

**Virtual Screening and Molecular Dynamics
Study of Approved Drugs Targeting RRM I
of the ALS-related Protein TDP-43**

**Gao Xufan
A0228869H**

**Undergraduate Research Opportunities in Science
PROJECT REPORT**

submitted to the

**Department of Biological Science
National University of Singapore**

LSM3288 (4 MC)

April, 2021

No. of words: 2957

Abstract: Amyotrophic lateral sclerosis (ALS) is a fatal motor neuron disease (MND) that causes muscle atrophy all over the body, difficulty in swallow and finally respiratory failure. TAR-DNA binding protein (TDP-43) is the central component in ALS-related proteinaceous inclusions. Its RNA-recognition motifs (RRMs) is found to be strongly stabilized by ATP, which prevents disease-causing aggregation/fibrillation of TDP-43. It's possible to search for ATP-alternative ligands as potential anti-ALS drugs to avoid interfering ATP-involved biochemical pathways. In this study, TDP-43 RRM1 is selected as the protein target. About 2500 approved drugs were put into a virtual screening procedure and filtered with binding energy and drug-like properties. Thirteen of them were selected for two rounds of molecular dynamics simulation to further validate the binding properties. The final candidates are: 1) azilsartan, which can strongly stabilize the target by decreasing its overall root mean square fluctuation; 2) indacaterol 8-O-Glucuronide, which also stabilize the target by decreasing its root mean square deviation and forming hydrogen bonds with it; 3) midostaurin induces special structural change in the target as well as stabilizing it. These compounds could be put forward into experimental validation to determine their binding properties with TDP-43 and explore their potential to be anti-ALS drugs.

Wordcount: 200

1 Introduction

Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's disease, was first described by Jean-Martin Charcot in 1874 (Rowland, 2001). ALS often starts locally but diffuses to different regions, where cumulative denervation of respiratory and neuronal systems typically limits survival to 3~5 years after symptom onset (Taylor et al., 2016). Given the fatal property of ALS, understanding its pathogenesis has emerged as a new theme for the last two decades.

Although the fundamental pathophysiological mechanisms underlying ALS have not been fully elucidated, the formation of plaque-like proteinaceous inclusions in motor neurons is the pathological hallmark of the disease (Hofmann et al., 2019). Further examinations indicated that TAR-DNA binding protein (TDP-43) is the main constituent of the inclusions. The aggregation/fibrillation of TDP-43 resulted in ~97% of patients with ALS (Guerrero et al., 2016; Mackenzie and Rademakers, 2008). Therefore, inhibiting the toxic aggregation/fibrillation of TDP-43 seems central to the therapeutics of ALS.

TDP-43 belongs to the heterogeneous nuclear ribonucleoproteins (hnRNP) family. It is encoded by TARDBP gene located on chromosome 1 and well conserved among *C.elegans*, *Drosophila*, mouse and human. TDP-43 is a 414-aa protein, which consists of an N-terminal domain (1~102) and a C-terminal domain (267~414) separated by two

RNA-recognition motifs (RRMs: RRM1 and RRM2). It is extensively reported that the two RRMs can function in ALS/FTD-causing aggregation/fibrillation (Zacco et al., 2018; Agrawal et al., 2019; Prakash et al., 2019). Therefore, finding small molecules stabilizing the RRMs may be an effective way to resist TDP-43 proteinopathy.

In our previous work, we discovered that, with the addition of ATP, the thermodynamic stability of TDP-43 RRMs was significantly increased from 49 to 53 degrees, and correspondingly, the amyloid formation kinetics was also largely reduced by monitoring with thioflavin T fluorescence (ThT) assay (Dang et al., 2021a). Based on the NMR study, it was revealed that ATP can specifically bind RRMs. The binding pocket comprising β -hairpins ($\beta3'/\beta3''$) as well as the linker region is partially overlapped with that used for binding nucleic acids. However, considering the multifunctional and degradation-prone properties of ATP in vivo, it is unlikely to directly deliver ATP molecules as a drug to target only TDP-43 protein but not affect other pathways. Therefore, finding/designing other ATP alternatives represents an initial but essential step for further development of therapeutic drugs for ALS/FTD treatment.

In this study, given the ATP binding affinity of tethered RRM1 is much stronger than that of tethered RRM2 (RRM1: 2.58 ± 0.31 vs. RRM2: 13.85 ± 0.86) (Dang et al., 2021a), we chose TDP-43 RRM1 (PDB: 2cqg) as the receptor for in silico virtual screening (VS) study. A 2500 compound-library was then chosen to be docked onto the NMR-derived ATP-binding pocket (Fig. 1). A total of 13 compounds were selected based on the energy scores and drug-like properties (Tab. 2). Next, with the combination of 10-ns molecular dynamic simulations, 5 out of 13 compounds were further filtered on the basis of binding energy. We then extended the simulation time to 50 ns to further investigate the dynamic properties of the 5 compounds. By systematically analyzing the RMSD, RMSF, etc, we found that compounds A and B are more promising due to the prominent influence on reducing the overall RMSD through changing C-terminal secondary structure.

To summarize, the current research provides the first systematic computational study for the TDP-43 RRM domain. Here we highlight that azilsartan and indacaterol 8-O-Glucuronide are potential ATP alternatives for antagonizing TDP-43 proteinopathy. Mi-

dostaurin also shows specifical binding modes. Further experimental studies should be included to confirm the effects of these molecules.

2 Methods

2.1 Establishment of small molecule library

Structures of small molecules were obtained from the "Approved" subset of Drugbank (Wishart et al., 2006) and "world-not-fda" subset of Zinc (Sterling and Irwin, 2015), which contain most of the approved drugs worldwide. Openbabel (O'Boyle et al., 2011) is a toolbox that generated 3D structure, performed energy minimization, added gasteiger charge and converted them to .pdbqt files. They carry no charge and only polar hydrogens were kept in the structure. Later, their drug-like properties were predicted with Python package RDkit (Landrum, 2016).

The pre-experiment showed that ligands with low molecular weight (MW) did not give good binding affinity; on the contrary, those with high MW won't show good drug-like properties. Then only ligands with MW between 300 and 650 Da were kept. There are 840 and 1661 ligands left in the two subsets.

2.2 Virtual screening

The structure of the target is retrieved from PDB (id: 2cqg). The first conformer (residues 102~185) is selected as the receptor. With AutoDockTools (Morris et al., 2009) in MGLTools 1.5.7, water deletion, polar hydrogen and Kollman charge addition as well as structural repair are conducted. According to (Lukavsky et al., 2013; François-Moutal et al., 2019; Dang et al., 2021b), residues that strongly contact with nucleic acids and might be affected upon binding to ATP are recognized as the potential active residues in binding. They related residues are labeled (Fig. 1b), and a grid box that restrains ligands' position is generated accordingly (Fig. 1a). The parameters inputted into AutoDock Vina (Oleg and Olson, 2012) is listed in Tab. 1.

Tab. 1: Parameters in Virtual Screening

term	value	term	value
center_x ¹	12	size_x	14
center_y	-7.5	size_y	20
center_z	0.5	size_z	16
exhaustiveness	24	number of modes ²	9

¹ Units of coordinates are Å.

² The number of conformers generated. Other settings are default.

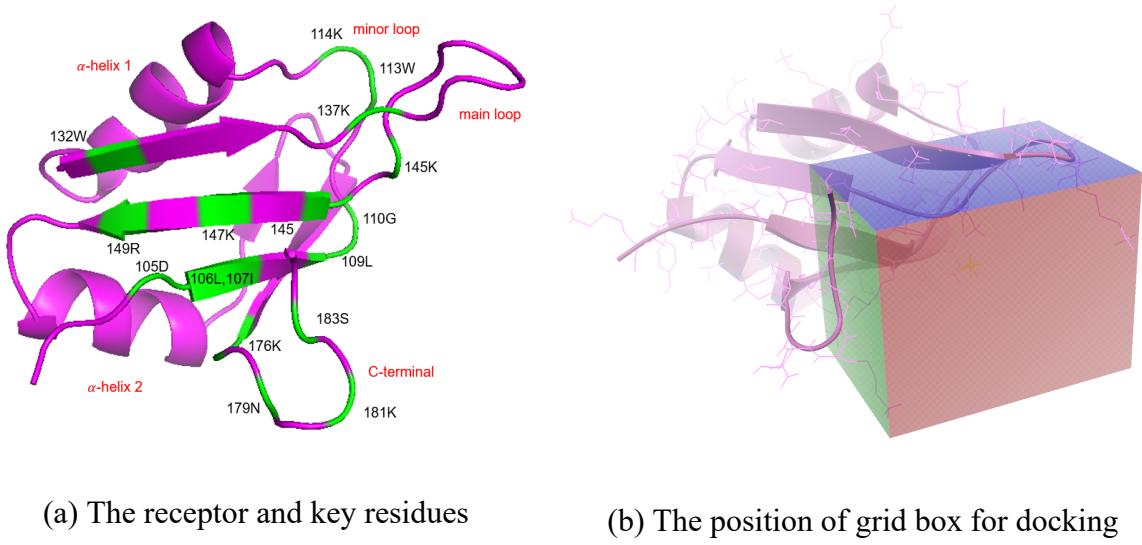


Fig. 1: The receptor structure and grid box

2.3 Molecular dynamics simulation

MD simulations were performed to further explore the structure of protein-ligand complexes. Protonation states of all ligands under pH 7.0 were determined and energy minimization performed in OpenBabel. RESP charge was calculated in Gaussian 16 (Frisch et al., 2016). AmberTools 2020 (Case et al., 2020) helped with combining the protein and ligand, and setting the periodic cubic box. The box was then filled with water molecules using TIP3P water model. Na^+ and Cl^- ions were added to it to neutralize the system and make ionic strength 0.15mM, whose quantities were calculated by SLTCAP (Schmit et al., 2018). AMBER ff99SB-ILDN (Lindorff-Larsen et al., 2010) and GAFF force field (Wang et al., 2004) were applied to the protein and ligands, respectively.

