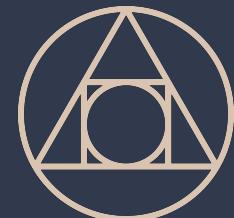


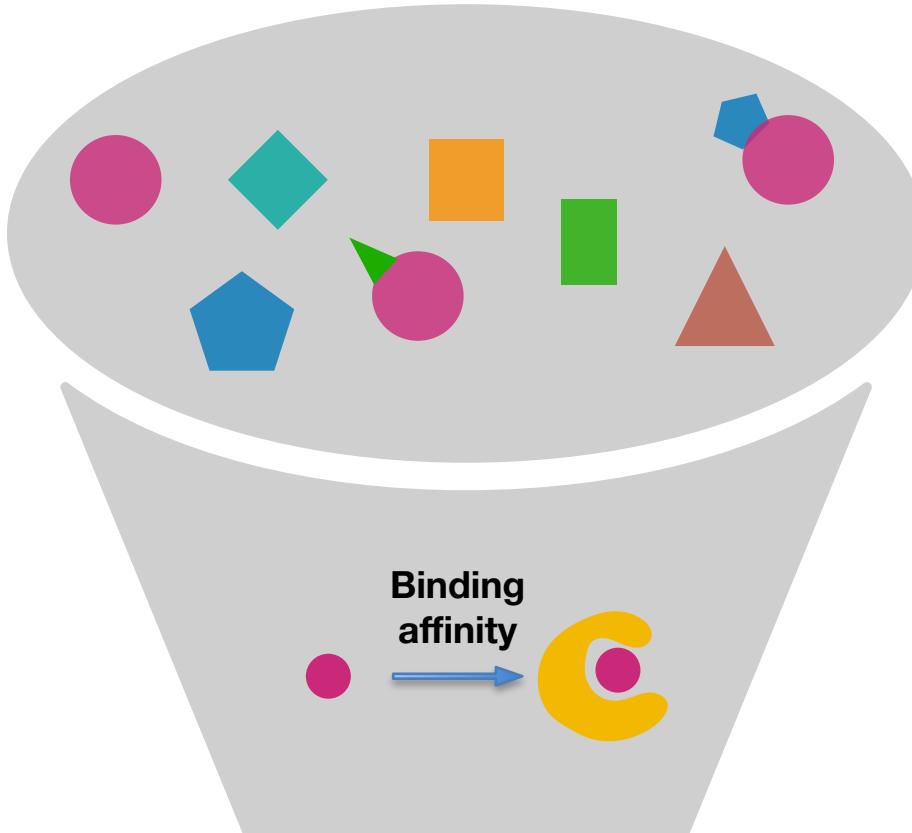
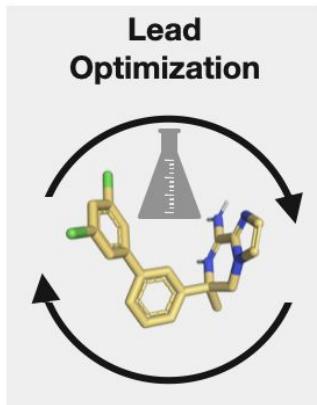
Breaking free from ligand similarity restrictions in binding free energy calculations

Hannah Baumann, OpenFreeEnergy (OMSF)
MDAnalysis UGM Lisbon, Sep 2023





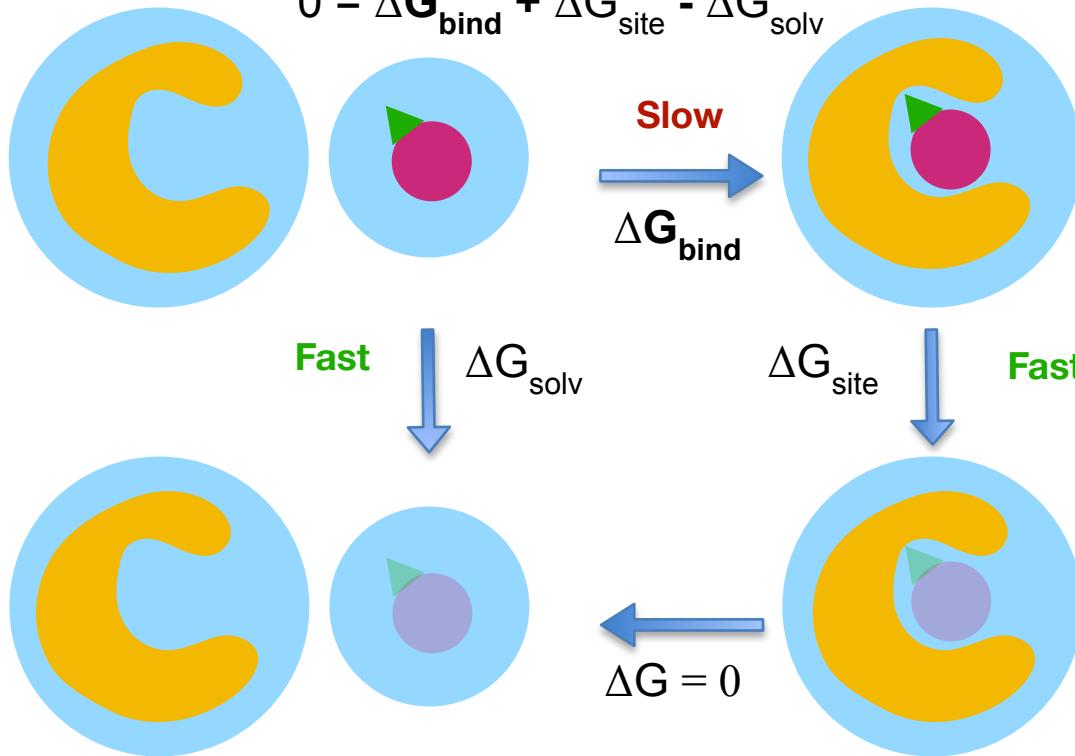
Binding free energy calculations play an important role in prioritizing the synthesis of novel drug candidates





