

Computational analysis and drug repositioning: Targeting the TDP-43 RRM using FDA-approved drugs

Siran Zhang¹, Longjiang Wu², Xiaoying Zhang^{2,3,4}, Mei Dang⁵

¹ Xi'an Middle School of Shaanxi Province, Xi'an, Shaanxi, China

² Qinba State Key Laboratory of Biological Resources and Ecological Environment, College of Biological Science and Engineering, Shaanxi University of Technology, Hanzhong, Shaanxi, China

³ Centre of Molecular & Environmental Biology, Department of Biology, University of Minho, Braga, Portugal

⁴ Department of Biomedical Sciences, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada

⁵ Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore

SUMMARY

TAR DNA binding protein-43 (TDP-43) aggregation is a hallmark for many neurodegenerative diseases including amyotrophic lateral sclerosis, frontotemporal dementia and Huntington's disease. These protein aggregates can disrupt neuronal function and contribute to neurodegeneration. Previous studies have uncovered that adenosine triphosphate (ATP) is a promising molecule to dock onto the TDP-43 RNA recognition motif (RRM) region to reduce amyloid-like aggregation. This provides a potential therapeutic strategy in which chemicals with similar binding properties could be selected as drugs. Under normal physiological conditions, TDP-43 RRM region mediates the binding of nucleic acid with TDP-43. Therefore, we hypothesized that molecules such as tyrosine kinase inhibitors (TKIs), which can bind to ATP-binding sites or competitively bind to other nucleic acid binding regions, including different variants of RRM domains, are of great screening interest. We conducted *in silico* simulations using molecular dynamic simulation and virtual screening, in which the ATP-binding pocket is introduced in docking model. Our results supported our hypothesis because five of ten selected binding chemicals were TKIs. From the result, we then selected the two molecules under maximum concentration in bloodstream by conducting further screening strategies such as long-term molecular dynamic simulation, and Lipinski's rules testing.

INTRODUCTION

Pathogenic deposits comprising TAR DNA-binding protein (TDP-43) are a distinguishing feature found in brain and spinal cord of individuals affected by a diverse neurodegenerative disease. These deposits have been acknowledged as the principal pathogenic determinant contributing to the etiology of these disorders, leading to their categorization as "TDP-43 proteinopathies" (1). Notably, cytosolic aggregation of TDP-43 is evident in up to 97% of patients with sporadic amyotrophic lateral sclerosis (ALS) and approximately 45% of patients with frontotemporal lobar degeneration (2). Consequently, research on TDP-43 offers an opportunity to gain insights into the progression of neurodegenerative diseases and

may facilitate development of chemical probes for potential therapeutic interventions.

TDP-43, a 43 kDa protein, was initially identified in 1995 as a suppressor of HIV-1 gene expression (3). Physiologically, TDP-43 is a pivotal DNA/RNA binding protein involved in multiple cellular processes. Its structure comprises an N-terminal domain and two tandem ribonucleic acid (RNA) recognition motifs (RRMs), RRM1 and RRM2, followed by a C-terminal prion-like domain (4). Of note, TDP-43 binds to nucleic acids via its RRM domains, contributing to RNA processing, including splicing, translation, and the formation of cytoplasmic stress granules that improve messenger RNA stability to minimize stress-related damage (5).

While the C-terminal prion-like domain has been extensively studied due to its prion-like activities in TDP-43 cytoplasmic aggregation, the RRM domains, which are abundantly found in eukaryotes, also bear responsibility for cytosolic aggregate formation and disease phenotypes (8). Notably, Shodai, et al. demonstrated that RRM1 readily acquires amyloidogenicity, which is the tendency to produce amyloid deposits, under physical stresses and contributes to pathogenic conversion of TDP-43 in ALS (9). Moreover, RRM1 drives the localization of TDP-43 to stress granules, whose excessive formation further facilitates cytoplasmic aggregation of TDP-43 in motor neurons (10). Both simulation and experimental studies have indicated that RRM1 exhibits structural and conformational instabilities, rendering it more susceptible to the formation of fibrils and aggregation (11). These findings highlight the significance of RRM1 in the formation of TDP-43 aggregates and suggest that targeting the RRM1 domain with small molecules holds promise in counteracting TDP-43 toxicity; however, current studies in this area are limited.

In our previous study, we discovered that ATP could not only increase the thermodynamic stability of TDP43 RRM1, but also suppressed the formation of amyloid-like aggregation (22). However, considering the multifunctional and degradation-prone nature of ATP *in vivo*, it is unsuitable for direct therapeutic application due to interference with other pathways. Therefore, the identification and design of ATP alternatives represents an essential initial step in the development of therapeutic drugs for TDP-43-related neurodegeneration disorders.

