BioExcel Centre of Excellence for Computational Biomolecular Research

pmx: free energy calculations

Vytautas Gapsys
vgapsys@mpibpc.mpg.de

BioExcel: Partners

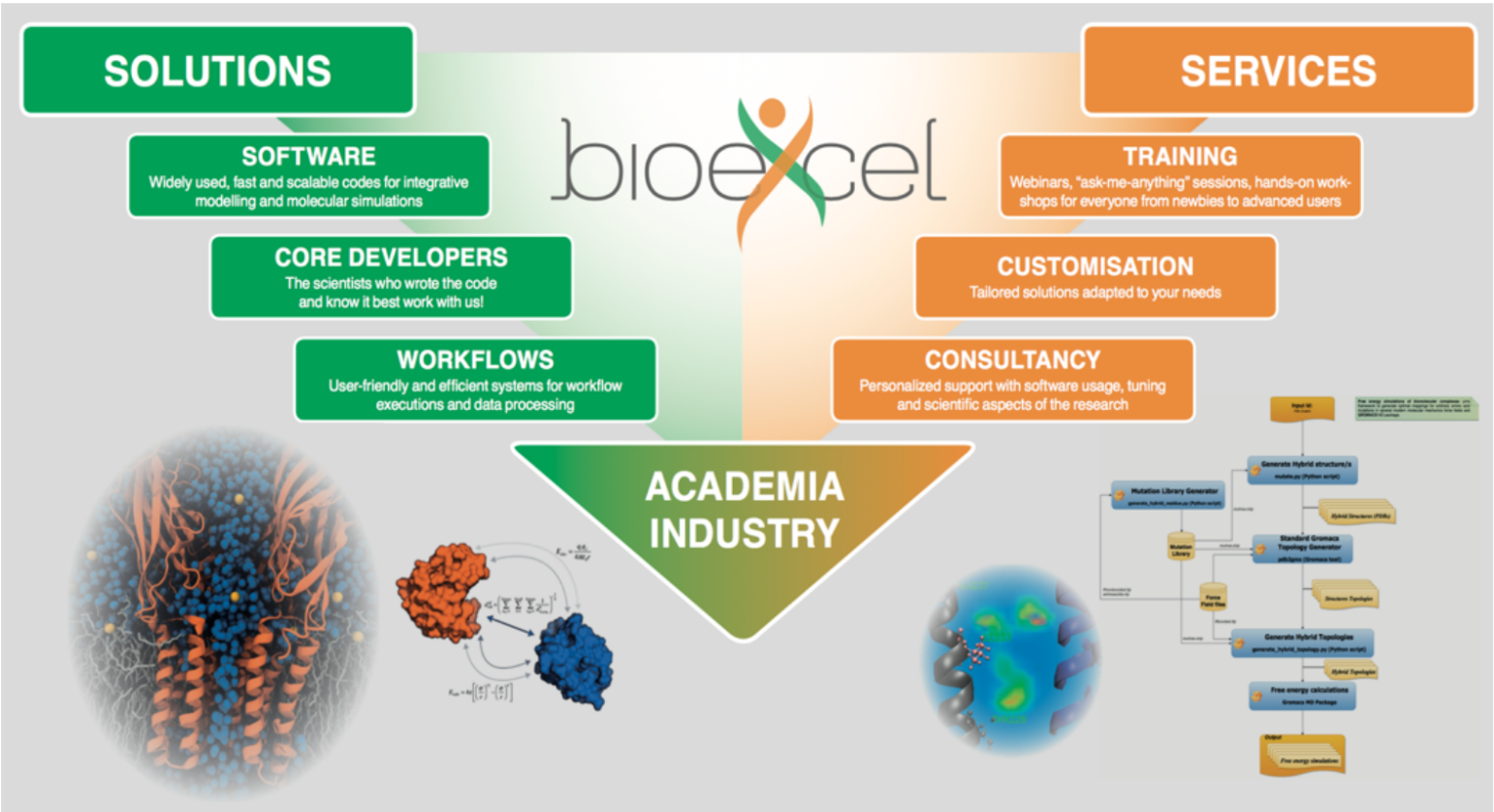
BioExcel Partners 2019

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
NORMAN CONSULTING
EMBL-EBI
Utrecht University
KTH VETENSKAP OCH KONST
THE UNIVERSITY OF EDINBURGH
JÜLICH FORSCHUNGSZENTRUM
NBD NOSTRUM BIODISCOVERY
IRB BARCELONA
INSTITUTO DE INVESTIGACIÓN BIOMÉDICA
Ian Harrow Consulting
acrosslimits
MAX-PLANCK-GESELLSCHAFT
MANCHESTER 1824
UNIVERSITY OF JYVÄSKYLÄ

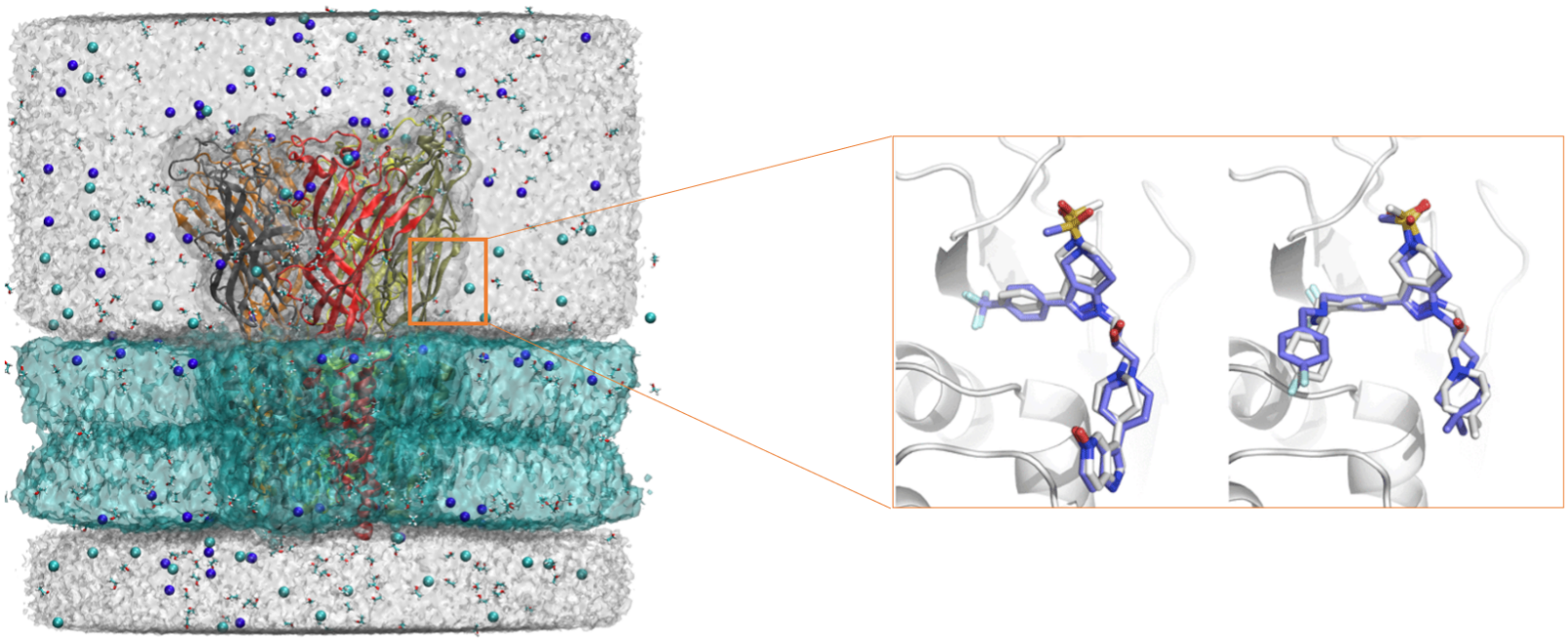
Horizon 2020
European Union Funding
for Research & Innovation

BioExcel is funded by the European Union Horizon 2020 program under grant agreements 675728 and 823830.

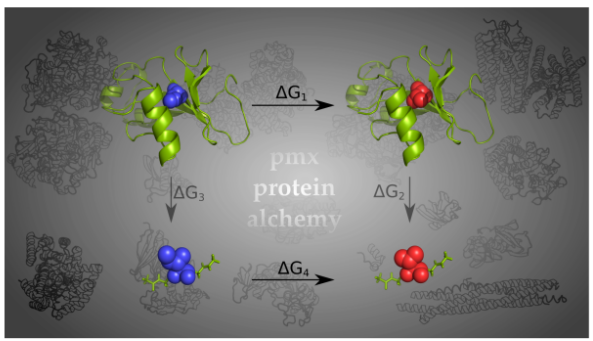
BioExcel: Overview

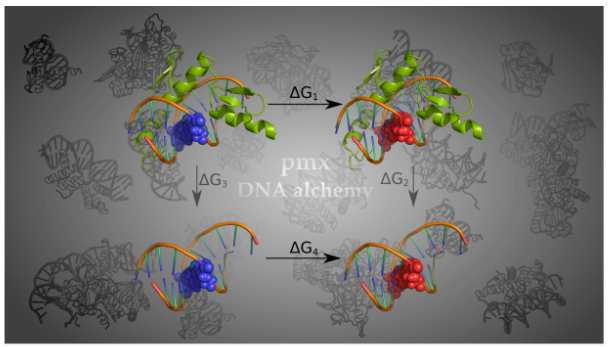


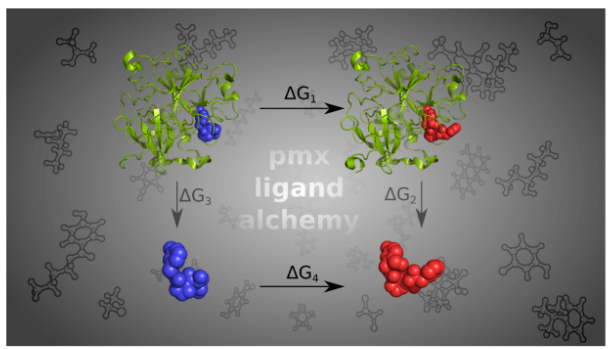
Molecular Simulations (GROMACS)



Free Energy Calculations (PMX)

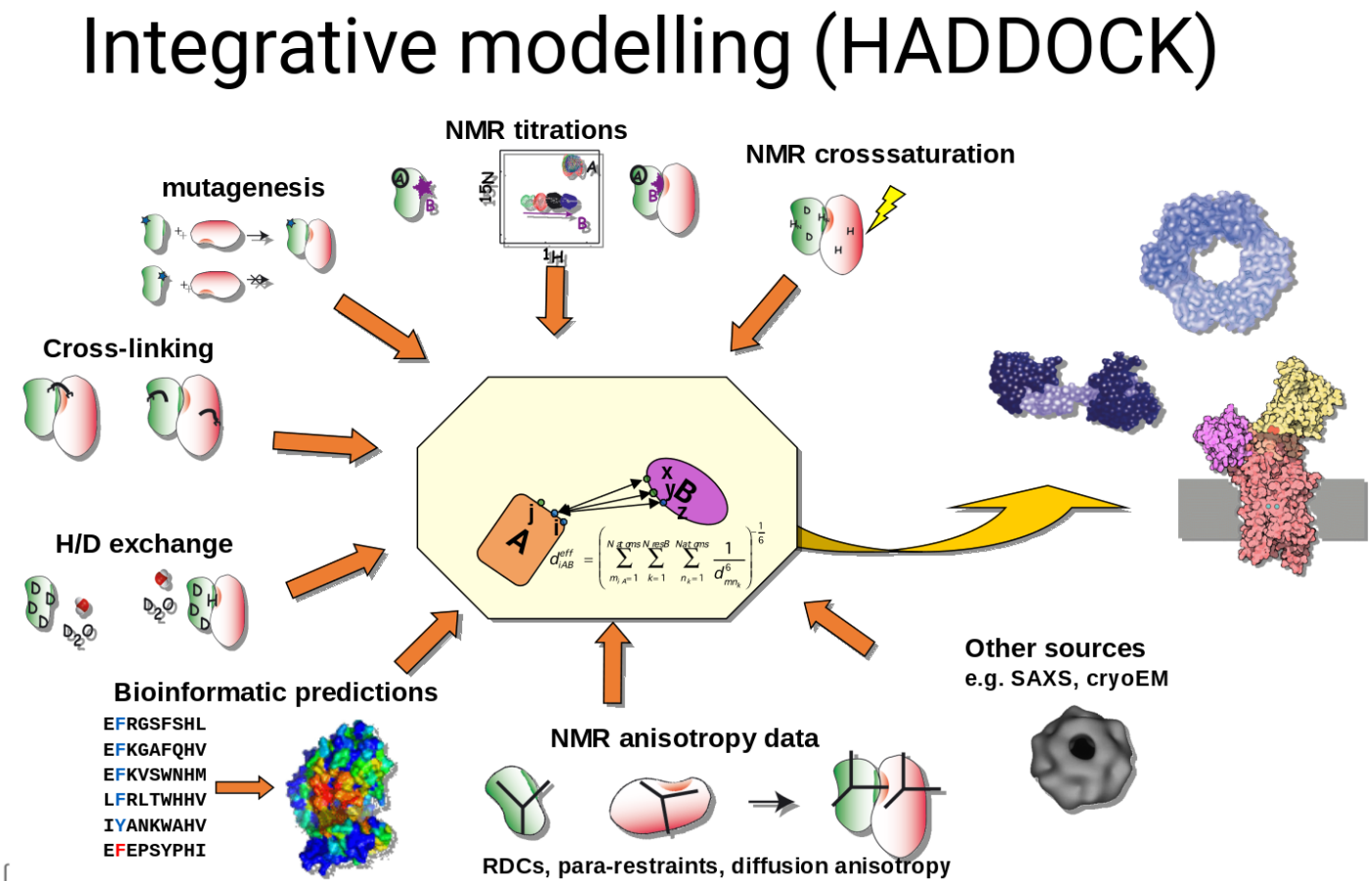




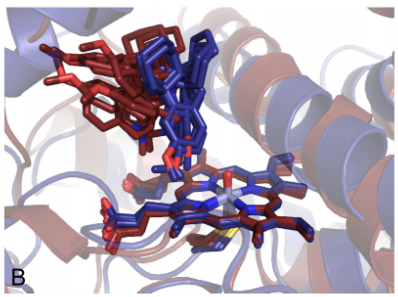


PMX provides an automated framework for the introduction of amino acid mutations in proteins and thus removes some of the most laborious and time consuming steps in traditional methods for free energy calculations

