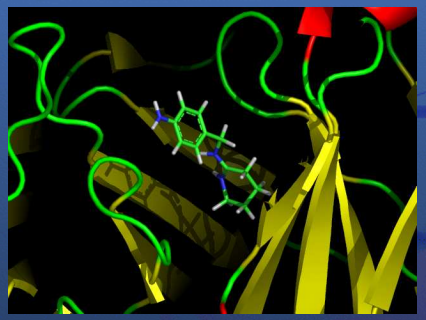
Computer Simulations of Biomolecular Systems

Jonathan W Essex

University of Southampton

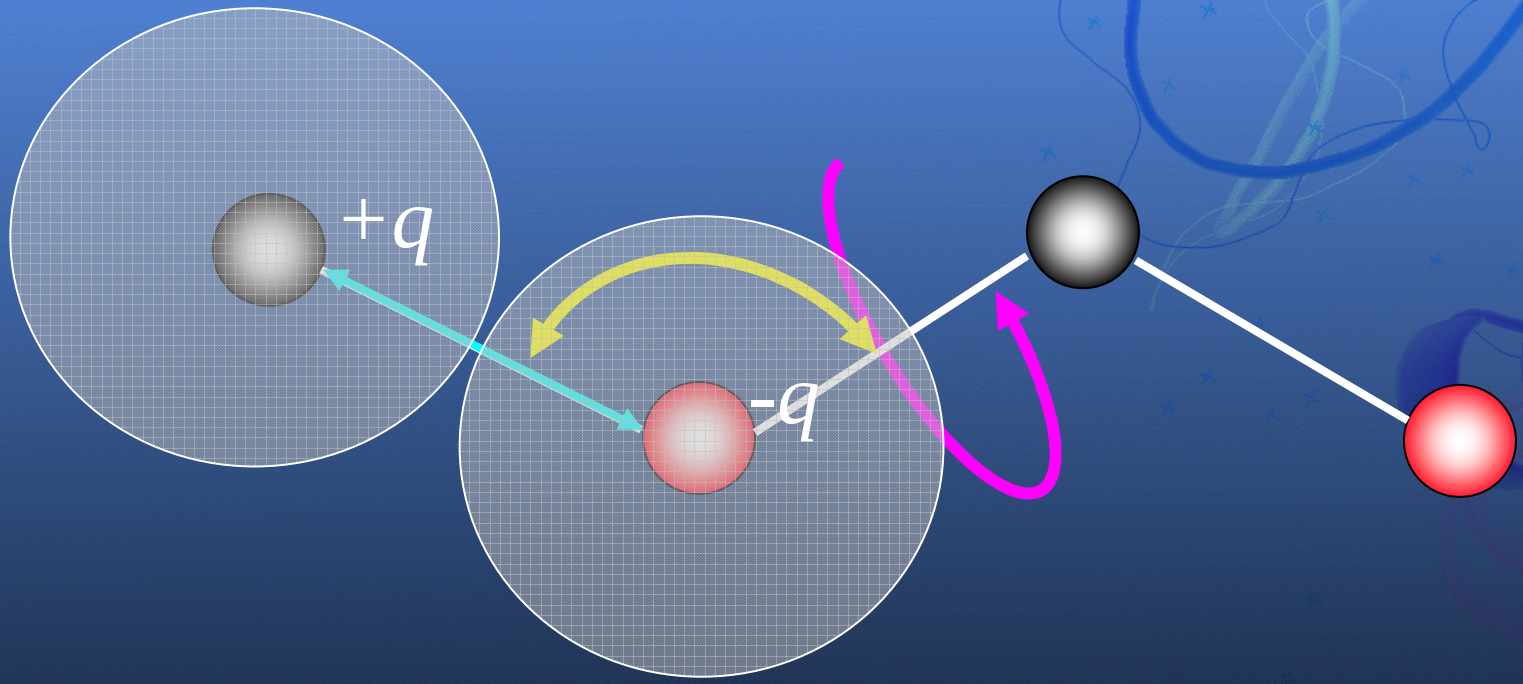
Introduction

- Structure-Based Virtual Screening (SVS) is a widely-used technique for lead discovery and optimisation
  - Protein-Ligand Docking:
    - Sampling - Geometry
    - Scoring - Energy



- Still significant room for improvement
  - Lots of efforts focused on the creation of novel scoring functions
    - Empirical
    - Knowledge based
    - Force field based
- In this presentation
  - Focus on the role of free energy calculations to score molecules

Molecular Mechanics



Molecular Mechanics: use of the potential energy

$$E_{potential} = E_{bond} + E_{angle} + E_{torsion} + E_{van\ der\ Waals} + E_{coulomb}$$

Biological Force Fields

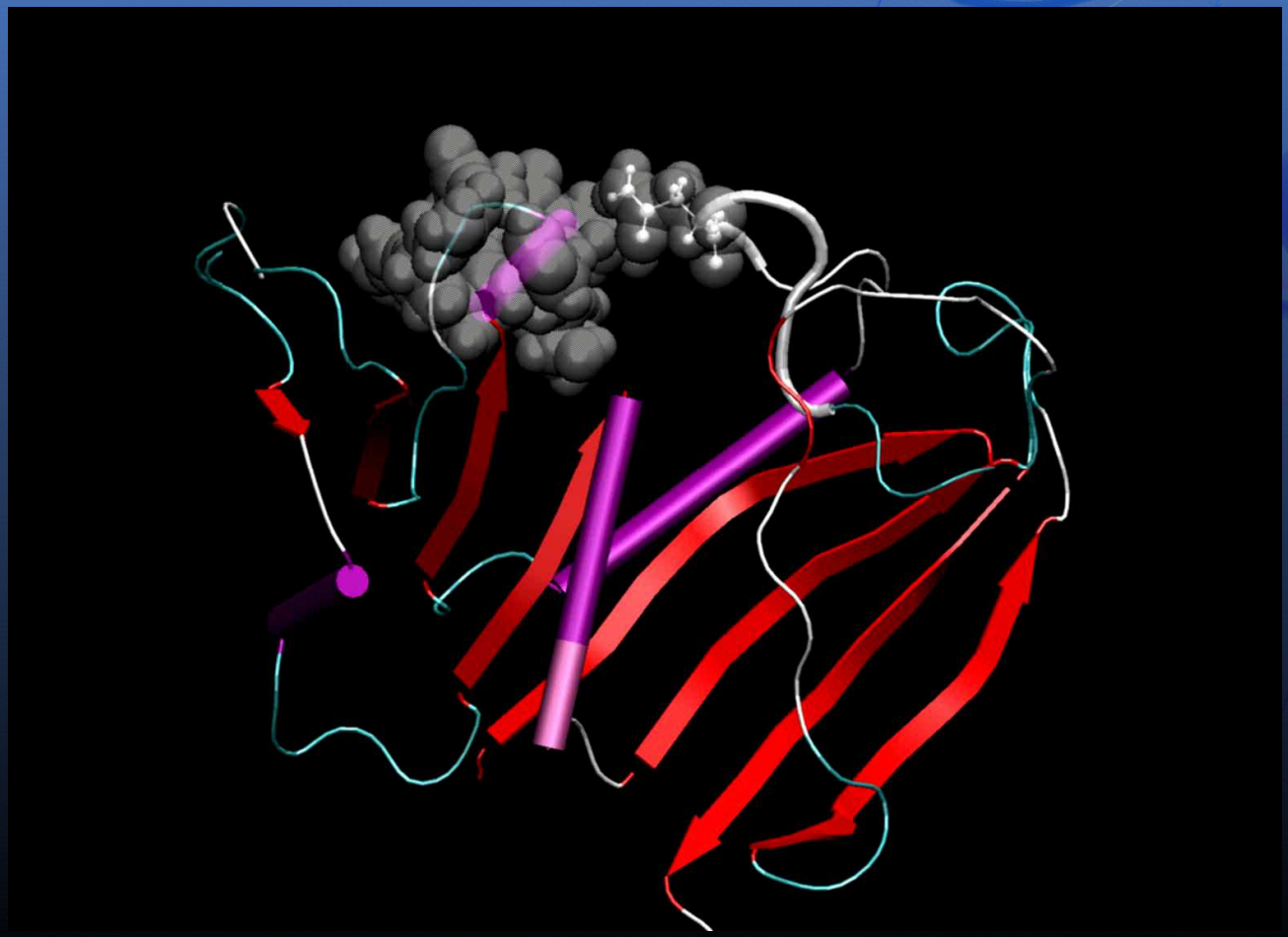
- Effective pair potentials
- Simple functional form

$$U = \sum_{\text{pairs}(i,j)} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

- More complex functional forms generally avoided

DHFR

- Molecular dynamics, with modifications to drive flexible loop



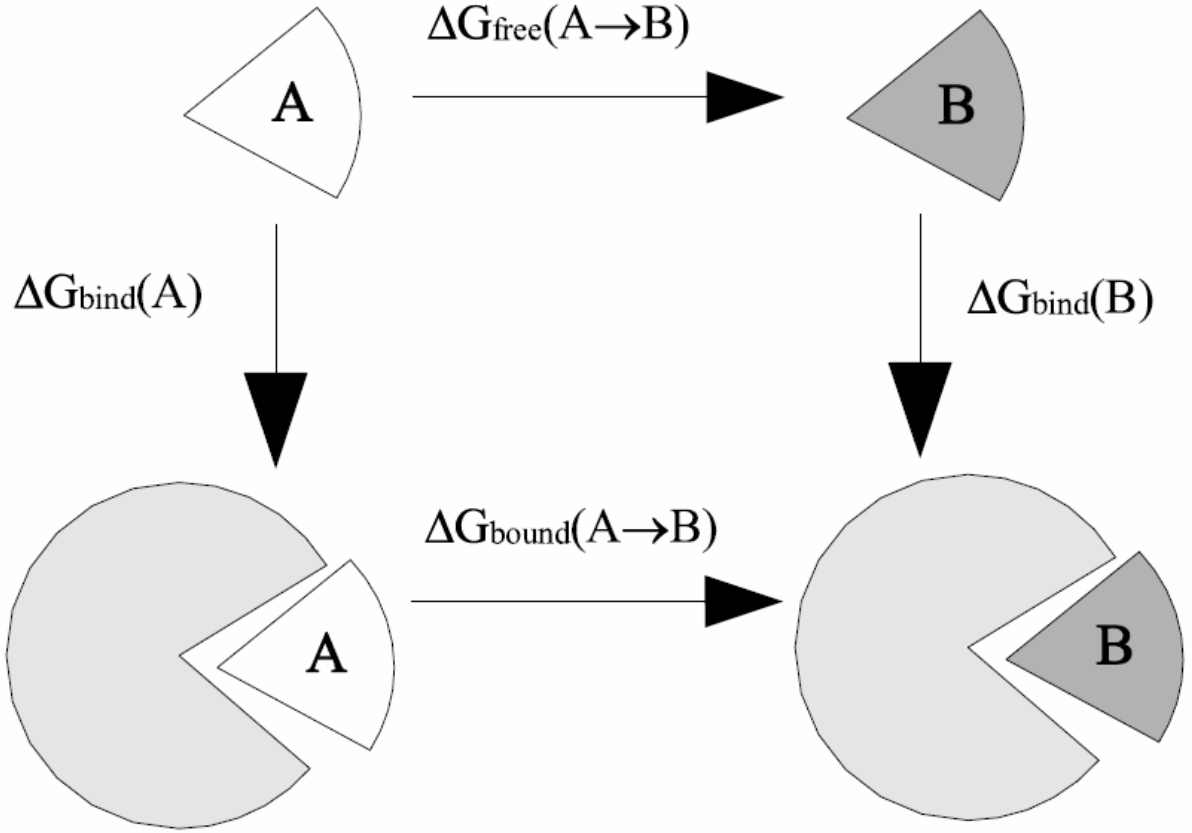
Challenges?

- Computational cost
  - For industrial use, need 10+ compounds per night
- Force field
  - Effective pair potentials
  - Accurate, but only with extensive parameterisation
  - Violates cost criterion!
  - QM/MM?
- Sampling
  - Satisfactory for a particular binding geometry
  - What if this changes though?
  - Novel, efficient sampling algorithms needed
- Applicability
  - Very similar compounds

Methodology

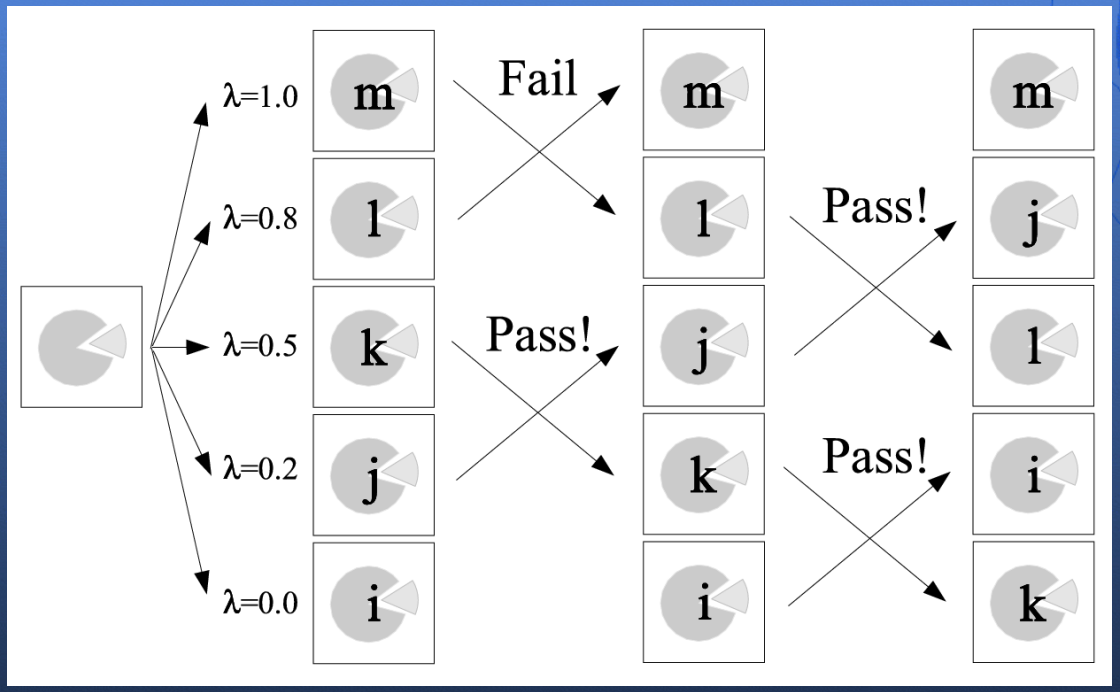
- Speed essential
- Statistical thermodynamics
  - Replica exchange
- Monte Carlo sampling
  - Flexible ligands and protein side chains
- Implicit solvent framework
  - GBSA parameterisation
  - Dual potential MC
- Large ligand differences
  - Dual topology