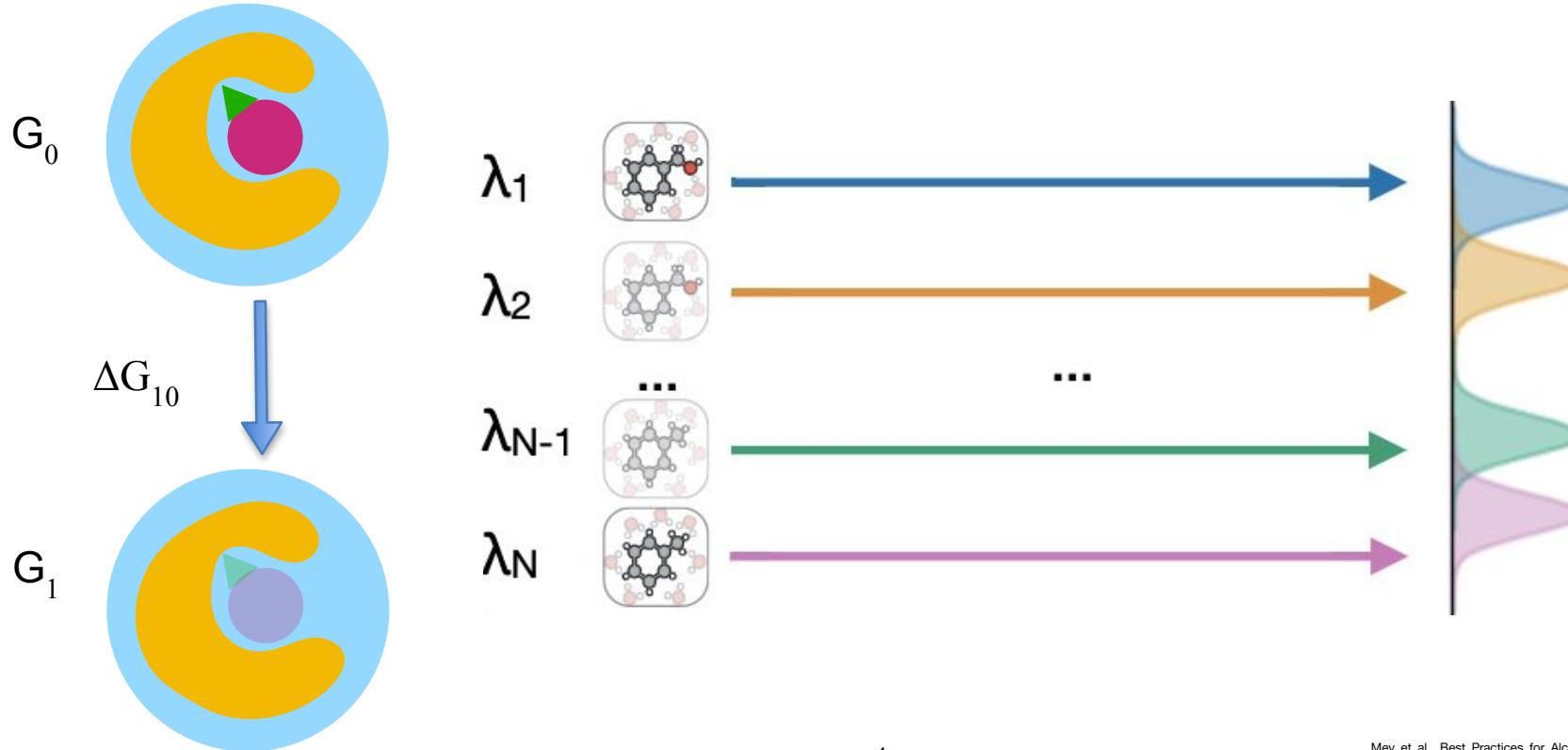
Absolute binding free energies (ABFE) can be obtained through a thermodynamic cycle

$$0 = \Delta G_{\text{bind}} + \Delta G_{\text{site}} - \Delta G_{\text{solv}}$$



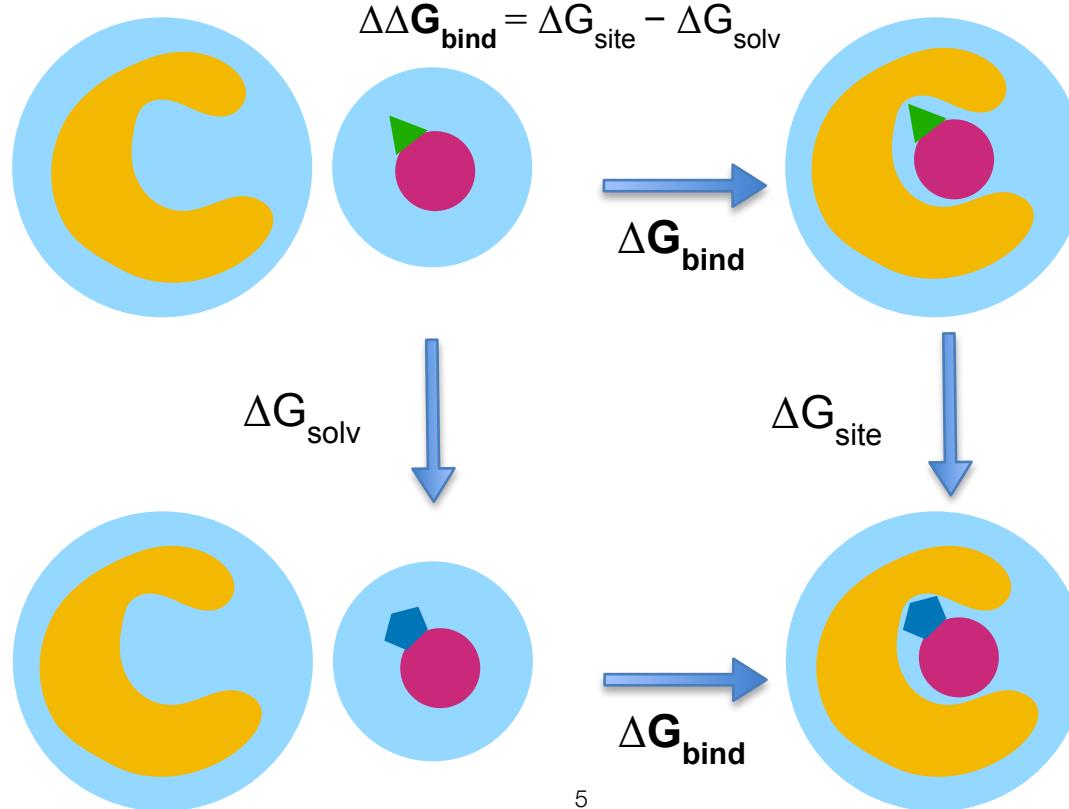


Intermediate states are introduced to achieve overlap between neighboring states





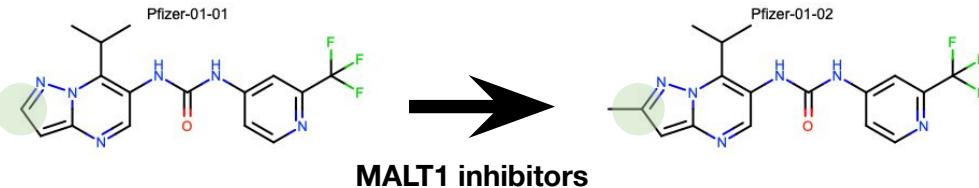
Relative binding free energies (RBFE) can be obtained through a thermodynamic cycle



