Gromacs/pmx for large scale alchemical protein-ligand binding affinity screening

Vytautas Gapsys
vgapsys@gwdg.de

Max Planck Institute for Multidisciplinary Sciences

Computational Biomolecular Dynamics Group (Dr. Bert de Groot)

April 21, 2022

Introduction: Alchemy

Computational Alchemy

- Molecular dynamics based method
- Free energy calculation

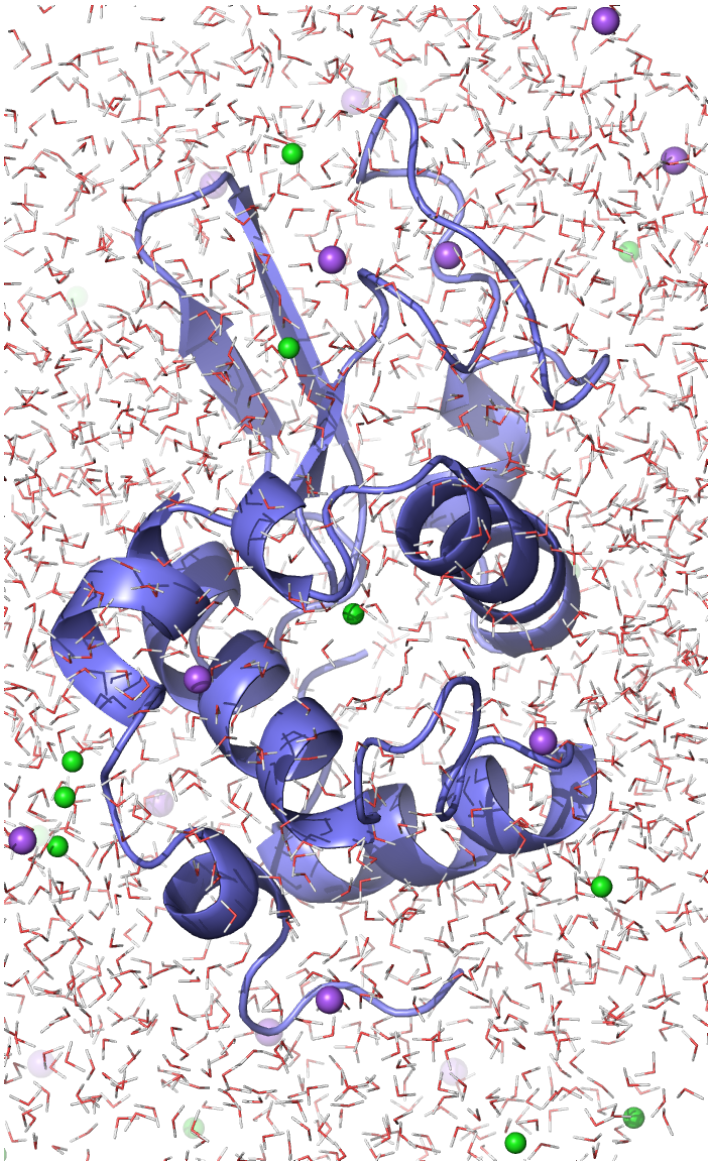
Introduction: Molecular Dynamics simulations

Molecular Dynamics simulations

- Thermostat: scale velocities
- Barostat: scale box dimensions
- Integrate equations of motion over time

System setup

- Fully atomistic representation
- Explicit solvation
- Ions



Introduction: Molecular Dynamics simulations

Introduction: Molecular Dynamics simulations

Molecular dynamics simulations

- Well defined statistical ensembles
- Atomistic resolution

Alchemical Methods

Aim

Calculate change in thermostability due to a mutation.

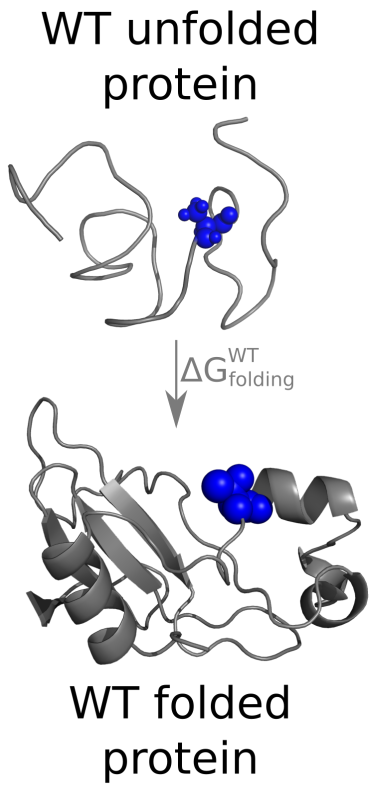
Alchemical Methods

Aim

Calculate change in thermostability due to a mutation.

Counting

- Simulate the process of folding the WT protein.
- Count folded/unfolded states.
- Extract $\Delta G^{WT}_{folding}$ from the frequencies.



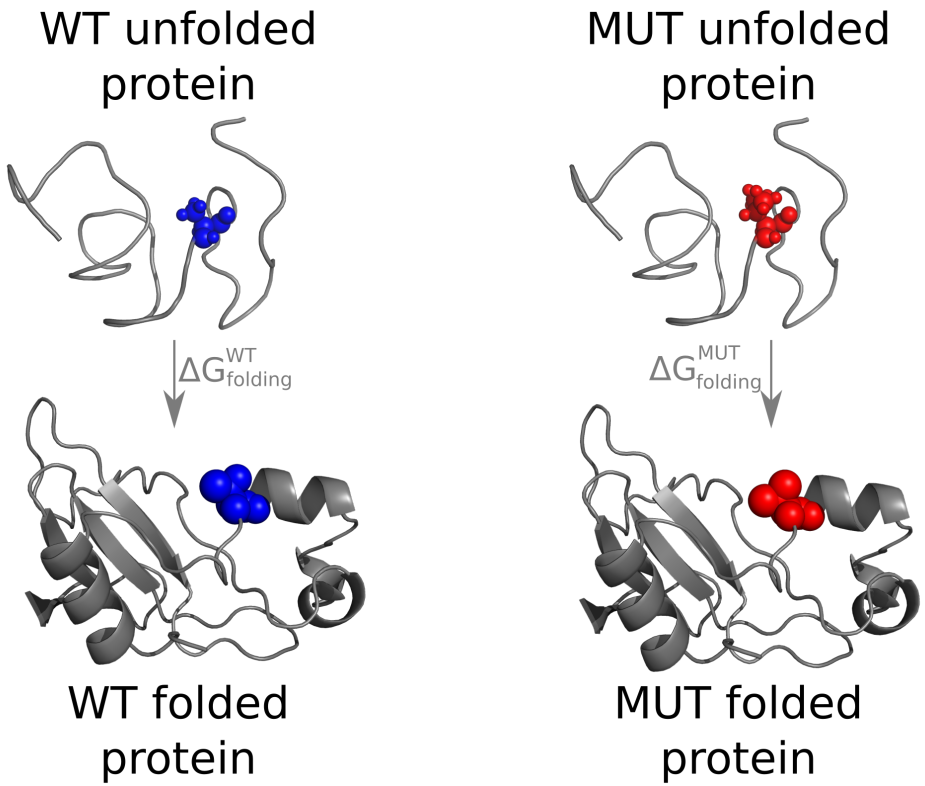
Alchemical Methods

Aim

Calculate change in thermostability due to a mutation.

Counting

- Do the same for the MUT protein:
  $\Delta\Delta G = \Delta G^{MUT}_{folding} - \Delta G^{WT}_{folding}$



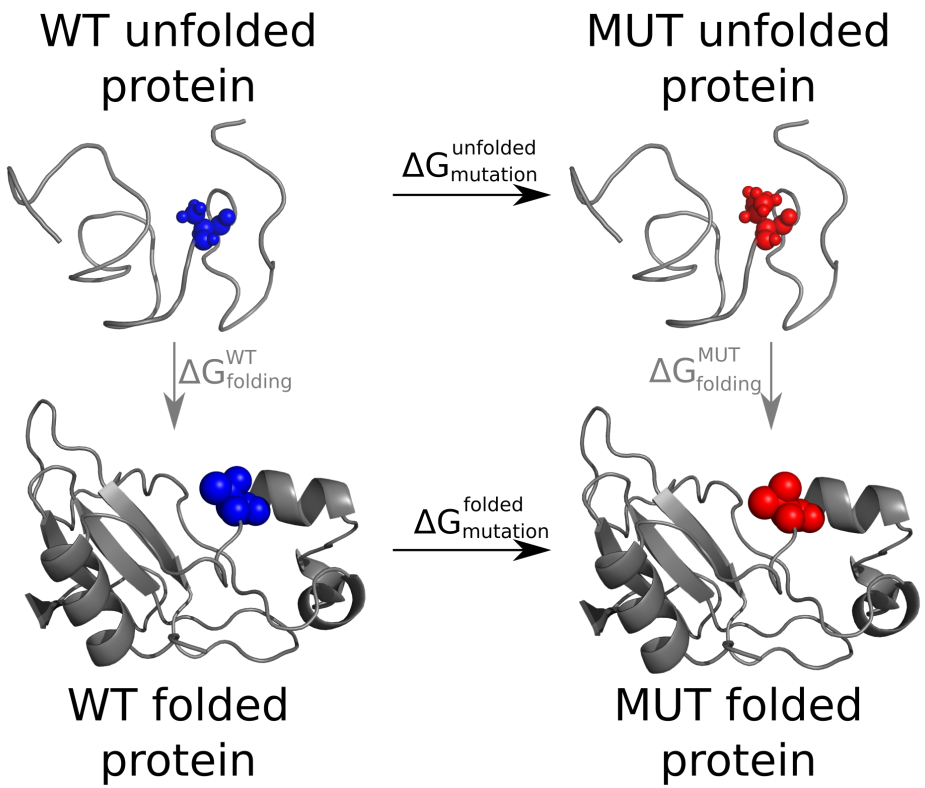
Alchemical Methods

Problem

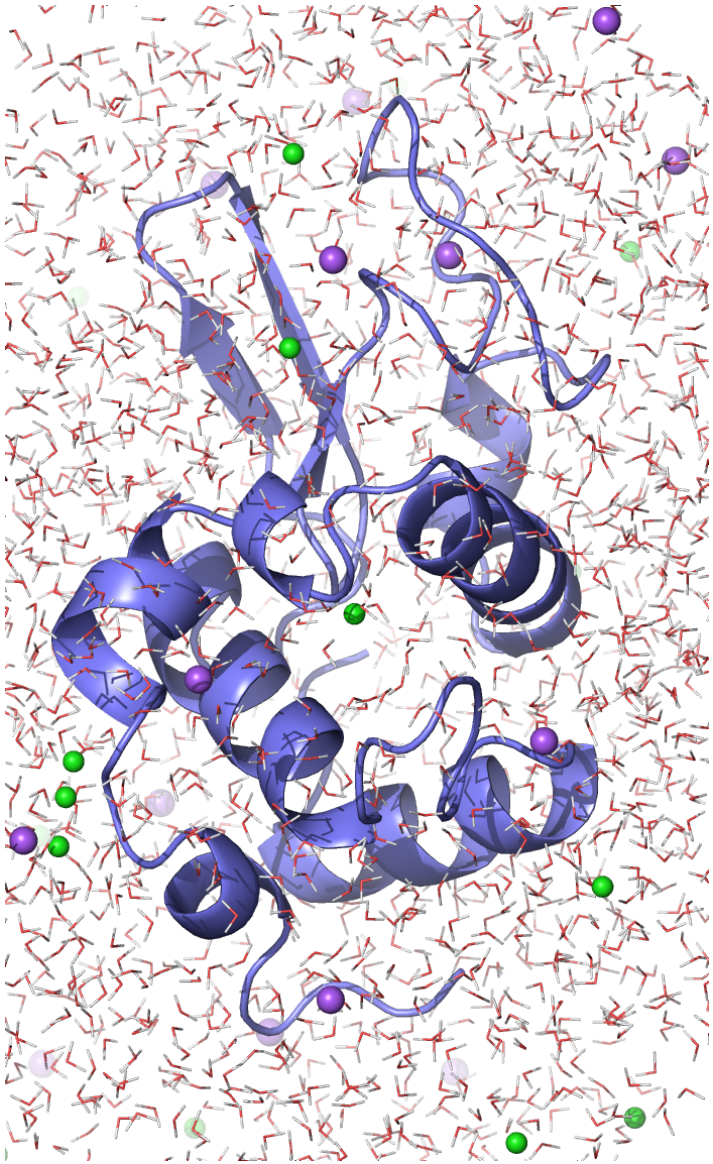
Folding simulations are computationally expensive.

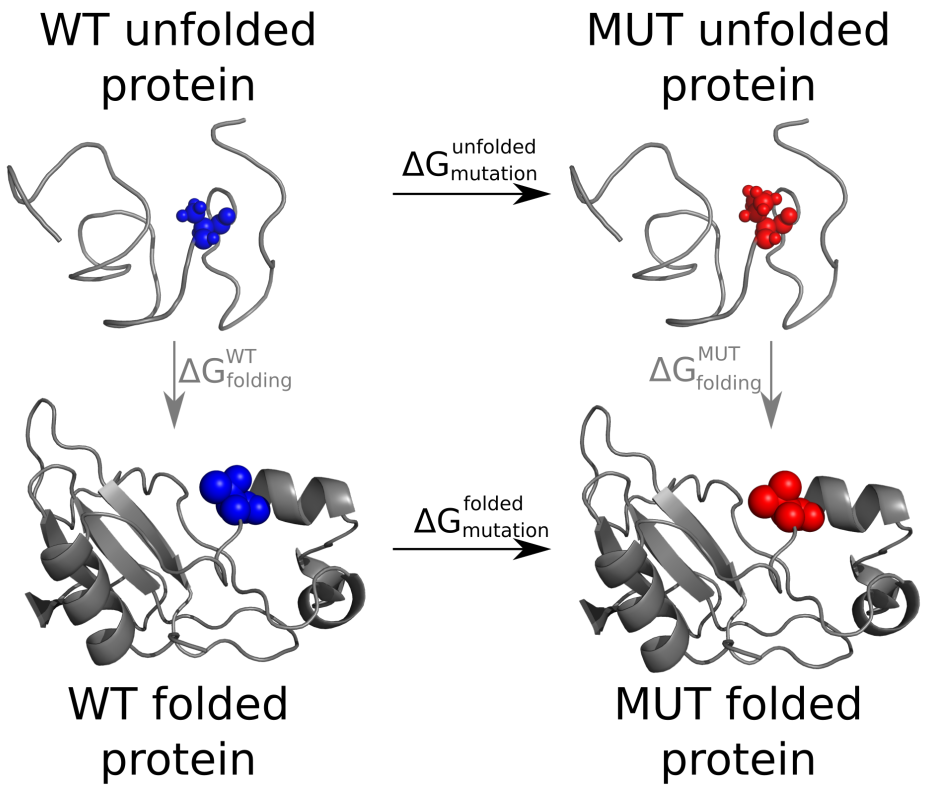
Alchemical solution

- Traverse the thermodynamic cycle in the horizontal directions.
- Use a modified hamiltonian: $H(\lambda) = (1 - \lambda)H_0 + \lambda H_1$.
- The same double free energy difference is recovered: $\Delta\Delta G = \Delta G^{folded}_{mutation} - \Delta G^{unfolded}_{mutation}$