Then .top, .gro and .ndx files were generated and Gromacs 2021 (Lindahl et al., 2021) performed the following steps. The whole system was subjected to 5000 steps of conju-

gate gradient energy minimization until the maximum force is under 200. Then 50000 2-femtosecond steps of NVT and NPT equilibration were performed in turn. Finally, production simulations were set up. All bonds are restrained by LINCS algorithm (Hess et al., 1997). Modeled with the particle-mesh Ewald method, the electrostatic cut-off was always 10Å, being the same as van der Waals cut-off. The reference temperature and pressure were 300.0 K and 1.0 bar, respectively.

In the first-round run, each complex is simulated for 10 ns at least three times independently. In the second-round run, the selected ligands are simulated for 50 ns. Except for the 13 complexes, we also set two control groups: the protein only (free) and protein with ATP⁴⁻.

All the average properties were based on the trajectory excluding the first few nanoseconds when the system might not reach equilibrium. For 10 ns simulations, the first 2 ns is excluded and 20 ns for 50 ns simulations. The properties like RMSD were extracted from the trajectory and calculated also in Gromacs. The trajectories of secondary structure were extracted by DSSP package in Biopython (Cock et al., 2009). MMGBSA Free energy calculations were done using package gmx_MMPSA (Tresanco et al., 2021). The molecular structures were visualized by PyMOL (Schrödinger, LLC, 2015) and Discovery Studio 2021 (Sysèmes, 2021). Ramachandran plot quality analysis is perfromed using MolProbity server (Williams et al., 2018).

3 Results and Discussion

3.1 Virtual screening and drug-like properties

All the results of selected ligands are shown in Tab. 2, and their molecular structure is shown in Fig. 2. More detailed interpretation of data could be found in Appendix A. The binding energy of ATP is shown in Tab. 3. Different databases provide different initial structures thus gives various but similar results.

Defined as the Gibbs free energy change of protein-ligand binding reaction, the binding energy is a commonly-used indicator of how strong their combining is. Comparing

with ATP, all selected ligands show much lower binding energy, manifesting potential binding on the target. Their mean binding energy are about 2 kcal/mol lower than that of ATP, which means over 25 times lower in K_d value.

To guarantee good absorbance and utilization of the drug in the human body, Lipinski's rules of five (Lipinski, 2004) and two additional rules were applied to filter the compounds ("criteria" in Tab. 2). The selected ligands generally have good drug-like properties. The violations concentrate on molecular weights, the majority of which is slightly higher than 500 Da. As a result, those which violates other requirements are excluded. Some of the selected ligands have too many aromatic rings, but another useful criteria, fraction of sp^3 carbon, is not applied, since many of the energy-favored ligands had been popped out.

Tab. 2: Properties of the selected ligands in virtual screening

ligand name	rank ¹	energy top (kcal/mol)	energy mean (kcal/mol)	molecular weight (Da)	number of H donor	number of H acceptor	$\log P$	number of rotatable bonds	Lipinski violations	number of aromatic rings
criteria		<-6.9		<500	<5 (≥ 1)	<10	>-2 and <5	≤ 10	≤ 2	≥ 1
ponatinib	d1	-7.6	-6.83	532.57	1	6	4.46	4	1	4
nilotinib	d2	-7.6	-7.43	529.53	2	7	6.36	6	2	5
dihydro-α-ergocryptine	d3	-7.5	-6.76	577.73	3	6	2.52	5	1	2
lumacaftor	d4	-7.4	-7.06	452.41	2	5	4.75	5	0	3
midostaurin	d5	-7.2	-6.89	570.65	1	6	5.91	3	2	6
azilsartan	d6	-7.1	-6.71	568.54	1	11	4.71	9	2	6
capmatinib	d7	-7.1	-6.66	412.43	1	6	3.43	4	0	5
imatinib	d8	-7.0	-6.71	493.62	2	7	4.59	7	0	4
ubrogepant	d9	-6.9	-6.61	549.55	2	5	3.53	4	1	3
dihydroergotoxine	z1	-7.3	-7.07	563.70	3	6	2.13	4	1	2
indacaterol 8-O-Glucuronide	z2	-7.2	-6.93	568.62	7	9	0.71	9	2	3
lurasidone	z3	-7.2	-6.70	508.69	1	7	3.23	5	1	2
brigitanib	do ²	-7.3	-6.44	584.11	2	9	5.09	8	2	3

¹ "d"/"z" means this ligand is from DrugBank/Zinc database respectively.

² this one was also included, because it's a TKI and showed relatively good binding energy in pre-experiment; but the data shown here is still those in the pre-experiment.

Tab. 3: Docking results for ATP

database	drugbank	pubchem	superdrug2	zinc	mean
energy top (kcal/mol)	-4.90	-5.00	-4.30	-5.10	-4.83
energy mean (kcal/mol)	-4.72	-4.83	-4.16	-4.94	-4.66

Among them, a family of compounds is attractive: tyrosine kinase inhibitor (TKI). They often competitively bind on the ATP binding pocket of tyrosine kinase, which infers that they may replace ATP as a stronger stabilizer of TDP-43. Their physiological function is usually associated with affecting signal transduction. Therefore, many TKIs have

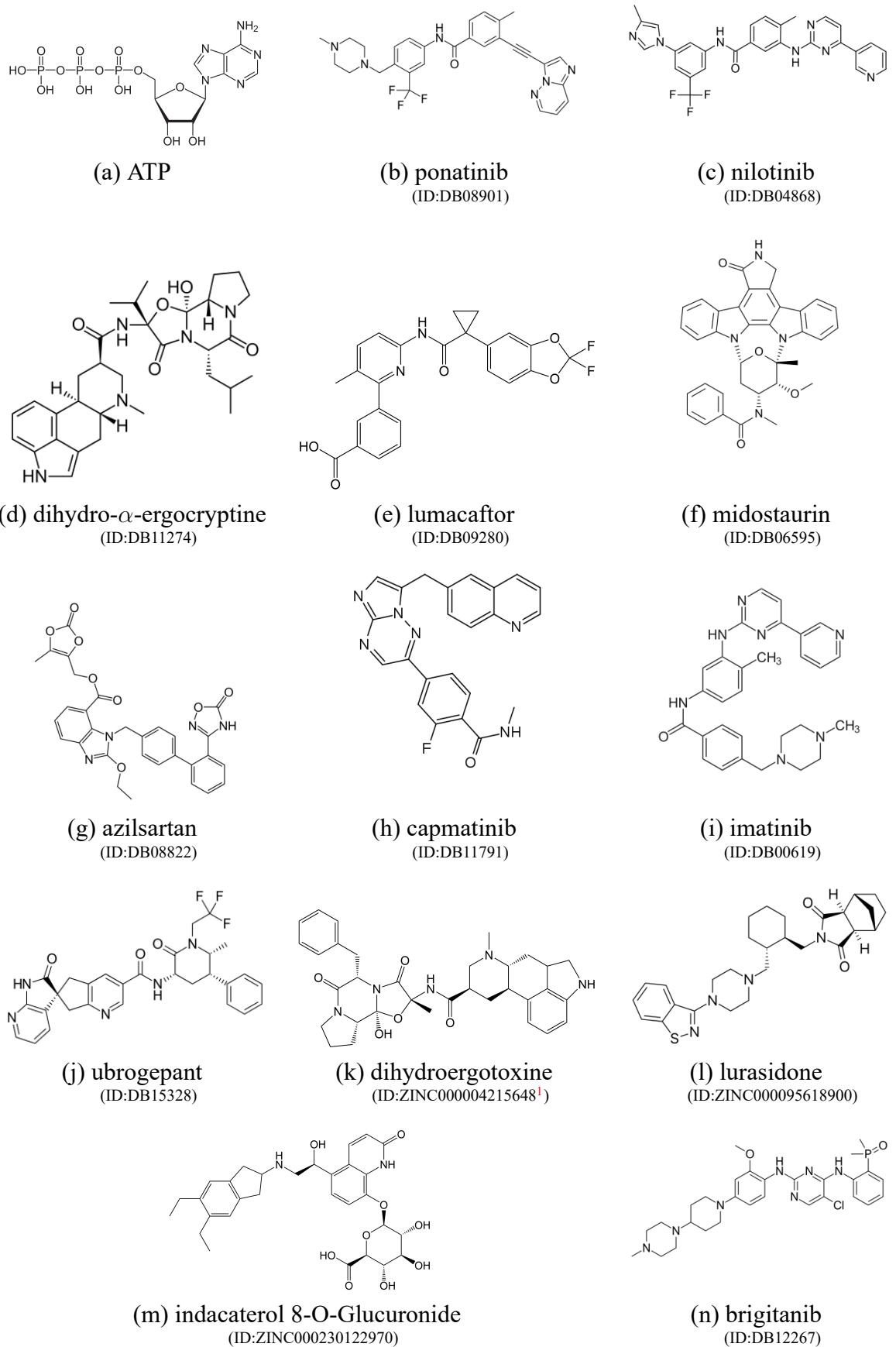


Fig. 2: Structure of selected ligands after virtual screening

¹ also DB12273

been developed into drugs for cancer treatment. In Fig. 2, imatinib (i), nilotinib (c) and ponatinib (b) are three generations of the chronic myelogenous leukemia (CML) targeting drug. They take the top two spots of the list. Except for tinib drugs, midostaurin (f) is also a good multitarget kinase inhibitor.

There's also some other types of ligands that are selected. A family of alkaloids, named ergoloid, also rank high. They look quite similar, which include d3, z1 and z4 (not selected). Lumacaftor is a chaperone which helps increases the number of CFTR. It has low MW and no lipinski violations. Azilsartan medoxomil is an ACE II receptor inhibitor to treat hypertension. Indacaterol 8-O-Glucuronide is a metabolite of indacaterol, which is a β -adrenoreceptor agonist for the treatment of asthma and chronic obstructive pulmonary disease.

Because virtual screening does not give a binding structure as accurate as MD simulation, the screening protocol only acts as a time-efficient filter to pick the potential ligands out of the big molecule library.