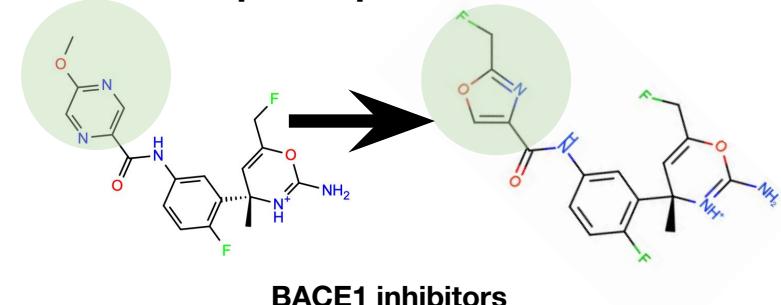


RBFE are limited to the comparison of structurally related ligands

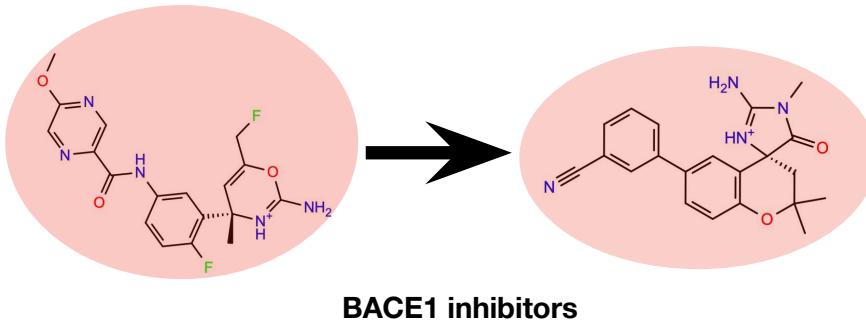
Small R-group modifications are feasible



Ring size changing transformations are feasible if special protocols are used



Large scaffold hopping transformations are outside the scope of the method





ABFE and RBFE have different strengths and weaknesses

ABFE



- No restriction on ligand structure
- No restriction on ligand binding mode

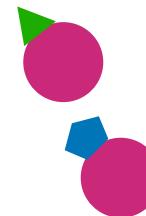


Sampling issues can be more severe than in RBFE



Computationally expensive

RBFE



- Ligands have to be similar
- Ligands have to have same binding mode
- Mostly R group modifications



Sampling of the apo system not necessary



Computationally efficient



Separated Topologies approach combines benefits of RBFE and ABFE



- No restriction on ligand structure
- No restriction on ligand binding mode



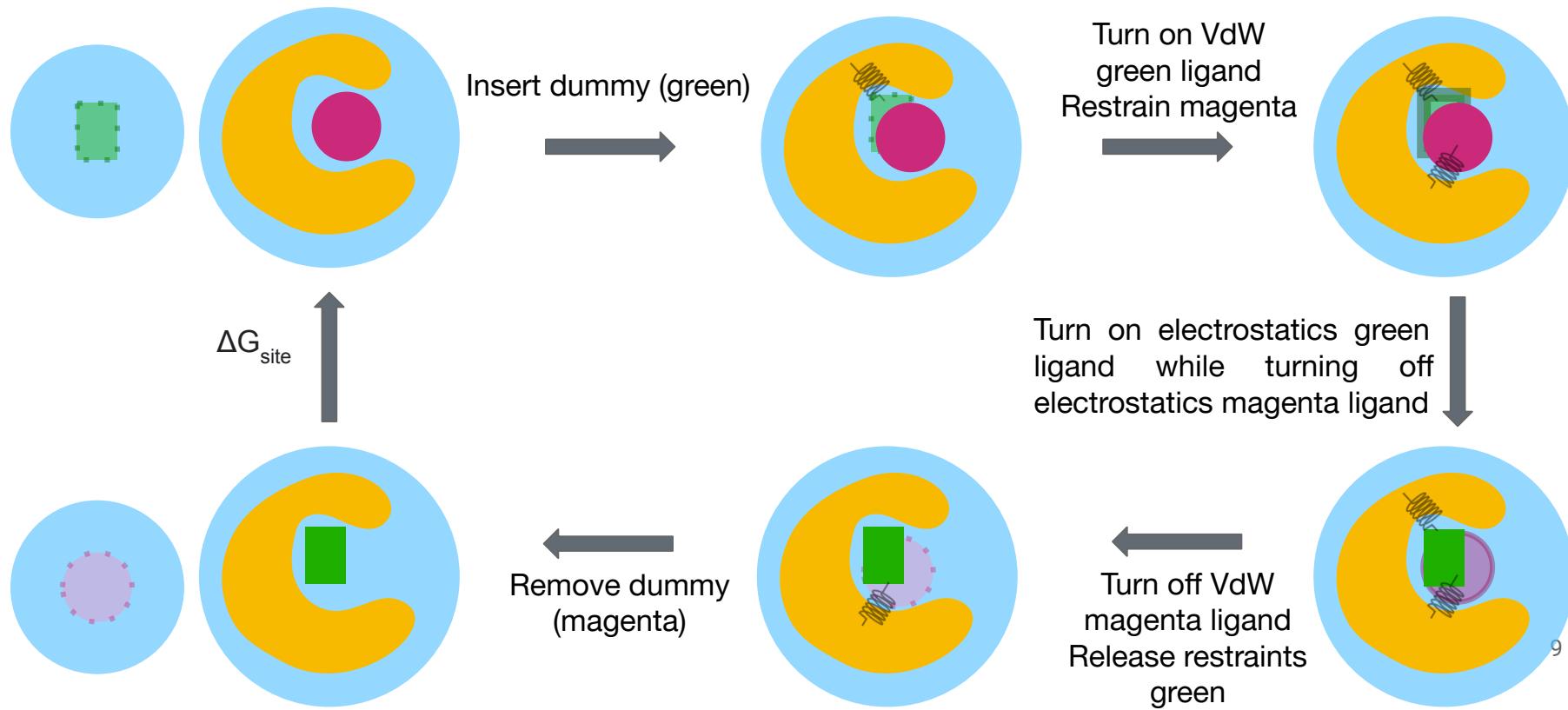
Sampling of the apo system not necessary