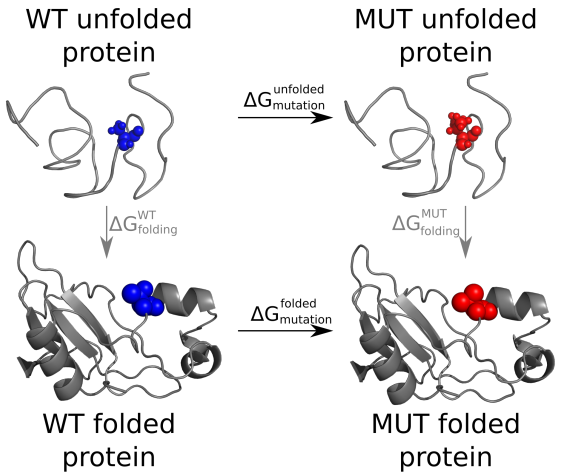


Alchemical Methods

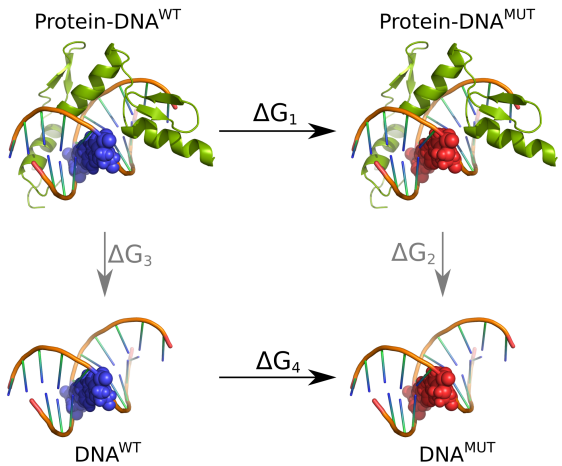




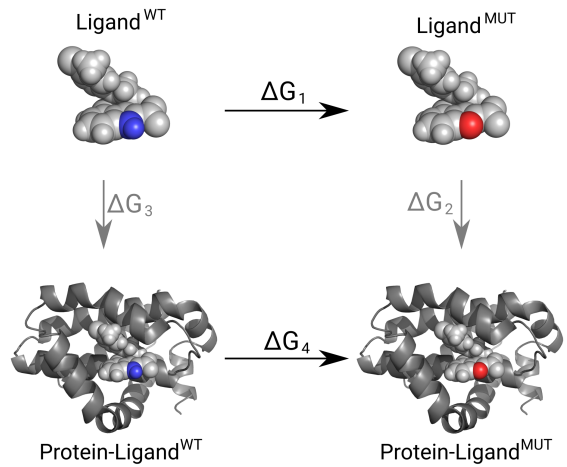
Gromacs/pmx based alchemistry



Amino acid mutations

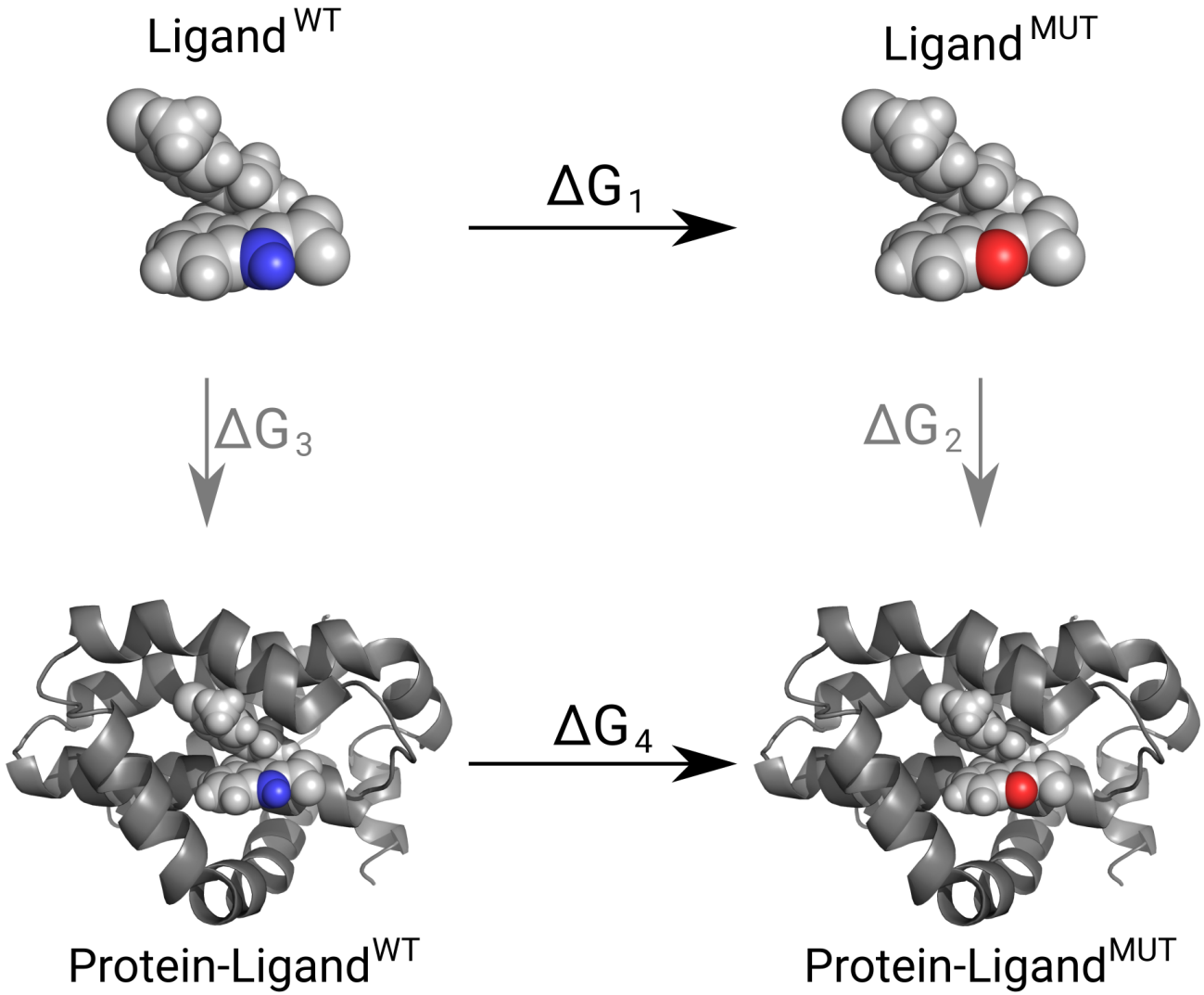


Nucleotide mutations

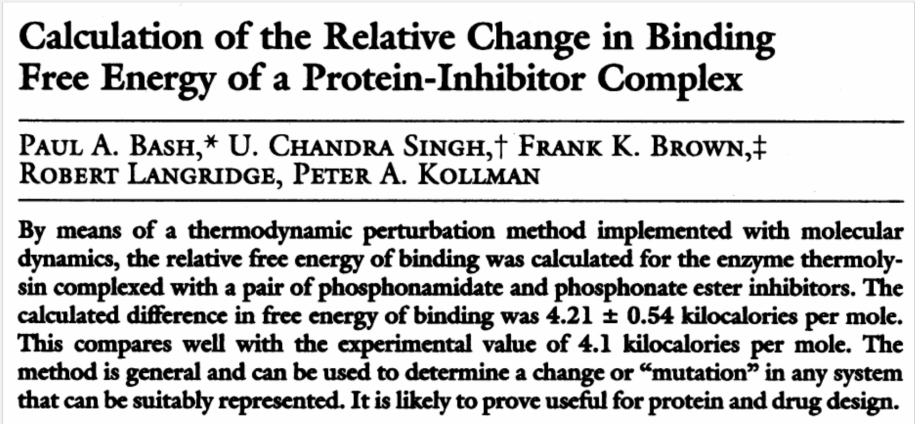


Ligand modifications

Relative Protein-Ligand binding free energy



Protein-Ligand binding

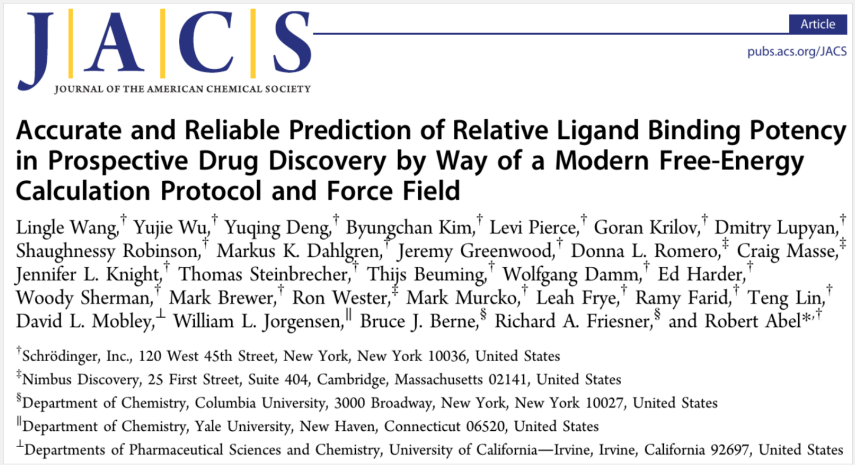


Calculation of the Relative Change in Binding Free Energy of a Protein-Inhibitor Complex

PAUL A. BASH,* U. CHANDRA SINGH,† FRANK K. BROWN,‡
ROBERT LANGRIDGE, PETER A. KOLLMAN

By means of a thermodynamic perturbation method implemented with molecular dynamics, the relative free energy of binding was calculated for the enzyme thermolysin complexed with a pair of phosphonamidate and phosphonate ester inhibitors. The calculated difference in free energy of binding was 4.21 ± 0.54 kilocalories per mole. This compares well with the experimental value of 4.1 kilocalories per mole. The method is general and can be used to determine a change or "mutation" in any system that can be suitably represented. It is likely to prove useful for protein and drug design.

TI for a single free energy value. Science, 1987



J|A|C|S JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

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

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

[†]Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States
[‡]Nimbus Discovery, 25 First Street, Suite 404, Cambridge, Massachusetts 02141, United States
[§]Department of Chemistry, Columbia University, 3000 Broadway, New York, New York 10027, United States
[∥]Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States
[⊥]Departments of Pharmaceutical Sciences and Chemistry, University of California—Irvine, Irvine, California 92697, United States

FEP for 330 relative free energies. JACS, 2015

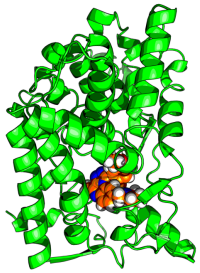
Protein-Ligand binding

Application:
482 ligand modifications in protein-ligand binding

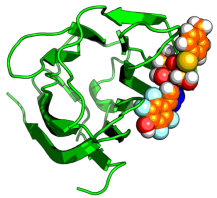
Gapsys, Perez-Benito, Aldeghi, Seeliger, van Vlijmen, Tresadern, de Groot,
Chemical Science, 2020

Protein-Ligand complexes

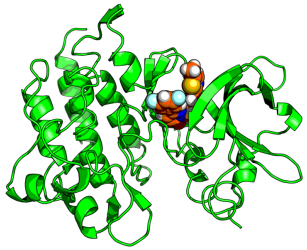
- 11 systems
- 482 mutations



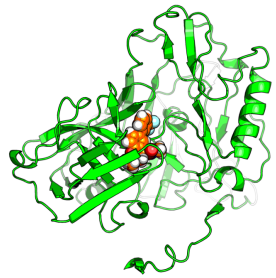
PDE2: 21 ligand
34 perturbations



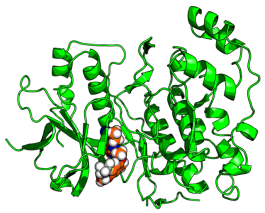
Galectin: 8 ligands
8 perturbations



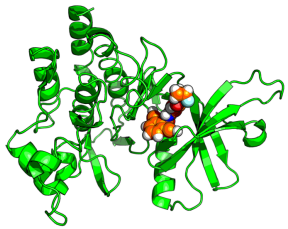
cMet: 12 ligands
25 perturbations



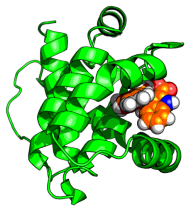
BACE: 80 ligands
144 perturbations
(divided in 3 sets)



