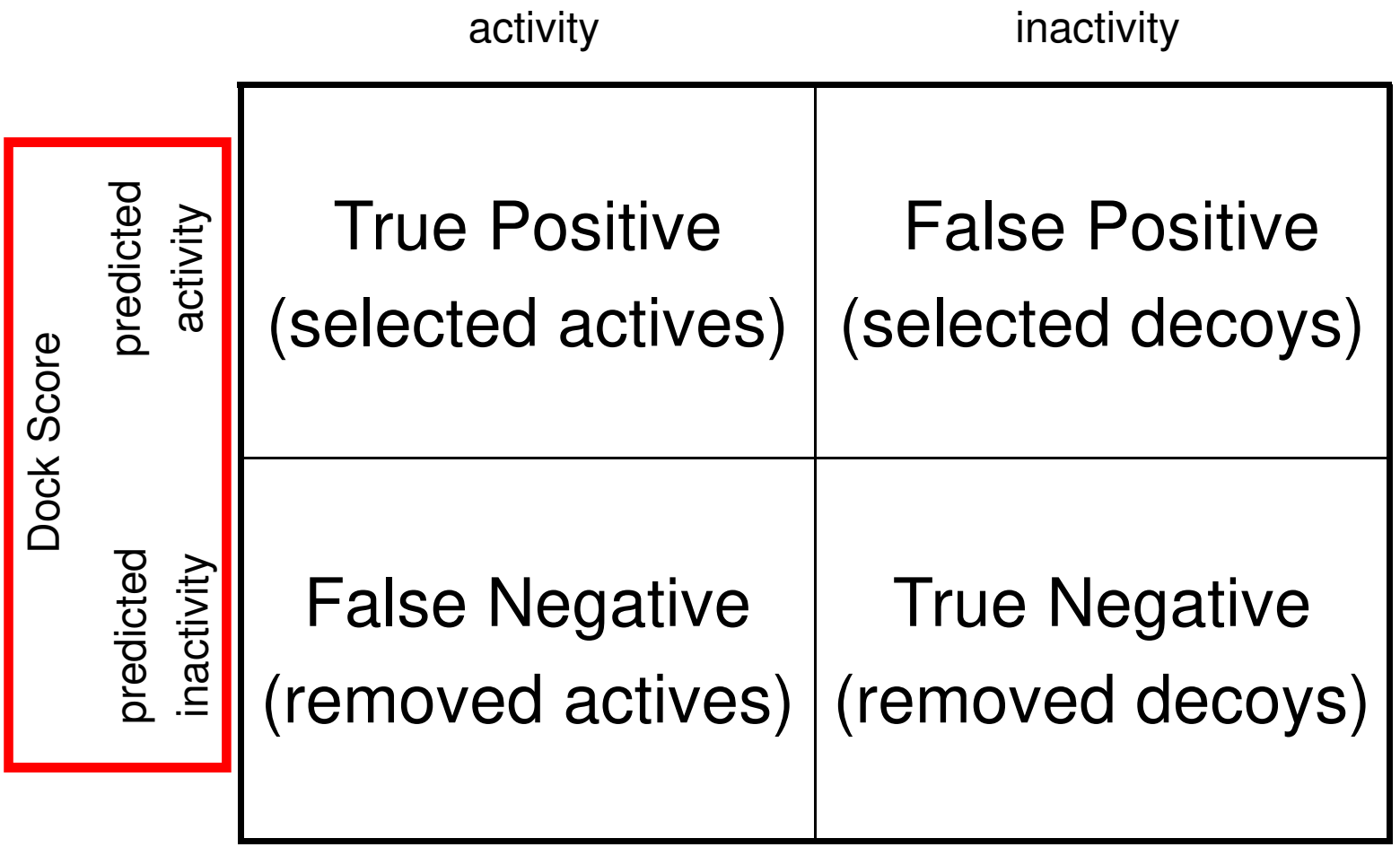
Retrospective testing
Enrich knowns over Decoys



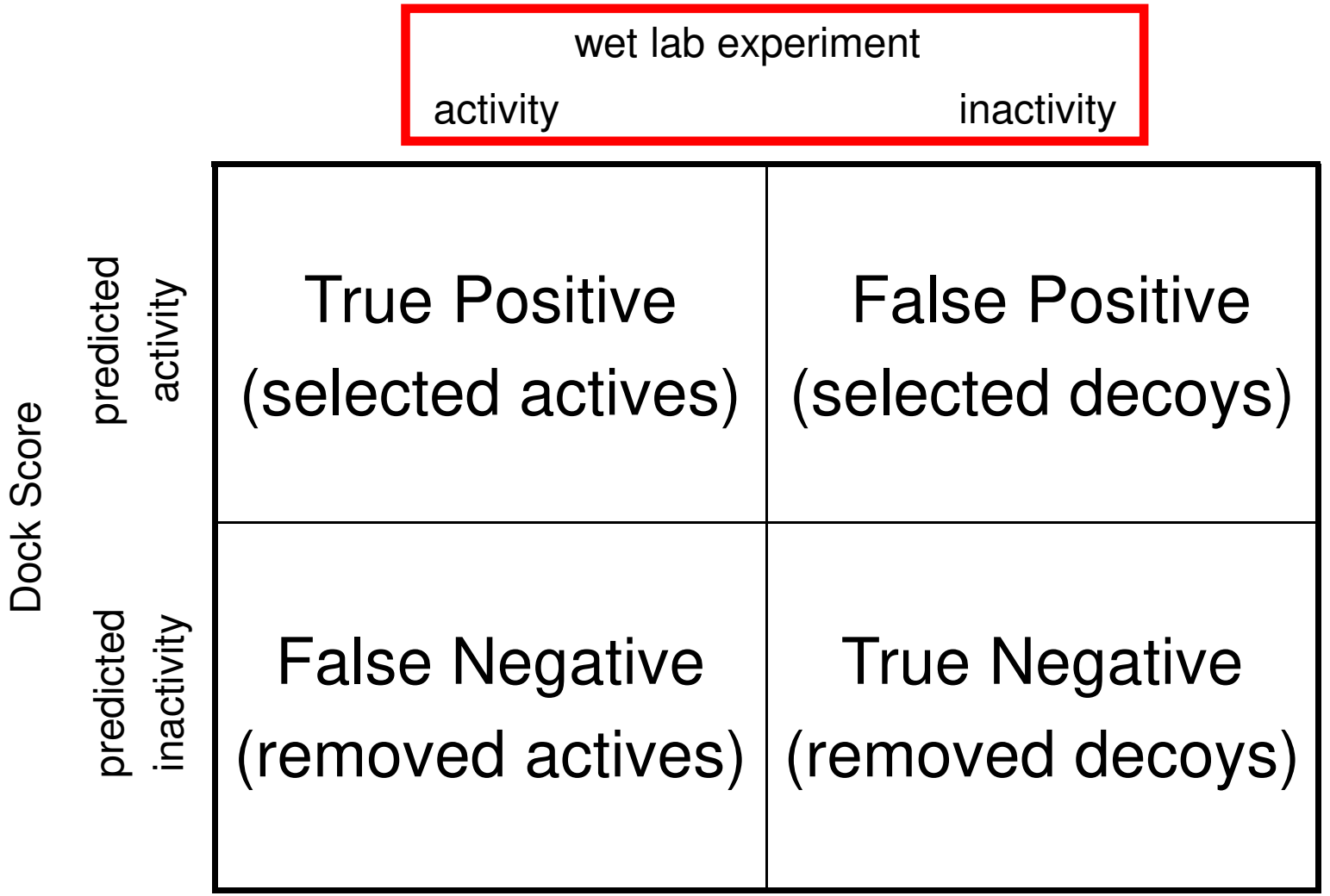
Computational Prediction vs. Experimental Evidenced



Computational Prediction vs. Experimental Evidenced



Computational Prediction vs. Experimental Evidenced

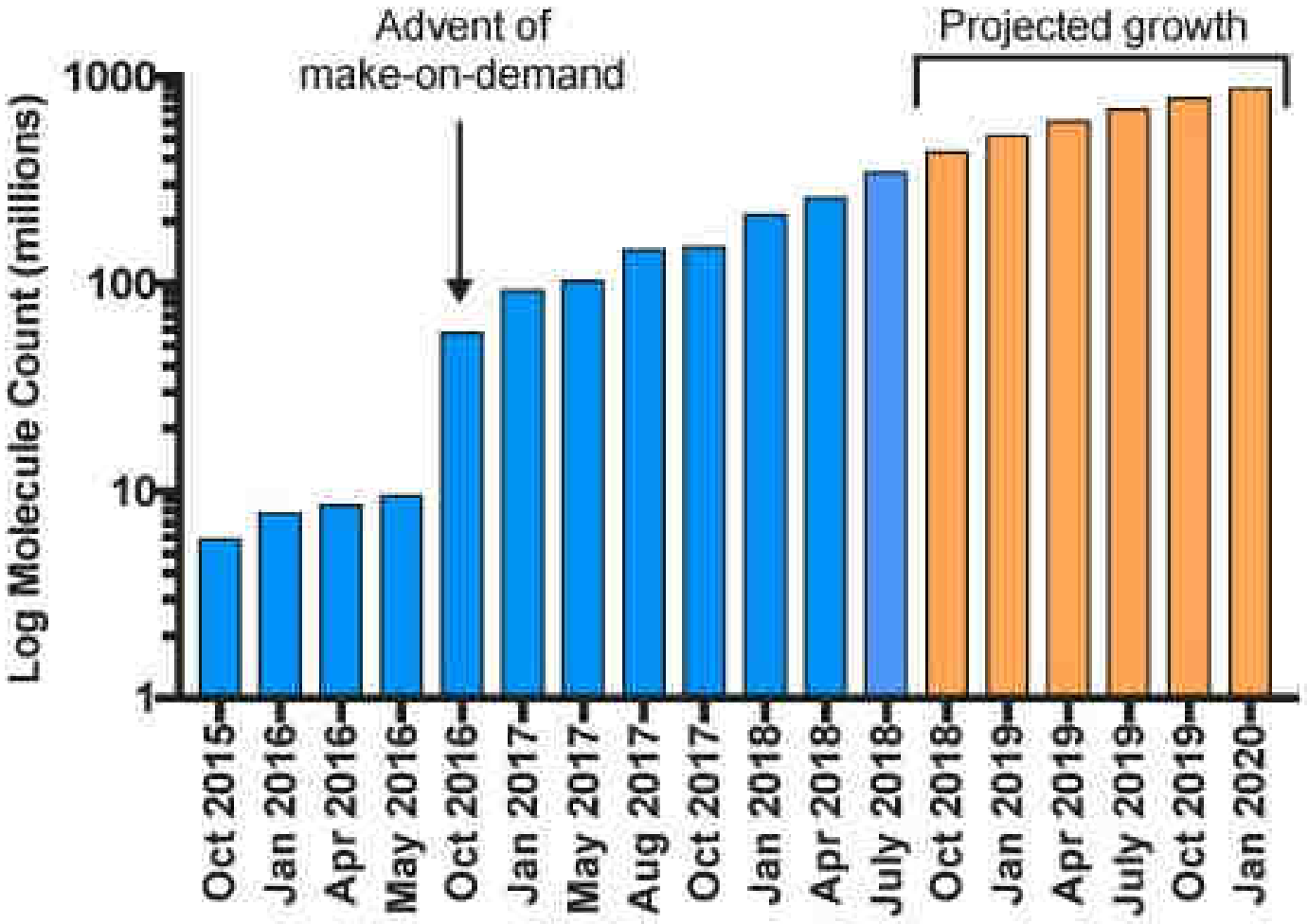


Outline

- The binding event using a thermodynamic cycle
- Introduction to molecular mechanics
- Introduction to molecular dynamics
- Introduction to molecular docking
- Applications from my work
    - Large-scale docking
    - Receptor flexibility
    - Receptor desolvation

Large-scale docking:
Screen all of purchasable space

Irwin's law: Docking libraries, crucial for chemical discovery, doubling every 2.5 years

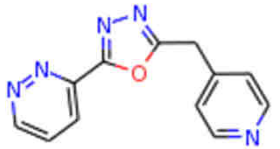


250 Million (now) -> 1B (next year) -> 10 Billion (next few years)!
How can we keep up in docking?

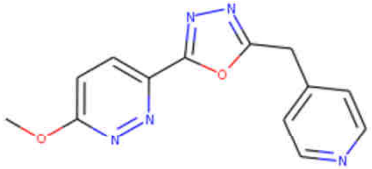
John Irwin

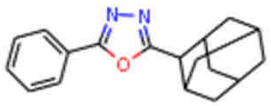
Database is larger and more diverse

The relationship between number of clusters and the cluster size (Bemis-Murcko scaffold)

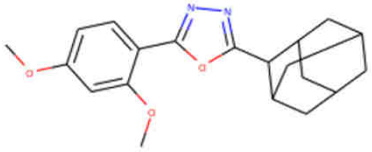


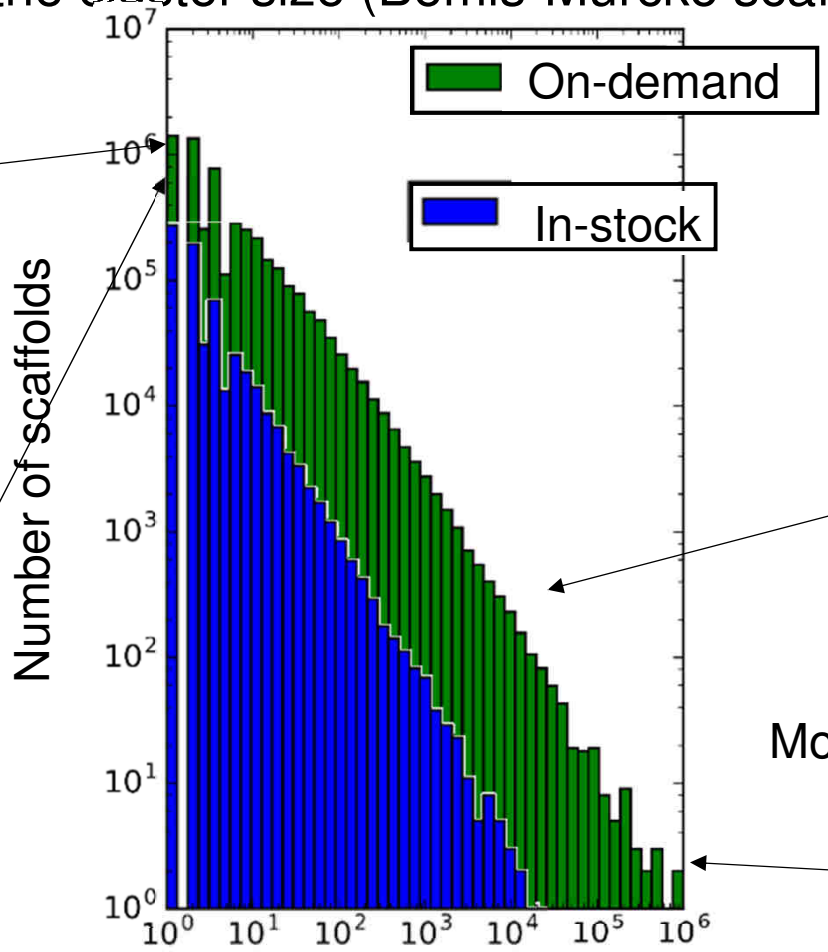
1 molecule

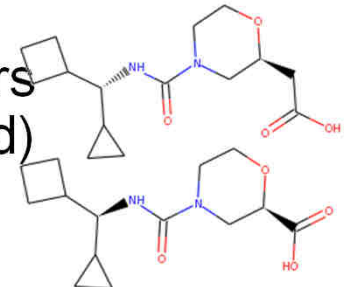




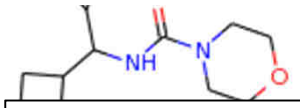
1 molecule







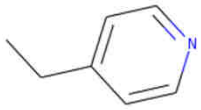
In-stock →
On-demand
# of scaffolds 16x
# of cpds 68x



Most

Screening this size diverse database is only possible though computational methods

2.3M molecules

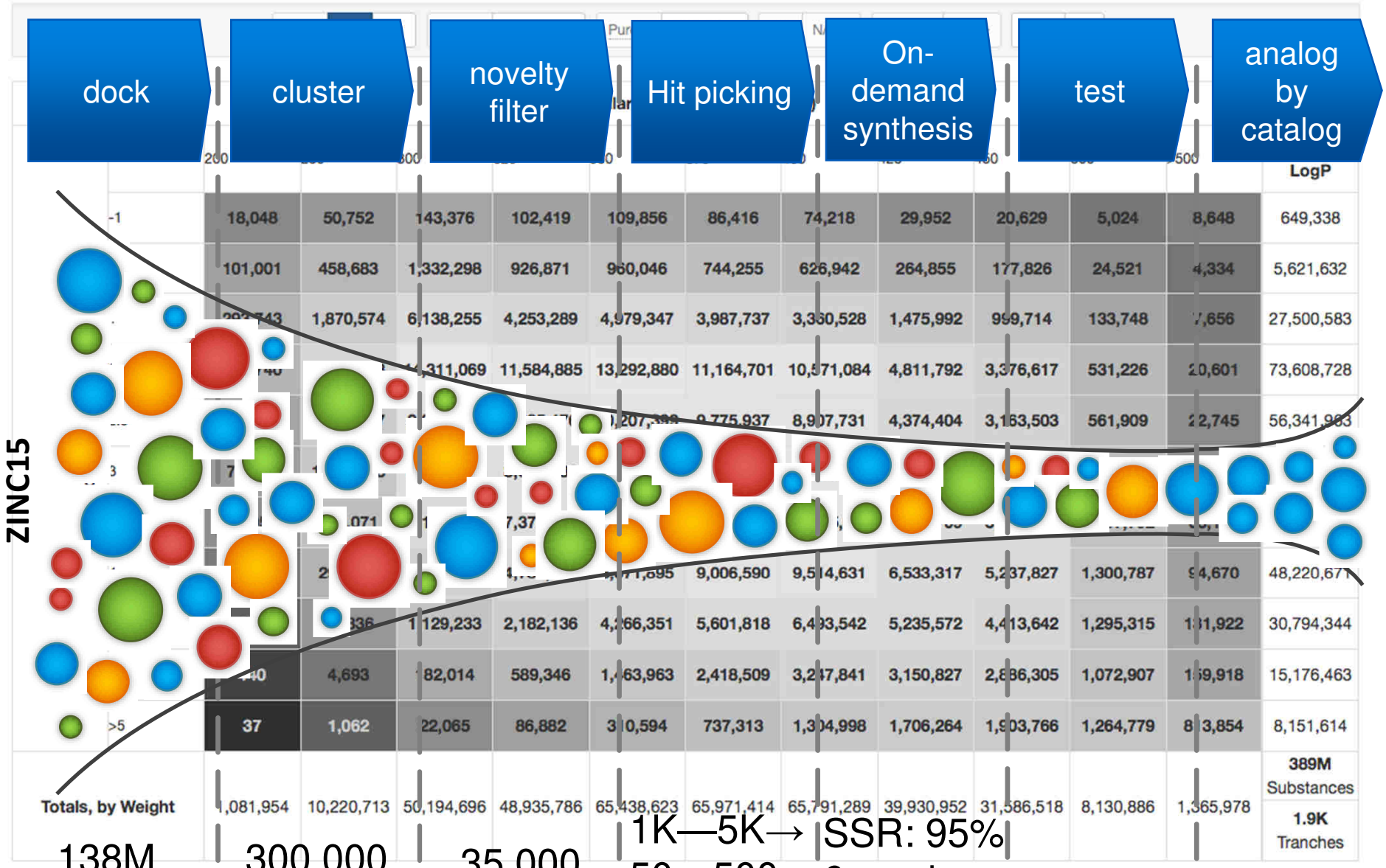


Jiankun Lyu & John Irwin

Populating new chemical space

Expanding around existing space

Large scale docking flow



Jiankun Lyu

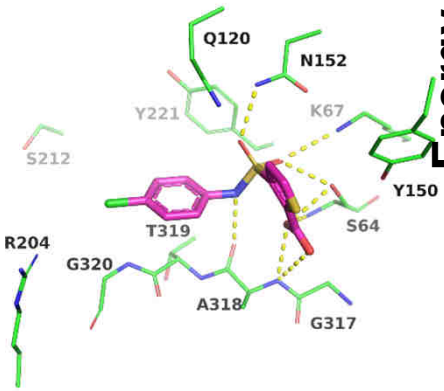
AmpC β-lactamase

We docked 100M molecules to **AmpC** β-lactamase

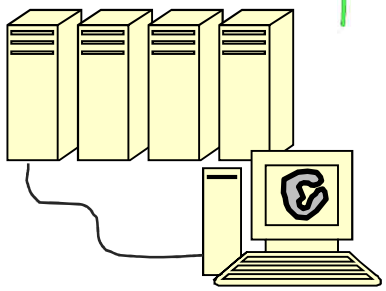
ZINC

100 million cmpds

β-lactamase



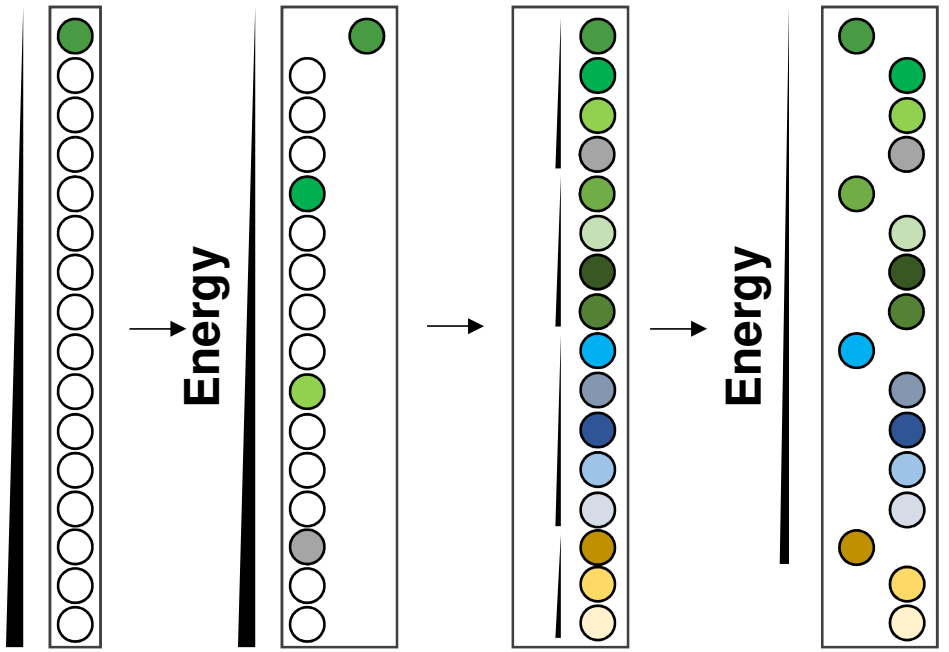
DOCK



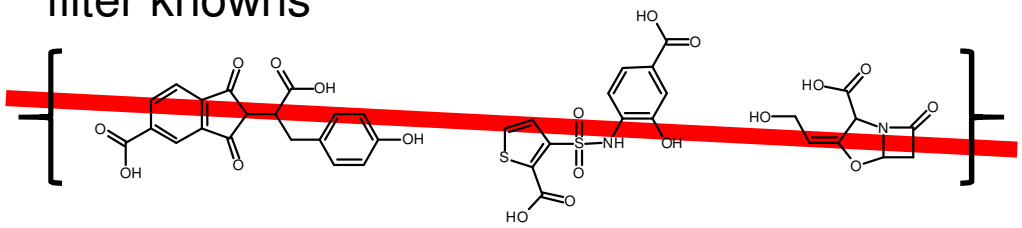
4.3x10^{13} complexes
4.1x10^{4} core hours
2 cluster days