Computational efficiency somewhere between RBFE and ABFE



SepTop does RBFE by running two ABFE in opposite directions

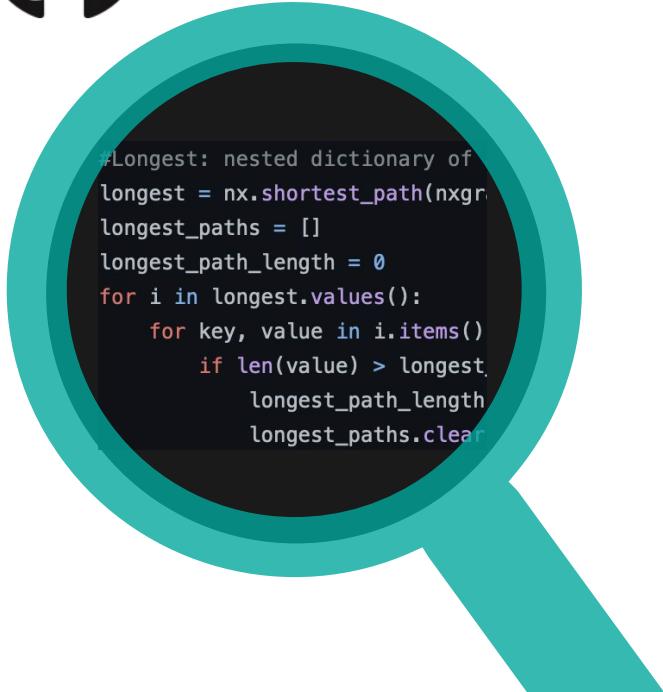




We developed a package to setup SepTop calculations



MobleyLab/[SeparatedTopologies](#)



```
#Longest: nested dictionary of
longest = nx.shortest_path(nxgr
longest_paths = []
longest_path_length = 0
for i in longest.values():
    for key, value in i.items():
        if len(value) > longest_
            longest_path_length
            longest_paths.clear()
```

**Now thanks to OpenFreeEnergy
RDKit compatible!!!**



Why “Separated Topologies” (SepTop)?

1. The topologies of the two ligands are treated separately

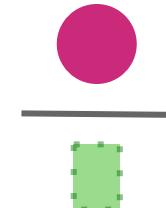
[atoms]

```
; nr      type resnr residue atom cgnr   charge     mass typeB   chargeB   massB
;residue  1 LIG
```

1	LIG1_C1	1	LIG C1	1	-0.23	12.01	scaled_LIG1_C1	0.0	12.01
2	LIG1_C2	1	LIG C2	2	-0.26	12.01	scaled_LIG1_C2	0.0	12.01

```
;residue  2 LIG
```

43	dLIG2_C1	2	LIG C1	43	0.0	12.01	scaled_LIG2_C1	-0.23	12.01
44	dLIG2_C2	2	LIG C2	44	0.0	12.01	scaled_LIG2_C2	-0.26	12.01



2. Exclude interactions between the two ligands

[nonbond_params]

```
; name i      name j    func   LJ-σ   LJ-ε
LIG1_C1      LIG2_C1    1        0       0
```



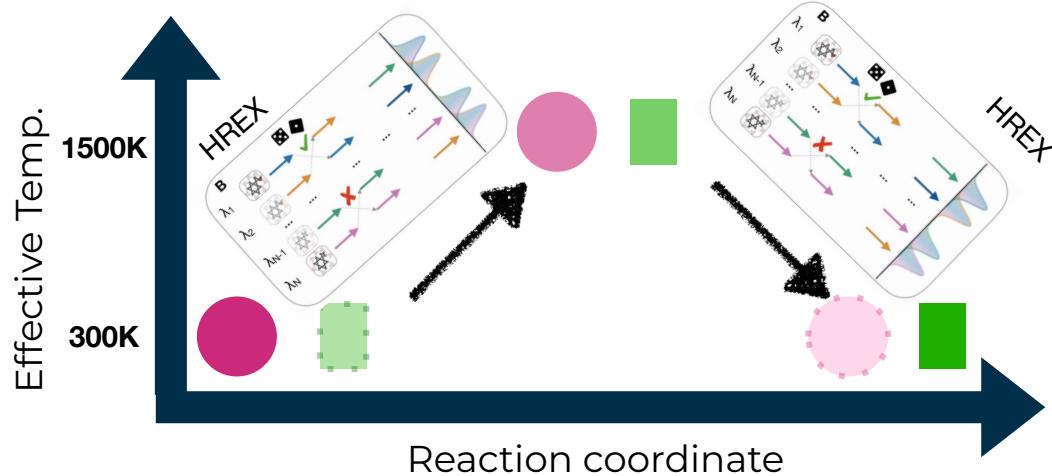


We scale down the LJ- ϵ to soften interactions and enhance sampling

ε -HREX:

- All replicas are run at the same temperature while the potential energy of every replica is scaled differently

$$V_{LJ}(\mathbf{r}_{ij}) = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$$



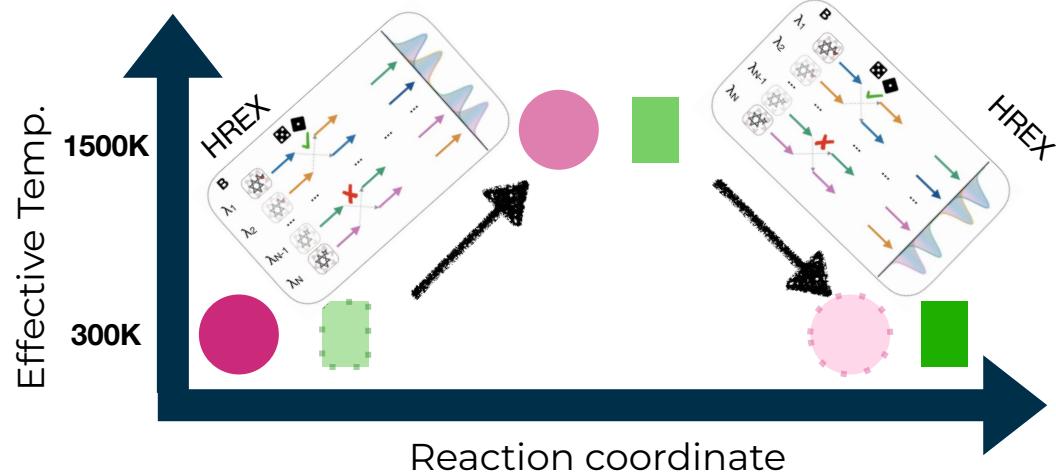


We scale down the LJ- ϵ to soften interactions and enhance sampling

ϵ -HREX:

