Tech assignment – protein interaction probing

Tiejun Wei

https://github.com/DerienFe/InstaD_pro_abc

File/repo

- Please refer to the Github repo:
- https://github.com/DerienFe/InstaD_pro_abc
- The main.sh also provides a markdown file:
- https://github.com/DerienFe/InstaD_pro_abc/blob/main/main.md

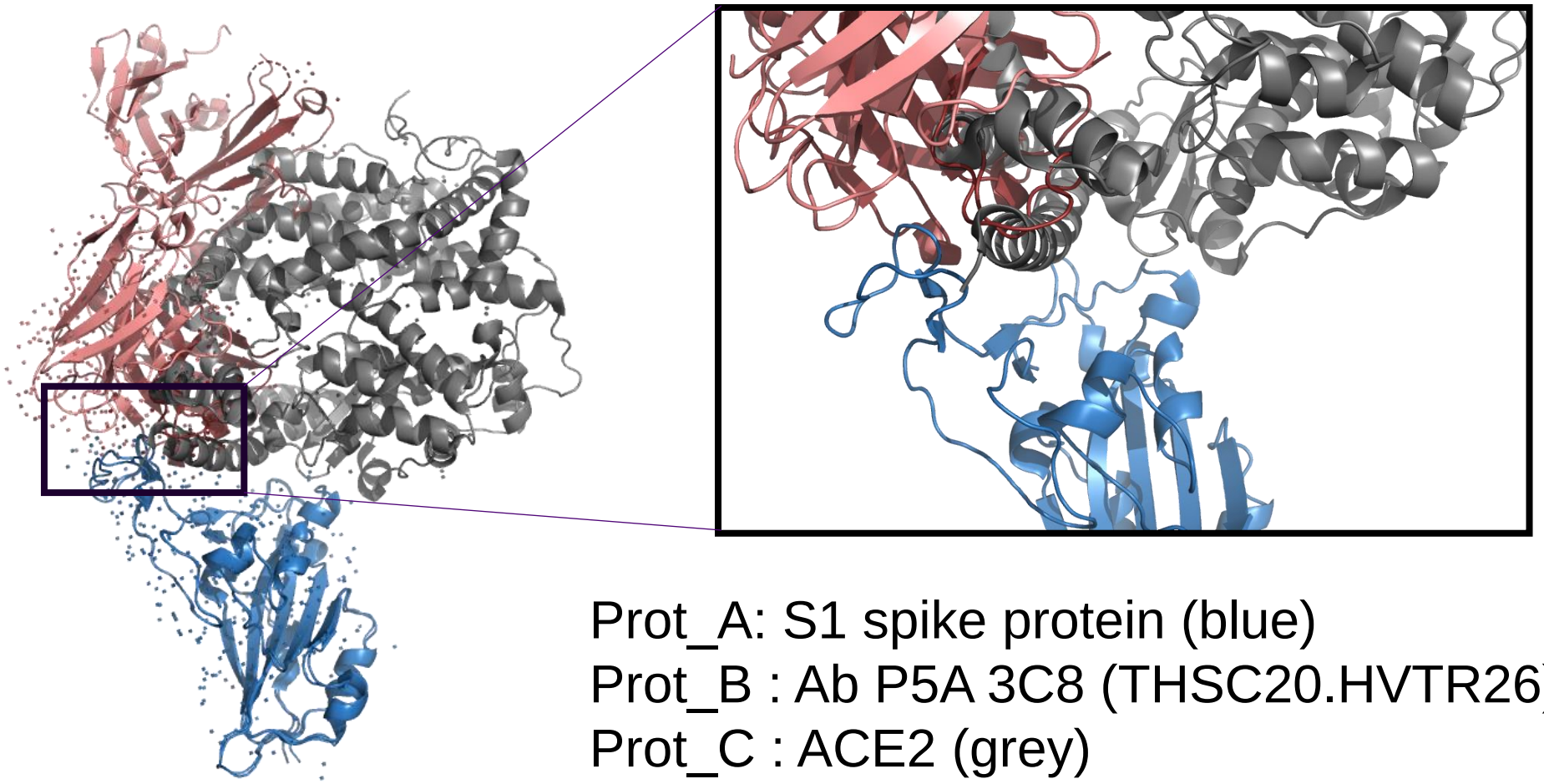
Intro

- Task:
- Probe the interface of Prot_A and Prot_B/Prot_C.