Thus, this study investigated small molecules with the potential to target the ATP-binding site on TDP-43 RRM1 to mitigate TDP-43 toxicity. Tyrosine kinase inhibitors (TKIs) typically have ATP-like structural features, often with a binding

mode similar to ATP, and can interact with the ATP binding pocket in the kinase binding site. Thus, we hypothesized that TKIs, which can alternatively bind to ATP-binding sites or competitively bind to other nucleic acid binding regions including different variants of the RRM domain, would be of great screening interest. Toward this end, we identified small molecules that interact favorably with the ATP-binding pocket *in silico*. A library of 2500 compounds was selected for docking onto the target site. Subsequently, the top five hits obtained from docking were subjected to further molecular dynamic simulation studies, leading to the identification of two compounds, indacaterol 8-O-glucuronide and lumacaftor, that possessed favorable binding properties. These results suggest their potential as ATP alternatives capable of antagonizing TDP43 proteinopathy. Our study serves as a compelling example of implementing a computational approach to discover small molecules that specifically target TDP-43 RRM1.

RESULTS

Identification of candidates binding to the TDP-43 RRM1-ATP interface

We docked 2500 compounds from DrugBank against the target TDP-43 RRM1 using AutoDockTools software. We then performed free energy calculations using gmx_MMPBSA package to rank their binding affinity. The top ten molecules with the highest binding energy to the TDP-43 RRM1-ATP interface were identified (Table 1). A detailed

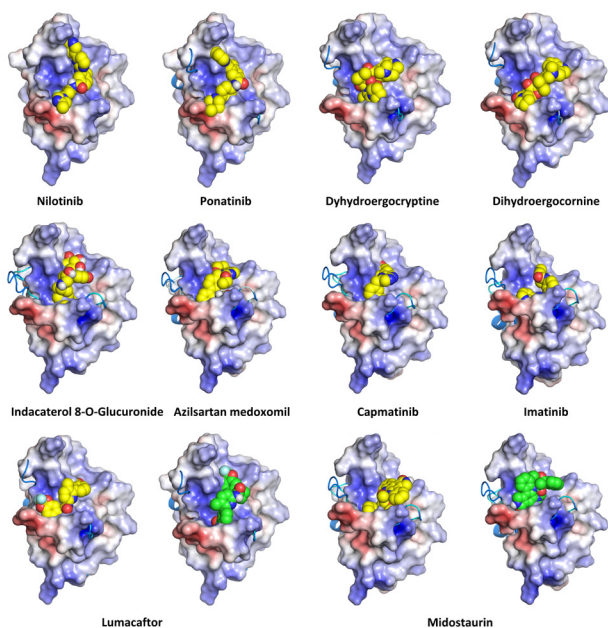


Figure 1: Electrostatic surface potentials of TDP-43 RRM1 in complex with various compounds. TDP-43 RRM1 is shown in electrostatic surface potential, as the color legend indicates, the red color (negative potential) represents negatively charged surface and the blue color (positive potential) occurs when the surface is positively charged. The corresponding ligand is displayed in spheres (yellow). The ligands that exhibited additional binding sites in docking structures, lumacaftor and midostaurin, are represented as green spheres.

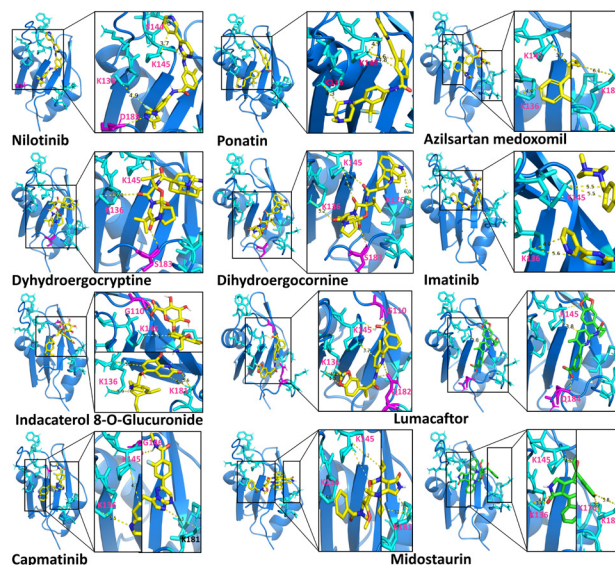


Figure 2: Molecular interactions between TDP-43 RRM1 and screened compounds. The diagram portrays TDP-43 RRM1 in cartoon, while ATP-perturbed residues are in sticks (cyan), and the ligands are in sticks (yellow or green). The inset provides detailed interactions between the protein and its corresponding compound.

examination revealed that ATP specifically binds to RRM1, partially overlapping with the nucleic acid-binding site, but penetrating deeper into the cavity due to its small size (22). The docking analysis revealed that the docked compounds displayed a great overlap with the ATP-binding surface of the TDP-43 RRM1, mimicking the interactions between RRM1 and ATP and occupying the positively charged cavity (Figure 1).