- All replicas are run at the same temperature while the potential energy of every replica is scaled differently

$$V_{LJ}(\mathbf{r}_{ij}) = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$$



Implementation in GROMACS:

- Create “scaled” atom types: multiply LJ- ϵ by scaling factor γ (default $\gamma=0.5$)

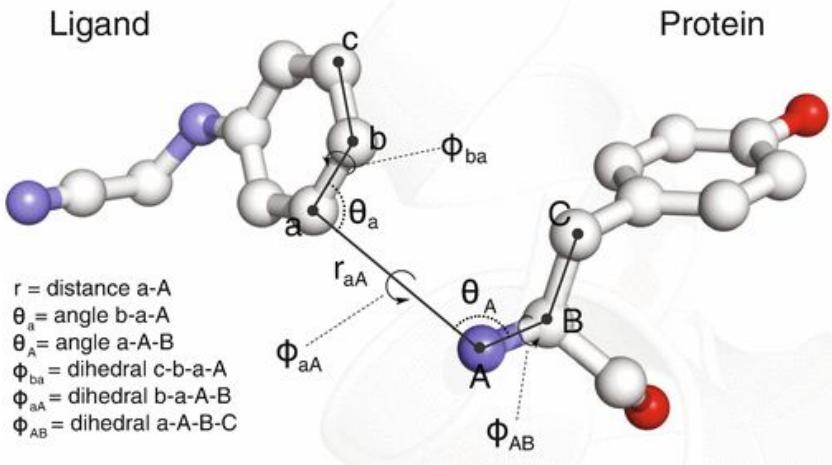
[atomtypes]

	name	at.num	mass	charge	ptype	LJ- σ	LJ- ϵ
	LIG1_L331	6	12.01	0.0	A	0.35	0.36
	scaled_LIG1_L331	6	12.01	0.0	A	0.35	0.18



We use orientational restraints to keep the non-interacting ligands in the binding site

Restraints are defined between 3 protein and 3 ligand atoms

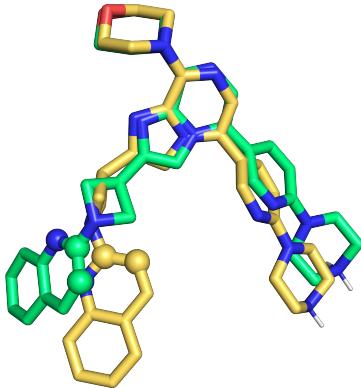


Absolute Binding Free Energies: A Quantitative Approach for Their Calculation, Boresch et al., 2003

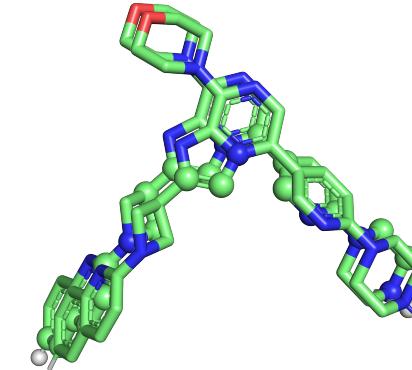


We developed heuristics for automatically picking suitable atoms for orientational restraints

Restraining “floppy” atoms can lead to slow convergence



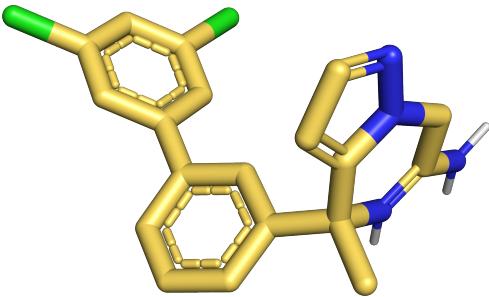
Central atoms are more likely to be stable



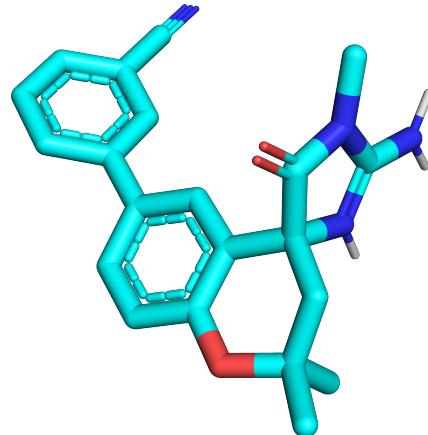


β -Secretase (BACE1) is an interesting test system with diverse ligand scaffolds

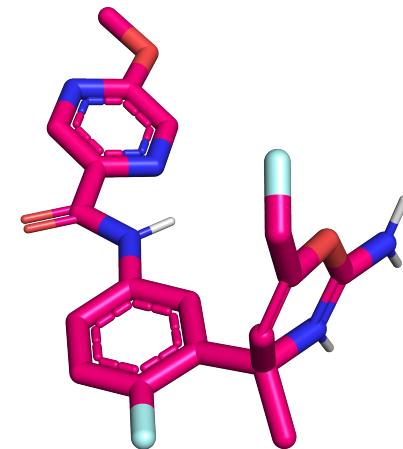
Biaryl series



Spirocyclic series

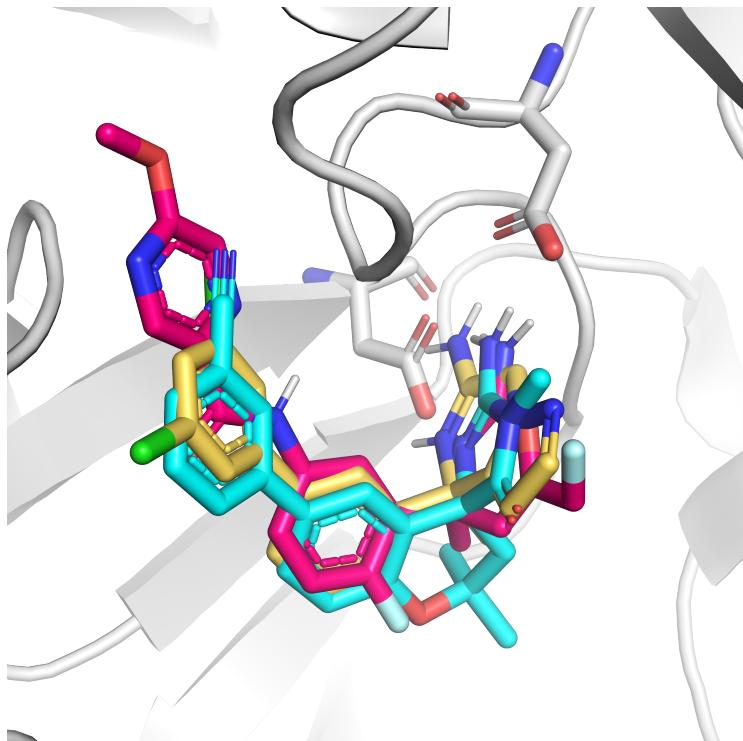


Amide series





BACE1 is expected to perform poorly with ABFE calculations due to protonation state change