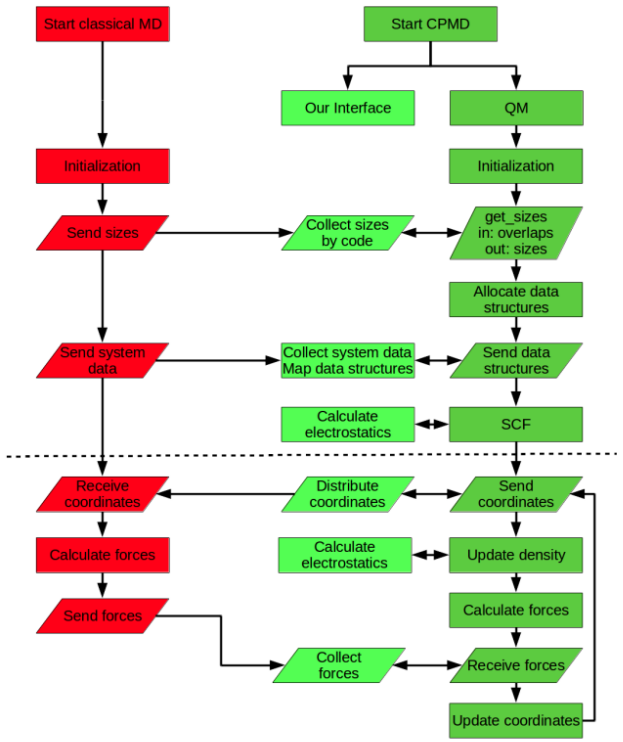
BioExcel: Haddock



Hybrid QM/MM calculations (CP2K)



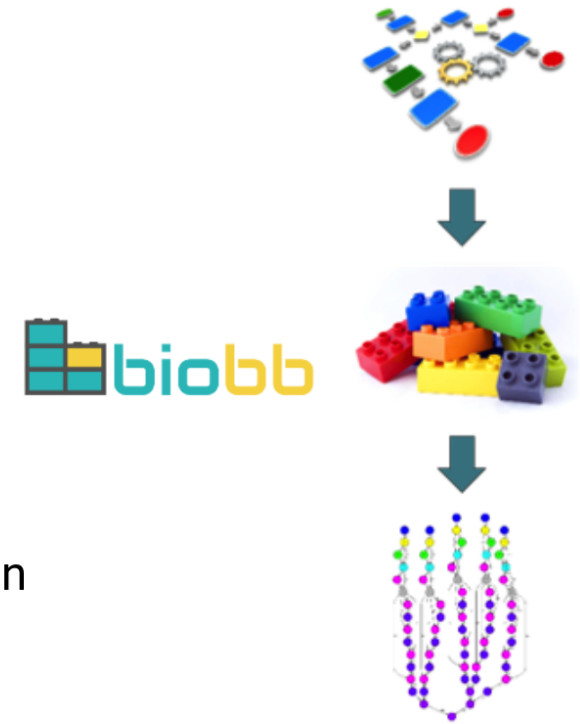
Workflow of our **MiMiC interface** in combination with **CP2K** and classical **MD** code.

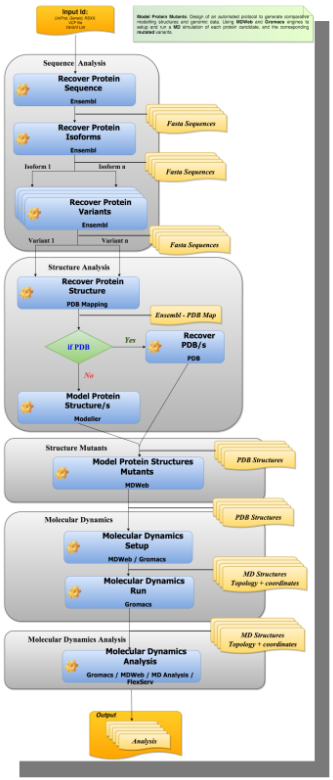


BioExcel: Workflows

Usability and Automation

- Make ICT technologies easier to use by biomolecular researchers, both in academia and industry
- Devise efficient workflow environments with associated data integration





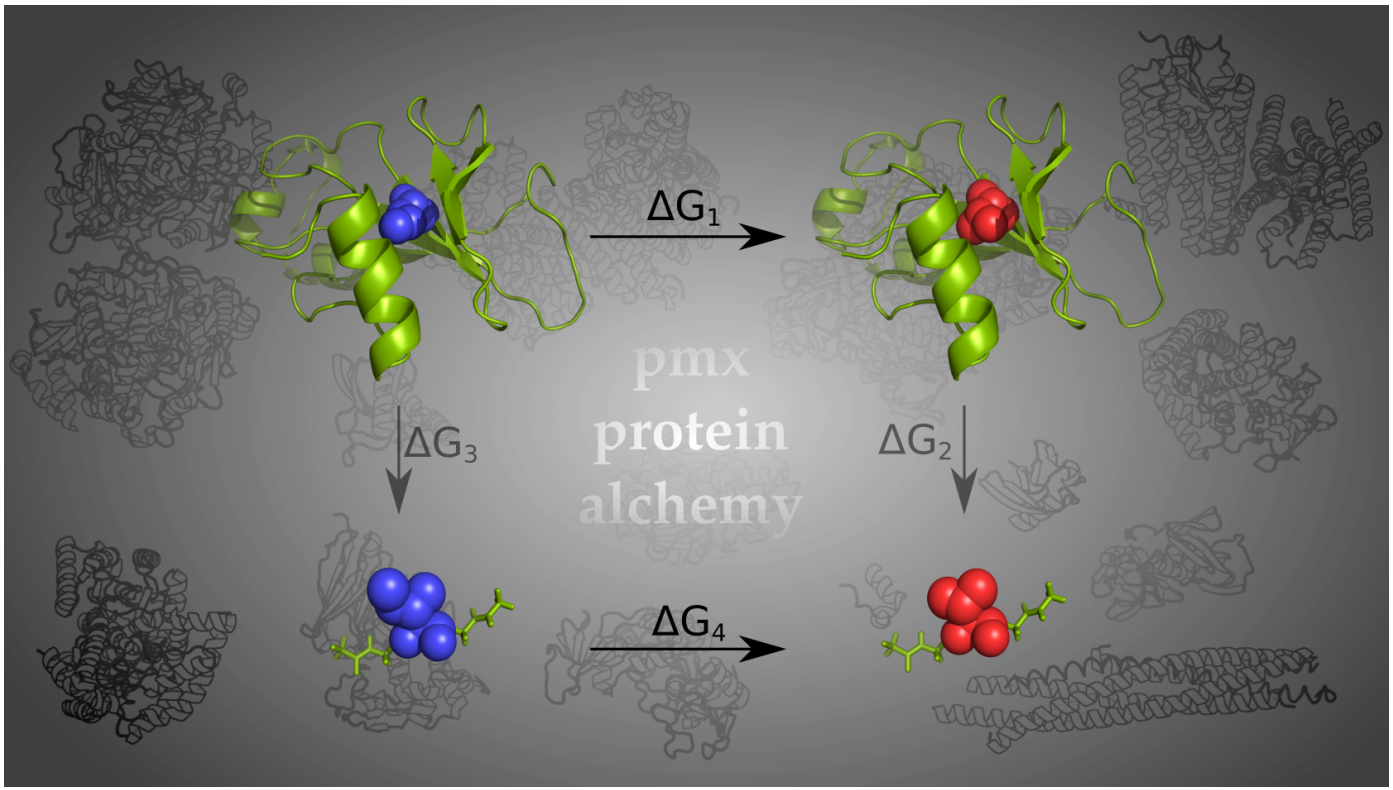
BioExcel: Training, Events, Consultancy, Webinars

Webinar Series *http://bioexcel.eu/webinars*

pmx: Free Energy Calculations

pmx Alchemistry for Proteins, DNA and Ligands

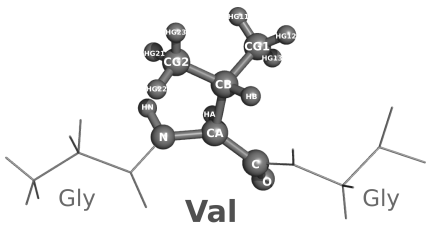
pmx Amino Acids



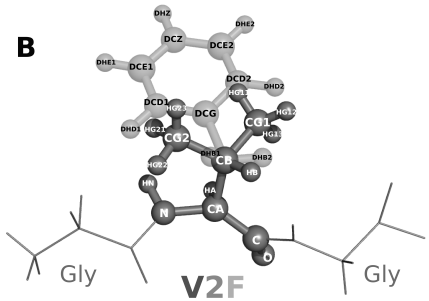
Amino Acid Mutations

Hybrid Structure/Topology

A

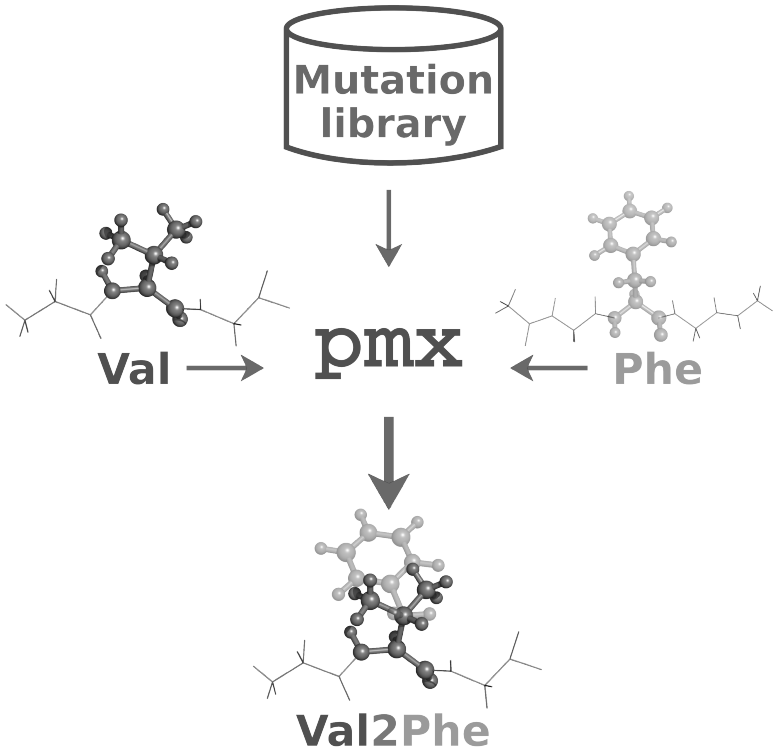


B



C

```
 [ atoms ]
;   nr       type   resnr residue  atom   cgnr    charge       mass       typeB    chargeB      massB
.............
    14         CT1      2    V2F     CB     14  -0.090000    12.0110         CT2  -0.180000    12.0110
    15          HA      2    V2F     HB     15   0.090000     1.0080      DUM_HA   0.000000     1.0000
.............
    26      DUM_HA      2    V2F   DHB1     26   0.000000     1.0000          HA   0.090000     1.0080
    27      DUM_HA      2    V2F   DHB2     27   0.000000     1.0000          HA   0.090000     1.0080
.............
 [ bonds ] ;harmonic potential
; atom_i  atom_j funct           b0        kb            b0B       kbB
.............
    14     15      1       0.111100  258571.2       0.111100  258571.2
    14     16      1       0.153800  186188.0       0.153800  186188.0
.............
 [ angles ] ;Urey-Bradley potential
; atom_i  atom_j atom_k funct  theta0   ktheta    r13       kub  theta0B  kthetaB   r13B      kubB
.............
    12     14     28      5     107.5 433.4624 0.0000     0.000    107.5 433.4624 0.0000     0.000   ; CA CB DCG
    15     14     16      5     110.1 288.6960 0.2179 18853.104    110.1 288.6960 0.2179 18853.104   ; HB CB CG1
    15     14     20      5     110.1 288.6960 0.2179 18853.104    110.1 288.6960 0.2179 18853.104   ; HB CB CG2
.............
 [ dihedrals ] ;periodic potential function
; atom_i  atom_j atom_k atom_l funct  phi  kphi mult    phiB kphiB multB
.............
    14     12     24     25     9     0  0.0000  1       0  5.8576  1        ; CB CA C O CT1 CT1 C O (AAAA->AAAA)
    14     12     24     25     9     0  5.8576  1       0  0.0000  1        ; CB CA C O CT1 CT1 C O (AAAA->AAAA)
.............
```

