Multi-Layer Computational Design of Small-Molecule Inhibitors Targeting Nck1 for Autoimmune Disease Treatment

Junyuan Chen *

The High School Affiliated to the Renmin University of China, Beijing, P.R. China

*Corresponding Author Email: jyandrewchen@163.com

Abstract. The adaptor protein Nck1 involved in the T cell CD3 signaling pathway is a promising therapeutic target for the treatment of immune system-related diseases. This study established a multi-layer virtual screening pipeline that integrates compound library generation, molecular docking, pan-assay inference compounds (PAINS) filtration, and free energy calculation from molecular dynamics (MD) simulation to design and sequentially identify promising novel small molecule inhibitors (SMIs) for autoimmune diseases treatment. The compound library consists of 10,149 potential SMIs by performing R group enumeration on a first-in-class SMI of Nck1, AX-024, proven effective in empirical experimentation. The compounds were first screened using molecular docking followed by PAINS analysis to narrow down the design pool for higher level computational modeling and simulations. The remaining compounds were subjected to MD simulations for binding free energy calculation of the SMI-Nck1 systems, which was used to propose top candidates with stable binding. For these candidates, extended MD trajectories were produced to reveal structural mechanics behind their strong performance. Binding mechanism analysis, RMSF profiling, and evaluation of R-group effects identified promising molecules, such as d6294, that consistently outperformed the reference AX-024 in simulations. These analyses revealed key interactions with the Nck1 protein, including salt bridges, cation-π interactions, and hydrogen bonds, that underpin their stability and potential effectiveness. Taken together, this study establishes a transferable pipeline for the *in silico* design of SMIs and identifies several top candidates with strong potential for subsequent in vitro validation for the treatment of Nck1-related immune system diseases.

Keywords: small molecule inhibitors (SMIs), molecular docking, molecular dynamics (MD), virtual screening, in silico design.

1. Introduction

Autoimmune diseases, such as autoimmune thyroid disease or type I diabetes (T1D), include a large range of disorders characterized by aberrant T cell and B cell reactivity towards normal host cells or structures.[1–3] The immune system mistakenly attacks normal tissues and results in progressive tissue or organ damage and eventually physiological dysfunction.[4–6] Although certain mechanisms group these conditions into a single category, the clinical progressions of different autoimmune diseases are highly varied.[1, 7] In recent years, efforts have been paid to understand the onset of these diseases and to develop therapeutics for them.[8] However, to date, curative treatments for most autoimmune diseases remain unavailable, and patients often depend on long-term immunosuppressive therapies that are not specific and/or do not fix the root causes of the illness, while sometimes carrying significant side effects.[5] Other immune system-related diseases like allergic asthma are also becoming increasing prevalent, affects the lives of many including children, and often lack definitive cures.[9, 10]

Given the substantial impact of autoimmune diseases on patient health and life quality, researchers have poured much effort into finding an effective and curative treatment for them.[8] Multiple therapeutic strategies aimed at modulating the immune system have been investigated, including immunotherapies targeting T cell-specific markers such as CD3[11, 12], CD4[13, 14], and CD8[15], and B cells[16, 17]. For some autoimmune diseases, antigen-specific therapy has also been explored to induce immune tolerance toward, in the case of T1D, for example, pancreatic β-cell antigens.[18–20] Collectively, these interventions aim to suppress or modify the autoimmune response responsible for self-destruction and thus to preserve or restore normal bodily functions like endogenous insulin

production.[21] A T-cell response to a foreign antigen is initiated when the T cell receptor (TCR)-CD3 complex recognizes and binds to a peptide antigen presented by antigen-presenting cells (APCs).[22] Subsequently, the CD3 complex recruits multiple intracellular proteins that are essential for signal transduction and T cell activation, including Lck[22, 23], ZAP-70[22, 24], and adaptor proteins such as Nck1[12, 22, 25, 26] and Nck2[22, 27]. In particular, Nck1 has been identified as an important positive regulator of T cell activation and facilitates the maturation of the immune synapse.[22] Targeting Nck1 with small molecules to disrupt its interaction with the complex has been shown to significantly impair T cell activation, highlighting its potential as a therapeutic target to suppress the autoimmune responses.[28]

In 2016, Borroto et al.[12] reported the discovery of a first-in-class orally available small molecule inhibitor (SMI), AX-024, that binds to SH3.1 domain of the Nck1 protein and inhibits its interactions with TCRs. Their *in vivo* study proved that AX-024 is effective in regulating abnormal immune responses and showed that AX-024 was effective in alleviating inflammation and allergic asthma in mice. While debate[29–31] was raised over whether the anti-inflammatory effects of AX-024 are primarily or exclusively mediated through its interaction with Nck1, emerging evidence[32, 33] supports the conclusion that AX-024 binds to Nck1 and contributes to T cell inhibition through this mechanism. In 2022, Ramirez et al.[33] supported the existence of effective binding between AX-024 and the Nck1/CD3 complex using a novel in-house experimental method, observing effective inhibition rates of 60% and 80% at 1:1 and 6:1 AX-024:GST-SH3.1 ratios, respectively. In this regard, optimization of the first-in-class AX-024 can potentially bring even higher inhibition rates at low concentrations, increasing the effectiveness of inhibition and hence aiding the treatment of autoimmune diseases and conditions.