- To provide insight how to redesign Prot_A such that it abrogate binding with Prot_B while maintain binding affinity for Prot_C.

- Practical use: Redesign the antibody such that
  - Antibody Target ✓
  - Antibody ADE target ✗

Intro: example structure overview



Prot_A: S1 spike protein (blue)
Prot_B : Ab P5A 3C8 (THSC20.HVTR26) (red)
Prot_C : ACE2 (grey)

https://www.rcsb.org/structure/6m0j

https://www.rcsb.org/structure/7Z0X

Analytical pipeline overview: functionality

- Input: [pro_a, pro_b, pro_c]
- Output:
  - Binding free energy on pair A-B, A-C;
  - FEP on selected CV space.

- Dependency:
  - Python with conda
  - GROMACS/AMBER/OpenMM as MD engine
  - Shell (execution and/or cluster submission)

Analytical pipeline overview: pipeline structure

- Phase 1: Preparation of structure
  - Search for binding pattern (skipped since we have bounded structures)
  - Water/Ion/Protonation state (skipped, H++ server down)
  - MD simulation system construction using tleap
  - Generation of classic MD run files (GROMACS)
  - Generation of complex file for MMPBSA (gmx_MMGBSA in conda)
- Phase 2: MD simulations
  - Consists of minimization/NVT/NPT and production run
- Phase 3: post-MD analysis
  - MD simulation validation and visualization
  - MMGBSA calculations
  - Analysis on per-residue decomposition
  - Select a CV and run metadynamics

Phase 1

Analytical pipeline: implementation (1)

$ main.sh

```
#this is the main sh file for whole automated process
#by TW 07th Dec 2023
#initialize all the folder structure

#########################
# 0. initialization
#########################

#!/bin/bash
mkdir ./mmgbsa
mkdir ./mmgbsa/ab
mkdir ./mmgbsa/ac

mkdir ./metaD
mkdir ./metaD/trajectory
mkdir ./metaD/aux_file_dir
```

- Automation of

Phase 1

Analytical pipeline: implementation (2)

```
18  ########################
19  # 1. system preparation
20  ########################
21
22  source /home/tj/miniconda3/etc/profile.d/conda.sh
23  conda activate gmxMMPBSA
24
25  pdb4amber -i ./pro/pro_a.pdb -o ./pro/pro_a.amber.pdb -y >> log.txt &&
26  pdb4amber -i ./pro/pro_b.pdb -o ./pro/pro_b.amber.pdb -y >> log.txt &&
27  pdb4amber -i ./pro/pro_c.pdb -o ./pro/pro_c.amber.pdb -y >> log.txt &&
28
29  python mmgbsa_tleap_gen.py >> log.txt &&
30
31  tleap -f ./mmgbsa/gen_complex.tleap >> log.txt &&
32
```

- Automation of pdb pre-processing for AMBER.
- Automated tleap generation -> construction of complex structure
- Execution of tleap.

Phase 1

Analytical pipeline: implementation (3)

```
33    #convert prmtop/inpcrd to gro since we only have access to gromacs...
34    python file_preparation.py >> log.txt &&
```

- File_preparation.py contains follows
  - For each working directory under
    - Generate .gro and .top file usi
    - Generate .mdp file for GROMA
    - Write the MMGBSA script give calculation type (qm/pb/gb)
    - Write the local run script that c commence the commands.

```
535  def file_preparation(mmgbsa_dir):
536  >   """...
541      workdirs = glob(f"{mmgbsa_dir}/*")
542      workdirs = [workdir for workdir in workdirs if os.path.isdir(workdir)]
543      for workdir in workdirs:
544          #conversion of prmtop/inpcrd to gro/top
545          prmtop_files = glob(f"{workdir}/complex.prmtop")
546          basenames = [os.path.basename(prmtop_file).split(".")[0] for prmtop_file in
547          for basename in basenames:
548              convert_gmx(workdir, basename)
549              print(f"converting {basename} to gmx format")
550
551          #generate mdp files for running MD gromacs
552          gmx_mdp_writer(workdir)
553          print(f"generating mdp files for {workdir}...")
554
555          #generate mmgbsa input files for running mmgbsa gromacs
556          mmgbsa_writer(workdir)
557          print(f"generating mmgbsa input files for {workdir}...")
558
559          #generate local .sh file to run gromacs on the generated mdp files
560          local_run_writer(workdir, basename)
561      return 0
```