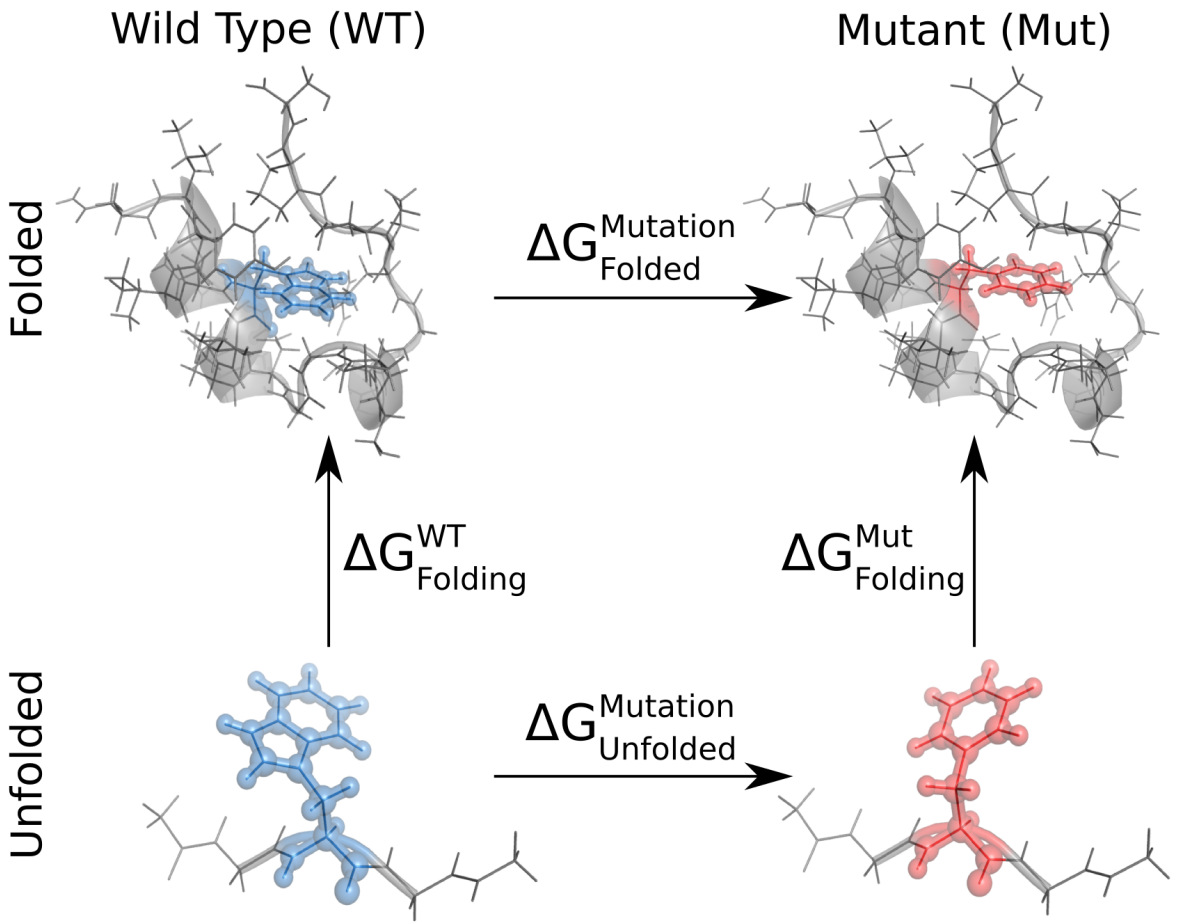


Gapsys, Michielssens, Seeliger, de Groot, JCC, 2015

Application: Protein Thermostability

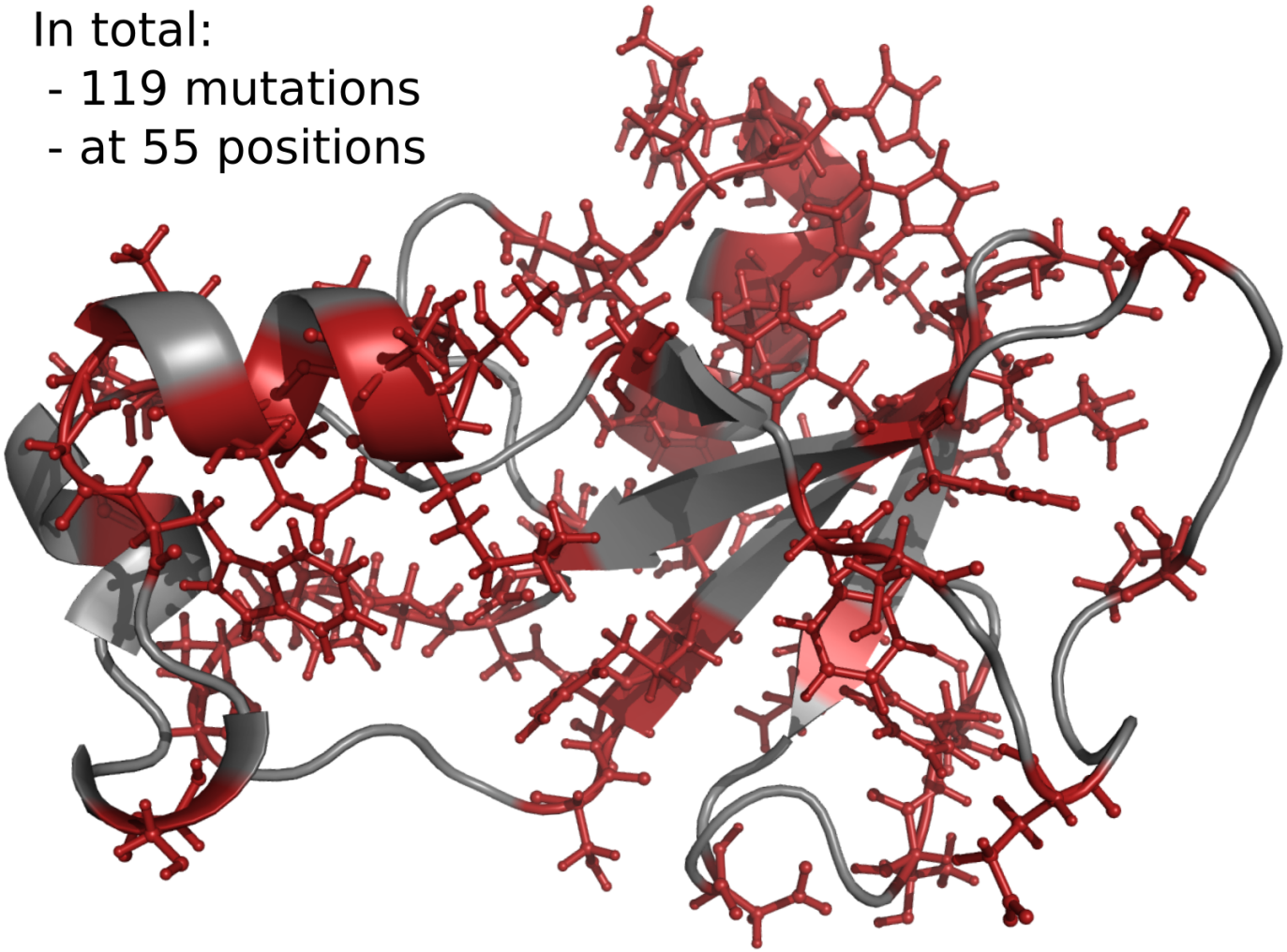
Application:
Protein Thermostability

Application: Protein Thermostability



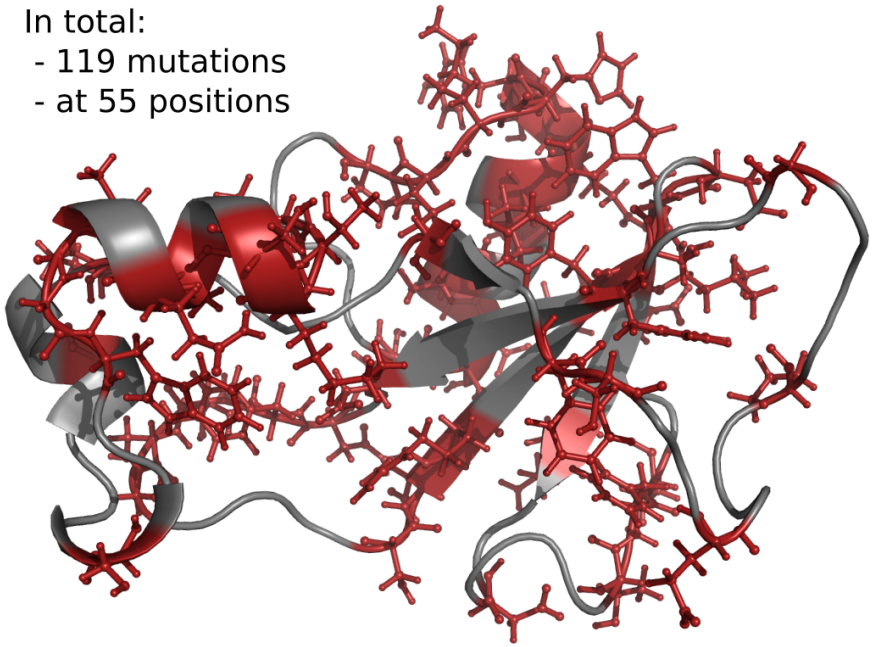
$$\Delta\Delta G^{Mutation}_{Folding} = \Delta G^{Mut}_{Folding} - \Delta G^{WT}_{Folding} = \Delta G^{Mutation}_{Folded} - \Delta G^{Mutation}_{Unfolded}$$

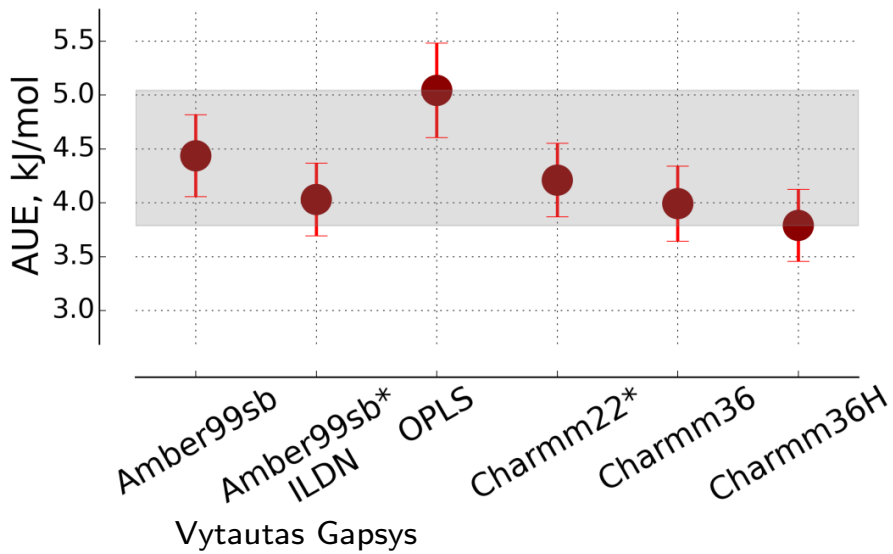
Application: Barnase Thermostability Mutation Scan



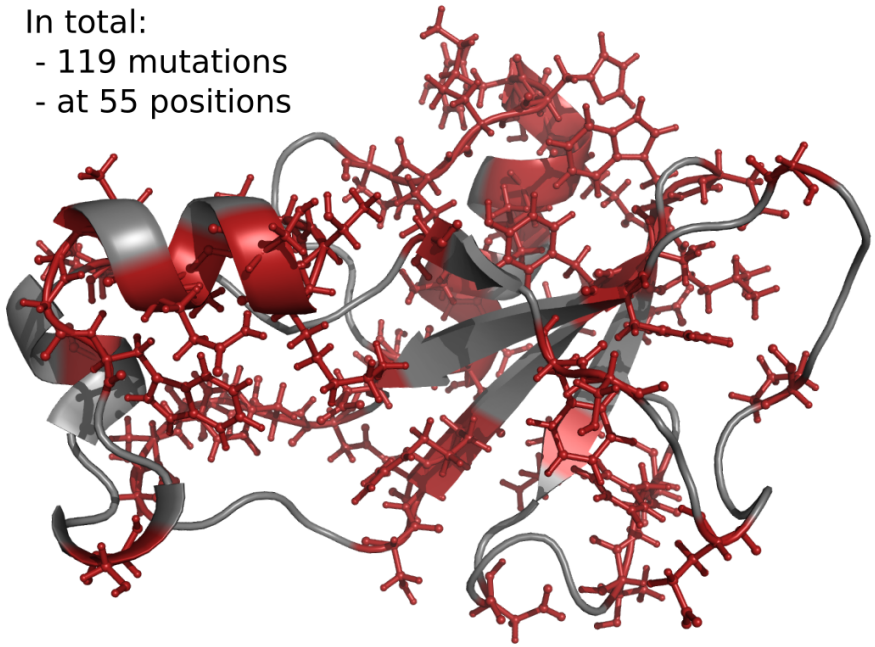
Gapsys, Michielssens, Seeliger, de Groot, Angewandte Chemie Int. Ed., 2016

