Large Scale Relative Ligand-Protein Binding Affinities Using Non-Equilibrium Alchemy

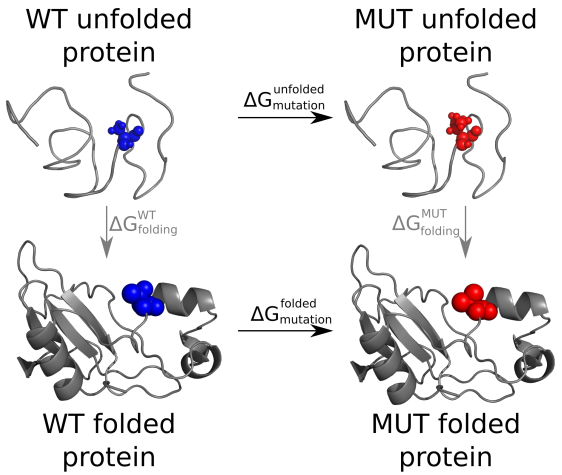
Vytautas Gapsys

Max Planck Institute for Biophysical Chemistry

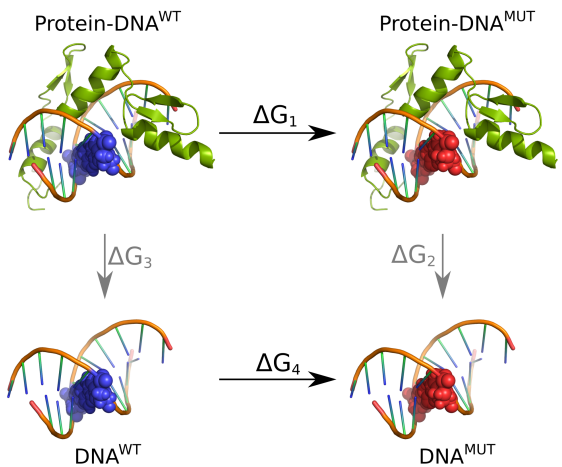
Computational Biomolecular Dynamics Group (Bert de Groot)

Aug 21, 2019

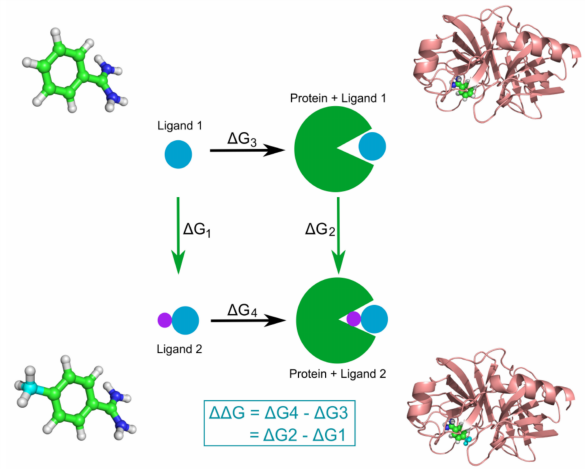
pmx based alchemistry



Amino acid mutations

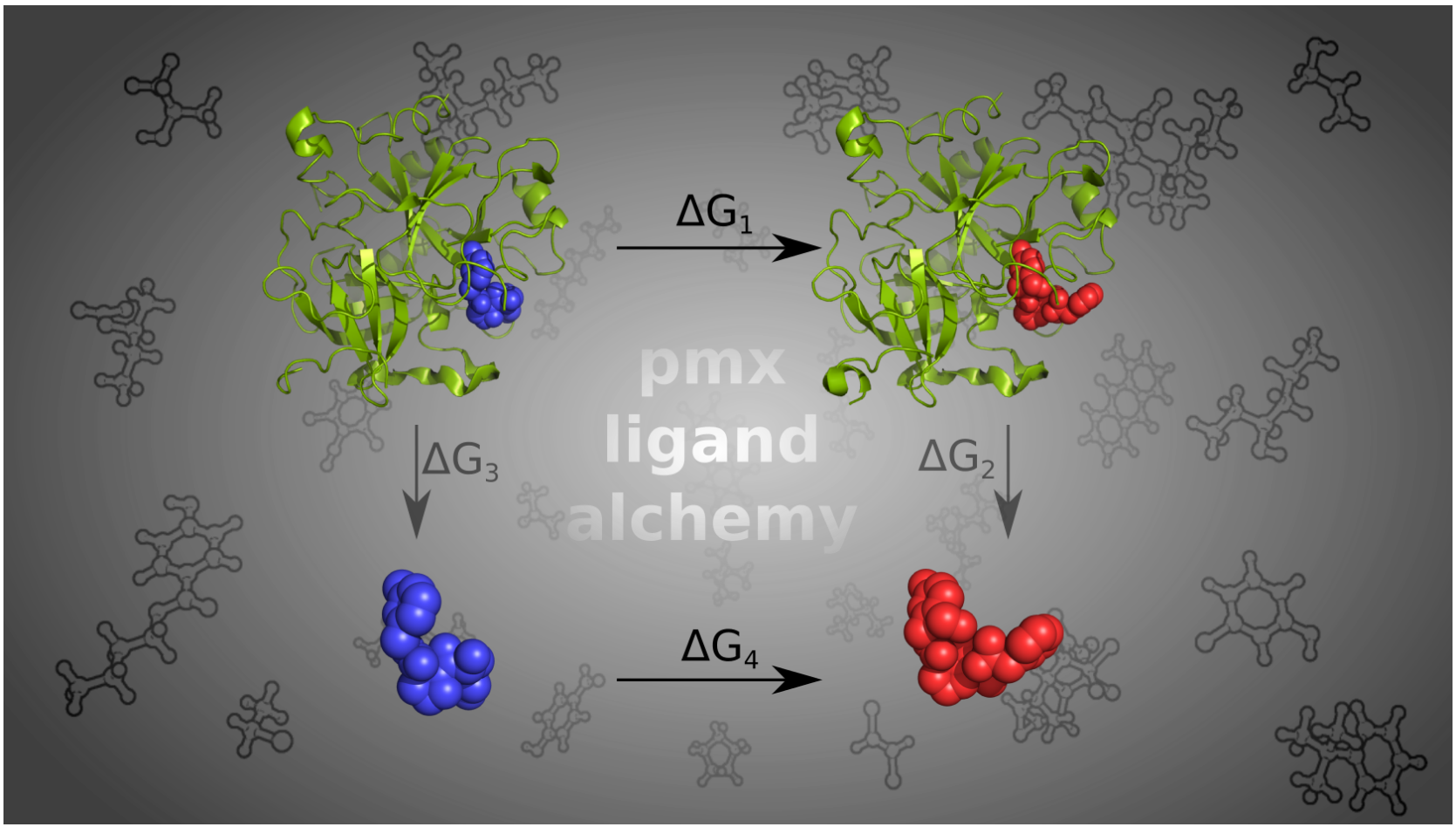


Nucleotide mutations



Ligand modifications

pmx: Ligands



pmx Ligands

atoms_to_morph.py

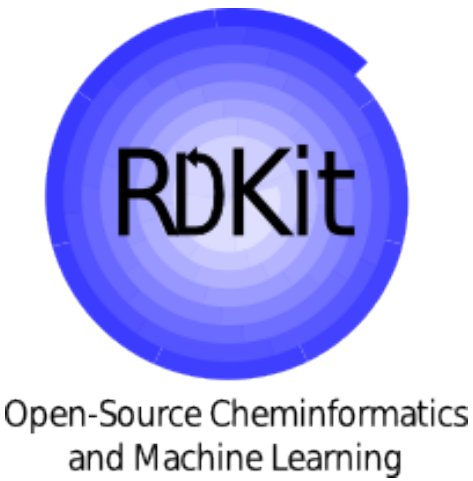
Identifies atoms to be morphed

make_hybrid.py

Builds hybrid topology

build_mst_graph.py

Suggests ligand pairs



Open source toolkit for cheminformatics

pmx Ligands

atoms_to_morph.py

Identifies atoms to be morphed

make_hybrid.py

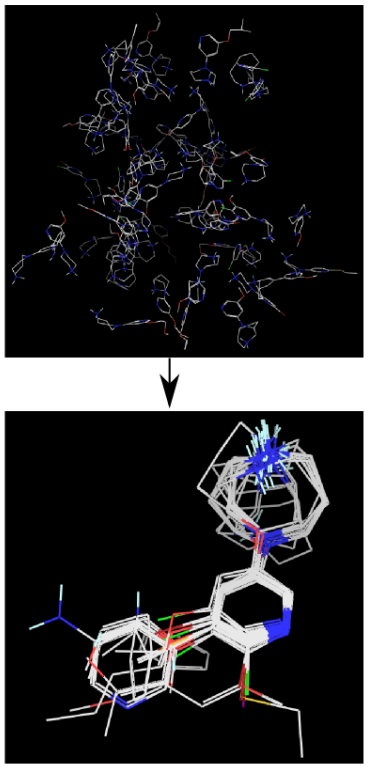
Builds hybrid topology

build_mst_graph.py

Suggests ligand pairs

Ligands: atom mapping

Tosco, 2013

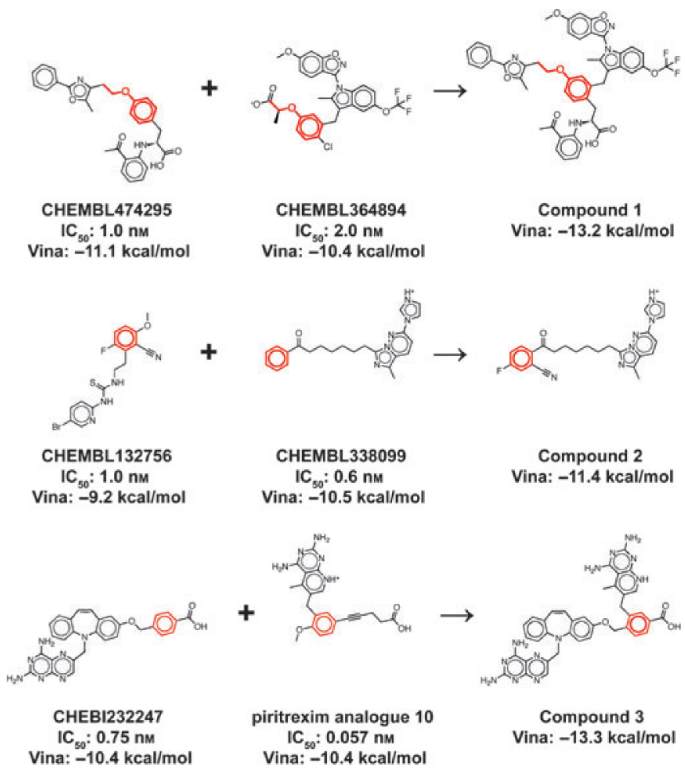


Open3DALIGN

Unsupervised ligand alignment and superposition

Linder et al., 2013

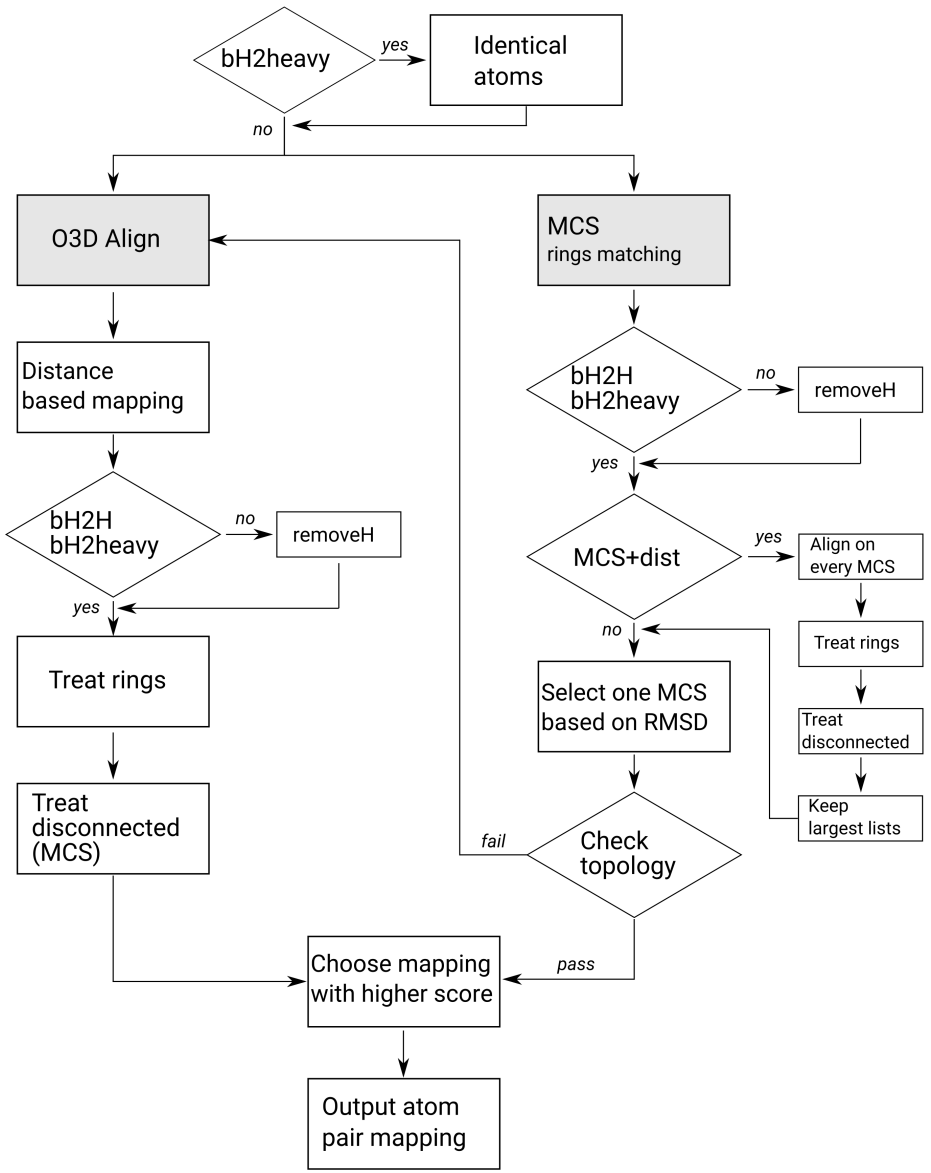
Andrew Dalke Scientific AB, 2012



MCS

Maximum Common Substructure

Ligands: atom mapping



Ligands: hybrid structure/topology

atoms_to_morph.py

Identifies atoms to be morphed

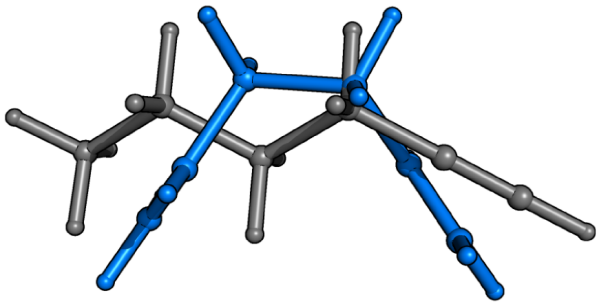
make_hybrid.py

Builds hybrid topology

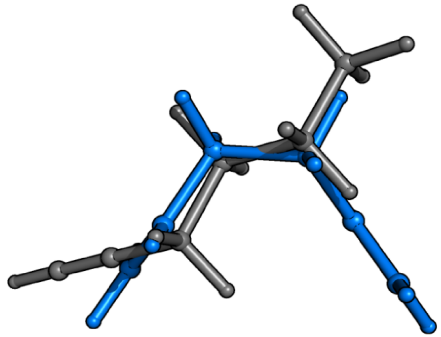
build_mst_graph.py

Suggests ligand pairs

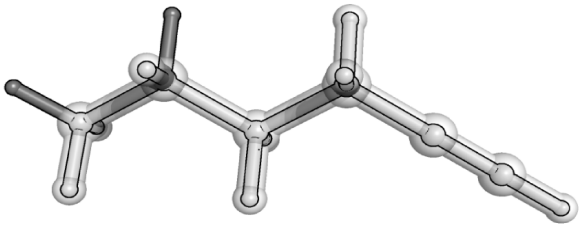
Ligands: hybrid structures/topologies

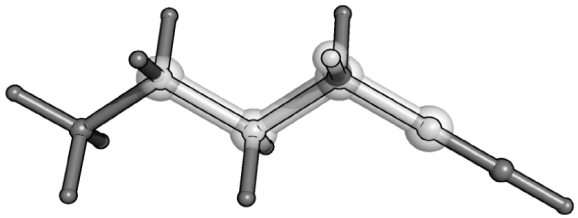


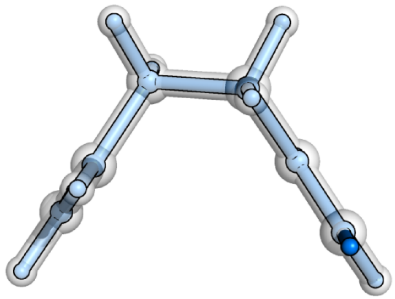
MCS (with H2H)
mol1: 14/16 atoms
mol2: 14/16 atoms