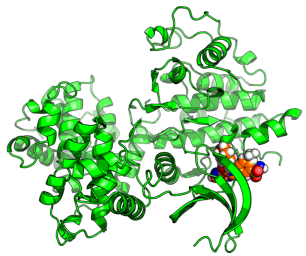
JNK1: 21 ligands
31 perturbations



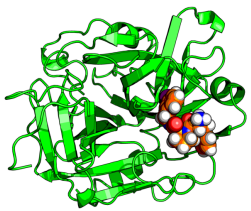
TYK2: 16 ligands
24 perturbations



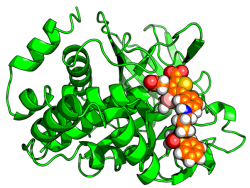
MCL1: 42 ligands
71 perturbations



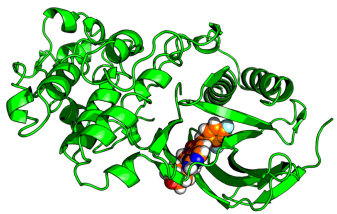
CDK2: 16 ligands
25 perturbations



Thrombin: 11 ligands
16 perturbations

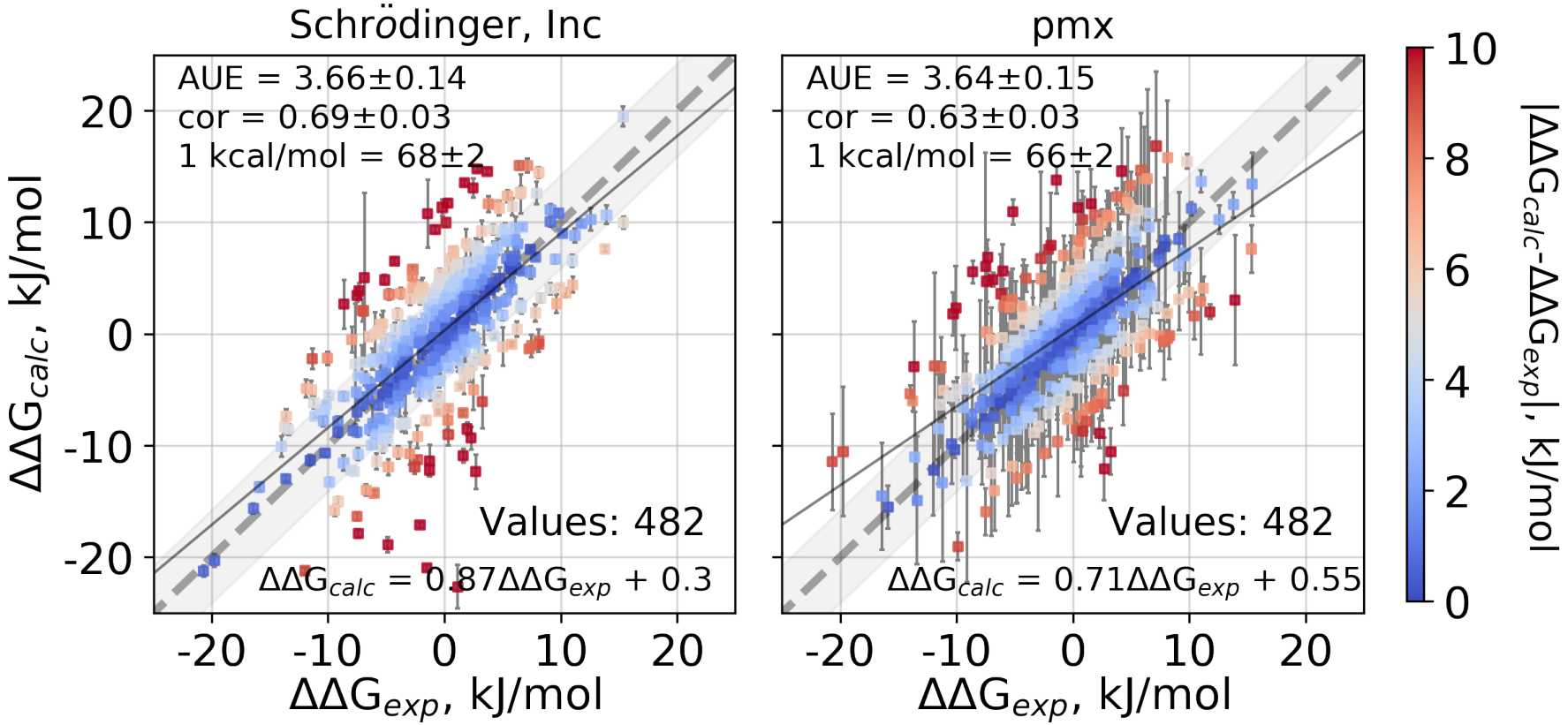


PTP1b: 23 ligands
49 perturbations

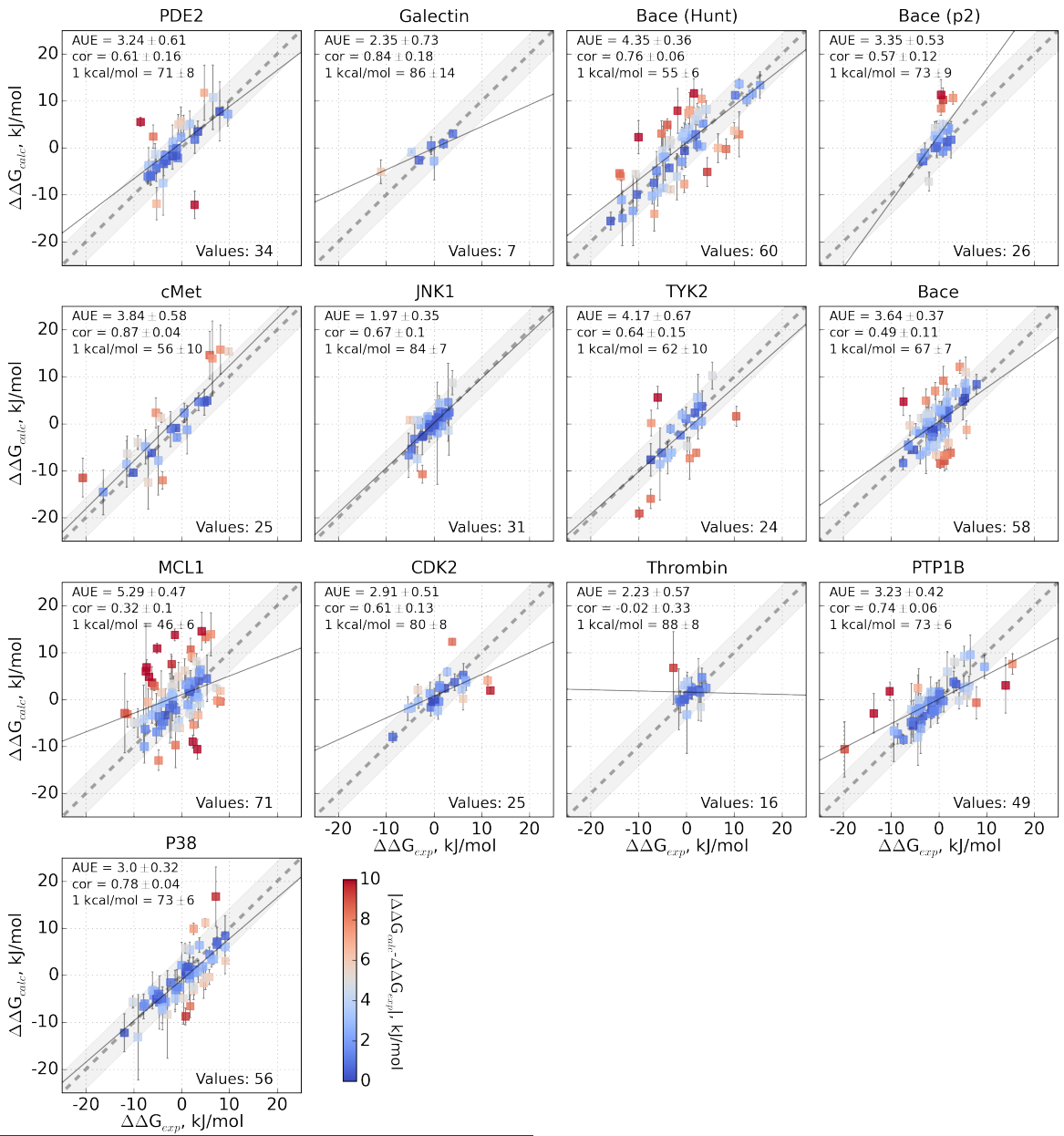


P38: 34 ligands
56 perturbations

Overall results



Results: by case



Ligands

High throughput ΔΔG calculations

Gapsys, Hahn, Tresadern, Mobley, Rampp, de Groot, JCIM, 2022

Kutzner, Kniep, Cherian, Nordstrom, Grubmüller, de Groot, Gapsys, JCIM, 2022

High throughput ΔΔG calculations

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

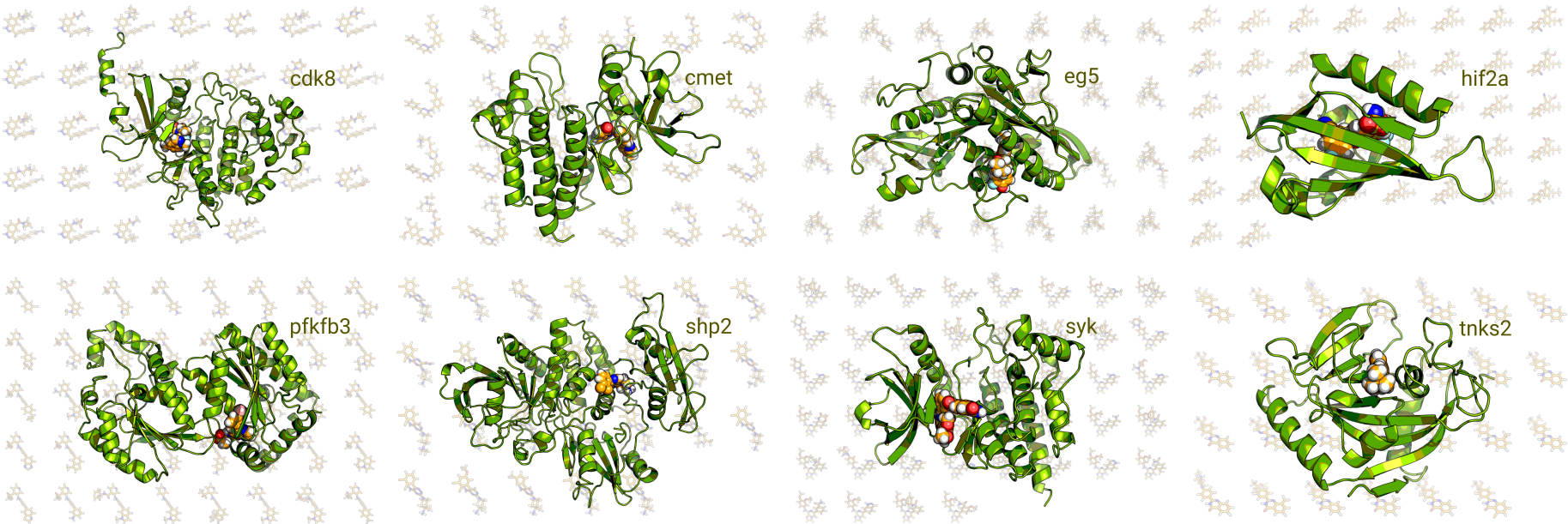
pubs.acs.org/jcim Article

Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects

Christina E. M. Schindler,* Hannah Baumann, Andreas Blum, Dietrich Böse, Hans-Peter Buchstaller, Lars Burgdorf, Daniel Cappel, Eugene Chekler, Paul Czodrowski, Dieter Dorsch, Merveille K. I. Eguida, Bruce Follows, Thomas Fuchß, Ulrich Grädler, Jakub Gunera, Theresa Johnson, Catherine Jorand Lebrun, Srinivasa Karra, Markus Klein, Tim Knehans, Lisa Koetzner, Mireille Krier, Matthias Leiendecker, Birgitta Leuthner, Liwei Li, Igor Mochalkin, Djordje Musil, Constantin Neagu, Friedrich Rippmann, Kai Schiemann, Robert Schulz, Thomas Steinbrecher, Eva-Maria Tanzer, Andrea Unzue Lopez, Ariele Viacava Follis, Ansgar Wegener, and Daniel Kuhn*

Benchmark dataset assembled by Merck KGaA

High throughput ΔΔG calculations

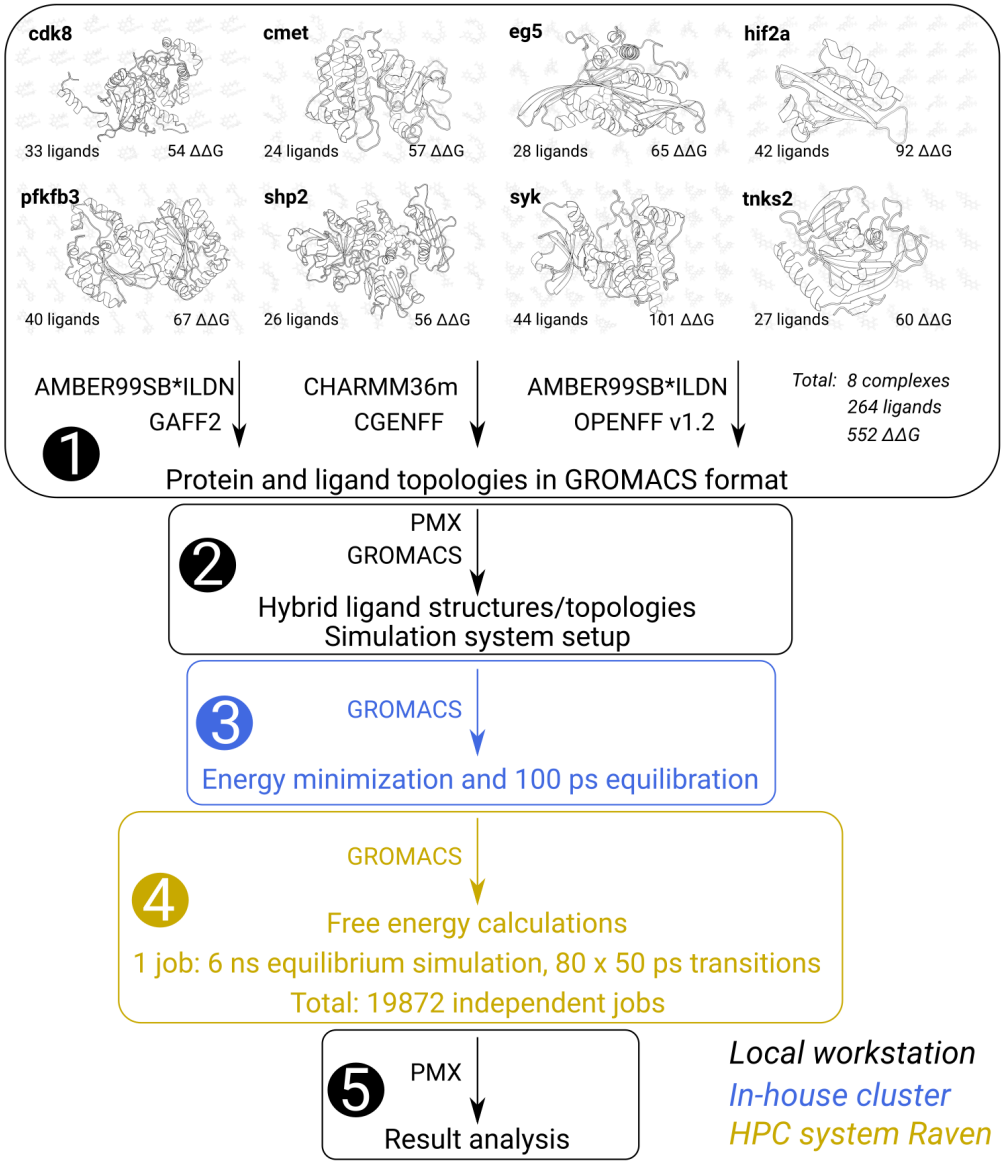


- 8 systems
- >500 ΔΔG
- >200 μs simulation
- 3 force fields
- 3 independent replicas for each calculation

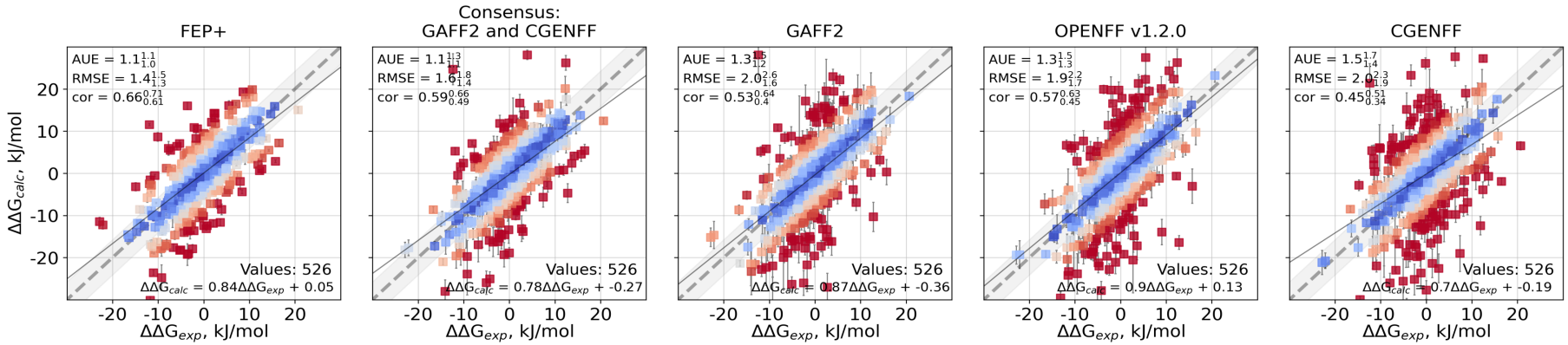
High throughput ΔΔG calculations

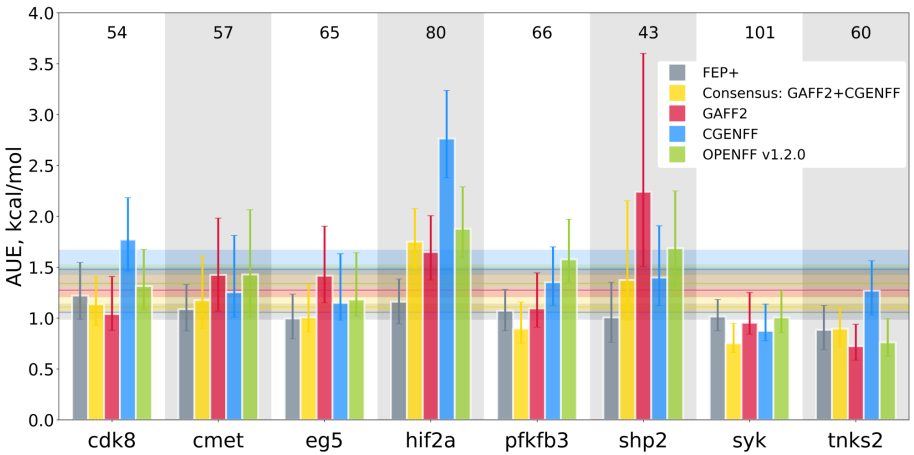
- MPCDF cluster Raven
- Using 480 nodes simultaneously (46k cores)
- 3 days of calculations: 3.4 million core hours