Predicting binding free energies



$$\Delta G_{bind}(B) - \Delta G_{bind}(A) = \Delta G_{bound}(A \rightarrow B) - \Delta G_{free}(A \rightarrow B)$$

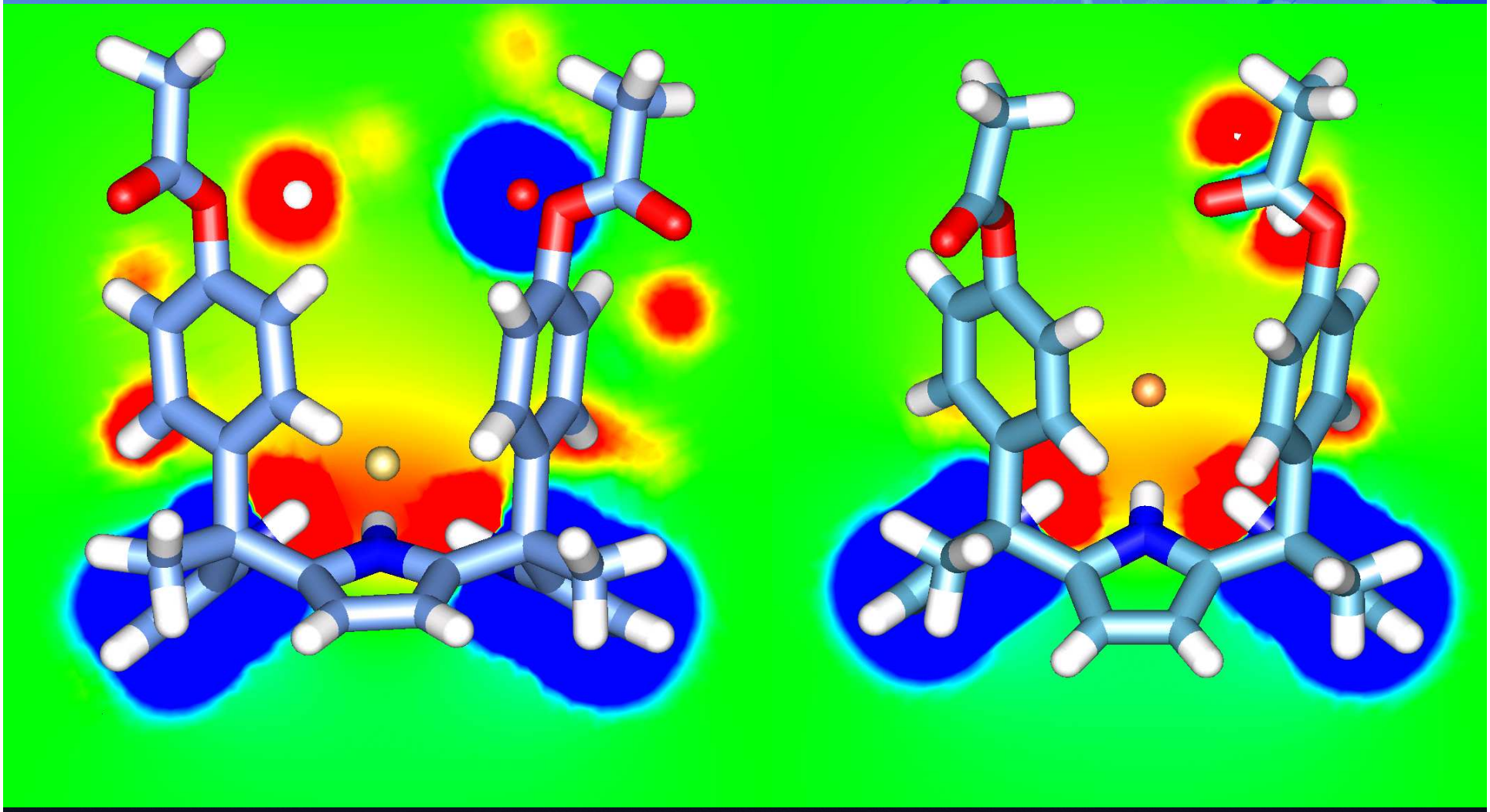
Replica-exchange free energy



$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

- Free energy methods that are applied between exchanges are the same as normal
- Exchanges require little extra computational cost

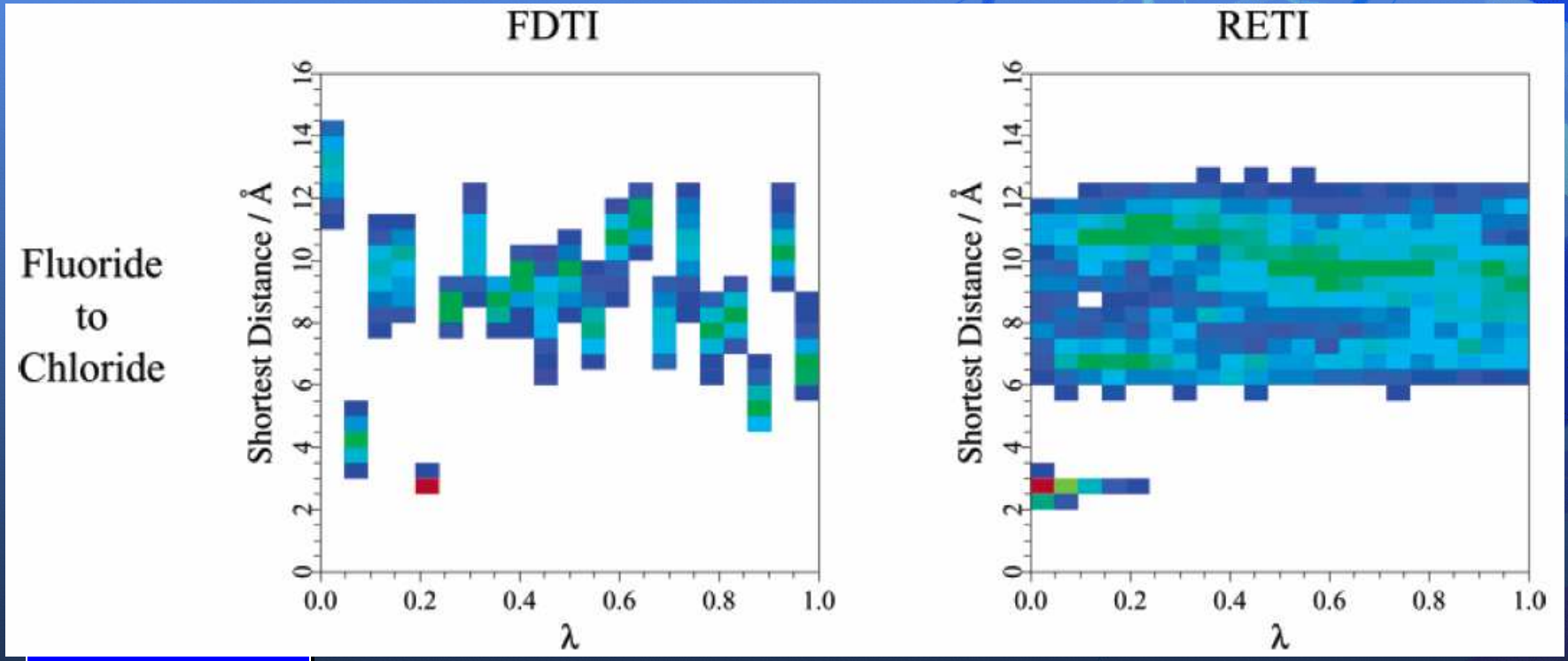
Calix[4]pyrrole/halide complexes



Fluoride bound structure

Chloride bound structure

Distances to contaminating water



Density

High

Medium

Low

RETI has weighted the available configurations across λ

J. Phys. Chem. B *107*, 13703-13710, **2003**

J. Phys. Chem. B *107*, 13711-13718, **2003**

Implicit solvent model

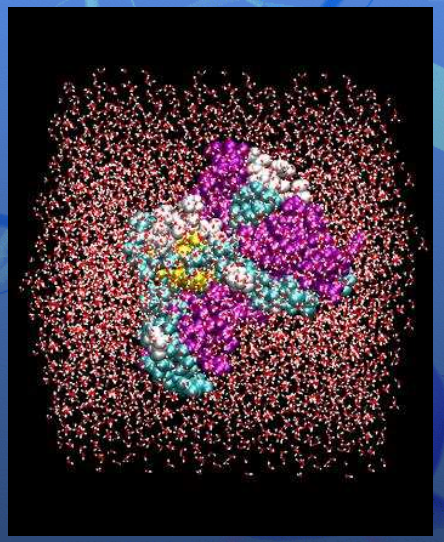
- Generalised Born
  - Pairwise Descreening Approximation
- Parameterisation
  - Dataset of small molecules

    AMBER/AM1-BCC
  - Optimisation

    Genetic Algorithm
  - Validation

    Cross-validation

    Potential of Mean Force calculations
- Approximations

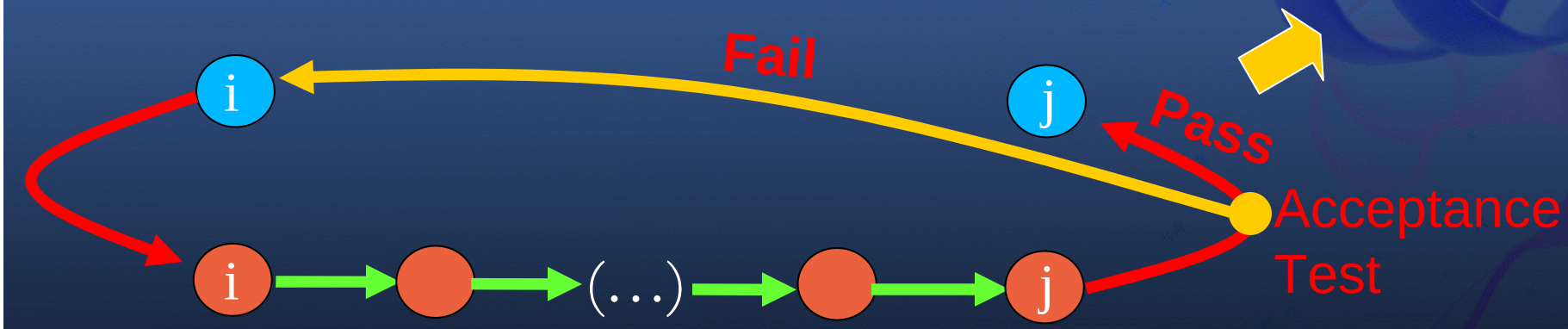


J. Comput. Chem. *25*, 1760-1770, **2004**

Simplified sampling potential

- Metropolis Monte Carlo : Random Walk

- Random Walk with two potentials



  - Sample N steps with the red potential
  - Add state j to the blue ensemble with probability

$$\chi = (\pi_j/\pi_i) * (\pi_i/\pi_j)$$

Simplified sampling potential: Application

- Reference Potential
  - $GBSA_{approx}$
    - GBSA with SA and approximated GB
- Two simplified sampling potentials
  - DDD
    - Distance Dependent Dielectric Model (no SA)
  - FastGB
    - GB model with low cutoffs (no SA)

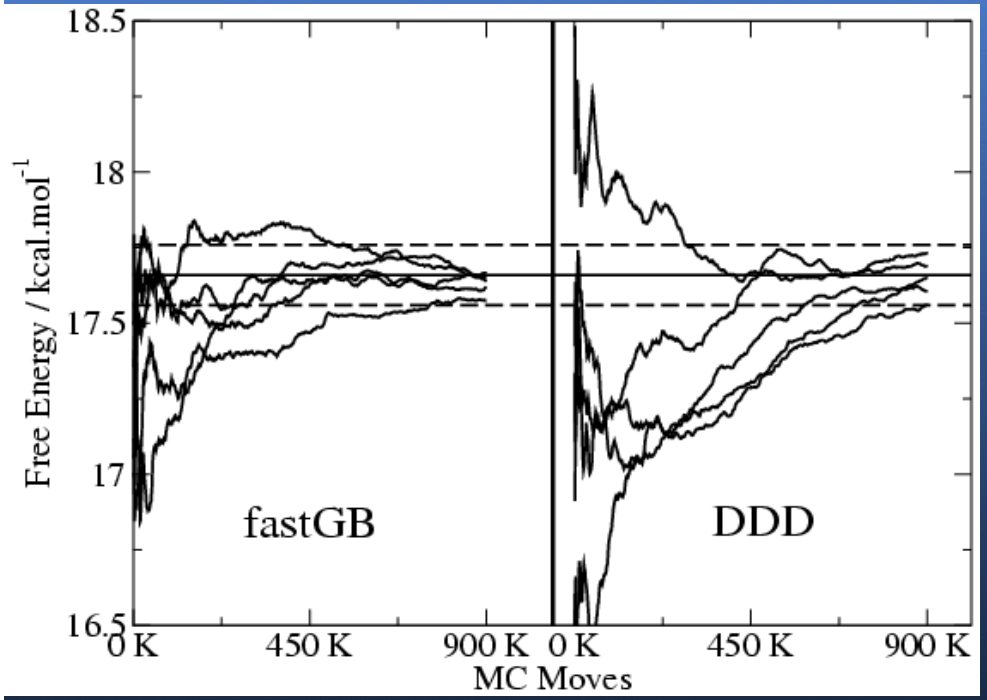
FastGB

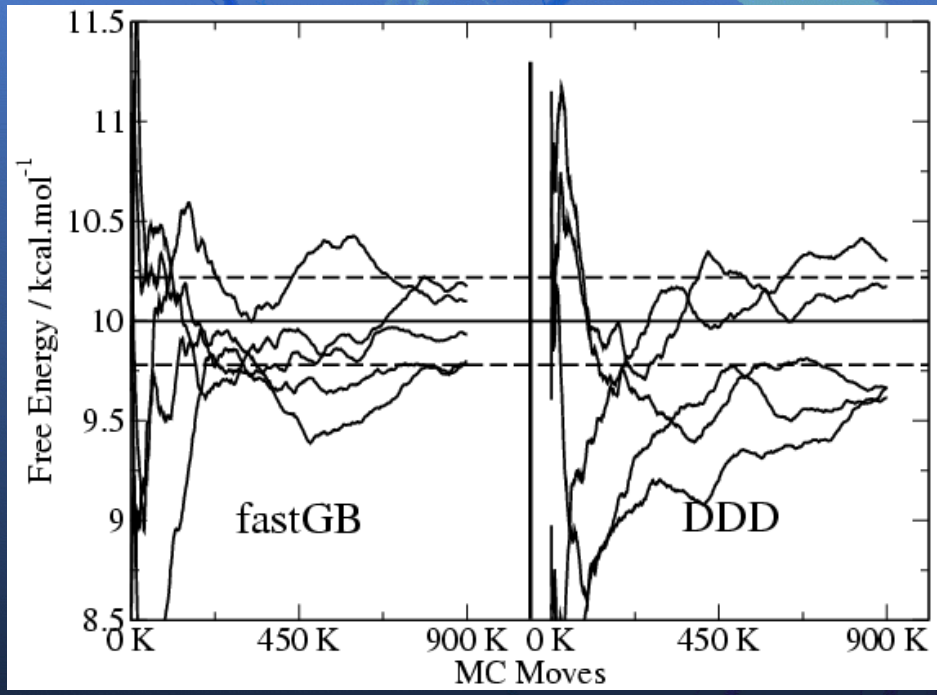
Speed up **x2.3**
Acceptance rate
at χ = **85%**