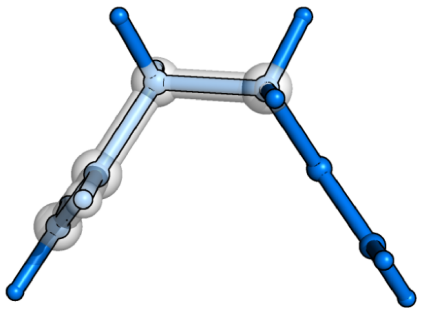


Alignment (with H2H)
mol1: 5/16 atoms
mol2: 5/16 atoms









Ligands: pairs of ligands to morph

atoms_to_morph.py

Identifies atoms to be morphed

make_hybrid.py

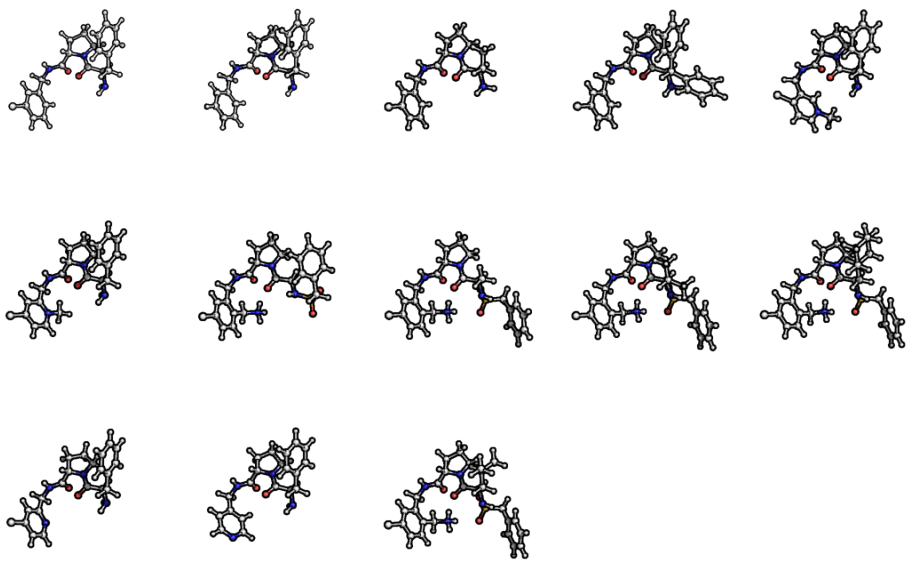
Builds hybrid topology

build_mst_graph.py

Suggests ligand pairs

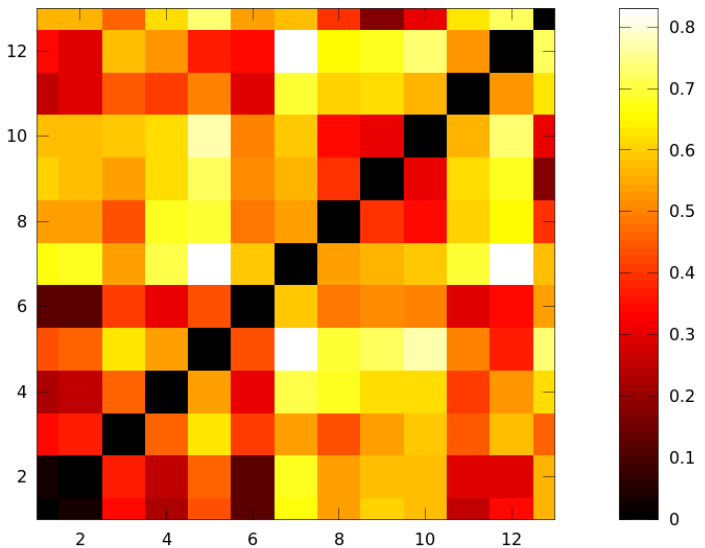
Ligands: pairs of ligands to morph

Starting with a set of ligand PDBs



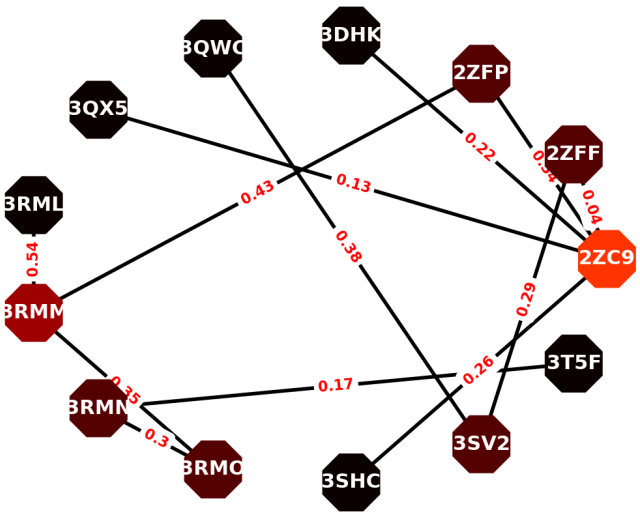
Run pairwise atoms_to_morph.py and score

$$d = 1 - \frac{n_1+n_2}{2(N_1+N_2)-(n_1+n_2)}$$

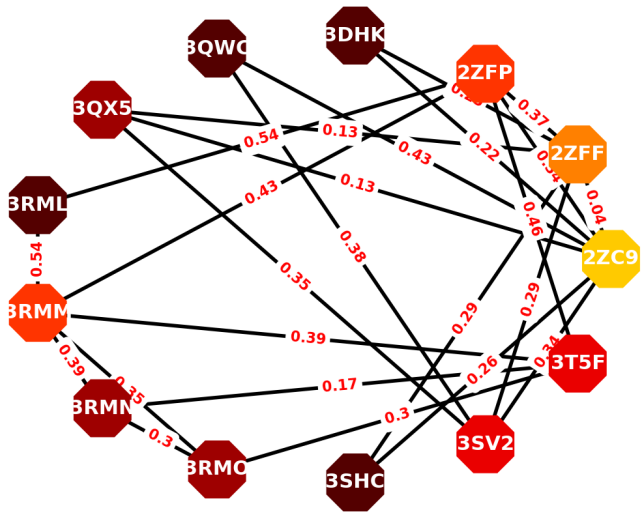


Obtain a distance matrix

Ligands: pairs of ligands to morph



Build a minimum spanning tree (MST)

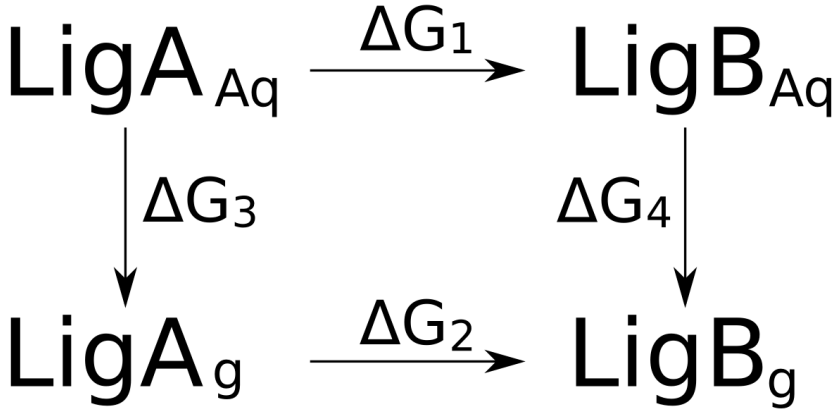


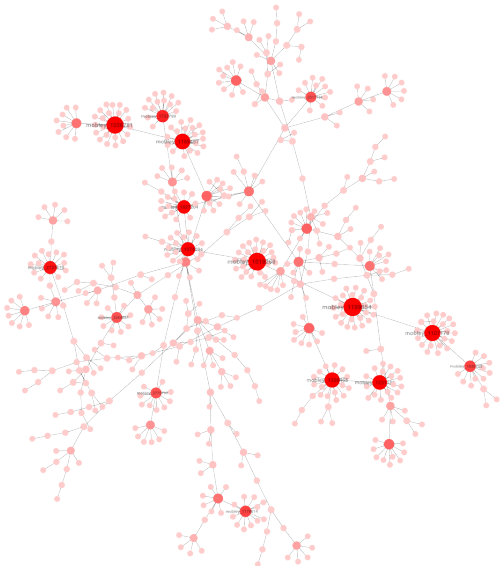
Build the second MST discarding all edges of the first MST

Validation: thermodynamic cycle for solvation free energies

Validation:
Thermodynamic Cycle for Solvation Free Energies

Validation: solvation free energies





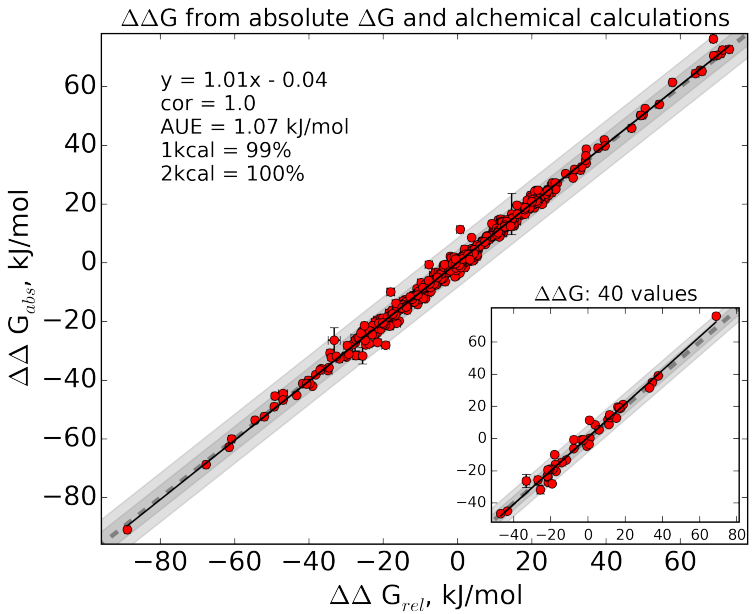
Full thermodynamic cycle

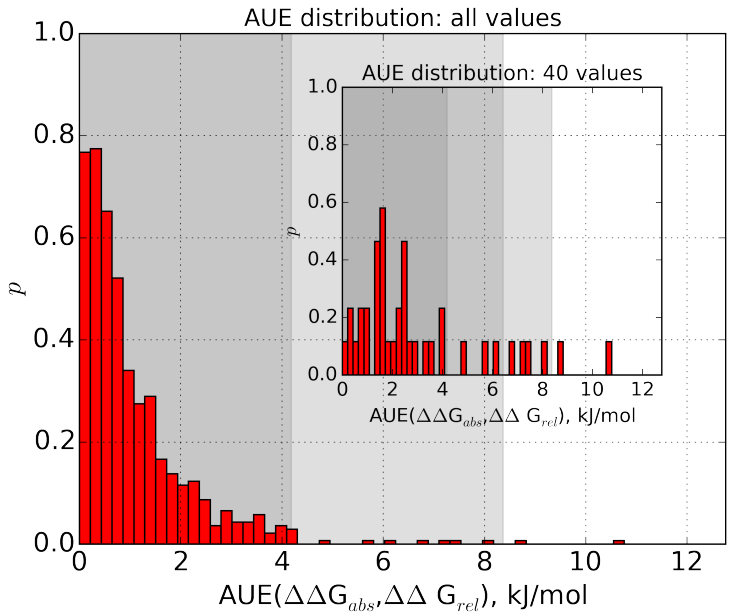
- ΔG_1 and ΔG_2 can be calculated using hybrid topology.
- ΔG_3 and ΔG_4 can be calculated by decoupling ligands from the solvent.

Mobley, Guthrie, 2014, JCAMD

- FreeSolv database
- >600 neutral ligand structures and topologies
- Experimental solvation free energies

Validation: solvation free energies





Only $\sim$ 1% of the cases have an AUE larger than 1 kcal/mol.