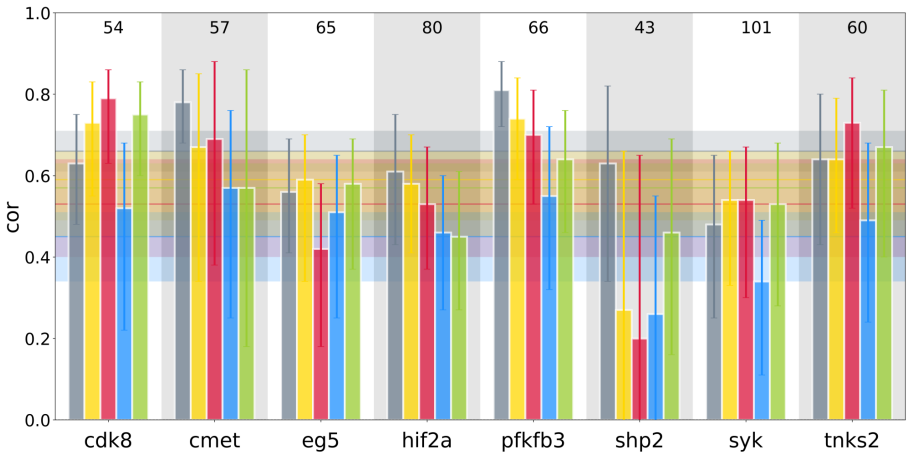
High throughput ΔΔG calculations



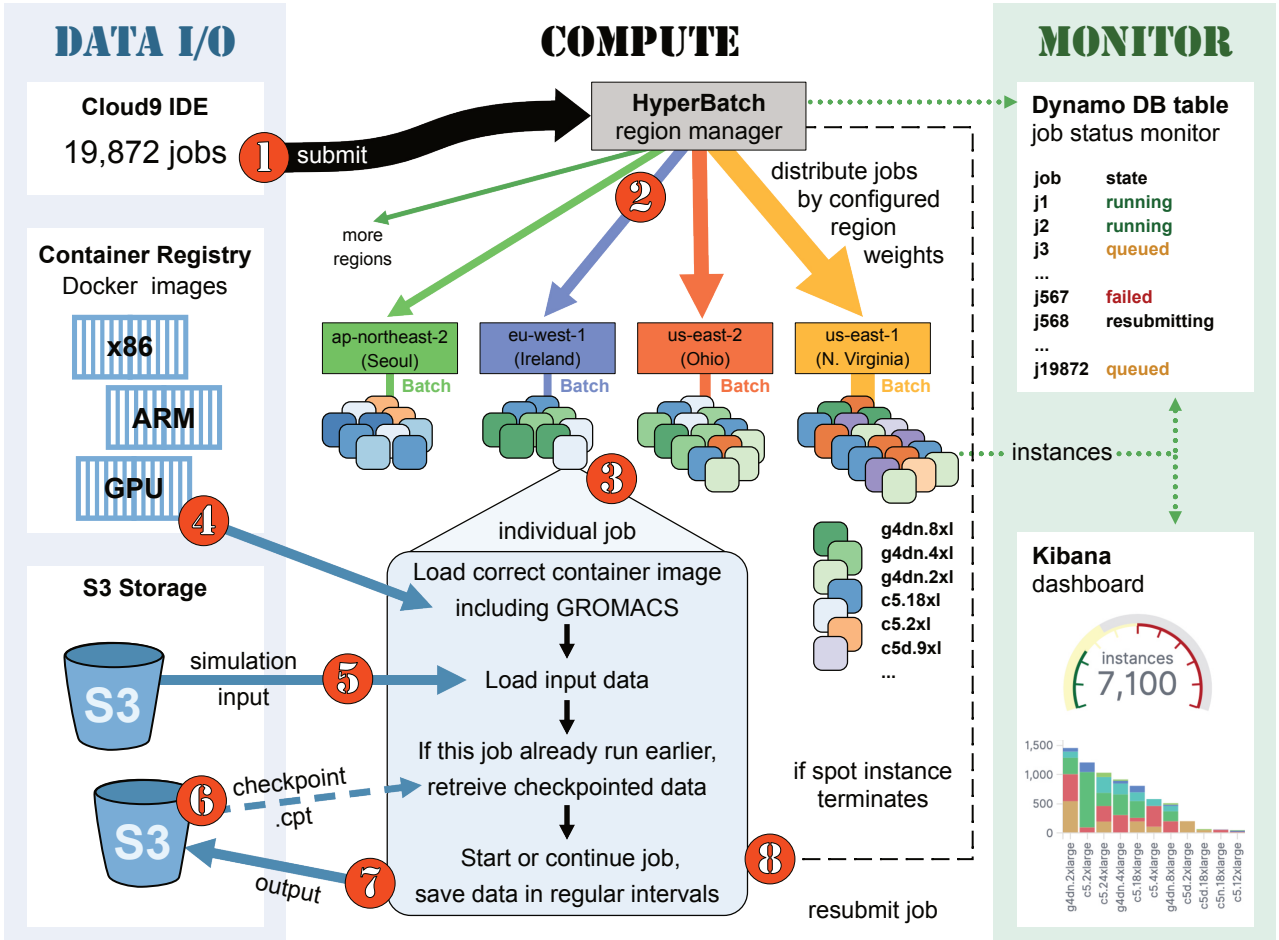
High throughput ΔΔG calculations



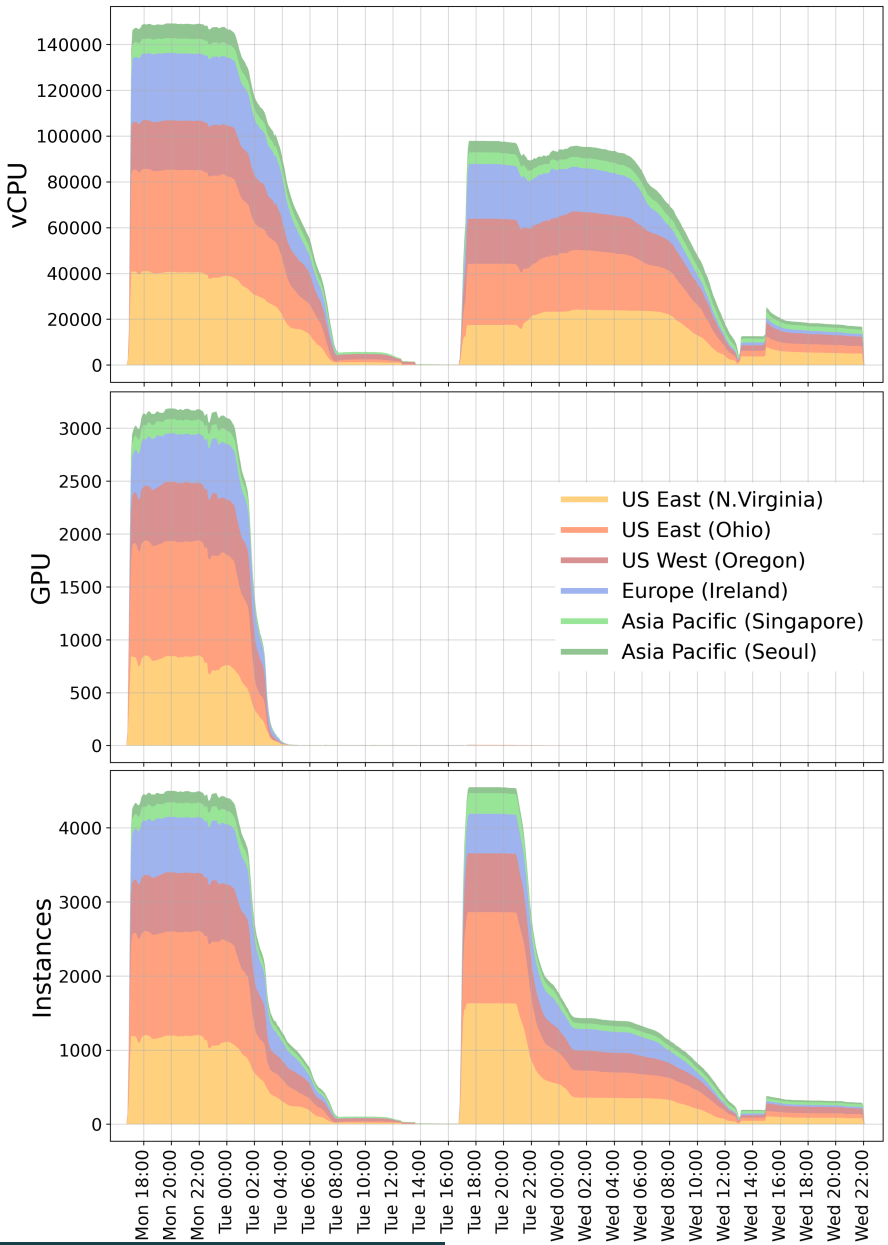




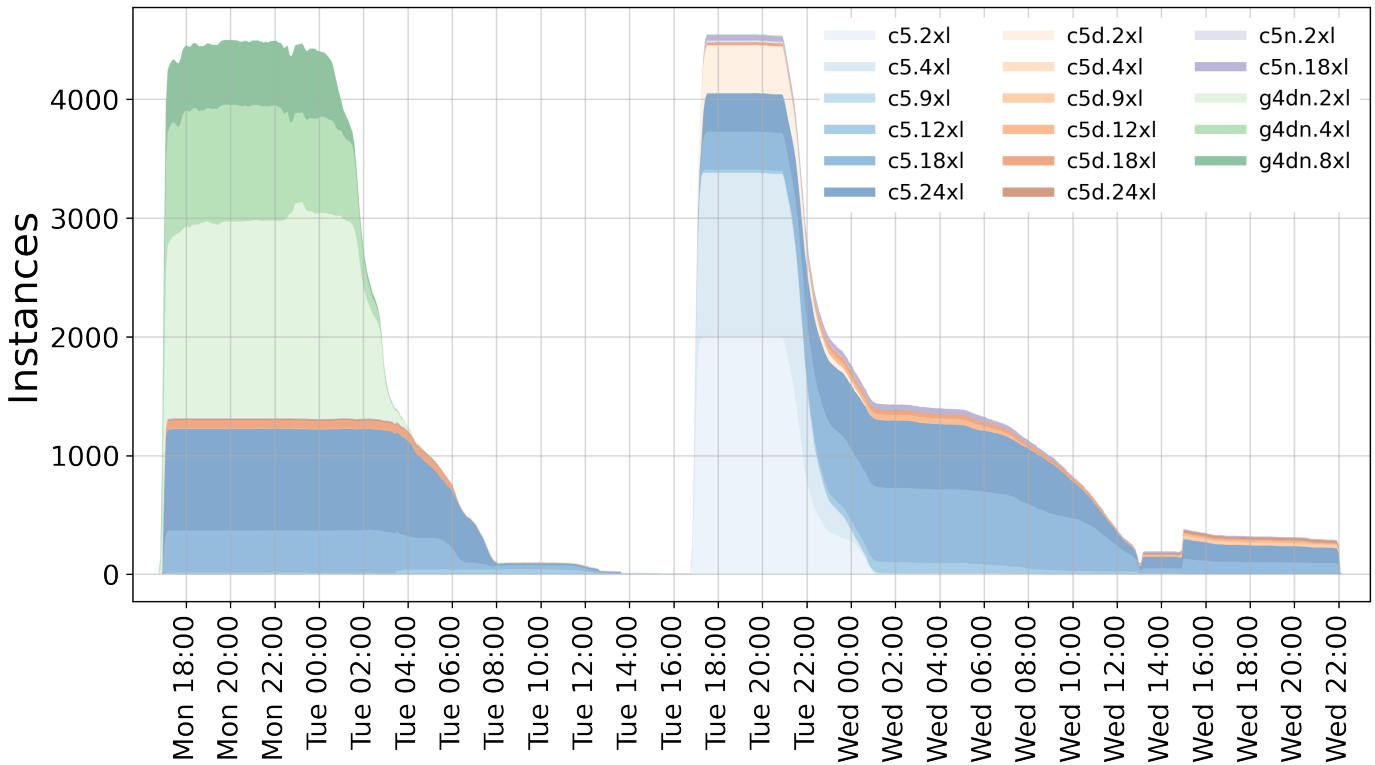
ΔΔG calculation in the cloud



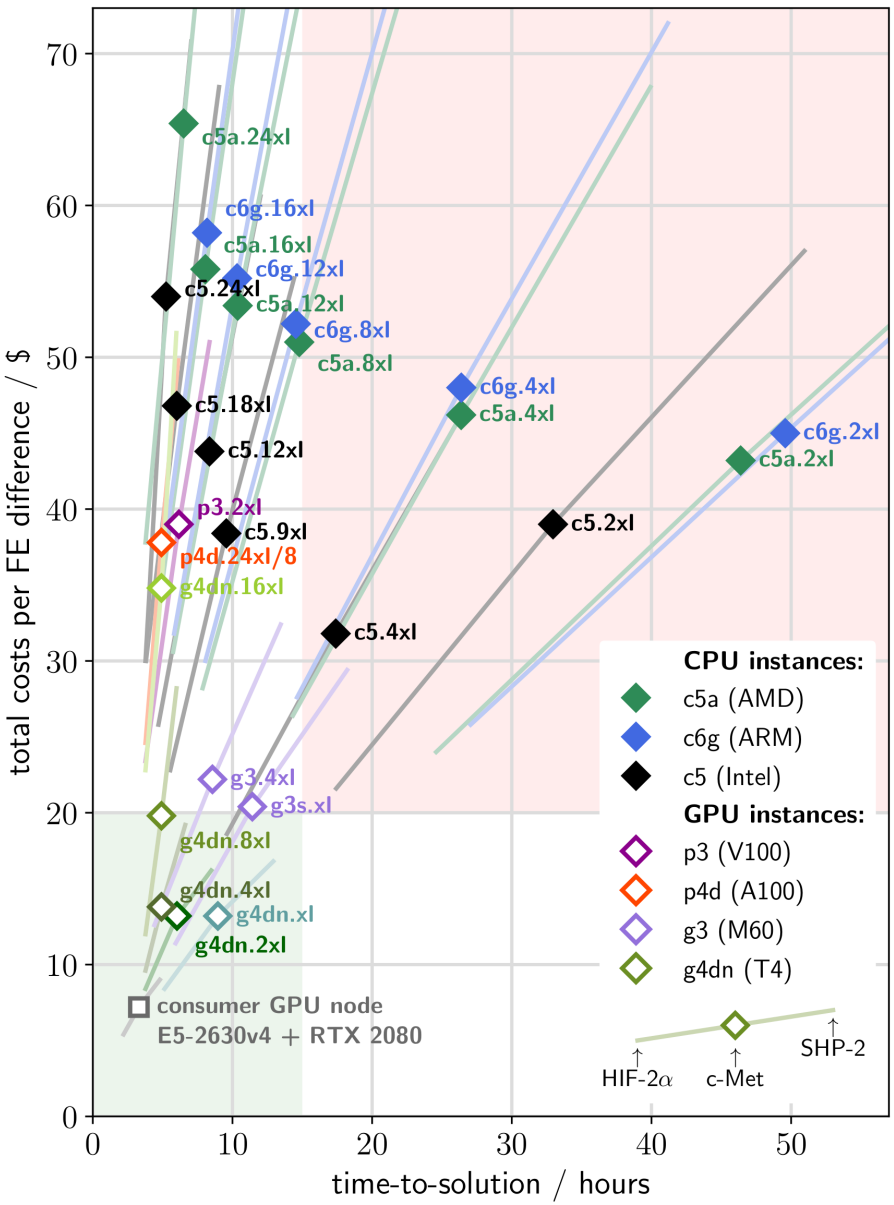
ΔΔG calculation in the cloud



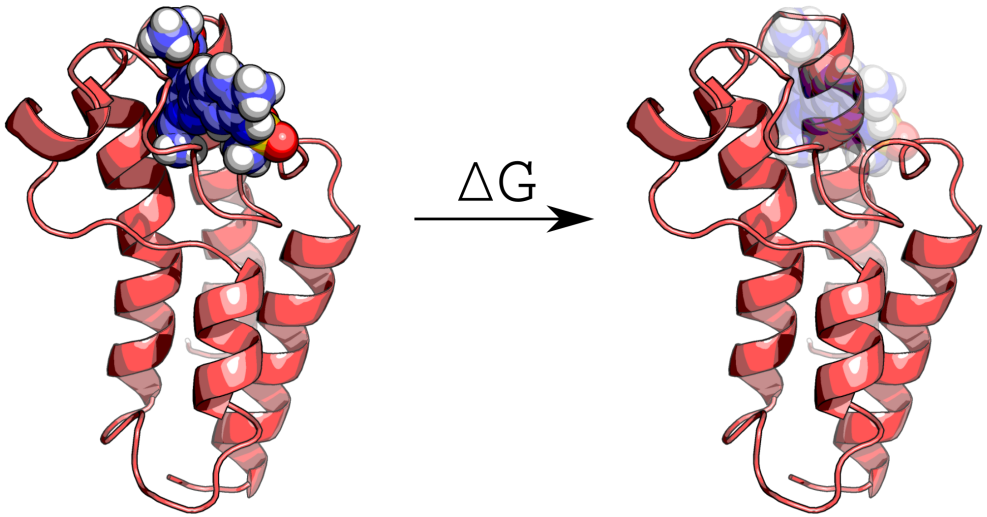
ΔΔG calculation in the cloud



AWS costs for ΔΔG calculation



Absolute protein-ligand binding ΔG



Absolute protein-ligand binding ΔG

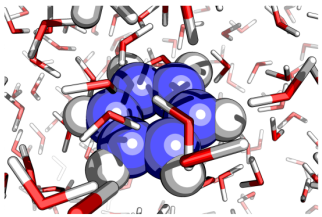
Absolute protein-ligand binding ΔG: Part 1

Explore applicability of FEP, HREX-FEP and Non-equilibrium approaches

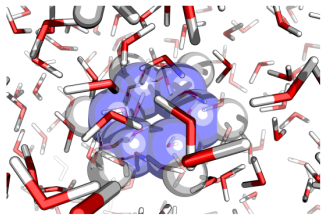
Gapsys, Yildirim, Aldeghi, Khalak, van der Spoel, de Groot, Commun. Chem., 2021

Non-equilibrium free energy calculation protocol

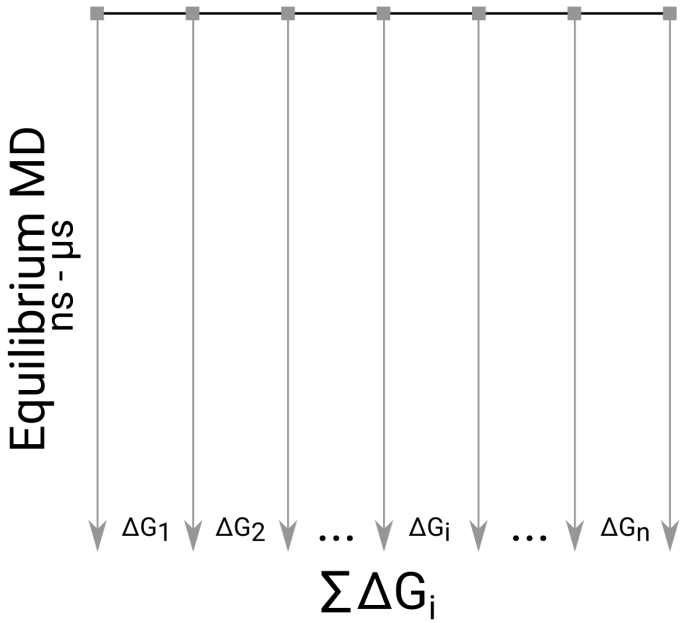
Equilibrium FEP



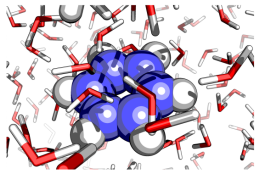
State A (λ=0)



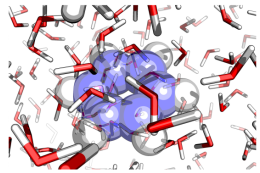
State B (λ=1)