Phase 2

Analytical pipeline: implementation (5)

```
36  ##########################
37  # 2. run simulation
38  ##########################
39
40  #deactivate gmxMMPBSA conda env. source the gromacs
41  #this is due to gmxMMPBSA was not complied with CUDA.
42  conda deactivate
43  source /usr/local/gromacs/bin/GMXRC
44
45  cd ./mmgbsa/ac
46  chmod +x run_local.sh
47  bash run_local.sh >> ../../log.txt &&
48
49  cd ../ab
50  chmod +x run_local.sh
51  bash run_local.sh >> ../../log.txt &&
52
53  cd ../../
```

- Initialize the correct environment /path
- Navigate to each directory and consecutively commence the MD simulations:
  - Minimization
  - NVT
  - NPT
  - Production

Phase 3

Analytical pipeline: implementation (6)

```
55  ##########################
56  # 3. post-MD analysis
57  ##########################
58
59  python mmgbsa_decomp_result_analysis.py >> log.txt
60
61  #result saved in fig/
62
63  #meta-Dynamics
64  conda deactivate
65  conda activate biophys_env
66
67  python metaD_sim.py >> log.txt
68
69  #result saved in metaD/
```

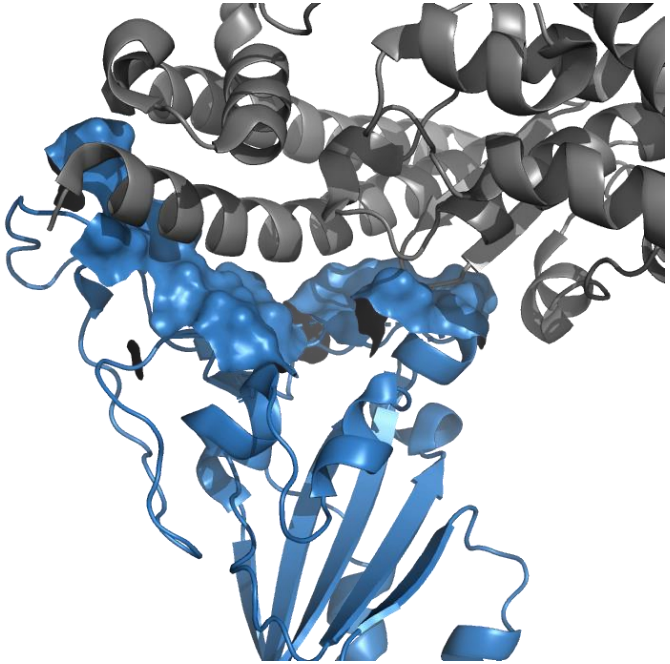
- Initialize the correct environment/path
- Perform mmgbsa per-residue decomposition
- Automated analysis for mmgbsa result.
- Given the residue index, perform the 1D distance based metadynamics.

Analytical pipeline summary

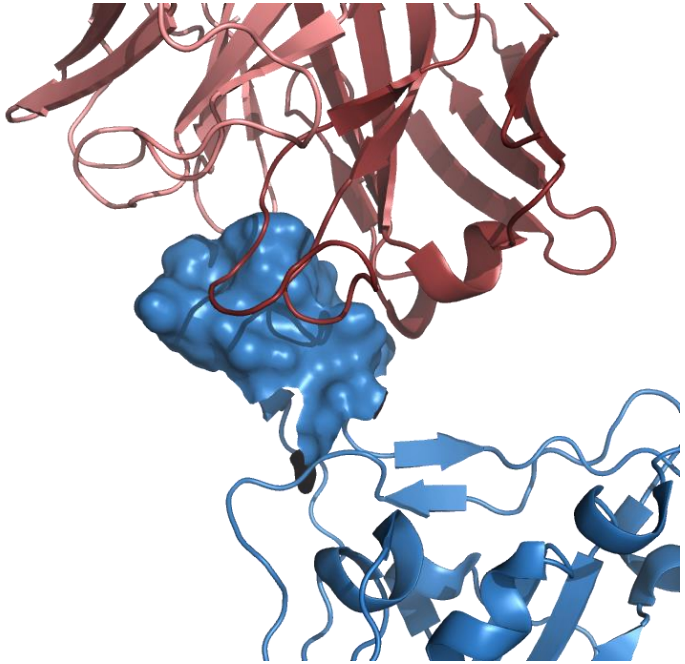
- Advantage:
  - Full automation
  - Good agreement with literature
  - Direct evidence which residue contributes the binding.
  - Provided insight into protein (antibody) redesign.
- Disadvantage:
  - Messy starting structure (Ion, water in pocket kept)
  - No protonation check (H++ server down)
  - Residue index hard coded to fix lost chainid info AMBER/GROMCAS
  - CV space too large and meta-dynamics very expensive.