3.2 Molecular dynamics simulation

Tab. 4: Abbreviations and expressions used in MD results

term	full name
RMSD	root mean square fluctuation
RMSF	root mean square deviation
SS	secondary structure
σ	standard deviation
R	correlation coefficient
ΔG	binding energy
SASA	solvent accessible surface area of the protein
R_g	radius of gyration of the protein
Δf_{Ct}	the average of difference of RMSF between ligand and "free" in residue 180~184
Δf_{loop}	the difference of RMSF between ligand and "free" in residue 140

3.2.1 first-round screening

The results of 10-ns MD simulation is shown in Tab. 5 and 8. More explanation of the calculated properties are available in Tab. 4 and Appendix B. Binding energies ΔG

are calculated from three independent repeated simulations. RMSD acts as an important reference and interpretation tool. Lower RMSD than free means structure stabilization, while larger RMSD means induction of some important structural change. Their correlation coefficient (R) is 0.02, which implies no linear relationship. Finally, ligands are ranked merely by average binding energy.

Tab. 5: ΔG and RMSD¹ in 10 ns MD simulation

ligand	ΔG (kcal/mol)				RMSD (Å)			
	run 1	run 2	run 3	mean	run 1	run 2	run 3	mean
indacaterol 8-O-Glucuronide	-33.52±0.64	-29.31±0.46	-33.75±0.45	-32.20	2.76±0.23	1.75±0.22	2.36±0.52	2.29
lumacaftor	-30.04±0.33	-27.04±0.45	-34.42±0.47	-30.50	1.70±0.27	1.38±0.27	1.57±0.22	1.55
midostaurin	-23.89±0.50	-27.55±0.47	-28.29±0.27	-26.58	1.52±0.39	1.44±0.34	1.31±0.23	1.42
azilsartan	-24.73±0.24	-25.39±0.22	-23.50±0.49	-24.54	1.32±0.18	1.39±0.22	1.66±0.22	1.45
ponatinib	-24.73±0.46	-21.73±0.34	-26.55±0.32	-24.34	1.18±0.13	1.36±0.23	1.28±0.25	1.27
lurasidone	-24.99±0.57	-22.41±0.50	-22.68±0.42	-23.36	1.54±0.22	2.12±0.70	1.39±0.19	1.69
dihydroergotoxine	-24.61±0.35	-23.78±0.30	-14.96±0.42	-21.12	2.01±0.86	1.17±0.23	1.21±0.20	1.46
nilotinib	-18.45±0.40	-15.66±0.55	-20.08±0.25	-18.07	2.19±0.42	1.80±0.28	1.47±0.19	1.82
dihydro- α -ergocryptine	-16.54±0.20	-18.90±0.30	-17.60±0.30	-17.68	1.21±0.24	1.48±0.18	1.42±0.36	1.37
capmatinib	-25.42±0.50	-17.89±0.38	-5.43±0.46 ²	-16.25	2.04±0.27	1.40±0.34	1.93±0.59	1.79
brigitanib	-15.28±0.24	-15.98±0.36	-16.85±0.33	-16.04	1.22±0.15	1.51±0.58	1.22±0.17	1.31
ubrogepant	-9.35±0.42	-15.64±0.38	-20.96±0.26	-15.32	2.14±0.29	1.83±0.17	2.03±0.52	2.00
ATP	-14.43±0.47	-5.29±0.56 ³	-17.88±0.47	-12.53	2.95±0.39	4.19±0.35	2.87±0.30	3.33
imatinib	-13.64±0.48	-3.44±0.36 ⁴	-13.20±0.33	-10.10	1.19±0.15	1.41±0.38	1.43±0.16	1.34
free	—	—	—	—	1.57±0.25	1.84±0.46	1.79±0.35	1.73

¹ Data format: ΔG : mean±error of mean; RMSD: mean±standard deviation

² In this run, the ligand went out of the binding pocket. So do note ³ and ⁴.

The top ligands generally show reasonable structures. Ponatinib (a) makes the lowest RMSD, but hardly changes other properties. It is interacting with 147F, 149F, 109L with its aromatic rings and CF_3 group. There is also special binding modes. Some of them just try to "gather" on the 147F, like azilsartan (b), nilotinib and ubrogepant. Azilsartan usually "gather" but shows low binding energy. In one trajectory, lumacaftor crosses the C-terminal and the β -sheet surface, which is shown in (c) and exhibits the lowest binding energy among all 10-ns runs. The third type of ligands may bind stably on the desired pocket in 3 runs, but neither make a difference in protein's structure nor show a good binding energy. An example is brigitanib (d). As for dihydro- α -ergocryptine, low RMSD is accompanied by unimpressive ΔG . There's also a few ligands which escape from the receptor: imatinib (2 out of 5 runs), capmatinib and ATP (2 out of 6 runs), which makes RMSD and ΔG extremely high (yellow curve in Fig. (a)). Even when they stay in the pocket, binding energies are still high.

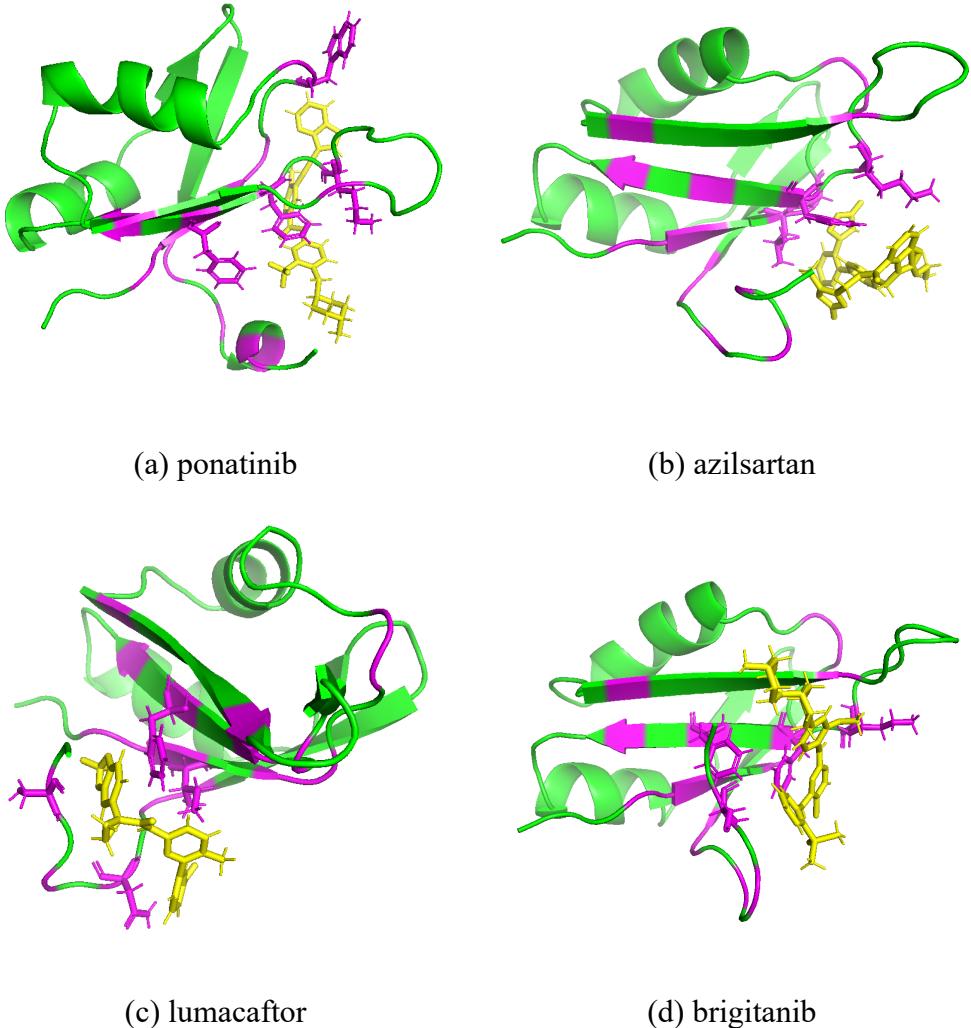


Fig. 3: Structure of complexes (10ns)

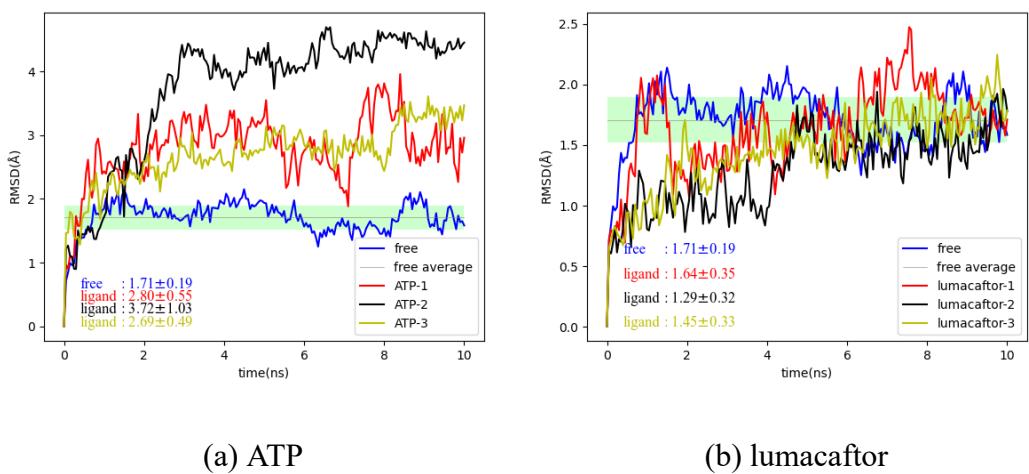


Fig. 4: RMSD-time plot

However, it is clear that some of the trajectories suggest that the system has not reached equilibrium. There is still big fluctuation in RMSD, some of which are even slowly increasing (in Fig. (b)). Thus, longer-time simulation is necessary for more accurate investigation of structure and properties. The top five ligands were selected for the 50-ns simulation. The initial structure comes from a 10-ns trajectory that shows the best binding effect.

3.3 Longer-time simulation

The important properties are summarized in Tab. 6, 7 and 5. Each ligand is separately analyzed. Details of terms can be seen in Tab. 4 and Appendix B, with Fig. 12 shows the secondary structure of "free" protein as a reference.