These ten molecules exhibited binding energies ranging from -6.643 to -7.996 kcal/mol, lower than that of ATP (-4.976 kcal/mol) (Table 1). Of the identified candidates, half were TKIs, including nilotinib, ponatinib, midostaurin, capmatinib, and imatinib, thus supporting our hypothesis. The druggability of molecules refers to the suitability of a compound or molecular target for development into a therapeutic drug, this was largely determined by their physicochemical properties. Criteria such as “Lipinski’s rule of five” (Ro5) were applied to quantify drug-like properties *in silico* to assess the pharmacokinetics and bioavailability profiles of the compounds (12). The Ro5 includes: a molecule with a molecular mass less than 500 Da (larger molecules may have difficulty crossing cellular membranes, limiting their absorption and distribution within the body), no more than 5 hydrogen bond donors, no more than 10 hydrogen bond acceptors (hydrogen bond donors can form interactions with water molecules, potentially reducing a compound’s ability to permeate lipid membranes), an octanol–water partition coefficient log P not greater than five (excessive lipophilicity may lead to poor aqueous solubility, and absorption, while over hydrophilicity may struggle to penetrate lipid barriers). A variation of the Ro5 proposed by Veber, 10 or fewer rotatable bonds, was also used to test metabolic stability and permeability of drugs.

Among the ten compounds, lumacaftor, capmatinib,

and imatinib fulfilled all four Ro5 rules, indicating their favorable pharmacokinetic properties. In contrast, nilotinib and azilsartan medoxomil violated two rules, primarily due to their high molecular weight (>500 g/mol) and octanol-water partition coefficient exceeding log P (Table 1). The remaining compounds violated only one rule, reflected in slightly higher molecular weights (500–600 g/mol). It should be noted that since all the selected compounds are FDA-approved and typically administered orally, strict adherence to Ro5 was not followed.

Determinants of binding affinity: role of H-bonds, pi-cation interactions, and binding sites

Among the top-ranked compounds, nilotinib demonstrated the highest affinity (-7.996 kcal/mol) towards RRM1. Detailed examination of the ligand-protein interactions was performed by PyMOL 2.5 showed that nilotinib forms two H-bonds with TDP-43 RRM1, which contribute to its high affinity, along with potential pi-cation interactions, and hydrophobic interactions. Ponatinib, the second-ranked compound, did not form any H-bonds but pi-cation interactions. Dihydroergocryptine and dihydroergocornine have similar hydrophobic interaction with TDP-43 RRM1. While dihydroergocryptine formed one more H-bonds with K145 and pi-cation interactions with K136, explaining its stronger binding affinity (-7.467 kcal/mol) compared to dihydroergocryptine (-7.223 kcal/mol) (Table 2). Indacaterol 8-O-glucuronide formed one H-bond and pi-cation interactions with K136, K145, and K181, resulting in a binding score of -7.161 kcal/mol (Figure 2).

Azilsartan medoxomil, the third-to last compound in binding energy (-6.942 kcal/mol), did not form any H-bonds but formed three pi-cation interactions. Capmatinib, the second last molecule in binding energy formed pi-cation interactions with the aforementioned three residues and an additional H-bond with G146, similarly to azilsartan medoxomil. However, the greater distance between the ligand and the three residues led to a slightly lower affinity (-6.923 kcal/mol). The molecule with least binding energy, imatinib, could not form any H-bonds and exhibited only two pi-cation interactions, thereby resulting the lowest binding affinity (-6.643 kcal/mol) (Table 2).

Notably, lumacafftor and midostaurin, each of them exhibited two potential ways to bind to RRM1 (Figure 1). Lumacafftor formed H-bonds with G110, K145, and Q182 on RRM1 and pi-cation interactions with K136. In comparison, the other binding configuration established H-bonds with K145 and Q184, suggesting that one of these two configurations may represent as a byproduct when one configuration is desired after the reaction, partially occupying the RRM1 and reducing the yield of main product. Unlike lumacafftor, midostaurin exhibited same pi-cation interactions in both binding sites. The residue composition of hydrophobic interaction clusters was similar in both states, making it difficult to determine the primary binding configuration.

In summary, the formation of H-bonds and pi-cation interactions played crucial roles in determining the binding affinity, with particular emphasis on K145 of RRM1 forming H-bonds with many molecules including nilotinib, ponatinib,