Interface review:

- Key binding res on pro_a to pro_c (left):
  - [417, 446, 449, 453, 455, 456, 475, 486, 487, 489, 493, 496, 498, 500, 501, 502, 505]
- Key binding res on pro_a to pro_b (right):
  - [484, 478, 486, 487]



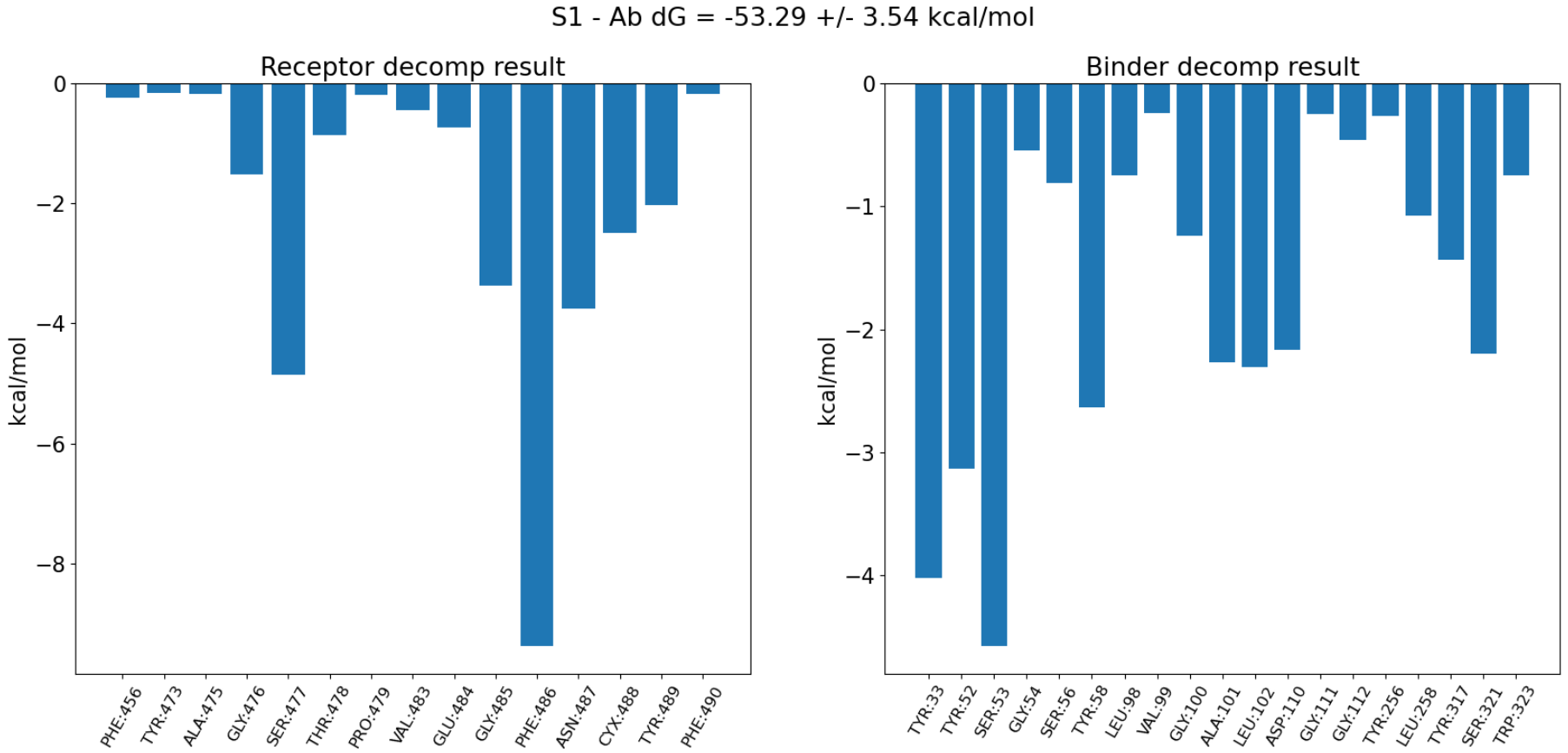
Binding of S1 spike protein(blue) with ACE2(grey). contacting residue on S1 protein are shown in blue surface.