filter and clusters

Best clusters



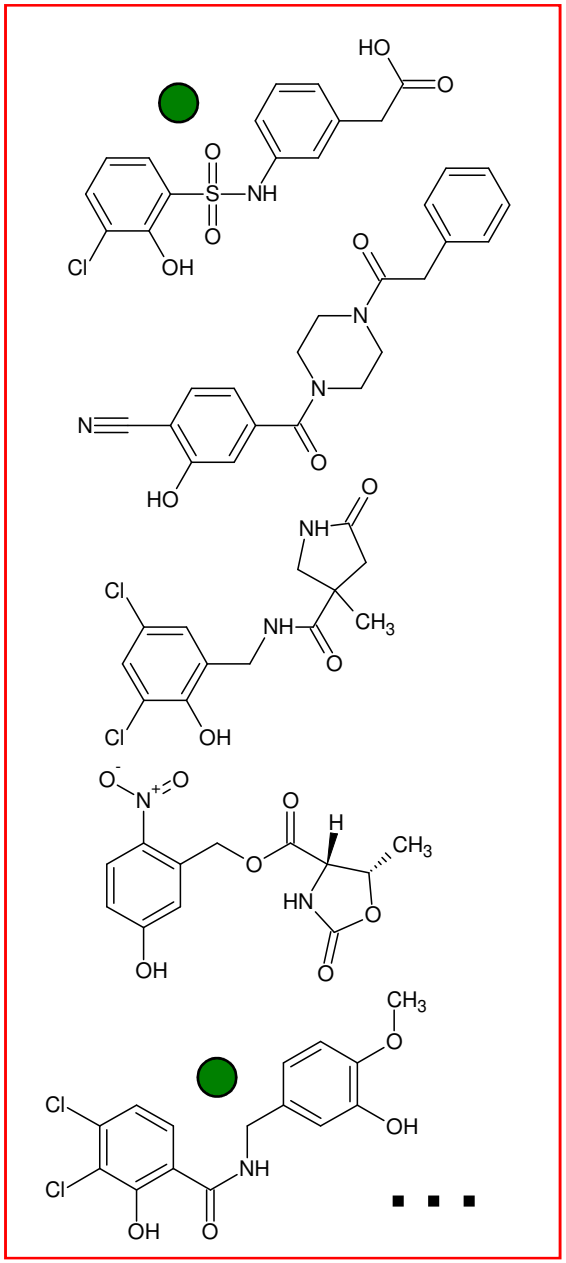
filter knowns

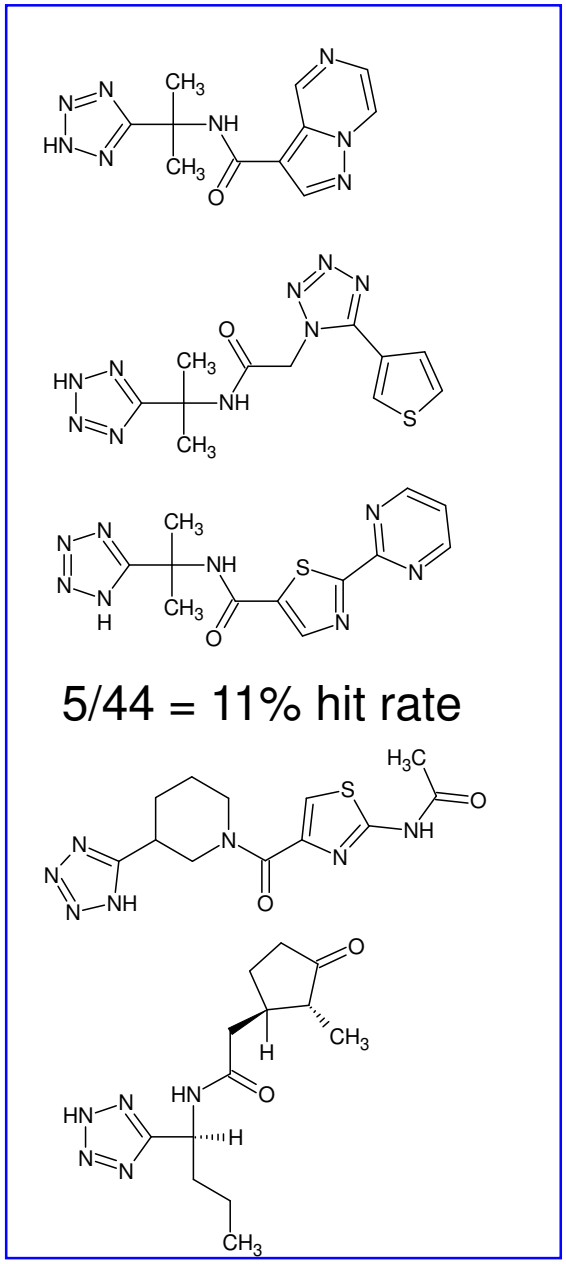


Buy 44 from top clusters

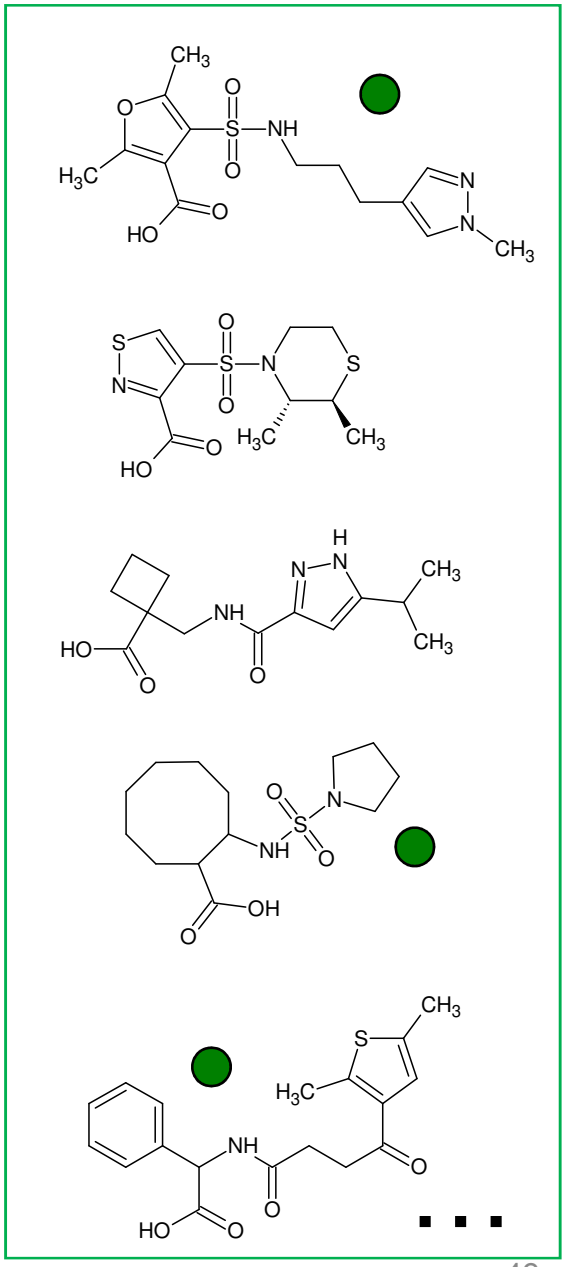
Jiankun Lyu & Anat Levit

We bought new chemotypes





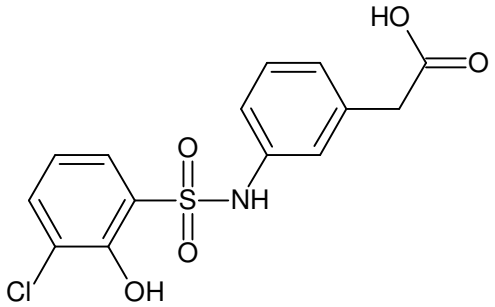
5/44 = 11% hit rate



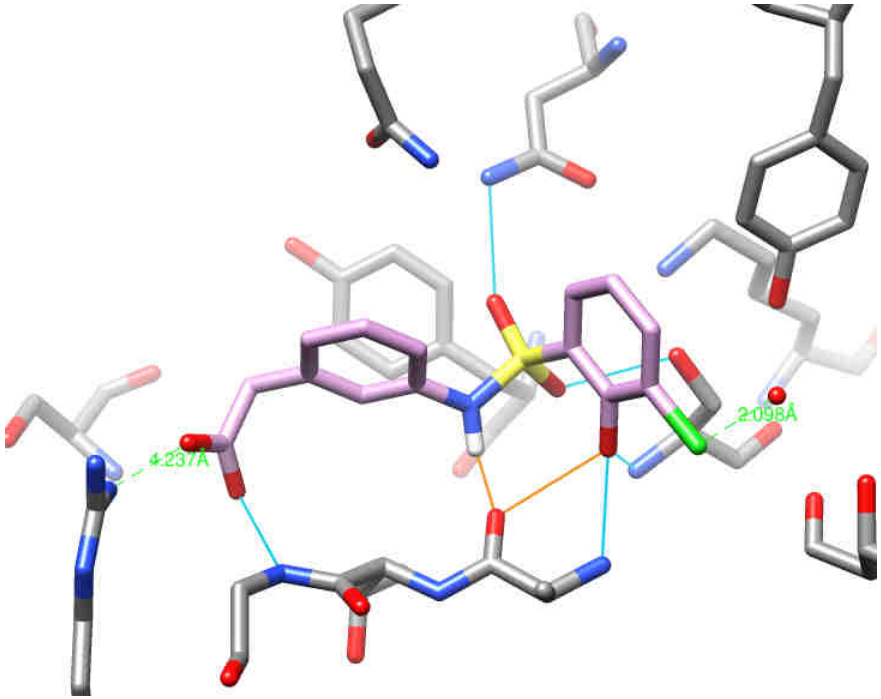
Isha Singh & Jiankun Lyu

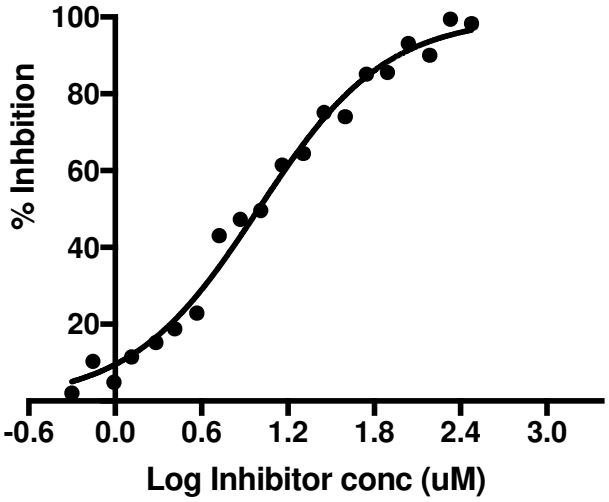
Best AmpC inhibitor found right out of docking (1 µM)

Compound 10 ZINC000339204163

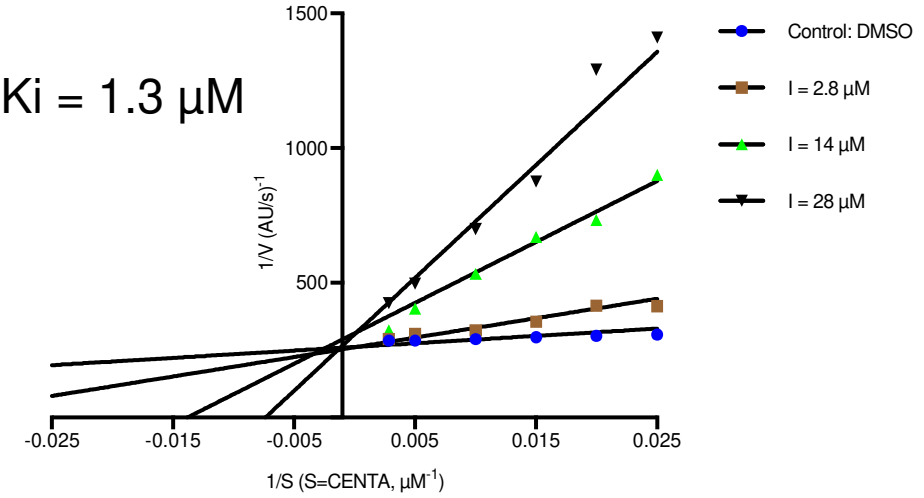


Best AmpC inhibitor found straight out of docking or any screening method





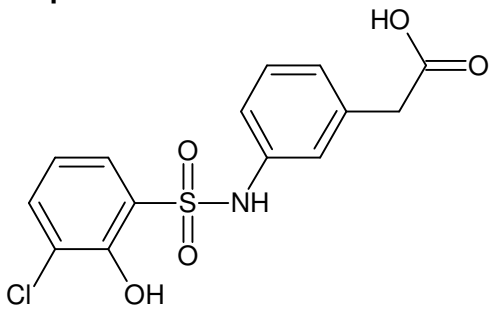
Ki = 1.3 µM



Isha Singh

Crystal structure confirms docked pose and phenolate

Cmpd 10

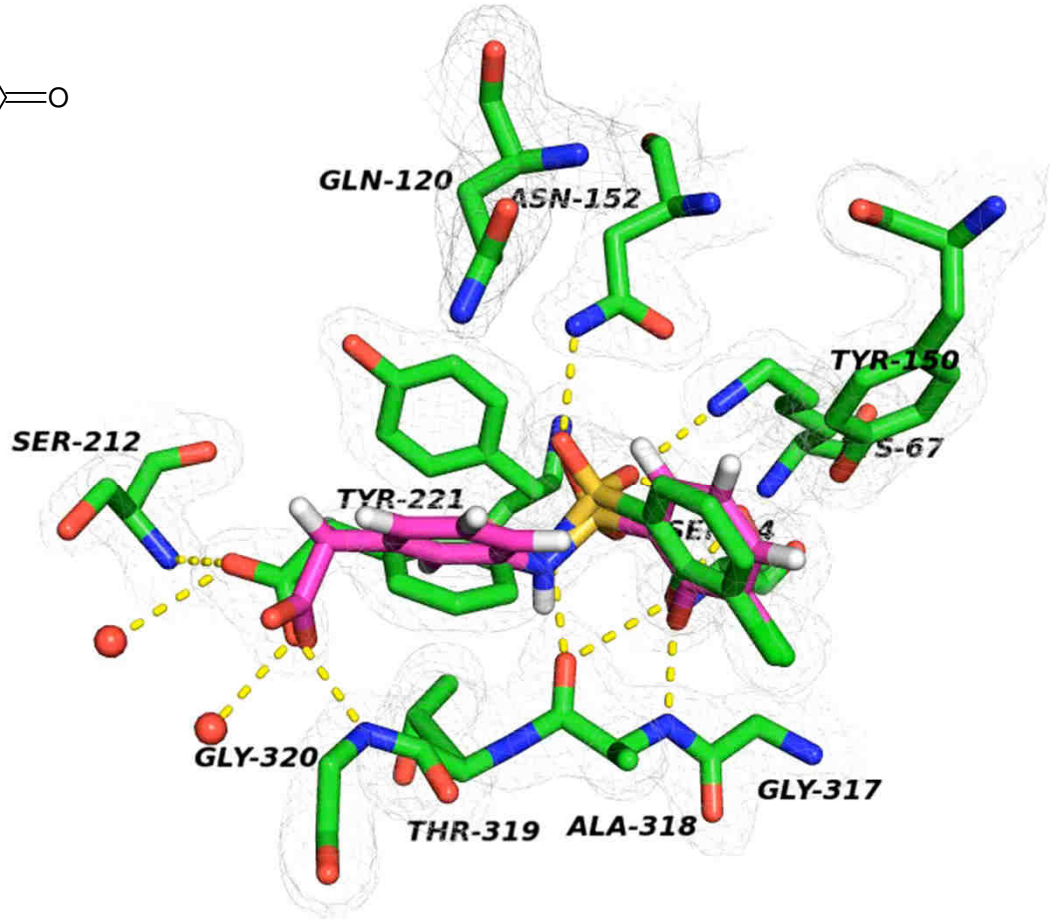


Ki=1.3 µM

rmsd=0.98Å

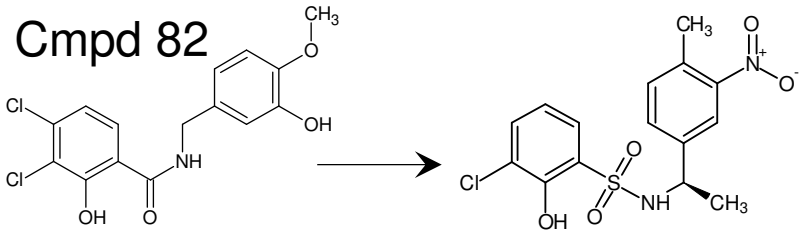
Resolution=1.9 Å

pKa of phenol is 5.8
Determined on a
carboxyl-less analog



5 hits (out of 45 molecules test) and their analogs

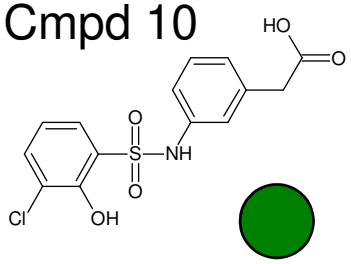
Cmpd 82



Ki = 210 µM

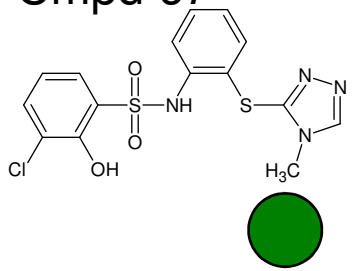
Ki = 100 µM

Cmpd 10



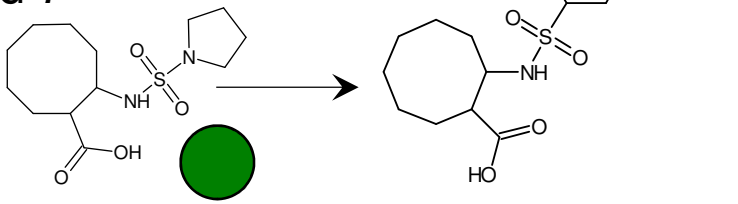
Ki = 1.3 µM

Cmpd 67



Ki = 77 nM

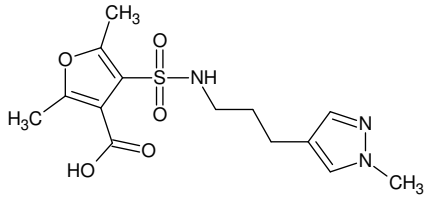
Cmpd 7



Ki = 260µM

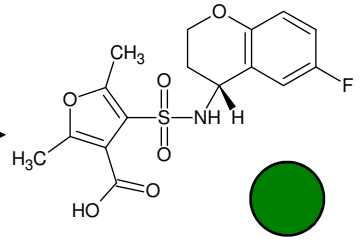
Ki = 80µM

Cmpd 5



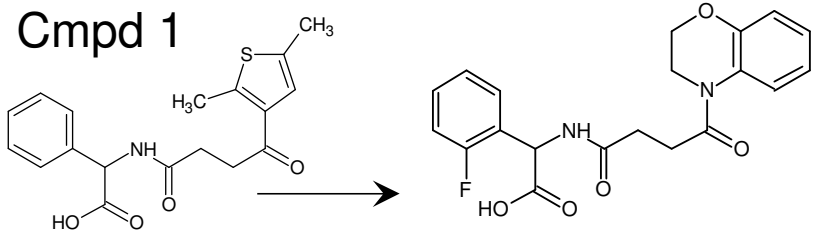
Ki = 190µM

Cmpd 44



Ki = 80 µM

Cmpd 1



Ki = 400µM

Ki =14µM

- 9/31 Cmpd1 Analog
- 8/15 Cmpd 5 Analogs
- 9/19 Cmpd 7 Analog
- 8/11 Cmpd 10 Analogs
- 8/14 Cmpd 82 Analogs

- Determined 4 crystal structures

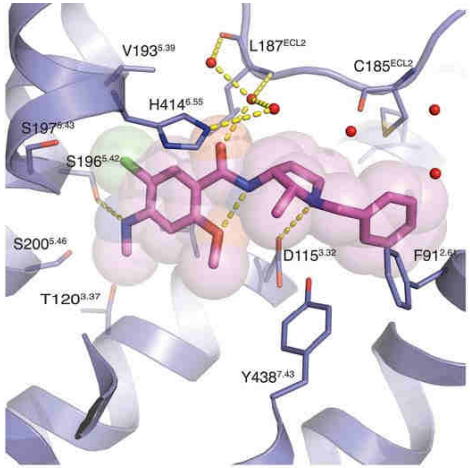
Isha Singh

D4 Dopamine receptor

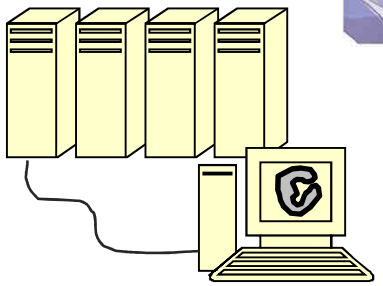
How many active ligands are in the library for **DRD4**?