However, *in vitro* optimization of such SMIs can be challenging, extremely time-consuming, and inefficient, especially when screening large libraries of structurally similar compounds without a strategy for prioritization. With the rise of computer modeling and simulation techniques, molecular dynamics (MD) serves as a popular tool for computation that offers direct sight of proteins' behavior on a molecular level with high accuracy and efficiency and has been proven effective in the design of SMIs and understanding of their binding mechanisms.[34–36] Molecular docking has also been widely used for the identification of hit SMIs by docking them into proteins and generating docking scores for effective selection of molecules from datasets with millions of molecules.[37] These *in silico* approaches significantly enhance the time and resource efficiency of SMI development by enabling the rational design and prioritization of candidate compounds before experimental validation. It is often standard practice to use molecular docking and MD simulations together as not a selection process but rather using docking to place the SMI in a roughly correct binding position before moving on to MD simulations, which can highly enhance the accuracy and efficiency of MD. For example, when researching the binding and inhibition mode of existing SMIs of influenza hemagglutinin in 2017, Guan et al.[38] used molecular docking before molecular dynamics simulations to find the best binding position of the small molecule in the binding pocket. While using molecular docking to provide starting positions for MD, I also employed both docking and MD simulations as screening methods of different accuracies and drug-likeness and/or false positive filtering (such as PAINS) to sequentially narrow down a large library of potential small molecules to the best molecules in an efficient and all-round way, a practice quite commonly used and proven effective in previous studies.[37, 39, 40] For instance, in 2023, Pathak et al.[41] used multiple layers of selection combining molecular docking using AutoDock Vina, drug-like properties filtration, and MD simulations to select top molecules from 97,999 natural compounds that could effectively bind to and inhibit the porcine epidemic diarrhea virus (PEDV) protease.

In this study, I employed a structure-based design strategy to generate a focused library of 10,149 compounds by permutating three functional groups on the scaffold of the known Nck1-inhibiting AX-024[12]. I established a high-throughput virtual screening (HTVS) pipeline that includes multiple layers of "sieving" or selection of the molecules to get down to the best-performing ones among the 10,149 potential SMIs (Figure 1). First, I used molecular docking to generate the binding poses of the

library to the Nck1 protein pocket and evaluate the scores of the potential SMIs. After applying a binding score cutoff to ensure a higher density of higher-affinity candidates, pan-assay interference compounds (PAINS) filters were used to eliminate molecules likely to yield false-positive results or have undesired harmful side effects in bioassays.[42–44] Subsequently, MD simulations were conducted to allow the calculation of binding free energies and further refine the ranking of the top candidates. From the narrowed chemical space, I selected 10 top-performing molecules after considering both performance and stability, and they were subjected to longer MD simulations and detailed structural analysis to examine the structure-property relationship and conformational stability of the protein-ligand complexes. From the screening pipeline, d6294 was identified as the most potent candidate, with a binding free energy of approximately -39.6 kcal/mol and consistently low ligand RMSD and RMSF profiles. It was observed that d6294 forms stable dual salt bridges along with supplementary π-π interactions, resulting in strong and persistent binding to the Nck1 pocket; most of the top compounds were able to achieve consistently better results than the reference (AX-024). These compounds are proposed to serve as novel lead SMIs targeting Nck1-mediated T cell activation, with the aim of identifying the most promising candidates for *in vitro* testing and guiding further optimization in future studies.