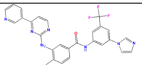
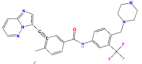
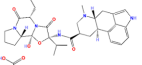
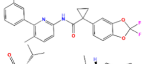
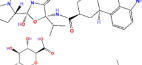
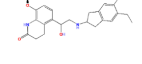
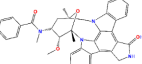
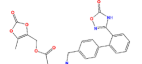
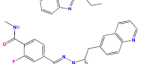
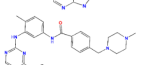
Compound	Structure	Binding Energy (kcal/mol)	Molecular Weight (g/mol)	H-bond Donor	H-bond Acceptor	Octanol-water partition coefficient (Log P)	Lipinski's rules	Rotatable bonds	Tyrosine Kinase inhibitor	Target disease
Nilotinib		-7.996	529.525	2	7	6.356	2	6	Yes	Chronic myeloid leukemia (CML)
Ponatinib		-7.561	532.570	1	6	4.456	3	4	Yes	Philadelphia chromosome-positive acute lymphoblastic leukemia/chronic myeloid leukemia
Dihydroergocryptine		-7.467	577.726	3	6	2.521	3	5	No	Parkinson's disease/ Alzheimer's disease
Lumacafftor		-7.428	452.414	2	5	4.747	4	5	No	Cystic fibrosis
Dihydroergocornine		-7.223	563.687	3	6	2.131	3	4	No	Hypertension/ peripheral vascular diseases
Indacaterol 8-O-Glucuronide		-7.161	568.623	7	9	0.714	2	9	No	Chronic obstructive pulmonary disease/asthma
Midostaurin		-7.083	570.649	1	6	5.907	2	3	Yes	Acute myeloid leukemia (AML)/aggressive systemic mastocytosis (ASM)/mast cell leukemia (MCL)
Azilsartan medoxomil		-6.942	568.542	1	11	4.705	2	9	No	Hypertension
Capmatinib		-6.923	412.428	1	6	3.429	4	4	Yes	Lung cancer/metastatic non-small cell lung cancer (NSCLC)/breast cancer malignant melanoma, colorectal cancer
Imatinib		-6.643	493.615	2	7	4.59	5	7	Yes	Systemic mastocytosis, myeloproliferative disease/dermatofibrosarcoma protuberans

Table 1: Drug-like properties of top 10 compounds selected by in silico screening.

Rank	Compound-Protein Complexes	ΔG (kcal/mol)
1	Indacaterol 8-O-Glucuronide-RRM1	-33.82 \pm 0.47
2	Lumacafftor-RRM1	-30.29 \pm 0.33
3	Midostaurin-RRM1	-27.28 \pm 0.45
4	Azilsartan medoxomil-RRM1	-26.39 \pm 0.32
5	Ponatinib-RRM1	-25.61 \pm 0.35
6	Nilotinib-RRM1	-22.61 \pm 0.42
7	Dihydroergotoxine-RRM1	-20.73 \pm 0.25
8	Dihydroergocryptine-RRM1	-18.56 \pm 0.43
9	Capmatinib-RRM1	-15.85 \pm 0.38
-	ATP-RRM1	-15.68 \pm 0.57
10	Imatinib-RRM1	-13.24 \pm 0.36
-	Free RRM1	-

Table 2: The binding free energies of the ligand-protein complexes in 10-ns MD simulations. ΔG values signify the average binding energies of the compound-TDP-43 RRM1 complexes, obtained from three independent 10-ns MD simulations. Control groups ATP-RRM1 and free RRM1 are excluded from the ranking for comparison. Conformers in which the ligand dissociated from the binding pocket were excluded from the analysis.

dihydroergocryptine, and lumacafftor, which contributed to the superior docking affinities of these molecules to others.

Binding affinities of compound-protein complexes revealed through MS simulations

To validate and quantify the stability of compound-protein complexes, we performed a preliminary 10-ns molecular dynamics (MD) simulation for all 10 complexes *in silico* (Figure 3). The binding free energy of individual compounds and ligand-protein complexes are shown in Table 1 and Table 2, respectively. With the exception of imatinib (-13.24 kcal/mol), all compounds hold stronger binding capabilities in complex towards RRM1 compared to ATP (-13.24 kcal/mol), with values ranging from -33.82 to -15.85 kcal/mol (Table 2). This is consistent with the above *in silico* studies, in which imatinib showed lowest binding capacity due to the absence of H-bonds and only two pi-cation interactions with RRM1. To further elucidate the conformational stabilization of the top potential compounds, we performed a more extensive 50-ns MD simulation for the top five selected compounds with the highest binding free energy, namely indacaterol 8-O-glucuronide, lumacafftor, midostaurin, azilsartan, and ponatinib. The binding free energies were calculated and ranged from -41.32 to -17.41 kcal/mol, with midostaurin displaying the highest and ponatinib the lowest affinity against TDP-43 RRM1 (Table 3). The relatively high binding energy of ponatinib could be attributed to its limited electrostatic interactions and non-polar interactions with the protein. However, electrostatic interactions are important for binding affinity of both indacaterol 8-O-glucuronide and lumacafftor, which ranked the first and second among the five compounds, respectively. While the binding capabilities resulting from polar interactions were relatively weaker for indacaterol 8-O-glucuronide and lumacafftor. The ΔG_{sol} values of indacaterol 8-O-glucuronide and lumacafftor

are relatively high with 143.12 kcal/mol and 105.86 kcal/mol respectively. This indicates the presence of solvent molecules also negatively influenced the binding affinity with RRM1, further inducing weaker binding affinities of these two compounds. Moreover, azilsartan medoxomil, the second-to-last compound in terms of binding energy, exhibited lower energy contributions from electrostatic interactions compared to ponatinib but higher than those of midostaurin, lumacafftor, and indacaterol 8-O-glucuronide.

Dynamic conformational changes of compound-protein complexes

To gain a dynamic perspective on conformational changes occurring during the 50-ns MD simulations, we computed the root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) for each complex, and individual compound without RRM1 (Table 4). The RMSD reflects the overall coordinate deviations at each time point, and these values of the complexes, were in the range of 1.38 to 2.57 Å. To assess the flexibility of residues during the MD simulations, the RMSF of C α -atoms was calculated for the five complexes to represent fluctuations of coordinates at residue level throughout the simulation. For both RMSD and RMSF values, high fluctuation may indicate instability or inconsistency in maintaining conformation, which is undesirable for durability and molecular docking.