ZINC

138 million cmpds

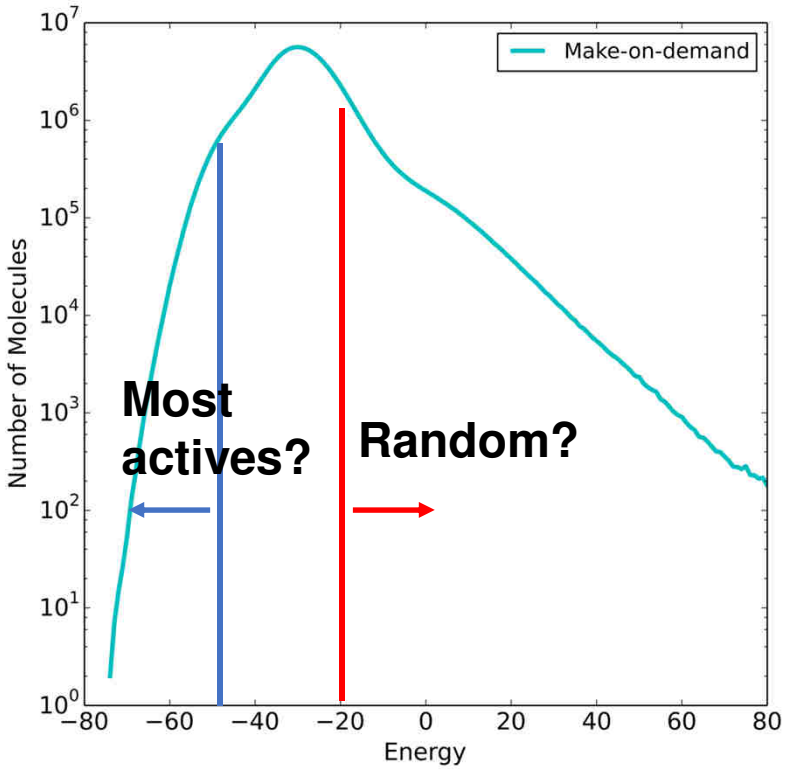


DOCK

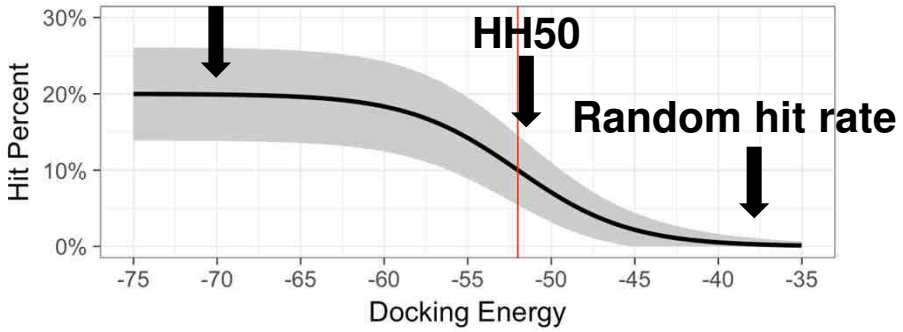


~~Test 50~~

Test 549



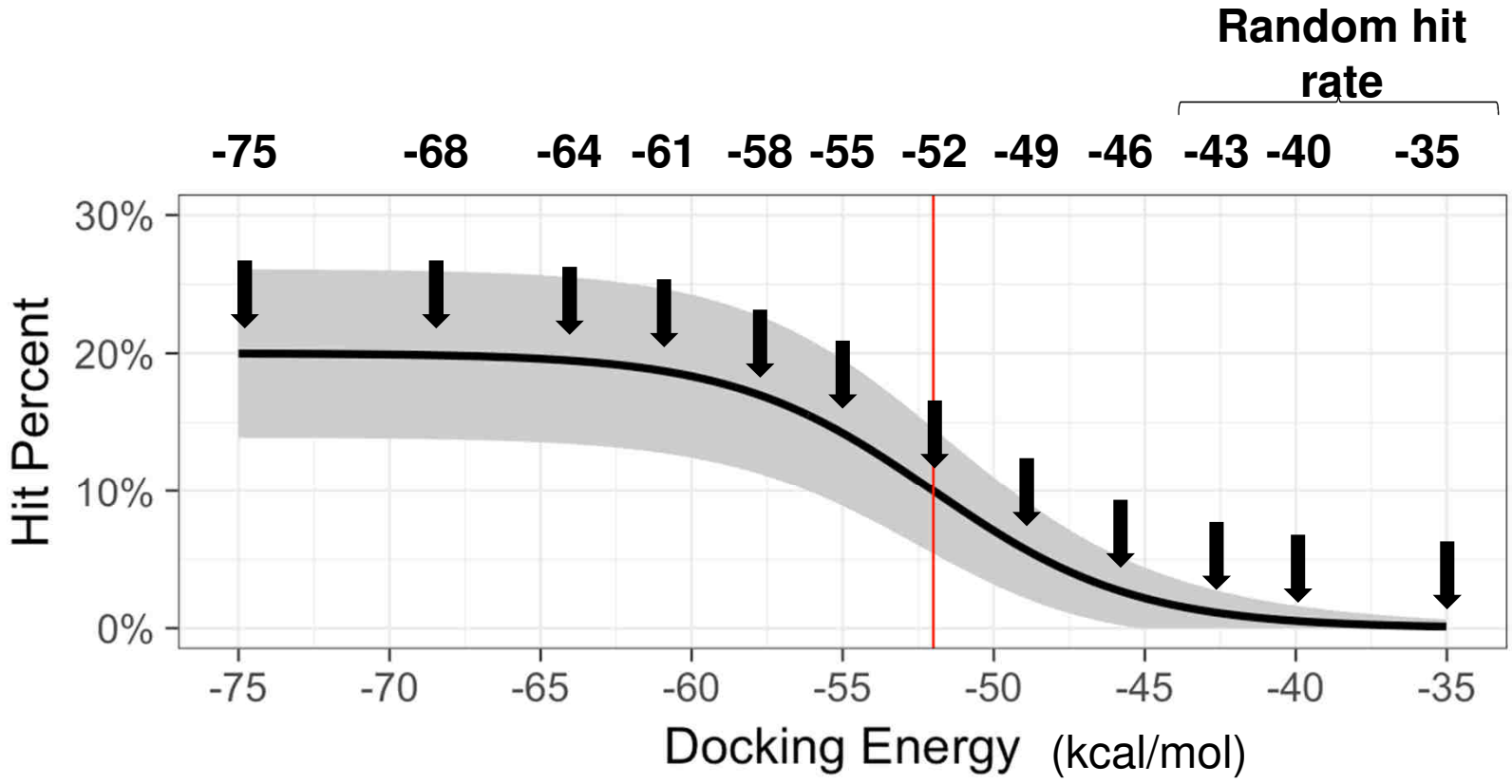
High hit rate



6.92 x10^{13} complexes
4.4x10^{4} core hours
1.8 cluster days

Jiankun Lyu, Matthew O'Meara

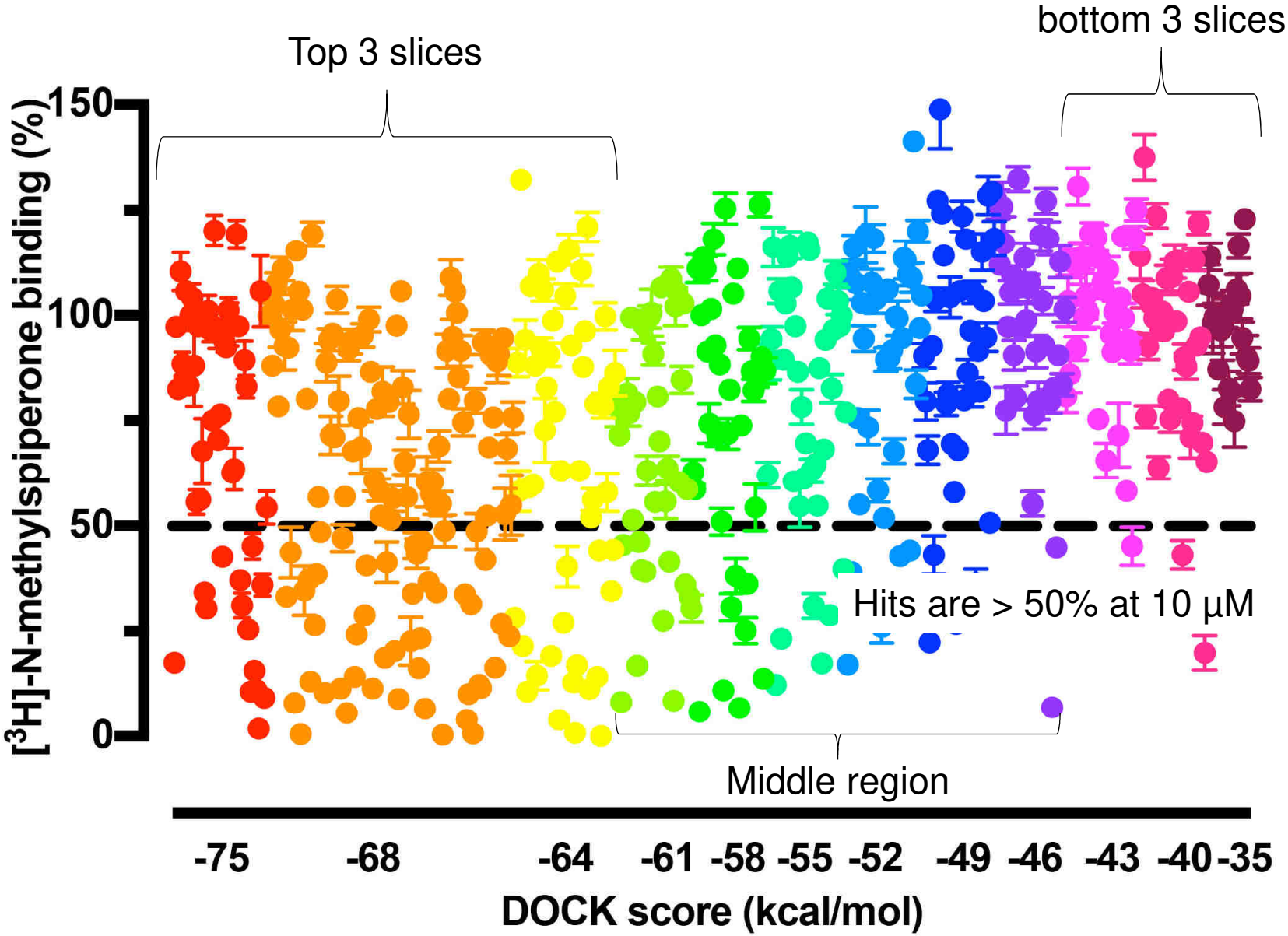
Bought 444 molecules to estimate the DOCKing hit rate curve for DRD4



444 molecules were picked automatically, with 35 to 40 molecules sampled at 12 energy windows from docking scores from -75 to -35 kcal/mol.

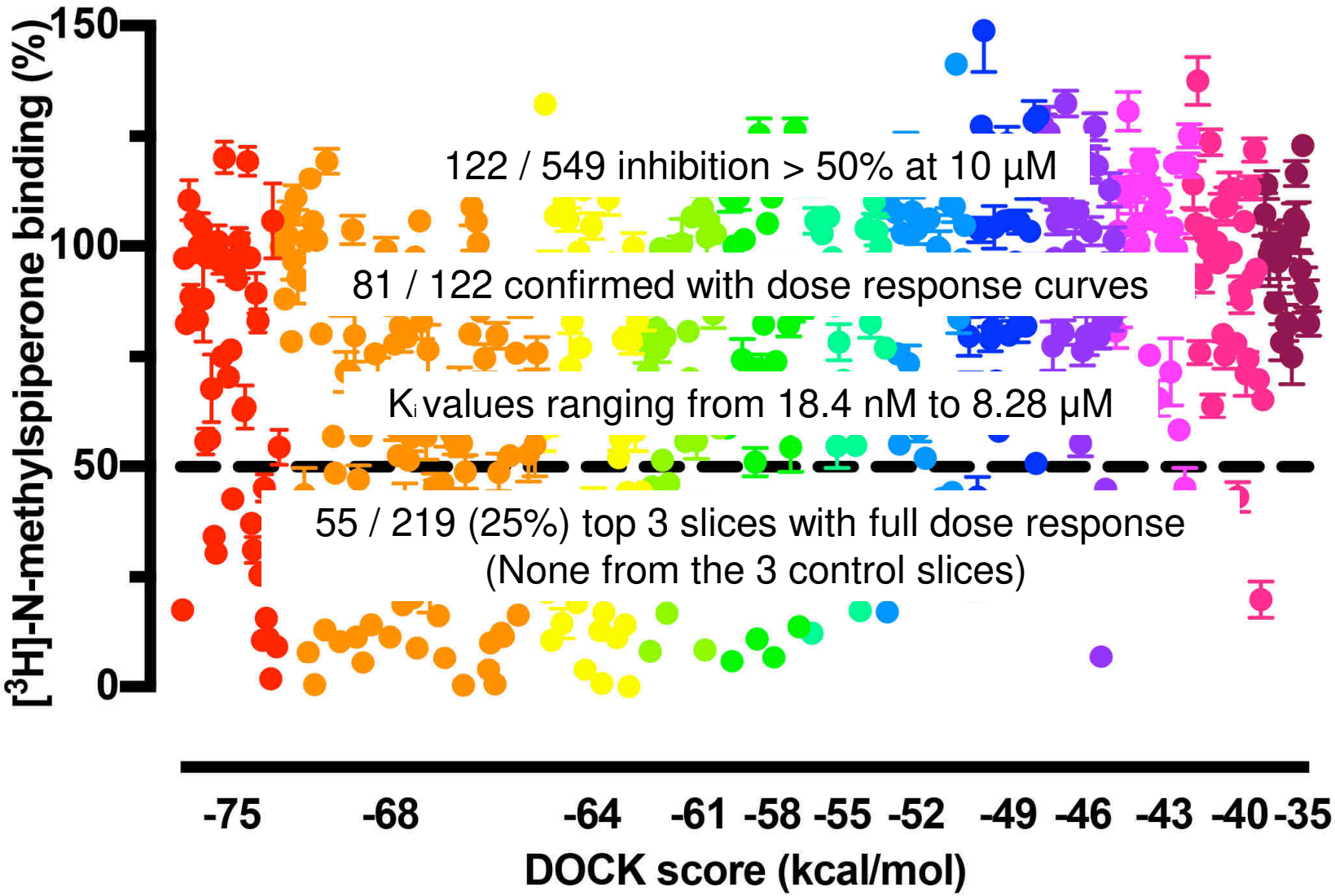
Matthew O'Meara

Great hit rate of 25% at top, poor hit rate of 0% on right



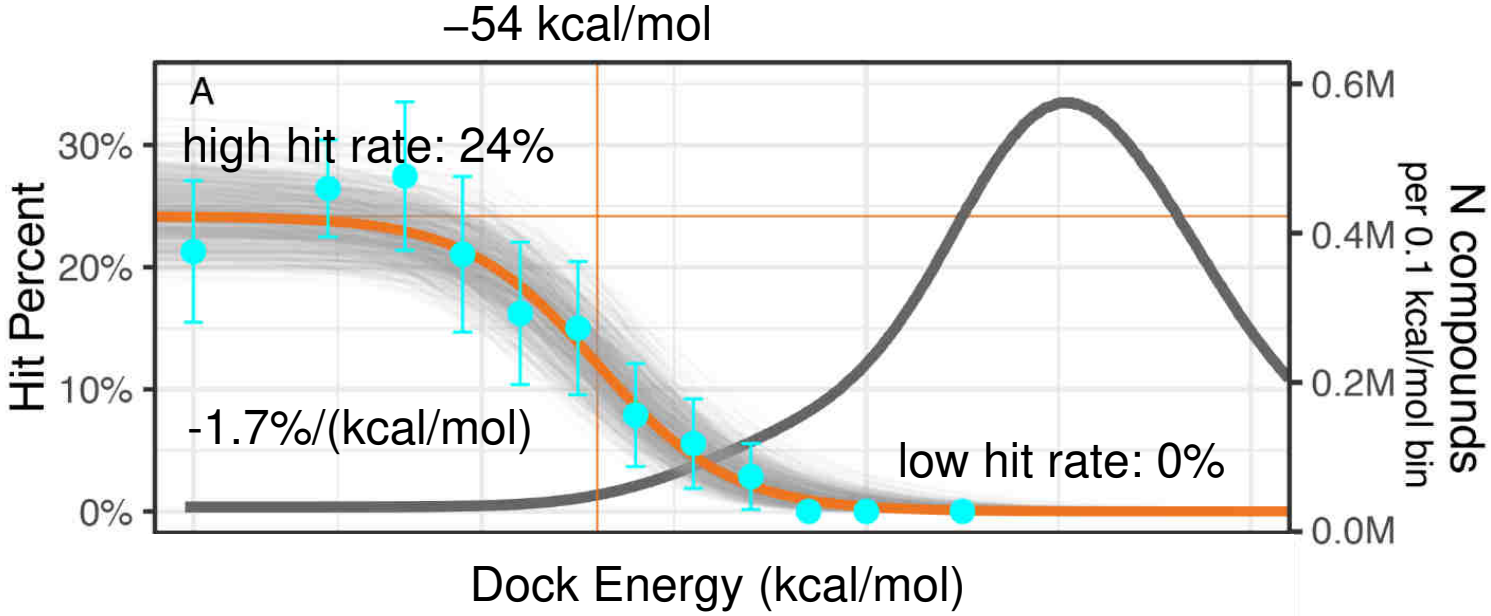
Sheng Wang, Jiankun Lyu

Great hit rate of 25% at top, poor hit rate of 0% on right



Sheng Wang, Jiankun Lyu

Among the 138 million molecule library there are calculated to be over 481,000 D4 active molecules

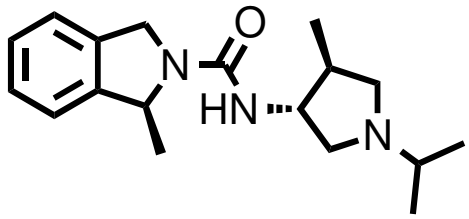


hit-rates fell almost monotonically with score

Matthew O'Meara

The 95% confidence: [209K,1,020K]