DDD

Speed up **x2.5**
Acceptance rate
at χ = **42%**

Simplified sampling potential: Results

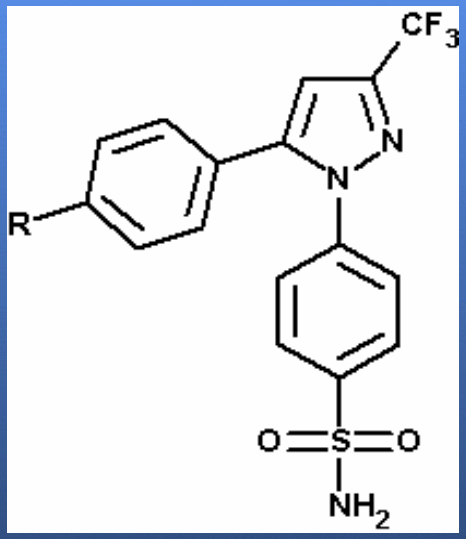




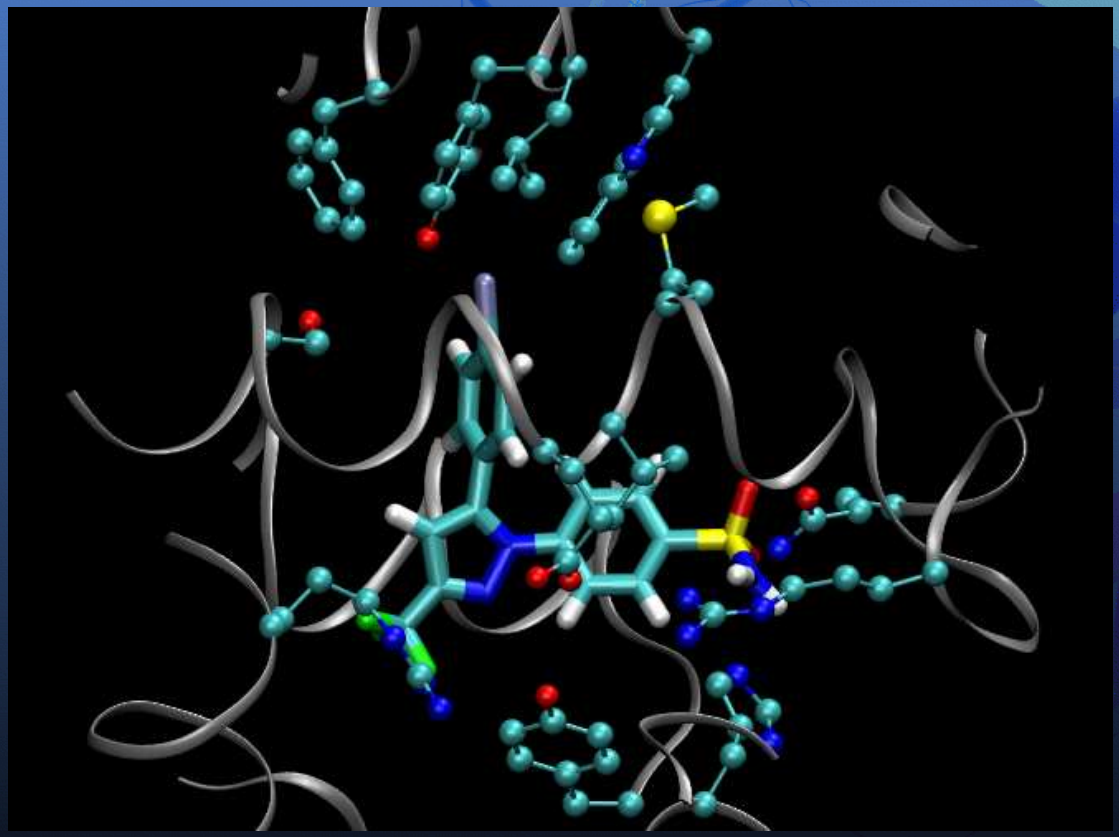
- Faster convergence of fastGB/GBSA$_{approx}$ over DDD/GBSA$_{approx}$

J. Chem. Theory Comput. *2*, 732-739, **2006**

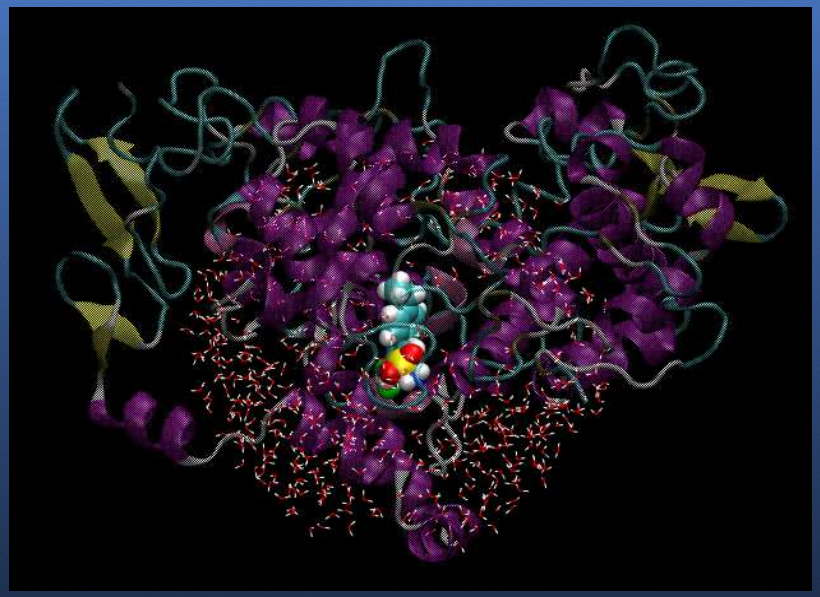
Case study: COX2

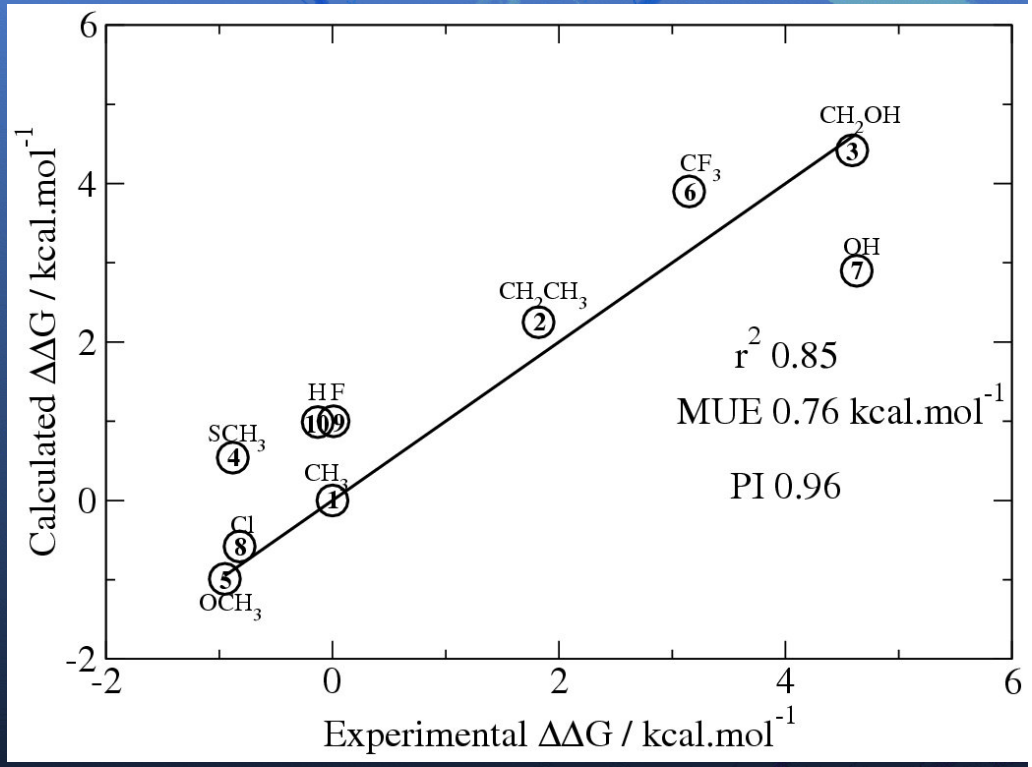


Compound	R	IC_{50} (μM)
1	CH_3	0.04
2	CH_2CH_3	0.86
3	CH_2OH	93.3
4	SCH_3	0.009
5	OCH_3	0.008
6	CF_3	8.23
7	OH	>100
8	Cl	0.01
9	F	0.041
10	H	0.032



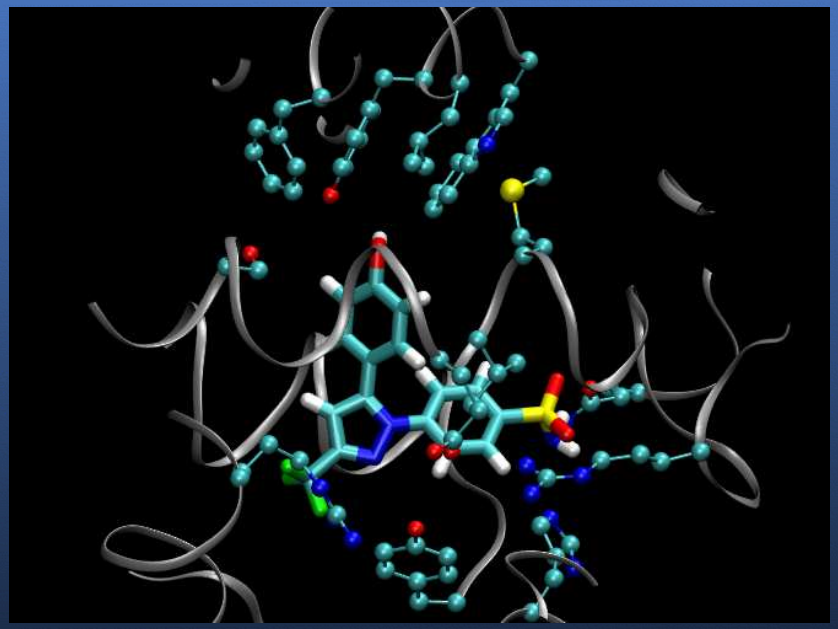
COX2: Explicit solvent results

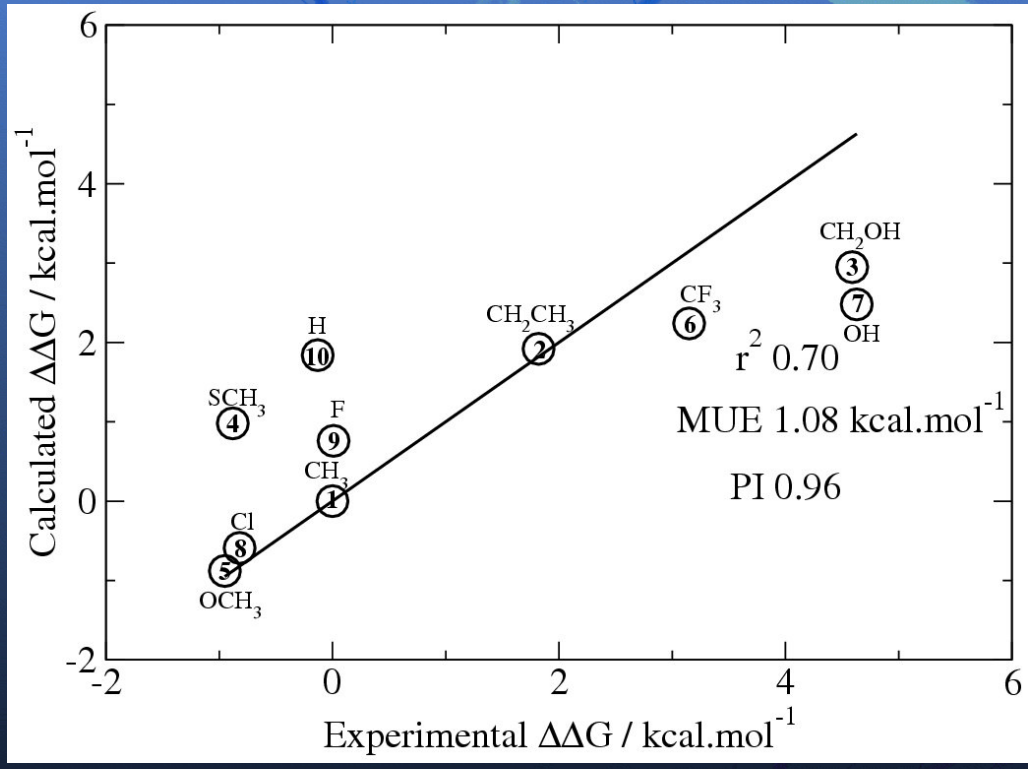




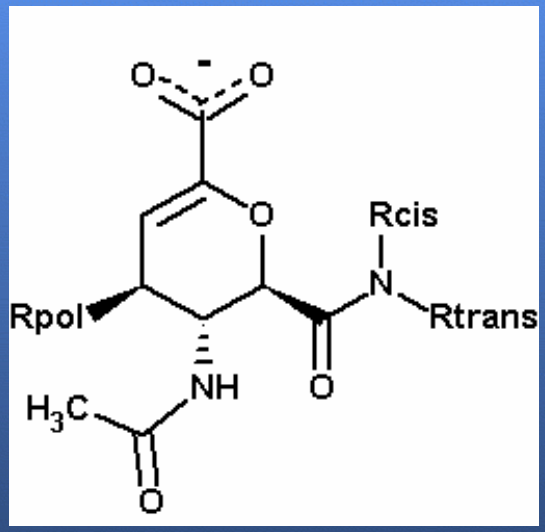
PI = Predictive index: quality of rank ordering