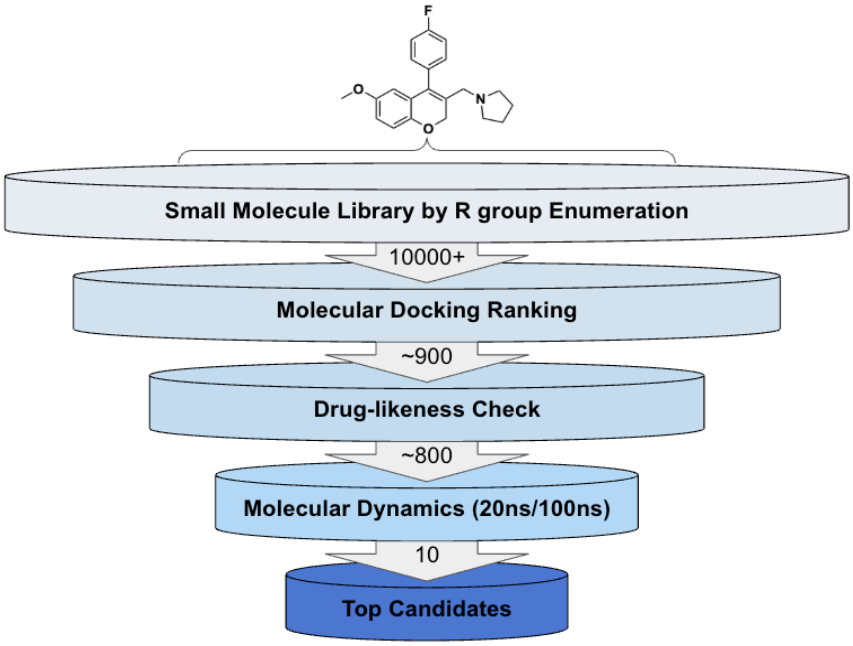

Figure 1. Schematic illustration of the multi-layer screening of SMIs for the Nck1 protein

2. Methods

2.1. Library Generation and Preparation

AX-024 is an effective SMI of the Nck1 protein according to empirical evidence first presented by Borroto et al[12] in 2016. I used its backbone structure (Figure 2) as a scaffold template and permuted different combinations of the R1, R2, and R3 substituents, including the original functional groups and similar functional groups I designed to preserve the binding mode of AX-024. Importantly, all permutations were designed to be chemically accessible, ensuring that the final top molecules are feasible for synthesis and hence downstream experimental evaluation. In total, I generated a compound library consisting of 10,149 novel molecules with structural and physicochemical similarity to AX-024, all of which were considered as potential binder to the Nck1 protein. For tracking purposes, the SMIs were labeled sequentially in the form of d001 or d011 for one- or two-digit numbers and in the form of d100, d1000, d10000 for three-, four-, and five-digit numbers respectively. Some indices were skipped because duplicate molecules were removed during library curation.

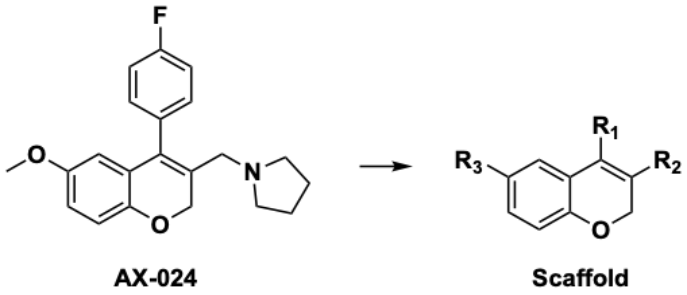

Figure 2. The molecular structure of AX-024 and its scaffold for compound library generation

2.2. Molecular Docking and Drug-Like Screening

The structure of the Nck1 protein was obtained from the RCSB PDB database (https://www.rcsb.org/, code: 2JW4) and curated by removing crystal water molecules and co-crystallized ligands. A cubic docking grid of 2 nm length was defined and centered at the center of mass of the residues W41, Q22, K39, and F53, the binding-site residues of Nck1 as presented in Borroto et al., 2016.[12] These residues were specified as flexible during the docking calculations to permit local conformational adjustments and account for induced-fit effects.

AutoDock Vina[45–47] was used to perform docking of all designed molecules and AX-024 as the reference. Since this brute-force docking step served primarily as an initial filter, the exhaustiveness parameter was set to 8 to balance computational efficiency and adequate conformational sampling. For each molecule, the lowest docking score was taken as the final docking score, and the corresponding binding pose was used for subsequent modeling and analysis. Once I obtained the docking score of the whole library, I applied a cutoff at -7.5 kcal/mol to prioritize the selected 923 molecules (9.10% of the library) as the top-scoring candidates for further screening and focused evaluation. This initial pass offers an assessment of binding capability across the entire library and establishes a baseline to narrow down the chemical space for subsequent filtering and prioritization. As in my expectations, all designed molecules were able to dock to the binding site of the Nck1 protein and adopted similar binding conformations.

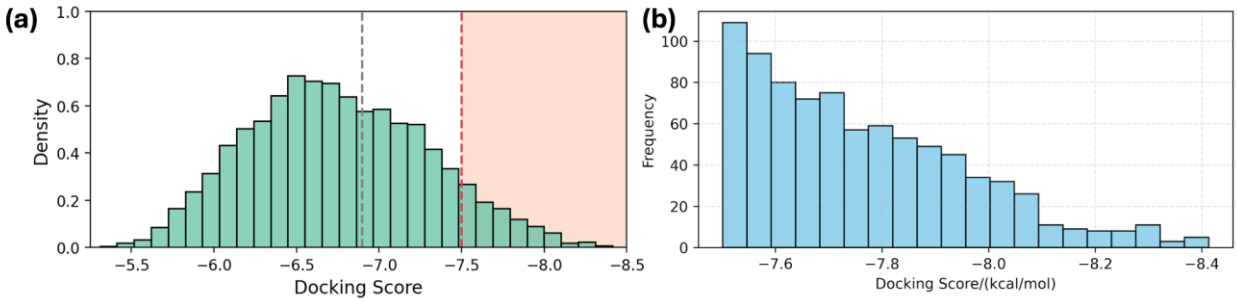

Figure 3. a. Distribution of docking scores of the 10,149 potential SMIs, with gray dotted line marking AX-024 (-6.898 kcal/mol), and area shaded in red marking selected section for further analysis; b. distribution of docking scores of the 923 SMIs with docking scores under -7.5 kcal/mol