Tab. 6: General properties in 50 ns MD simulation

ligands	RMSD	RMSD of ligand	RMSF	Δf_{Ct}	Δf_{loop}	SASA	Rg	H bond count
free	1.89±0.33	—	0.92	—	—	60.08±1.61	12.82±0.14	—
ATP	1.91±0.18	1.27±0.18	0.72	1.81	0.49	60.47±1.23	12.90±0.07	3.29±1.16
azilsartan	1.67±0.15	1.77±0.09	0.64	2.29	0.62	58.12±1.13	12.71±0.07	1.94±0.63
lumacaftor	2.21±0.35	2.13±0.40	0.82	0.96	0.20	60.32±1.20	12.82±0.09	2.58±1.32
ponatinib	1.36±0.16	1.73±0.38	0.66	2.36	0.23	58.76±1.14	12.69±0.06	0.57±0.67
midostaurin	2.55±0.21	0.47±0.10	0.72	1.53	0.09	62.82±1.12	13.01±0.07	1.73±0.94
indacaterol 8-O-Glucuronide	1.35±0.22	0.95±0.29	0.81	1.75	-0.32	61.41±1.18	12.89±0.06	3.04±1.31

¹ the unit of RMSD, RMSF, Rg are Å; that of SASA is nm²; that of Δf_{Ct} and Δf_{loop} are kcal/mol.

Tab. 7: Free energy components in 50 ns MD simulation

ligands	ΔG	ΔG_{vdw}	ΔG_{ele}	ΔG_{gas}	ΔG_{polar}	$\Delta G_{nonpolar}$	$\Delta G_{solvation}$	σ_{decomp} ¹
ATP	+32.99±0.68	-22.74	-80.63	-103.38	140.22	-3.85	136.37	1.54
azilsartan	-24.96±0.17	-34.88	-17.93	-52.81	32.23	-4.39	27.85	0.39
lumacaftor	-34.27±0.29	-38.35	-101.78	-140.13	110.64	-4.77	105.86	1.62
ponatinib	-17.41±0.23	-27.78	-1.71	-29.49	15.88	-3.80	12.08	0.51
midostaurin	-41.32±0.21	-49.45	-36.04	-85.50	50.29	-6.11	44.18	2.55
indacaterol 8-O-Glucuronide	-29.82±0.31	-37.17	-135.76	-172.94	148.45	-5.33	143.12	0.92

¹ the standard deviation of ΔG decomposition on every residue.

² all units are kcal/mol; all data regarding energy components are average values.

3.3.1 Indacaterol 8-O-Glucuronide

This ligand (Fig. 6) binds very stably, with fairly lower RMSD (1.35±0.22 kcal/mol) than "free". It mainly stabilizes C-terminal dramatically ($\Delta f_{Ct}=1.75\text{\AA}$), and another end extends to the minor loop. It seems to induce the extension of α -helix 1 since 25ns, and

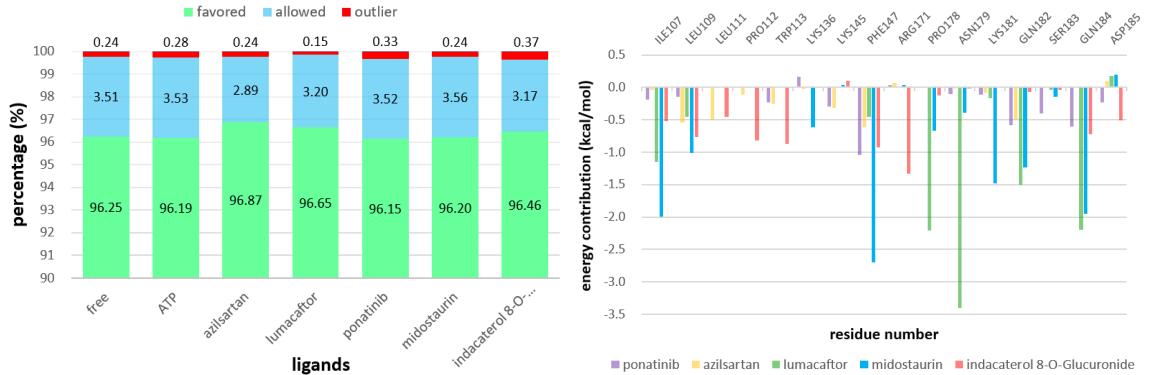
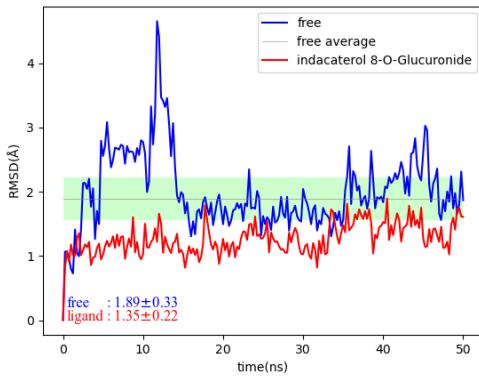


Fig. 5: Micelleous properties in 50-ns MD simulation

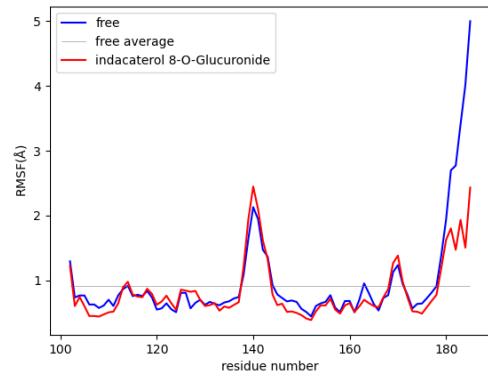
the fraction of favored conformation is promoted. With the hydroxyl-rich ring with a negative-charged carboxyl group, it forms many hydrogen bonds with 113W, 143H, 171R (also electrostatic interaction) near the loops, as well as 184Q. It is probably also this group that makes the electrostatic potential decrease so much (about 135 kcal/mol). Its aromatic rings also interact with 110G and 147F. These two kinds of interactions contribute to most of the ΔG .

3.3.2 Lumacaftor

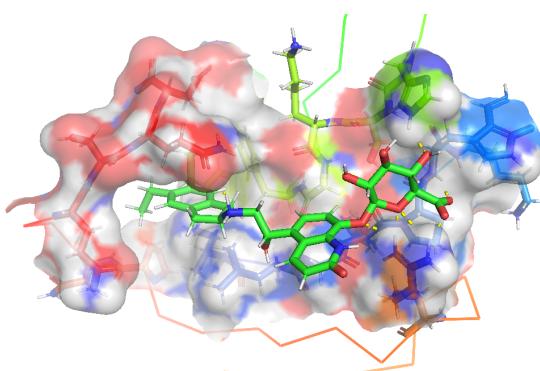
Lumacaftor (Fig. 7) still shows the special "crossing" conformation as mentioned in the 10-ns simulation. It clears much more Ramachandran outliers than other ligands do not, implying a more reasonable 3D structure. Though its binding energy is the second-lowest, no special changes in RMSF, SS, or other properties appear on the protein. The high protein RMSD ($2.21 \pm 0.35 \text{ \AA}$) and σ of RMSD is mainly due to the peak at around 25ns, but later the ligand is not affecting the protein structure. This might be explained by its binding site being far on the C-terminal. It interacts with CTD residues which induce dramatic ΔG decreases (over 2 kcal/mol per residue). The major interaction between them is hydrogen bonding, with van der Waals interaction with 107L additionally.



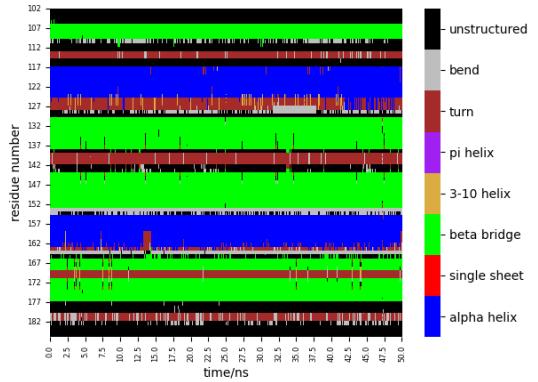
(a) RMSD-time plot



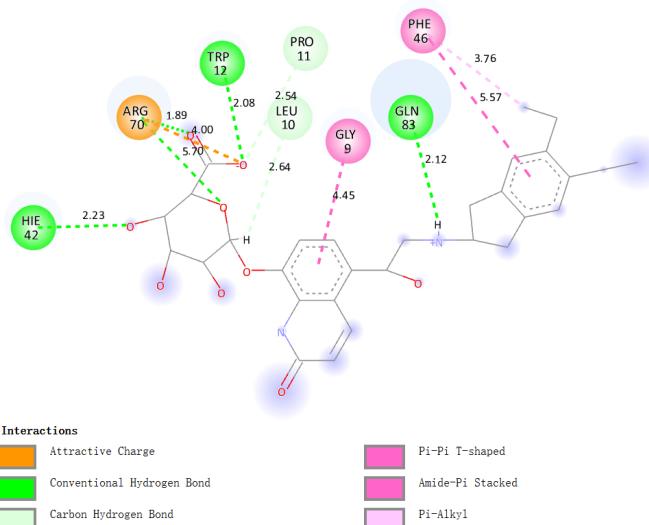
(b) RMSF-residue plot



(c) binding structure



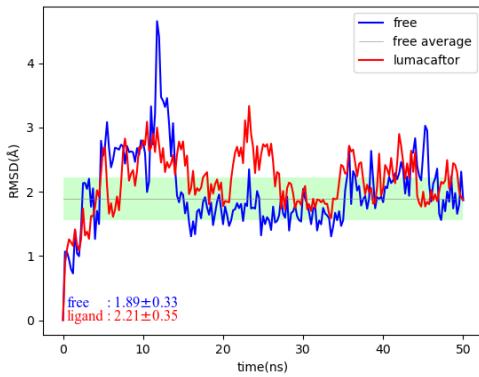
(d) secondary structure



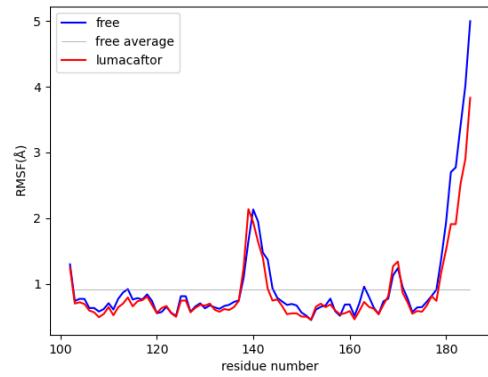
(e) 2d interaction plot¹

Fig. 6: Results of indacaterol 8-O-Glucuronide (50ns MD simulation)

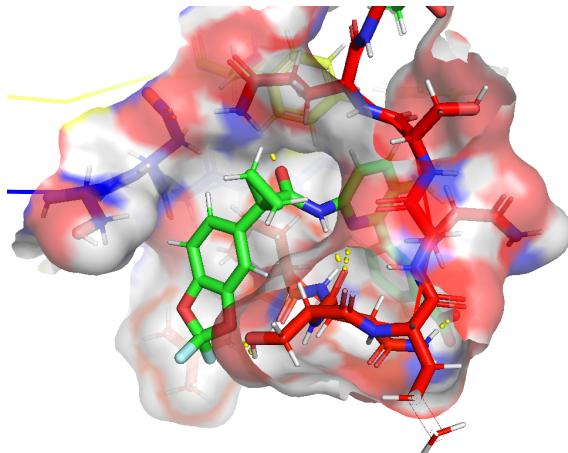
¹ For the renumbering issue, the actual residue number in the plot should plus 101.