180 pM Gi-biased, selective, full agonist, among the most potent sub-type selective agonists known for this receptor



ZINC621433144

$K_{i,DRD4}$ = 4.32 nM

$K_{i,DRD2}$ > 10,000 nM

$K_{i,DRD3}$ > 10,000 nM

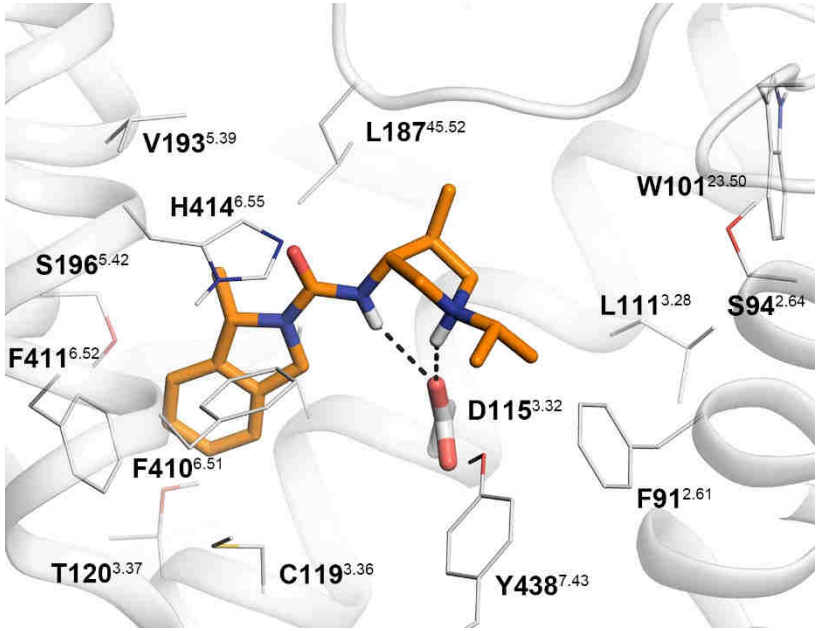
cAMP EC_{50} = 0.18 nM

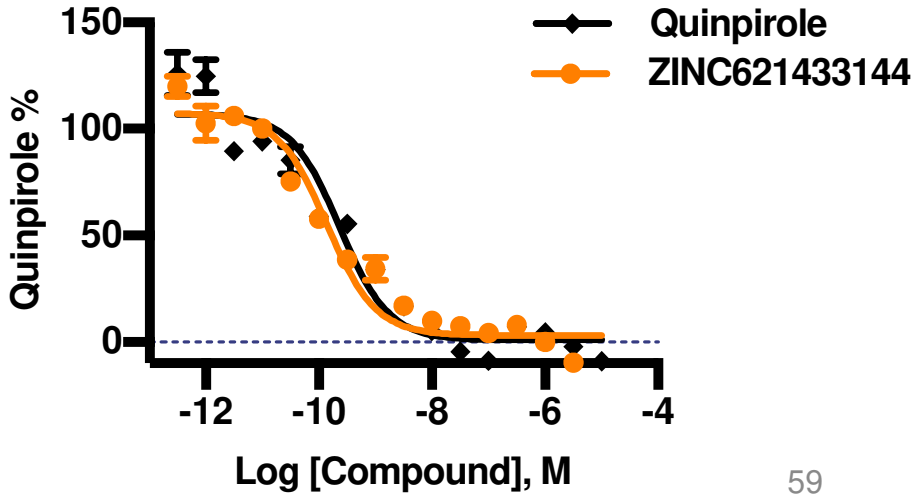
Tango EC_{50} = 57.3 nM

G_i BRET EC_{50} = 0.56 nM

Arrestin BRET EC_{50} = 2.3 nM

Bias factor = 17 to G protein



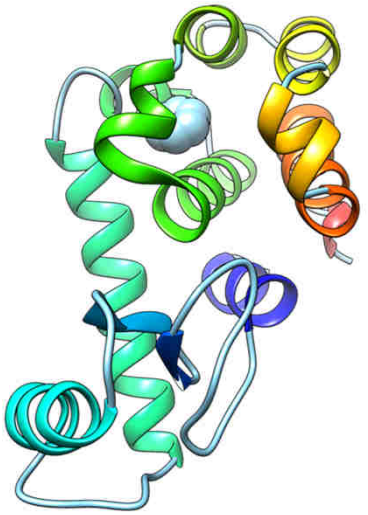


Tao Che, Bryan Roth

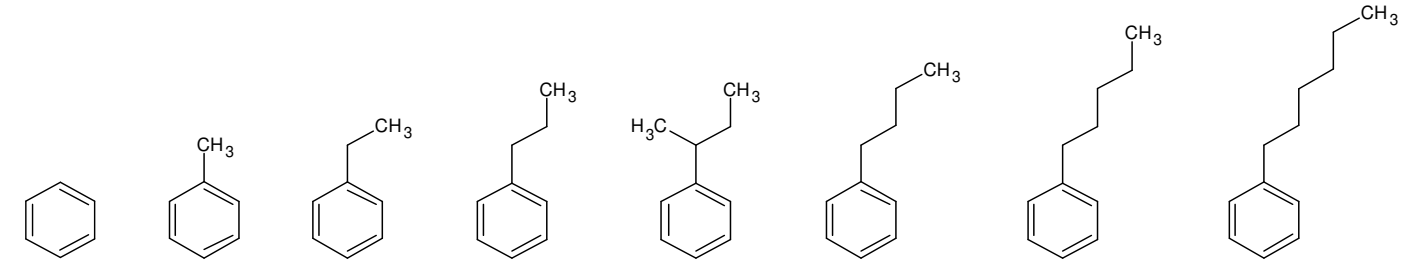
Outline

- The binding event using a thermodynamic cycle
- Introduction to molecular mechanics
- Introduction to molecular dynamics
- Introduction to molecular docking
- Applications from my work
  - Large-scale docking
  - Receptor flexibility
  - Receptor desolvation

T4 lysozyme L99A



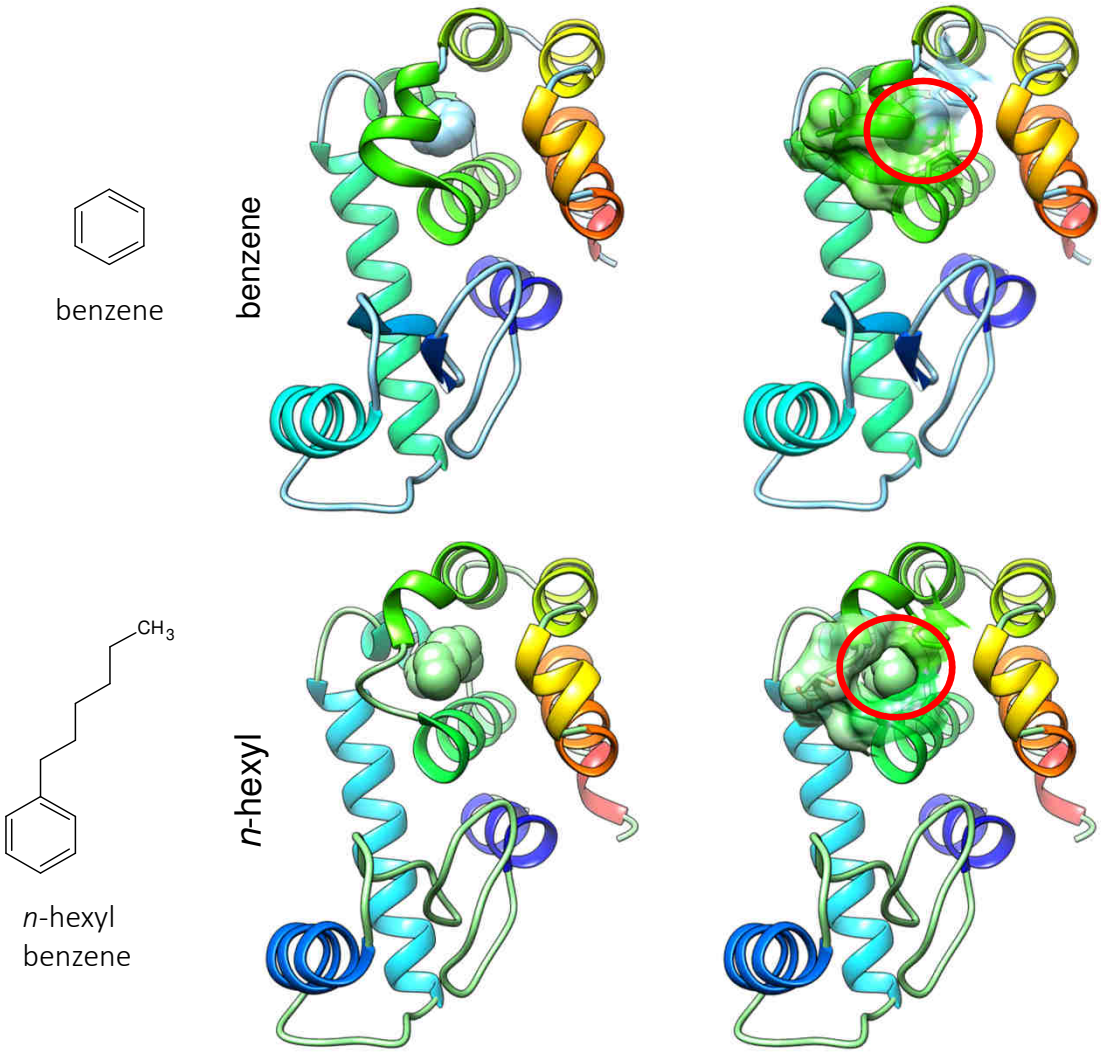
How does T4-lysozyme bind a congeneric ligand series



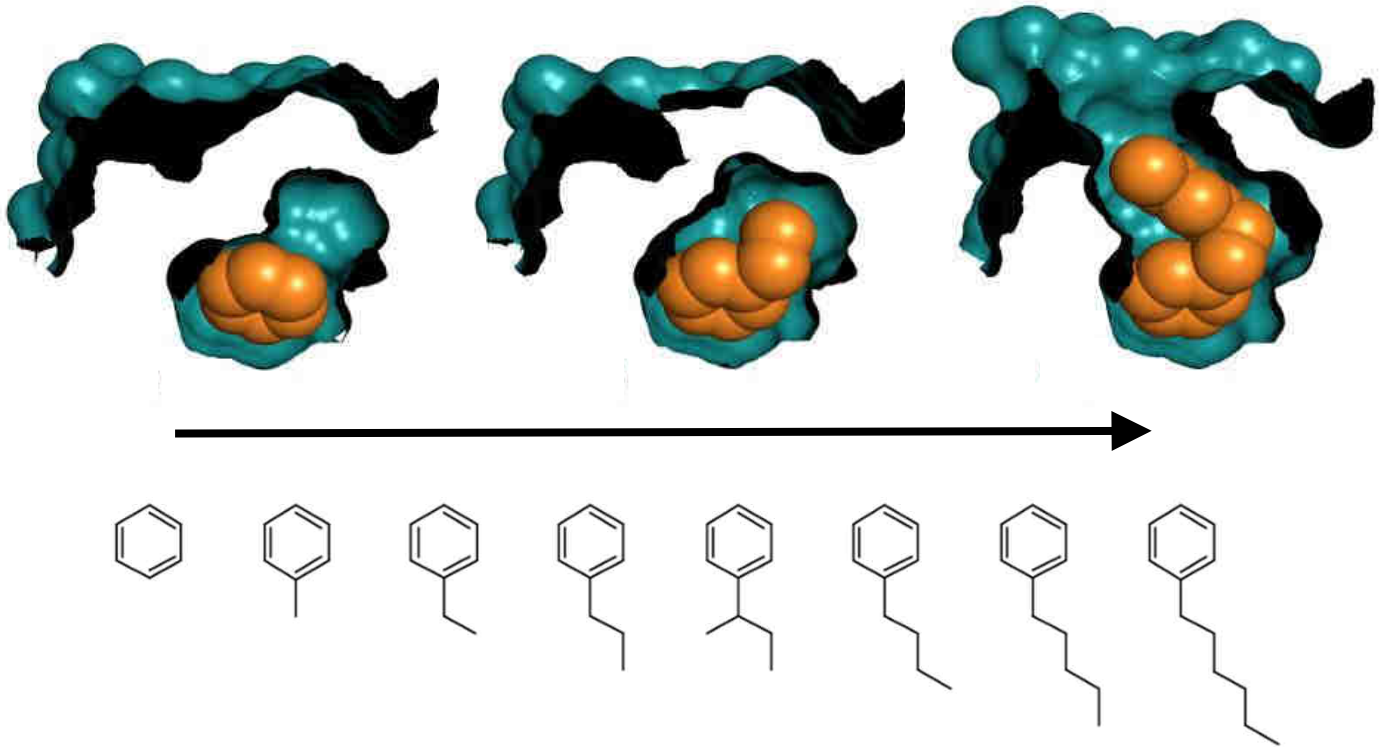
benzene toluene ethyl benzene *n*-propyl benzene *sec*-butyl benzene *n*-butyl benzene *n*-pentyl benzene *n*-hexyl benzene

Merski, Fischer, Balius et al 2015 PNAS 112(16):5039-44
Morton, et al. 1995 Biochemistry 34(27):8564–8575
Morton & Matthews 1995 Biochemistry 34(27):8576–8588

T4 lysozyme L99A opens



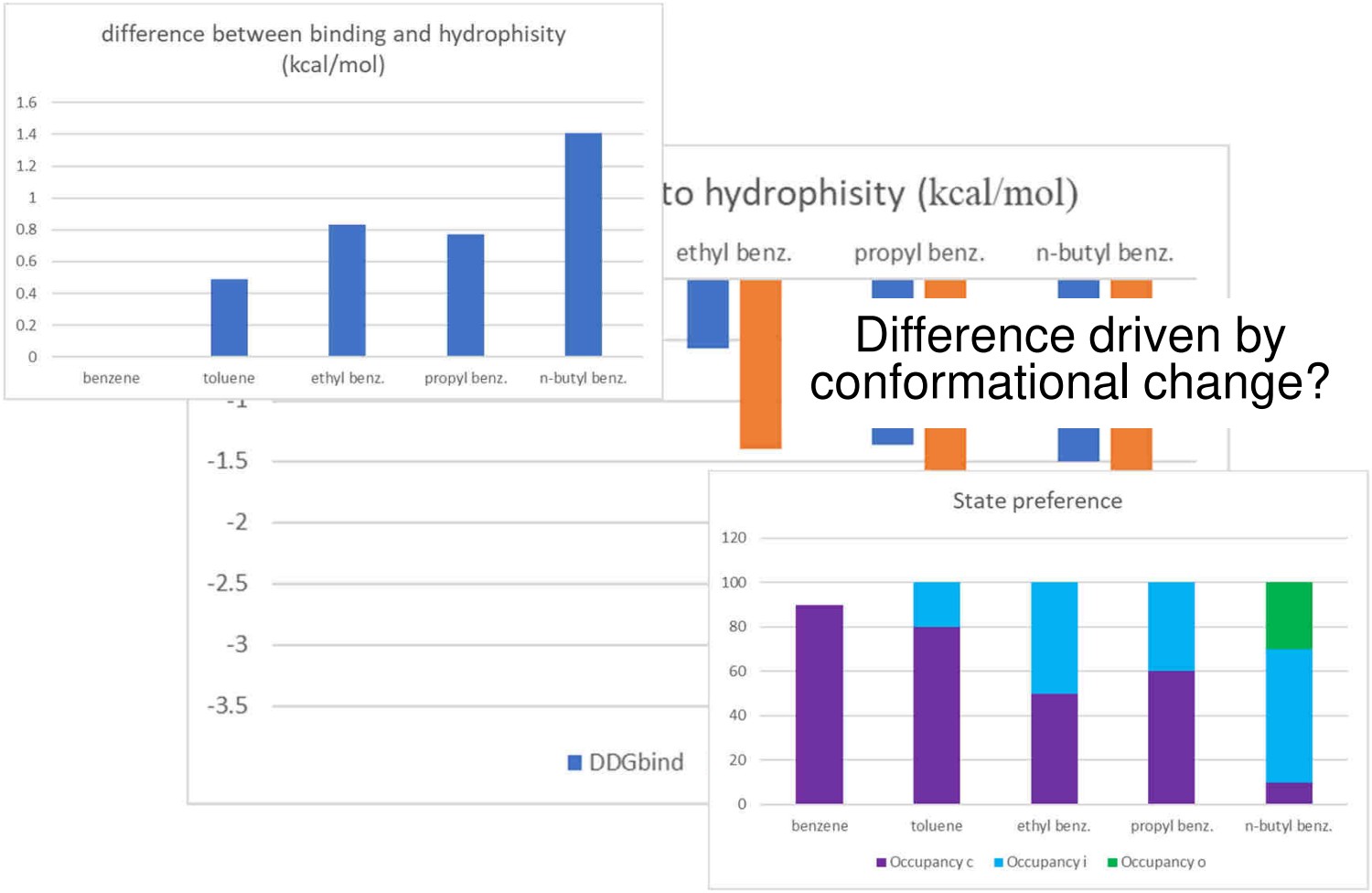
Probing receptor conformational change induce by ligand series



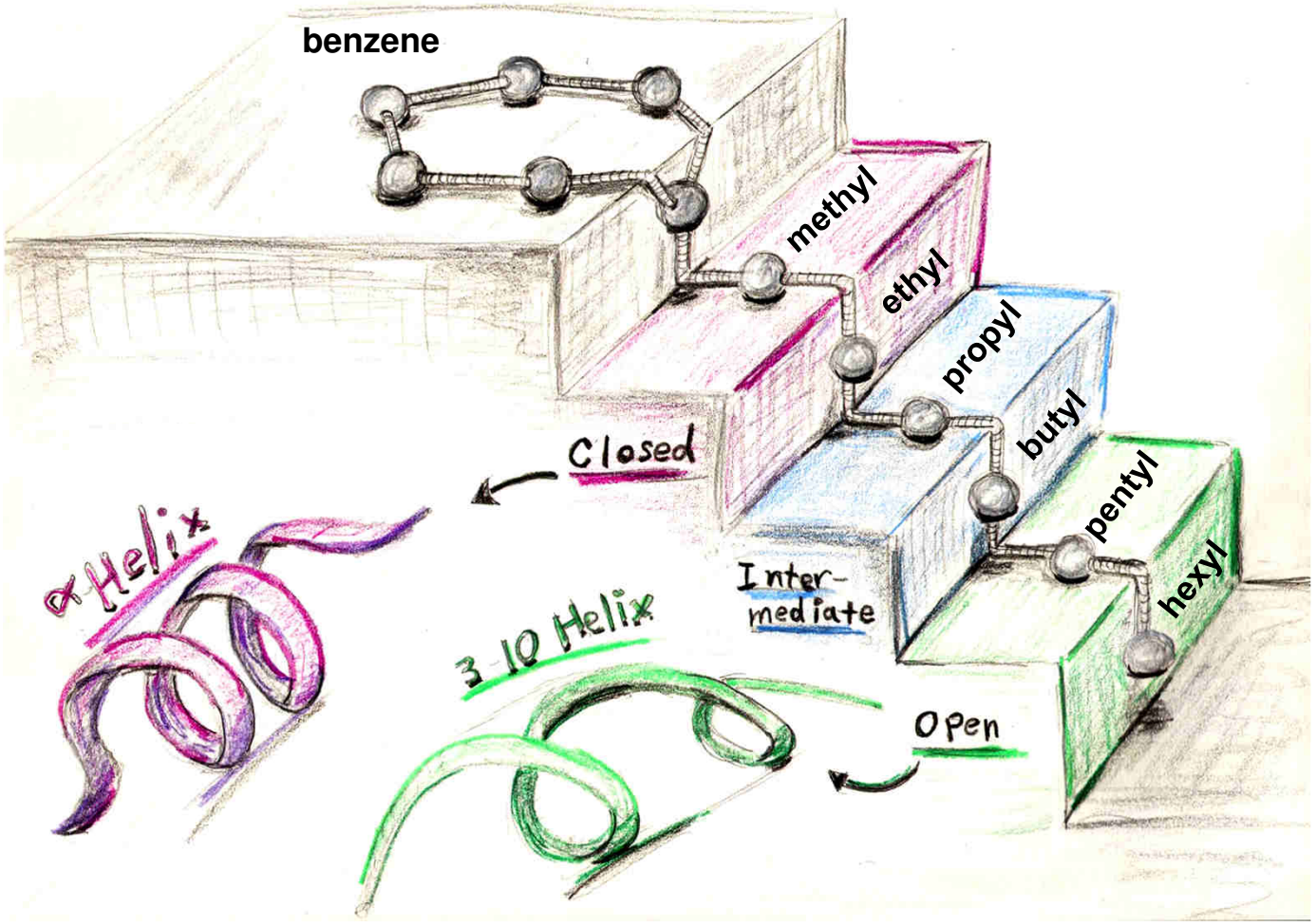
Is this response of the receptor to ligand size discreet or continuous?