2.3. Pan-Assay Interference Compounds Filtering

Prior to MD simulations for the filtered compounds, I applied the Pan-Assay Interference Compounds (PAINS) filter implemented in RDKit[42] introduced by Baell and Holloway in their 2010 study[43] to exclude molecules that are considered unfavorable for drug development. From the analysis, only one PAINS moiety emerged in the subset, anil_di_alk_C(246) (Figure 4(a)). It represents a redox-active aniline that is prone to nonspecific covalent or oxidative reactions with biomolecules such as proteins. Such assay interference motifs can produce false-positive signals in biochemical assays and/or cause harm to the human body and compromise the reliability and/or results of downstream drug development efforts.[44] This motif occurred primarily at the R3 position, and two R-group designs containing this functionality were consequently removed from the library

(Figure 4(b)). Applying this filter, I obtained 840 compounds for the next layer of selection - MD simulations to further refine their rankings and select the top candidates.

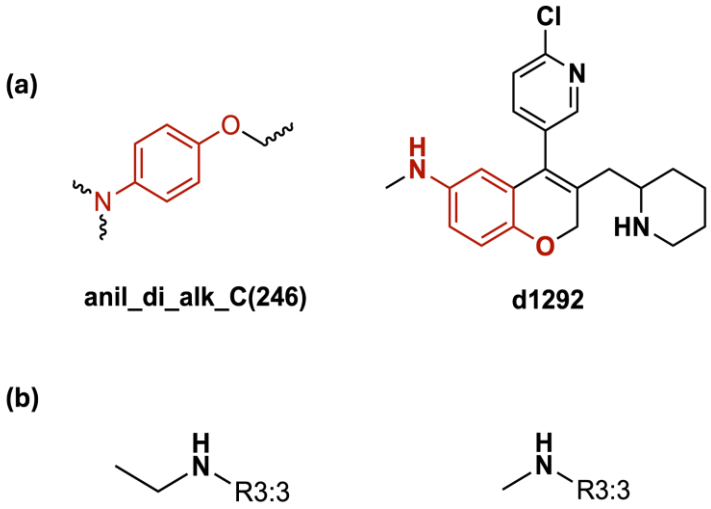

Figure 4. (a) The molecular pattern anil_di_alk_C(246) in the PAINS library, the one that is identified in my library (left) and an example SMI contains the pattern that was removed (right); (b) the two R3 structures that are in effect deleted after PAINS filtering

2.4. MD Simulations

Before simulations, the selected small molecules were re-docked into the protein with a higher exhaustiveness of 128 to obtain more accurate initial binding positions. All MD simulations were performed using the GROMACS 2024.4 package.[48] The AMBER99SB force field was used for the protein and the second-generation General AMBER Force Field (GAFF2) was employed for the ligand atoms using ACPYPE.[49–51] Each protein-ligand complex was placed in a cubic box with a length of 7 nm and then solvated using the TIP3P water model.[52] Sodium and chloride ions were added at a concentration of 0.15 M to mimic a physiological ion strength while ensuring charge neutrality. Energy minimization was carried out using the steepest descent algorithm to remove any interatomic forces larger than 1000 kJ/mol/nm^2, followed by a 200 ps NVT equilibration. Subsequently, a 20 ns MD simulation was performed in an NPT ensemble to collect samples for free energy calculation. The leap-frog integrator was employed to integrate Newton's equations of motion with a 2 fs time step. The long-range electrostatics were treated with the particle mesh Ewald (PME) algorithm.[53] A cut-off distance of 1.0 nm was used for van der Waals and short-range electrostatic interactions. All bonds involving hydrogen atoms were constrained using the LINCS (Linear Constraint Solver) algorithm.[54] In all stages, the temperature and pressure were maintained at 300 K and 1 bar using the a velocity-rescale thermostat and a cell-rescale barostat, respectively. Energy and simulation trajectories were saved every 100 ps. Through sequential docking and PAINS filters, the total group was reduced from 10,149 candidates to 840 compounds for MD relaxation and free energy calculations. Binding free energies were estimated by the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) theory in the gmx_MMPBSA package from the final 5 ns simulation consisting of 50 frames.[55, 56] A multiple frame strategy was employed to obtain both the average values and uncertainties. These calculations were successfully implemented for both the reference ligand AX-024 and the 840 selected molecules. Both the mean binding energy and its standard deviation were evaluated to identify SMIs with the most favorable and stable binding profiles. In addition, the root mean squared deviation (RMSD) was calculated using MDAnalysis from the trajectories as an additional indicator of the binding stability of ligands to further filter the molecules.[57, 58] Finally, the final proposed top 10 candidates were subjected to MD simulations extended to 100 ns to ensure the convergence of the free energy estimation. After extension, I conducted further examination of the binding poses, RMSD, root mean square fluctuation (RMSF) and other properties of the top 10 molecules.