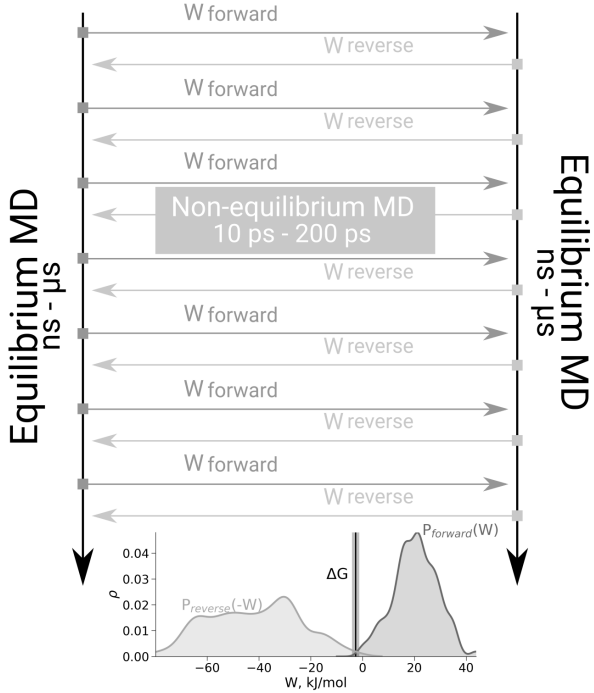
Non-Equilibrium TI



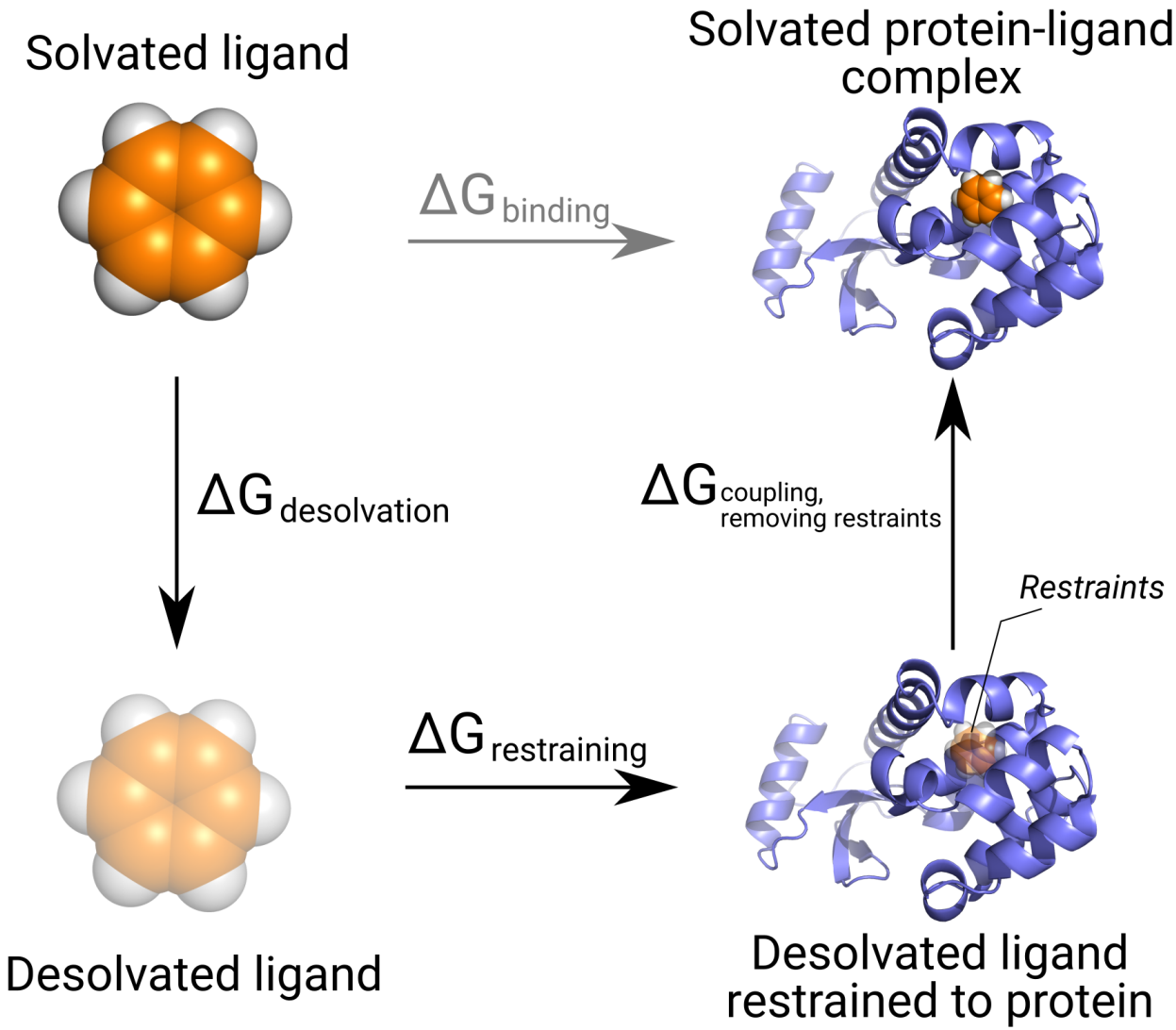
State A (λ=0)



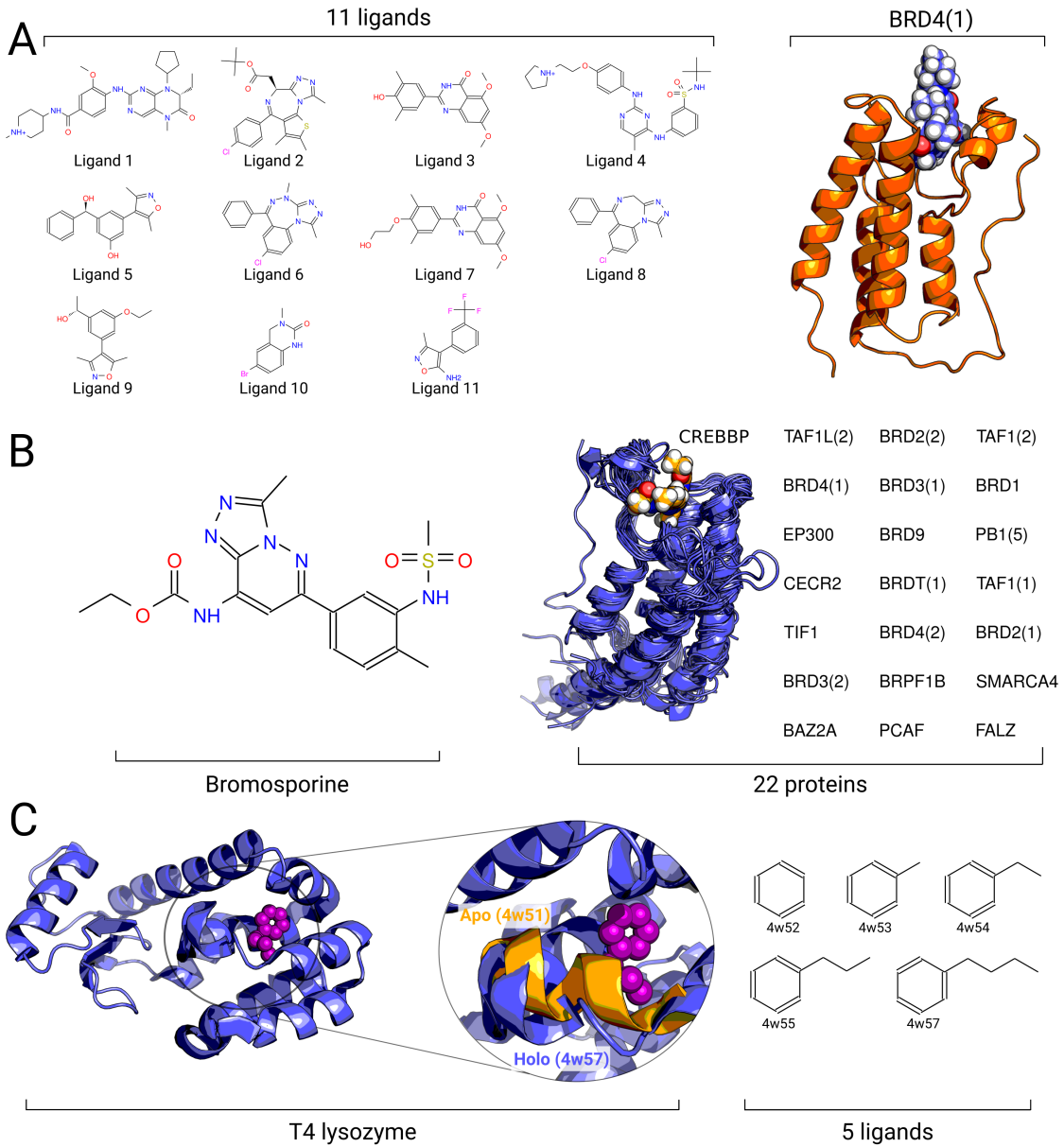
State B (λ=1)



Absolute ΔG

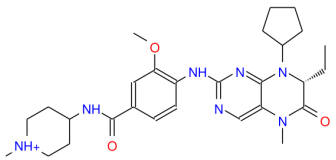

Solvated ligand
Solvated protein-ligand complex
ΔG binding
ΔG desolvation
ΔG coupling, removing restraints
Restraints
ΔG restraining
Desolvated ligand
Desolvated ligand restrained to protein

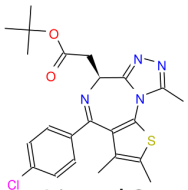
Studied systems

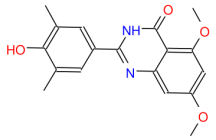


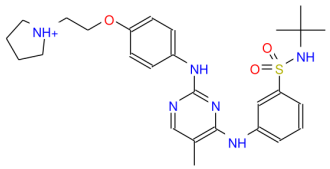
BRD4(1) specificity

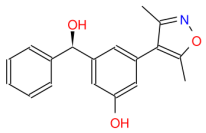
11 ligands

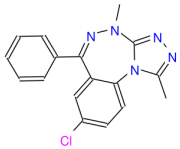

Ligand 1

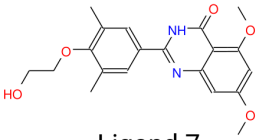

Ligand 2

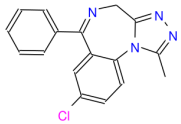

Ligand 3

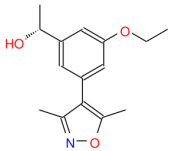

Ligand 4

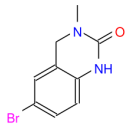

Ligand 5

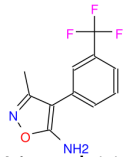

Ligand 6


Ligand 7

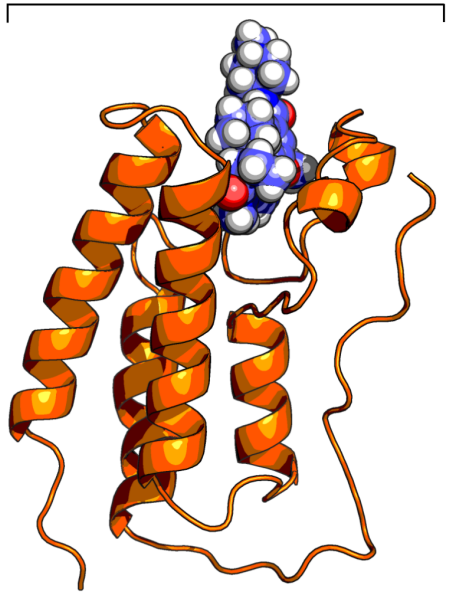

Ligand 8


Ligand 9

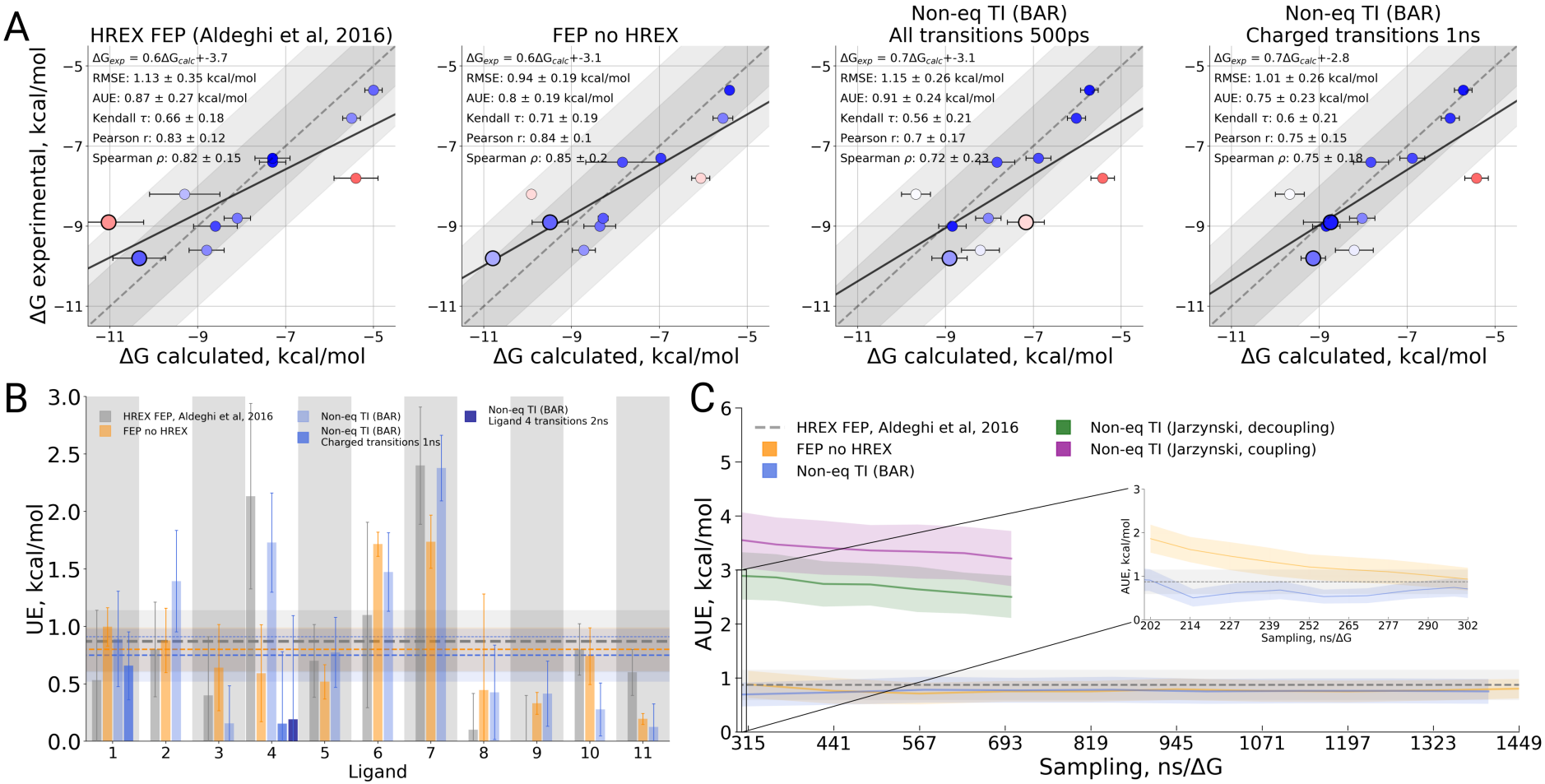

Ligand 10


Ligand 11

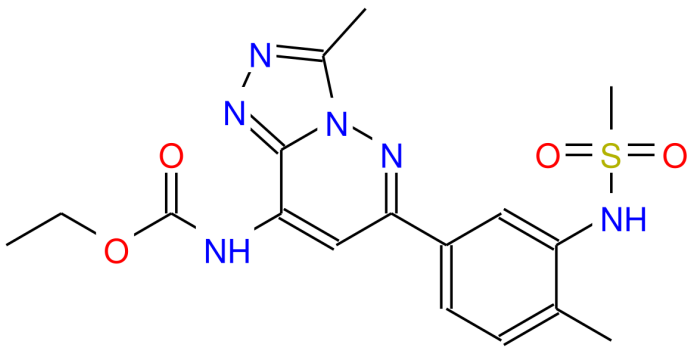
BRD4(1)