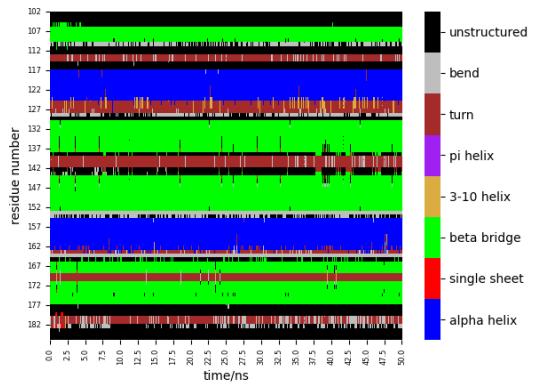
(a) RMSD-time plot



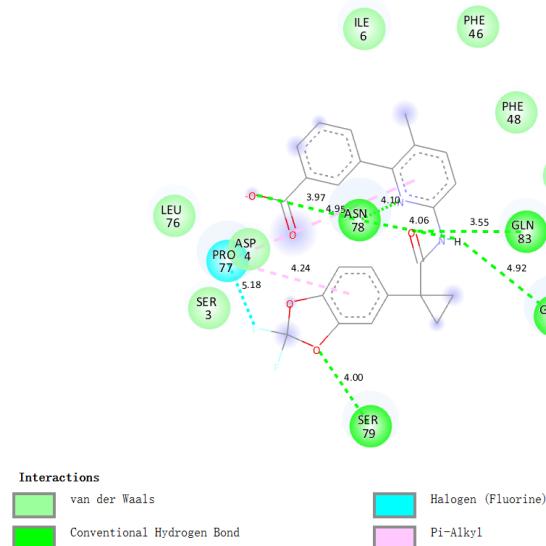
(b) RMSF-residue plot



(c) binding structure



(d) secondary structure



(e) 2d interaction plot

Fig. 7: Results of lumacaftor (50ns MD simulation)

3.3.3 Midostaurin

It shows the lowest ΔG (-41.32 ± 0.21 kcal/mol), where van der Waals contribution becomes the major component (about -50 kcal/mol) because of its numerous aromatic rings. A lot of residues are involved in the hydrophobic contact: 107I, 109L, 147F and 149F (strong $\pi-\pi$ interaction), 176K and 181K, among which 109L and 147F contribute a lot to ΔG . Moreover, it seems to be stuck between the β -sheets and C-terminal, enhancing the contacts and making the ligand's RMSD the lowest (0.47 ± 0.10 Å). However, the properties RMSD (2.55 ± 0.21 Å), SASA (62.82 ± 1.12 nm 2) and Rg (13.01 ± 0.07 Å) of protein are the highest among ligands and change the most from "free", while their σ is not high. Overall RMSF even decreased pretty much. All these indicate that this ligand quickly induced important structural change and stabilizes the changed structure. The protein becomes a little looser and more exposed, but nothing is shown in the SS plot.

3.3.4 Azilsartan

Though the RMSD does not decrease as much as ponatinib and Indacaterol 8-O-Glucuronide, azilsartan induces the sharpest incline in overall RMSF (mean: 0.65 Å), especially those of the main loop and C-terminal ($\Delta f_{Ct} = 2.29$ Å, $\Delta f_{loop} = 0.62$ Å). It makes the lowest average σ of RMSD (0.15 Å), average RMSF (0.64 Å), SASA (0.58 nm 2) and highest fraction of favored conformers (96.87%), exhibiting extraordinary stabilization. The most astonishing effect is that it induces the formation of α -helix in C-terminal from 15ns. Checking the structure, the ligand is always located between the β -sheet, C-terminal and the main loop. The benzimidazole ring has hydrophobic interaction with 145K, 147F and 181K. Residue 109L, 111L and 183Q may contribute to van der Waals interaction. The two five-member rings form several hydrogen bonds, and other aromatic rings are involved in even one benzene is "stacking" with the diester ring. It has the lowest σ_{decomp} (0.39 kcal/mol), which means the interaction between residues is more uniform even for terminal residues. In summary, azilsartan undoubtfully ranks the first in stabilizing the protein structure.

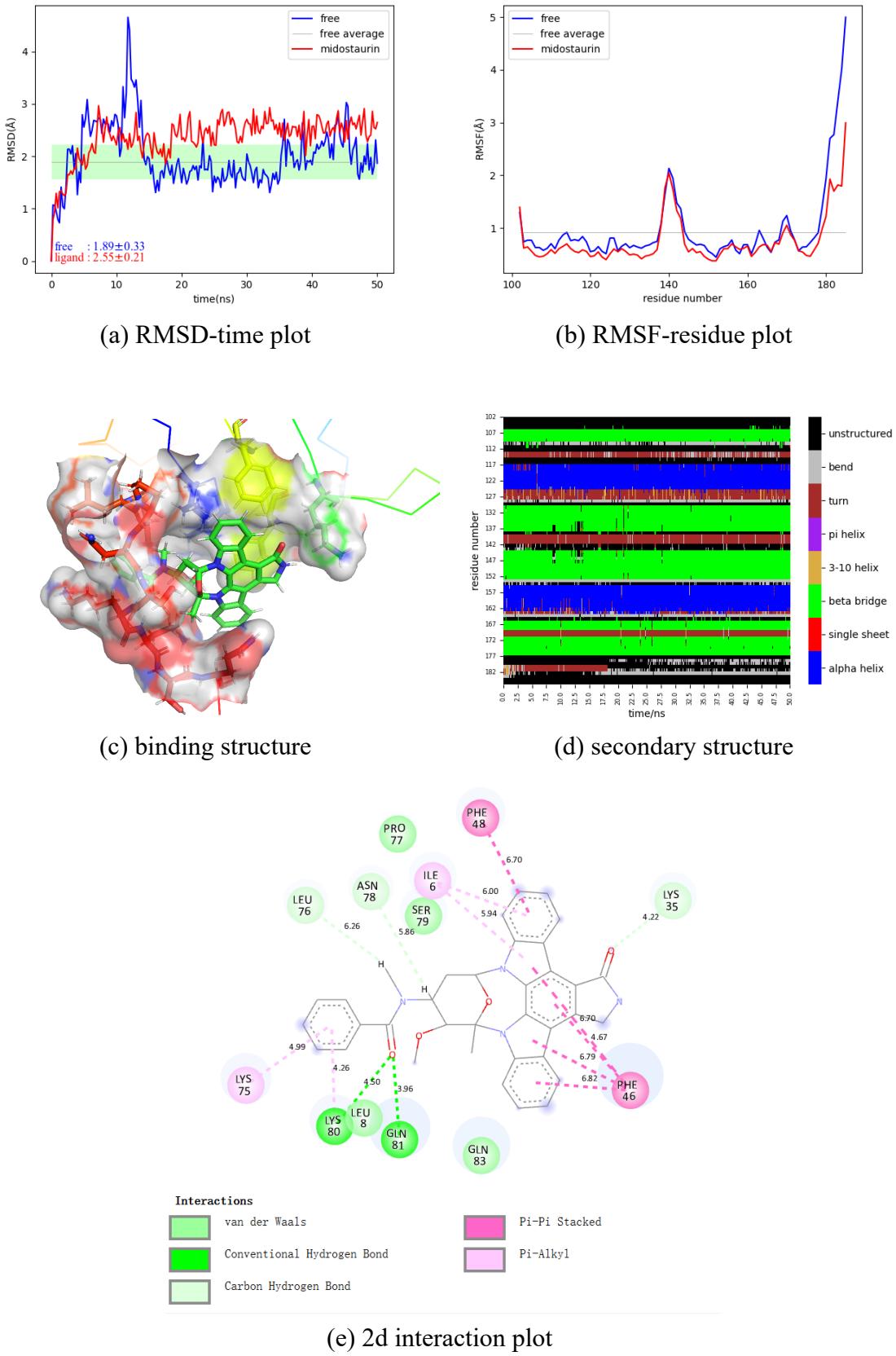


Fig. 8: Results of midostaurin (50ns MD simulation)