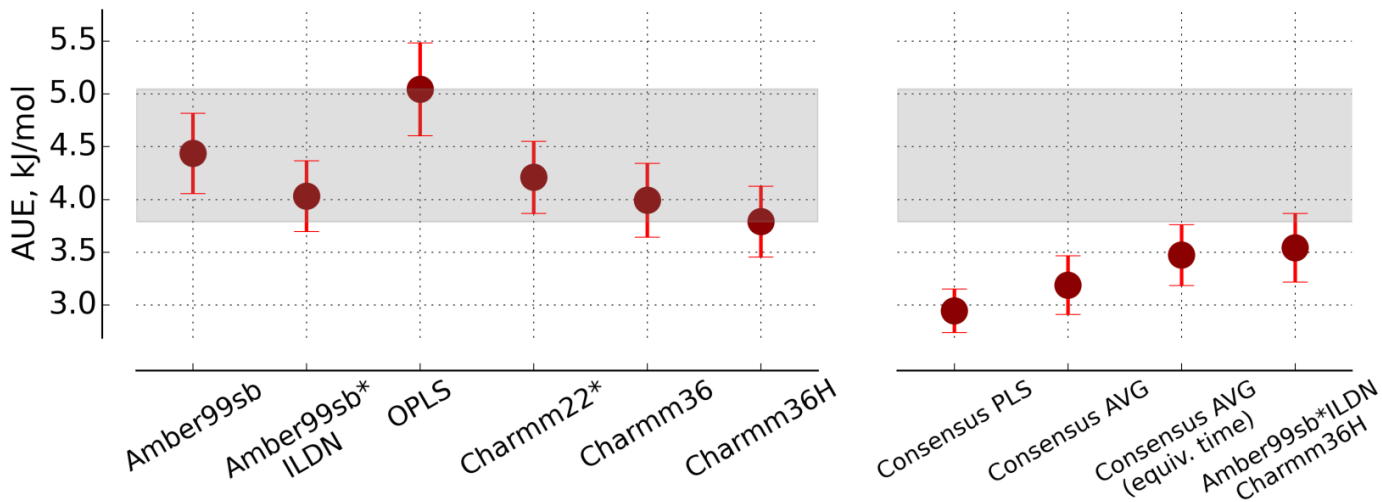
Application: Force Field Comparison





Application: Force Field Combination





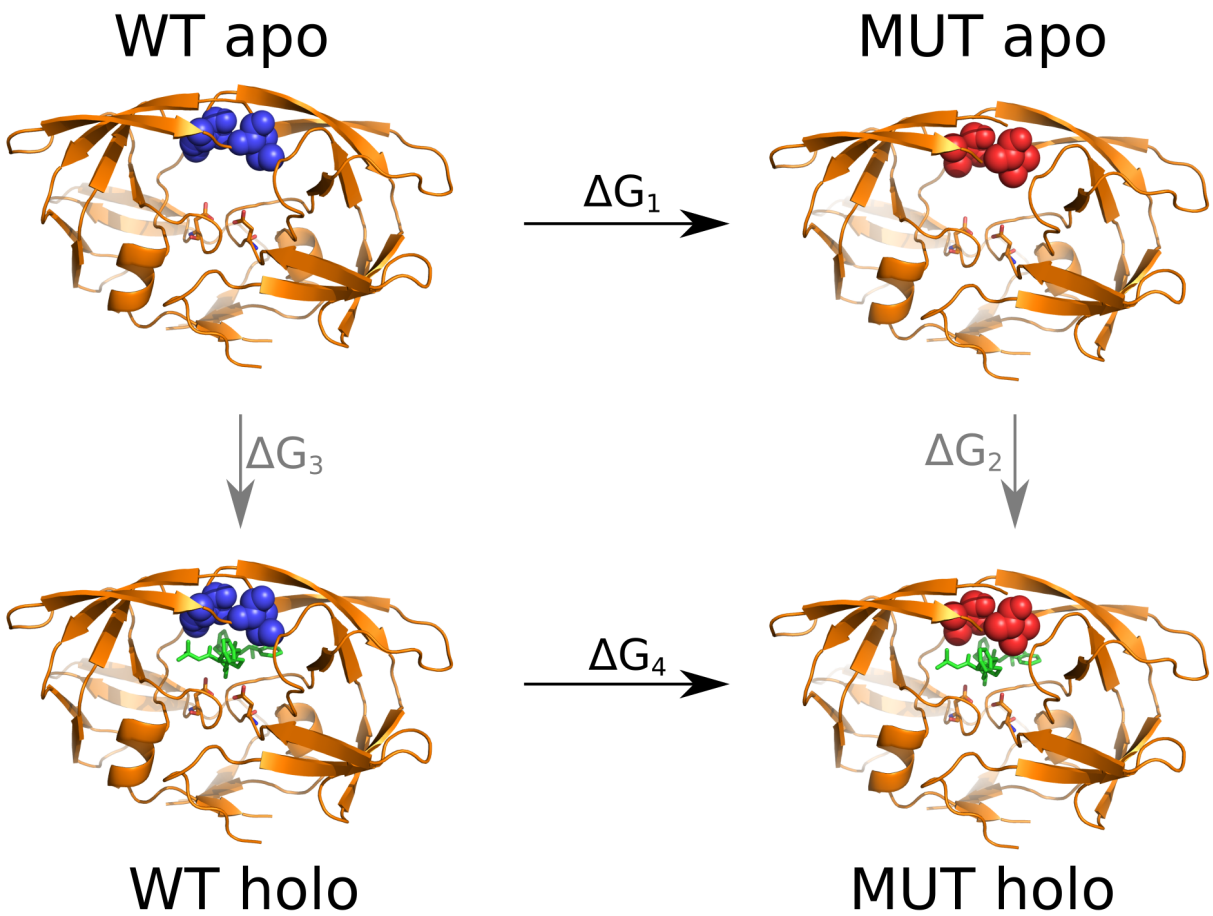
Application: Drug Resistant Mutations

Application:
Large Scale Drug Resistant Mutation Scan

Aldeghi, Gapsys, de Groot, ACS Central Science, 2018

Aldeghi, Gapsys, de Groot, ACS Central Science, 2019

Application: Drug Resistant Mutations

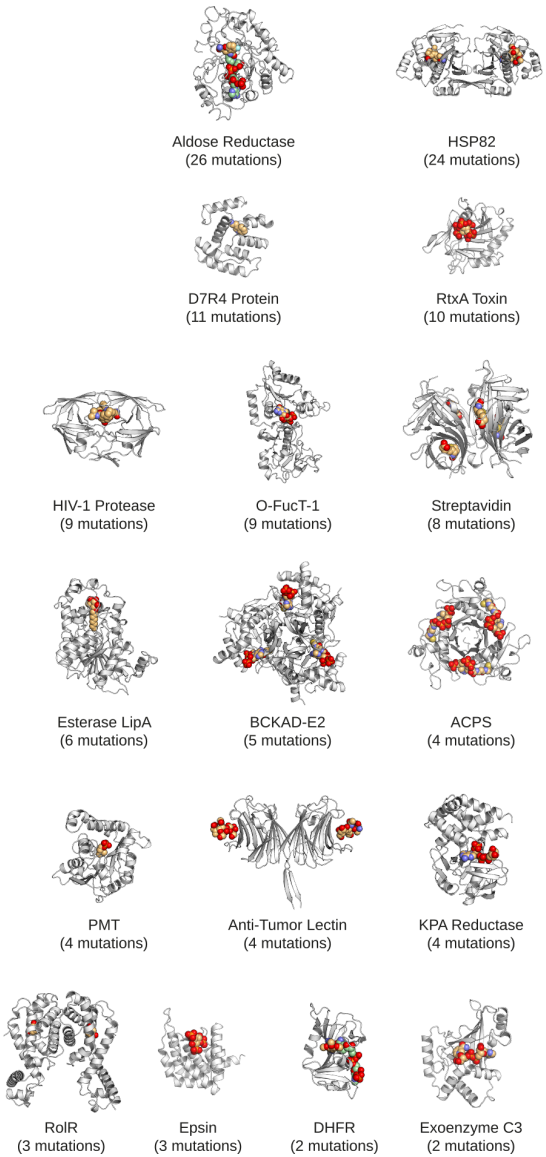


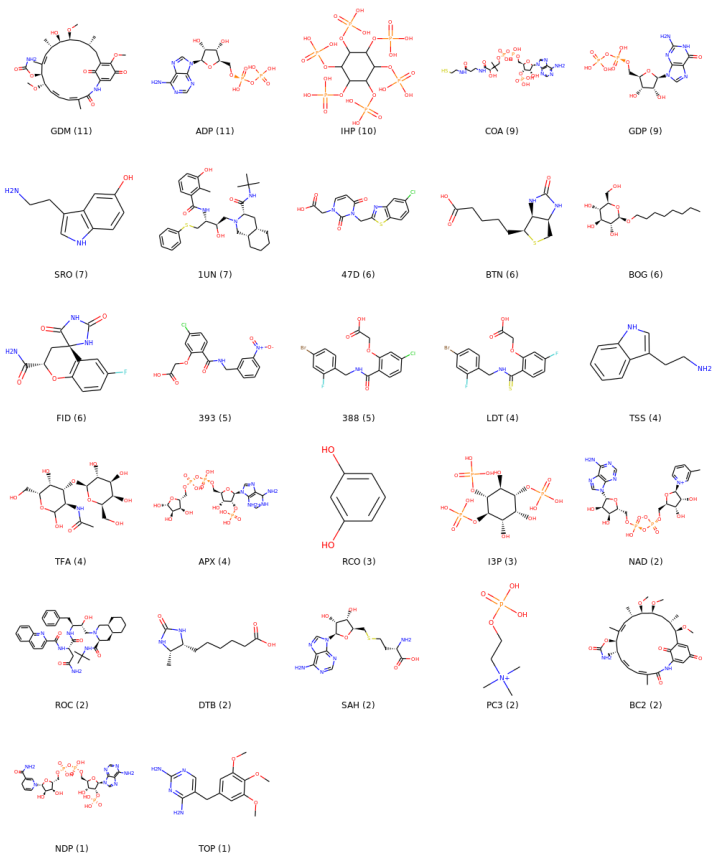
Change in the ligand binding free energy upon an amino acid mutation

$$\Delta\Delta G_{mutation} = \Delta G_2 - \Delta G_3 = \Delta G_4 - \Delta G_1$$

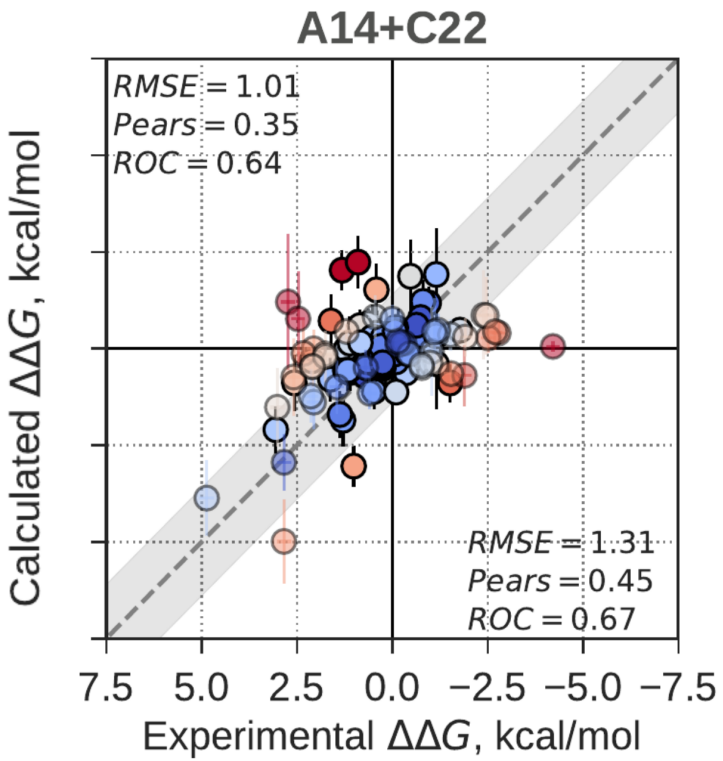
Application: Drug Resistant Mutations

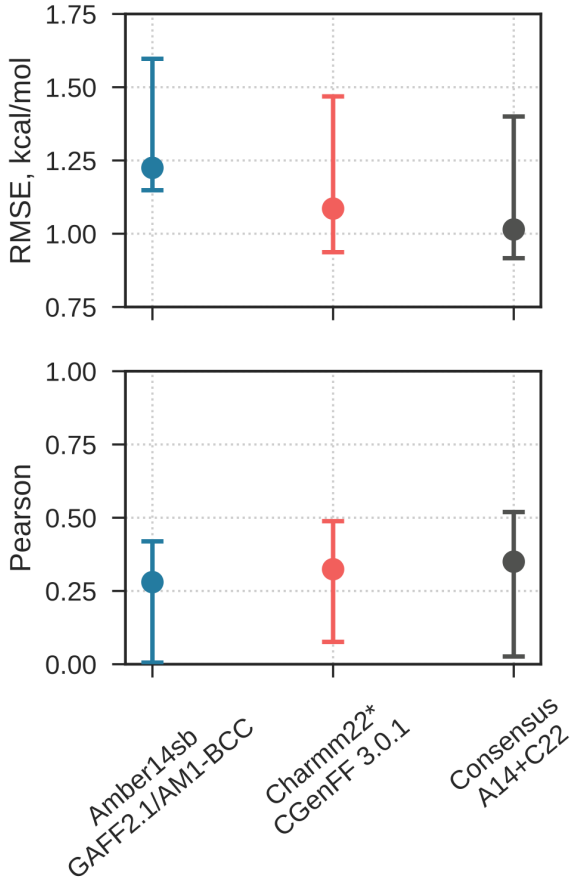
- 17 systems
- 134 mutations
- 27 ligands





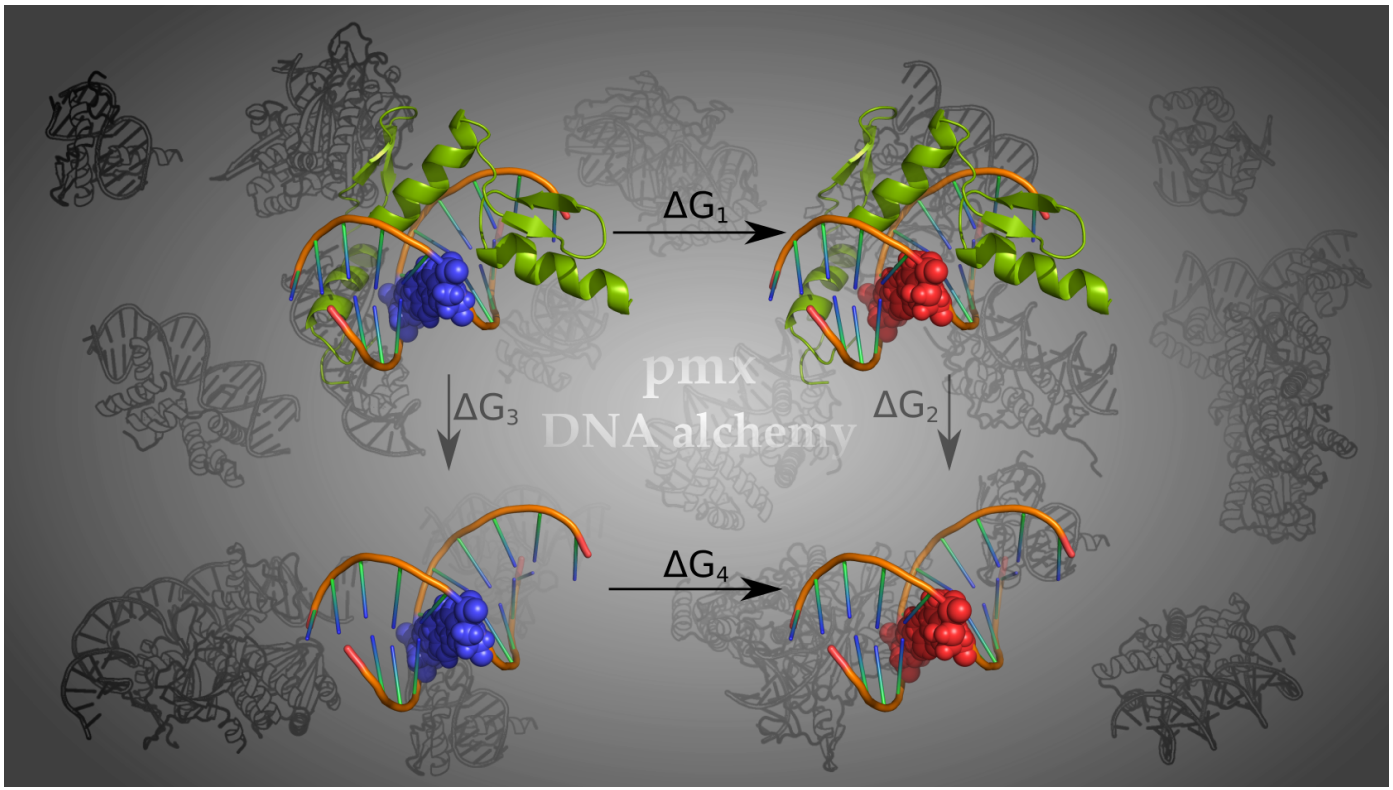
Application: Drug Resistant Mutations





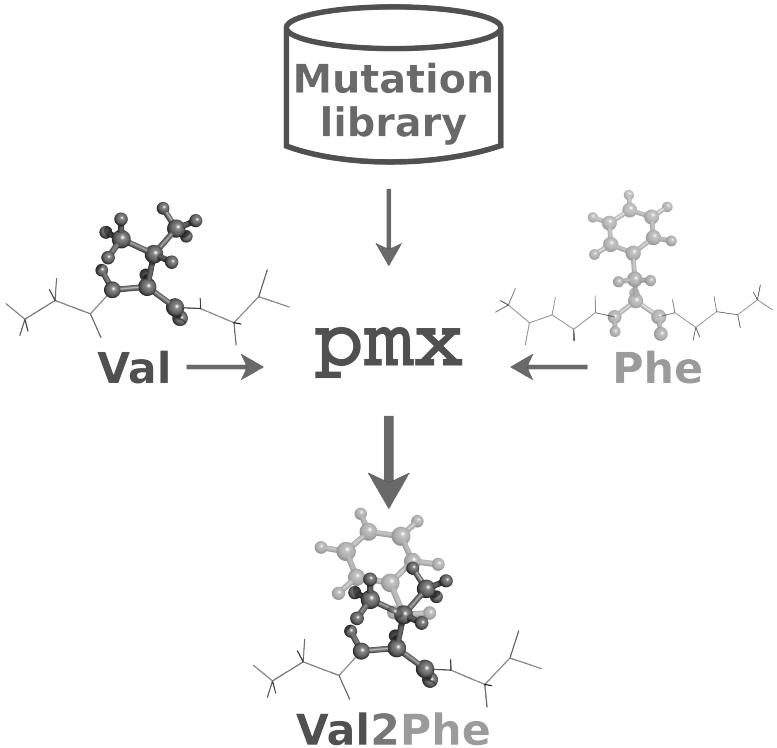
The overall averaged accuracy in terms of RMSE reaches 1-1.2 kcal/mol