Notably, both the midostaurin-RRM1 and lumacafftor-RRM1 complexes had average RMSD values higher than that of the ATP complex (1.86 Å) (Table 4). Despite midostaurin exhibiting the lowest binding energy, midostaurin-RRM1 had the highest RMSD, showing significant fluctuations after 12 ns compared to the unbound state. Lumacafftor-RRM1, characterized by the second largest RMSD value of 2.32 Å, induced noticeable fluctuations within the 20-30 ns timeframe. The distinct orientations adopted by lumacafftor, as observed in the docking results, may contribute to these

Compound-protein complexes	ΔG_{bind}	polar contributions			nonpolar contributions		ΔG_{sol}
		ΔG_{ele}	ΔG_{pol}	ΔG_{vdw}	ΔG_{nonpol}		
Midostaurin-RRM1	-41.32 \pm 0.21	-36.04	50.29	-49.45	-6.11	44.18	
Lumacafftor-RRM1	-34.27 \pm 0.29	-101.78	110.64	-38.35	-4.77	105.86	
Indacaterol 8-O-Glucuronide-RRM1	-29.82 \pm 0.31	-135.76	148.45	-37.17	-5.33	143.12	
Azilsartan medoxomil-RRM1	-24.96 \pm 0.17	-17.93	32.23	-34.88	-4.39	27.85	
Ponatinib-RRM1	-17.41 \pm 0.23	-1.71	15.88	-27.78	-3.80	12.08	
ATP-RRM1	32.99 \pm 0.68	-80.63	140.22	-22.74	-3.85	136.37	

Table 3: The binding free energies of the ligand-protein complexes in 50-ns MD simulations. ΔG_{bind} values represents the binding free energies of the five-top compound-TDP-43 RRM1 complexes, averaged from three independent 50-ns MD simulations. ΔG_{ele} : electrostatic free energy, ΔG_{vdw} : van der Waals free energy, ΔG_{pol} : electrostatic polar components of the solvation free energy, ΔG_{nonpol} : non-polar component of the solvation free energy, ΔG_{sol} : solvation energy.

Compound-protein complexes	RMSD	RMSD (compound)	RMSF	Rg	ΔfCt	SASA	NumH-bonds
Midostaurin-RRM1	2.57 ± 0.35	0.53 ± 0.34	0.84 ± 0.48	12.97 ± 0.18	1.59	63.14 ± 0.96	1.69 ± 1.24
Lumacafftor-RRM1	2.32 ± 0.16	2.32 ± 0.43	0.82 ± 0.59	12.78 ± 0.12	0.94	60.46 ± 0.89	2.54 ± 0.56
Indacaterol 8-O-Glucuronide-RRM1	1.41 ± 0.28	0.98 ± 0.39	0.70 ± 0.67	12.92 ± 0.09	1.78	61.38 ± 0.77	2.98 ± 0.67
Azilsartan Medoxomil-RRM1	1.77 ± 0.67	1.64 ± 0.22	0.79 ± 0.26	12.73 ± 0.06	2.32	58.36 ± 1.03	1.78 ± 0.24
Ponatinib-RRM1	1.38 ± 0.34	1.79 ± 0.78	0.68 ± 0.37	12.65 ± 0.16	2.39	58.69 ± 1.59	0.60 ± 0.45
ATP-RRM1	1.86 ± 0.49	1.32 ± 0.11	0.69 ± 0.60	12.94 ± 0.23	1.79	60.53 ± 0.83	3.26 ± 1.18
Free RRM1	1.79 ± 0.21	-	0.96 ± 0.57	12.86 ± 0.13	-	60.12 ± 0.61	-

Table 4: The topology and dynamics variations of the compound-protein complexes through 50-ns simulations. ΔfCt denotes the average differences between the complexes and the apo state, calculated across residues 180-184. Rg and SASA are abbreviations of radius of gyration and solvent accessible surface area, respectively. The descriptors are utilized to assess the conformational changes, flexibility, compactness, and surface accessibility of compound-protein complexes.

notable fluctuations. Lumacafftor itself had the highest RMSD value compared with other compounds. The RMSD values of RRM1 in the complex with indacaterol 8-O-glucuronide and ponatinib were comparable, with values of 1.41 and 1.38 Å, respectively. However, for compounds alone without RRM1, the RMSD value of indacaterol 8-O-glucuronide was lower than that of ponatinib, indicating larger fluctuations of the latter compound in the complex. The RMSD value of RRM1 in complex with azilsartan medoxomil was slightly higher than those of the Indacaterol 8-O-glucuronide and ponatinib complexes, but lower than that of the midostaurin and lumacafftor complexes. Throughout the simulations, the RMSD of RRM1 in the complex with indacaterol 8-O-glucuronide and ponatinib remained relatively stable.