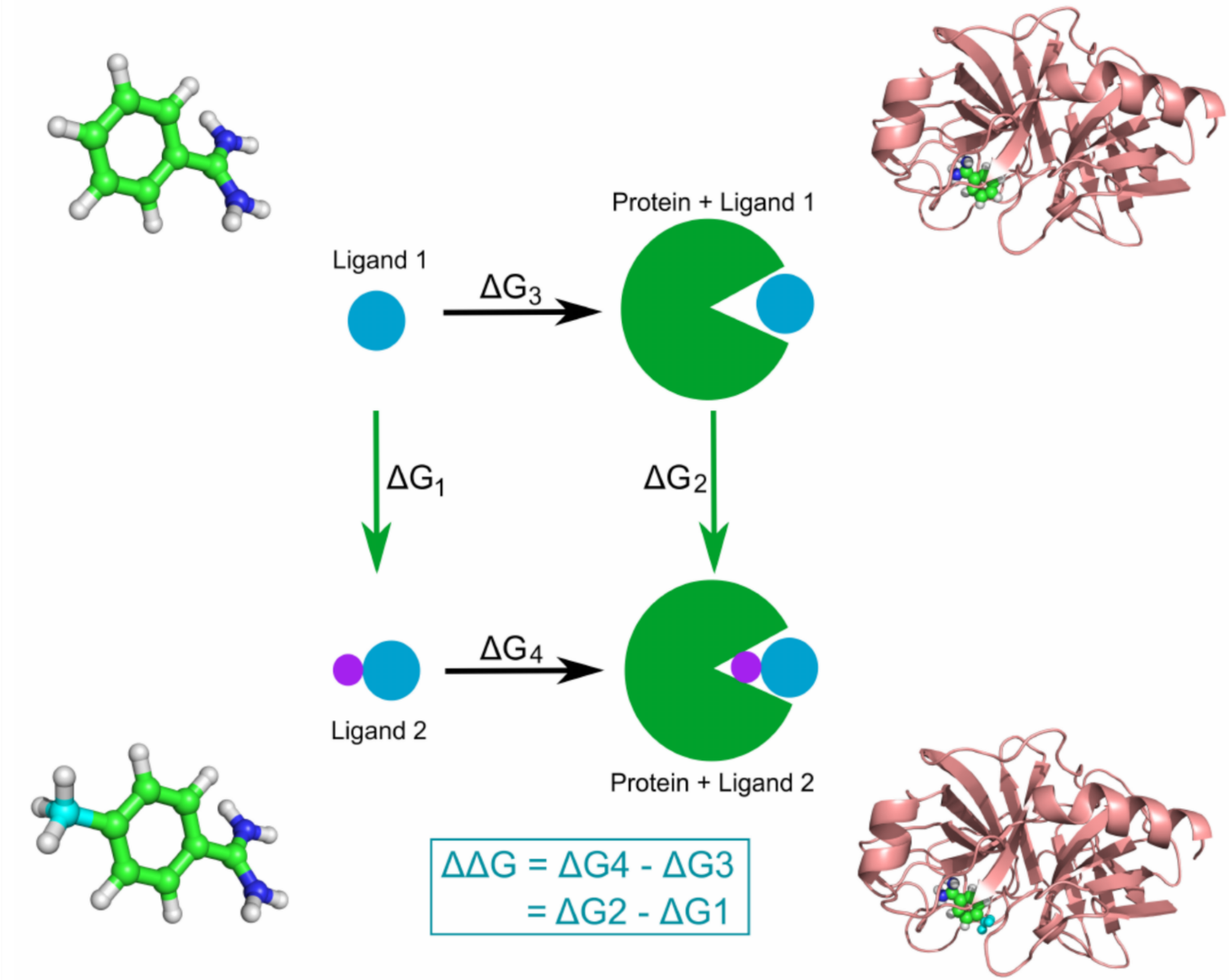
Protein-Ligand binding

Application:
482 ligand modifications in protein-ligand binding

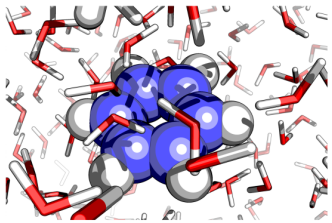
Gapsys, Perez-Benito, Aldeghi, Seeliger, van Vlijmen, Tresadern, de Groot,

submitted

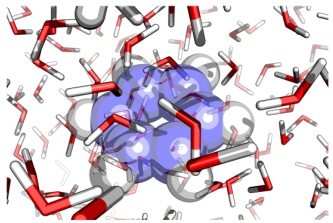
Methods (briefly): thermodynamic cycle



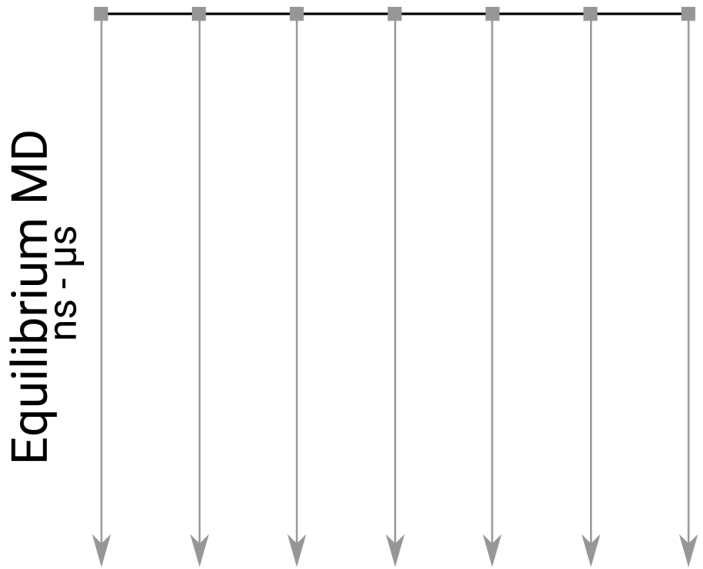
Methods (briefly): FEP protocol



State A (λ=0)

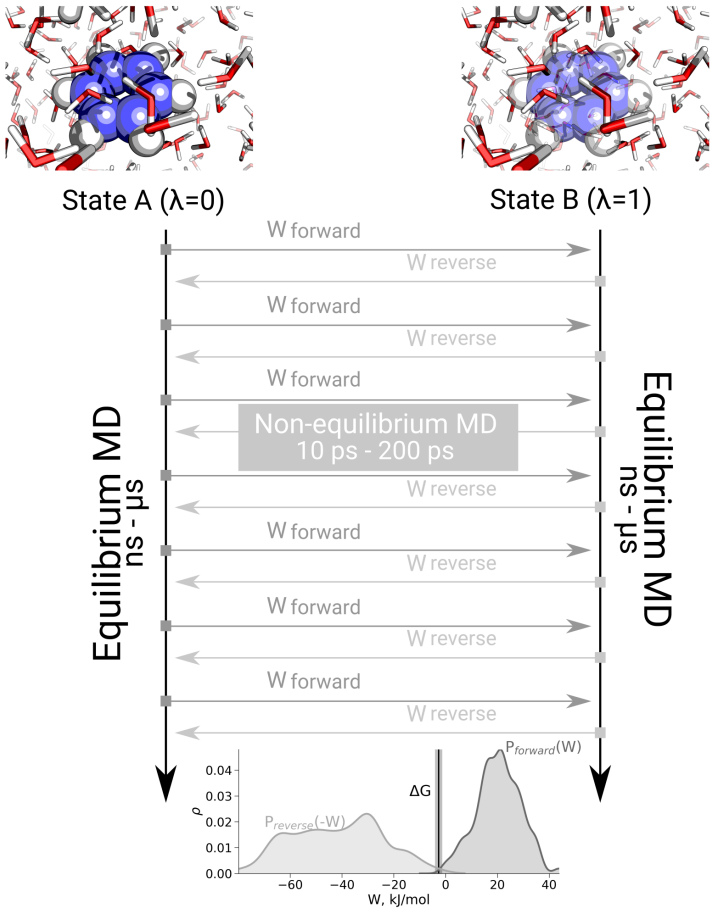


State B (λ=1)



Free energy perturbation (FEP)

Methods (briefly): non-eq TI protocol

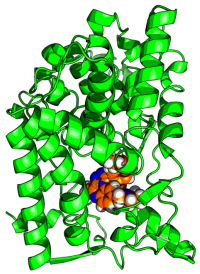


Crooks Fluctuation Theorem:

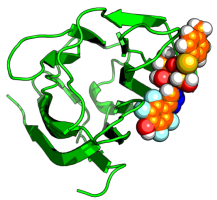
$$\frac{P_f(W)}{P_r(-W)} = e^{\beta(W-\Delta G)}$$

Protein-Ligand complexes

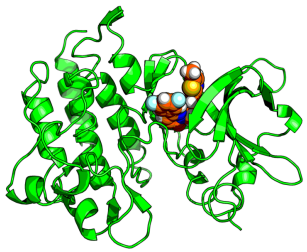
- 11 systems
- 482 mutations



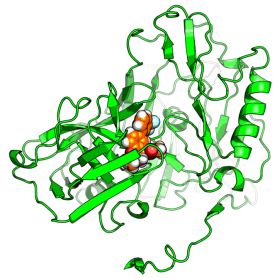
PDE2: 21 ligand
34 perturbations



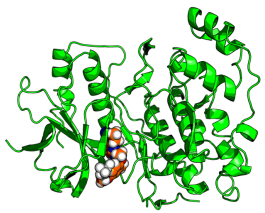
Galectin: 8 ligands
8 perturbations



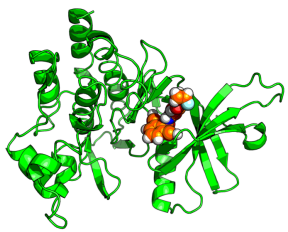
cMet: 12 ligands
25 perturbations



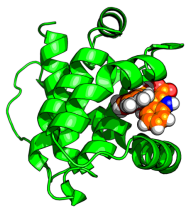
BACE: 80 ligands
144 perturbations
(divided in 3 sets)