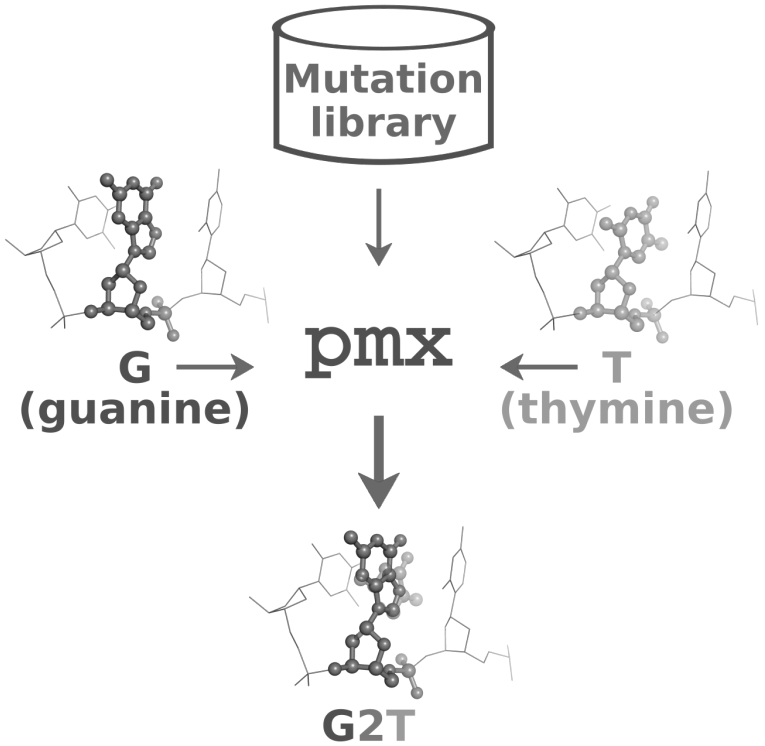
pmx DNA



DNA Mutations

DNA: Nucleic Acid Mutations



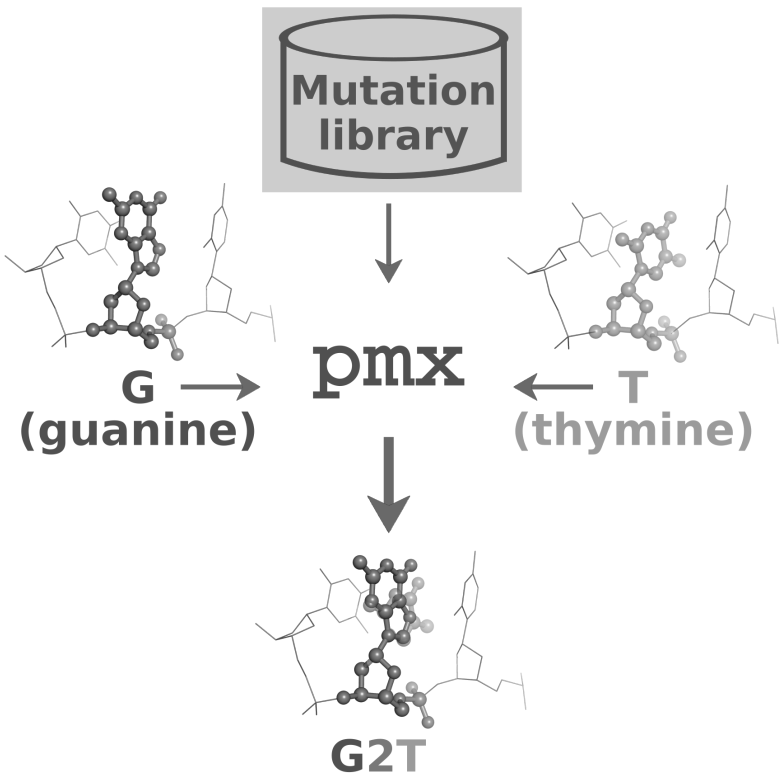


Amino acid mutations

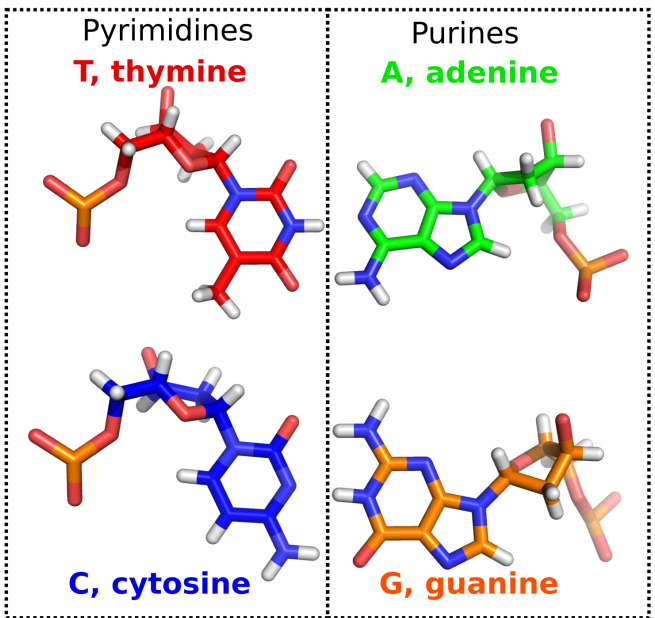
DNA nucleotide mutations

Gapsys, de Groot, JCTC, 2017

DNA: Nucleic Acid Mutations



- A2G, A2C, A2T
- G2A, G2C, G2T
- C2A, C2G, C2T
- T2A, T2C, T2G

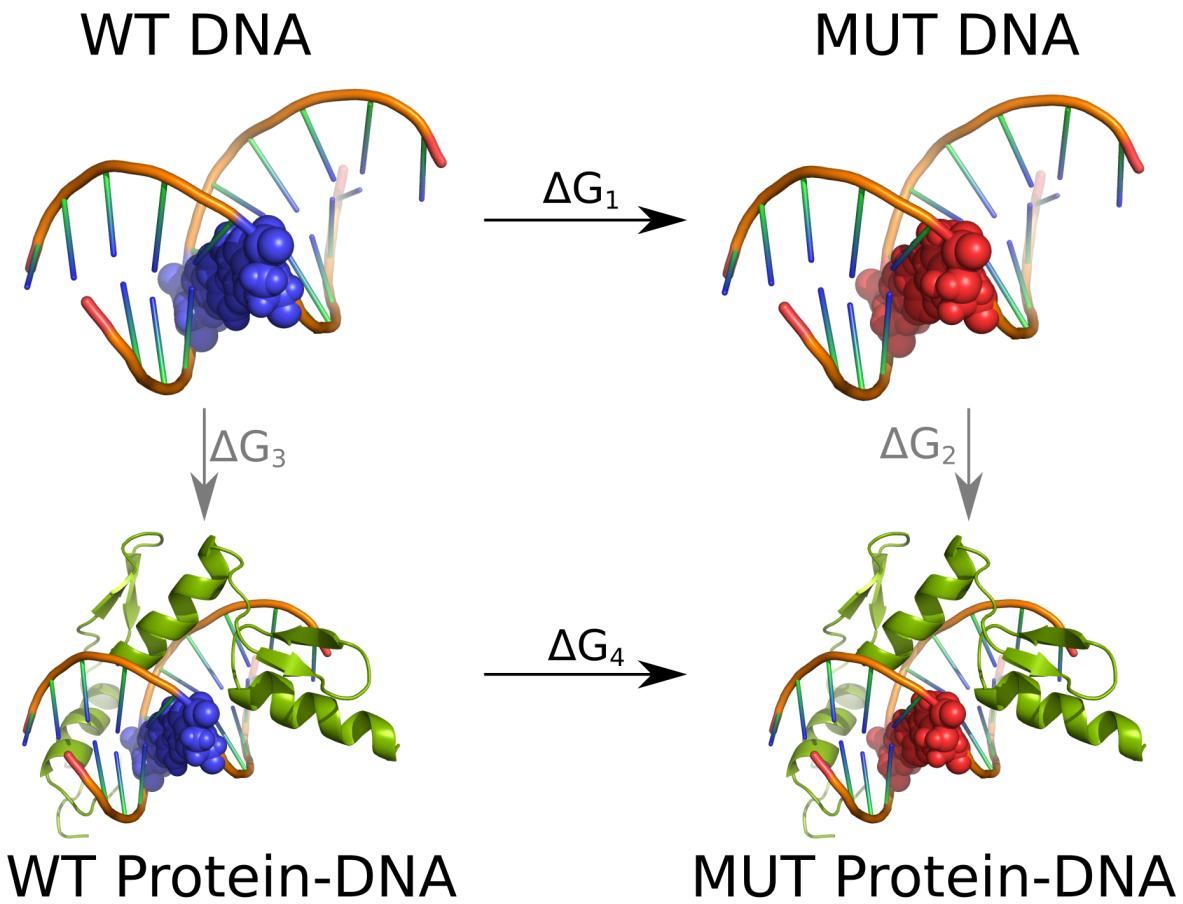


Application: Protein-DNA Binding

Application: Protein-DNA Binding

Gapsys, de Groot, JCTC, 2017

Application: Protein-DNA Binding

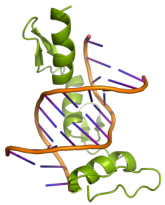


Change in the Protein-DNA binding free energy upon nucleotide mutation

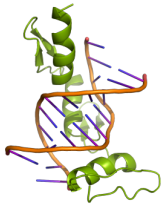
$$\Delta\Delta G_{mutation} = \Delta G_2 - \Delta G_3 = \Delta G_4 - \Delta G_1$$

Application: Protein-DNA Binding

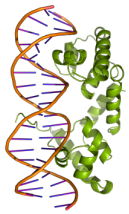
- 16 systems
- 397 mutations



Zif268: zinc finger (1aay)
21 mutations

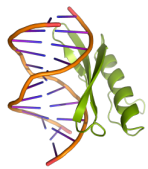


Zif268: zinc finger D20A (1jkl)
6 mutations

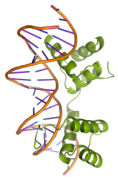


λR: λ repressor (1lmb)
51 mutations

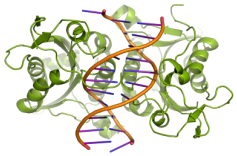
trpR: trp repressor (1tro)
9 mutations



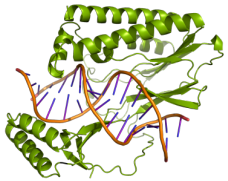
AtERF1: ethylene responsive transcription factor (1gcc)
21 mutations



MAT **a**1-α2: homeodomain (1yrn)
54 mutations

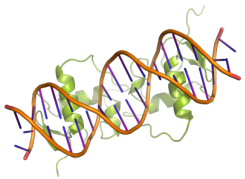


BamHI: endonuclease (1bhm)
23 mutations

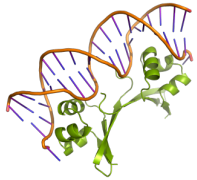


Tus: Tus protein-*Ter* DNA (1ecr)
38 mutations

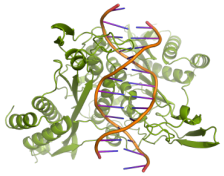
LacR: *Lac* repressor (1efa)
5 mutations



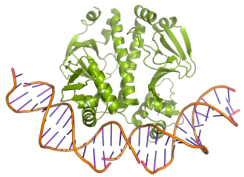
ER: estrogen receptor α (1hcq)
7 mutations



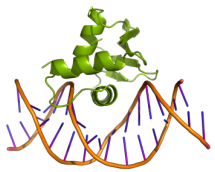
CroR: CRO repressor (6cro)
56 mutations



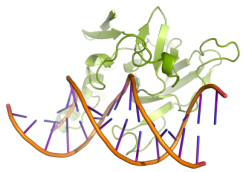
EcoRI: endonuclease (1ckq)
13 mutations



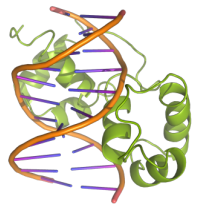
CAP: catabolite gene activator protein (1run)
15 mutations