The RMSF values for all complexes were below 1 Å, ranging from 0.68 to 0.84 Å. It should be noted that all compounds were able to stabilize the fluctuations of the C-terminal region. Although lumacafftor had a weaker stabilizing effect on the C-terminal region, it behaved almost the same as when it was unbound. Consistent with the RMSD analysis, the RMSF of RRM1 in complex with ponatinib and indacaterol 8-O-glucuronide showed comparable and the lowest values. In the case of indacaterol 8-O-glucuronide, it showed slightly weaker stability around K140, which is located in the large loop formed by $\beta 2$ and $\beta 3$. Further analysis of the secondary structure unraveled that midostaurin and azilsartan medoxomil induced structural changes over residues 179–182 transitioning from a turn-to-unstructured and turn-to-alpha helix, respectively.

DISCUSSION

In prior literature, the selection of ligand binding sites has often been left open to any region within the entirety of the target molecule (18). The present model distinguishes

itself by precisely designating the ATP-binding pocket as the focal point of interest. This strategic approach delimits the search scope for potential binding candidates and further defines the intended ligand function, which prevent TDP-43 aggregation through coverage of the β -strands within the ATP-binding pocket. The outcome of this study underscores the effectiveness of this approach, with a noteworthy observation that approximately half of the identified molecules are classified as TKIs. This observation attests to the robustness of the model, suggesting promising avenues for the construction of future virtual screening models aimed at identifying alternative ligand drugs for established protein-ligand complexes. The selected drugs indeed comprise a significant proportion of TKIs. They competitively binding to the ATP-binding site of tyrosine kinases (17). This binding action effectively curbs the aberrant behavior of mutated TKs, thereby hindering dysregulated signal cascades driven by ATP interactions. Given the high degree of conservation in ATP binding sites across various ATP-binding proteins and the overlapping binding cavities shared between kinases and TDP-43, coupled with the orientation of the virtual screening model toward the ATP-binding pocket, the emergence of TKIs as a substantial portion of the repurposed drug results is rational. This outcome can be attributed to the intrinsic alignment of the model's design and the functional attributes of TKIs.

Through virtual screening and MD simulations, two desirable molecules, lumacafftor and indacaterol 8-O-glucuronide, were identified. Of note, lumacafftor shows inferior drug properties than indacaterol 8-O-glucuronide. This is mainly because lumacafftor has two distinct conformations to dock on TDP-43 RRM1; it is therefore predicted that this ligand could waver between two conformations, which may partially compromise its stability and MD *in vivo*, resulting in its inferior druggability compared to indacaterol 8-O-glucuronide.

The colocalization and physical association among FUS, RBM45, and TDP-43 in ALS has been documented, implying that the RRM-binding agents elucidated in this study may possess therapeutic potential across a broader spectrum of neurodegenerative disorders (15,16). Although our study yields valuable insights, it is essential to accentuate the need for extended exploration. This entails not only in-depth preclinical and clinical investigations but also a meticulous examination of conformational dynamics, allosteric sites, and the application of a broader structural context for virtual screening experiments. Such endeavors are pivotal to advancing our understanding of potential therapeutic interventions and their translational implications.

To summarize, neurodegenerative diseases marked by TDP-43 proteinopathy necessitate targeted interventions. Employing a computational approach, we investigated small molecules binding to the ATP-binding site on TDP-43 RRM1. Utilizing *in silico* screening and subsequent simulations, two compounds, indacaterol 8-O-glucuronide and lumacafftor, hold significant promise as lead compounds, offering valuable structural insights into the design of anti-TDP-43 aggregation agents, and expediting the development of effective interventions for TDP-43-associated neurodegeneration. This work underscores the potential for computational methods in drug discovery and provides a foundation for developing anti-TDP-43 aggregation agents for neurodegenerative disease intervention.

MATERIALS AND METHODS

Preparation of the compound library and protein structure

We sourced a library consisting of 2500 FDA-approved drugs from the DrugBank and Zinc databases, and then converted drugs into three-dimensional (3D) structures using the Open Babel package (version 2.3.1) (18,19). We retrieved the solution structure of TDP-43 RRM1 structure from the Protein Data Bank (ID 2CQG) as the template for our study, the topology of which is very similar to those in the complex with nucleic acids. We selected the ATP-binding pocket within the first conformer of this structure as the primary binding site of interest (**Figure 3**), which is composed of W113, K114, K136, K137, K145, K176, K179, and K181.

Structure-based *in silico* screening

To identify the potential candidates targeting the TDP-43 RRM1, we employed a rigorous and systematic structure-based *in silico* molecular docking approach. Control groups including apo RRM1, which is the inactive or unbound form of the protein RRM1, as well as RRM1 in complex with ATP were also included for comparison. We utilized the AutoDockTools software, an integral component of the MGLTools 1.5.7 package, for optimizing and analyzing structures. The initial steps of the docking preparation involved adding missing atoms, assigning charges, and defining the region of interest within the protein for ligand binding. We then performed the molecular docking simulations using AutoDock Vina, a widely used protein-ligand program with predefined search space and docking parameters. Specifically, we docked the refined 3D structures of drugs sourced from DrugBank against TDP-43 RRM1. To ensure accurate sampling of the binding pocket, specific parameters for the 3D grid box were defined as follows: box size (Å) (x, y, z) = (14, 20, 16) and center coordinates (Å) (x, y, z) = (12, -7.5, 0.5). We set the exhaustiveness to 24, allowing for thorough exploration of the conformational space, while generating 10 conformers to capture diverse ligand binding orientations.