3. Results and Discussion

3.1. AX-024 – Nck1 Complex and Compound Library Construction

Prior to constructing the compound library, I first examined the binding mode and molecular interactions of AX-024 in the Nck1 protein in detail. The bound complex was constructed using molecular docking with AutoDock Vina which successfully reproduced the reported binding mode from Borroto's work in 2016[12] (Figure 5(a)). The aromatic R1 substituent of AX-024 forms extensive π-π stacking with Phe53 and Trp41, and the central phenyl core stacks with Phe53. It underscores the importance of aromatic contacts in stabilizing the binding of AX-024 to the protein pocket. The basic amine in the R2 group is protonated and forms a positive charge center which establishes a strong hydrogen bond with the amide oxygen of Gln20 and a salt bridge with the carboxylate of Glu23. Moreover, the protonated amine also participates in the formation of cation-π interactions, which, along with all other favorable interaction modes, constitutes the important molecular forces that promote the binding of AX-024 onto the Nck1 protein pocket. This analysis suggests that both an aromatic substituent at R1 and a protonatable nitrogen at R2 are preferred features for productive binding.

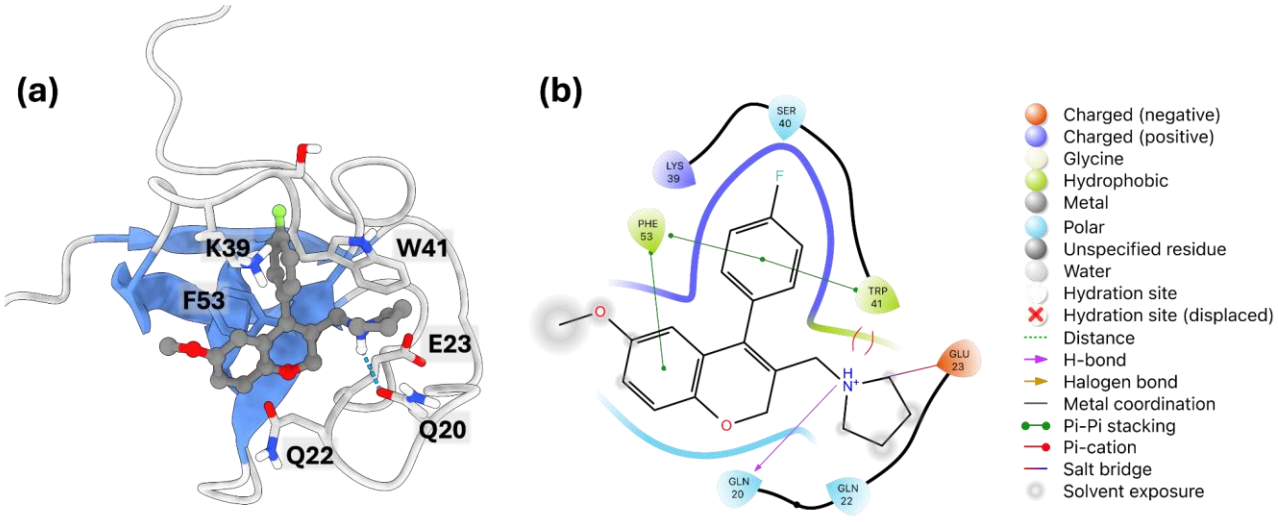

Figure 5. Binding interactions of the AX-024 with the Nck1 protein; (a) 3D structural view showing the ligand (stick representation) bound to the protein pocket (cartoon structure). The key residues, K39, W41, F53, E23, Q20, Q22, are labeled and highlighted; (b) 2D interaction diagram depicting the molecular interactions

Guided by these insights, I constructed my library by enumerating R-group modifications around the AX-024 scaffold with diverse functional groups while preserving AX-024's key interaction features. In this way, the resulting SMIs were expected to retain the general binding mode of AX-024; meanwhile, the introduced structural diversity could allow for the discovery of changes to the R groups that allow for the high potency of binding. In the R1 position, I introduced aromatic substituents that covered a range of ring structures as well as different halogen atoms in place of AX-024's fluorine. In the R2 position, I predominantly employed an aliphatic ring with at least one sp^3 nitrogen as the basic amine for protonation and promoting hydrogen bonding and/or salt-bridge formation with residues Q20 and E23. Beyond these motifs, I added substitutions (e.g., cyclohexanyl or pyridinyl rings) to further expand the explored chemical space and more precisely resolve the structure-property relationship by emphasizing the impact of the loss of aforementioned interactions in the binding mode. Full lists of the R groups are shown in Figures S1-S3. The R groups' permutations generated a library of 10,149 molecules.