Merski, Fischer, Balius, et al 2015 PNAS 112(16):5039-44

Binding affinity is less than hydrophobic burial



Merski, Fischer, Balius, et al 2015 PNAS 112(16):5039-44


benzene
methyl
ethyl
propyl
butyl
pentyl
hexyl
Closed
Inter-
mediate
Open
α Helix
3 10 Helix

Outline

- The binding event using a thermodynamic cycle
- Introduction to molecular mechanics
- Introduction to molecular dynamics
- Introduction to molecular docking
- Applications from my work
  - Large-scale docking
  - Receptor flexibility
  - Receptor desolvation

Adding a Term to the Scoring Function

DOCK 3.7 scoring function:

Charge complementarity

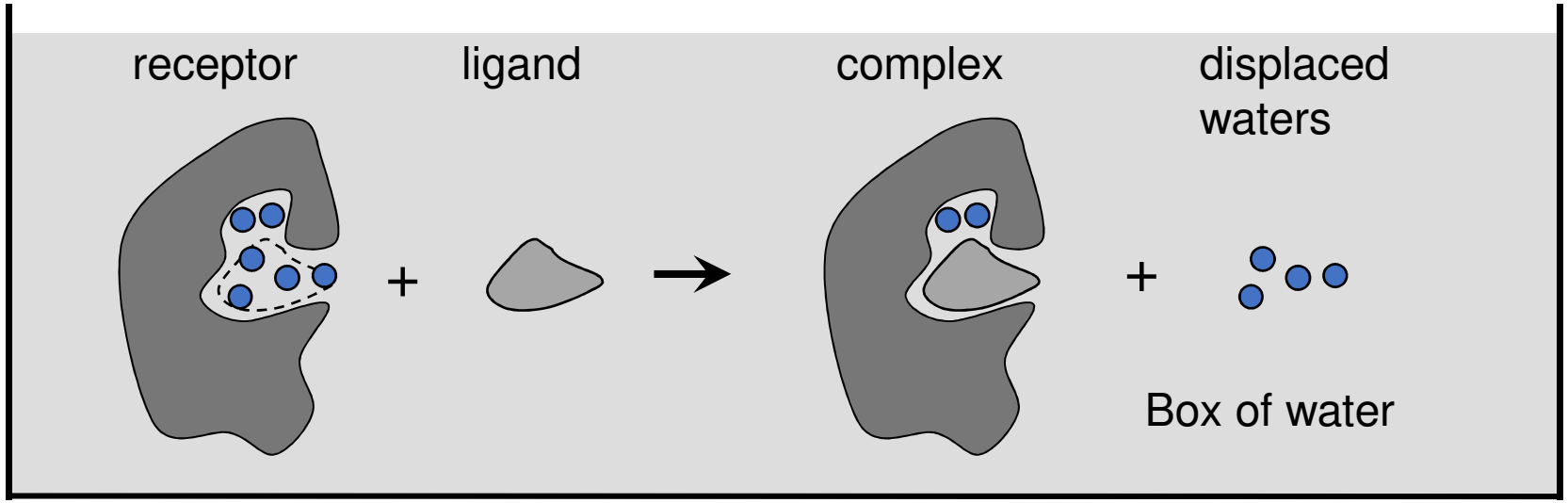
What's missing?

$$E_{score} = E_{VDW} + E_{ES} + E_{lig,desol}$$

Shape complementarity

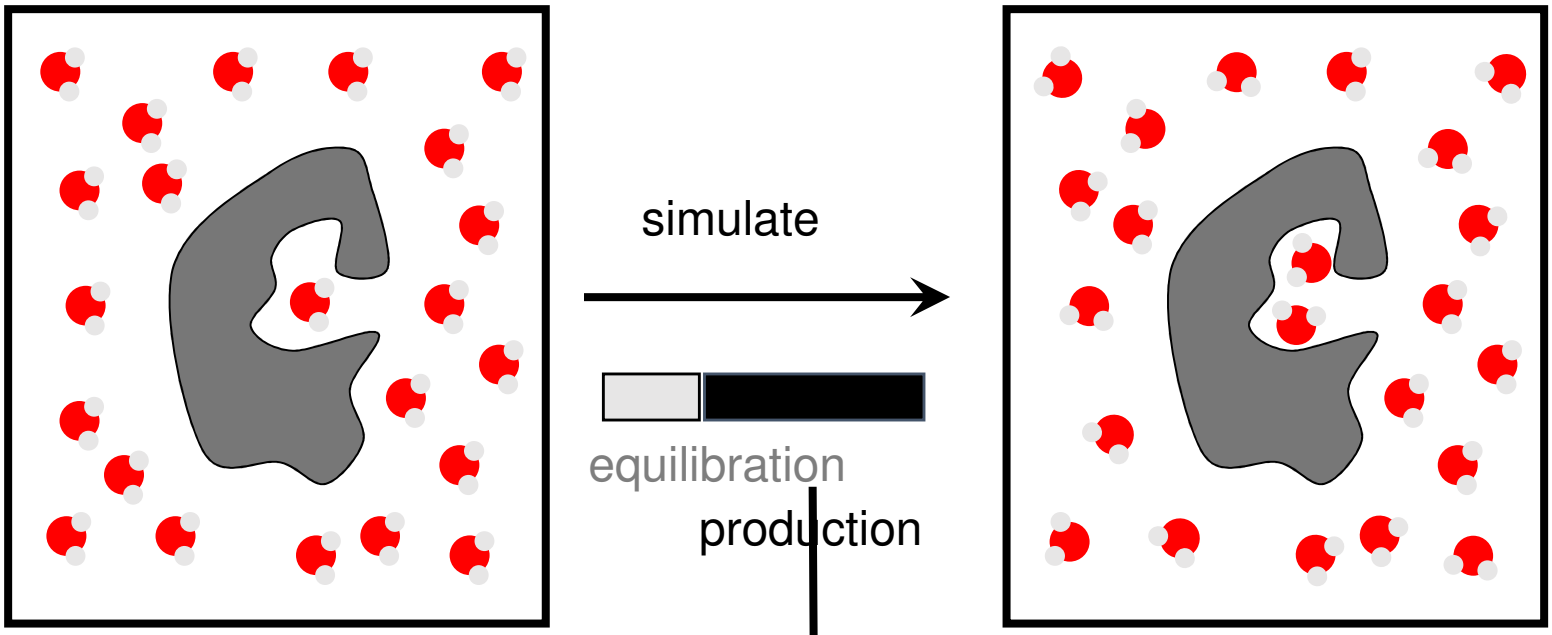
Cost of removing water from **ligand**

Cost of removing water from **receptor**



Meng, et al J. Comput. Chem. 1992, 13, 505– 524
Mysinger and Shoichet J Chem Inf Model. 2010, 50(9):1561-73

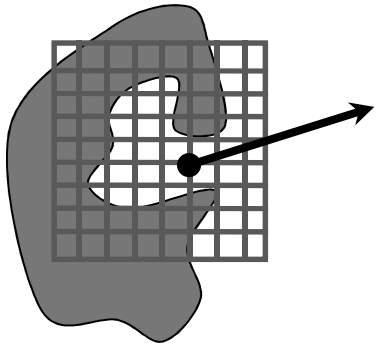
Calculating Water Energetics with Receptor



Process simulation

GIST is implemented
in AmberTools14 in CPPtraj

Grid Inhomogeneous
Solvation Theory or GIST

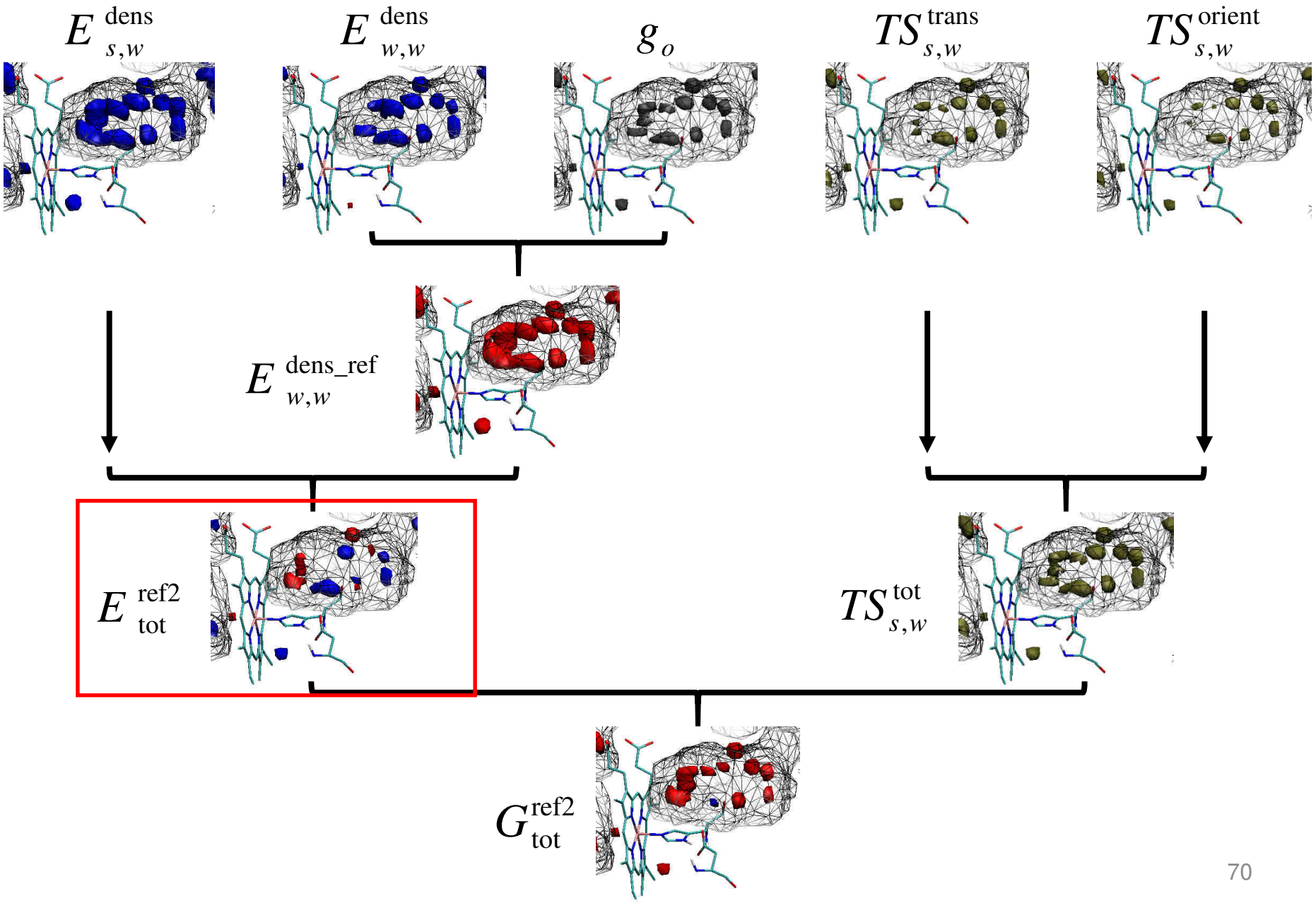


Each grid point is a voxel
4 values:
DE_{sw} ; DE_{ww} ;
$TDS_{sw,orient}$; $TDS_{sw,trans}$

T. Lazaridis, J. Phys. Chem. B, 1998, 102, 3531-3541.
T. Lazaridis, J. Phys. Chem. B, 1998, 102, 3542-3550

C.N. Nguyen, et al, J. Chem. Phys. 2012, 137, 044101

Combining GIST Grids

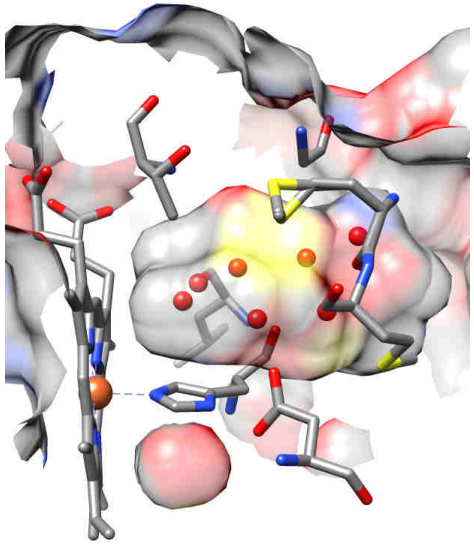


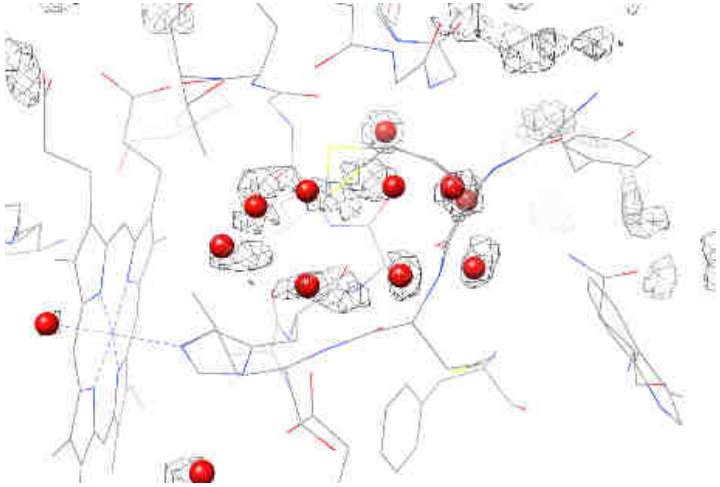
Testing DOCK+GIST Using a Model Cavity

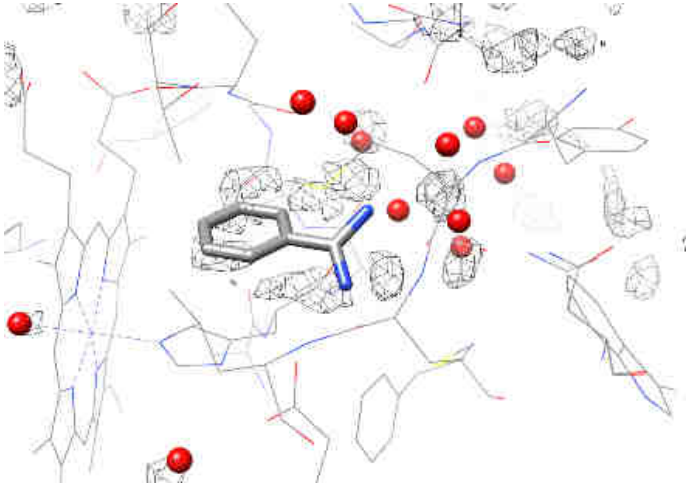
- Model systems are simple engineered cavities
- They are dominated by 1-2 interaction terms, allowing us to disentangle various energetic contributions in docking

Cytochrome *c* peroxidase gateless mutant:

- Mutations/deletions result in solvent-exposed binding site (~8 water molecules)
- Alternative loop conformations (residues 186-194)
- Contains one anionic residue (Asp233)
- Almost exclusively binds small monocations
- Straightforward binding assay and crystallography







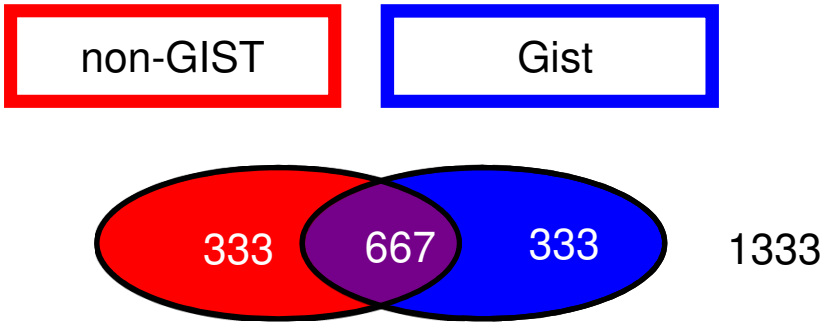
(Fischer et al., 2014; Rosenfeld et al., 2002)

Prospective Screens

We make a computational prediction and test it experimentally

Selecting Molecules to Test with Differences

- We screened up to 1.8 million fragment molecules to the CcP-gateless mutant
  - Preformed 2 screens: Non-GIST and GIST
  - We are interested in differences



- Comparing GIST to standard screening, molecules were chosen based on:

Large Rank Differences

Different Poses

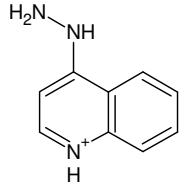
New Chemotypes

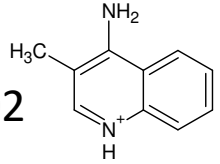
Water-mediated Interactions

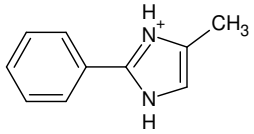
- 17 compounds (**14 pro-GIST** and **3 anti-GIST**) have been bought and tested

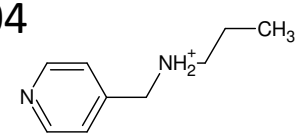
We Tested 17 Molecules

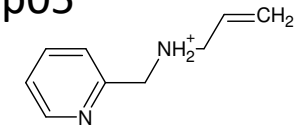
14 pro-GIST, 3 anti-GIST

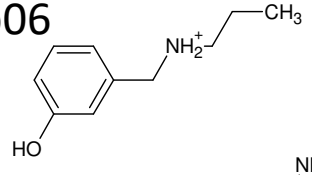
cmp01 

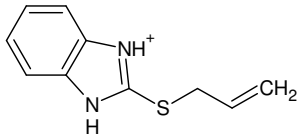
cmp02 

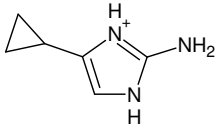
cmp03 

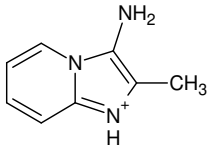
cmp04 

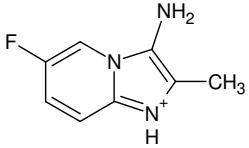
cmp05 

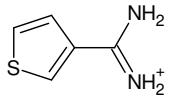
cmp06 

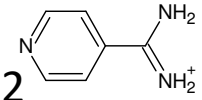
cmp07 

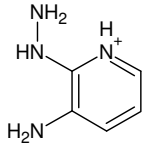
cmp08 

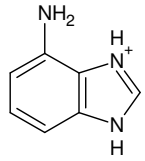
cmp09 

cmp10 

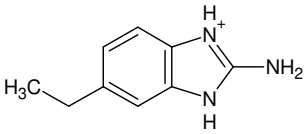
cmp11 

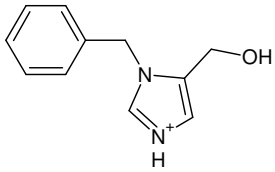
cmp12 

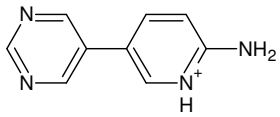
cmp13 

cmp14 

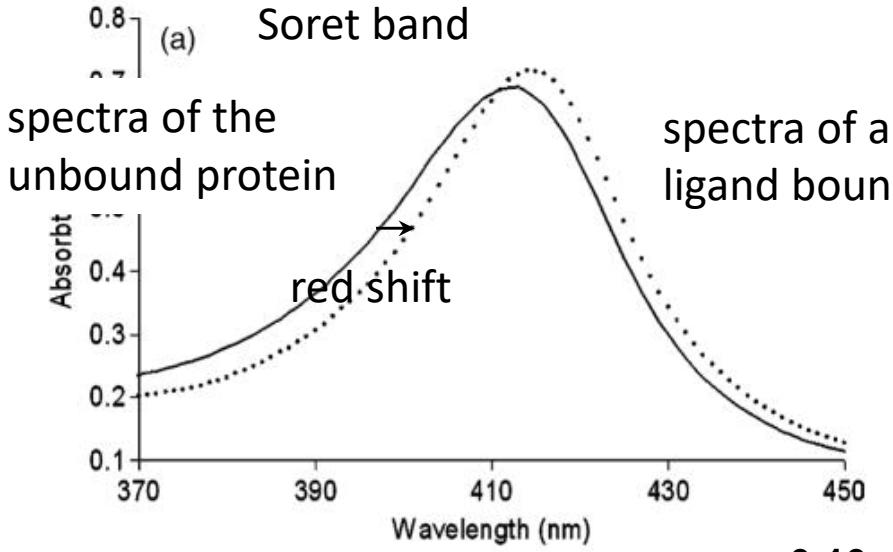
3 Anti-GIST:

cmp15 

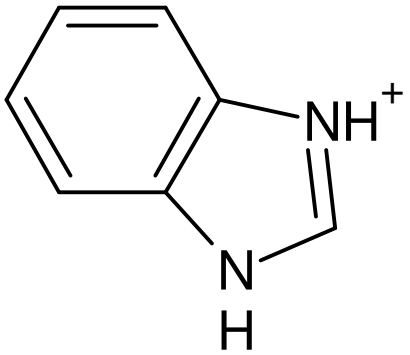
cmp16 

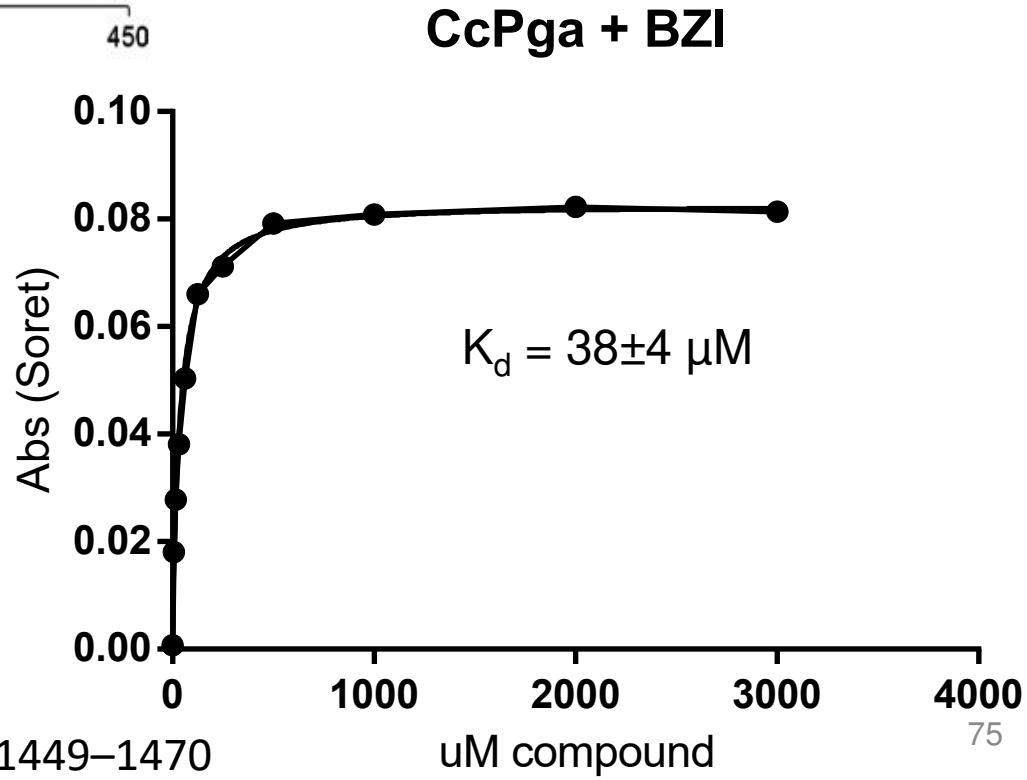
cmp17 

Experimental Assay to Detect Binding



Benzimidazole (control)