Binding of S1 spike protein(blue) with antibody THSC20.HVTR26 (red). contacting residue on S1 protein are shown in blue surface.

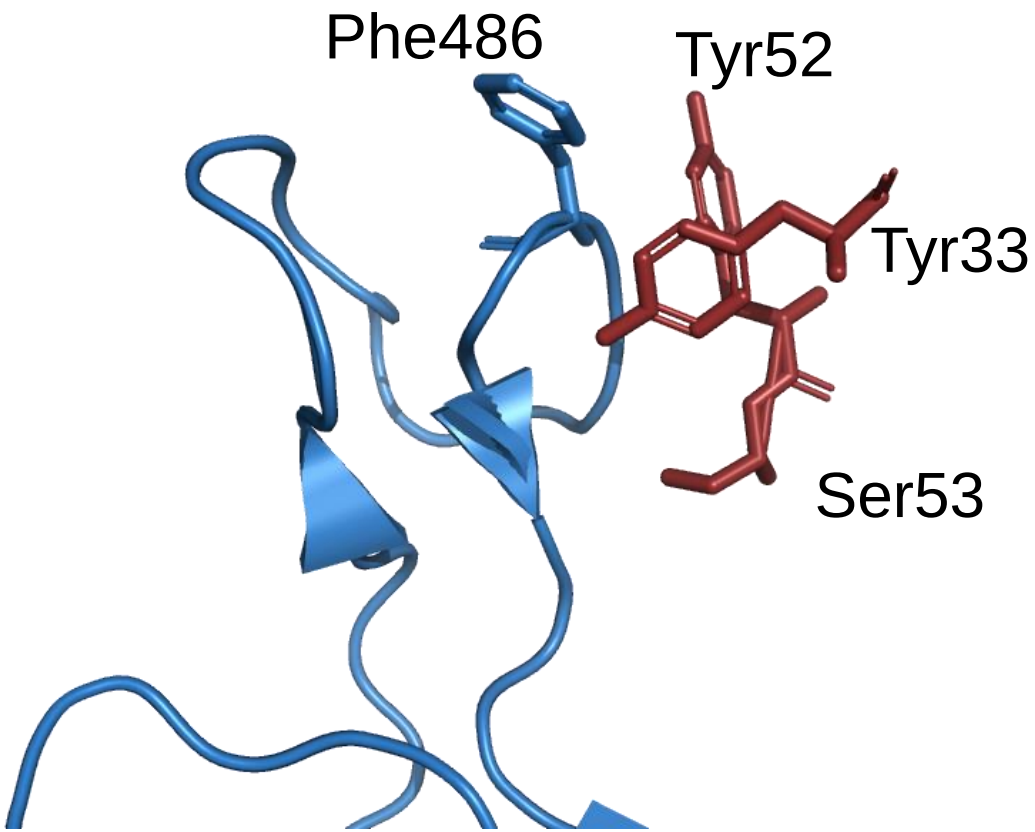
https://www.nature.com/articles/s41586-020-2180-5

MMGBSA per-residue decomposition S1-Ab



Suggest the residue 486 on S1 protein contribute the most (~-9 kcal/mol), while the Ab has several key residues Tyr 33, Tyr52 and Ser53 etc..