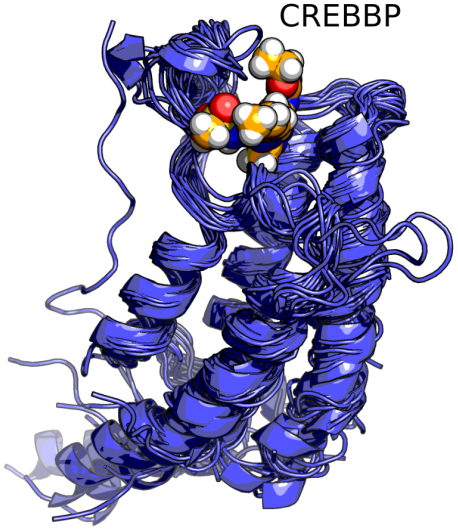
BRD4(1) specificity: overall accuracy



Bromosporine



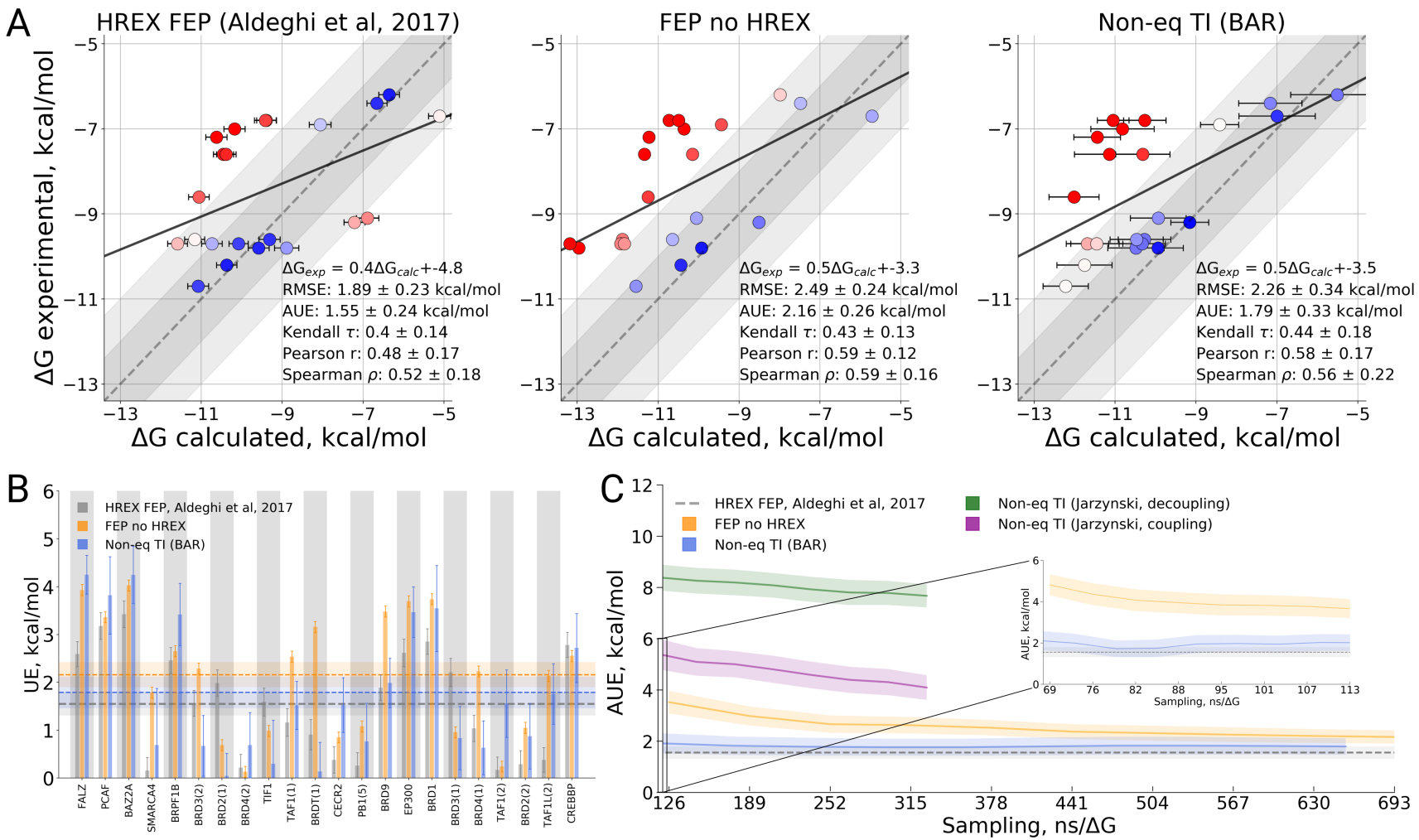
Bromosporine



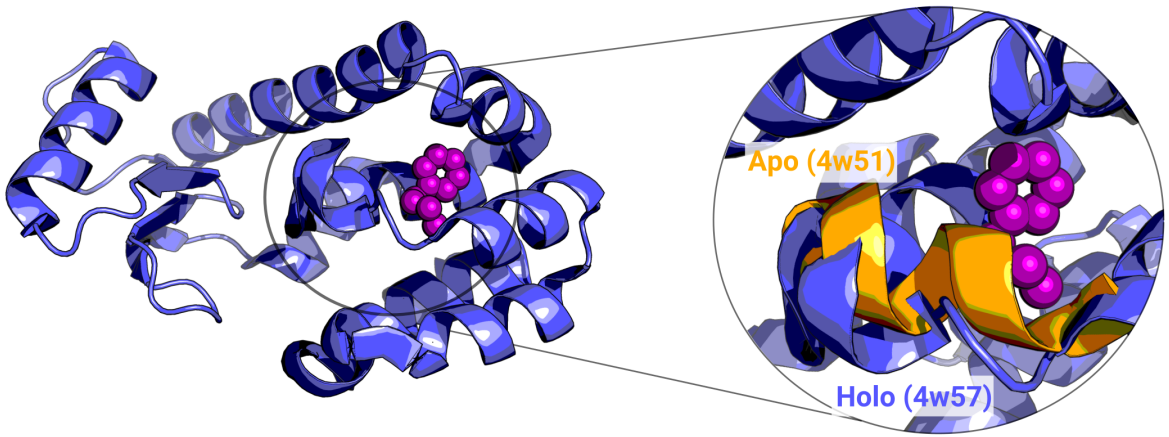
CREBBP	TAF1L(2)	BRD2(2)	TAF1(2)
	BRD4(1)	BRD3(1)	BRD1
	EP300	BRD9	PB1(5)
	CECR2	BRDT(1)	TAF1(1)
	TIF1	BRD4(2)	BRD2(1)
	BRD3(2)	BRPF1B	SMARCA4
	BAZ2A	PCAF	FALZ

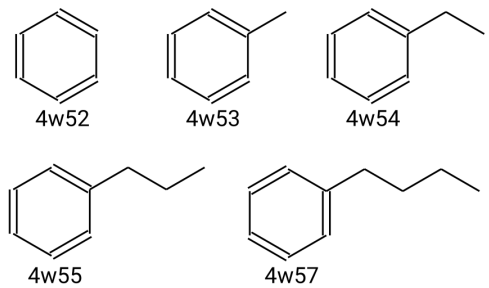
22 proteins

Bromosporine: overall accuracy



T4 Lysozyme (L99A)



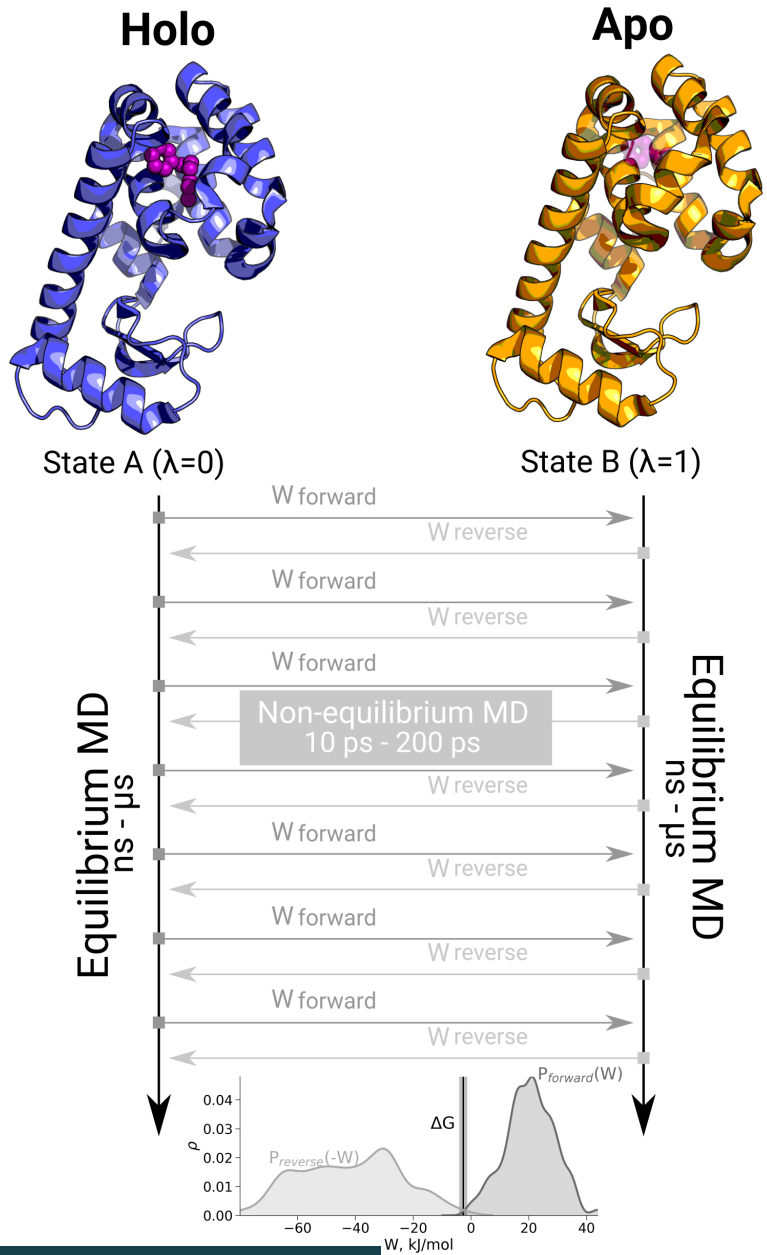


T4 lysozyme

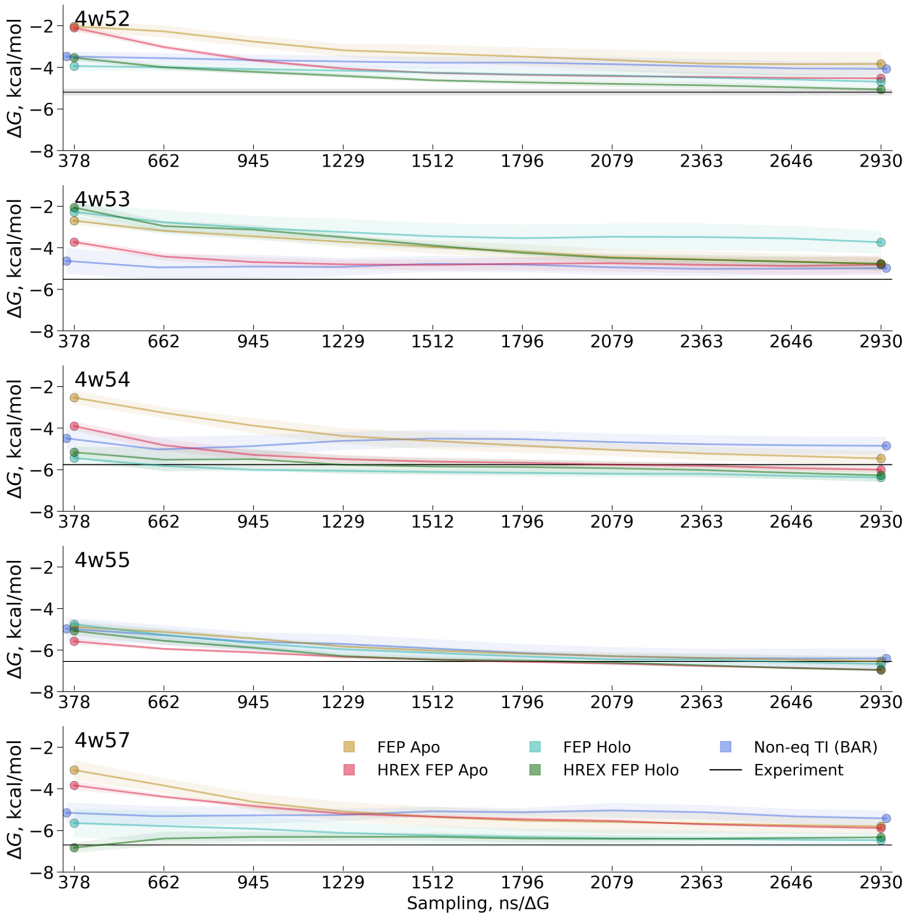
5 ligands

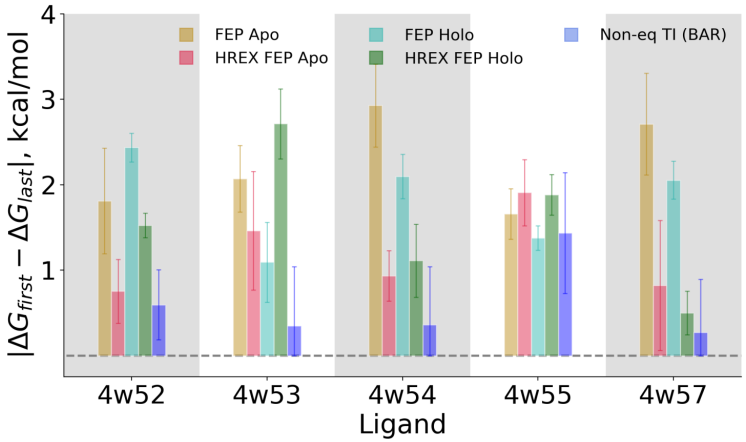
- Apo and Holo states differ significantly
- *Lim, Wang, Abel, Mobley, 2016, JCTC* observed that long FEP simulations needed to converge $\Delta\Delta G$
- Non-equilibrium method allows combining different Apo and Holo states

T4 Lysozyme


Holo
Apo
State A (λ=0)
State B (λ=1)
W forward
W reverse
W forward
W reverse
W forward
Non-equilibrium MD
10 ps - 200 ps
W reverse
W forward
W reverse
W forward
W reverse
W forward
W reverse
Equilibrium MD
ns - µs
Equilibrium MD
ns - µs
0.04
0.03
0.02
0.01
0.00
ρ
P_reverse(-W)
ΔG
P_forward(W)
−60
−40
−20
0
20
40
W, kJ/mol

T4 Lysozyme


4w52
4w53
4w54
4w55
4w57
ΔG, kcal/mol
−2
−4
−6
−8
378
662
945
1229
1512
1796
2079
2363
2646
2930
Sampling, ns/ΔG
FEP Apo
HREX FEP Apo
FEP Holo
HREX FEP Holo
Non-eq TI (BAR)
Experiment


$|\Delta G_{first} - \Delta G_{last}|$, kcal/mol
4
3
2
1
FEP Apo
HREX FEP Apo
FEP Holo
HREX FEP Holo
Non-eq TI (BAR)
4w52
4w53
4w54
4w55
4w57
Ligand

Summary: absolute ΔG Part 1

Non-equilibrium approach

- Can be used for absolute ΔG
- Converges faster than FEP, comparably to HREX-FEP
- Allows taking into account Apo and Holo conformations

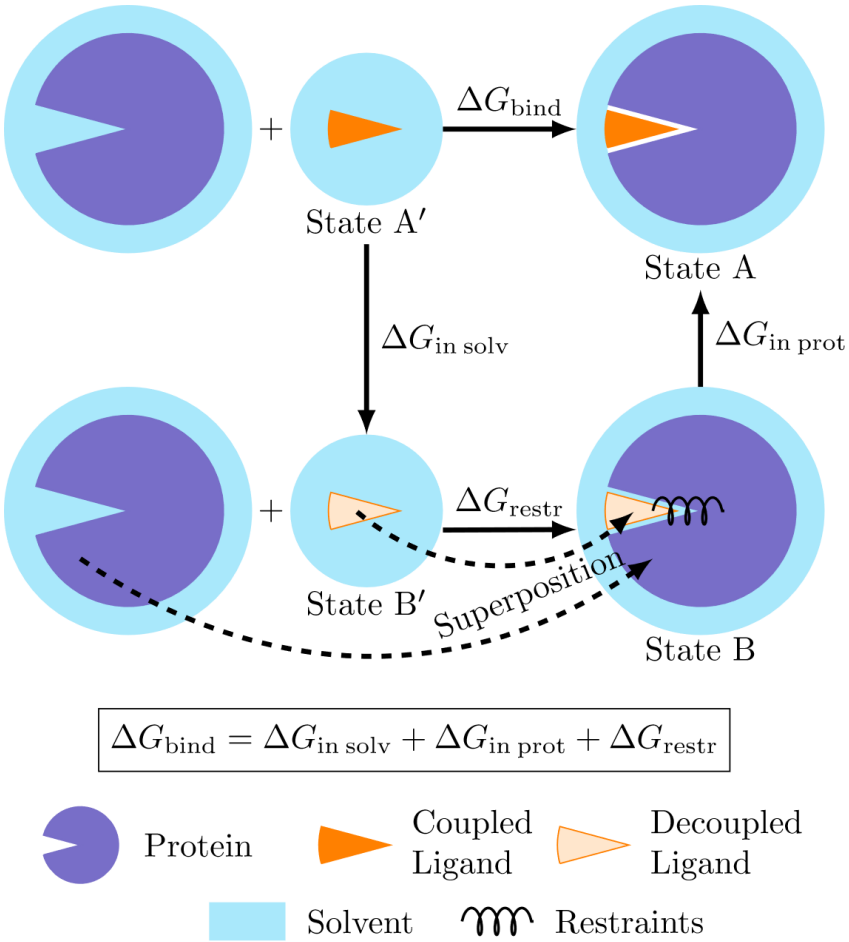
Absolute protein-ligand binding ΔG: Part 2

Large scale absolute protein-ligand binding free energies

Khalak, Tresadern, Aldeghi, Baumann, Mobley, de Groot, Gapsys, Chem. Sci., 2021

Method

- Do not simulate protein-ligand complex with the decoupled ligand
- Place a proper ligand ensemble into the apo protein ensemble