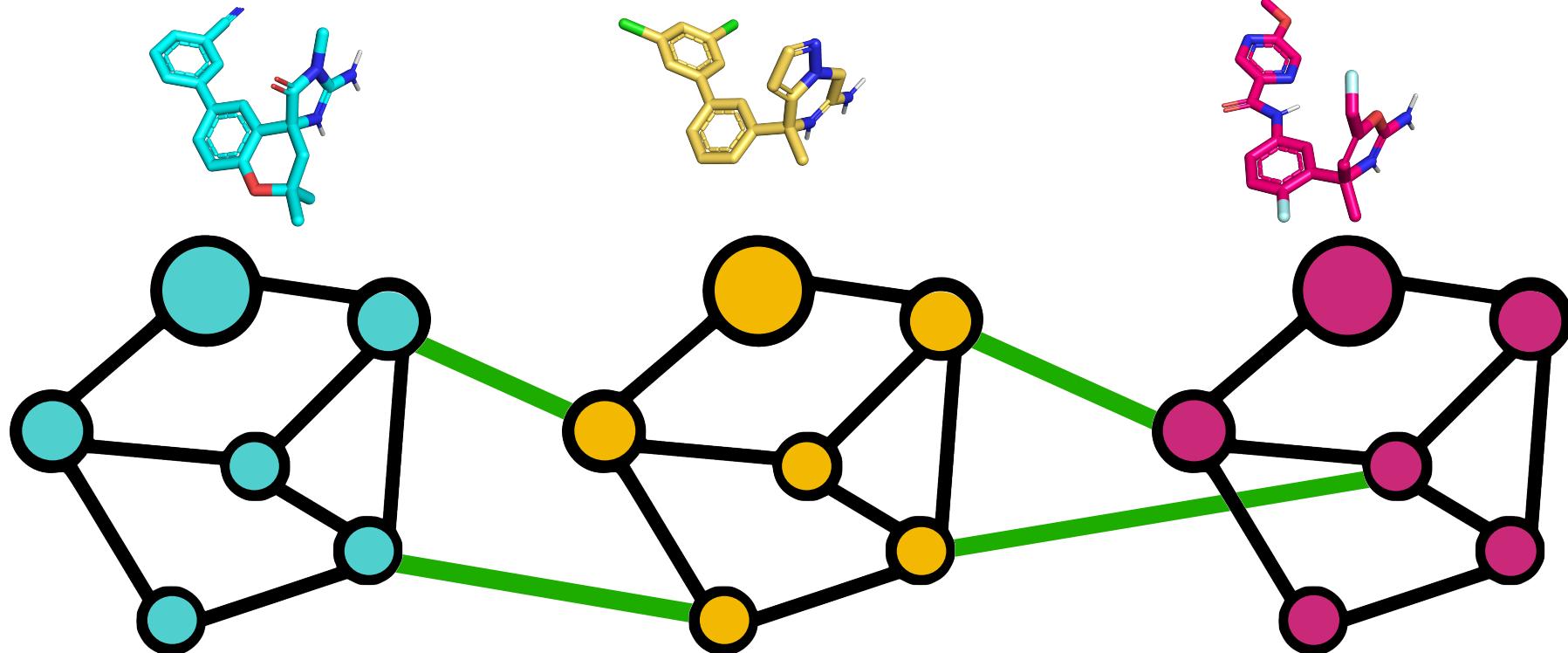


Ligands make key contacts with catalytic dyad, stabilize deprotonated state

In unbound state, must change protonation or binding site will be unstable and induce protein rearrangement and severe sampling problems

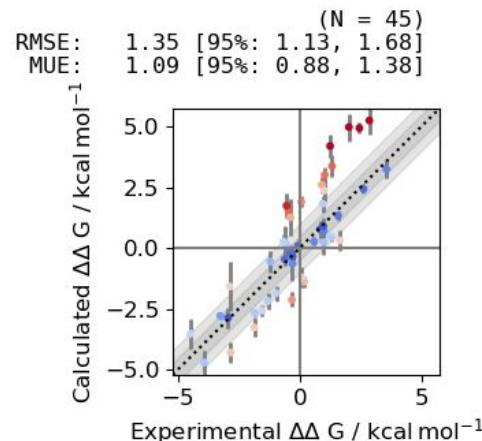
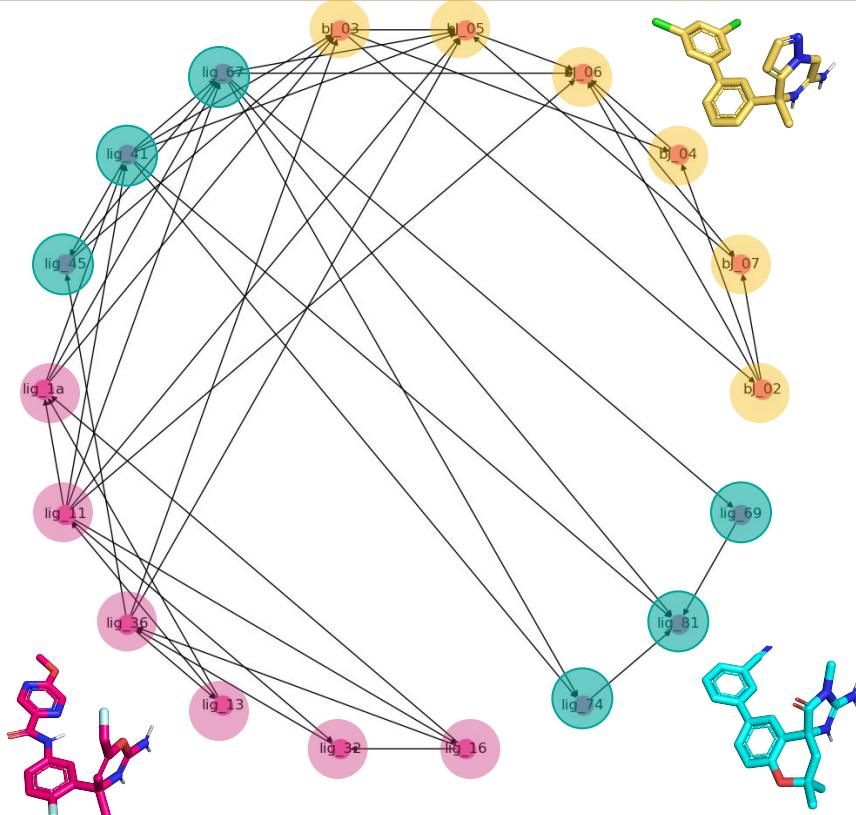