3.2. Results of Candidate SMIs

3.2.1. Correlation Between Docking Score and Free Energies

Compared with docking, free energy calculations from MD simulations offer a more reliable measure of ligand-protein interactions. Using the calculated ΔG values, I first examined the correlation between the free energies and the docking scores. Figure 6(a) reflects the distribution of docking scores and ΔG's of the 840 molecules that remained after two layers of selection. The ΔG values exhibit significant aggregation around -20 to -15 kcal/mol, while more favorable candidates are evenly distributed between -20 and -32 kcal/mol. No such aggregation exists for docking scores (Figure 3(a)). Although no clear direct linear or any other form of a strong simple relation between docking scores and ΔG values exists nor does any aggregation of ΔG and docking scores correlate, I observe that, apart from a dip at -8.3 kcal/mol that may result from a specific group of molecules favored in docking not being favored in MD, lower docking score groups generally correspond to lower average free energies (see Figure 6(b)), indicating that they generally contain more favorable candidates compared to the weaker SMIs clustering near -17 kcal/mol. These results prove that docking score is still a relevant parameter and remains practical for prioritizing well-performing SMIs when computational power is limited. Hence, as illustrated in Figure 6(b), selecting candidates with lower scores generally improves the likelihood of identifying effective SMIs and enhances the efficiency of the selection process.

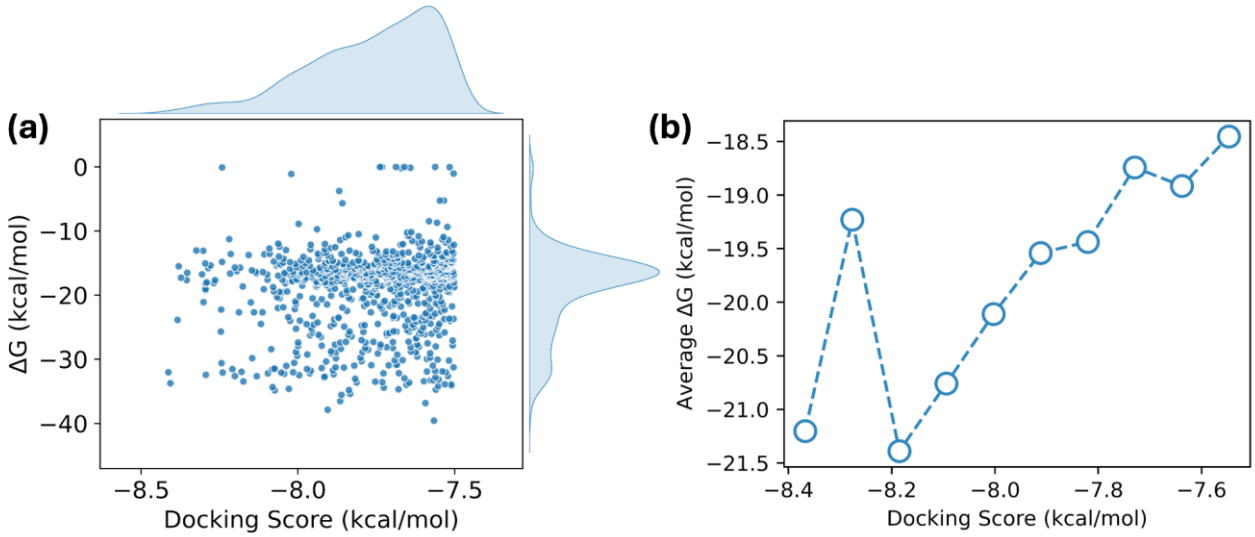

Figure 6. (a) Scatter of all the docking scores and ΔG pairs; (b) Relationship of score interval and average ΔG of SMIs in that docking score interval