COX2: Implicit solvent results

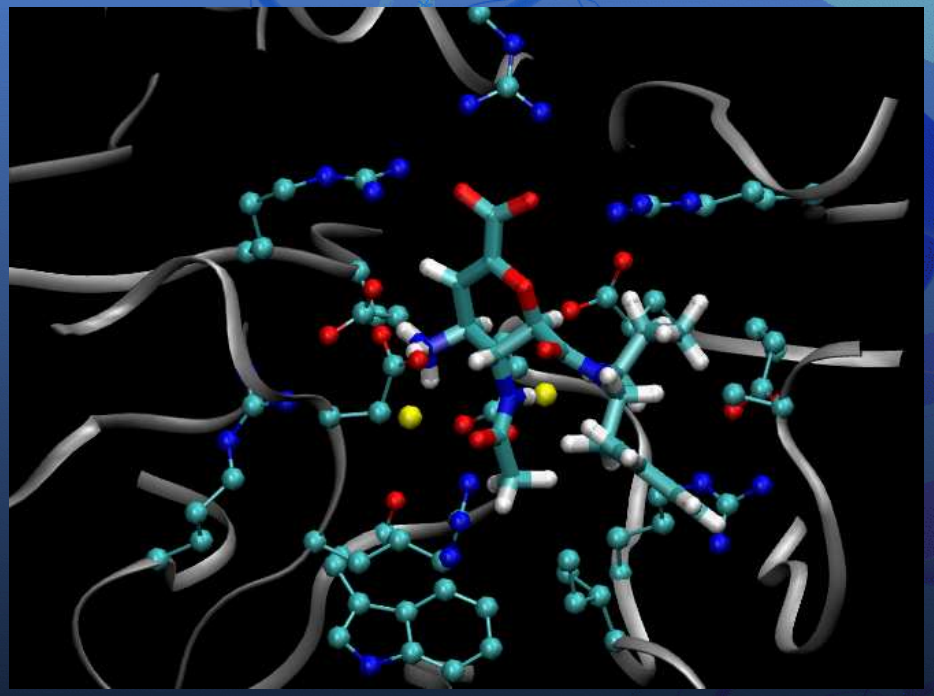




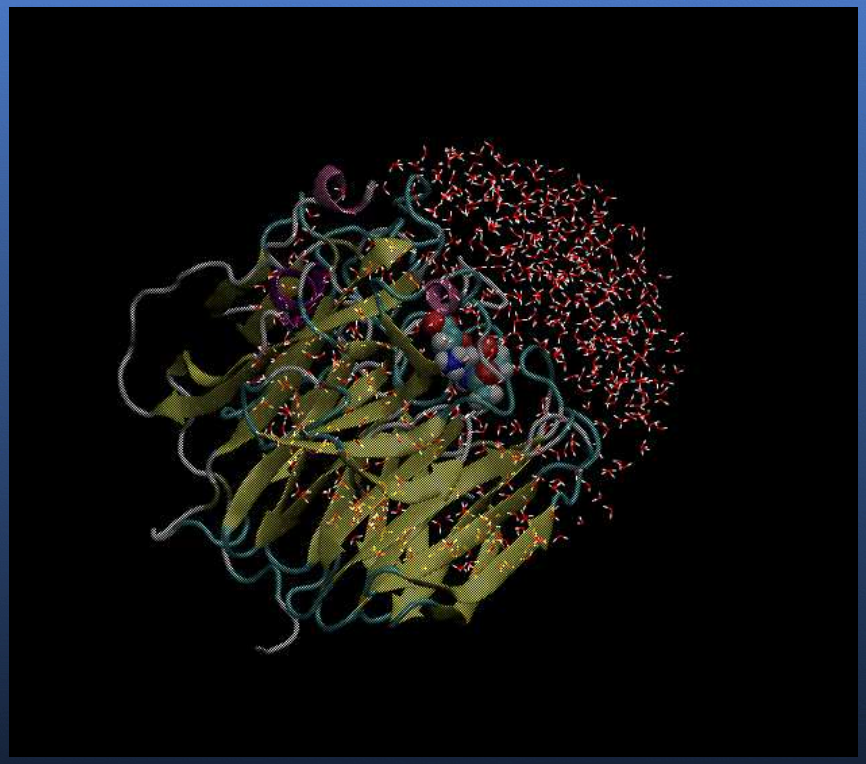
Case study: Neuraminidase

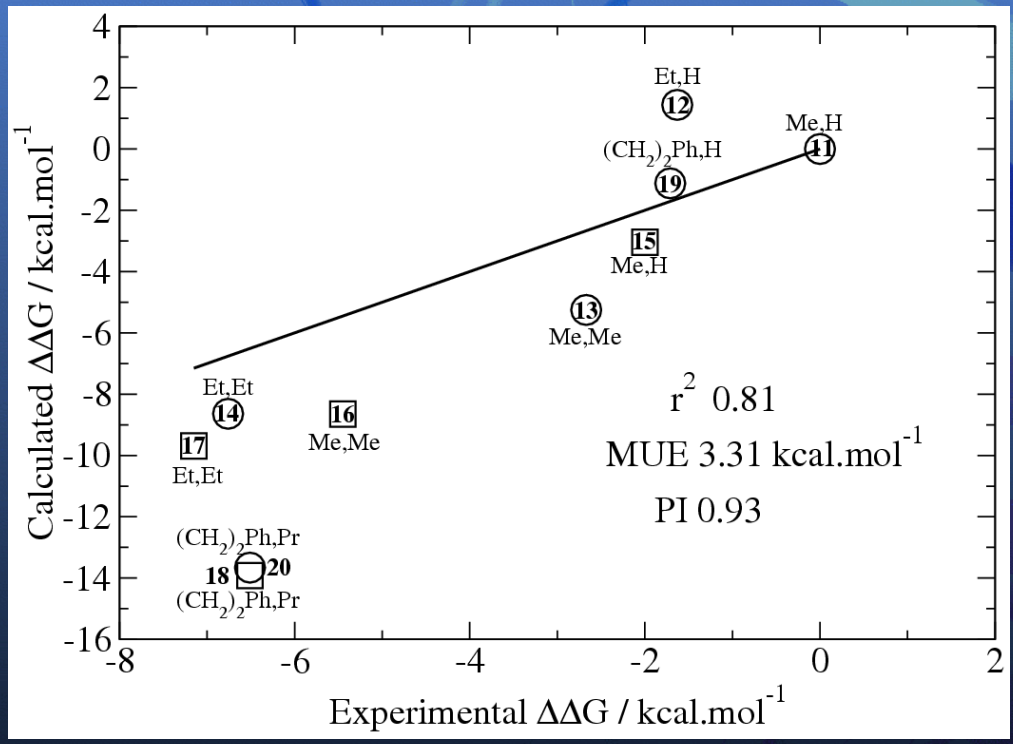


Compound	R_{trans}	R_{cis}	R_{pol}	IC_{50} (μM)
11	Me	H	NH_3+	190
12	Et	H	NH_3+	13
13	Me	Me	NH_3+	2.4
14	Et	Et	NH_3+	0.003
15	Me	H	NH_3+	7
16	Me	Me	$NHC(NH_2)_2$+	0.025
17	Et	Et	$NHC(NH_2)_2$+	0.001
18	$(CH2)_2Ph$	Pr	$NHC(NH_2)_2$+	0.005
19	$(CH2)_2Ph$	H	$NHC(NH_2)_2$+	12
20	$(CH2)_2Ph$	Pr	$NHC(NH_2)_2$+	0.005