PU.1 ETS: ETS domain of a transcription factor (1pue)
25 mutations

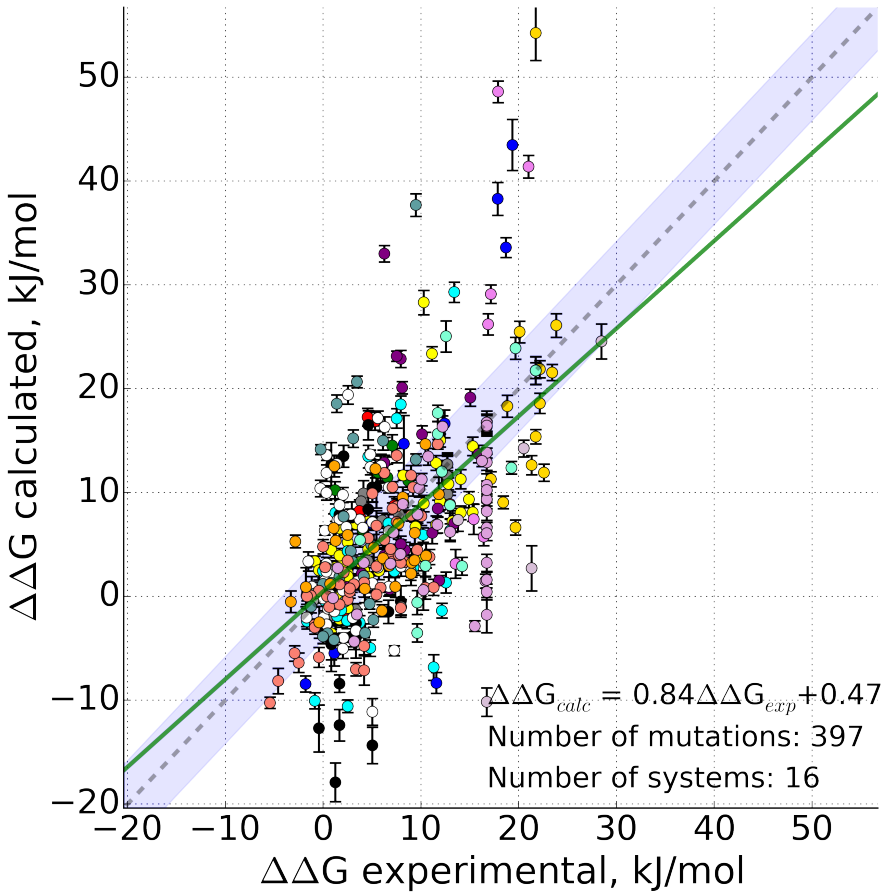


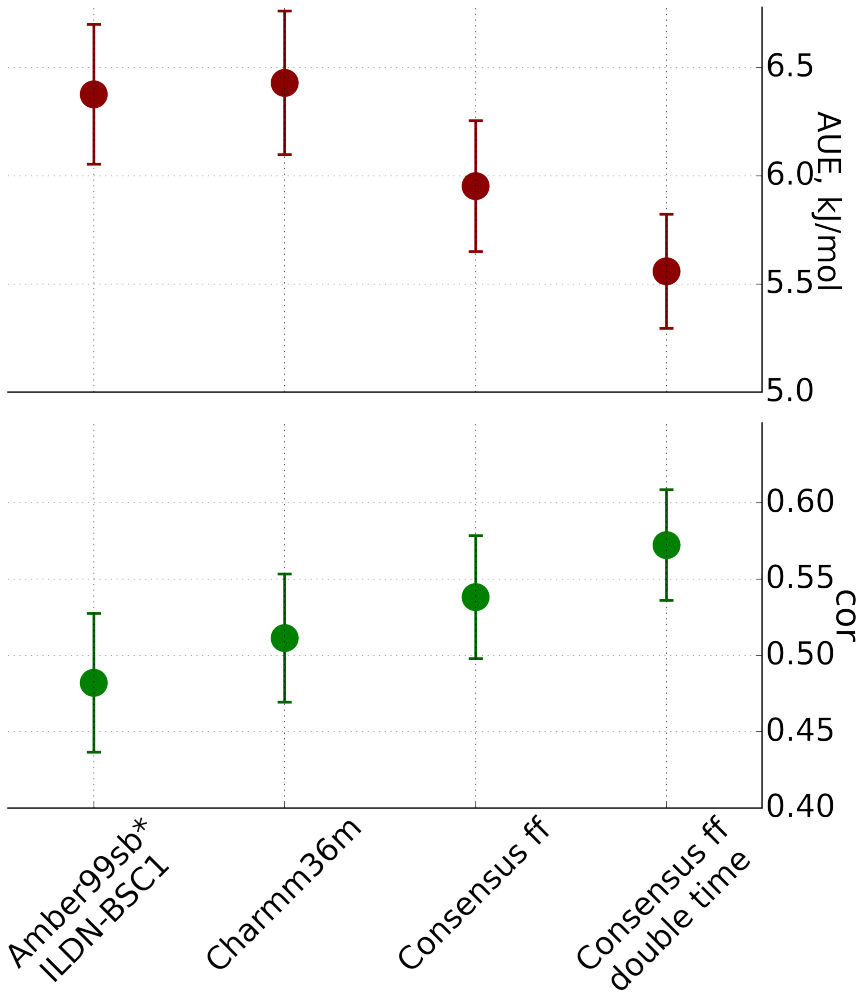
Ndt80: transcriptional activator (1mnn)
26 mutations



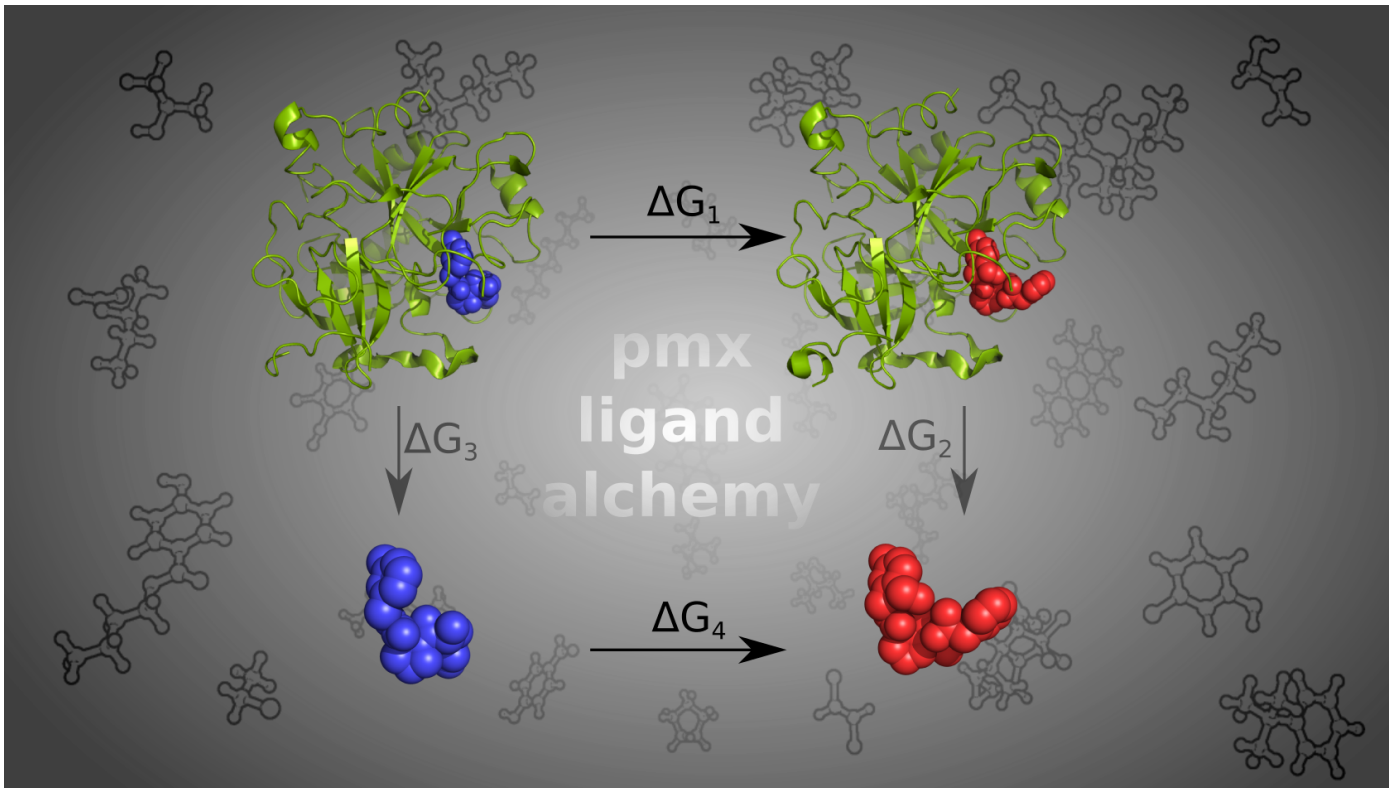
c-Myb: protooncogene (1mse)
27 mutations

Application: Protein-DNA Binding





pmx Ligands



Ligand Modifications

pmx Ligands

atoms_to_morph.py

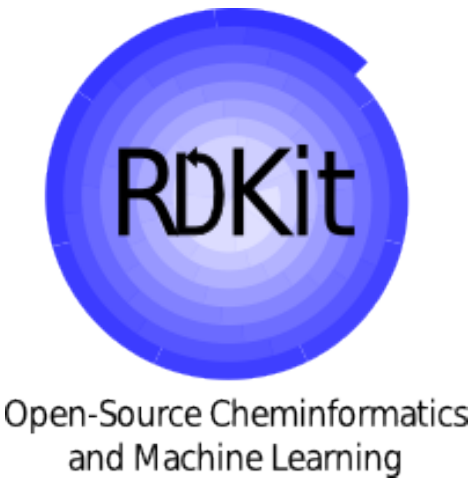
Identifies atoms to be morphed

make_hybrid.py

Builds hybrid topology

build_mst_graph.py

Suggests ligand pairs



Open source toolkit for cheminformatics

Protein-Ligand Binding

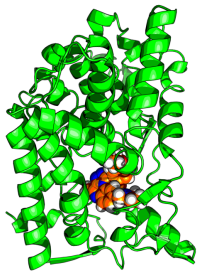
Application:
482 ligand modifications in protein-ligand binding

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

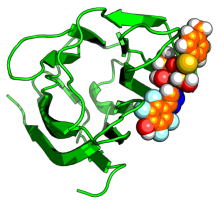
under review

Protein-Ligand Complexes

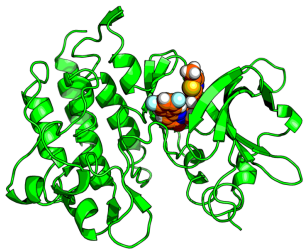
- 11 systems
- 482 modifications



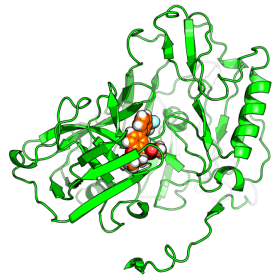
PDE2: 21 ligand
34 perturbations



Galectin: 8 ligands
8 perturbations

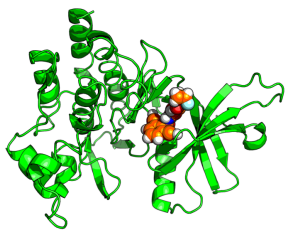


cMet: 12 ligands
25 perturbations

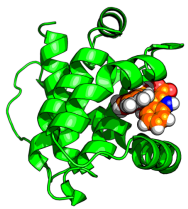


BACE: 80 ligands
144 perturbations
(divided in 3 sets)