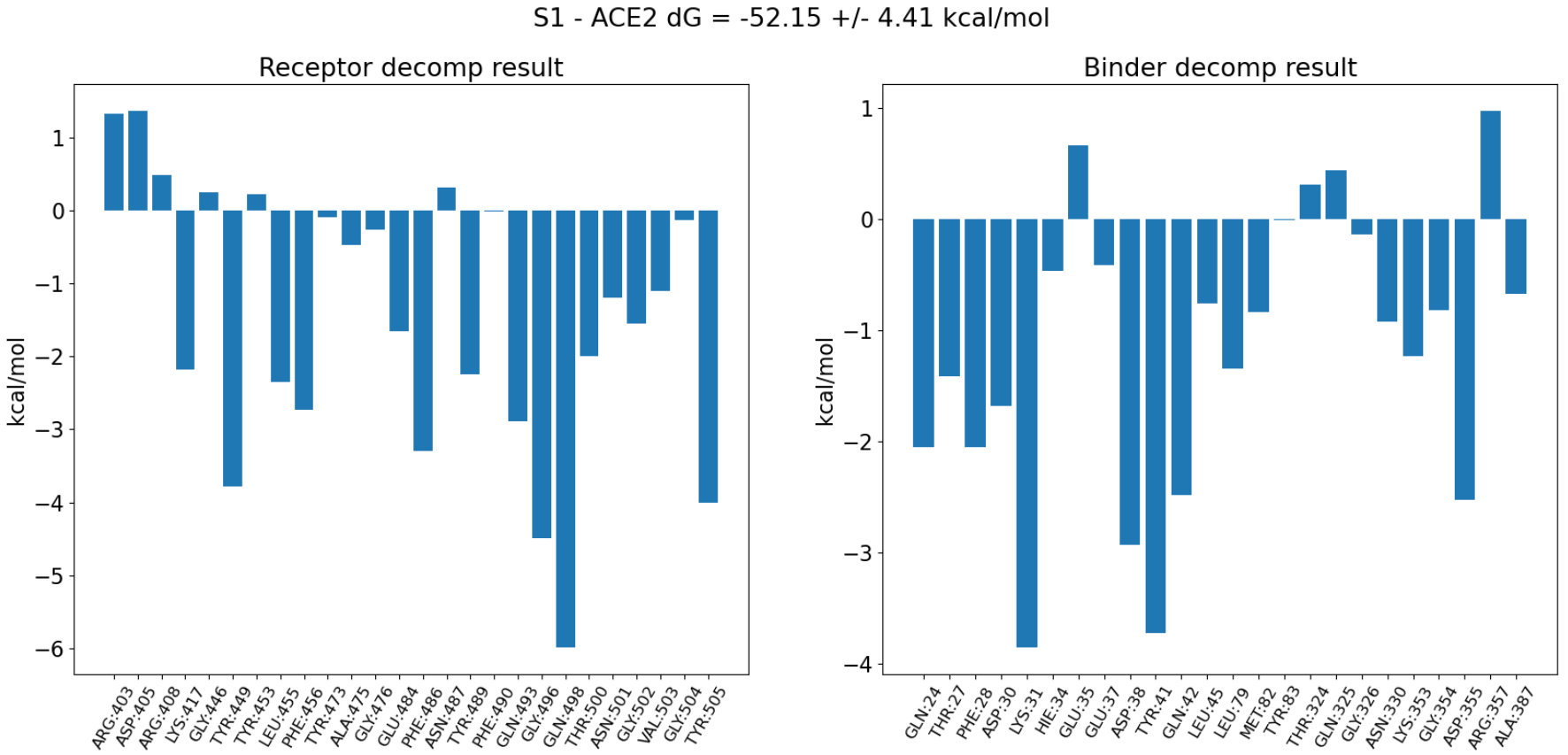
MMGBSA per-residue decomposition S1-Ab visualization



Good agreement with the literature.

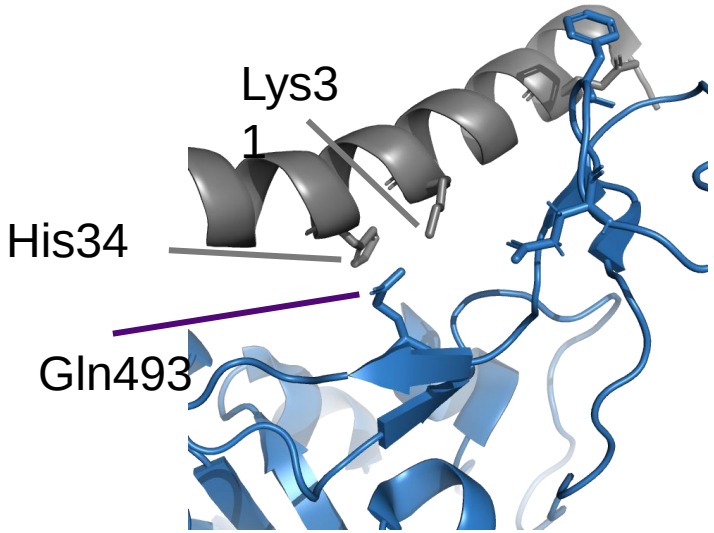
Close contact:
between the Phe486 on S1 protein (blue)
with Tyr52, Tyr33 and Ser53 on Antibody P5A3C8 (THSC20.HVTR26) (red)