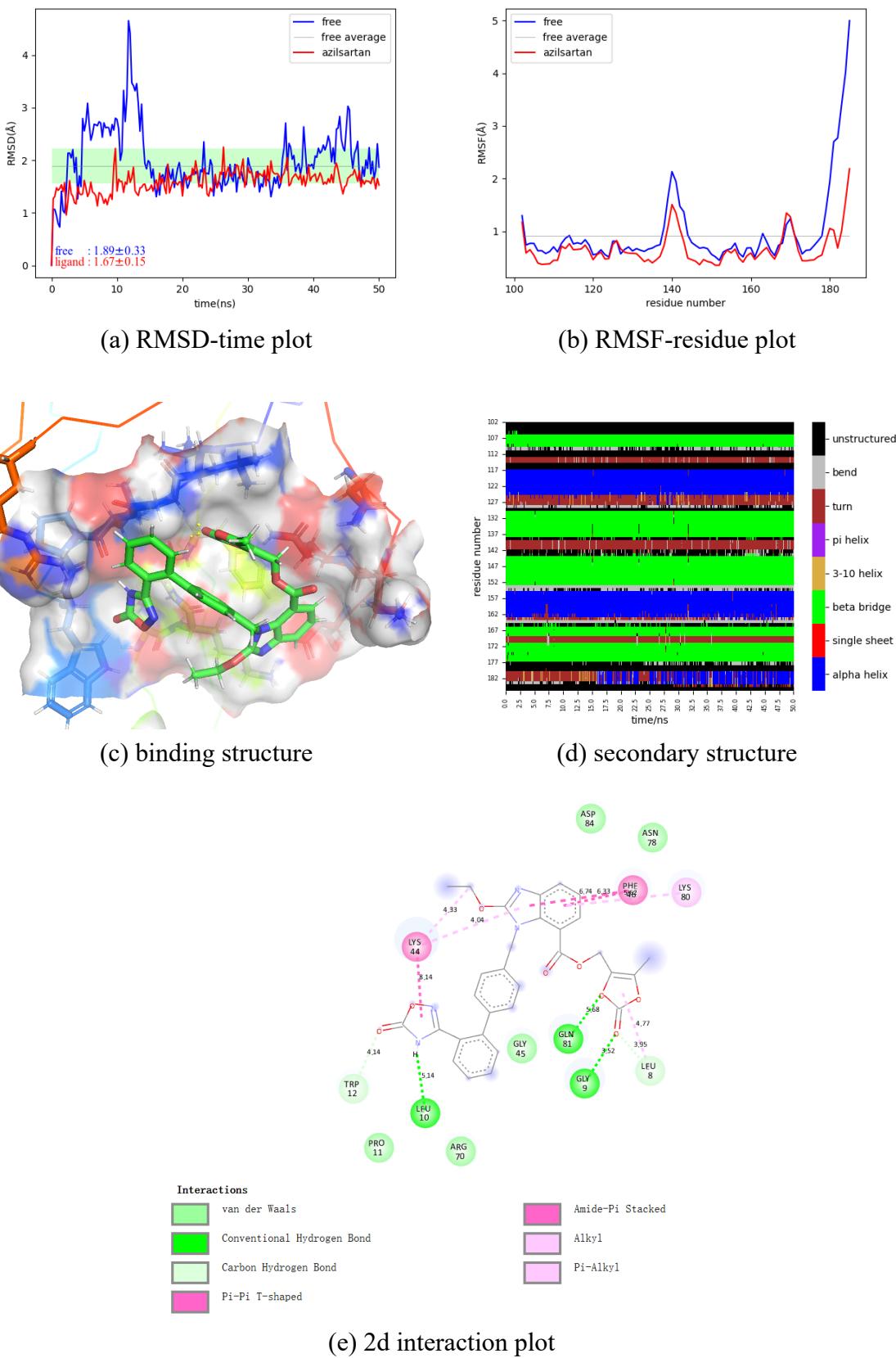
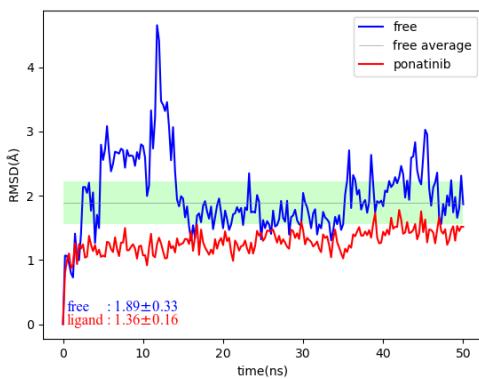
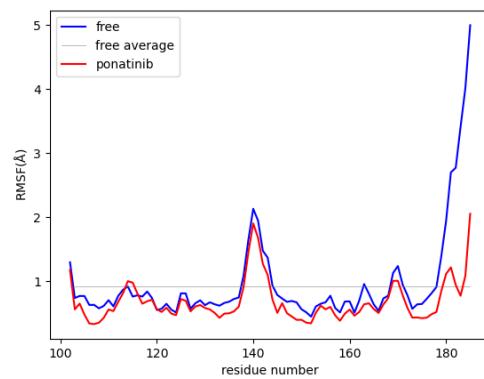


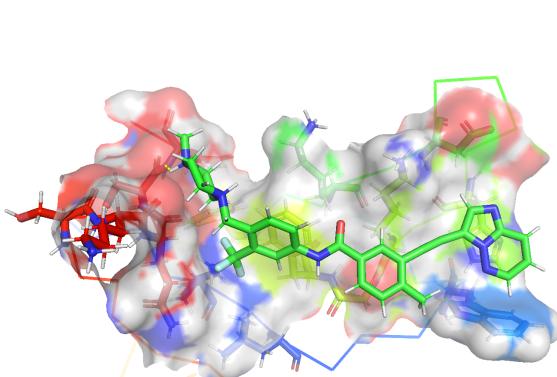
Fig. 9: Results of azilsartan (50ns MD simulation)



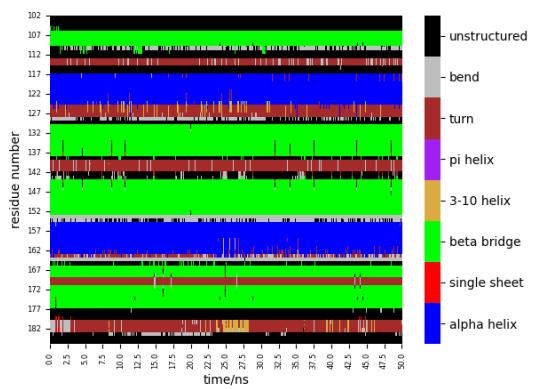
(a) RMSD-time plot



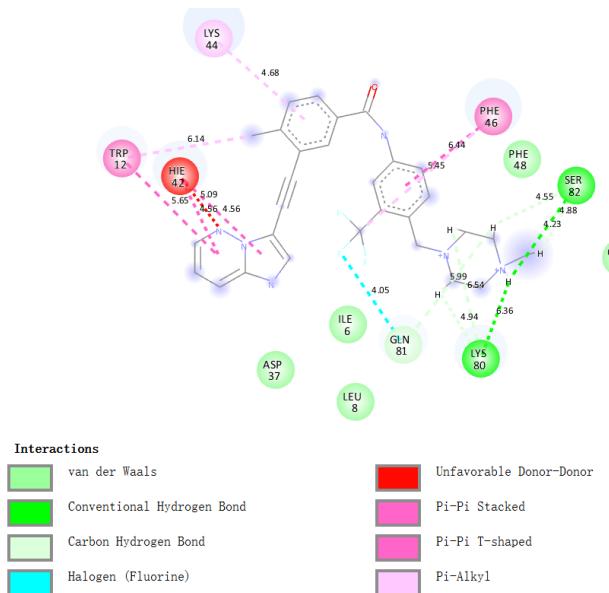
(b) RMSF-residue plot



(c) binding structure



(d) secondary structure



(e) 2d interaction plot

Fig. 10: Results of ponatinib (50ns MD simulation)

3.3.5 Ponatinib

Though its position fluctuates, the RMSD of protein is the lowest and stable (1.36 ± 0.16). The RMSF, SASA and Rg all look similar to azilsartan's, only except that stabilization on the main loop disappears. Another difference is the lack of hydrogen bonds, which only the terminal piperazine may form. As a result, the ΔG (-17.41 ± 0.23) is about 7 kcal/mol higher than azilsartan. Ramachandran analysis even implies a destabilization effect. The double ring stacks with 113W and 143H, while the other two aromatic rings' interaction with 145K and 147F contributes a lot to ΔG . Its σ_{decomp} is pretty low (0.51 kcal/mol) but still not as well as azilsartan.

4 Conclusion

After performing MD simulation and analysis, we recommend azilsartan and indacaterol 8-O-glucuronide as compounds for the next-step experimental validation. Azilsartan shows the strongest stabilizing effect, while indacaterol 8-O-glucuronide shows good overall properties. The wide range of stabilizing is the key reason they are selected.

Lumacaftor and ponatinib might also be recommended but with less priority. Though lumacaftor shows fairly good properties, no decrease in RMSF and its binding site become major problems. Ponatinib is considered inferior to azilsartan, but other newer TKI drugs might also have the potential to be the candidate inhibitor of TDP-43. Ponatinib and midostaurin decrease the RMSF of 169D where a well-known ALS-causing mutation occurs.

Midostaurin shows the most special structural change and moderately good stabilization. It is not excluded from candidates. Its dynamic behavior needs more validation to further determine how it can change the protein structure. Moreover, lumacaftor and midostaurin can adopt other binding modes in the 10-ns simulation. To explore more possibilities, 50-ns simulations for them are also worth trying.

5 Acknowledgements

I would like to extend my sincerely gratitude to Prof. Song Jianxing, my supervisor, who kindly provided the valuable research topic for me to step into structural biology research. During the three months, I learned a lot of techniques and concepts, e.g. the relationship between experiment and computing research. This project is the first step in the future long research. Also, thanks to the computing facilities and a confortable students' office, the virtual experiments and discussion can go smoothly.

During the project, the graduate in our lab, Dang Mei, offered her massive and devoted help on the experiment design and result analysis. She is always providing valuable suggestions and sparing her time guiding us junior students. I'm deeply indebted to her for the selfless help with both academic issues and my life in Singapore!