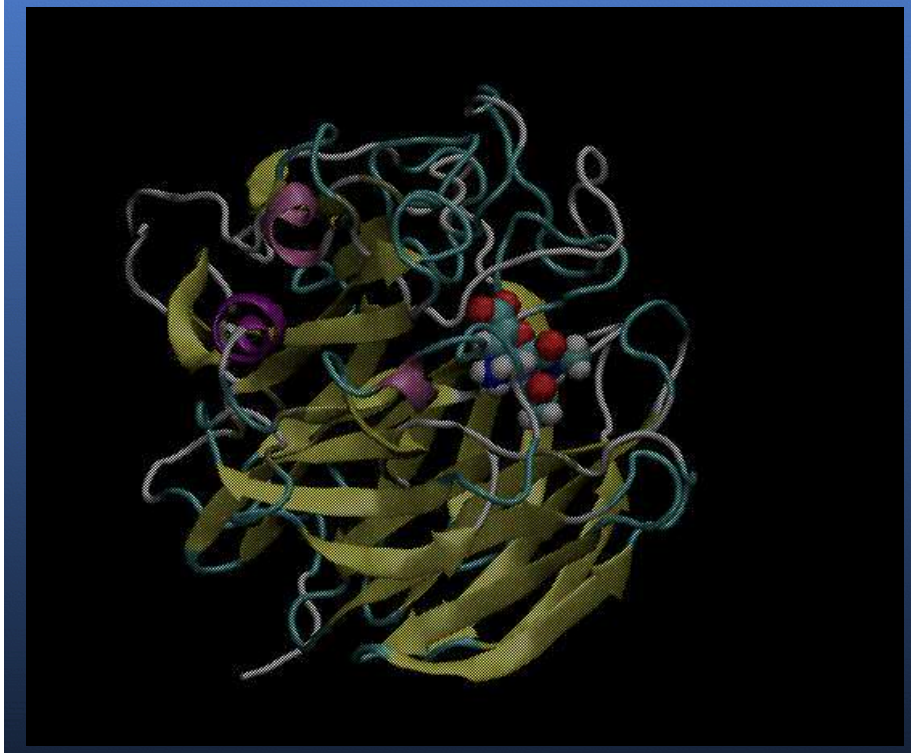


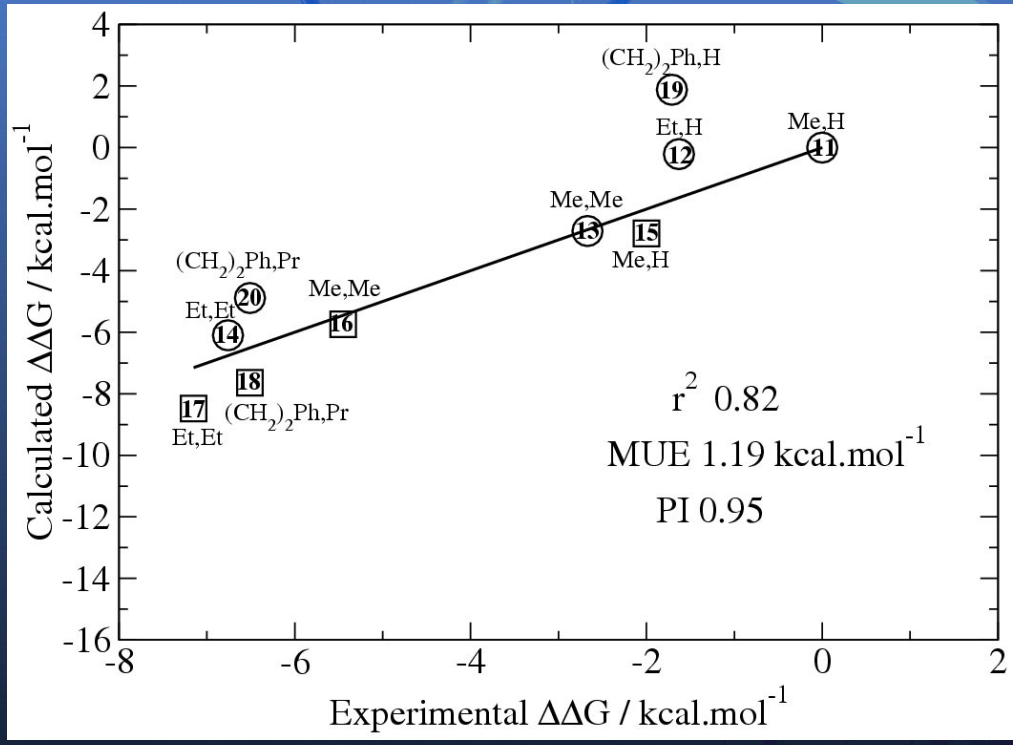
Neuraminidase: Explicit solvent results





Neuraminidase: Implicit solvent results



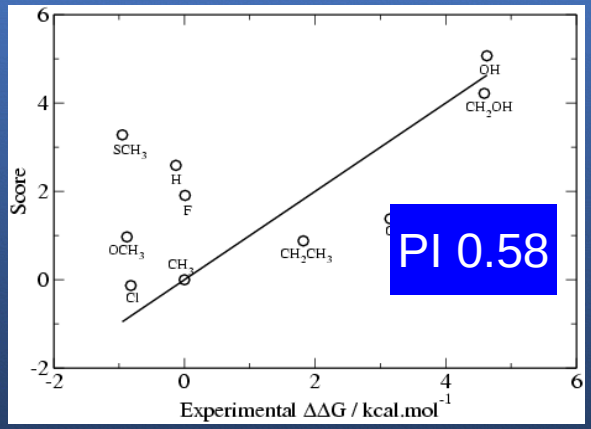


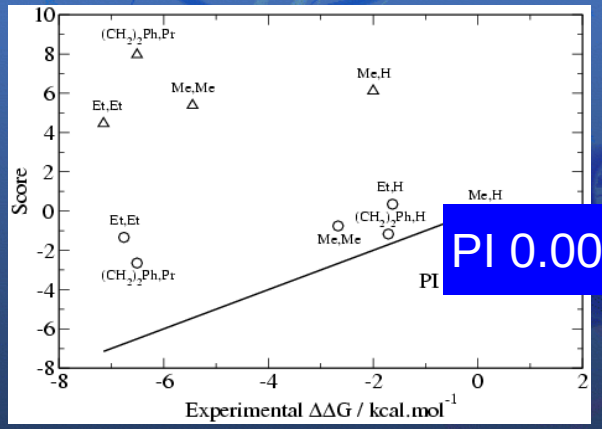
Ranking: Empirical scoring functions

COX2

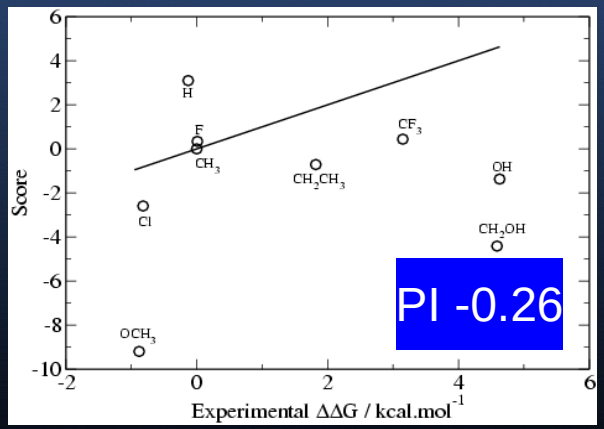
Neuraminidase

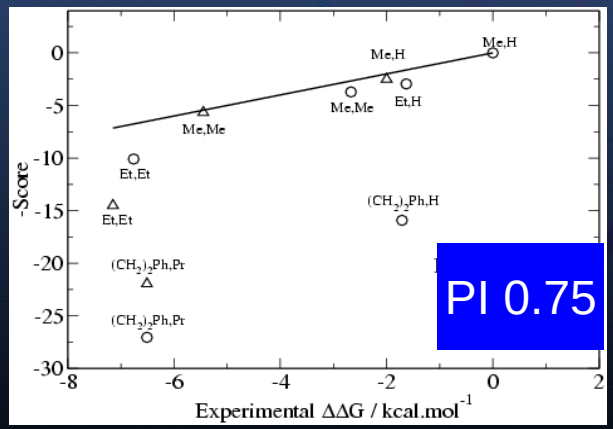
Chemscore





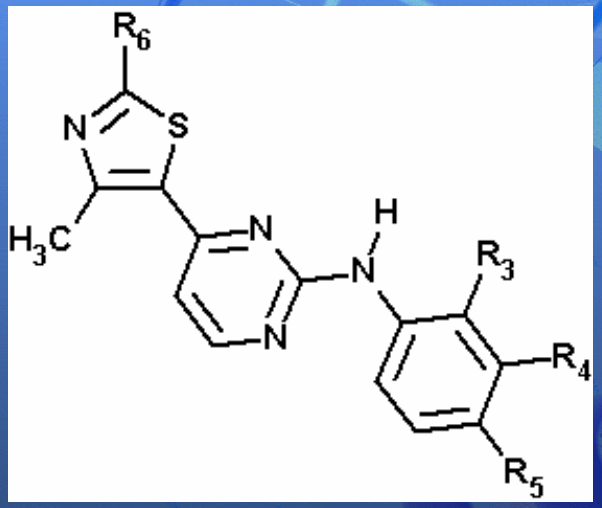
Goldscore

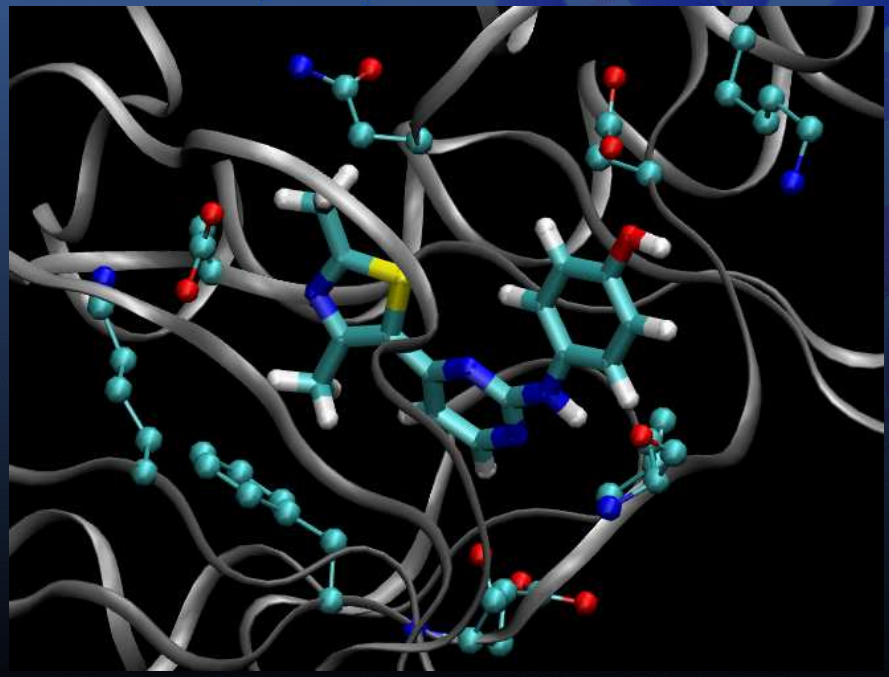




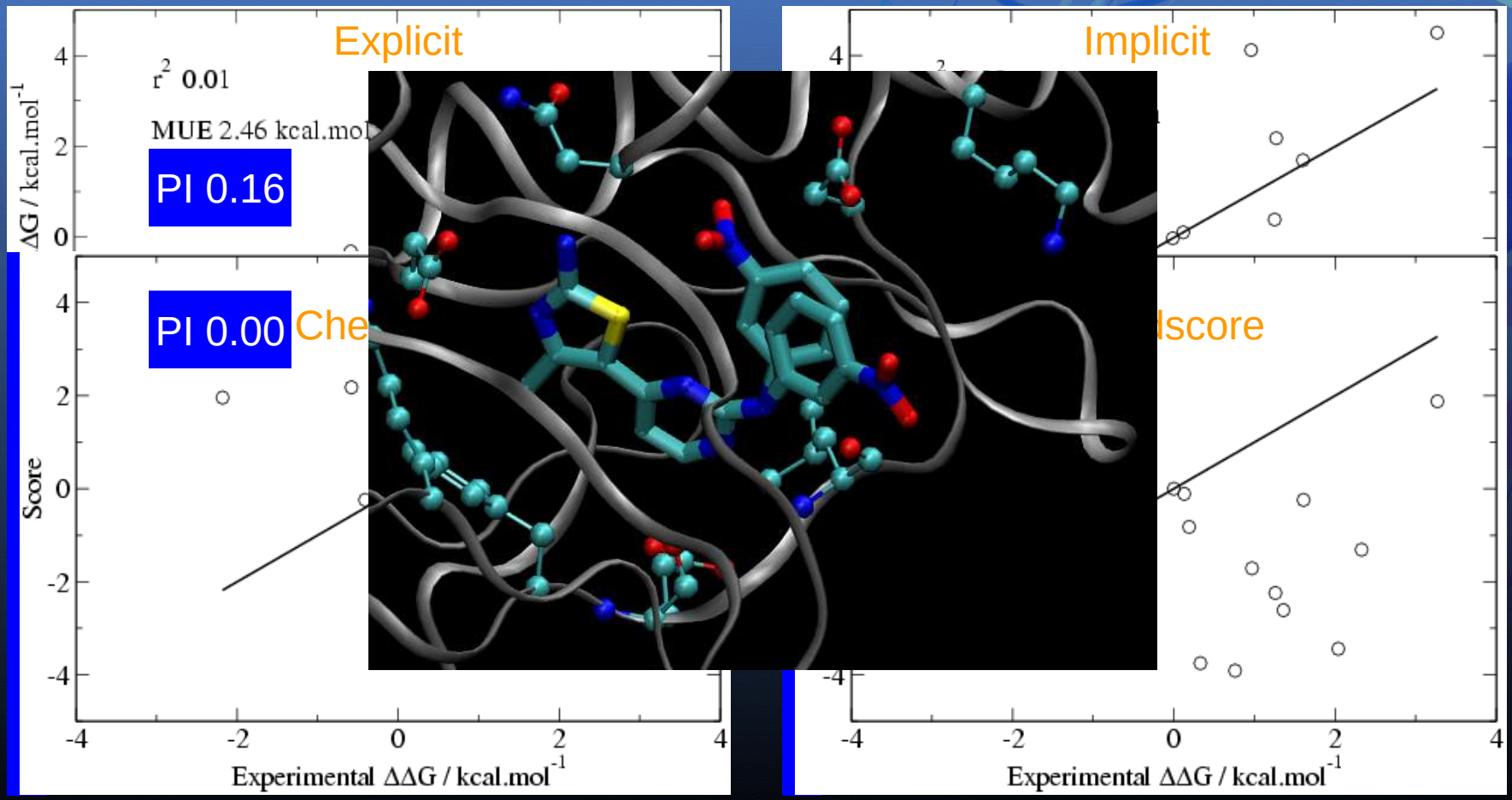
Case study: CDK2

Compound	R_3	R_4	R_5	R_6	IC_{50} (μM)
21	H	H	H	Me	0.08
22	Cl	H	H	Me	>20
23	H	Cl	H	Me	0.67
24	H	H	Cl	Me	2.5
25	F	H	H	Me	1.2
26	H	F	H	Me	0.1
27	H	H	F	Me	0.04
28	H	H	CF3	Me	0.29
29	H	OH	H	Me	0.06
30	H	H	OH	Me	0.14
31	H	H	OH	NHMe	0.07
32	H	H	OH	NH2	0.03
33	H	NO2	H	Me	0.11
34	H	NO2	H	NHMe	0.8
35	H	NO2	H	NH2	0.002
36	H	H	NO2	Me	4.1
37	H	NH2	H	Me	0.4
38	H	H	H	NMe2	0.7



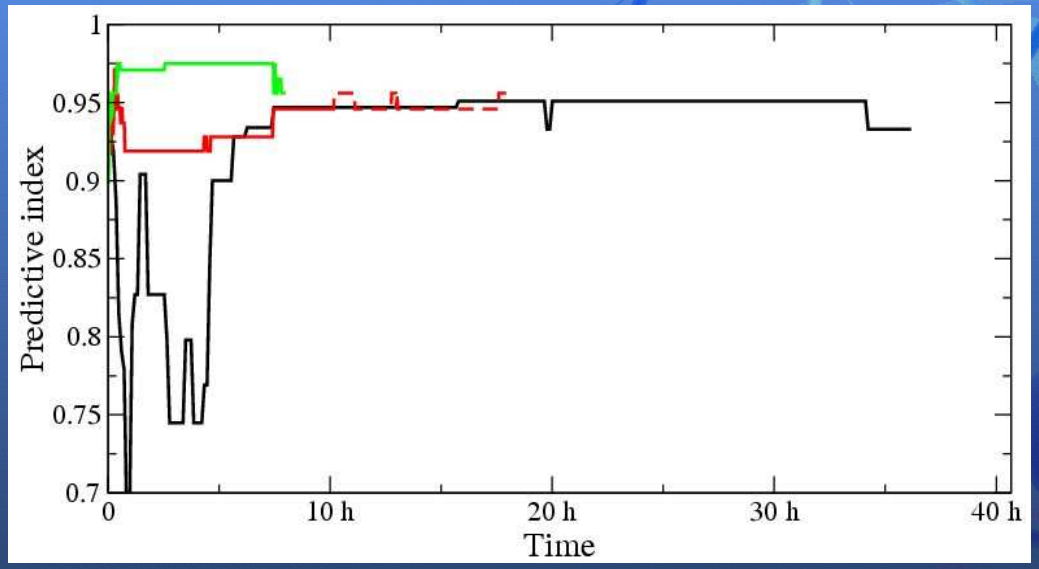


CDK2: Results

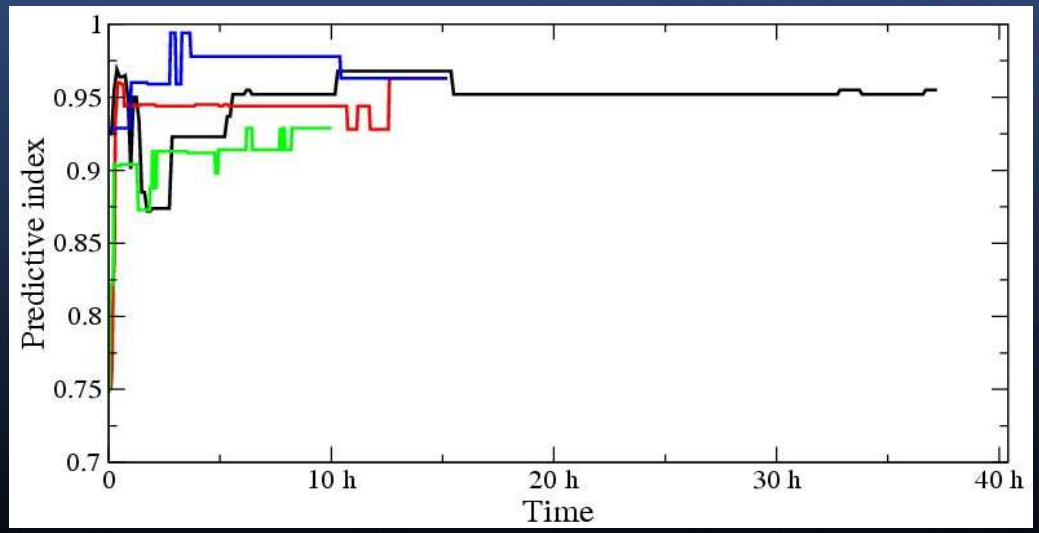


Convergence

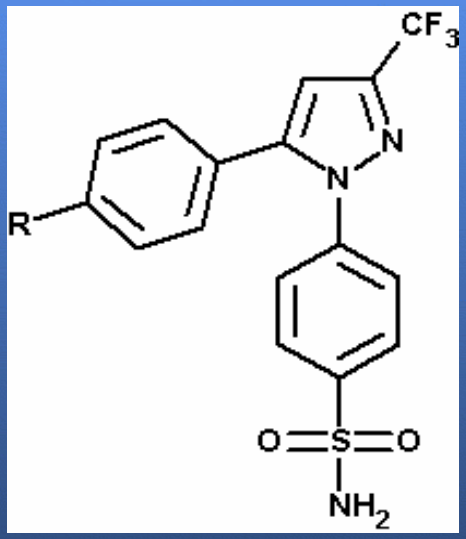
Neuraminidase



COX2



Free energy calculations in drug design



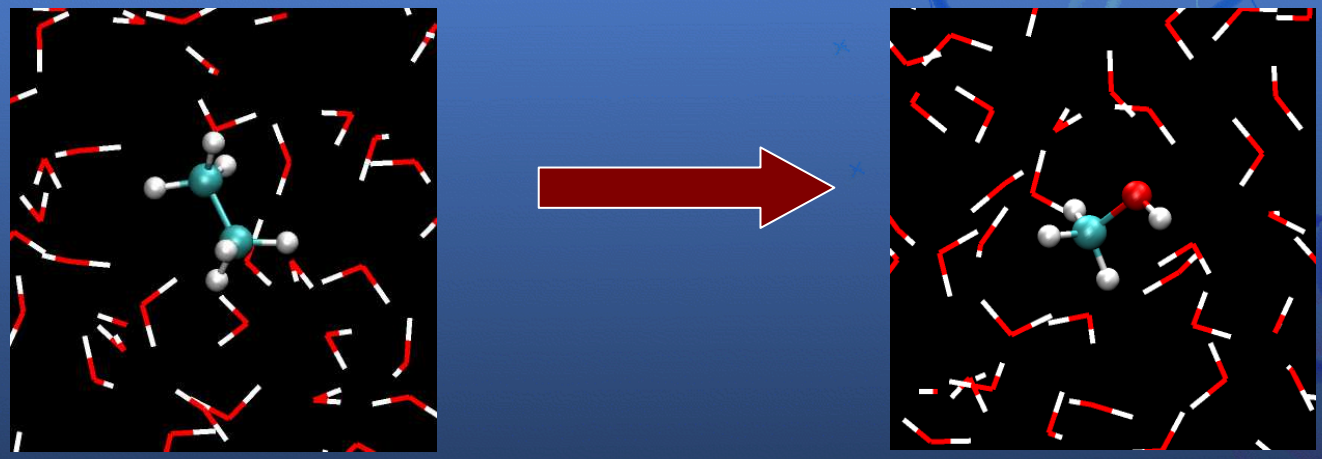
J. Med. Chem., *49*, 7427-7439, **2006**

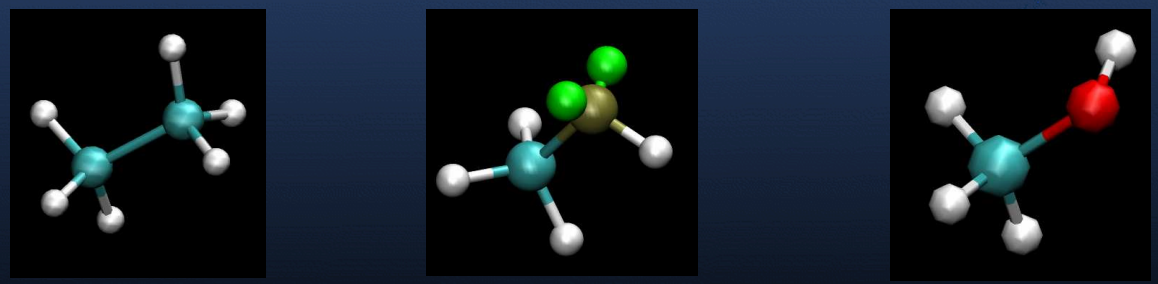
Compound	R	IC_{50} (μM)
1	CH_3	0.04
2	CH_2CH_3	0.86
3	CH_2OH	93.3
4	SCH_3	0.009
5	OCH_3	0.008
6	CF_3	8.23
7	OH	>100
8	Cl	0.01
9	F	0.041
10	H	0.032

- Restricted to similar (congeneric) ligands
- Fast, but fast enough?