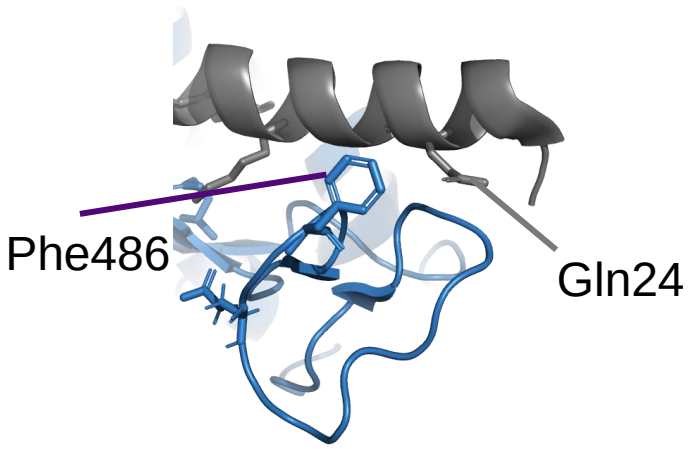
MMGBSA per-residue decomposition S1-ACE2

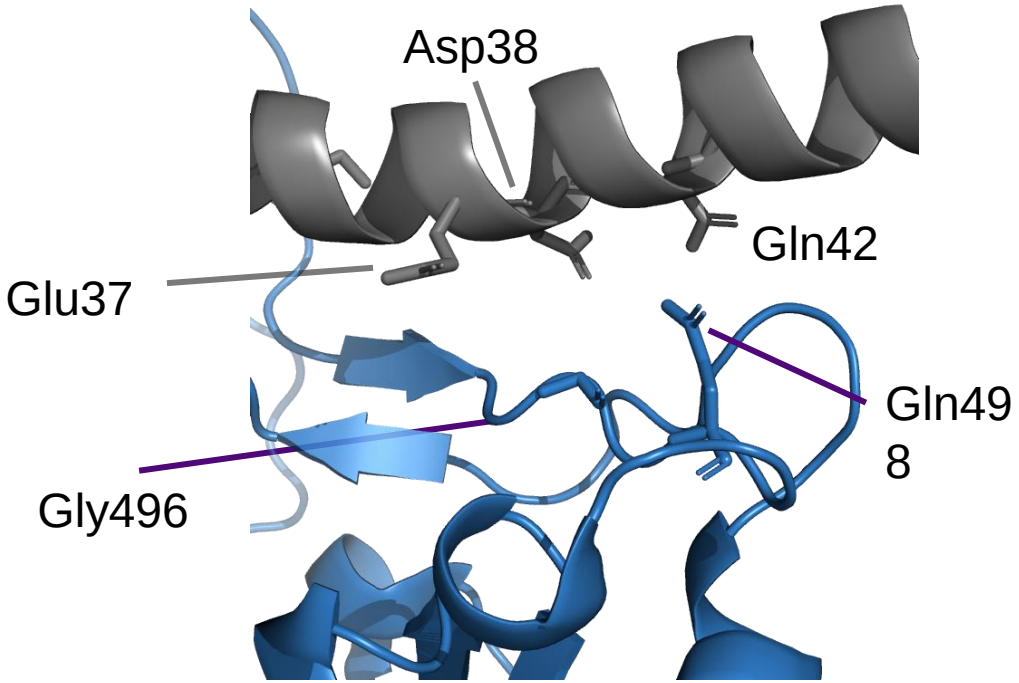


Phe486 contribute ~-3.4 kcal/mol, less than S1-Ab scenario. While the Gln498 contributes the most ~-5.9 kcal/mol. On the other hand, ACE has 3 residues contribute the most: Lys31 (~-3.8 kcal/mol), Asp41 (~-2.9 kcal/mol) Tyr38 (~-3.7 kcal/mol).