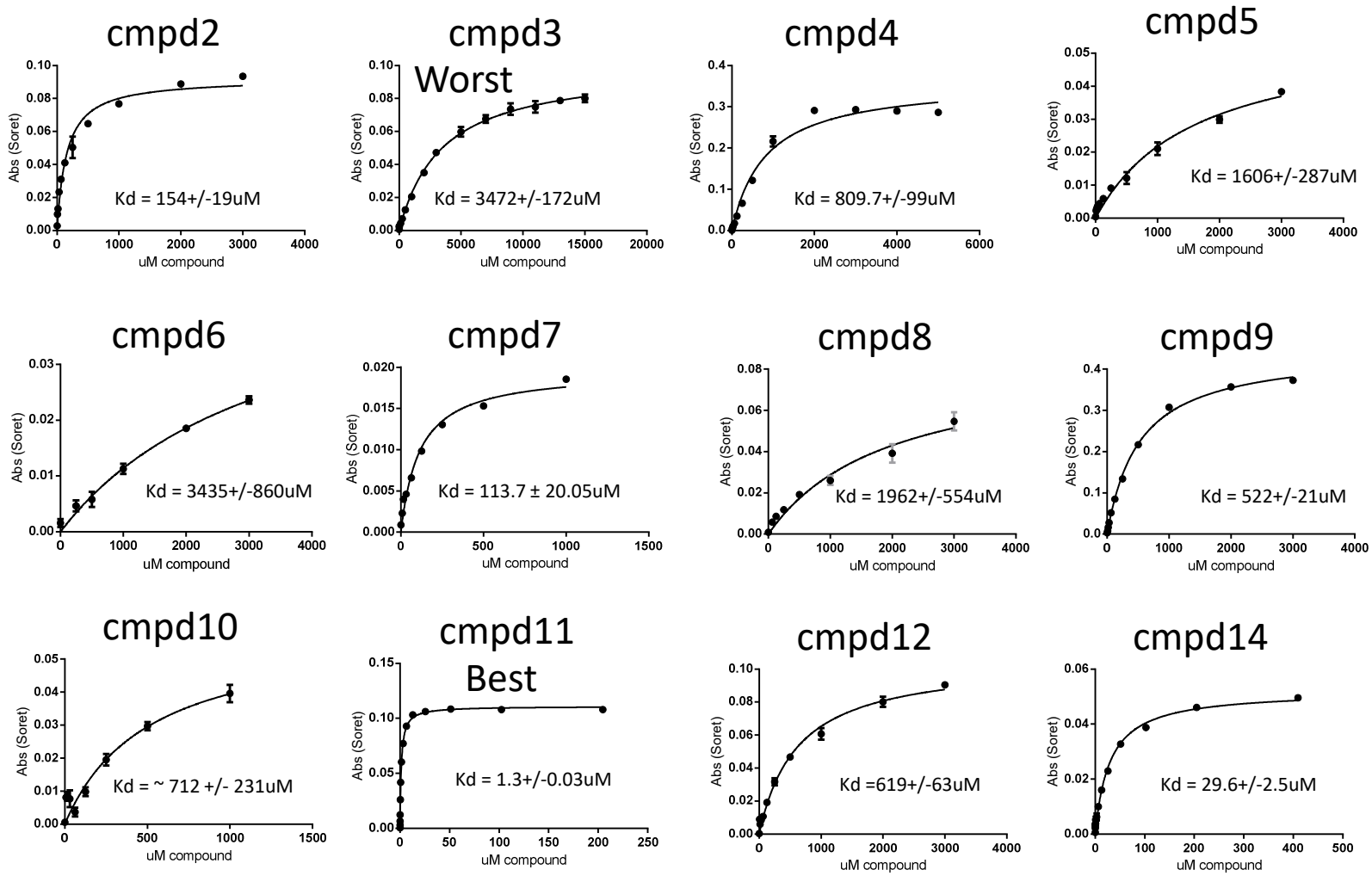




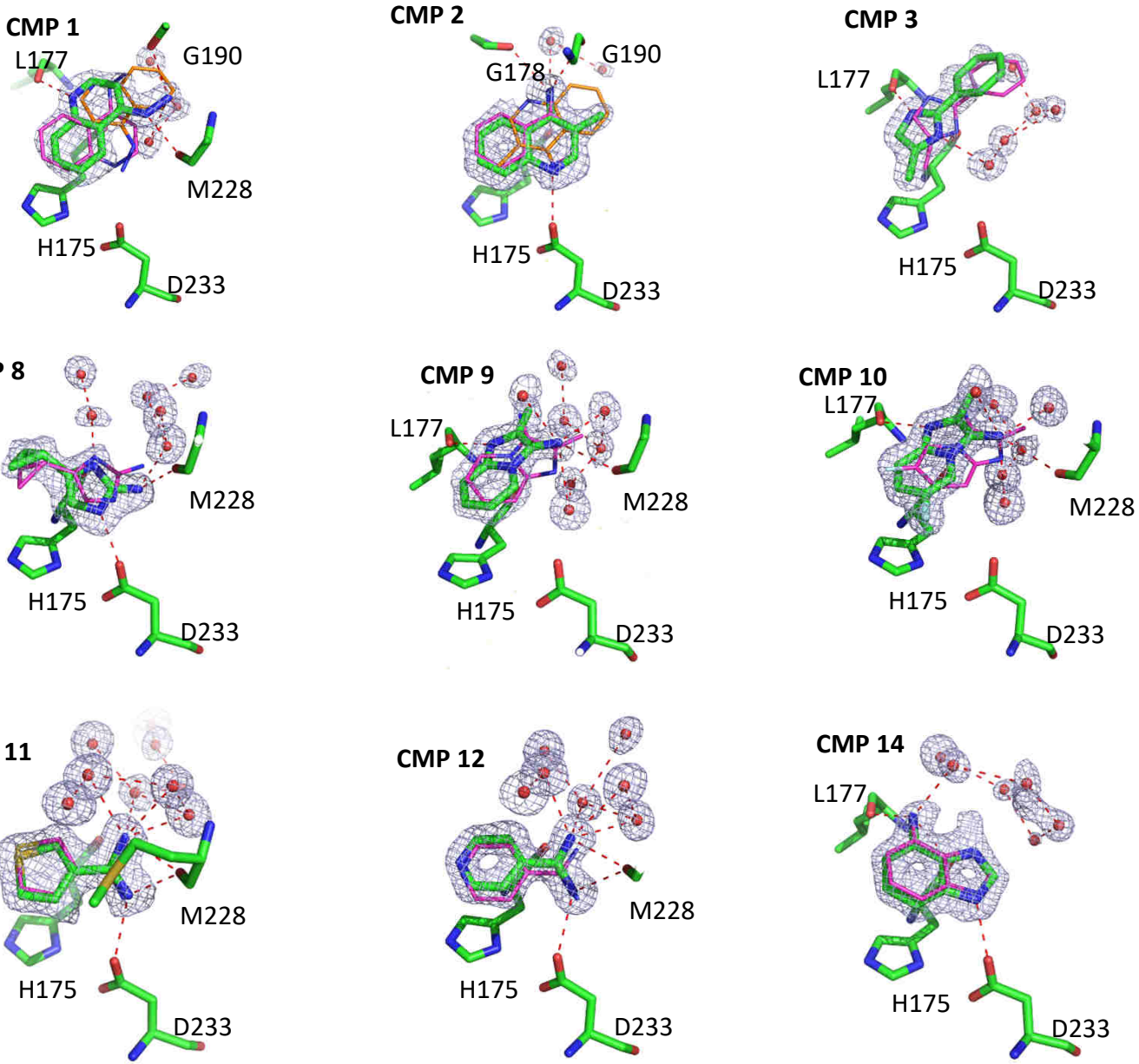
Brenk, J Mol Biol. 2006 Apr 14; 357(5): 1449–1470

We determined Affinities for 12 molecules

Affinities range from 1µM to 3.5mM



9 Crystal Structures for Pose Comparisons



Prospective Summary

14 pro-GIST, 3 anti-GIST

8 Binder <1mM

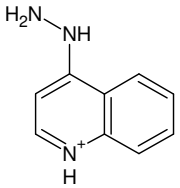
4 Binder [1mM to 4mM]

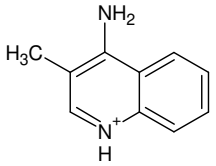
6 Xtal pose, dock is right

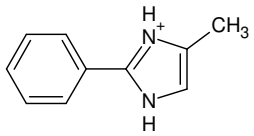
3 Xtal pose, dock is wrong

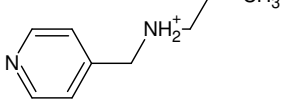
Non-determined

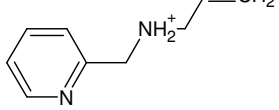
Non-binder

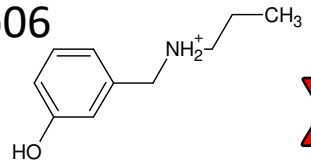
cmp01 

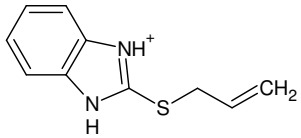
cmp02 

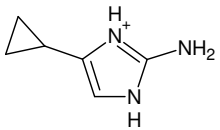
cmp03 

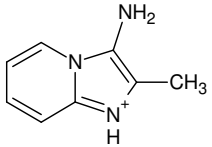
cmp04 

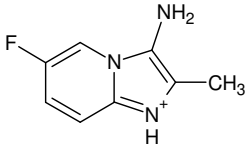
cmp05 

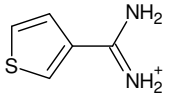
cmp06 

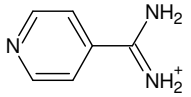
cmp07 

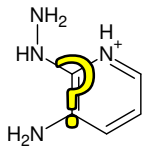
cmp08 

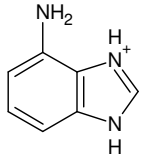
cmp09 

cmp10 

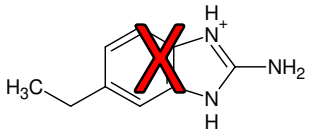
cmp11 

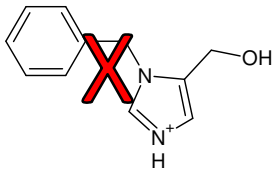
cmp12 

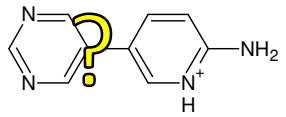
cmp13 

cmp14 

3 Anti-GIST:

cmp15 

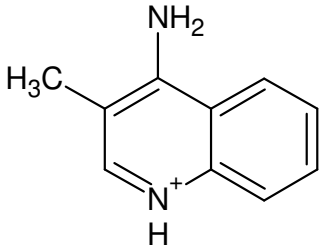
cmp16 

cmp17 

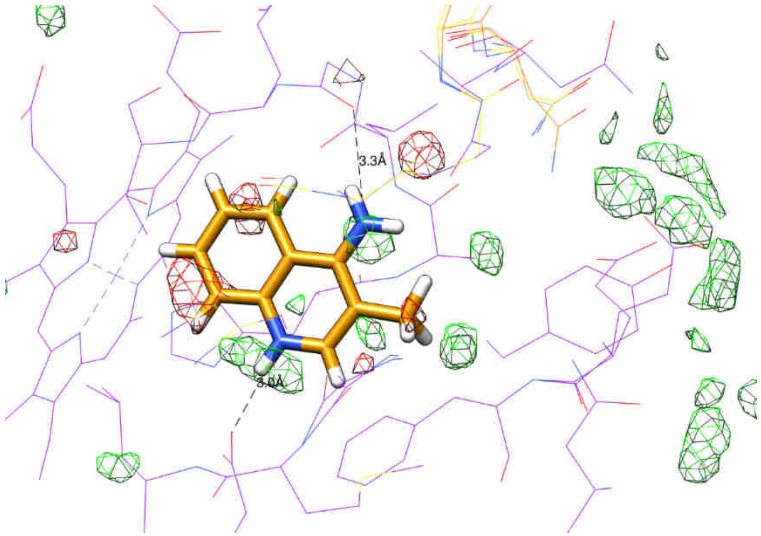
Extra slides.