With SepTop, we can calculate the relative binding free energy of ligands within a series and across series

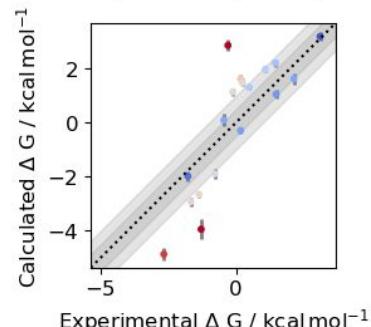




With SepTop, we obtained good results for this BACE1 dataset

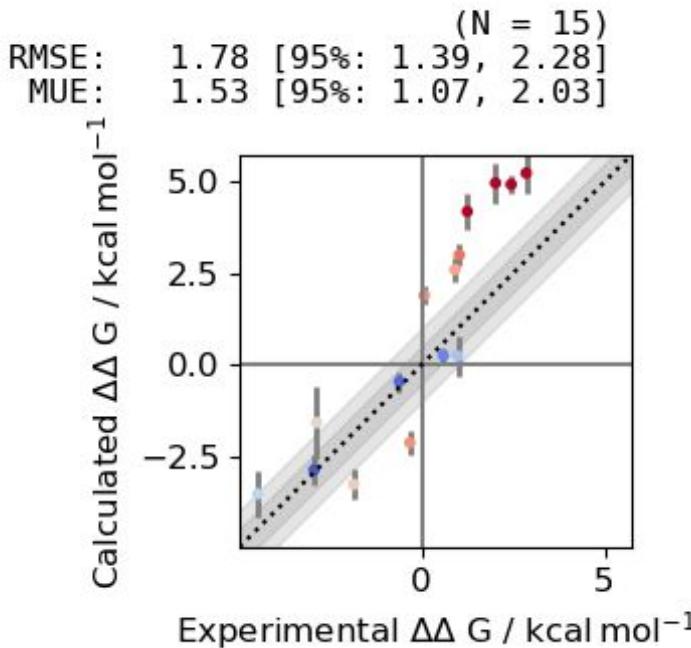
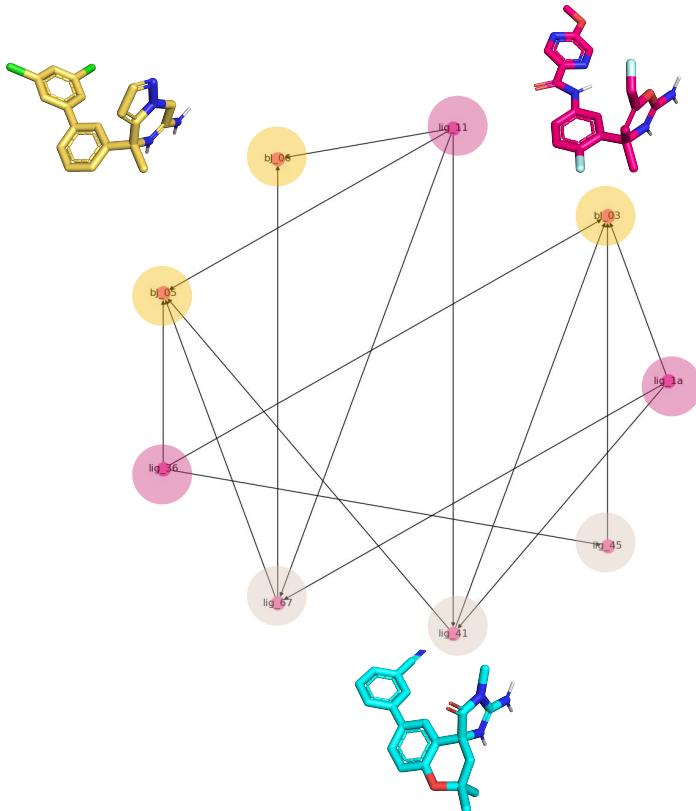


RMSE: 1.39 [95%: 0.92, 1.80]
MUE: 1.13 [95%: 0.79, 1.55]
R2: 0.71 [95%: 0.43, 0.89]
rho: 0.84 [95%: 0.67, 0.94]
RAE: 0.98 [95%: 0.61, 1.75]
KTAU: 0.64 [95%: 0.33, 0.83]





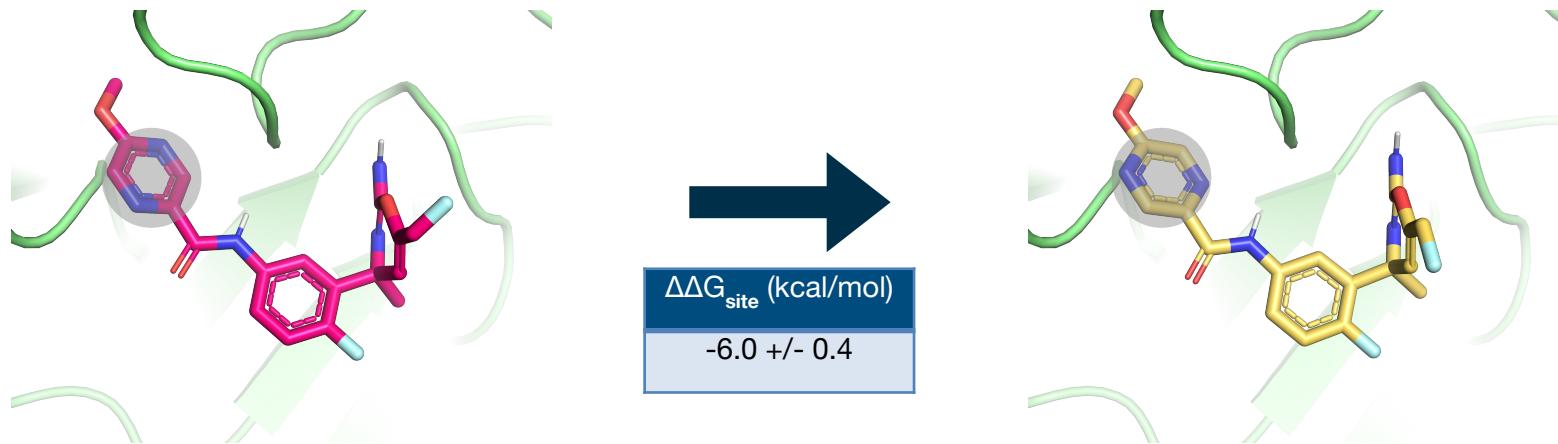
Error statistics for transformations across series look slightly worse, however, still promising considering the size of the changes





Adequate sampling of the binding mode can pose challenges

- The orientation of the pyrazine is experimentally unknown
- Running SepTop between the two poses shows that one pose is more favorable than the other





Creating a custom Reader for alchemistry



We create trajectory data that describe multiple replicas at different (and exchanging) lambda states.