A general methodology that can handle structurally diverse ligands would be very useful

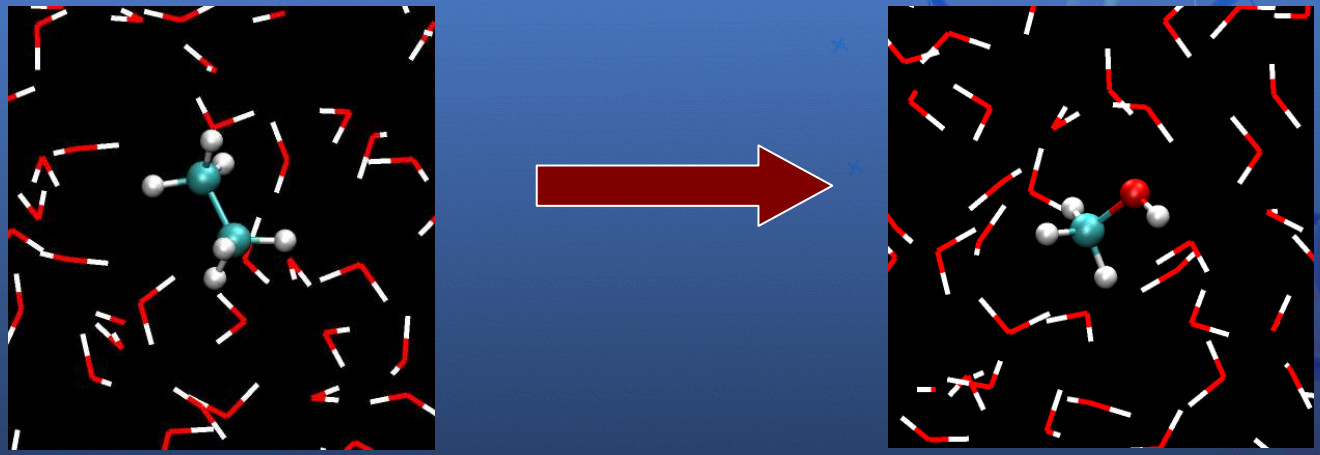
Ethane to methanol: single topology

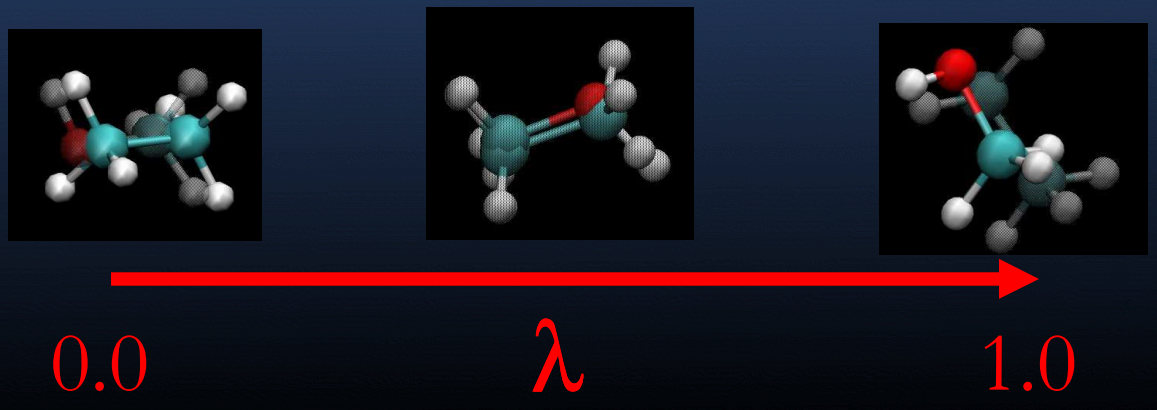




0.0 λ 1.0

Ethane to methanol: dual topology





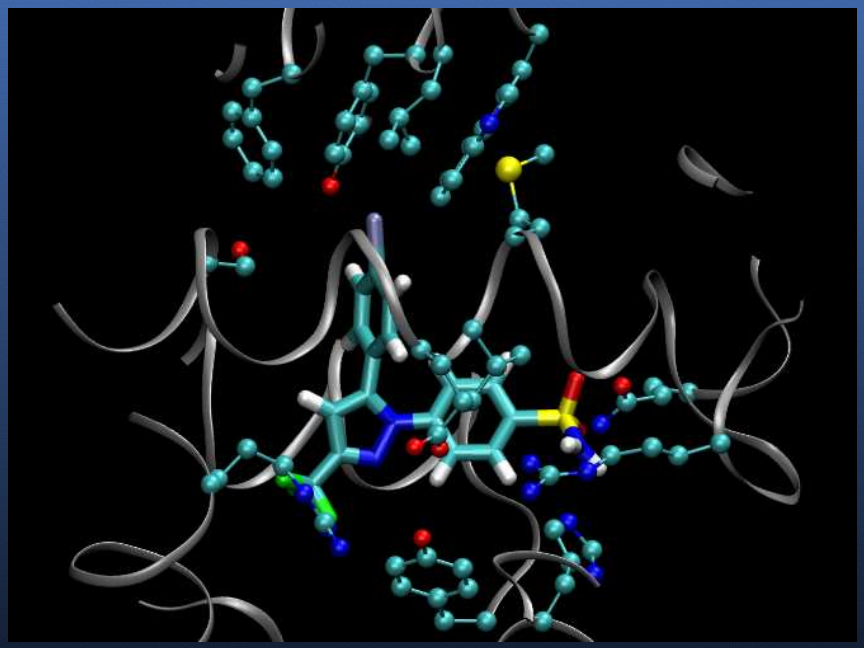
Dual topology: implementation

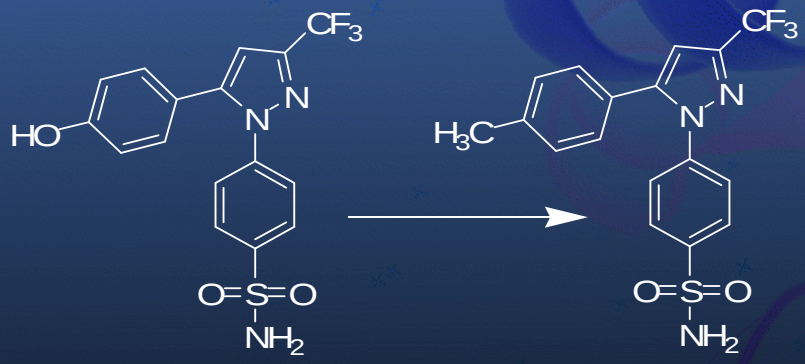
- Complete dual topology
  - Perturb molecules with no common structure
- Flexible softcore energy function

$$U_{nonbonded,\lambda} = (1-\lambda)4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}^{12}}{\left(\lambda\delta\sigma_{ij}+r_{ij}^{2}\right)^{6}}\right)-\left(\frac{\sigma_{ij}^{6}}{\left(\lambda\delta\sigma_{ij}+r_{ij}^{2}\right)^{3}}\right)\right]+\frac{(1-\lambda)^{n}q_{i}q_{j}}{4\pi\varepsilon_{0}\sqrt{\left(\lambda+r_{ij}^{2}\right)}}$$

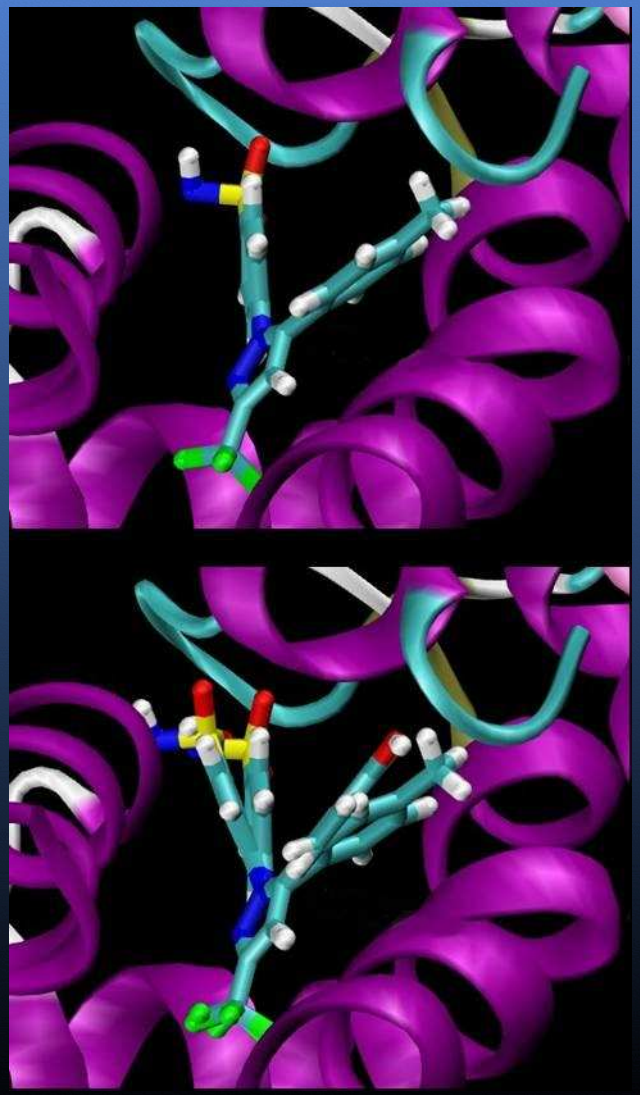
- Coupling of solutes' translation and rotation
  - Vanishing solute does not drift away
- Intramolecular terms not coupled
  - Decoupled solute is transferred to the gas phase

Congeneric inhibitors of COX2





COX2 inhibitors: results



$\Delta\Delta G_{bind}$ in kcal.mol^{-1}

Protocol	$\langle\Delta\Delta G\rangle$	S_E
dual topology	-2.7	0.6
single topology	-3.0	0.1

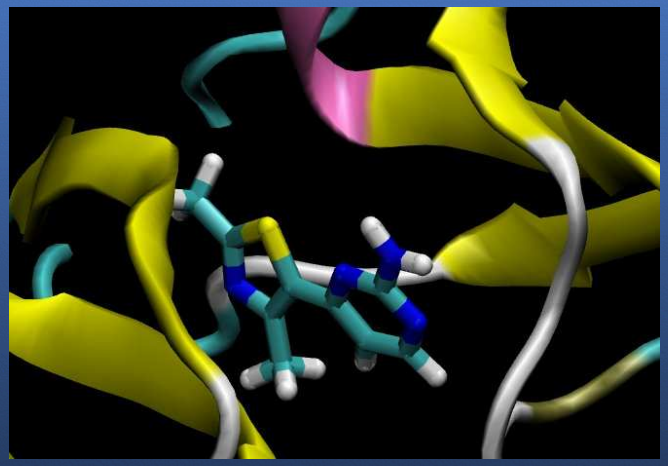
$\Delta\Delta G_{solv}$ in kcal.mol^{-1}

Protocol	$\langle\Delta\Delta G\rangle$	S_E
dual topology	4.6	0.4
single topology	4.5	0.1

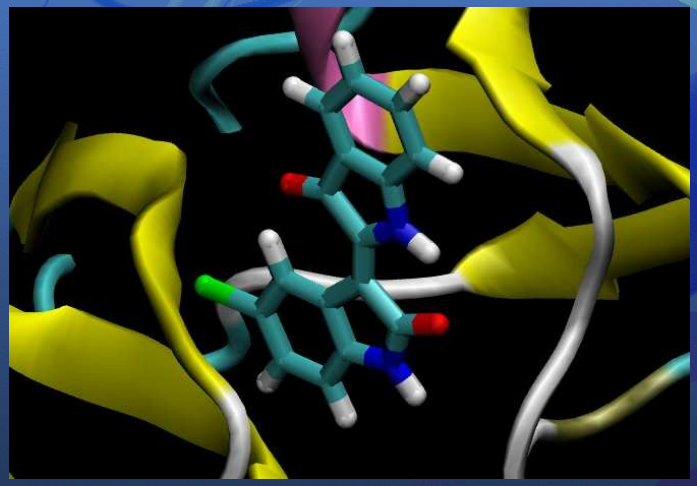
- Single topology simulations more precise
- When both methods are applicable, single topology preferred over dual topology

Two CDK2 scaffolds

activated CDK2 / CK hit

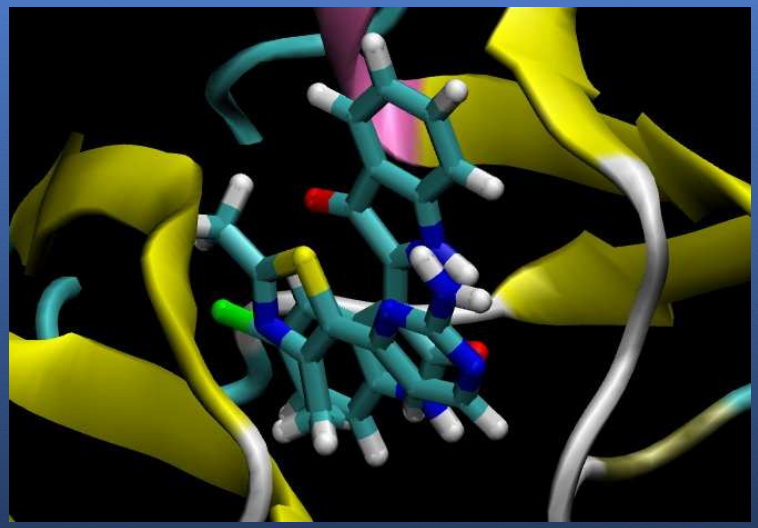


activated CDK2 / 5-bromoindirubine



- No common structural features
- Very difficult to handle by a single topology approach