3.2.2. Top 10 Candidates and Ligand Stability

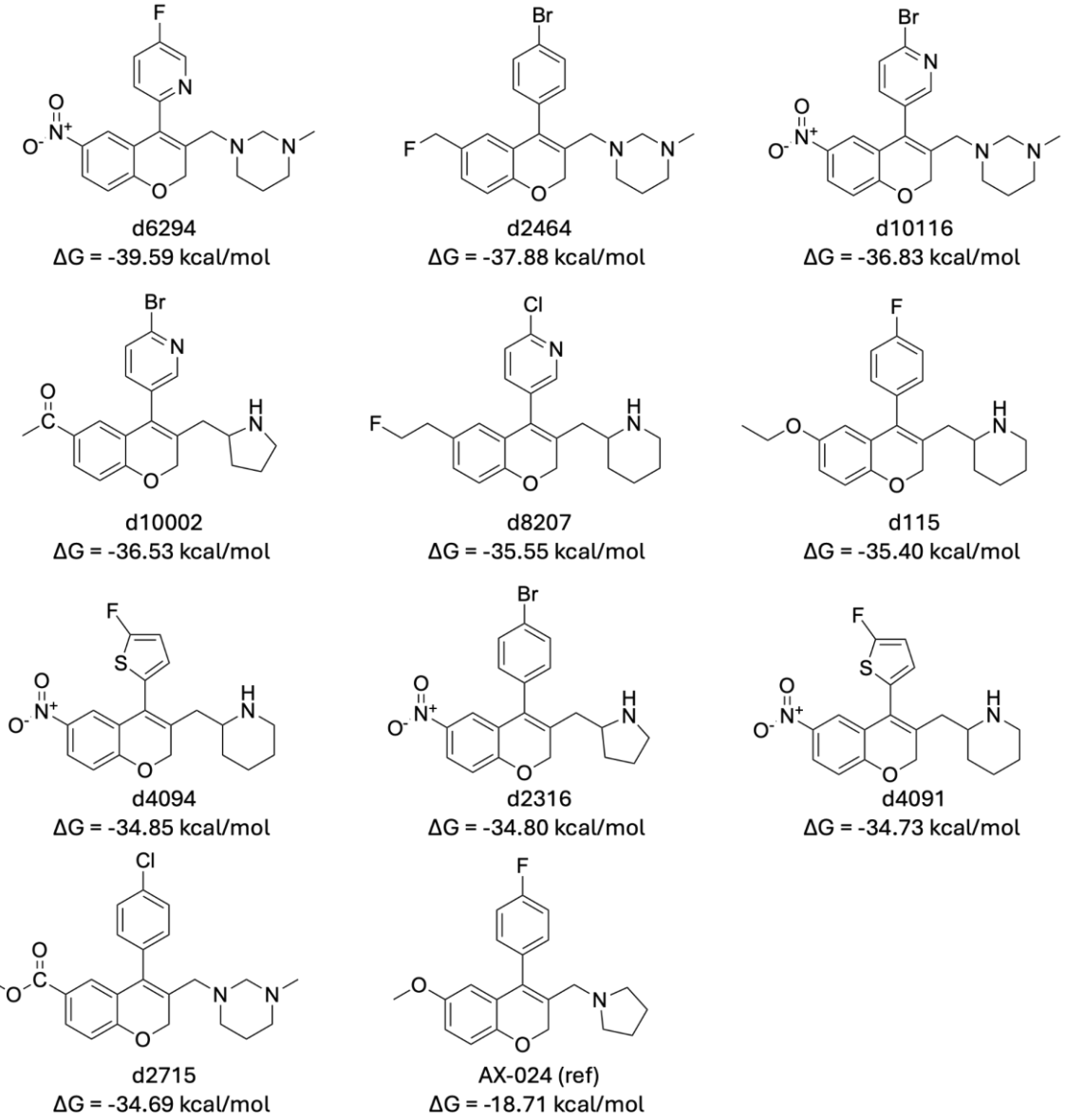

Figure 7. The chemical structures of the top 10 molecules with the lowest binding free energy from 20 ns MD simulations; AX-024 is listed as well as reference

To further analyze the outcomes of the screening of this stage, I selected the top ten molecules with the lowest predicted binding free energies after excluding all molecules with a ΔG standard deviation greater than 5 kcal/mol (structures presented in Figure 7). Among the top candidates, halogen substituents positioned on distal positions on aromatic rings relative to the core in the R1 group are observed frequently. To my anticipation, all candidates incorporate basic amines at the R2 position which enables favorable cation-π or hydrogen-bonding interactions with the Nck1 binding pocket. Consistent with these structural features, all top molecules exhibited highly favorable binding free energies, with several candidates such as d6294 bettering AX-024 by as much as 20 kcal/mol.

Next, I extracted RMSD profiles of the top ligands from the simulation trajectories to compare their conformation flexibility (Figure 8(a)). RMSDs were calculated excluding the initial 5 ns to allow each SMI to relax into the binding pocket from the approximate pose suggested by AutoDock Vina. Since the Nck1 binding pocket is located on the protein surface (see Figure 5(a)), SMIs bound in this region may exhibit high mobility, leading to comparatively elevated RMSD values. Nevertheless, the selected 10 SMIs display relatively low average RMSDs values lower than 8 Å, with a list of molecules possessing values lower than 4 Å, indicating stable binding of the ligands (Figure 8(b)).