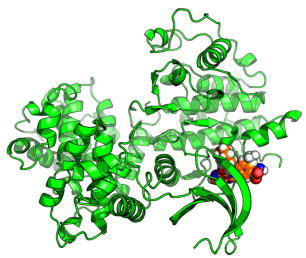
JNK1: 21 ligands
31 perturbations



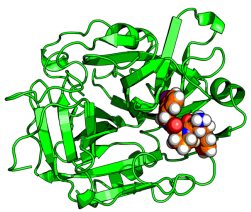
TYK2: 16 ligands
24 perturbations



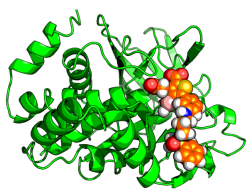
MCL1: 42 ligands
71 perturbations



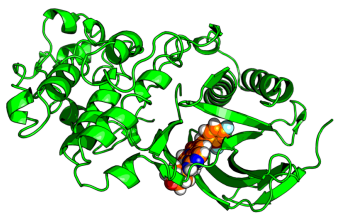
CDK2: 16 ligands
25 perturbations



Thrombin: 11 ligands
16 perturbations



PTP1b: 23 ligands
49 perturbations

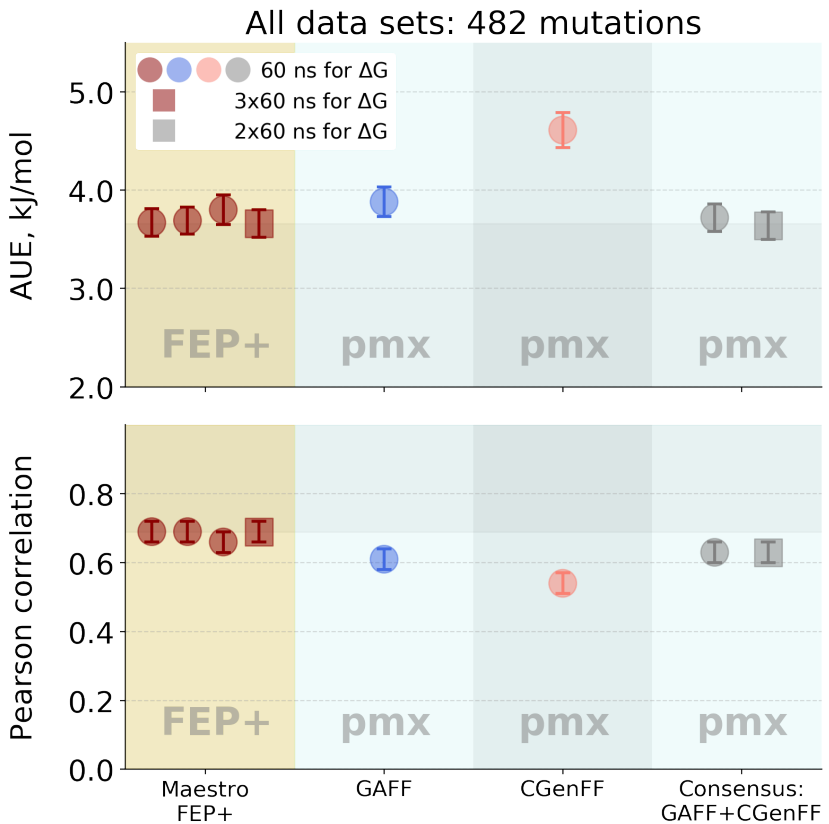


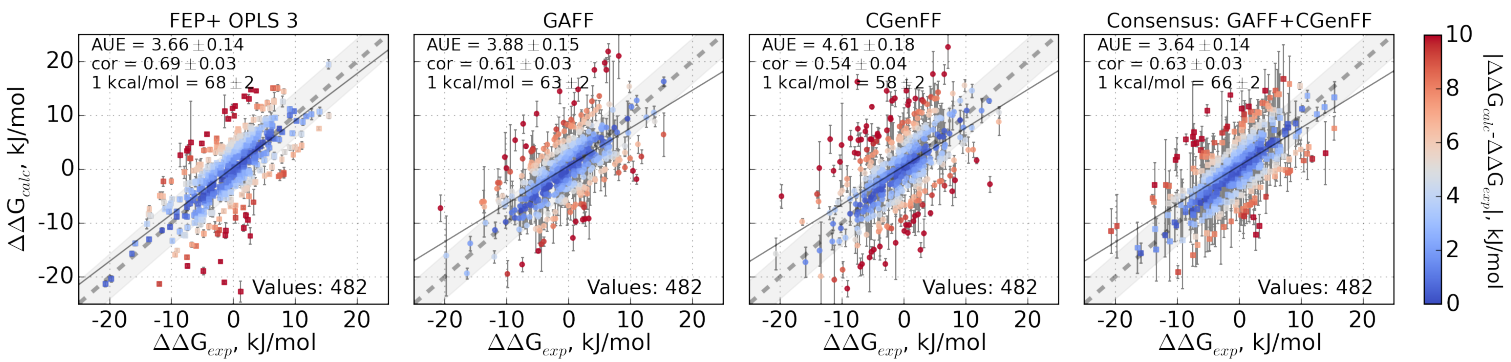
P38: 34 ligands
56 perturbations

Overall results

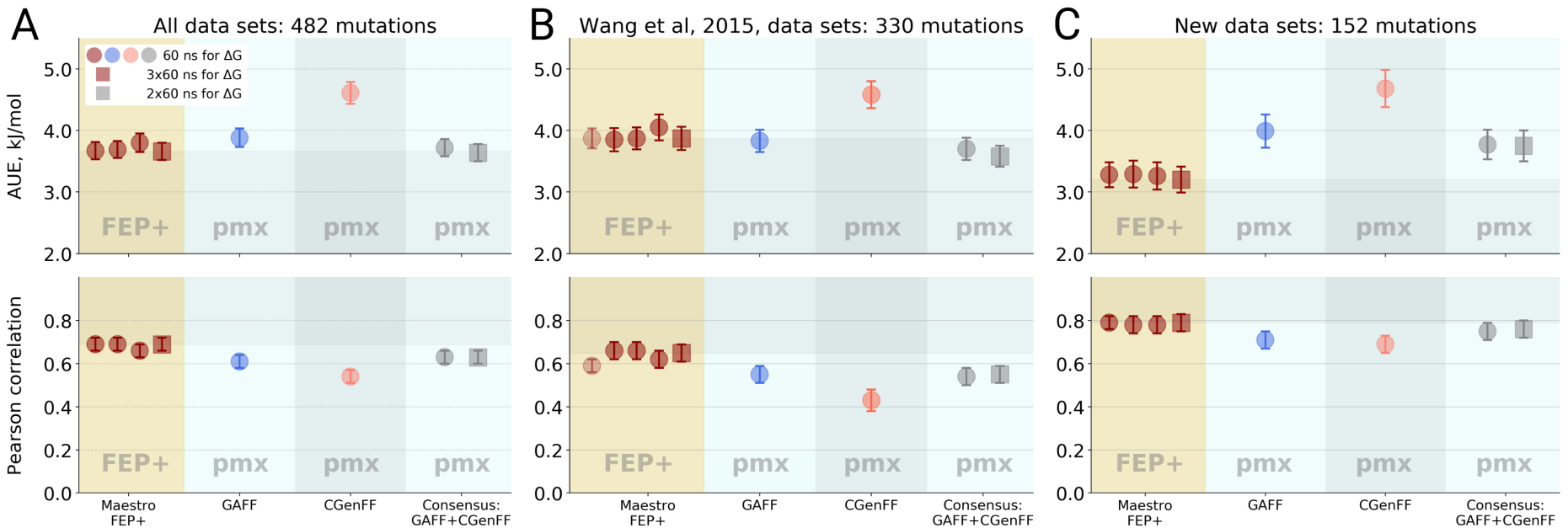
Overall results

Overall results

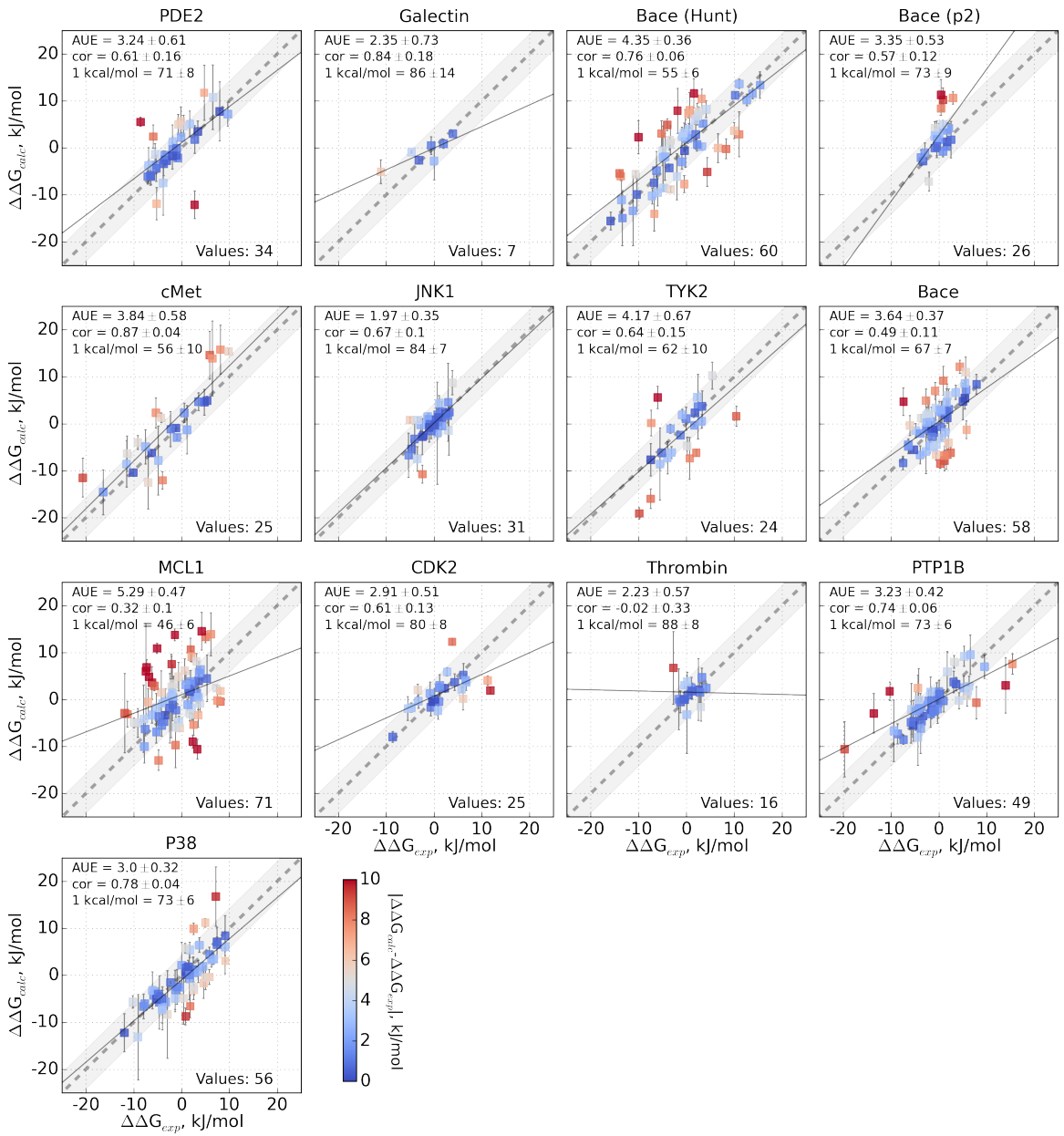




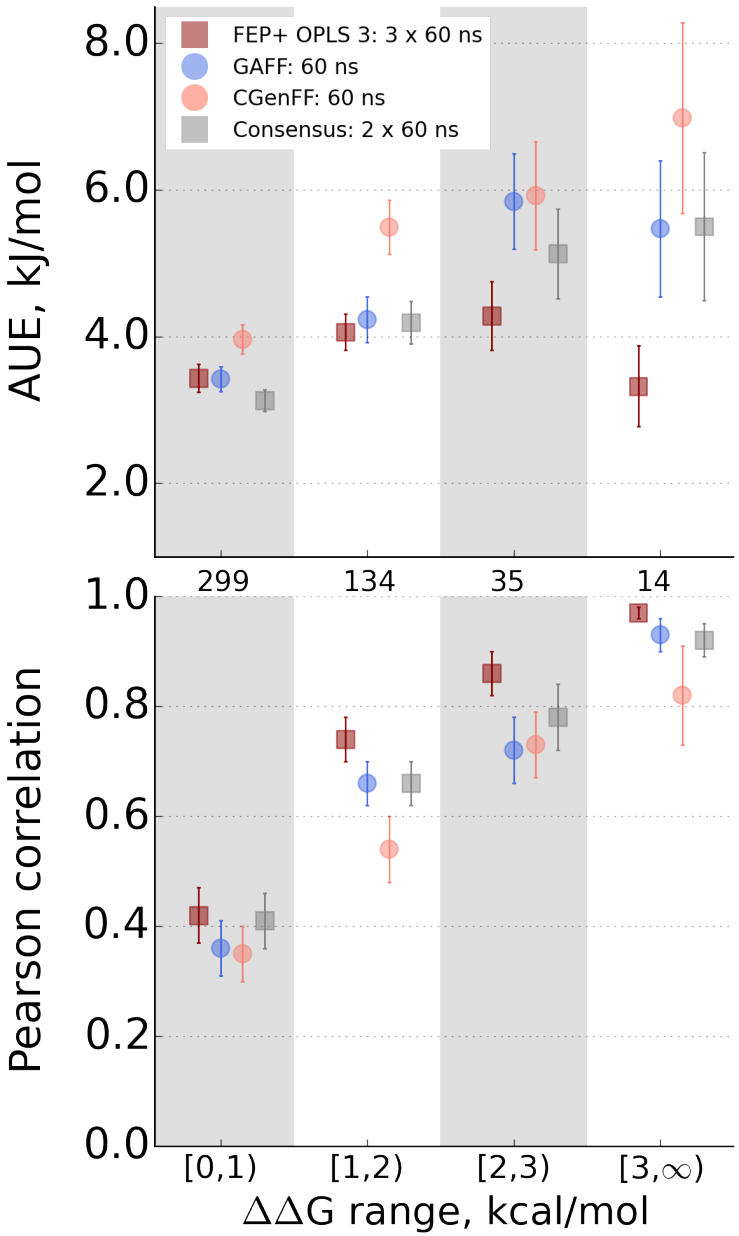
Results: subsets



Results: by case



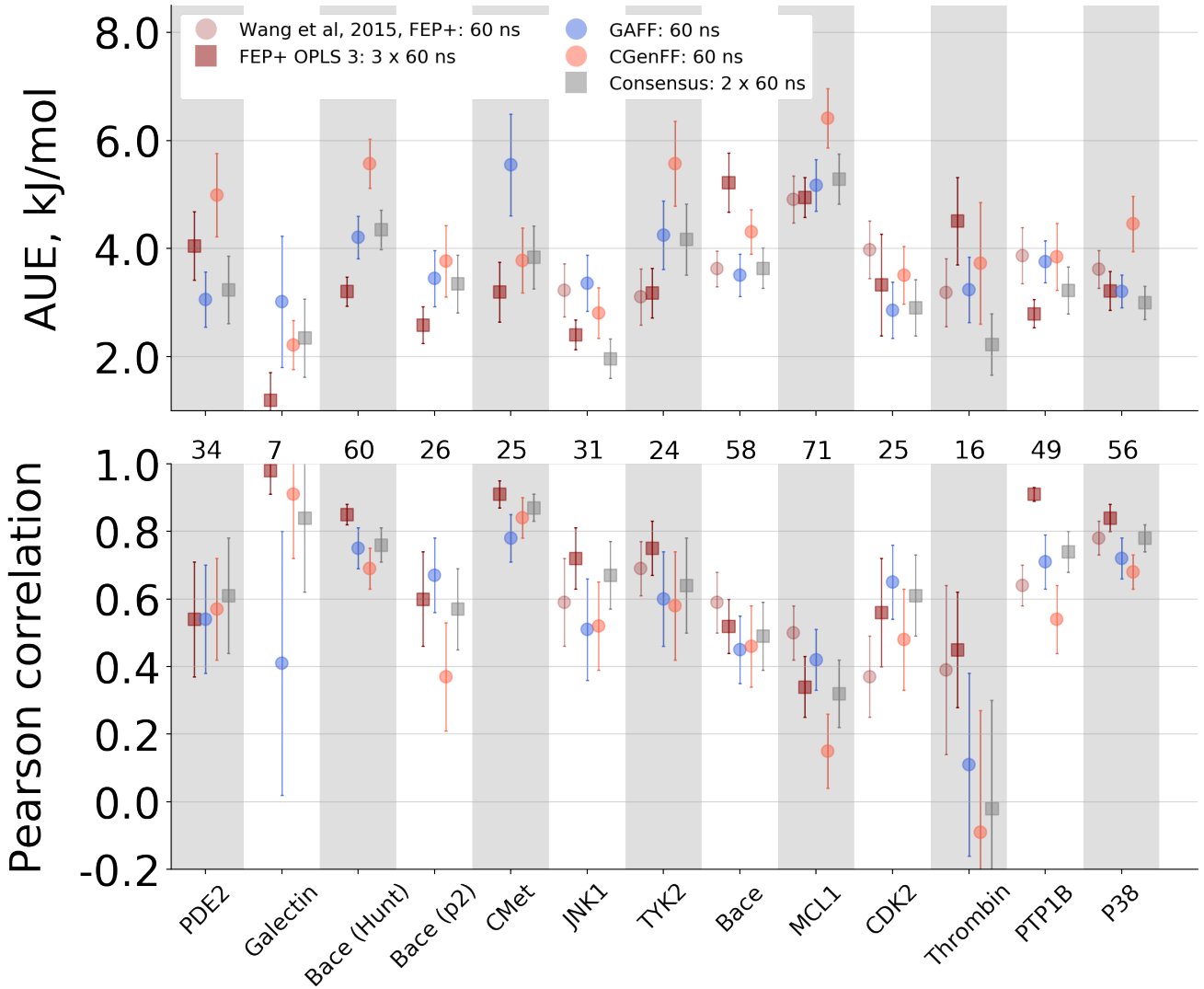
Results: by ΔΔG range



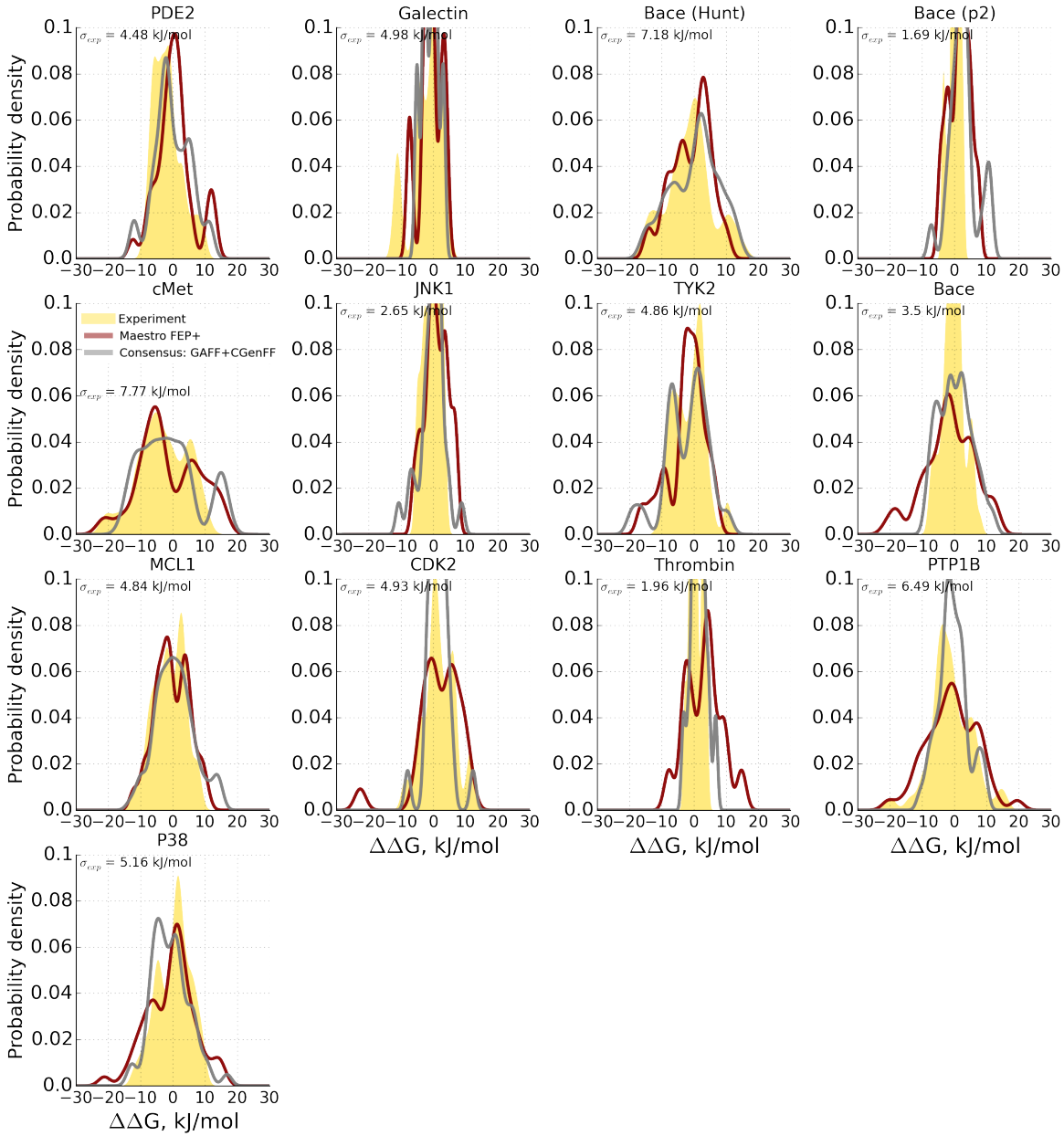
FEP+ performs better in predicting large ΔΔG changes:

- FEP vs non-eq TI
- enhanced sampling (REST)

Results: by case



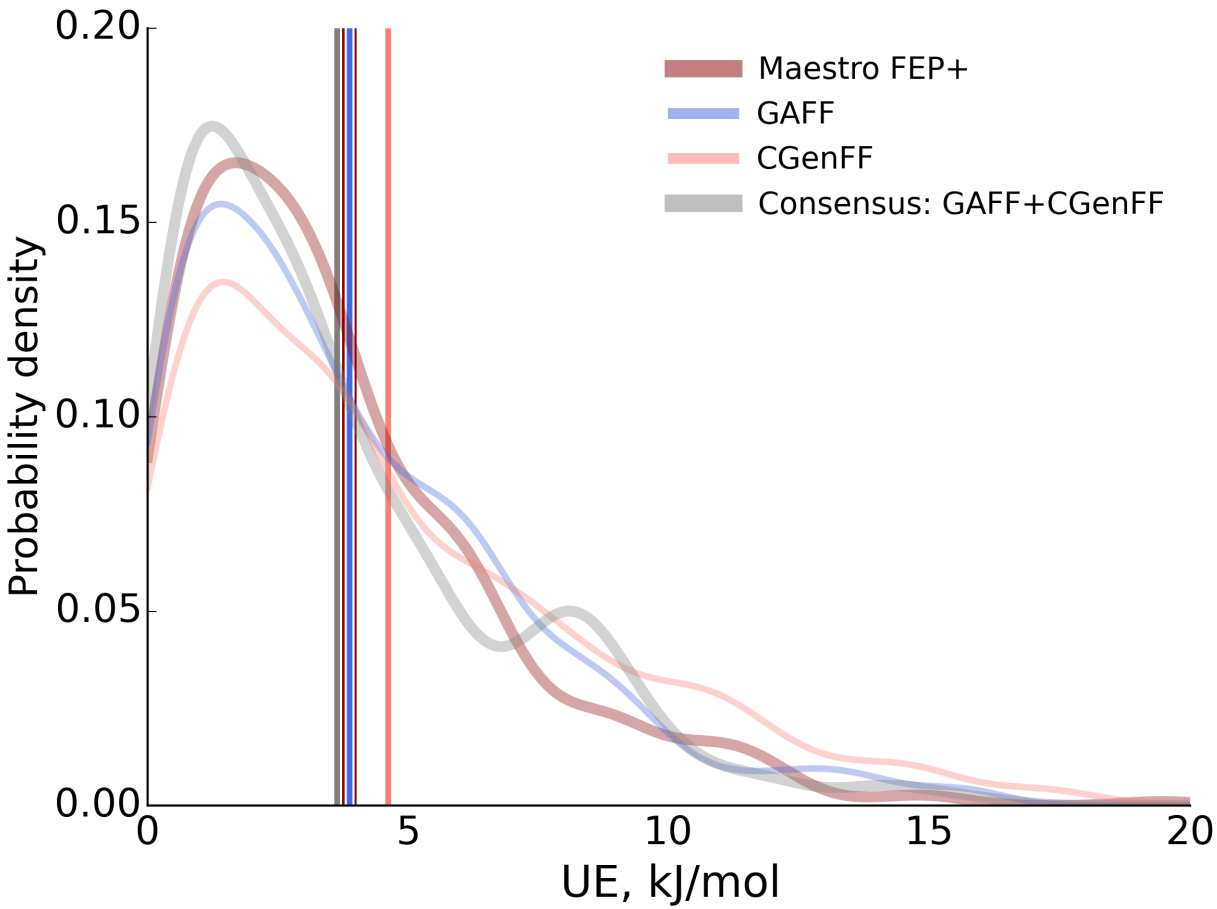
Results: ΔΔG range



Accuracy and precision

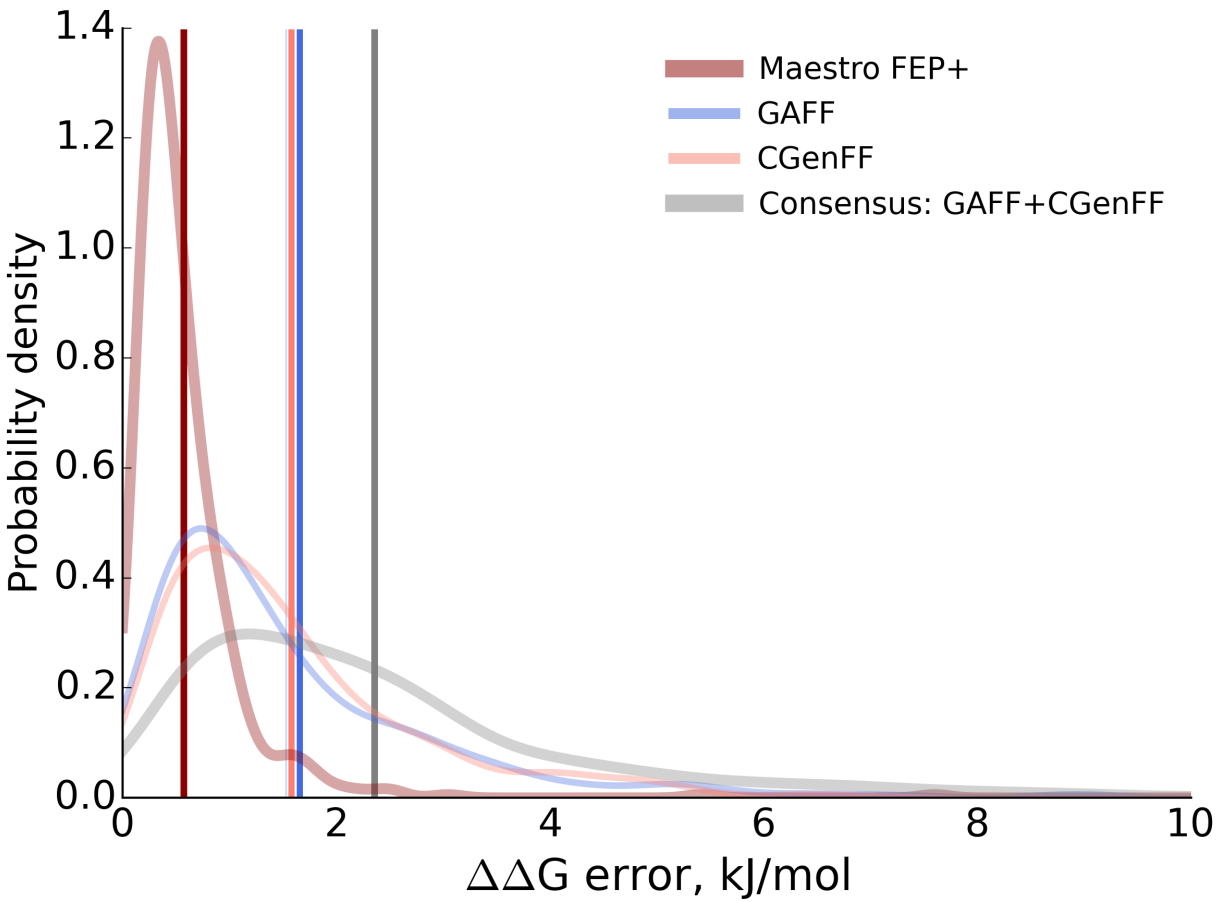
Accuracy and precision

Accuracy



Different approaches reach comparable accuracies.

Precision

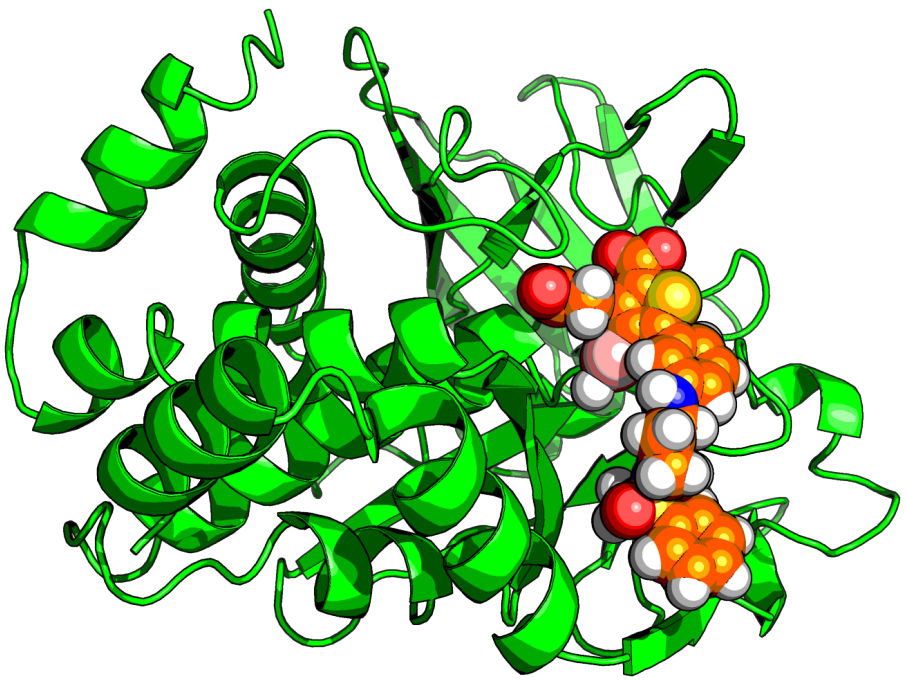


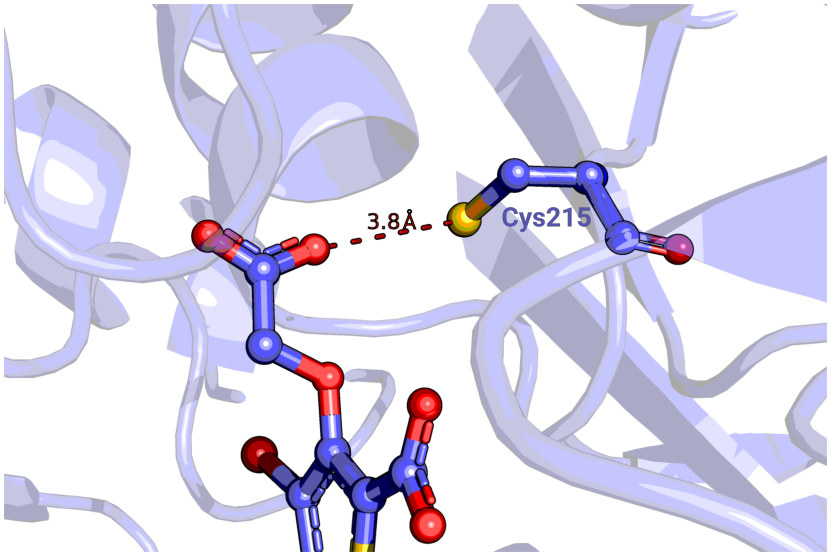
FEP+ has higher precision.

Determinants of prediction accuracy

Determinants of prediction accuracy

Case: ptp1b



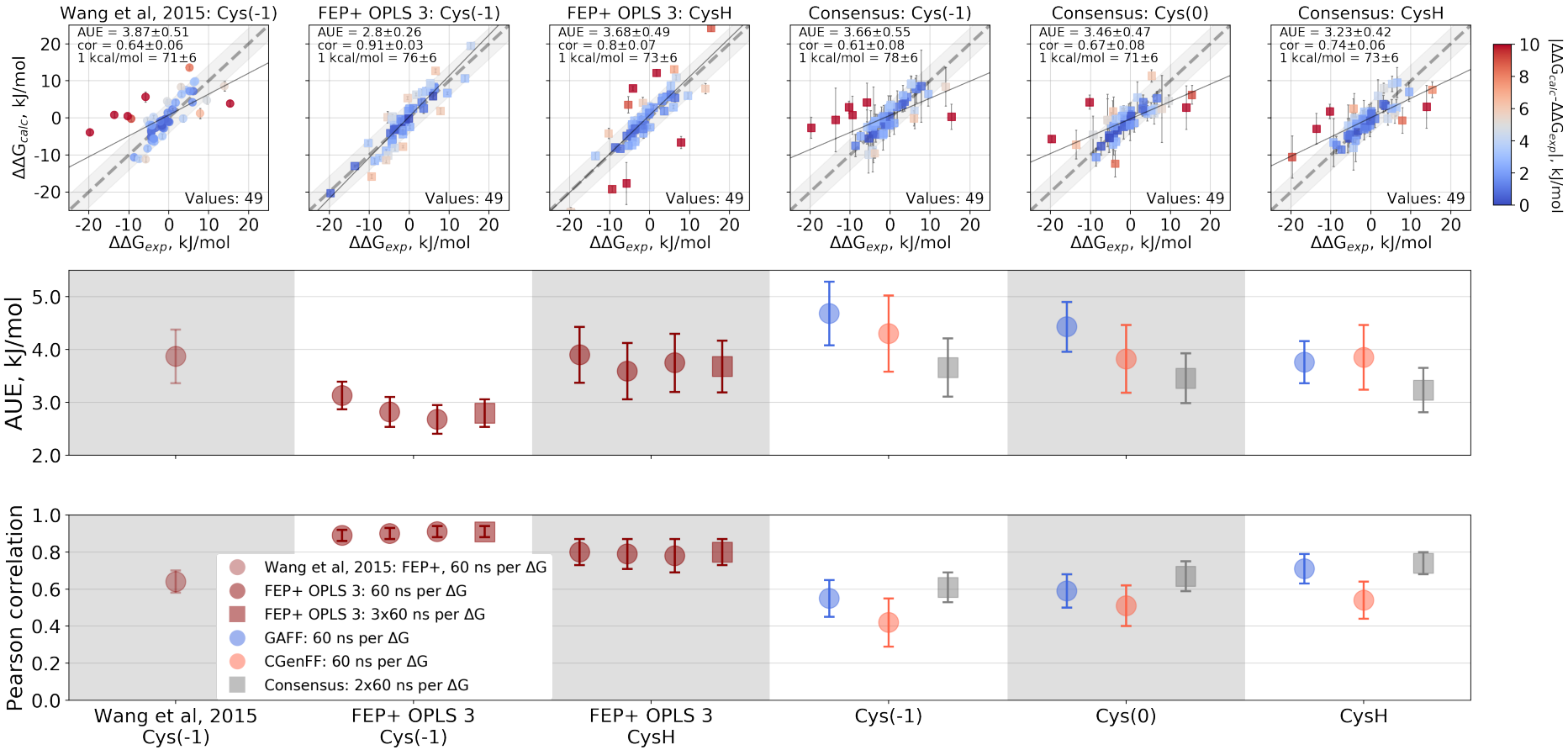


ptp1b

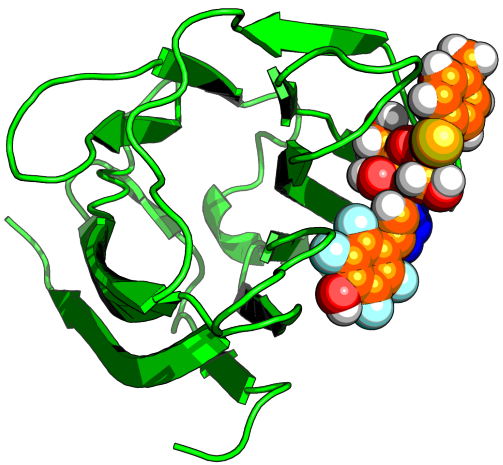
Cys215 in apo state has pKa of 5.4.