To validate the binding efficiencies, we conducted a redocking experiment. We re-docked the obtained binding configuration of the ligand onto the same active binding site

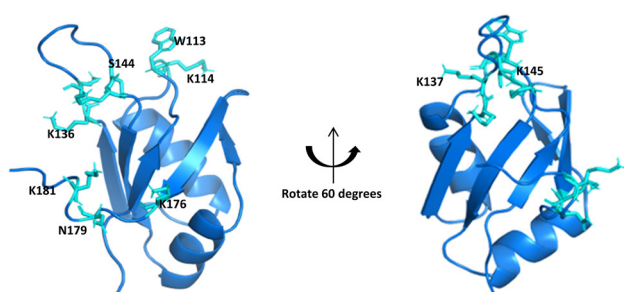


Figure 3: The binding site of interest on TDP-43 RRM1 used in computational screenings. The protein structure employed in computational screening was extracted from TDP-43 RRM1 solution structure (PDB ID 2CQG, model 01). The residues implicated in the binding site of interest are labeled and depicted in cyan sticks.

of the protein to ensure the consistency and reliability of the docking results. All docking protocols remained unchanged, and the grid parameters were kept constant. We shortlisted the ligands with favorable binding energies for further molecular dynamic simulation studies. The detailed examination of ligand-protein interactions was performed by PyMOL 2.5.

Molecular dynamics simulation

We conducted a preliminary 10 nanoseconds MD simulation using GROMACS v5.1.4 for the top 10 complexes with the lowest binding energy. Then we performed more extensive 50 nanoseconds simulations for top five complexes with the lowest binding energy. To create the simulation environment, we embedded the complexes in a predefined TIP3P water model within a periodic cubic box, where the minimal distance of 12 Å was maintained between any ligand atom and the box edge to prevent interactions with periodic images throughout the MD simulations. During the simulations, we employed Newton's classical equation of motion to calculate the atomic movements over time. In addition, we determined electrostatic potentials using the widely used 6-31G (d,p) basis set in Gaussian 16 and subsequently converted into partial charges through the Antechamber program. The protein and ligand were subjected to the AMBER99SB-IDLN all-atom force field and GAFF force field, respectively. The system was then neutralized by adding Na⁺ and Cl⁻ ions using GROMACS genion tool, achieving an ionic strength of 150 mM. Following neutralization, the system underwent an initial energy minimization using the steepest descent algorithm to achieve force convergence of <1000 kcal/mol/nm. Once the initial minimization was completed, we equilibrated the whole system for 5 ns at 300K degree and 1 bar pressure using canonical and the isothermal-isobaric ensembles, allowing for keeping the box volume minimized.

We computed the long-range electrostatic interactions using the particle-mesh Ewald summation, and constrained bond lengths through LINCS algorithm. Each compound-protein complex underwent three independent runs. In addition to the five complexes, two critical control groups were incorporated, namely the protein in apo form and the protein complexed with ATP. We conducted the trajectories analysis by excluding the initial 20 ns for system equilibrium. We evaluated the stability of the simulations by measuring RMSD and RMSF throughout the trajectories. We analyzed the secondary structure behaviors, and performed free energy calculations (20,21).

Received: August 24, 2023

Accepted: January 4, 2024

Published: April 26, 2024

REFERENCES

1. De Boer, Eva M. J., et al. "TDP-43 proteinopathies: a new wave of neurodegenerative diseases." *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 92, no. 86–95, Nov. 2020, <https://doi.org/10.1136/jnnp-2020-322983>.
2. Mehta, Puja R., et al. "The era of cryptic exons: implications for ALS-FTD." *Molecular Neurodegeneration*, vol. 18, no. 16, Mar. 2023, <https://doi.org/10.1186/s13024-023-00608-5>
3. Ou, S. H., et al. "Cloning and characterization of a

- novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs." *Journal of Virology*, vol. 69, no. 6, pp. 3584-3596, Jun. 1996, <https://doi.org/10.1128/jvi.69.6.3584-3596.1995>
4. François-Moutal, Liberty., et al. "Structural Insights Into TDP-43 and Effects of Post-translational Modifications." *Frontiers in Molecular Neuroscience*, vol. 12, no. 301, Dec. 2019, <https://doi.org/10.3389/fnmol.2019.00301>.
 5. Wolozin, Benjamin, and Pavel Ivanov. "Stress granules and neurodegeneration." *Nature Reviews Neuroscience*, vol. 20, no. 11, pp. 649-666, Nov. 2019, <https://doi.org/10.1038/s41583-019-0222-5>.
 6. Srivastava, Ankit., et al. "Prion-like behavior of TDP-43 aggregates and its implication to disease." *TDP-43 and Neurodegeneration*, edited by Vijay Kumar and Manoj Kumar Jaiswal. pp. 97-115, Jan. 2021, <https://doi.org/10.1016/B978-0-12-820066-7.00004-7>
 7. Wishart, D. S., et al. "DrugBank 5.0: a major update to the DrugBank database for 2018." *Nucleic Acids Research*, vol. 3, no. 766, Apr. 2012, <https://doi.org/10.1038/ncomms1766>.
 8. Ash, Peter E A et al. "Neurotoxic effects of TDP-43 overexpression in *C. elegans*." *Human Molecular Genetics*, vol. 19, no. 16, pp. 3206 - 3218, Aug. 2010, <https://doi.org/10.1093/hmg/ddq230>.
 9. Shodai, A., et al. "Aberrant assembly of RNA recognition motif 1 links to pathogenic conversion of TAR DNA-binding protein of 43 kDa (TDP-43)." *Journal of Biological Chemistry*, vol. 288, pp. 14886-14905, May. 2013, <https://doi.org/10.1074/jbc.M113.451849>.
 10. Aulas, Anaïs, and Christine Vande Velde. "Alterations in stress granule dynamics driven by TDP-43 and FUS: a link to pathological inclusions in ALS?" *Frontiers in Cellular Neuroscience*, vol. 9, no. 423, Oct. 2015, <https://doi.org/10.3389/fncel.2015.00423>
 11. Prakash, A., Kumar, V., Meena, N. K., Hassan, M. I., & Lynn, A. M. (2019). "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*, vol. 37, pp. 178-194, Jan. 2019, <https://doi.org/10.1080/07391102.2017.1422026>.
 12. Radwa, E., Noor, A., "Chapter Three - Using filters in virtual screening: A comprehensive guide to minimize errors and maximize efficiency." *Annual Reports in Medicinal Chemistry*, vol. 59, pp. 99-136, Nov. 2022, <https://doi.org/10.1016/bs.armc.2022.09.002>
 13. Kitamura, A., et al. "Hetero-oligomerization of TDP-43 carboxy-terminal fragments with cellular proteins contributes to proteotoxicity." *bioRxiv*, May 2022, <https://doi.org/10.1101/2022.05.22.493003>.
 14. Liu, W., et al. "Wang, Y., & Chen, G. (2021). Insights into the aggregation mechanism of RNA recognition motif domains in TDP-43: a theoretical exploration." *Royal Society Open Science*, vol. 8, no. 210160, Aug. 2021, <https://doi.org/10.1098/rsos.210160>.
 15. Collins, M., et al. "The RNA-binding motif 45 (RBM45) protein accumulates in inclusion bodies in amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration with TDP-43 inclusions (FTLD-TDP) patients." *Acta Neuropathologica*, vol. 124, no. 5, pp. 717-732, Nov. 2012, <https://doi.org/10.1007/s00401-012-1045-x>.
 16. Kang, J., et al. "ATP binds and inhibits the neurodegeneration-associated fibrillization of the FUS RRM domain." *Communications Biology*, vol. 2, no. 223, Jun. 2019, <https://doi.org/10.1038/s42003-019-0463-x>.
 17. Pottier, C., et al. "Tyrosine Kinase Inhibitors in Cancer: Breakthrough and Challenges of Targeted Therapy." *Cancers (Basel)*, vol. 12, no. 3, no. 731, Mar. 2020, <https://doi.org/10.3390/cancers12030731>.
 18. Afzaal, H., et al. "Virtual screening and drug repositioning of FDA-approved drugs from the ZINC database to identify the potential hTERT inhibitors." *Frontiers in Pharmacology*, vol. 13, no. 4816, Nov. 2022, <https://doi.org/10.3389/fphar.2022.1048691>.
 19. Wishart, D. S., et al. "DrugBank 5.0: a major update to the DrugBank database for 2018." *Nucleic Acids Research*, vol. 46, pp. D1074-D1082, Jan. 2018, <https://doi.org/10.1093/nar/gkx1037>.
 20. Cock, Peter J A., et al. "Biopython: freely available Python tools for computational molecular biology and bioinformatics." *Bioinformatics (Oxford, England)*, vol. 25, no.11, pp. 1422-1423, Jun. 2009, <https://doi.org/10.1093/bioinformatics/btp163>.
 21. Valdés-Tresanco., et al. "gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS." *Journal of Chemical Theory and Computation*, vol. 17, no. 10, pp. 6281-6291, Oct. 2021, <https://doi.org/10.1021/acs.jctc.1c00645>.
 22. Dang, M., et al. "ATP biphasically modulates LLPS of TDP-43 PLD by specifically binding arginine residues." *Communications Biology*, vol. 4, no. 714, Jun. 2021, <https://doi.org/10.1038/s42003-021-02247-2>.

Copyright: © 2024 Zhang, Wu, Zhang, and Dang. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.