To ingest this into MDAnalysis, we wrote a **custom Reader class**. This allows us to read a non-supported format.

The extensible Reader plugin system allows this without modifying the MDAnalysis source.

See:

https://github.com/OpenFreeEnergy/openfe_analysis/blob/main/src/openfe_analysis/reader.py

```
class FEReader(ReaderBase):
    """A MDAnalysis Reader for nc files created by openfe RFE Protocol

    Looks along a multistate .nc file along one of two axes:
    - constant state/lambda (varying replica)
    - constant replica (varying lambda)
    """
    _state_id: Optional[int]
    _replica_id: Optional[int]
    _frame_index: int
    _dataset: nc.Dataset
    _dataset_owner: bool

    format = 'openfe RFE'

    def __init__(self, filename, convert_units=True, **kwargs):
        """



```



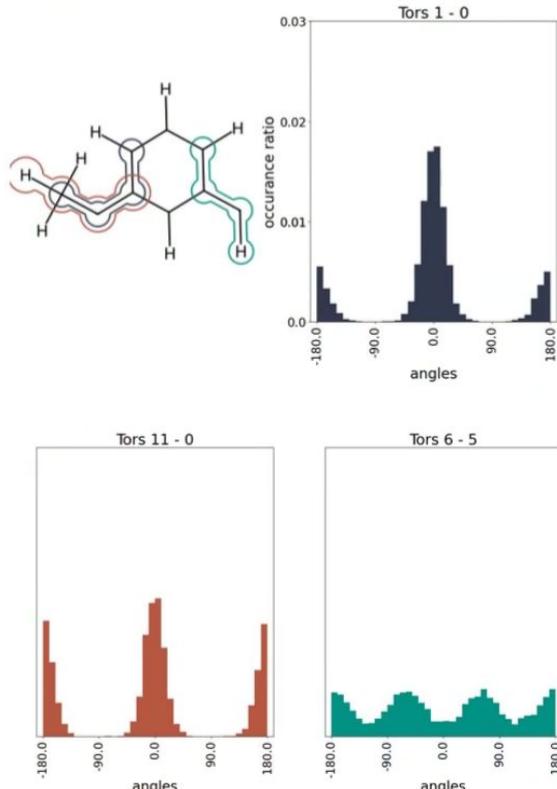
Dihedral Analysis using RDKit conversion



Using the `AtomGroup.convert_to('rdkit')` method, (Thanks Cedric!), we can **create a RDKit Mol of our AtomGroup**.

As this preserves atom ordering we can do cheminformatics (here finding rotatable bonds) and transfer these indices back to MDAnalysis. We then use the Dihedral analysis class to generate distributions of dihedrals.

Similarly, we can draw a nice representation of the dihedral using rdkit, converting our AtomGroup for this.





Ligand RMSD using on-the-fly Transformations



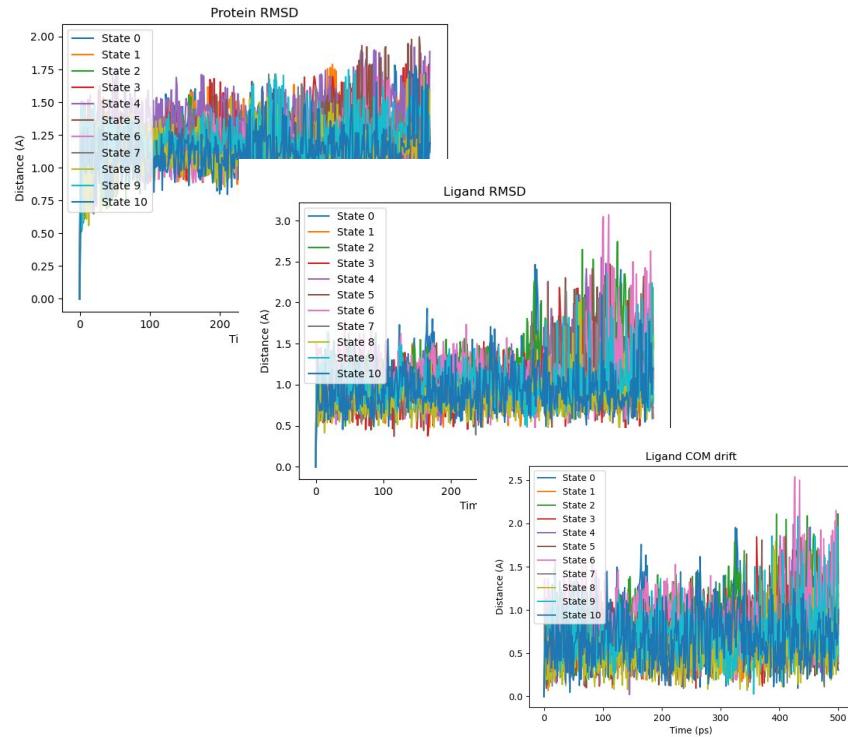
We use **on-the-fly transformations** to center and align our entire system relative to the protein. This allows us to correctly calculate the ligand RMSD.

We also use **transformations** to force the binding small molecule into the same image as the binding pocket, allowing us to easily calculate ligand COM drift (to detect bad simulations).

Using Transformations here made such automatic pre-processing far easier.

See:

https://github.com/OpenFreeEnergy/openfe_analysis/blob/main/src/openfe_analysis/transformations.py





Acknowledgements



MobleyLab:

David Mobley
Chris Zhang
Martin Amezcuia
Oanh Tran
Trevor Gokey
Anjali Dixit
Mary Pitman
Pavan Behara
Swapnil Wagle
Ana Caldaruse
Ak Nandkeolyar
Meghan Osato



OpenFreeEnergy:

Richard Gowers
Irfan Alibay
Benjamin Ries
David Swenson
Mike Henry
James Eastwood

Pfizer for funding & collaboration on the SepTop work:

Eric Dybeck
Alan Mathiowitz
Christopher McClendon

Janssen for providing the BACE structures:

David Hahn
Laura Pérez Benito
Gary Trasadern
Vytautas Gapsys

OpenEye for licenses and scientific discussions:

Christopher Bayly
Gaetano Calabró