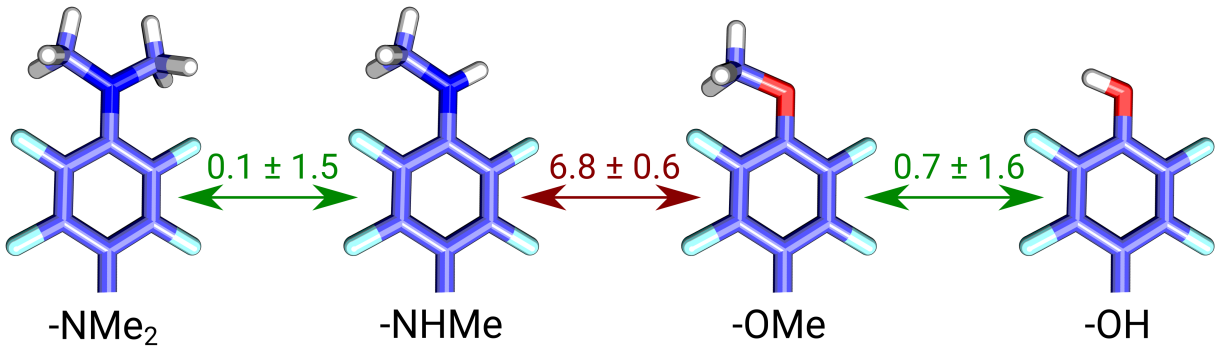
pKa in holo state is unknown.

Case: ptp1b


Wang et al, 2015: Cys(-1)
FEP+ OPLS 3: Cys(-1)
FEP+ OPLS 3: CysH
Consensus: Cys(-1)
Consensus: Cys(0)
Consensus: CysH
AUE = 3.87±0.51
cor = 0.64±0.06
1 kcal/mol = 71±6
AUE = 2.8±0.26
cor = 0.91±0.03
1 kcal/mol = 76±6
AUE = 3.68±0.49
cor = 0.8±0.07
1 kcal/mol = 73±6
AUE = 3.66±0.55
cor = 0.61±0.08
1 kcal/mol = 78±6
AUE = 3.46±0.47
cor = 0.67±0.08
1 kcal/mol = 71±6
AUE = 3.23±0.42
cor = 0.74±0.06
1 kcal/mol = 73±6
Values: 49
ΔΔG_calc, kJ/mol
ΔΔG_exp, kJ/mol
|ΔΔG_calc-ΔΔG_exp|, kJ/mol
AUE, kJ/mol
Pearson correlation
Wang et al, 2015: FEP+, 60 ns per ΔG
FEP+ OPLS 3: 60 ns per ΔG
FEP+ OPLS 3: 3x60 ns per ΔG
GAFF: 60 ns per ΔG
CGenFF: 60 ns per ΔG
Consensus: 2x60 ns per ΔG
Wang et al, 2015 Cys(-1)
FEP+ OPLS 3 Cys(-1)
FEP+ OPLS 3 CysH
Cys(-1)
Cys(0)
CysH

Case: galectin



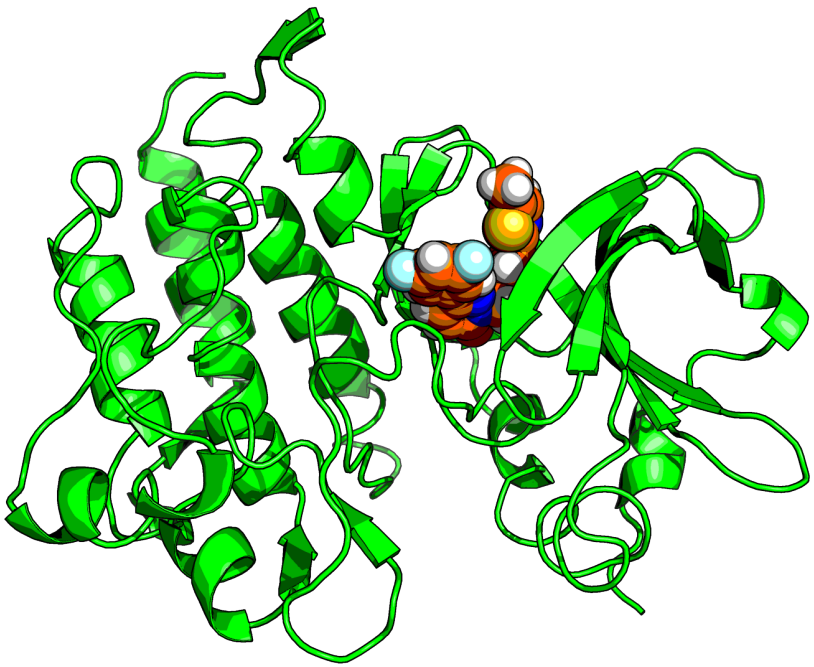


Galectin: UE from experiment, kJ/mol

ΔΔG estimates are accurate for changes within chemical groups.

Inter-chemical group ΔΔG values are less accurate.

Case: cMet

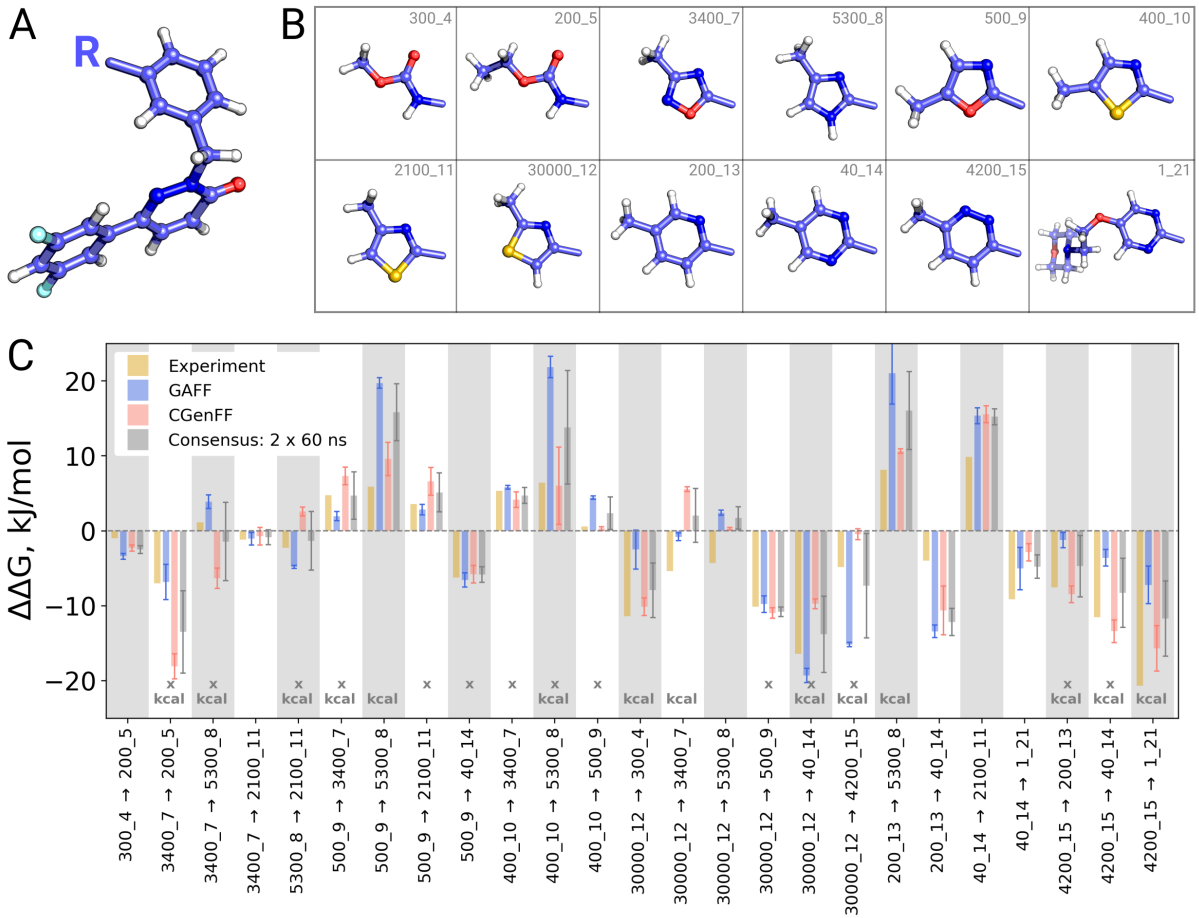


cMet

For cMet, in 56% of $\Delta\Delta G$ cases GAFF and CGenFF pointed in the opposite directions from experiment.

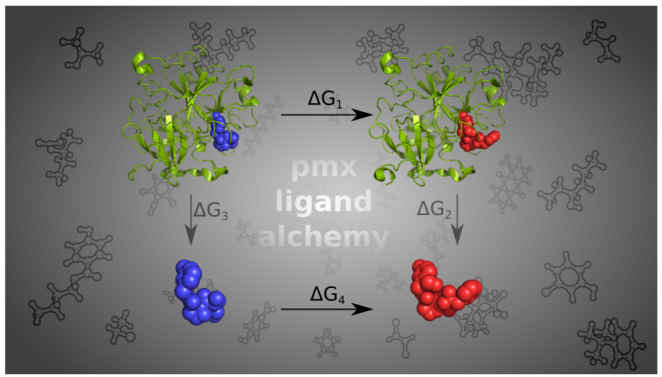
A good candidate to look for trends in force field performance.

Case: cMet

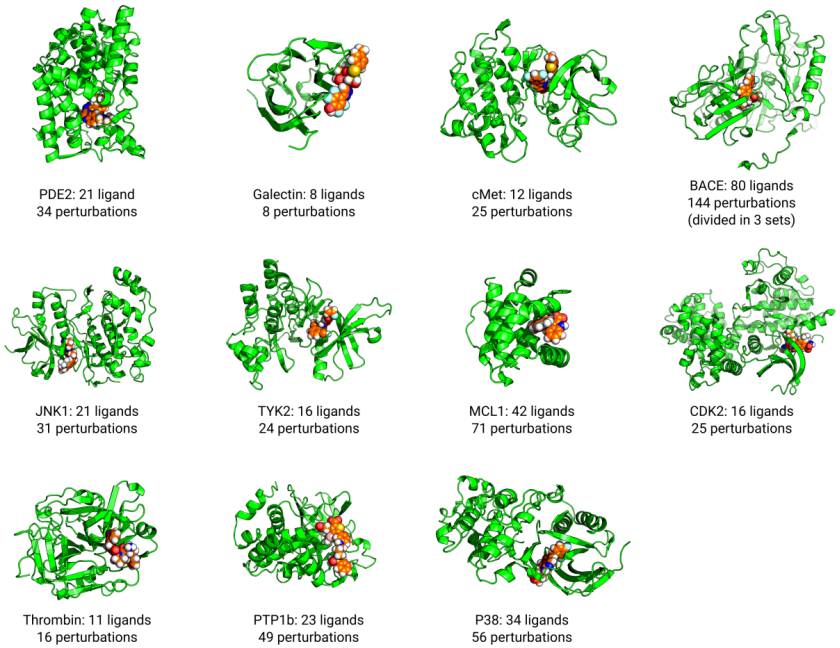


For example, GAFF overestimates binding affinity of the ligands 4200_15 and 400_10 in comparison to CGenFF and experiment.