ΔΔG of two CDK2 scaffolds: results



$\Delta\Delta G_{bind}$ / kcal.mol^{-1}

Protocol	<ΔΔG>	σ
in TIP4P	-0.48	1.12
in GBSA	-5.62	0.21

$\Delta\Delta G_{solv}$ / kcal.mol^{-1}

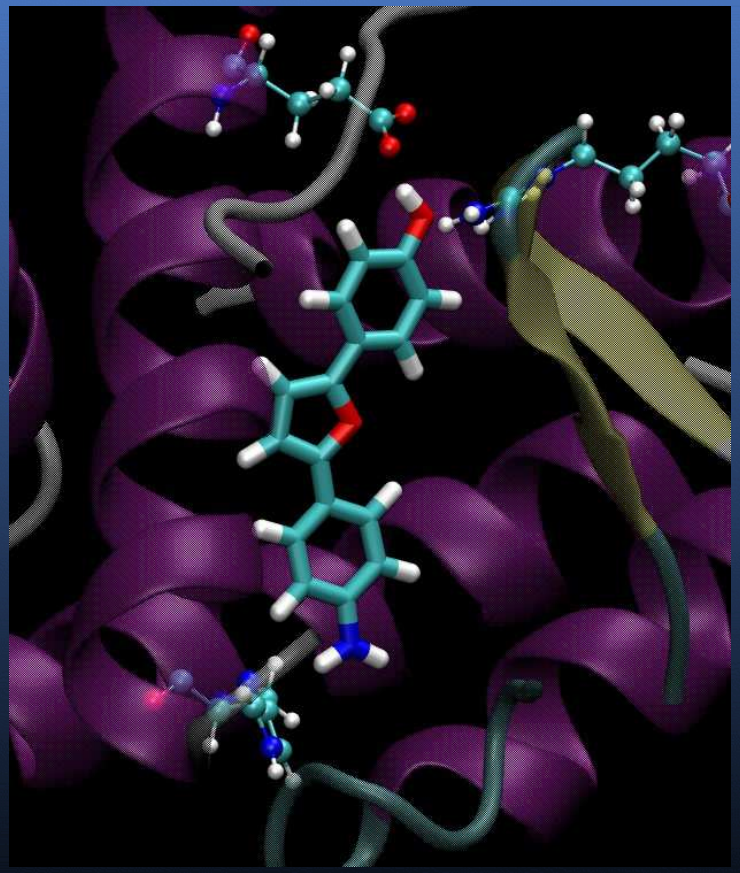
Protocol	<ΔΔG>	σ
in TIP4P	3.07	0.68
in GBSA	7.14	0.01

- Experiment?

J. Chem. Theory Comput., *3*, 1645-1655, **2007**

- Implicit solvent simulations more precise
- Sampling and force field

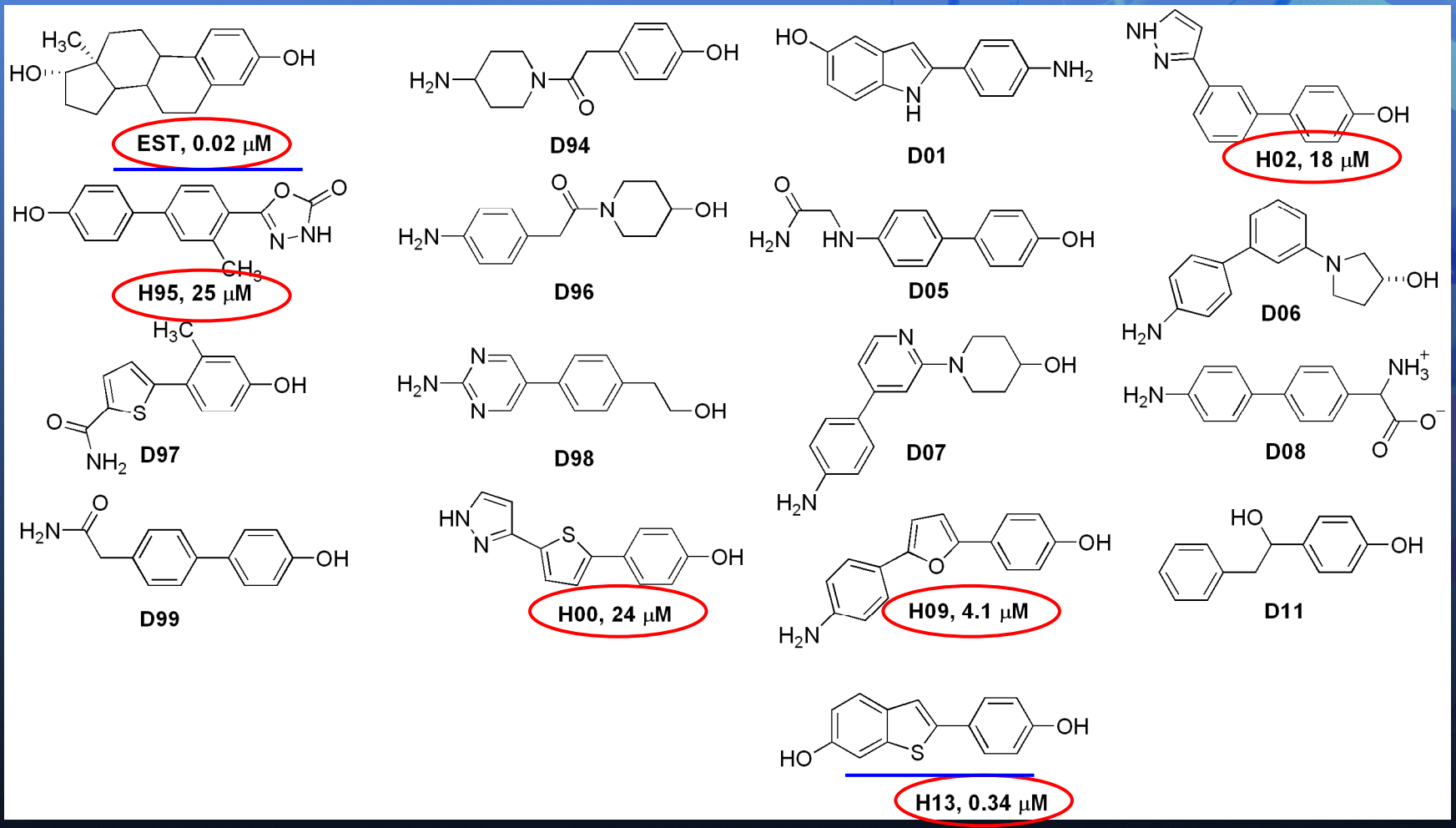
Case study: the estrogen receptor α



Aims

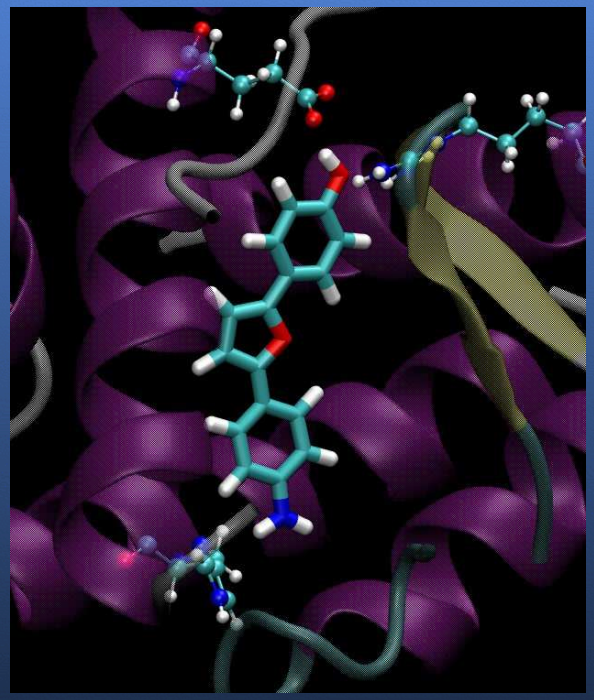
- Test the methodology
- Demonstrate **if**, **how** and **when** free energy simulation techniques can complement existing modelling tools for drug design purposes

Estrogen receptor α: Ligands



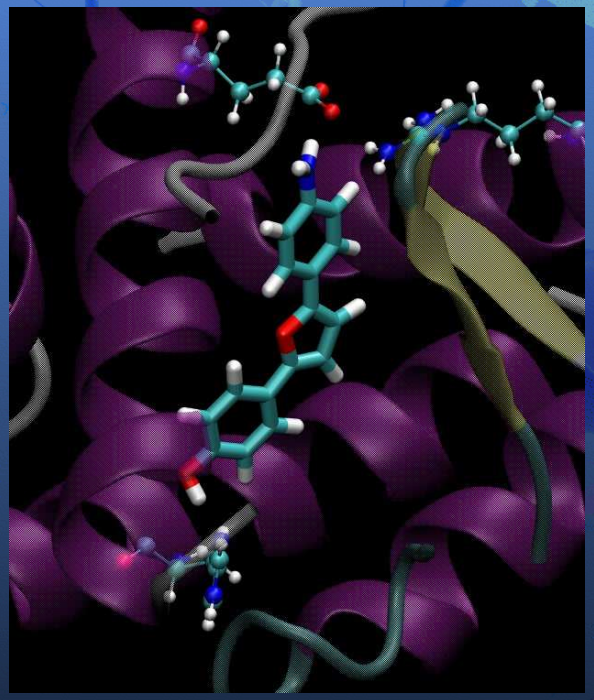
Firth-Clark et al., J. Chem. Inf. Model. *46*, 642-647, **2006**

Uncertainties in the binding mode



“top”

“bottom”



Binding mode A

Binding mode B

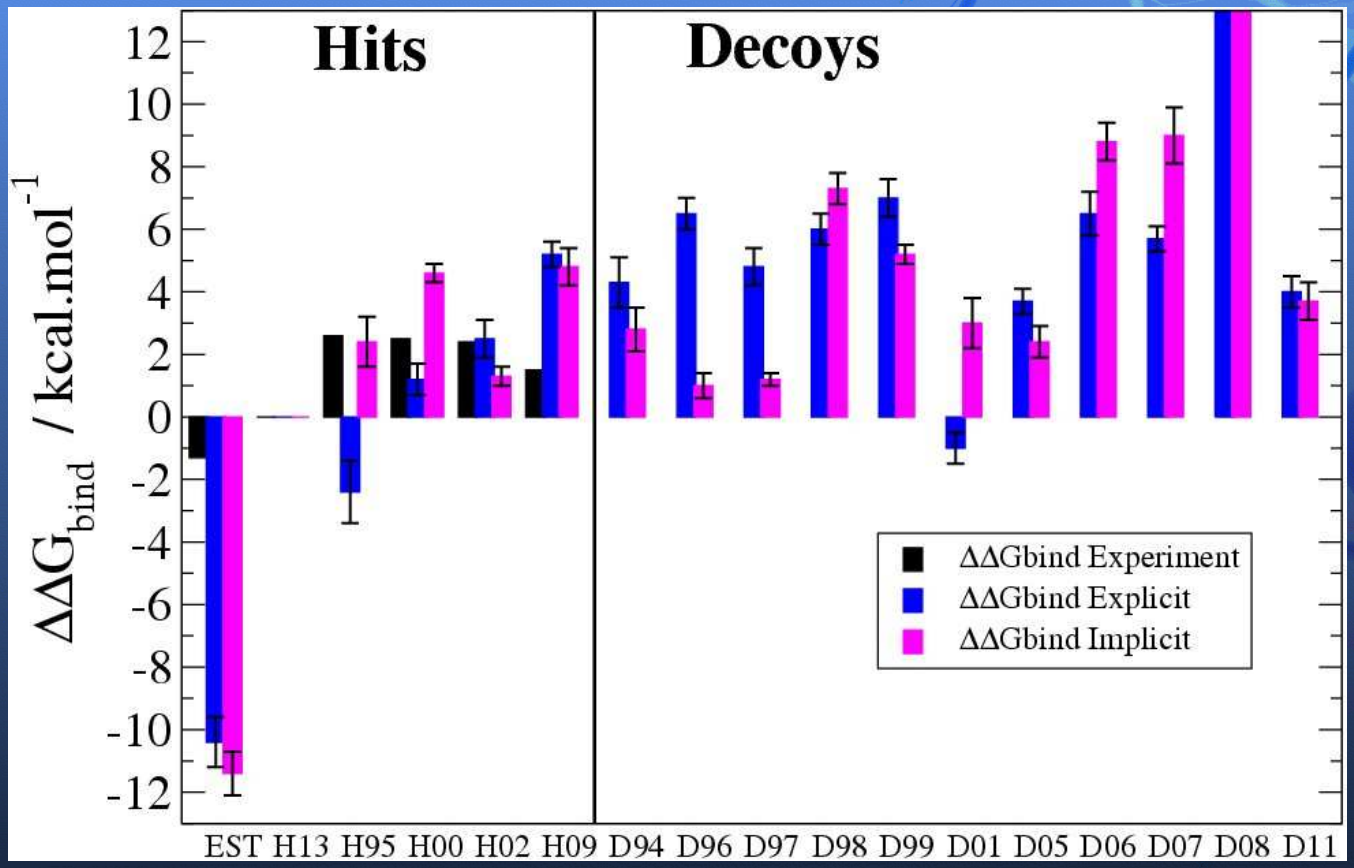
- Hydrogen donors/acceptors can be satisfied both ways
- Flips observed in some crystal structures

Binding mode predictions

Compound	GOLD	Explicit	Implicit
D94	top	top	top
H95	both	top	top
D96	top	bottom	bottom
D97	top	bottom	both
D98	top	bottom	bottom
D99	top	top	top
H00	top	top	top
D01	top	top	both
H02	top	bottom	bottom
D05	top	top	top
D06	top	bottom	bottom
D07	top	top	top
D08	top	top	top
H09	top	bottom	bottom
D11	top	top	top

- GOLD predicts (almost) always that the hydroxyl group will interact with Glu^{353}/Arg^{394}
- Alternative orientations are predicted favourable by free energy simulations 6 times
- Implicit solvent simulations suggest both orientations are possible for 2 compounds