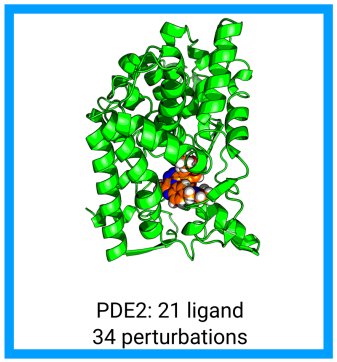


Protein-Ligand complexes

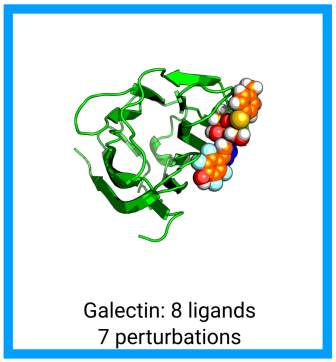
- 7 systems
- 128 ligands

Data sets from:

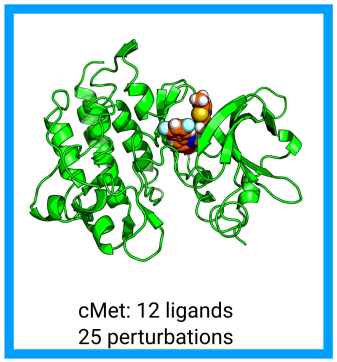
Gapsys, Perez-Benito, et al, Chemical Science, 2020



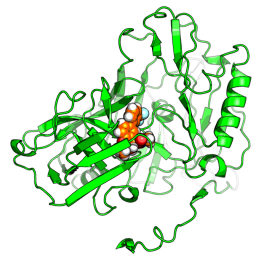
PDE2: 21 ligand
34 perturbations



Galectin: 8 ligands
7 perturbations

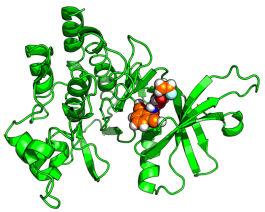


cMet: 12 ligands
25 perturbations

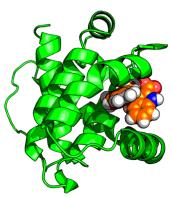


BACE: 80 ligands
144 perturbations
(divided in 3 sets)