Summary



pmx on github: https://github.com/deGrootLab/pmx
pmx webserver: http://pmx.mpibpc.mpg.de



Further topics (if time permits):

- *Absolute free energies*
- *Equilibrium FEP vs Non-Equilibrium TI*

pmx and free energies

Dr. Servaas Michielssens

Dr. Daniel Seeliger

Dr. Matteo Aldeghi

Dr. Yuriy Khalak

Professor Dr. Bert de Groot

Small molecule study

Dr. Laura Benitez

Dr. Gary Tresadern

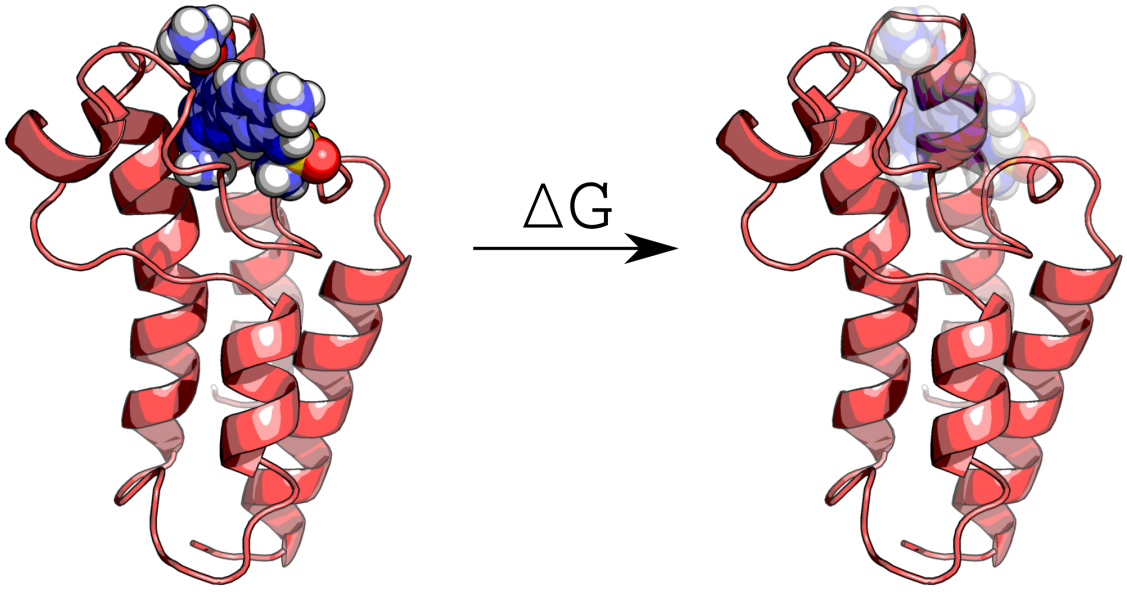
Professor Dr. Herman van Vlijmen

Funding

Absolute ΔG

Absolute ΔG calculation

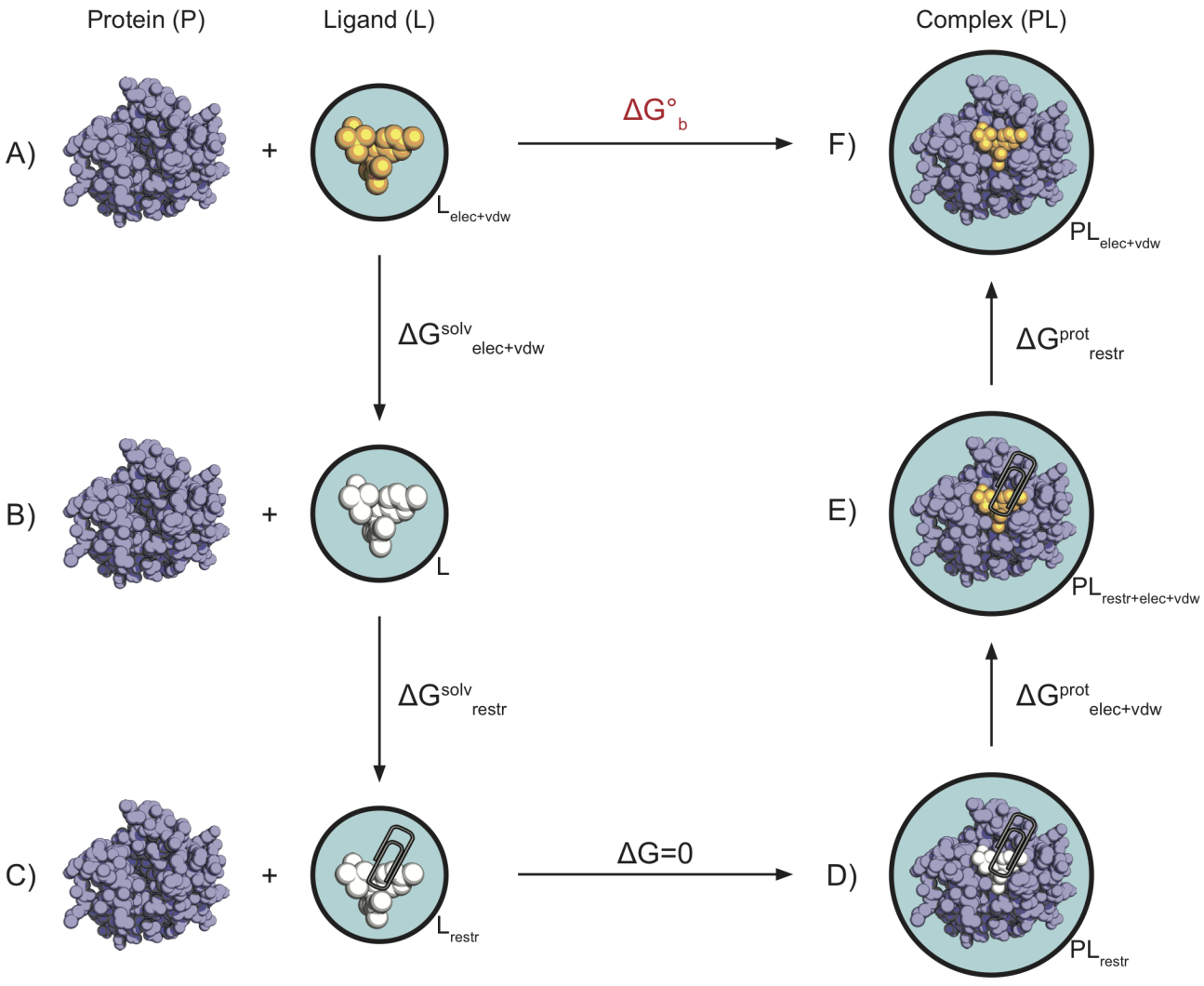
Absolute ΔG



Decoupling whole ligand

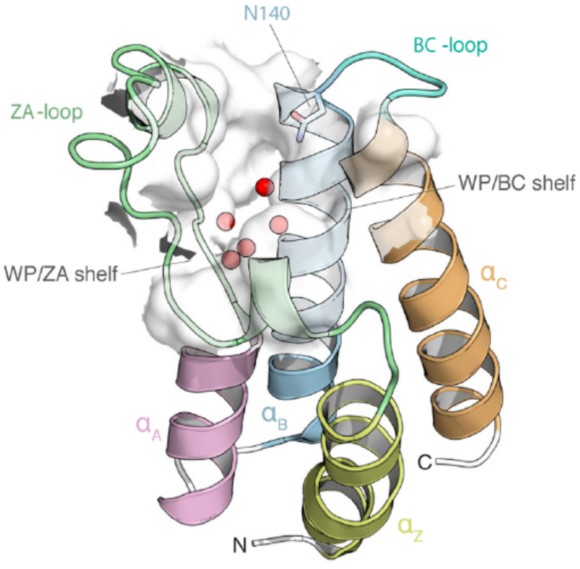
- Requires restraints
- Large perturbation
- Convergence is slow

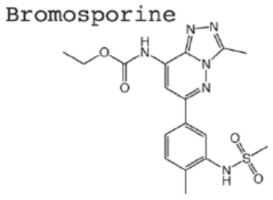
Absolute ΔG

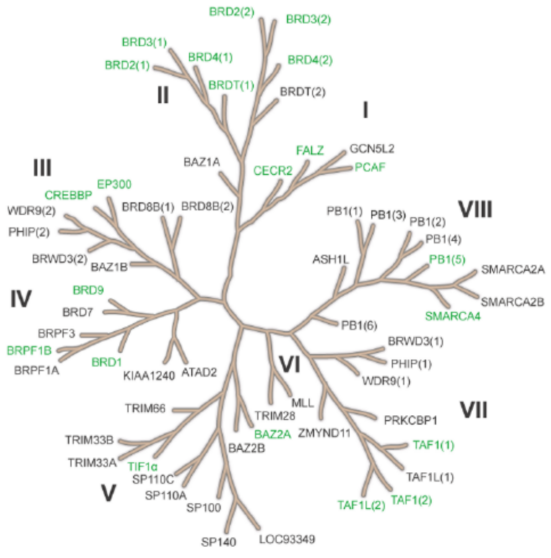


Reproduced from Aldeghi *et al.*, Chem. Sci., 2016, **7**, 207-218. © The Royal Society of Chemistry

Absolute ΔG: bromodomains



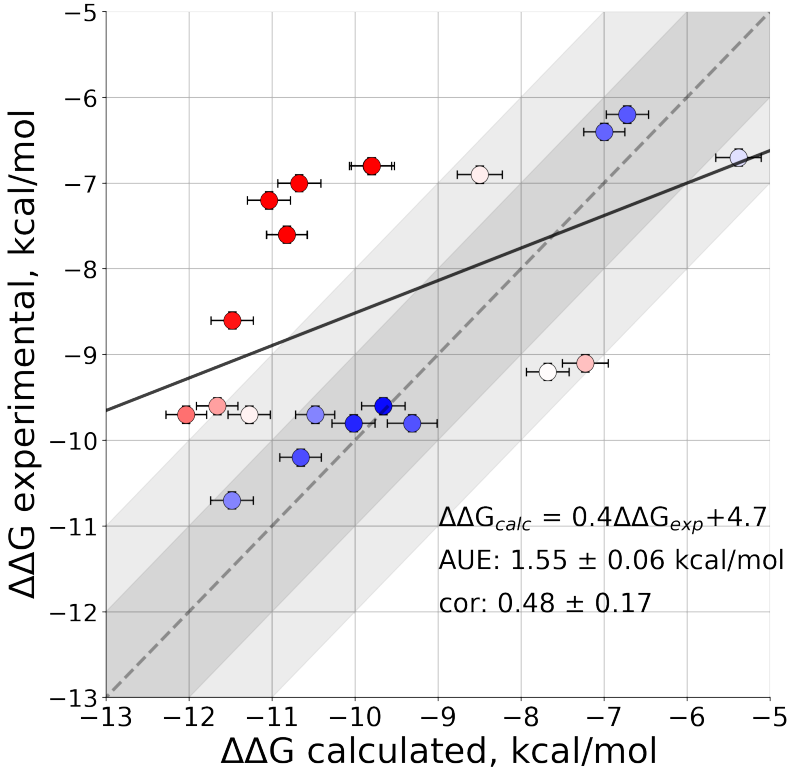




Aldeghi et al, JACS, 2017

- Bromosporine binding to the bromodomain proteins
- 22 different bromodomains

Absolute ΔG: bromodomains

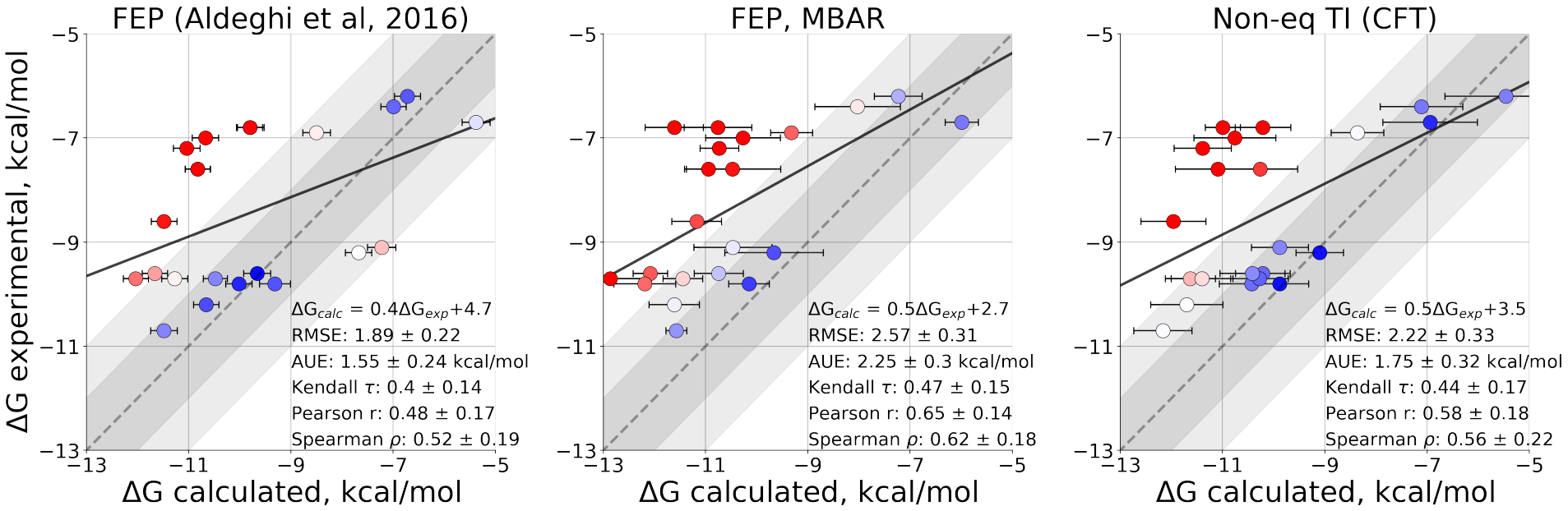


- Accuracy lower in comparison to the ΔΔG calculations
- More than 600 ns for a single ΔG value required for convergence