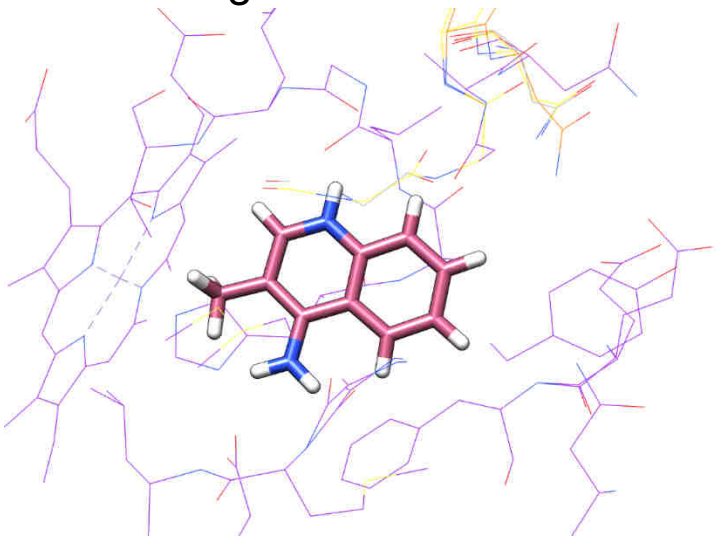
CMP 2: Binder, GIST Pose is Right

name	GIST	nonGIST	logrankdiff	rmsd
ZINC000006557114	664	740	0.047	4.62



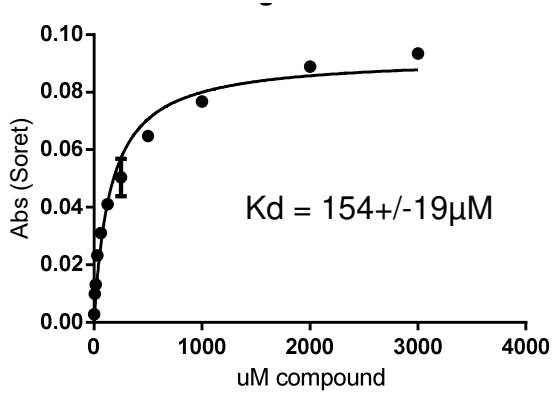
Pose change

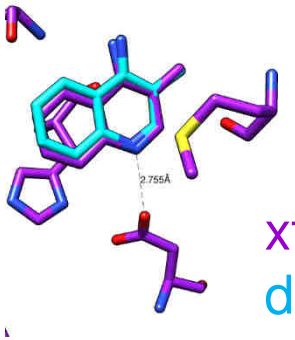




GIST

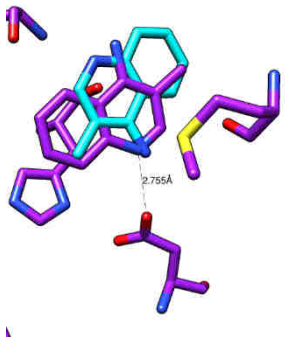
Non-GIST





xtal

docked

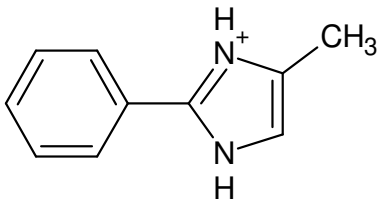


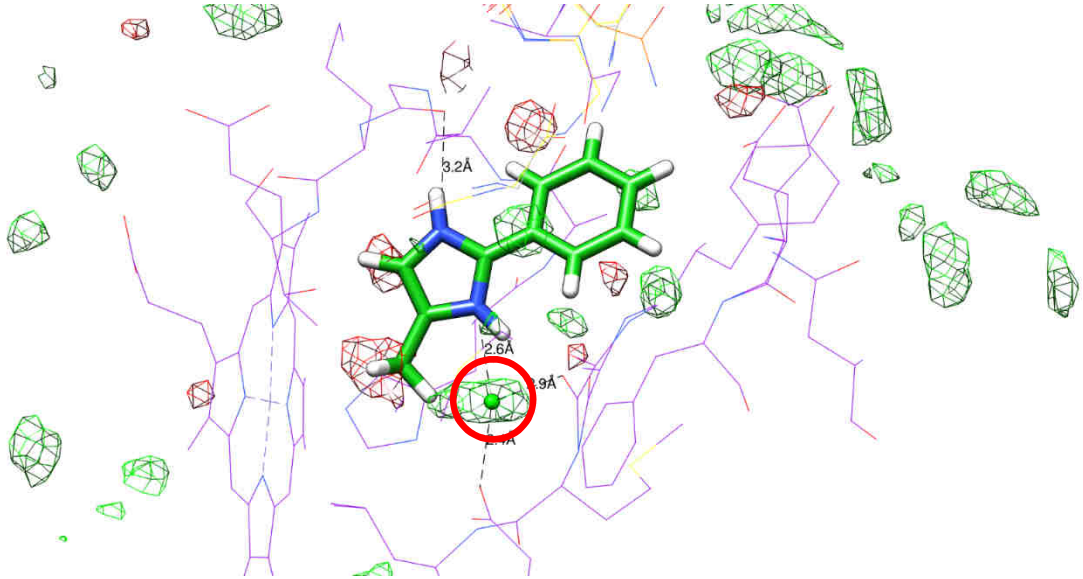
RMSD = 0.34

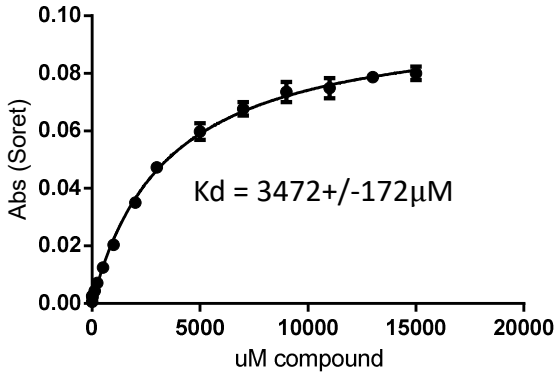
RMSD = 3.23

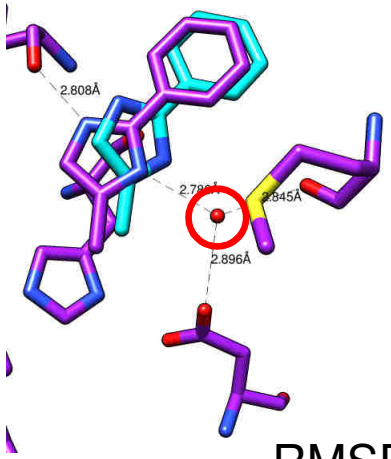
CMP 3: Water-Mediated, Weak Binder

name	GIST	nonGIST	logrankdiff	rmsd
ZINC000004705523	13	249	1.28	0





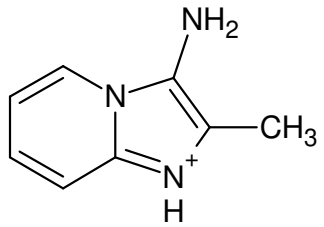


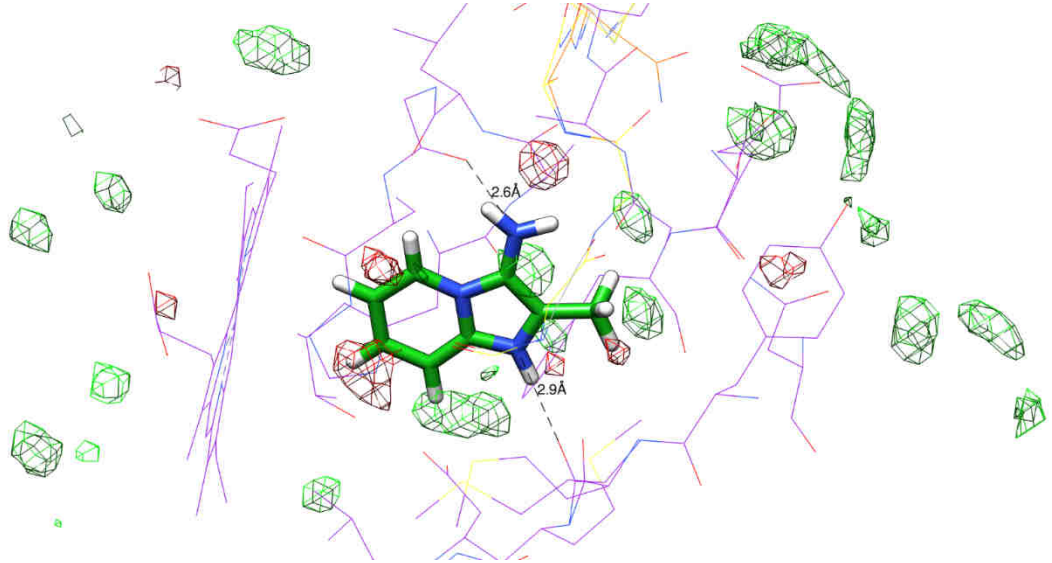


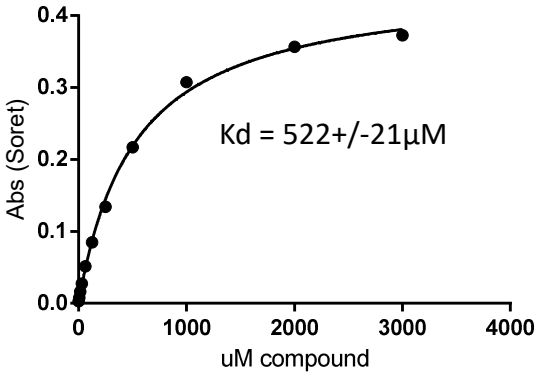
RMSD = 1.32

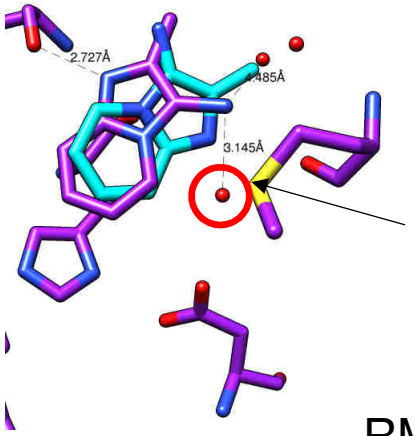
CMP 9: binder, wrong pose

name	GIST	nonGIST	logrankdiff	rmsd
ZINC000020357620	98	745	0.88093	0







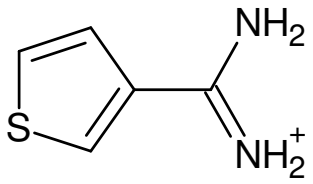


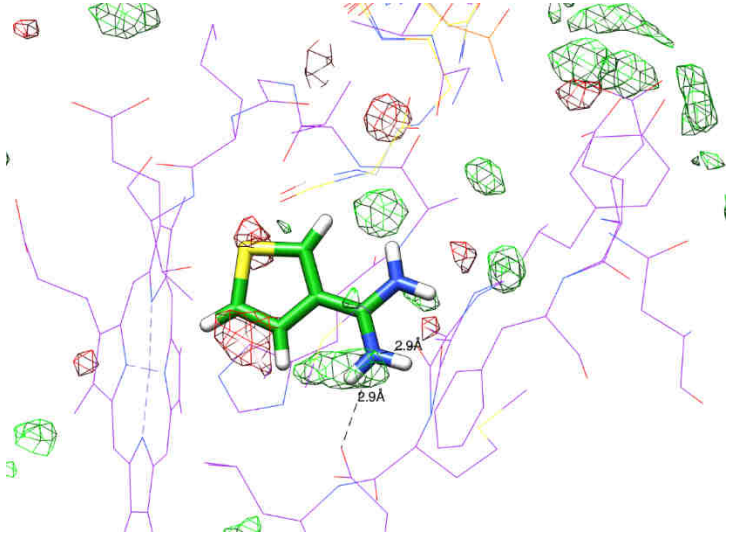
Failed to predict water-mediated interaction

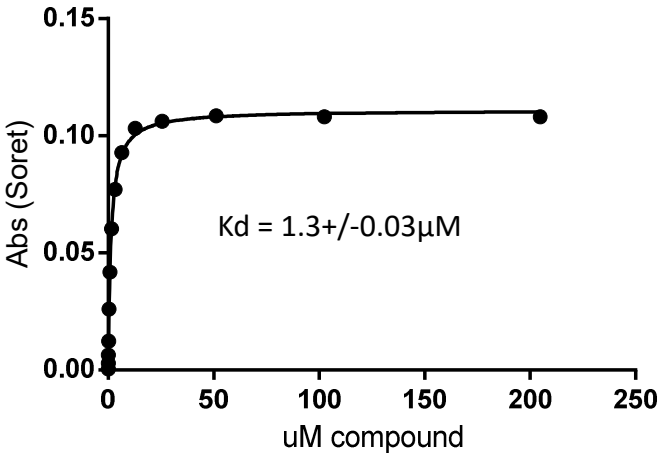
RMSD = 1.73

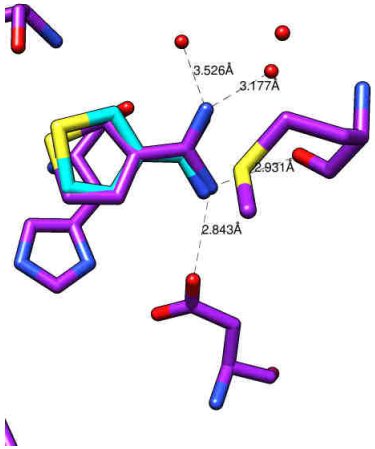
CMP 11: best binder

name	GIST	nonGIST	logrankdiff	rmsd
ZINC000000161834	358	1212	0.52962	0



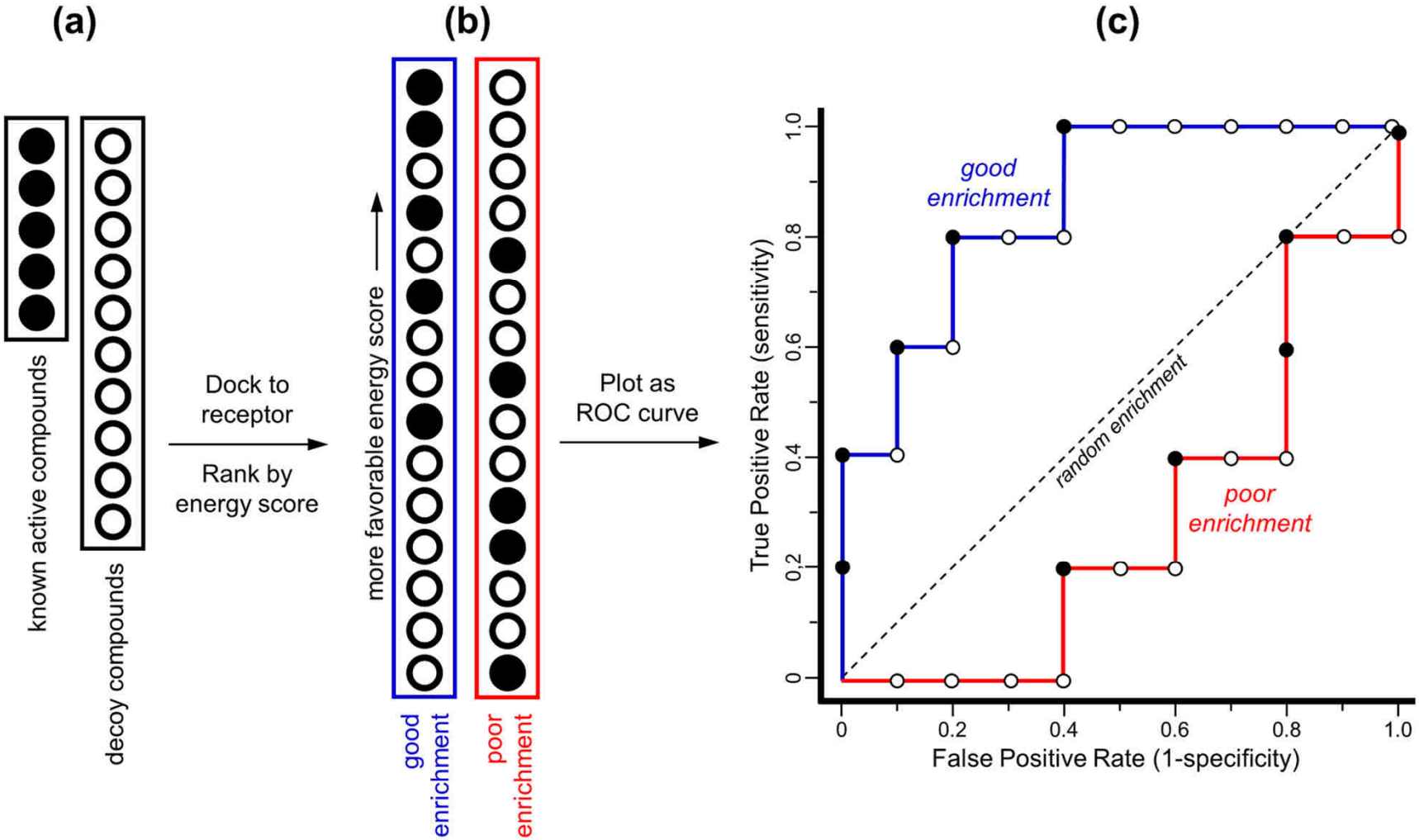




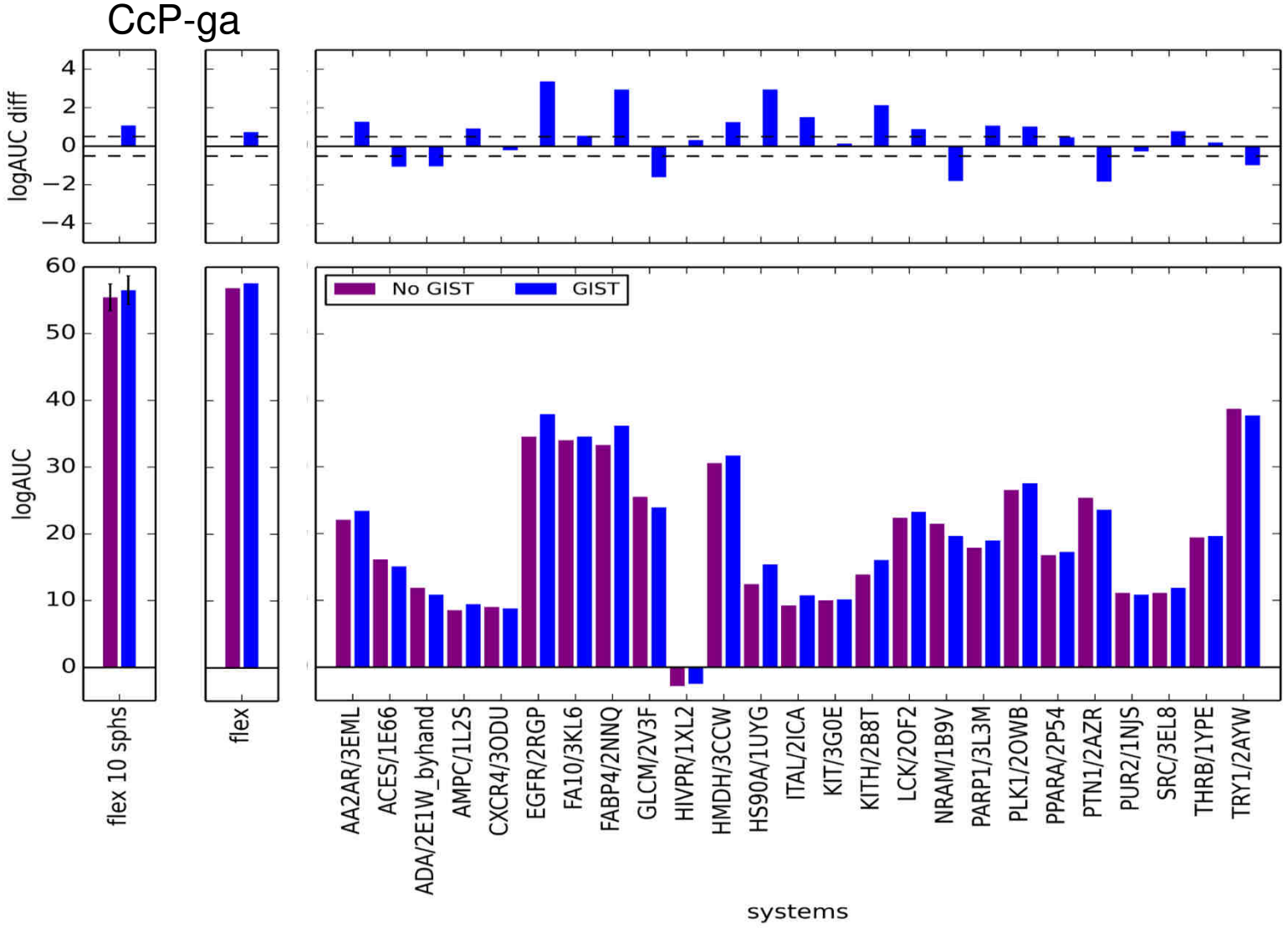


RMSD = 0.4448

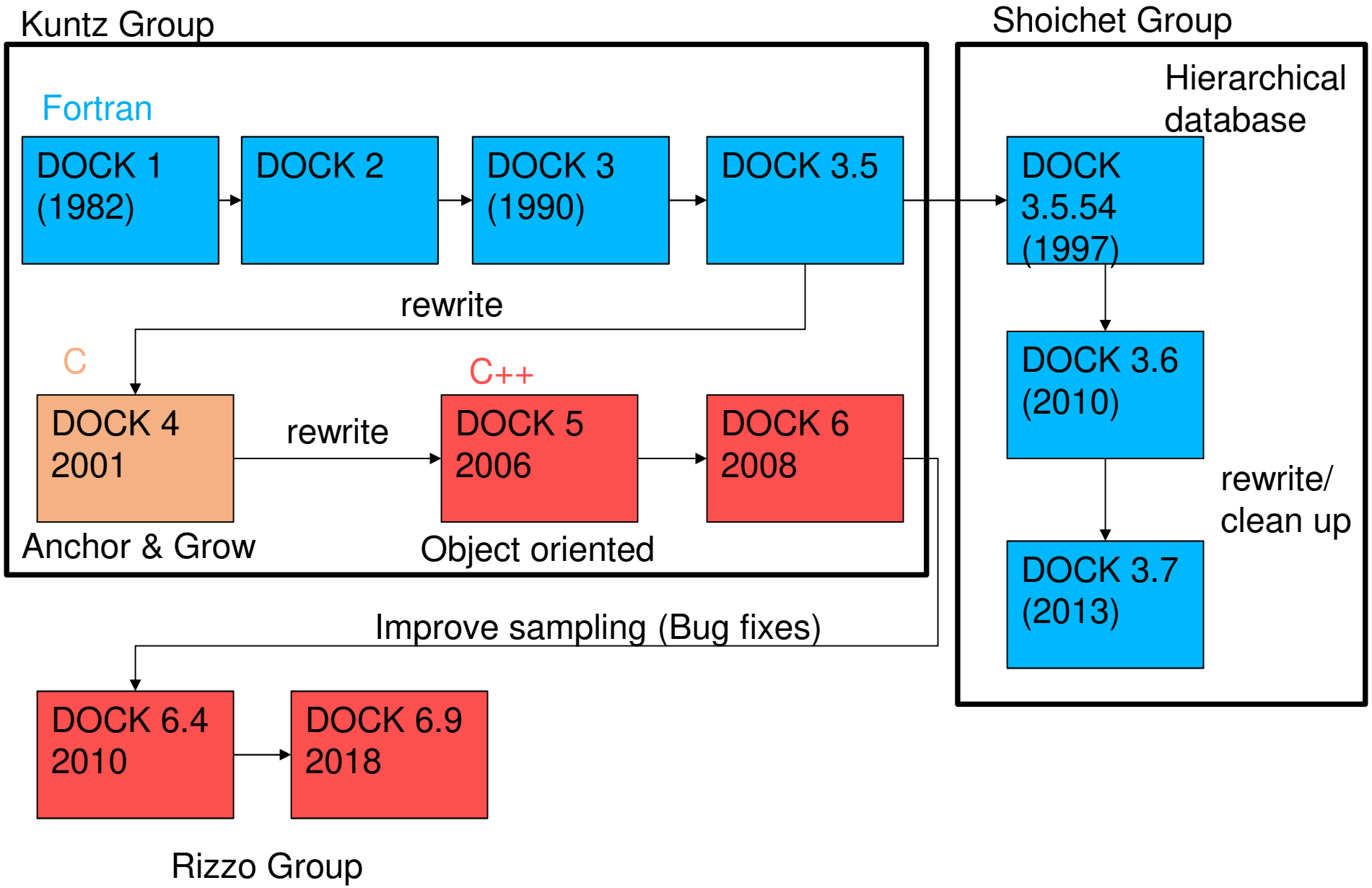
Retrospective testing
Enrich knowns over Decoys



GIST has little effect on retrospective analysis of 25 DUDE systems



DOCK: A History



Docking is Important in Ligand Discovery

Applications of docking

- Virtual Screening: given a protein and database of molecules find those that bind.
- Pose prediction: given a molecule and a protein predict how that they bind

Docking Tasks

- Sampling
  - generate all the possibilities including finding the correct geometry
- Scoring
  - of all the possibilities, rank the correct pose first
  - also, rank the binders better than decoys
- Balance of speed and accuracy, docking has to be fast.

How to evaluate docking methods

- Pose reproduction, reproduce the crystallographic poses
- Enrichment calculations, make sure ligand found in the top of the rank orders lists.
- Prospective testing on model cavities, make a predication, and test it!

It's a miracle docking ever works

DOCK 3.7 scoring function:

Charge complementarity

$$E_{score} = E_{VDW} + E_{ES} + E_{lig,desol}$$

What's missing?