References

- Sashank Agrawal, Pan-Hsien Kuo, Lee-Ya Chu, Bagher Golzarroshan, Monika Jain, and Hanna S Yuan. RNA recognition motifs of disease-linked RNA-binding proteins contribute to amyloid formation. *Scientific Reports*, 9(1):6171, 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-42367-8. URL <https://doi.org/10.1038/s41598-019-42367-8>.
- D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, and P.A. Kollman. AMBER 2020, 2020.
- Peter J A Cock, Tiago Antao, Jeffrey T Chang, Brad A Chapman, Cymon J Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, and Michiel J L de Hoon. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, 2009. ISSN 1367-4803. doi: 10.1093/bioinformatics/btp163. URL <https://doi.org/10.1093/bioinformatics/btp163>.
- Mei Dang, Jian Kang, Liangzhong Lim, Yifan Li, Lu Wang, and Jianxing Song. ATP is a cryptic binder of TDP-43 RRM domains to enhance stability and inhibit ALS/AD-associated fibrillation. *Biochemical and Biophysical Research Communications*, 522(1):247–253, 2021a. doi: 10.1016/j.bbrc.2019.11.088. URL <https://www.sciencedirect.com/science/article/pii/S0006291X19322144>.
- Mei Dang, Yifan Li, and Jianxing Song. Tethering-induced destabilization and ATP-binding for tandem RRM domains of ALS-causing TDP-43 and hnRNPA1. *Scientific Reports*, 11(1):1–17, 2021b. ISSN 20452322. doi: 10.1038/s41598-020-80524-6. URL <https://doi.org/10.1038/s41598-020-80524-6>.
- Liberty François-Moutal, Razaz Felemban, David D. Scott, Melissa R. Sayegh, Victor G. Miranda, Samantha Perez-Miller, Rajesh Khanna, Vijay Gokhale, Daniela C. Zarnescu, and May Khanna. Small Molecule Targeting TDP-43’s RNA Recognition Motifs Reduces Locomotor Defects in a Drosophila Model of Amyotrophic Lateral Sclerosis (ALS). *ACS Chemical Biology*, 14(9):2006–2013, 2019. ISSN 15548937. doi: 10.1021/acscchembio.9b00481.
- M J Frisch, G W Trucks, H B Schlegel, G E Scuseria, M A Robb, J R Cheeseman, G Scalmani, V Barone, G A Petersson, H Nakatsuji, X Li, M Caricato, A V Marenich, J Bloino, B G Janesko, R Gomperts, B Mennucci, H P Hratchian, J V Ortiz, A F Izmaylov, J L Sonnenberg, D Williams-Young, F Ding, F Lipparini, F Egidi, J Goings, B Peng, A Petrone, T Henderson, D Ranasinghe, V G Zakrzewski, J Gao, N Rega, G Zheng, W Liang, M Hada, M Ehara, K Toyota, R Fukuda, J Hasegawa, M Ishida, T Nakajima, Y Honda, O Kitao, H Nakai, T Vreven, K Throssell, J A Montgomery Jr., J E Peralta,

F Ogliaro, M J Bearpark, J J Heyd, E N Brothers, K N Kudin, V N Staroverov, T A Keith, R Kobayashi, J Normand, K Raghavachari, A P Rendell, J C Burant, S S Iyengar, J Tomasi, M Cossi, J M Millam, M Klene, C Adamo, R Cammi, J W Ochterski, R L Martin, K Morokuma, O Farkas, J B Foresman, and D J Fox. Gaussian 16 Revision C.01, 2016.

Erika N Guerrero, Haibo Wang, Joy Mitra, Pavana M Hegde, Sara E Stowell, Nicole F Liachko, Brian C Kraemer, Ralph M Garruto, K S Rao, and Muralidhar L Hegde. TDP-43/FUS in motor neuron disease: Complexity and challenges. *Progress in Neurobiology*, 145-146:78–97, 2016. ISSN 0301-0082. doi: <https://doi.org/10.1016/j.pneurobio.2016.09.004>. URL <https://www.sciencedirect.com/science/article/pii/S0301008216300740>.

Berk Hess, Henk Bekker, Herman J.C. Berendsen, and Johannes G.E.M. Fraaije. LINCS: A Linear Constraint Solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997. ISSN 01928651. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

Jeffrey W Hofmann, William W Seeley, and Eric J Huang. RNA Binding Proteins and the Pathogenesis of Frontotemporal Lobar Degeneration. *Annual Review of Pathology: Mechanisms of Disease*, 14(1):469–495, 2019. doi: 10.1146/annurev-pathmechdis-012418-012955. URL <https://doi.org/10.1146/annurev-pathmechdis-012418-012955>.

Greg Landrum. RDKit: Open-Source Cheminformatics Software. 2016. URL https://github.com/rdkit/rdkit/releases/tag/Release_2016_09_4.

Lindahl, Abraham, Hess, and van der Spoel. GROMACS 2021 Manual, jan 2021. URL <https://doi.org/10.5281/zenodo.4420784>.

Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L. Klepeis, Ron O. Dror, and David E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics*, 78(8):1950–1958, 2010. ISSN 08873585. doi: 10.1002/prot.22711.

Christopher A. Lipinski. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discovery Today: Technologies*, 1(4):337–341, 2004. ISSN 17406749. doi: 10.1016/j.ddtec.2004.11.007.

Simon C Lovell, Ian W Davis, W Bryan 3rd Arendall, Paul I W de Bakker, J Michael Word, Michael G Prisant, Jane S Richardson, and David C Richardson. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*, 50(3):437–450, feb 2003. ISSN 1097-0134 (Electronic). doi: 10.1002/prot.10286.

Peter J. Lukavsky, Dalia Daujotyte, James R. Tollervey, Jernej Ule, Cristiana Stuani, Emanuele Buratti, Francisco E. Baralle, Fred F. Damberger, and Frédéric H.T. Allain. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nature Structural and Molecular Biology*, 20(12):1443–1449, 2013. ISSN 15459993. doi: 10.1038/nsmb.2698. URL <http://dx.doi.org/10.1038/nsmb.2698>.

Ian R.A. Mackenzie and Rosa Rademakers. The role of transactive response DNA-binding protein-43 in amyotrophic lateral sclerosis and frontotemporal dementia. *Current Opinion in Neurology*, 21(6):693–700, 2008. ISSN 13507540. doi: 10.1097/WCO.0b013e3283168d1d.

Garrett M. Morris, Huey Ruth, William Lindstrom, Michel F. Sanner, Richard K. Belew, David S. Goodsell, and Arthur J. Olson. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16):2785–2791, 2009. ISSN 01928651. doi: 10.1002/jcc.21256.

Noel M O’Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1):33, 2011. ISSN 1758-2946. doi: 10.1186/1758-2946-3-33. URL <https://doi.org/10.1186/1758-2946-3-33>.

Trott Oleg and Arthur J. Olson. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*, 23(1):1–7, 2012. ISSN 15378276. doi: 10.1002/jcc.21334. AutoDock. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf>.

Amresh Prakash, Vijay Kumar, Naveen Kumar Meena, Md. Imtaiyaz Hassan, and Andrew M Lynn. Comparative analysis of thermal unfolding simulations of RNA recognition motifs (RRMs) of TAR DNA-binding protein 43 (TDP-43). *Journal of Biomolecular Structure and Dynamics*, 37(1):178–194, 2019. doi: 10.1080/07391102.2017.1422026. URL <https://doi.org/10.1080/07391102.2017.1422026>.

Lewis P Rowland. How Amyotrophic Lateral Sclerosis Got Its Name: The Clinical-Pathologic Genius of Jean-Martin Charcot. *Archives of Neurology*, 58(3):512–515, 2001. ISSN 0003-9942. doi: 10.1001/archneur.58.3.512. URL <https://doi.org/10.1001/archneur.58.3.512>.

Jeremy D. Schmit, Nilusha L. Kariyawasam, Vince Needham, and Paul E. Smith. SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation. *Journal of Chemical Theory and Computation*, 14(4):1823–1827, 2018. ISSN 15499626. doi: 10.1021/acs.jctc.7b01254.

Schrödinger, LLC. The PyMOL Molecular Graphics System, Version~1.8. nov 2015.

Teague Sterling and John J Irwin. ZINC 15 –Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–2337, 2015. doi: 10.1021/acs.jcim.5b00559. URL <https://doi.org/10.1021/acs.jcim.5b00559>.

Dassault Systèmes. BIOVIA Discovery Studio, 2021. URL <https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/>.

J Paul Taylor, Robert H Brown Jr, and Don W Cleveland. Decoding ALS: from genes to mechanism. *Nature*, 539:197+, mar 2016. ISSN 00280836. URL <https://link.gale.com/apps/doc/A469646969/AONE?u=nuslib&sid=AONE&xid=4898c83a>.

Mario Sergio Valdés Tresanco, Mario E Valdes-Tresanco, Pedro A Valiente, and Ernesto Moreno Frías. gmx_MMPBSA, mar 2021. URL <https://doi.org/10.5281/zenodo.4626383>.

Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general Amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004. ISSN 01928651. doi: 10.1002/jcc.20035.

Christopher J. Williams, Jeffrey J. Headd, Nigel W. Moriarty, Michael G. Prisant, Elizabeth L. Videau, Lindsay N. Deis, Vishal Verma, Daniel A. Keedy, Bradley J. Hintze, Vincent B. Chen, Swati Jain, Steven M. Lewis, W. Bryan Arendall, Jack Snoeyink, Paul D. Adams, Simon C. Lovell, Jane S. Richardson, and David C. Richardson. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*, 27(1):293–315, 2018. ISSN 1469896X. doi: 10.1002/pro.3330.

David S. Wishart, Craig Knox, An Chi Guo, Savita Shrivastava, Murtaza Hassanali, Paul Stothard, Zhan Chang, and Jennifer Woolsey. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research*, 34(Database issue): 668–672, 2006. ISSN 13624962. doi: 10.1093/nar/gkj067.

Elsa Zacco, Stephen R Martin, Richard Thorogate, and Annalisa Pastore. The RNA-Recognition Motifs of TAR DNA-Binding Protein 43 May Play a Role in the Aberrant Self-Assembly of the Protein. *Frontiers in Molecular Neuroscience*, oct 2018. doi: <http://dx.doi.org/10.3389/fnmol.2018.00372>. URL http://libproxy1.nus.edu.sg/login?url=https://www.proquest.com/scholarly-journals/rna-recognition-motifs-tar-dna-binding-protein-43/docview/2309515990/se-2?accountid=13876http://bb2sz3ek3z.search.serialssolutions.com/directLink?&atitle=The+RNA+Recognition+Motifs+of+TAR+DNA+Binding+Protein+43+May+Play+a+Role+in+the+Aberrant+Self+Assembly+of+the+Protein&author=Zacco%2C+Elsa%3BMartin%2C+Stephen+R%3BThorogate%2C+Richard%3BPastore%2C+Annalisa&issn=&title=Frontiers+in+Molecular+Neuroscience&volume=&issue=&date=2018-10-09&spage=&id=doi:10.3389%2Ffmol.2018.00372&sid=ProQ_ss&genre=article.