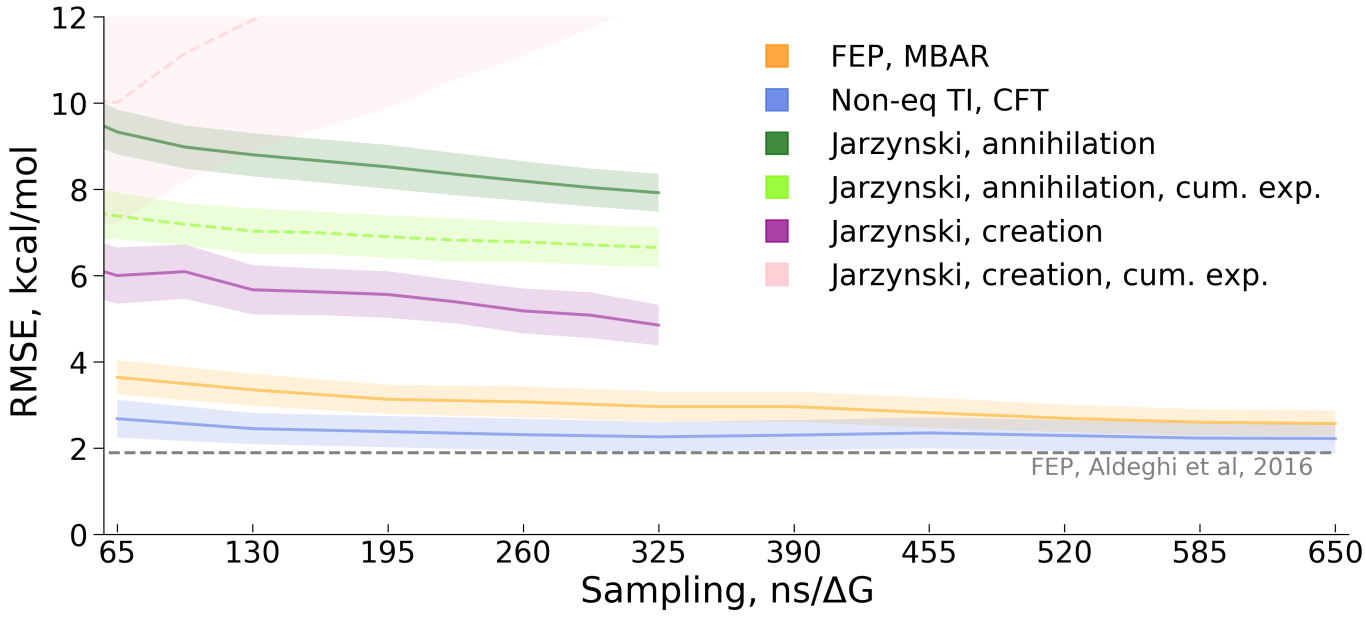
Aldeghi et al, JACS, 2017

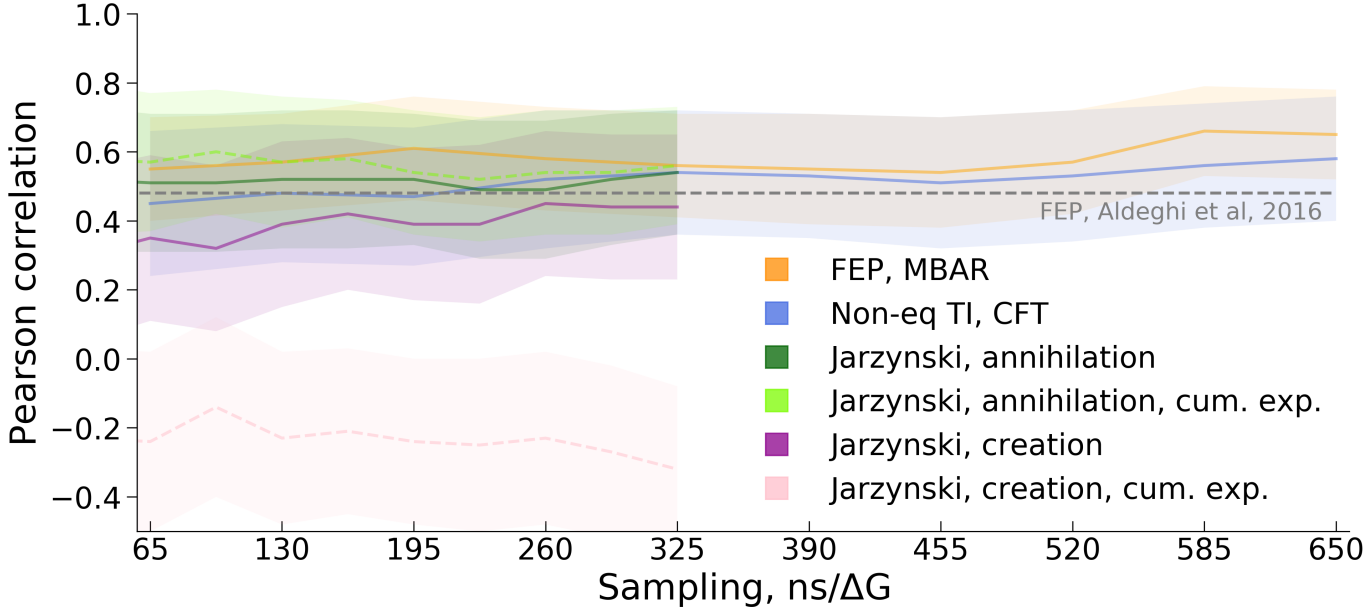
Equilibrium free energy calculation

Absolute ΔG: Non-Equilibrium TI



Absolute ΔG: Convergence





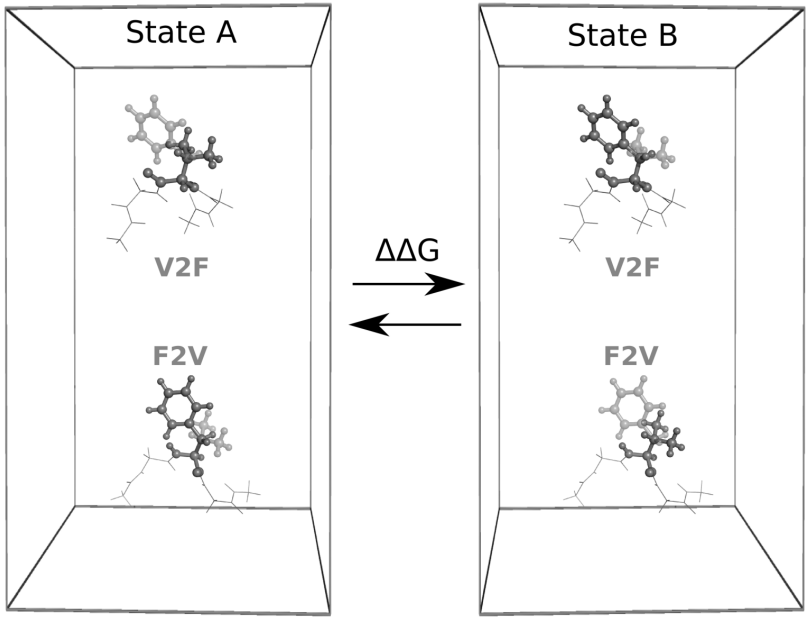
FEP vs Non-Eq TI

FEP vs Non-Eq TI

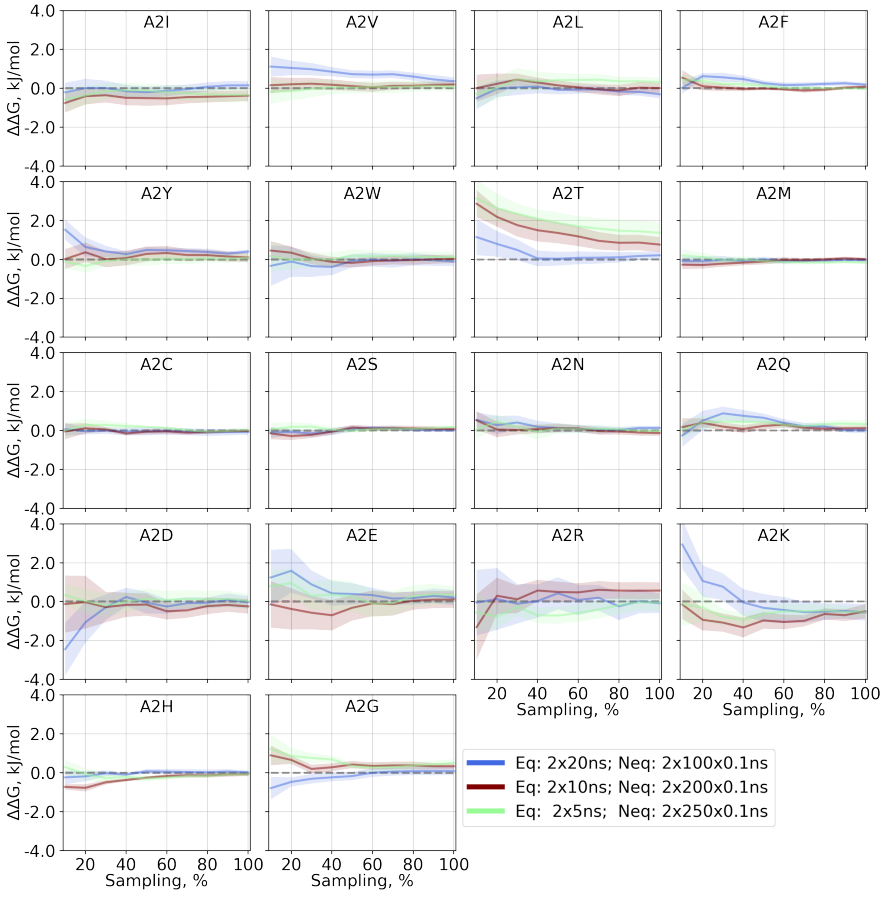
Comparison Strategy

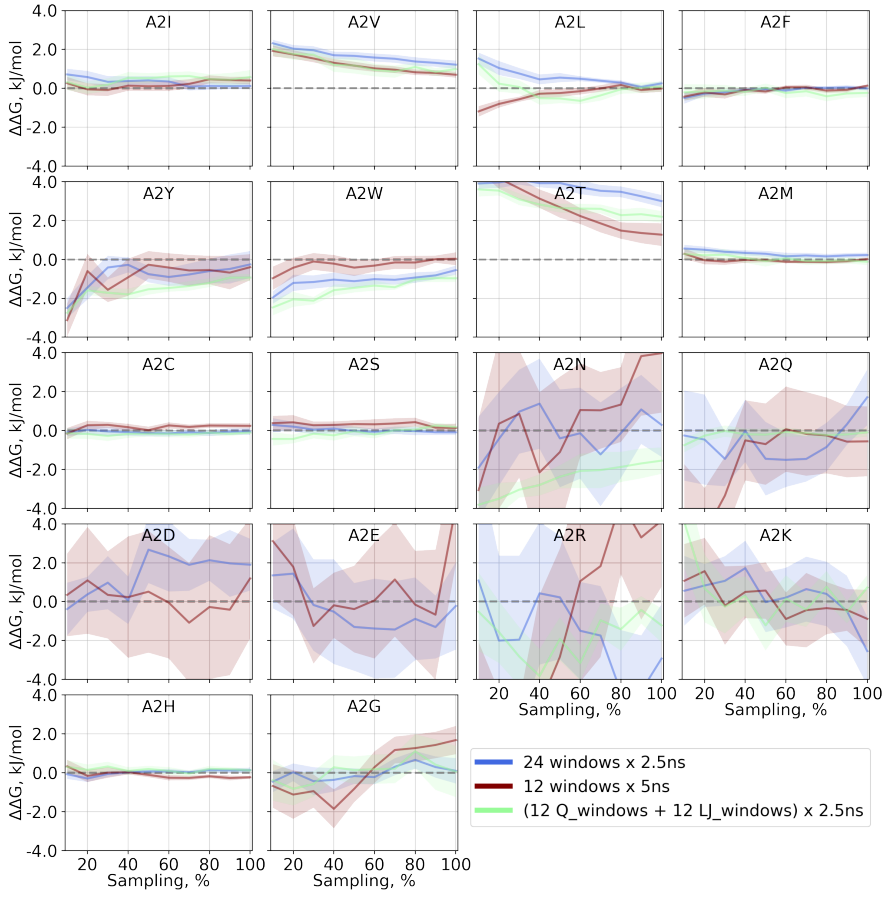
Requirement

- Large number of realistic perturbations
- Equivalent conditions and sampling time
- Known target value



Amino Acids

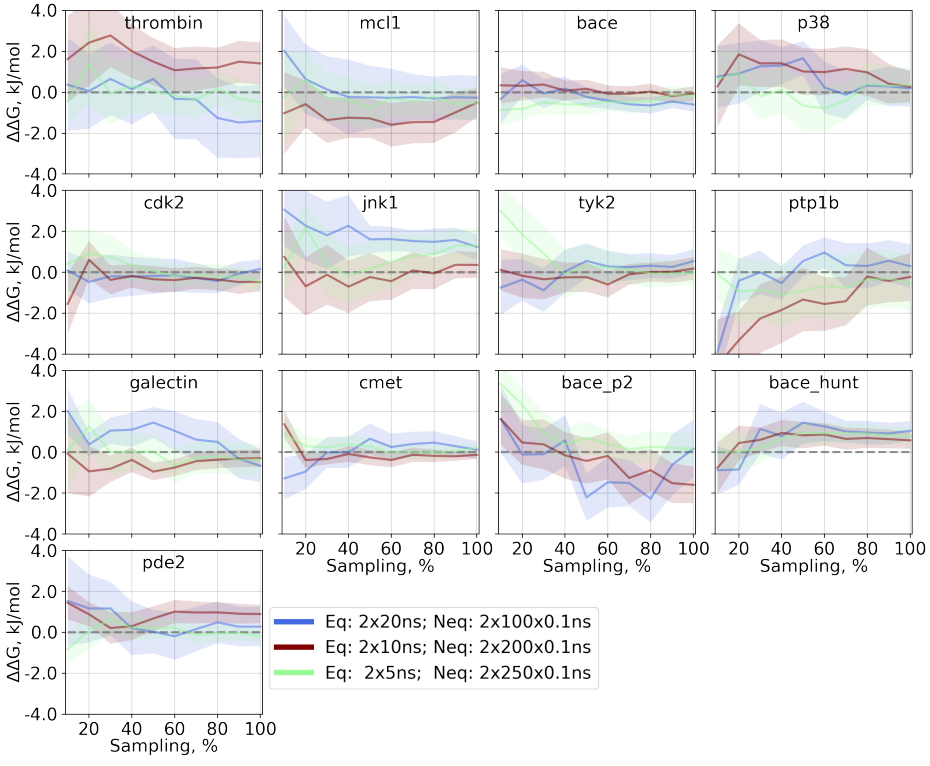


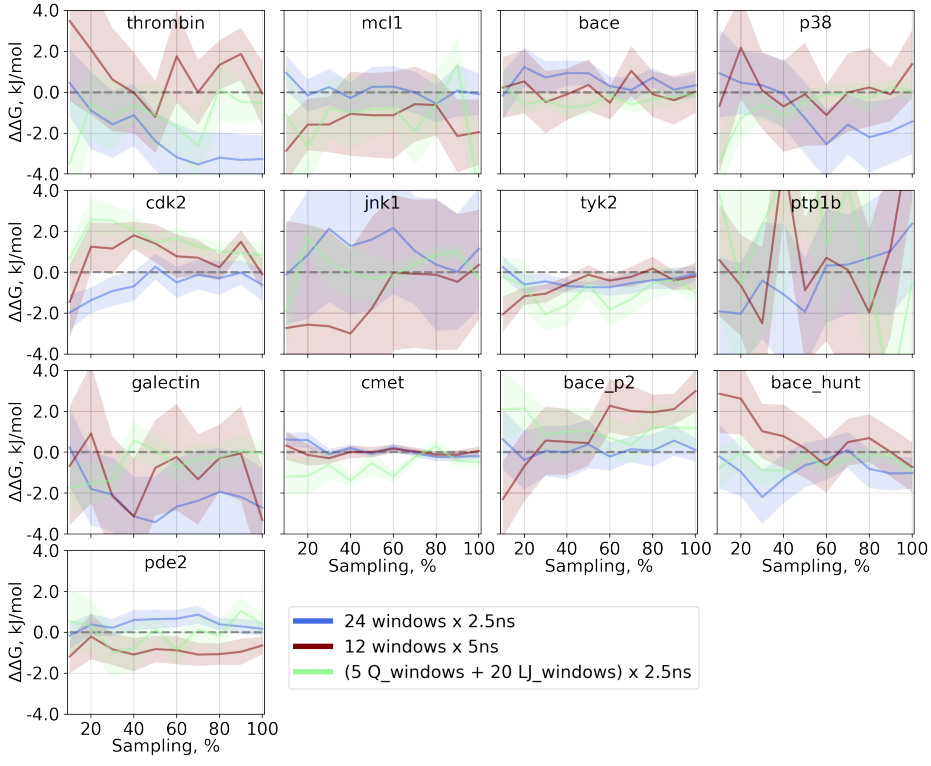


Non-equilibrium TI

Equilibrium FEP

Ligands





Non-equilibrium TI

Equilibrium FEP