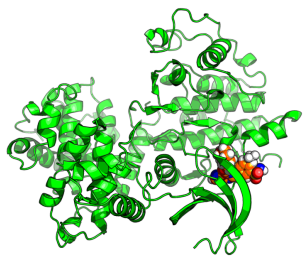
JNK1: 21 ligands
31 perturbations



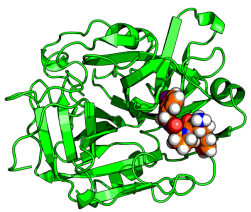
TYK2: 16 ligands
24 perturbations



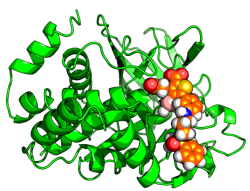
MCL1: 42 ligands
71 perturbations



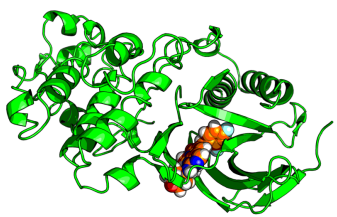
CDK2: 16 ligands
25 perturbations



Thrombin: 11 ligands
16 perturbations

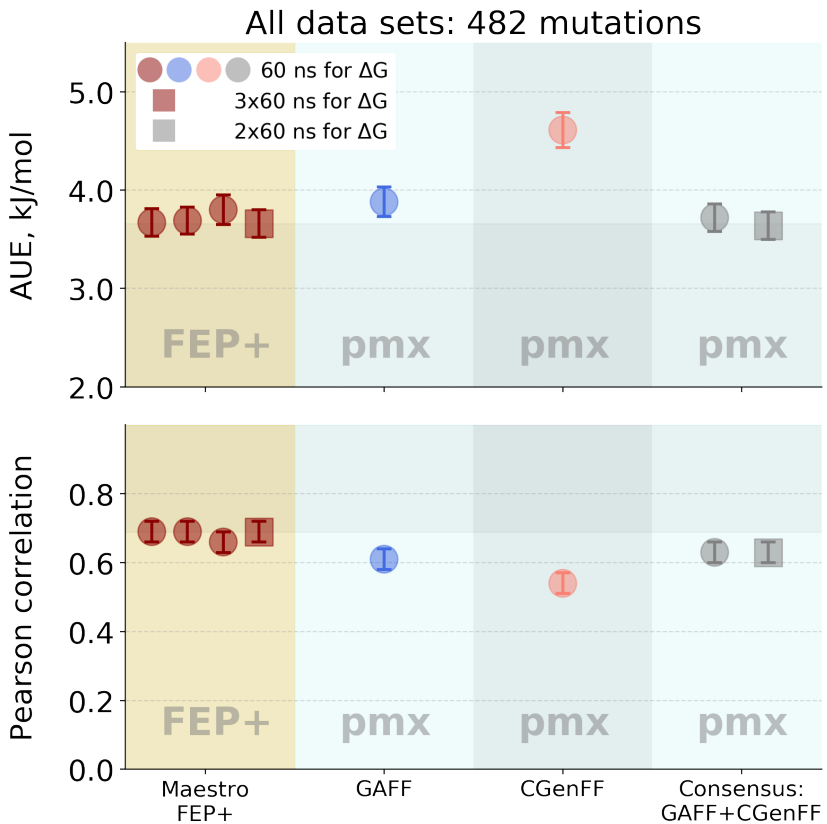


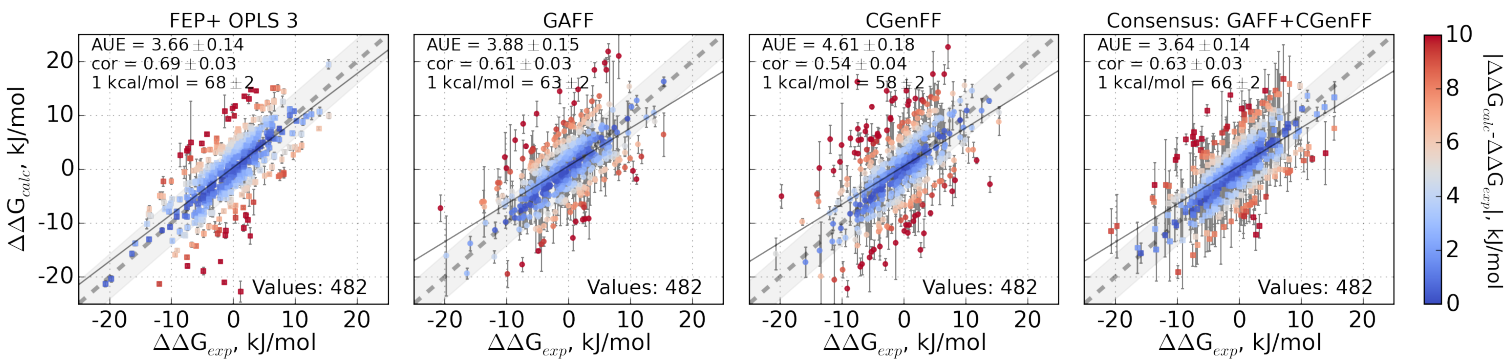
PTP1b: 23 ligands
49 perturbations



P38: 34 ligands
56 perturbations

Overall Results





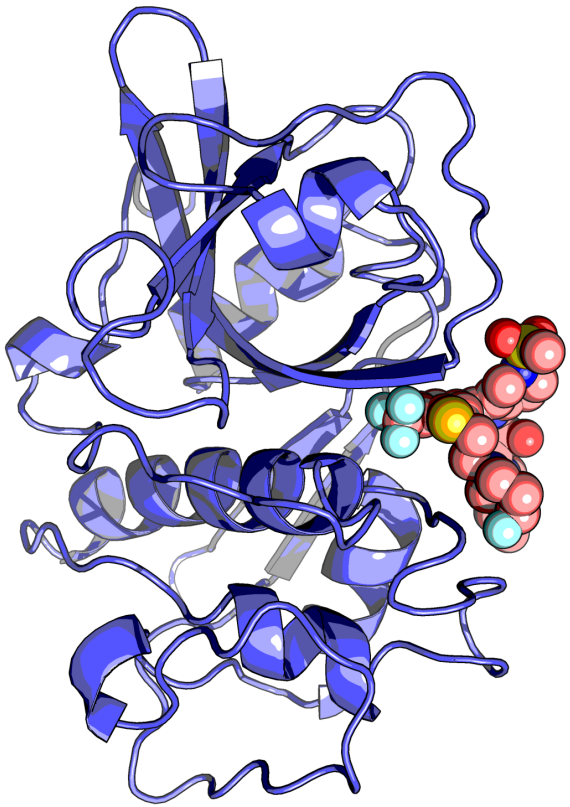
D3R: Cathepsin S

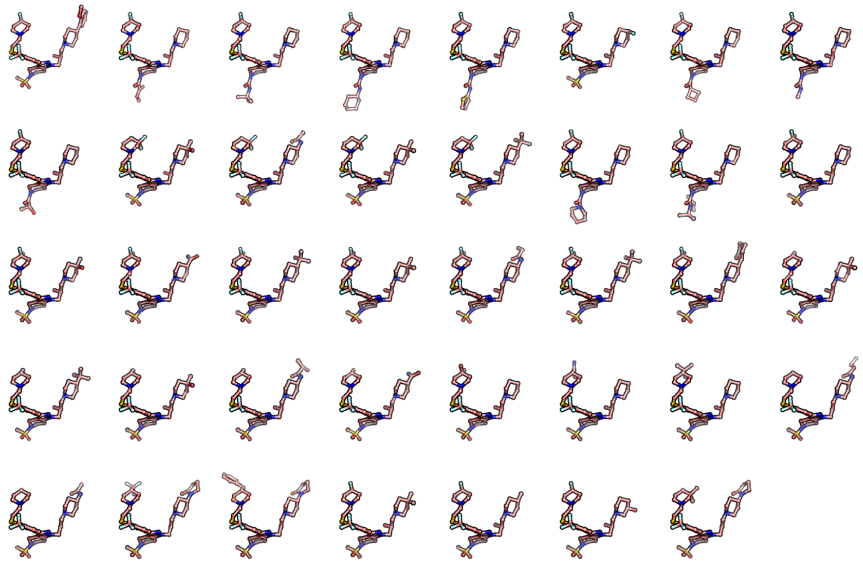
D3R: Free Energy Calculations Cathepsin S

Elisée, Gapsys, Mele, Chaput, Selwa, de Groot, Iorga

in revision

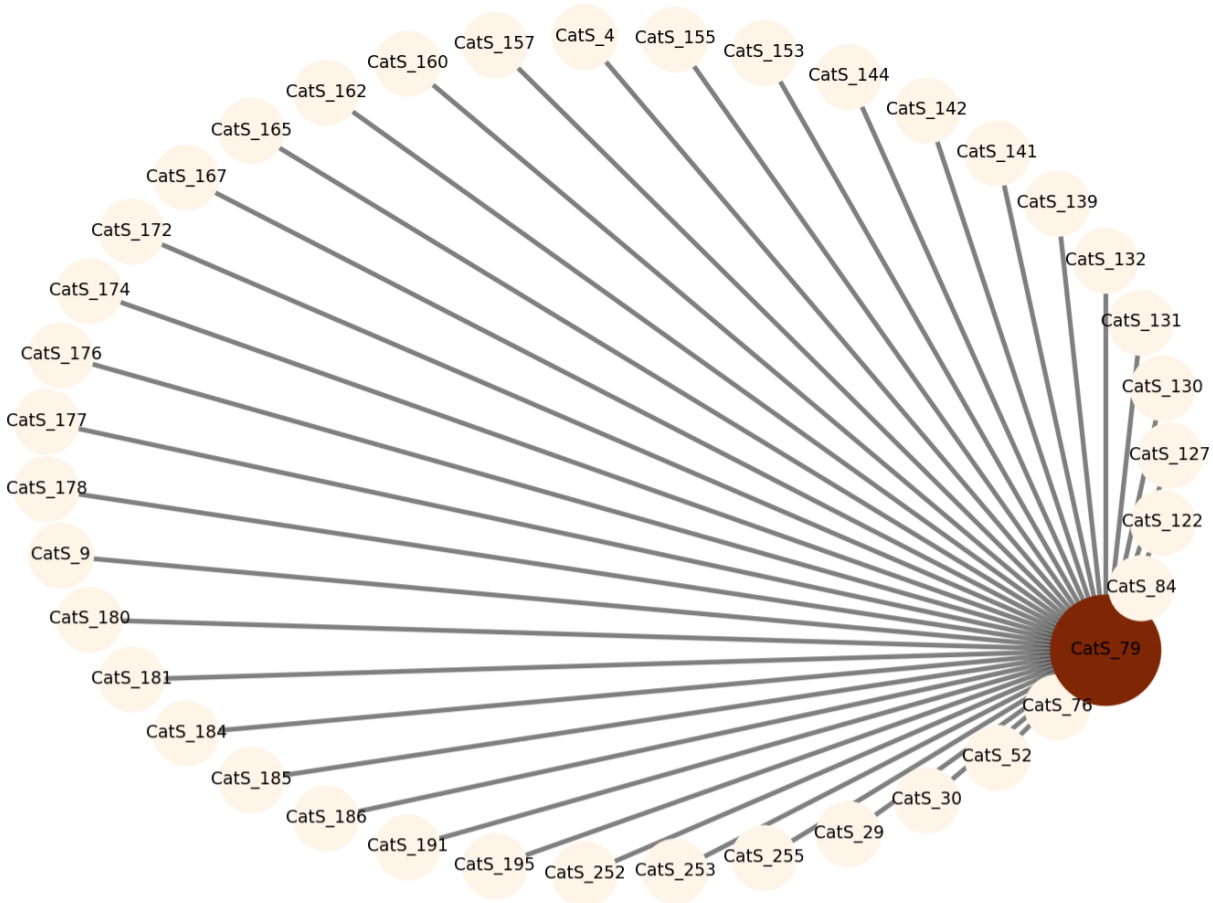
D3R: Cathepsin S





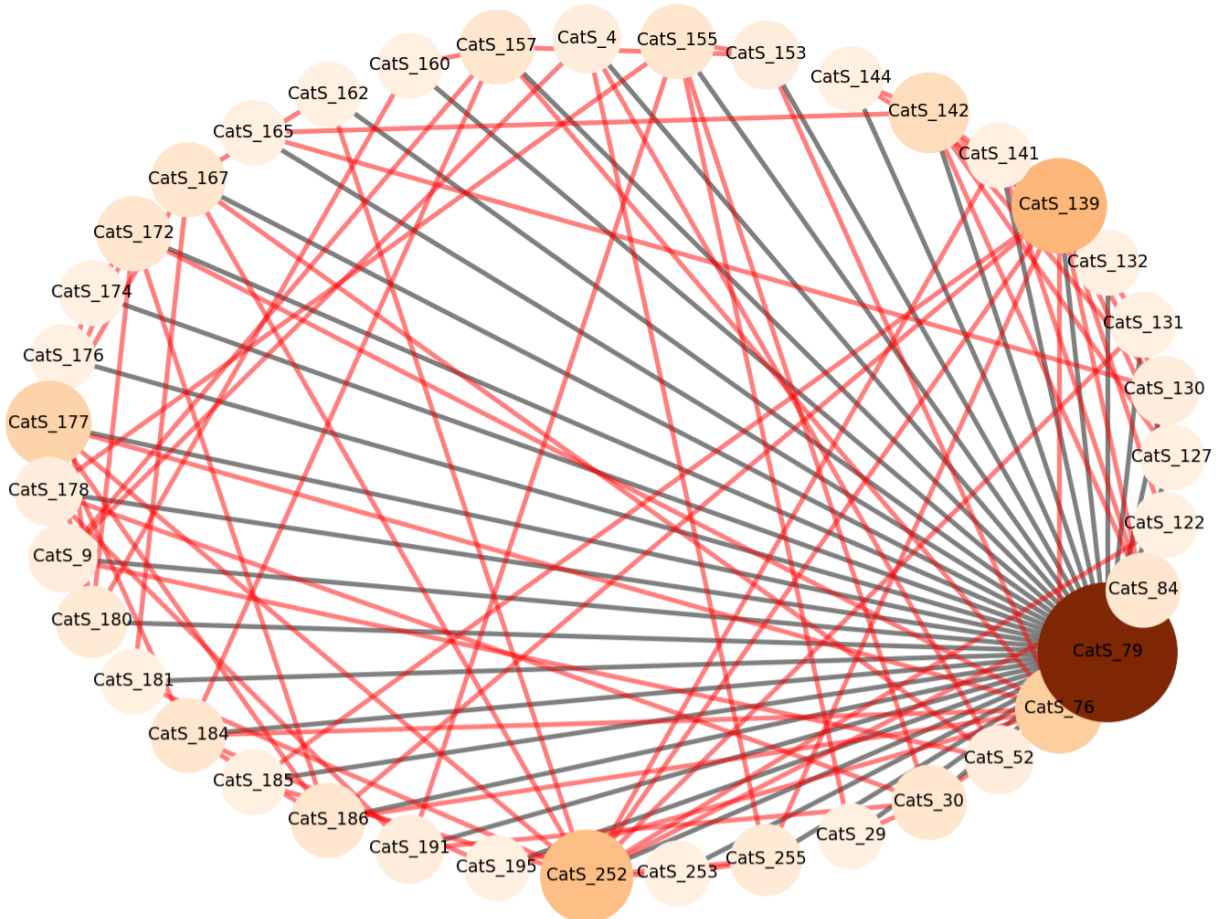
Poses generated by Eddy Elisée based on the crystallographic poses released in the previous D3R-GC3

D3R: Ligand Mapping



One ligand was selected as a reference.

D3R: Ligand Mapping

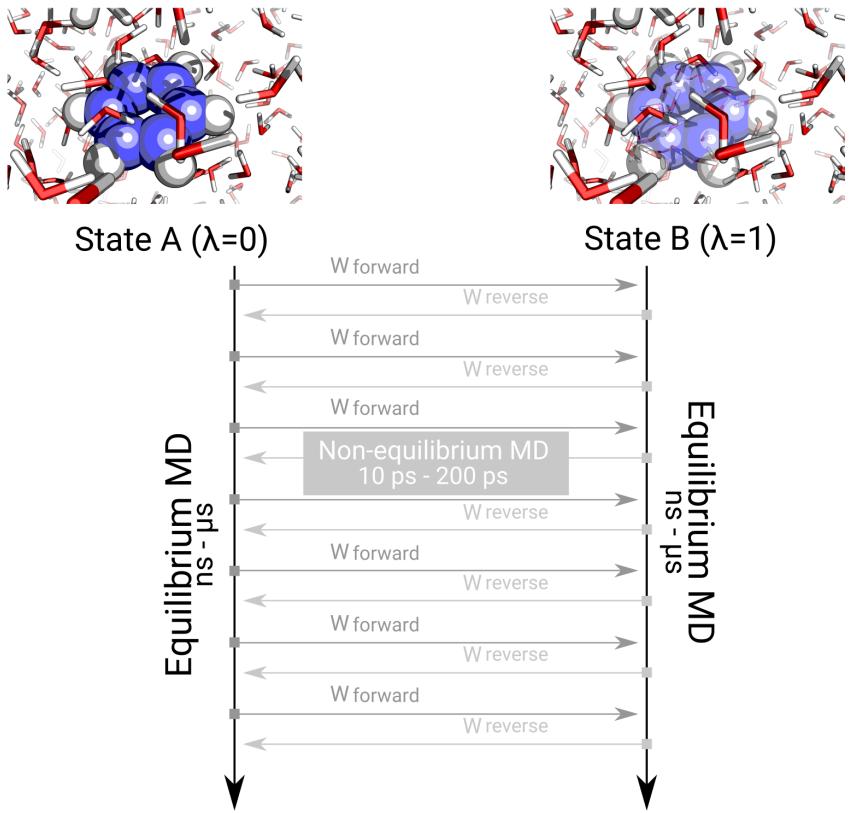


Redundancies were included.