Binding affinity predictions



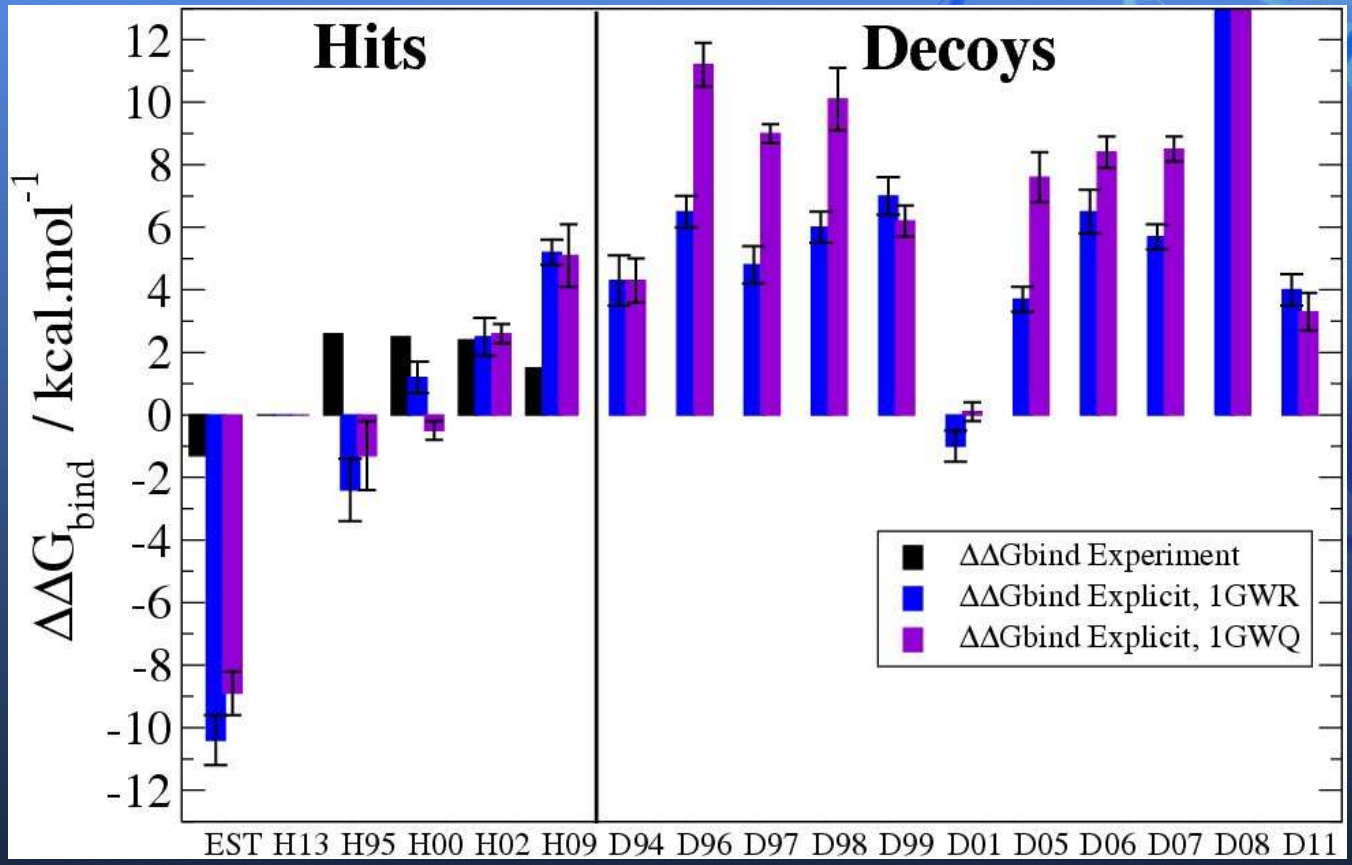
- Explicit solvent: 5 hits in top 6 ligands

  EST<**H95**<D01<**H13**<**H00**<**H02**<D05

- Implicit solvent: 3/4 hits in top 6 ligands

  EST<**H13**<D96~=**H02**~=D97<D05 ~=**H95**

Influence of protein structure



- Explicit solvent, 1GWR: 5 hits in top 6 ligands

EST<H95<D01<H13<H00<H02<D05

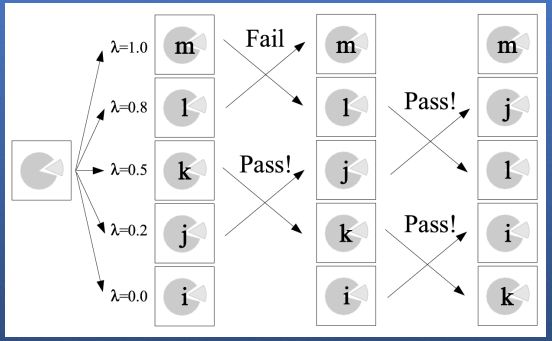
- Explicit solvent, 1GWQ: 5 hits in top 6 ligands

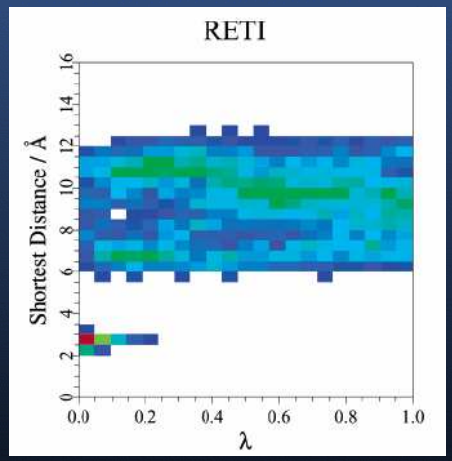
EST<H95<H00~=H13~=D01<H02<D11

J. Med. Chem.
in press

Summary

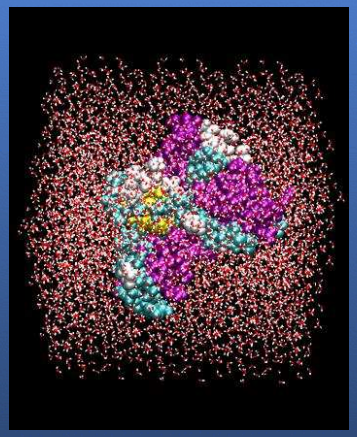
- A binding free energy calculation method
  - Replica exchange





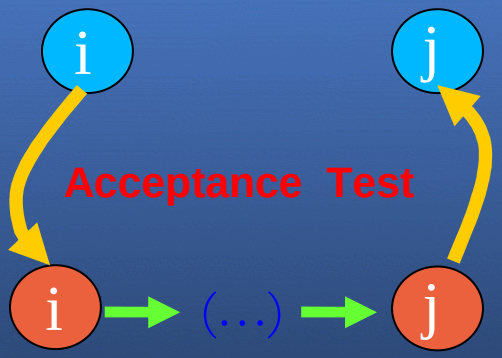
Summary

- A binding free energy calculation method
  - Replica exchange
  - Implicit solvent



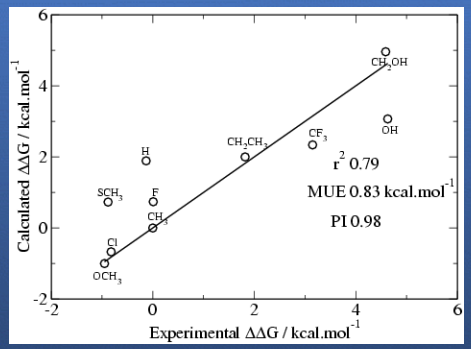
Summary

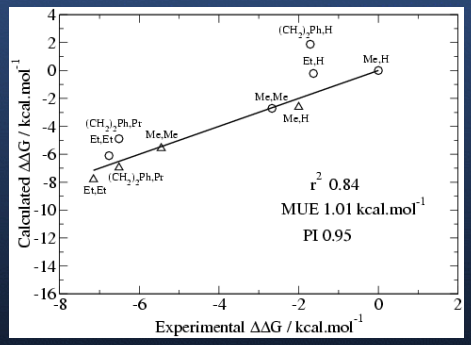
- A binding free energy calculation method
  - Replica exchange
  - Implicit solvent
  - Simplified sampling potential



Summary

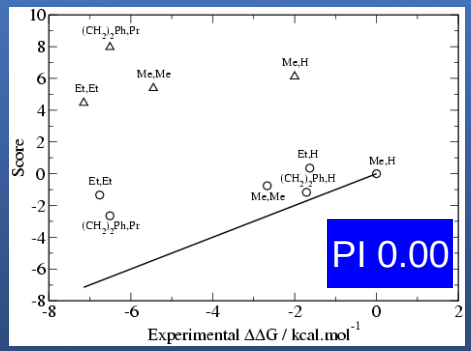
- A binding free energy calculation method
  - Replica exchange
  - Implicit solvent
  - Simplified sampling potential
- Case studies : COX2, Neuraminidase, CDK2, homologous ligands
  - Implicit solvent simulations of comparable or better accuracy than explicit solvent simulations

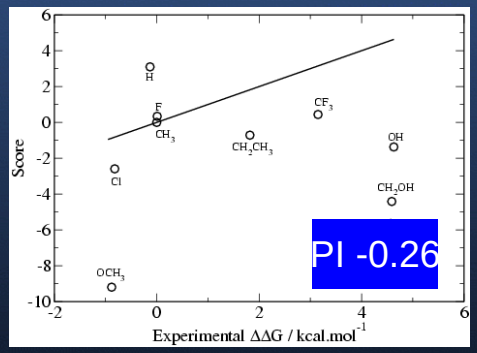




Summary

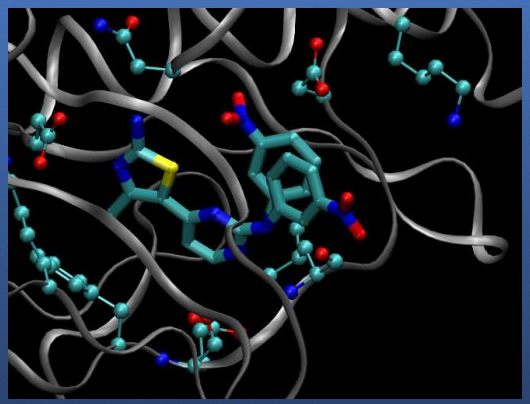
- A binding free energy calculation method
  - Replica exchange
  - Implicit solvent
  - Simplified sampling potential
- Case studies : COX2, Neuraminidase, CDK2, homologous ligands
  - Implicit solvent simulations of comparable or better accuracy than explicit solvent simulations
  - Better than common scoring functions

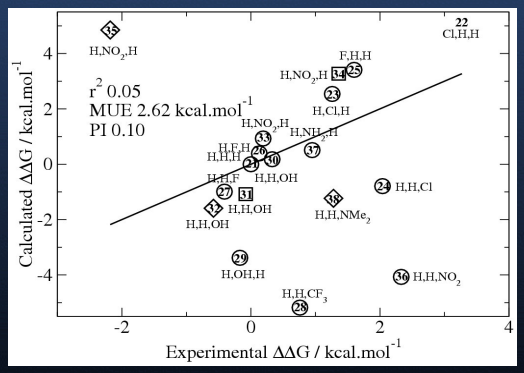




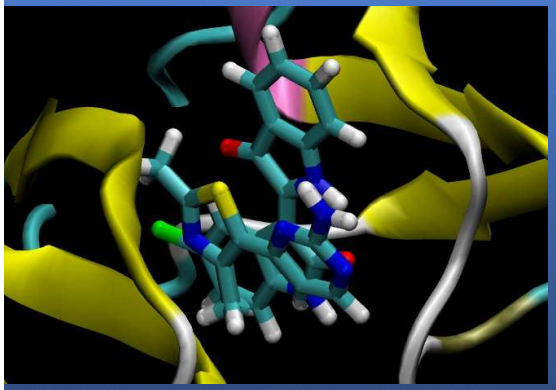
Summary

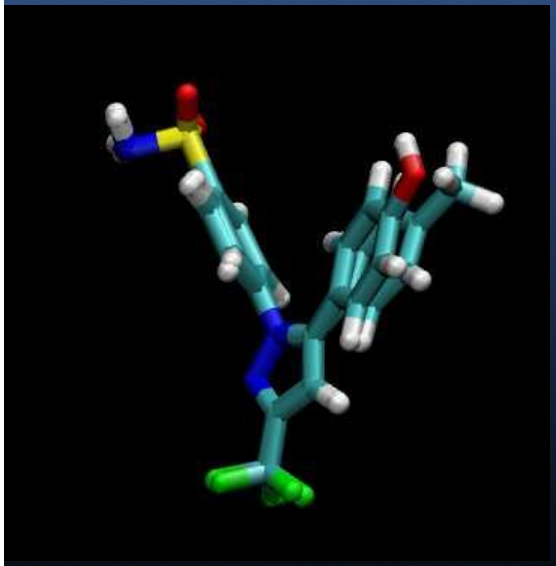
- A binding free energy calculation method
  - Replica exchange
  - Implicit solvent
  - Simplified sampling potential
- Case studies : COX2, Neuraminidase, CDK2, homologous ligands
  - Implicit solvent simulations of comparable or better accuracy than explicit solvent simulations
  - Better than common scoring functions
  - Problem systems





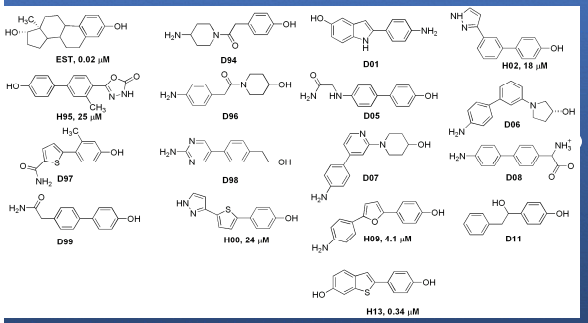
Summary

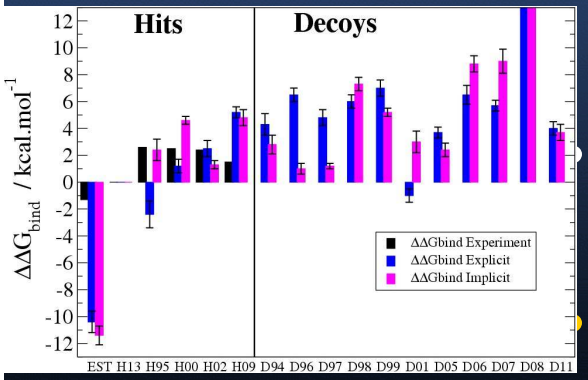




- A binding free energy calculation method
  - Replica exchange
  - Implicit solvent
  - Simplified sampling potential
- Case studies : COX2, Neuraminidase, CDK2, homologous ligands
  - Implicit solvent simulations of comparable or better accuracy than explicit solvent simulations
  - Better than common scoring functions
  - Problem systems
- Dual topology – very different molecules
  - Force field and sampling more critical – cost!

Summary





- A binding free energy calculation method
  - Replica exchange
  - Implicit solvent
  - Simplified sampling potential
- Case studies : COX2, Neuraminidase, CDK2, homologous ligands
  - Implicit solvent simulations of comparable or better accuracy than explicit solvent simulations
  - Better than common scoring functions
  - Problem systems
- Dual topology – very different molecules
  - Force field and sampling more critical – cost!
- Broader scope of the methodology
  - Binding mode prediction
  - Scaffold selection

Acknowledgements

- RETI
  - Chris Woods
  - Mike King, Celltech-UCB

- Julien Michel
- Richard Taylor, Marcel Verdonk, Chris Murray, Astex Therapeutics

- Astex Therapeutics
- EPSRC
- BBSRC