MMGBSA per-residue decomposition S1-ACE2: visualization







Meta-Dynamics and other Enhanced Sampling techniques

- Thus in this case we will use distance of Phe487 to Tyr52 as our CV.
  - $\xi = \vec{r}(N487, Y52)$
- For detailed introduction please see:
  - http://docs.openmm.org/7.5.0/api-python/generated/simtk.openmm.app.metadynamics.Metadynamics.html
  - https://doi.org/10.1103/PhysRevLett.100.020603

- Due to the large system size the calculation is not done, but a run-able script has been given out at:
  - metaD_sim.py

A bit more on ES…

I've been working on a MSM optimized biasing technique that sample the target state at least magnitudes faster than meta-dynamics…

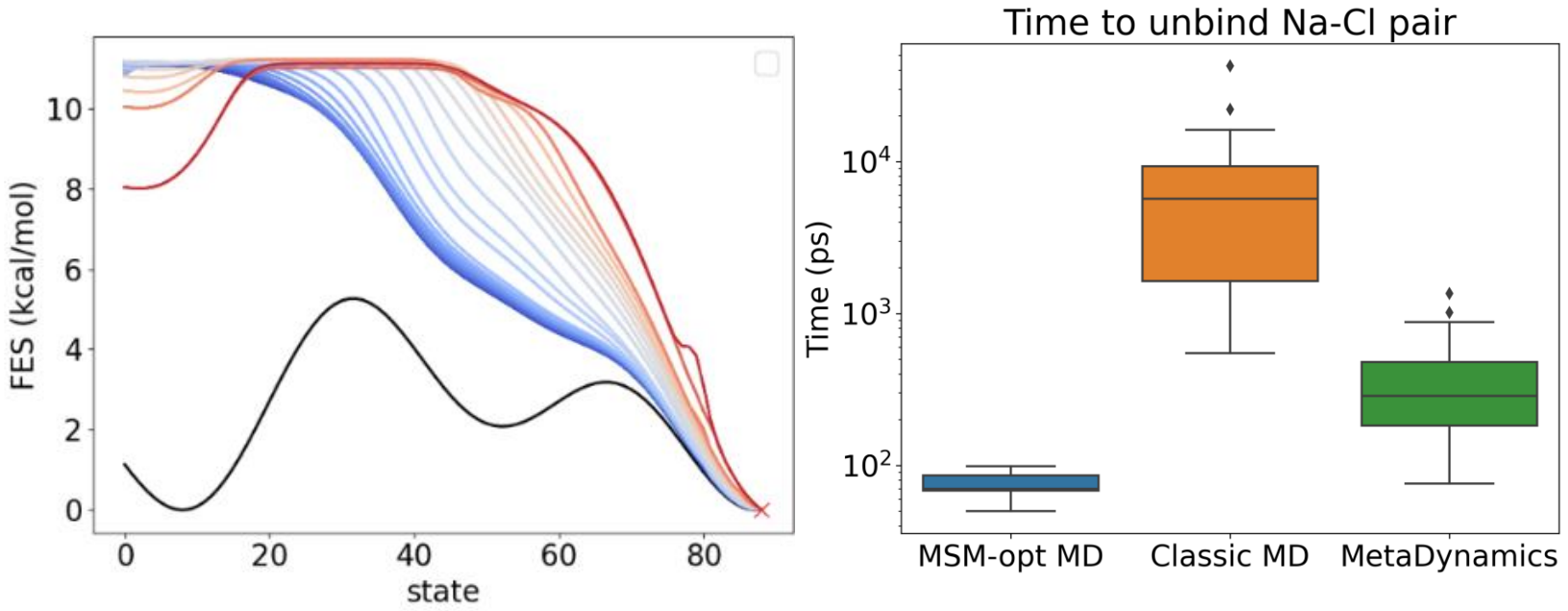


Figure 1 Left: Illustration of adaptive MSM guided bias generation, color lines (blue to red): the biased free energy surface, black solid line: the original FES. Right: preliminary benchmark result on time needed for Na^+-Cl^- ion pair separation task.