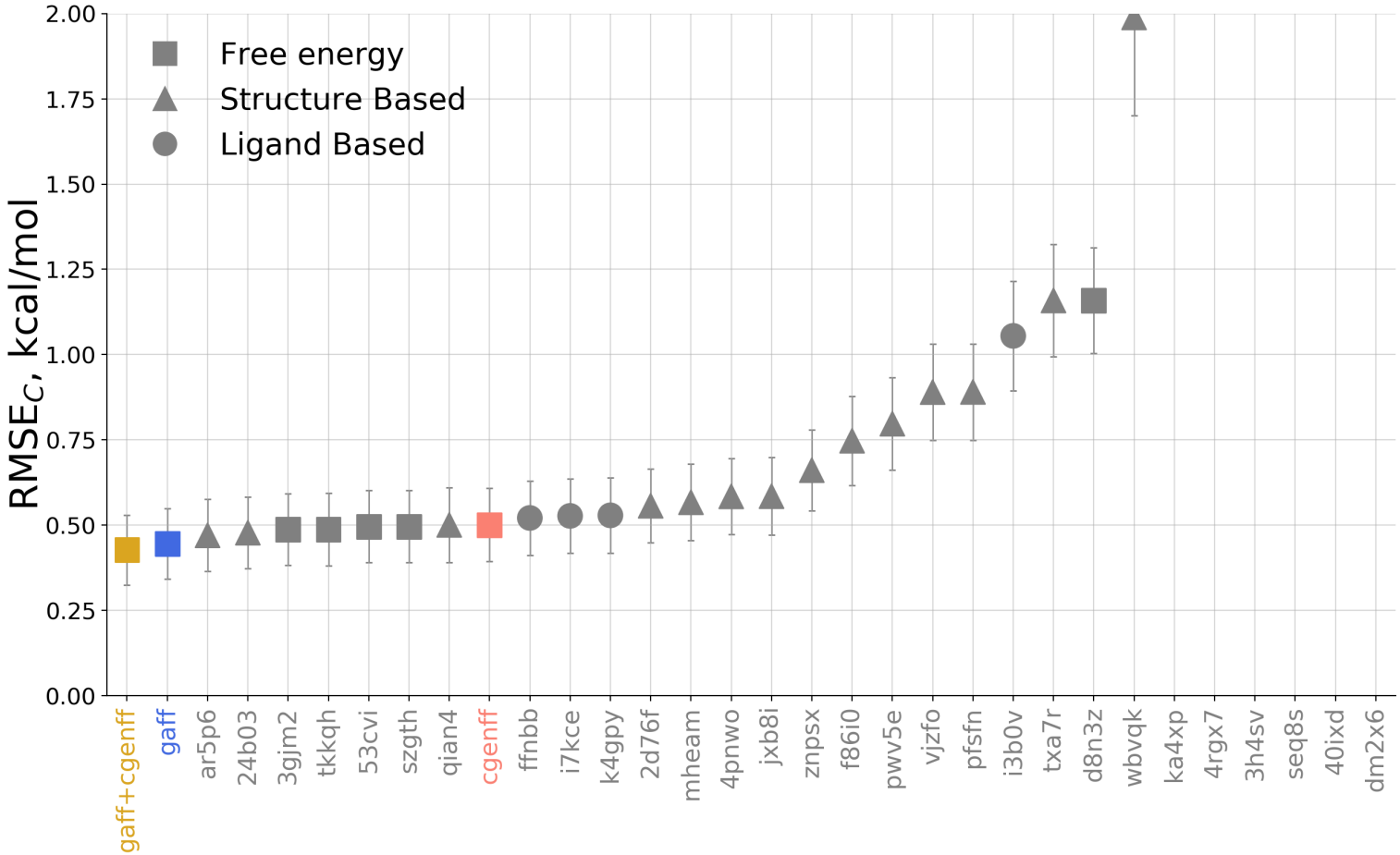
D3R: Calculation

Setup

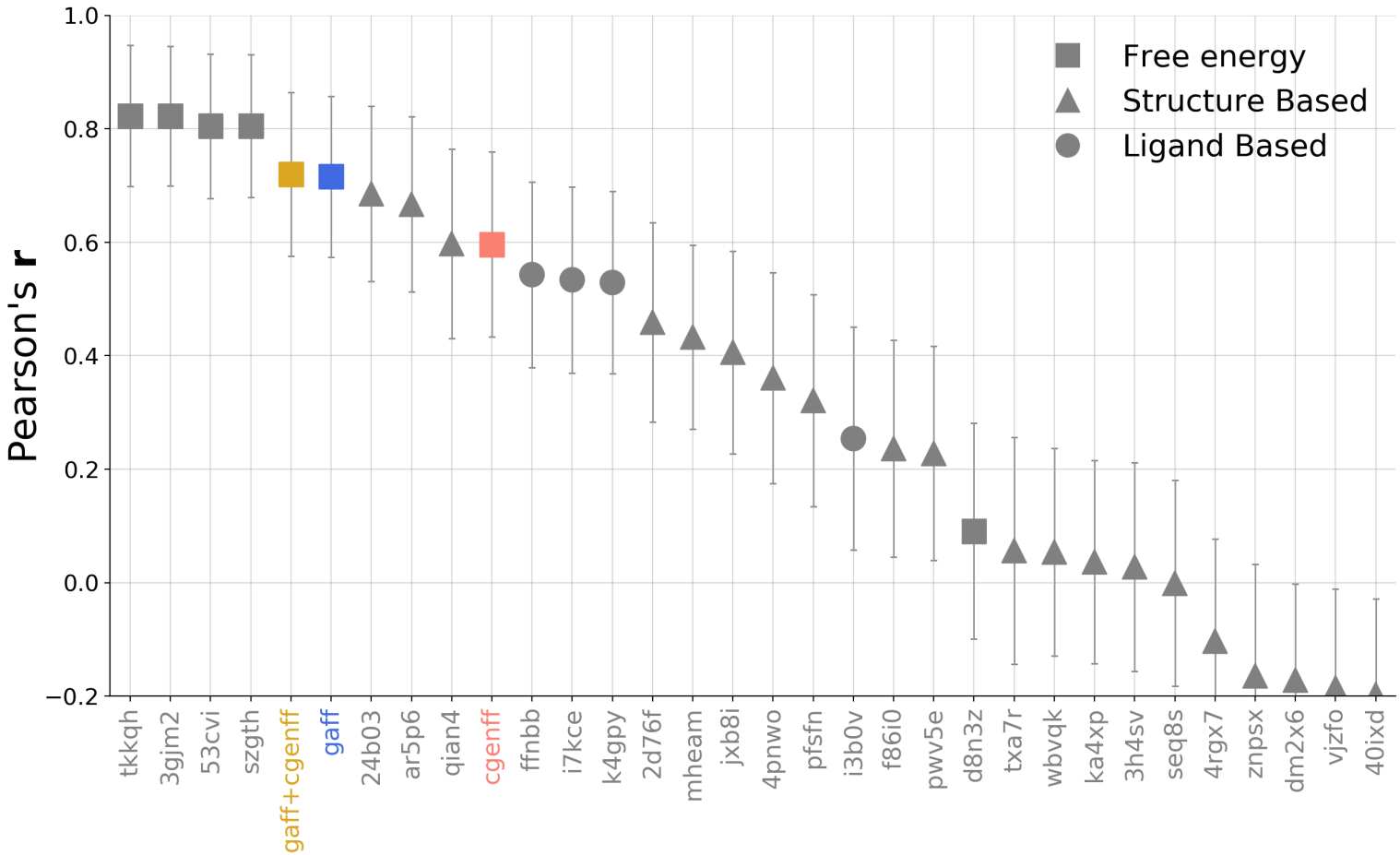
- Non-equilibrium protocol
- 109 edges
- 3 repeats per edge
- 44 ns per repeat
- 2 force fields: Gaff 2.1, CGenFF 4.1
- Gromacs2018 + pmx



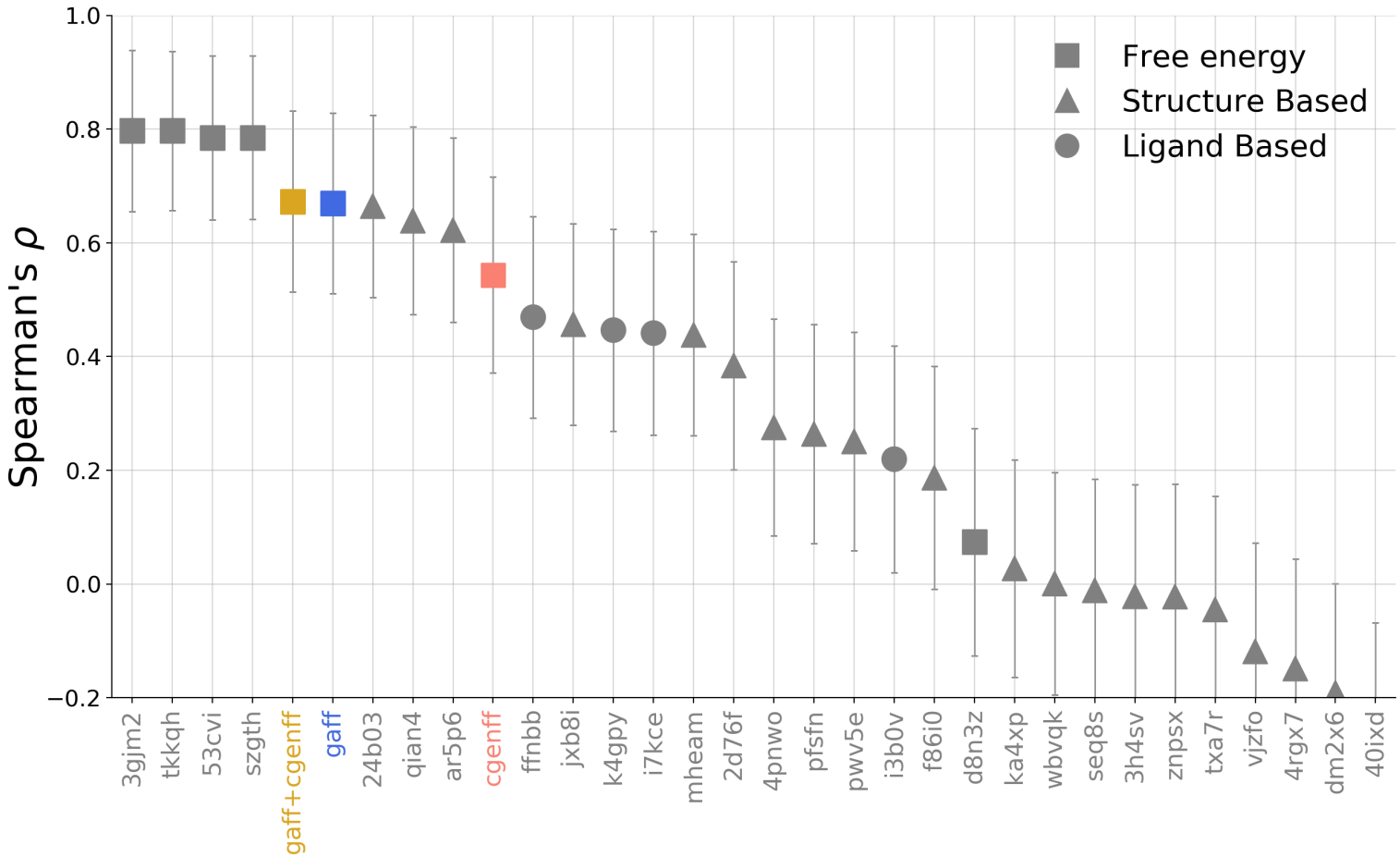
D3R: Results (RMSE)



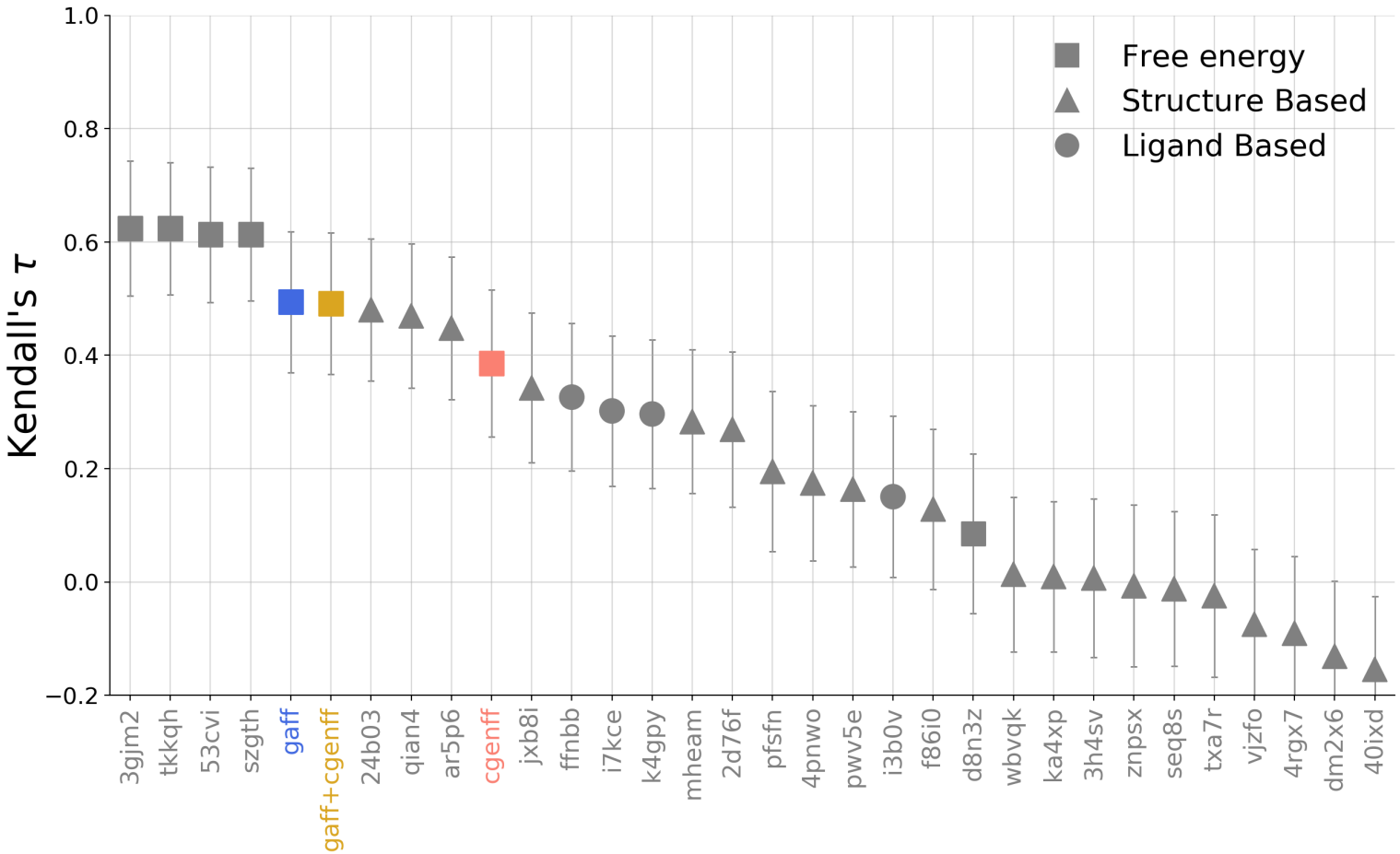
D3R: Results (Pearson correlation)



D3R: Results (Spearman correlation)

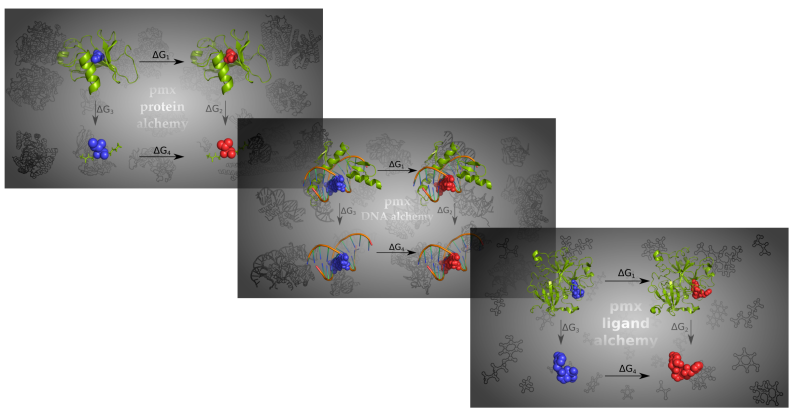


D3R: Results (Kendall correlation)



Summary

GROMACS
FAST. FLEXIBLE. FREE.



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

D3R: Cathepsin S

Dr. Bogdan I. Iorga
Eddy Elisee
Dr. Nawel Mele
Dr. Ludovic Chaput
Dr. Edithe Selwa

Funding