Appendix

A Virtual screening and drug-like properties

The AutoDock Vina program generated nine conformers that are most energy-favored for each ligand and calculated binding energy for each conformer. Ligands in two subsets were separately ranked according to the lowest binding energy. The threshold is set to -6.9 kcal/mol because about 10 ligands are desired to be picked. However, the computed binding energy, including those calculated from MD simulation, will usually not be compared with the experimental data for limit of accuracy. They are used for ranking the *in silico* screened ligands.

Lipinski's rules of five is generally a limitation of molecular properties. Take $\log P$ as an example, a drug molecule should have appropriate solubility in both water and lipids in order to flow with body fluid and pass through the cell membrane respectively. Additionally, two extra rules are added to the filter so that the ligand could mimic ATP: there should be at least one hydrogen donor and one aromatic ring in the molecule. Based on these criteria, up to 13 ligands stand out.

The pre-experiment shows that binding energy calculated by the Autodock Vina program is correlated to the molecular weight of the small molecule (Fig. 11a). Large MW ligands do not always show low binding energy, while small MW ligands can impossibly show very low binding energy. This is a disadvantage of the program, as well as the reason low MW ligands are also excluded.

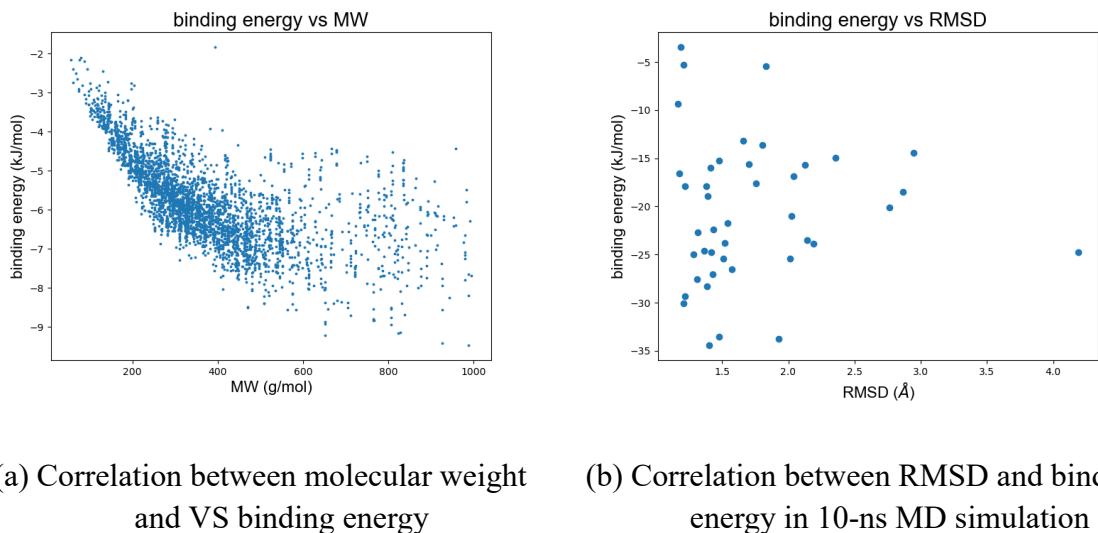


Fig. 11: Correlation study

B Properties in molecular dynamics simulation

RMSD in this report quantifies the average distance of C_α atoms from the original structure, indicating the structural change. If the RMSD value seems to converge to a

constant, and the fluctuation or standard deviation is low, the protein-ligand complex is structurally stable.

RMSF in this report is defined as the average distance of C_α atom in each residue from the average position. If the RMSF value for a residue is high, this residue may be involved in wild positional fluctuation. RMSD and RMSF is most widely-used indexes to depict stability.

Solvent accessible surface area (SASA) is generally the area of the outer surface of a molecule, which describes how the ligand can induce conformational change in the protein. Larger SASA might imply exposure of hydrophobic residues. Radius of gyration (R_g) is derived from the motion of the protein. Larger R_g suggests a more compact and stable structure. Change in these two properties are often dependent on the specific structure of protein. Number of hydrogen bonds between protein and ligand is another common indicator, which highly relies on ligand structure. Thus they only act as references rather than "screening criteria".

MMPBSA and MMGBSA are common binding energy calculation tools in MD simulations. A brief introduction to the principles is here. MMGBSA is thought to be more simplified and less time-consuming than MMPBSA. They mainly differ in polar part of the solvation free energy calculation, where they use different implicit solvent model. MMPBSA is believed to be closer to experimental data, while MMGBSA is known to be suitable for comparing the binding energy of various ligands on the same protein. Evenly-spaced frames are extracted from the trajectory, and binding energy is calculated for each frame. Thus the statistical average and standard deviation is obtained, which can evaluate the binding energy more realistically.

However in this study, the calculated binding energy of ligands doesn't seem reasonable comparing to experimental value obtained from our lab previously. And the difference between MMPBSA and MMGBSA is really big. Maybe it's due to the details of this particular system. Despite this, they are still usually used in current research papers.

As Tab. 8 and Fig. 11b indicates, it is found that there is not significant linear correlation between averaged bind energy and other properties.

Tab. 8: more properties in 10 ns MD simulation

ligand	ΔG (kcal/mol)				RMSD ¹ (Å)	RMSF (Å)	SASA (nm ²)	R_g (Å)	H bond count with ligand
	run 1	run 2	run 3	average					
indacaterol 8-O-Glucuronide	-34.52 ± 0.81	-28.29 ± 0.52	-34.26 ± 0.54	-32.35	2.29	0.78	60.66	12.93	3.14
lumacaftor	-29.65 ± 0.37	-26.45 ± 0.57	-35.53 ± 0.55	-30.54	1.55	0.77	60.87	12.85	1.65
midostaurin	-22.97 ± 0.59	-27.53 ± 0.51	-28.23 ± 0.34	-26.24	1.42	0.77	61.90	12.92	0.37
azilsartan	-24.77 ± 0.28	-25.58 ± 0.25	-24.49 ± 0.59	-24.95	1.45	0.71	59.54	12.82	1.01
ponatinib	-24.22 ± 0.55	-20.89 ± 0.33	-26.74 ± 0.38	-23.95	1.27	0.69	59.41	12.79	1.31
lurasidone	-25.44 ± 0.69	-21.43 ± 0.60	-22.39 ± 0.48	-23.09	1.69	0.84	61.59	12.94	0.76
dihydroergotoxine	-25.04 ± 0.43	-24.58 ± 0.30	-14.02 ± 0.49	-21.21	1.46	0.84	60.04	12.82	0.94
dihydro-alpha-ergocryptine	-16.74 ± 0.25	-18.35 ± 0.36	-17.96 ± 0.38	-17.68	1.37	0.69	59.47	12.83	0.85
nilotinib	-17.32 ± 0.41	-15.26 ± 0.71	-20.23 ± 0.28	-17.60	1.82	0.77	59.34	12.83	0.57
brigitanib	-15.42 ± 0.27	-16.31 ± 0.45	-17.90 ± 0.38	-16.54	1.31	0.70	59.26	12.77	0.37
capmatinib	-25.13 ± 0.52	-19.36 ± 0.38	-03.21 ² ± 0.43	-15.90	1.79	0.86	59.50	12.80	0.94
ubrogepant	-07.56 ± 0.37	-16.71 ± 0.45	-20.80 ± 0.32	-15.03	2.00	0.75	59.72	12.78	0.82
ATP	-15.09 ± 0.57	-04.59 ± 0.69 ³	-19.70 ± 0.48	-13.13	3.33	0.94	60.61	12.95	3.44
imatinib	-15.12 ± 0.55	-04.68 ± 0.41 ⁴	-12.71 ± 0.40	-10.84	1.34	0.72	59.24	12.74	0.86
free	—	—	—	—	1.73	0.84	59.37	1.280	—

¹ Only mean values are provided for the following properties.

² In this run, the ligand went out of the binding pocket. So do note ³ and ⁴.

All the values in the tables and figures are excluding the first 20 ns. The grey horizontal line refers to the average of this property of free protein, while the lime rectangle around the line is a region between average (μ) \pm standard deviation (σ). Standard deviation of RMSD is picked to measure the stability of this trajectory.

The relationship between energy components are:

$$\begin{aligned}\Delta G &= \Delta G_{gas} + \Delta G_{solvation} \\ \Delta G_{gas} &= \Delta G_{vdW} + \Delta G_{ele} \\ \Delta G_{solvation} &= \Delta G_{polar} + \Delta G_{nonpolar}\end{aligned}$$

But this formula for ΔG in tables means a little different: average \pm standard error of average, which equals to $\frac{\sigma}{\sqrt{n}}$. n is the number of data points. This is used to measure how close the calculated average is to the "real" average.

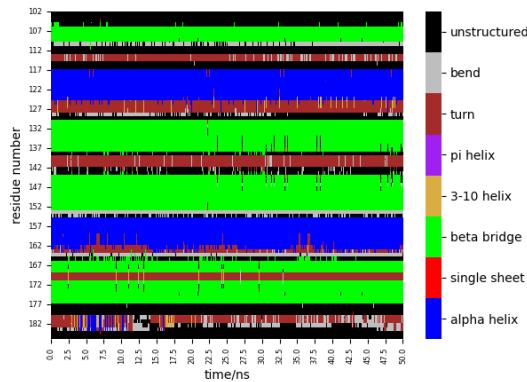


Fig. 12: SS plot of "free"

Second structure (SS) analysis were conducted to observe the effect of ligand binding on the protein. The plot illustrates which type of secondary structure is assigned to each residue in every moment, thus shows the dynamic change and protein structure stability. Fig. 12 is an example.

Ramachandran plot is a common-used representation of protein conformation and criteria for quality of protein structures solved by X-ray crystallography and NMR. In this study, several frames are extracted to calculate all ϕ and ψ angles of all residues excluding the first and last two. Only when the angles are in a certain region specified by residue type can this conformer be considered "reasonable" or "favored"; otherwise, there might be certain combination of angles that leads to sterical repulsion between atoms, which means instability. It is called "outliers". Criteria proposed by Lovell et al. (2003) is adopted. The fractions of all three types are averaged overall all frames and shown in Fig 5a.