Shape complementarity

Cost of removing water from **ligand**

$$p_A(g) = \sum_{b \in rec} \frac{A_b}{r^{12}}$$

To make it fast we precompute potentials, ligand properties, displacement

interpolate

$$E_{VDW} = \sum_{a \in lig} A_a * p_A(r_a) - B_a * p_B(r_a)$$

$$E_{ES} = \sum_{a \in lig} q_a * \Psi(r_a) \qquad E_{lig,desol} = \sum_{a \in lig} s_a * D(r_a)$$

Cost of removing water from **receptor**

$$E_{score} = E_{VDW} + E_{ES} + E_{lig,desol} + E_{rec,desol}$$
$$E_{lig,wat} + E_{rec,conf} + E_{lig,strain}$$

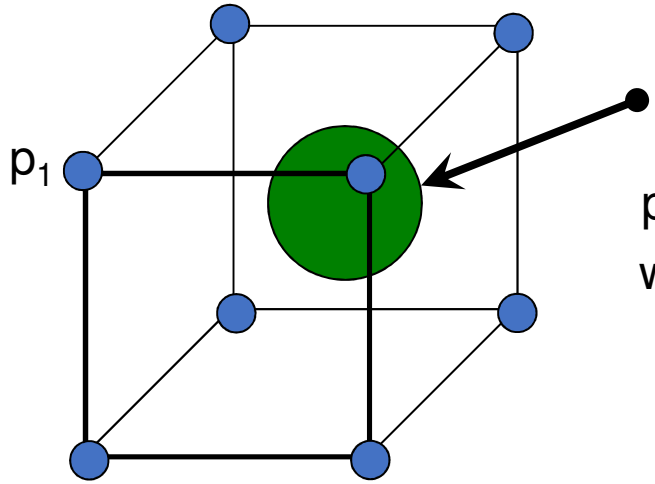
Meng, et al J. Comput. Chem. 1992; Mysinger and Shoichet JCIM, 2010; Fischer et al, Nature Chem. 2014; Balius et al, PNAS, 2017

$$E = \sum_{i \in L} \left(\sqrt{A_{i,i}} \sum_{j \in R} \frac{\sqrt{A_{j,j}}}{r_{i,j}^{\,a}} - \sqrt{B_{i,i}} \sum_{j \in R} \frac{\sqrt{B_{j,j}}}{r_{i,j}^{\,b}} + 332 q_i \sum_{j \in R} \frac{q_j}{D r_{i,j}} \right)$$

$$G_{av}(p) = \sum_{l \in R} \frac{\sqrt{A_{l,l}}}{r_{p,l}^{\,a}}$$

$$G_{rv}(p) = \sum_{l \in R} \frac{\sqrt{B_{l,l}}}{r_{p,l}^{\,b}}$$

$$G_{es}(p) = 332 \sum_{l \in R} \frac{q_l}{D r_{p,l}}$$



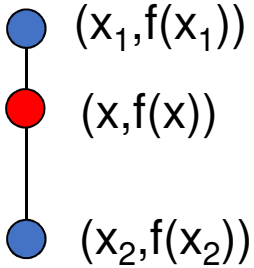
http://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm#Grid

$$E \approx \sum_{i \in L} \begin{pmatrix} \sqrt{A_{i,i}}\,\mathrm{interp}[G_{av}(p_1), \cdots, G_{av}(p_8)] \\ -\sqrt{B_{i,i}}\,\mathrm{interp}[G_{rv}(p_1), \cdots, G_{rv}(p_8)] \\ +332 q_i\,\mathrm{interp}[G_{es}(p_1), \cdots, G_{es}(p_8)] \end{pmatrix}$$

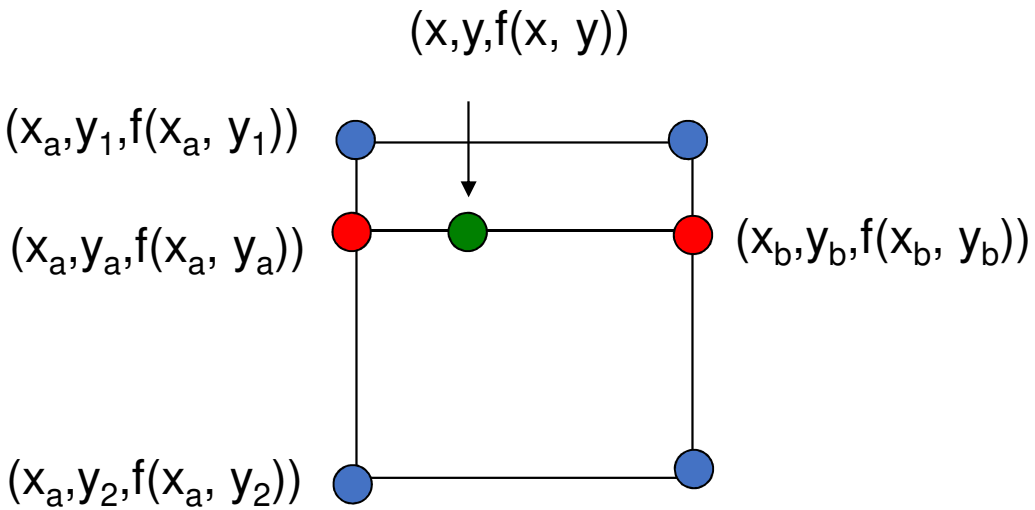
Interpolation

linear

$$f(x) \approx \frac{(x - x_1) f(x_2) + (x_2 - x) f(x_1)}{(x_2 - x_1)}$$



bilinear: Perform 3 linear Interpolations: 2 to calculate red (from cyan); and 1 to calculate green (from red)



Trilinear: for a cube, perform 7 linear interpolations: 4 to calculate red (from the cyan); 2 to calculate green (from red); and 1 to calculate the atomic approximation (from green)

http://en.wikipedia.org/wiki/Trilinear_interpolation

ROC curves

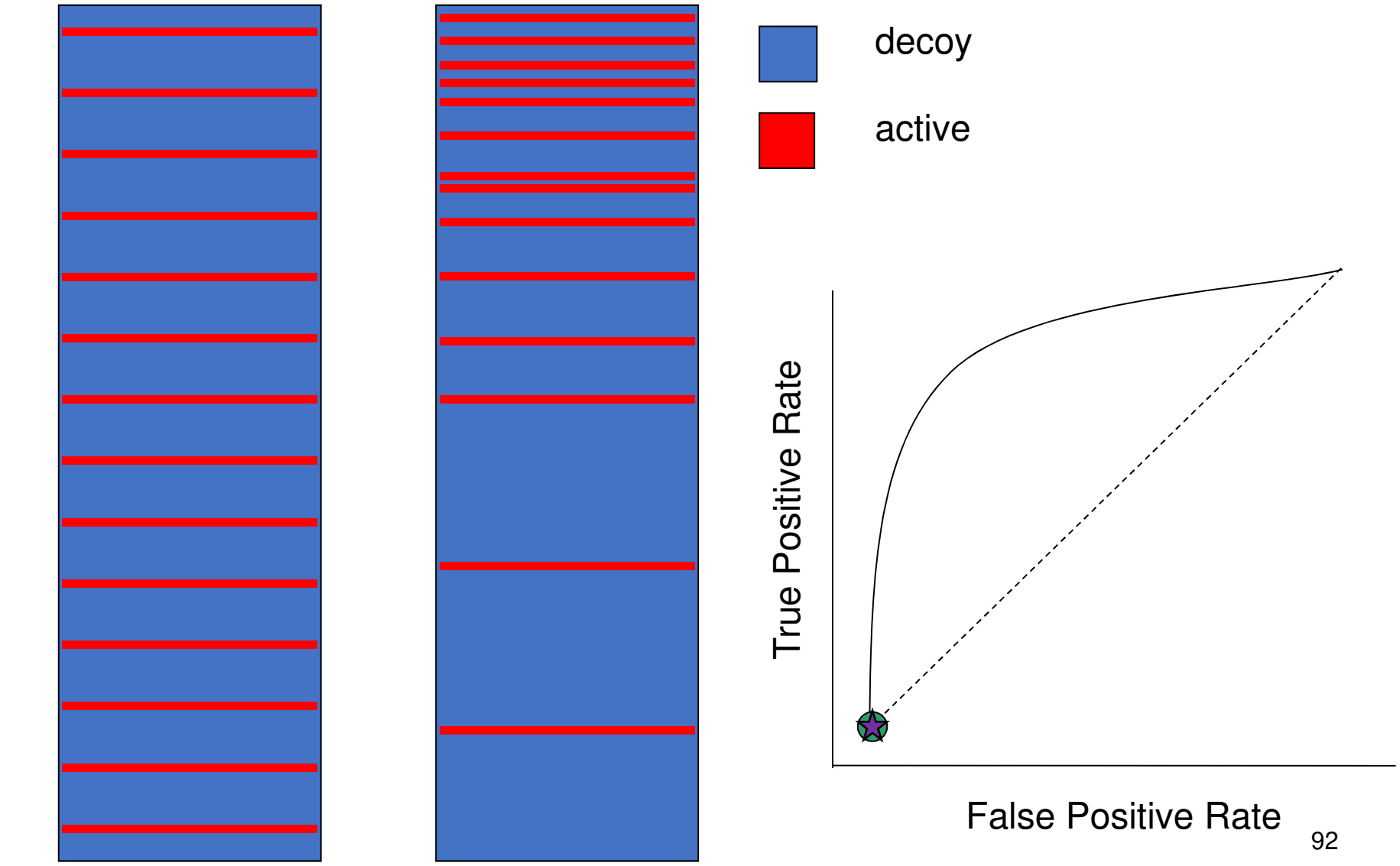
$$TP_{Rate} = Se_{\text{subset}} = \frac{ligands_{\text{selected}}}{ligands_{\text{total}}}$$

$$FP_{Rate} = (1 - Sp)_{\text{subset}} = \frac{decoys_{\text{selected}}}{decoys_{\text{total}}}$$

Se - Sensitivity, Sp - Specificity

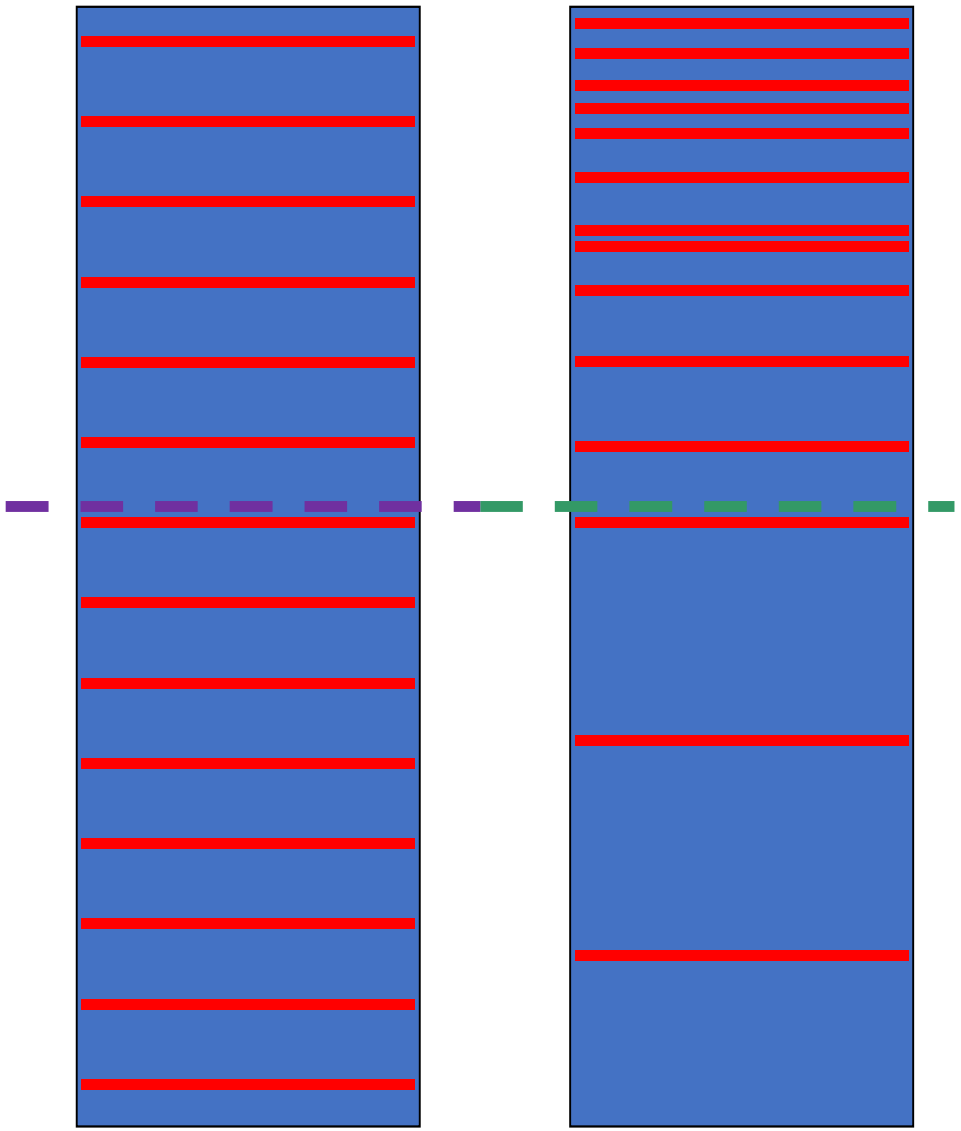
ROC Curves

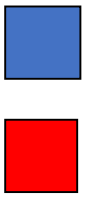
Small molecule database seed with actives



ROC Curves

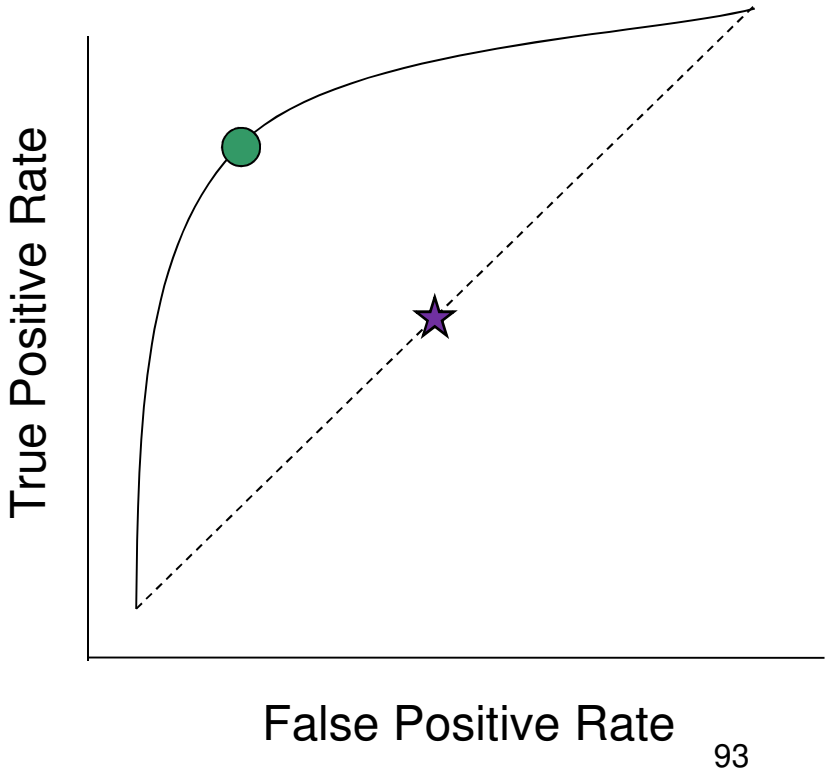
Small molecule database seed with actives





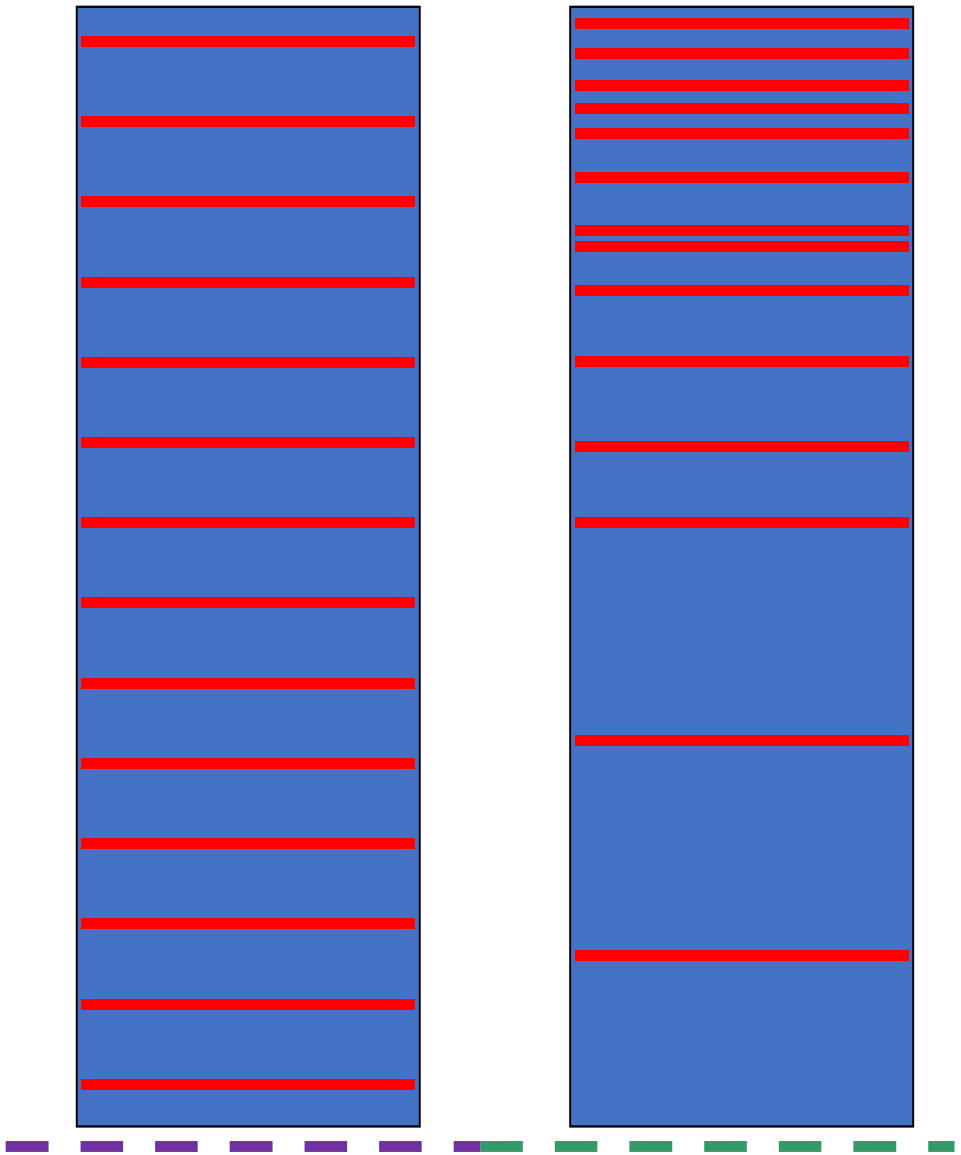
decoy

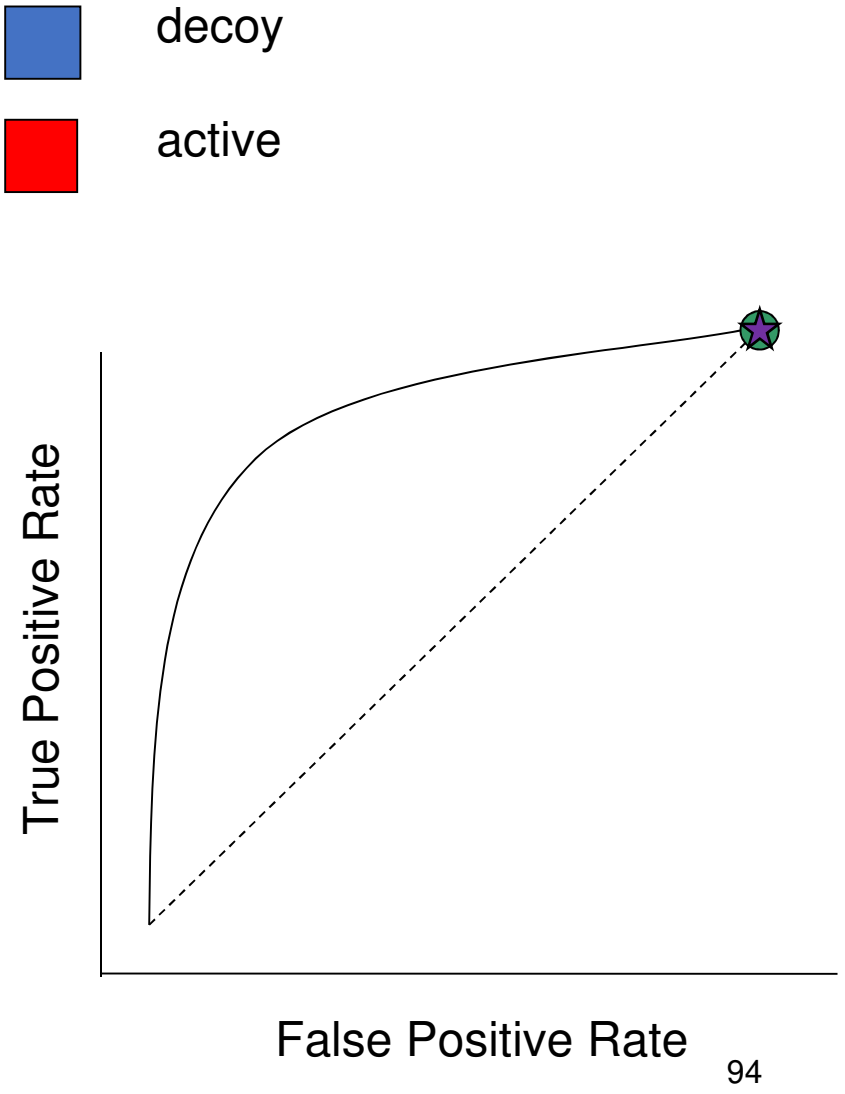
active



ROC Curves

Small molecule database seed with actives





Example ROC Curves