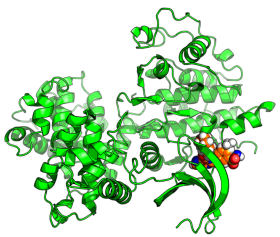
JNK1: 21 ligands
31 perturbations



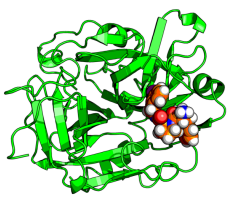
TYK2: 16 ligands
24 perturbations



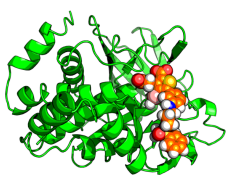
MCL1: 42 ligands
71 perturbations



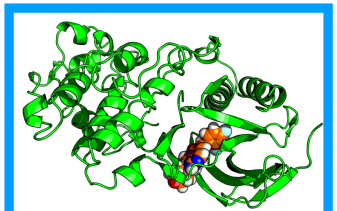
CDK2: 16 ligands
25 perturbations



Thrombin: 11 ligands
16 perturbations

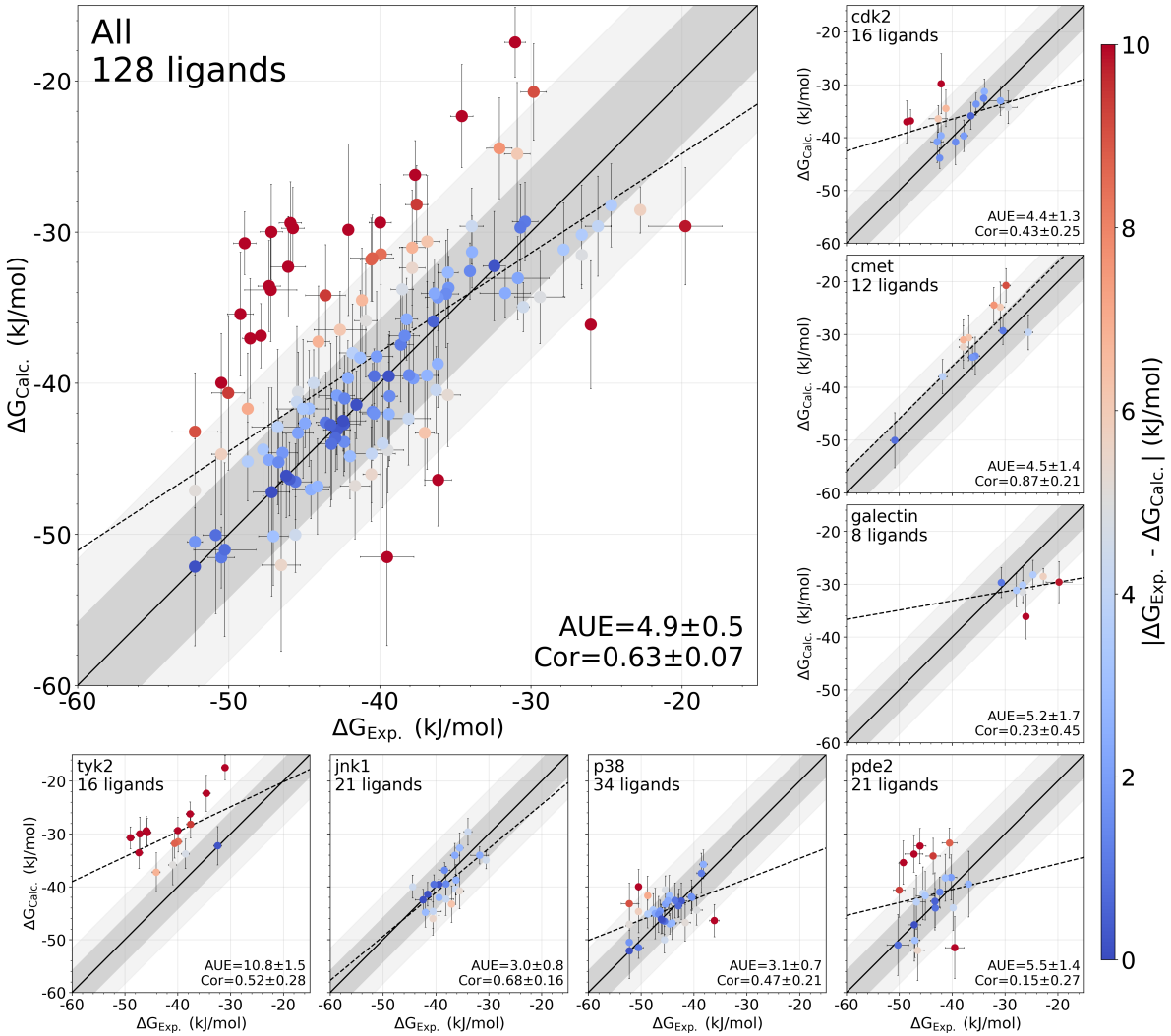


PTP1b: 23 ligands
49 perturbations

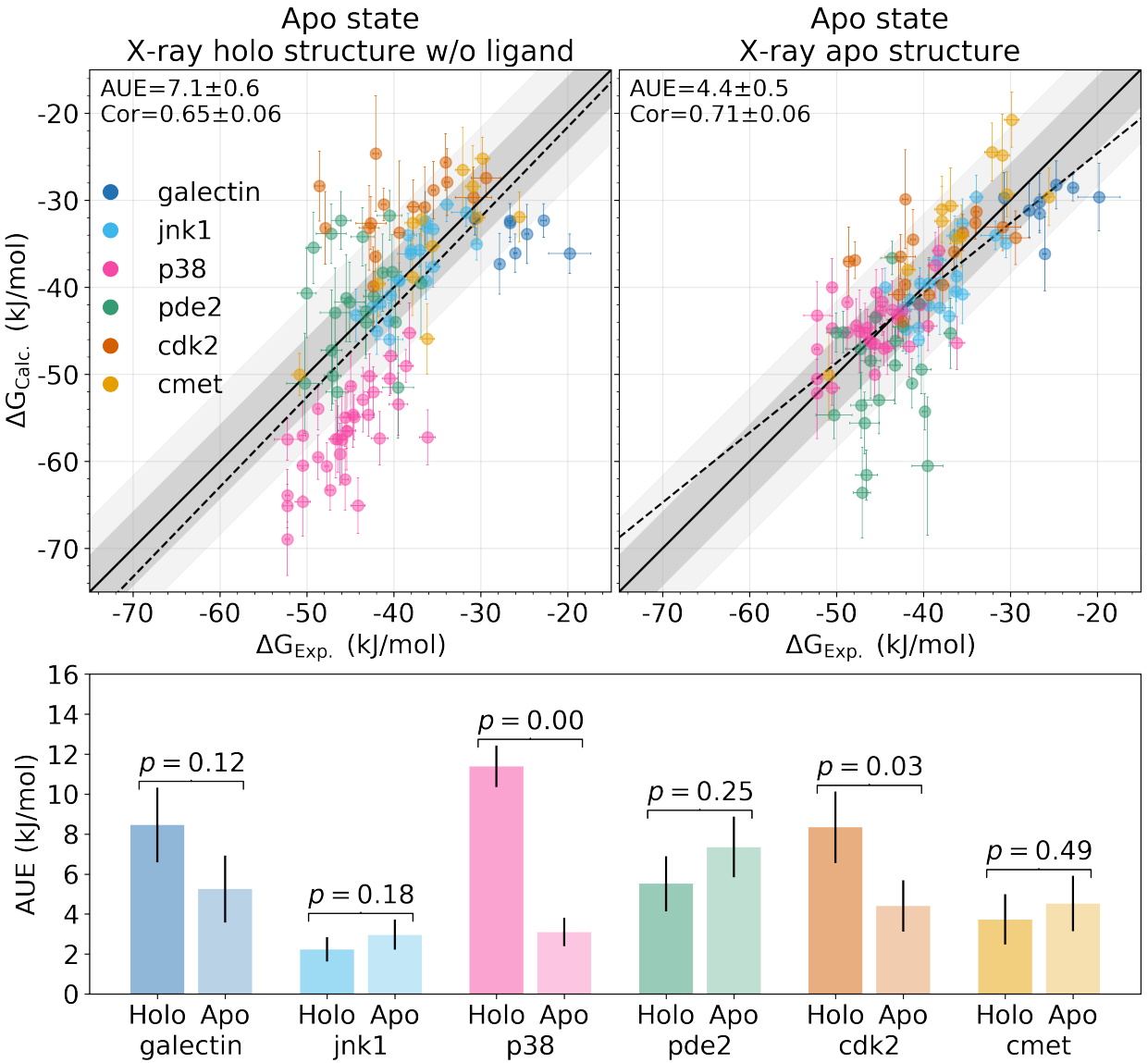


P38: 34 ligands
56 perturbations

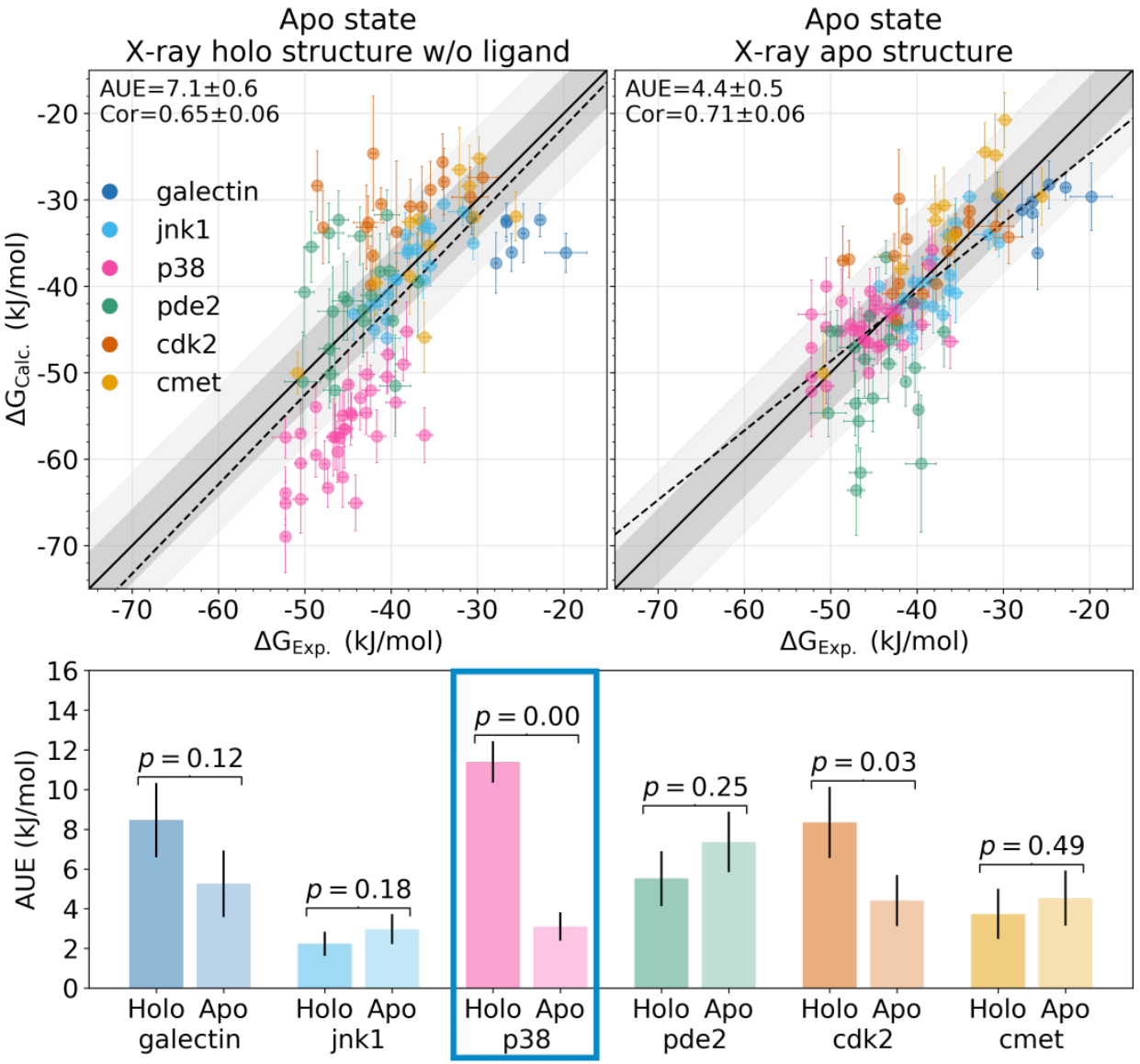
Overall results



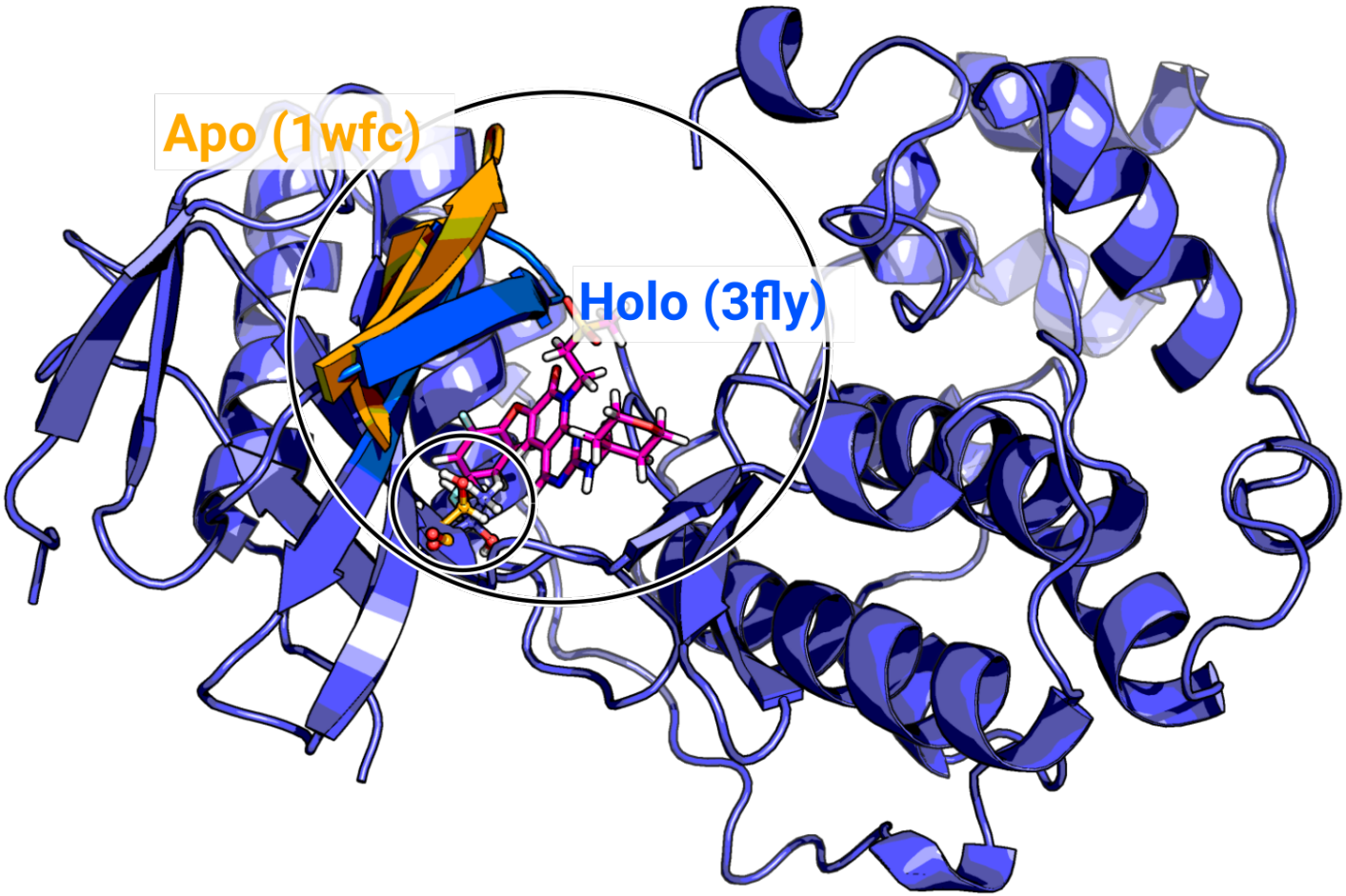
The importance of the apo structure



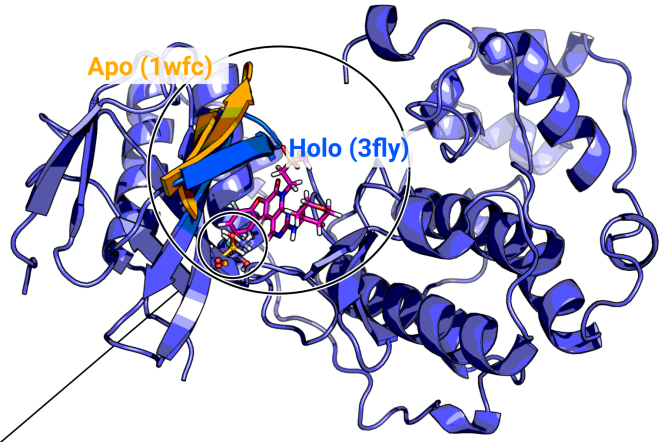
The importance of the apo structure



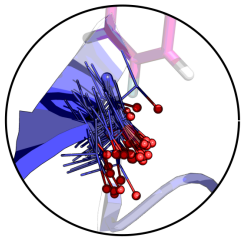
A case study: p38

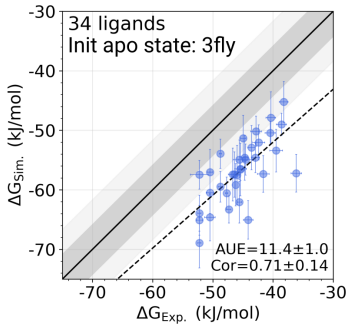


A case study: p38

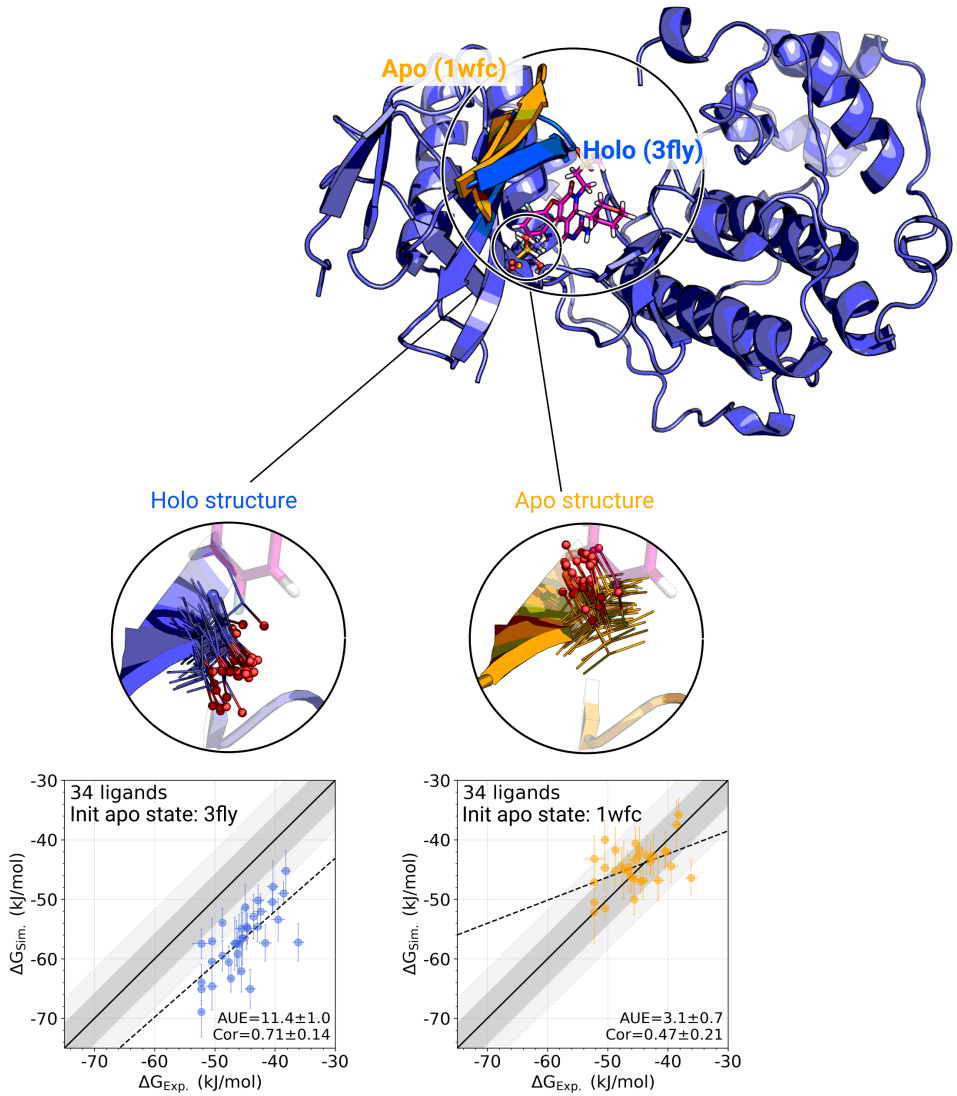


Holo structure

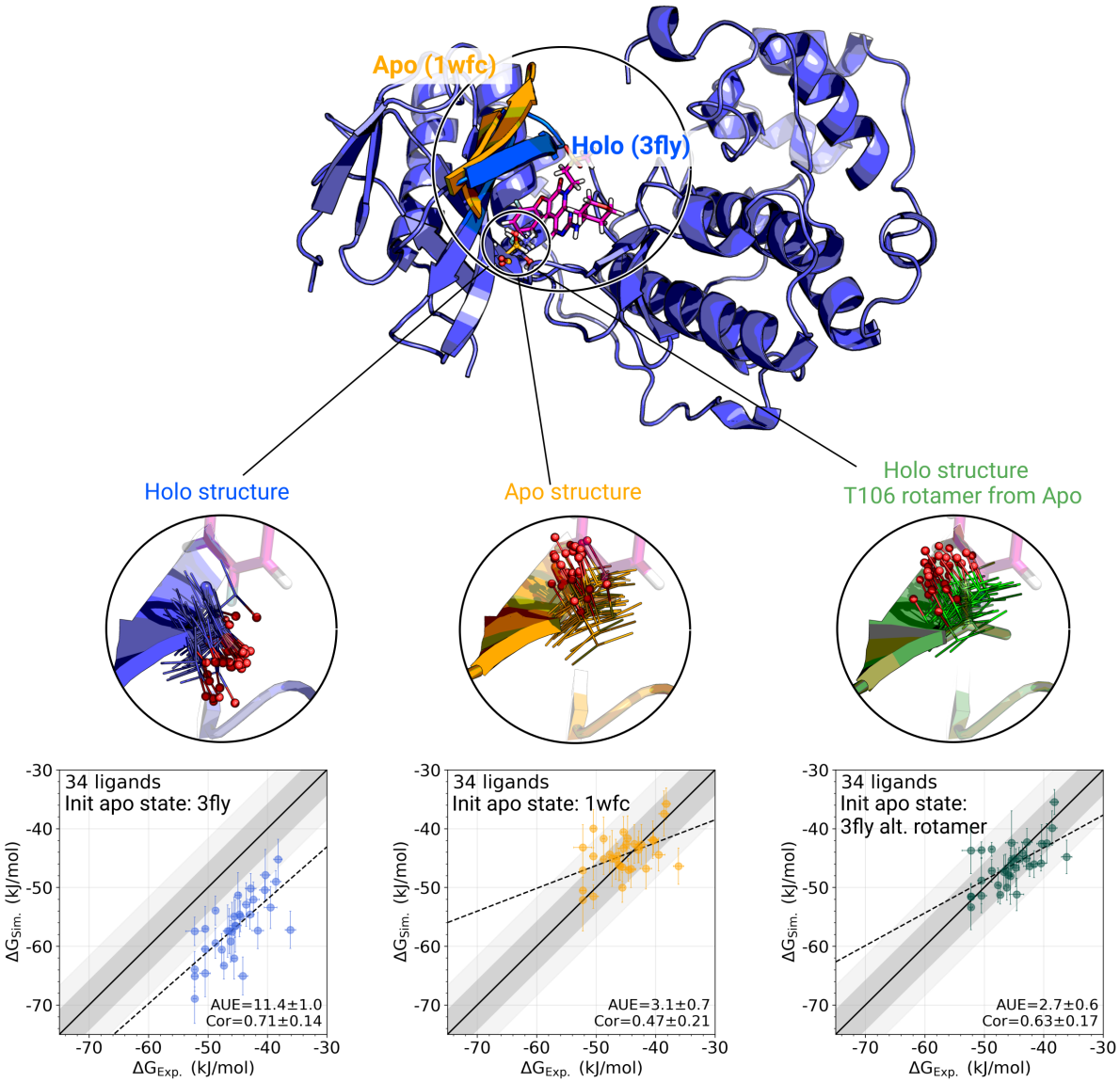




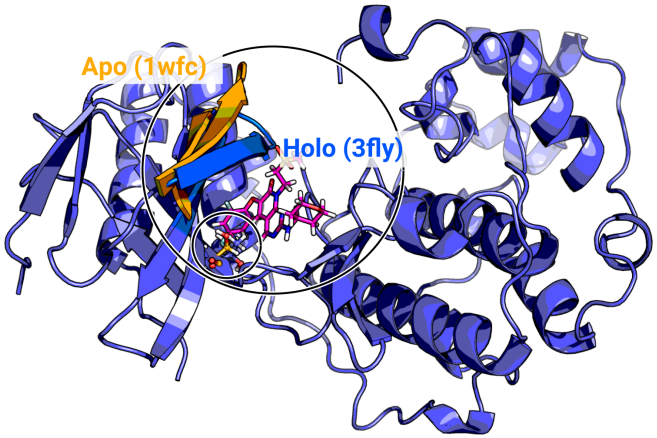
A case study: p38

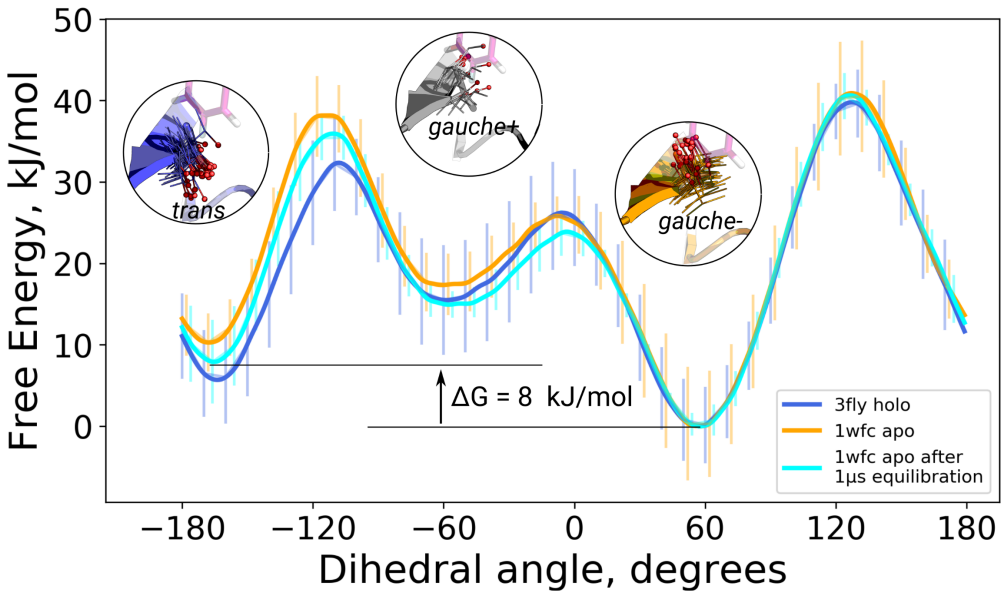


A case study: p38



A case study: p38

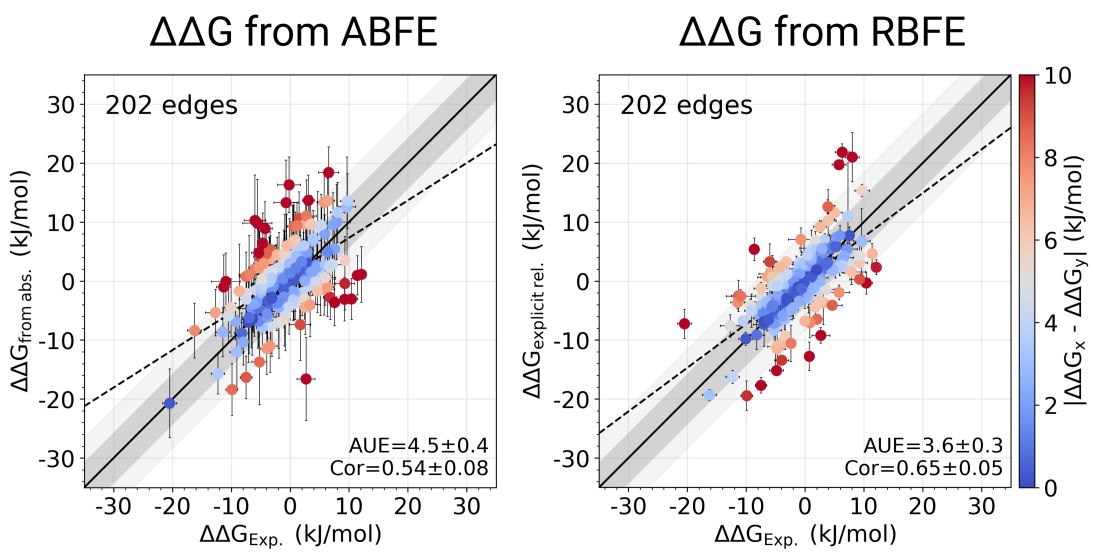




Summary: absolute ΔG Part 2

ABFE

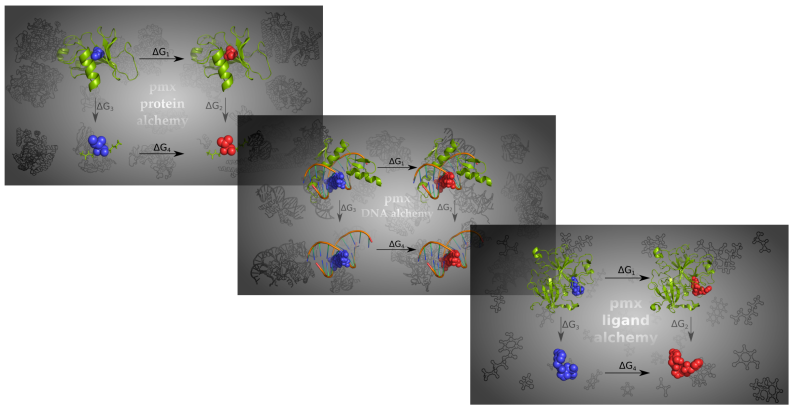
- It is essential to capture ΔG between apo and holo protein conformers
- Absolute ΔG accuracy comparable to that of RBFE
- Comes at a higher computational cost



Summary

GROMACS

FAST. FLEXIBLE. FREE.



pmx

Dr. Daniel Seeliger

Dr. Matteo Aldeghi

Dr. Servaas Michielssens

Dr. Yuriy Khalak

Professor Dr. Bert de Groot

Protein-Ligand binding

Dr. Laura Perez-Benito

Dr. Gary Tresadern

Dr. David Hahn

Dr. Ahmet Yildirim

Hannah Baumann

Professor Dr. David Mobley

Professor Dr. David van der Spoel

Professor Dr. Herman van Vlijmen

AWS and HPC

Dr. Carsten Kutzner

Dr. Markus Rampp

Christian Kniep

Austin Cherian

Ludvig Nordstrom

Funding

bioexcel

MAX-PLANCK-GESELLSCHAFT

Janssen

Boehringer Ingelheim