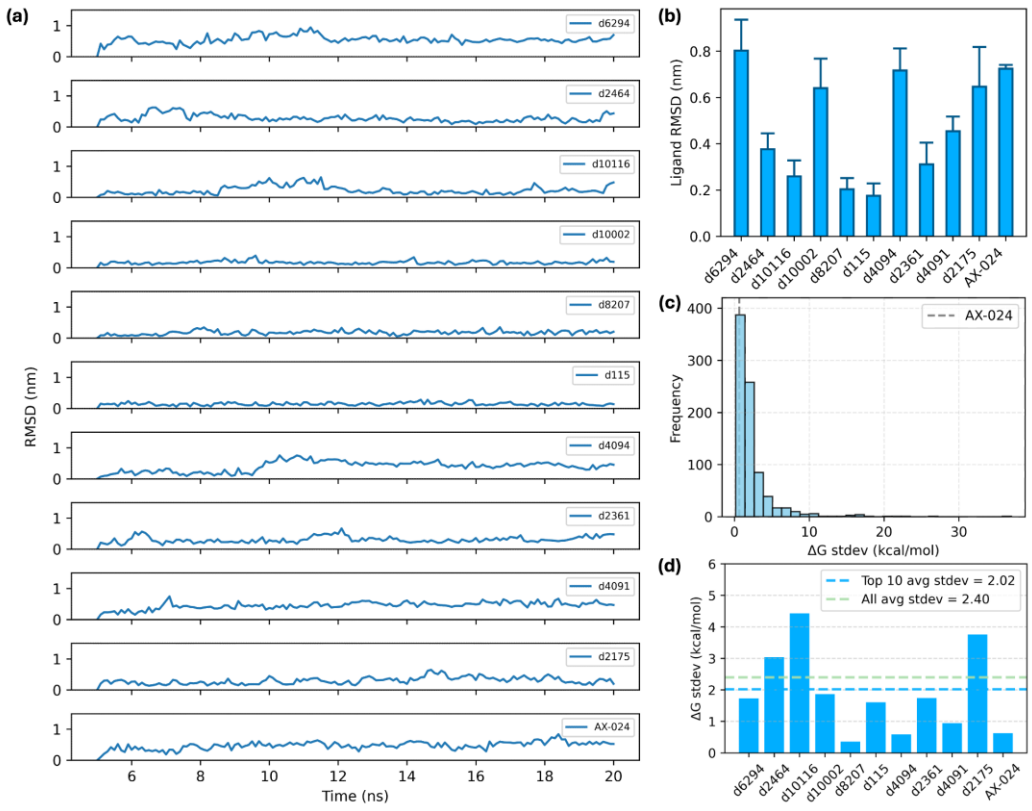

Figure 8. (a) RMSD profiles of the top 10 molecules in the 20 ns time period. b. Average RMSD values of the top 10 molecules; (c) Standard deviation of ΔG in the 200 frames used to calculate free energies by gmx_MMPBSA for the 840 SMIs; (d) Standard deviation of ΔG of the top 10 molecules

A closer examination of their ΔG standard deviations further supports their stability. Figure 8(c) shows that the vast majority of the candidates exhibit relatively low standard deviations, indicating that most of the molecules adopt stable binding poses in the binding pocket and that the obtained binding energies are reliable reflecting physical interactions rather than transient fluctuations. Notably, while compound d10116 shows a ΔG standard deviation slightly above 4 kcal/mol, near the upper acceptable limit, its consistently low RMSD across the simulation suggests that it remains a reliable candidate among the top SMIs (Figure 8(d)). By ranking the SMIs using ΔG values and excluding those with a ΔG standard deviation greater than 5 kcal/mol, these top 10 candidates with the lowest binding free energies passed the filtering and were retained to extended MD simulations to further refine the results and analyze the binding interactions.

3.3. Extended Analysis of Top Molecules

After extending the MD simulations to 100 ns, I recalculated ΔG values and standard deviations for the top 10 molecules using 50 frames from the final 5 ns of the trajectories, thereby ensuring a fair comparison of uncertainties. For the majority of the candidates, the ΔG values decreased compared to those from the 20 ns simulations (Figure 9(a)). Standard deviations in the free energy measure also declined, especially for compounds that initially displayed high uncertainties (>3 kcal/mol) (Figure 9(b)). These reflected reduced temporary fluctuations and stabilized binding poses over longer timescales. Importantly, several candidate molecules, including d6294, d2361, d10002, d4091, and d8207, showed consistent ΔG values across both 20 ns and 100 ns simulations, underscoring their reliability as potent binders to Nck1. Ligand RMSD values increased slightly in general but remained within a stable range, confirming their persistent binding within the protein pocket (Figure 9(c)). Collectively, these results identify a subset of molecules with converged favorable ΔG values and low uncertainties, supporting their status as the most promising SMIs from my library for subsequent experimental validation.

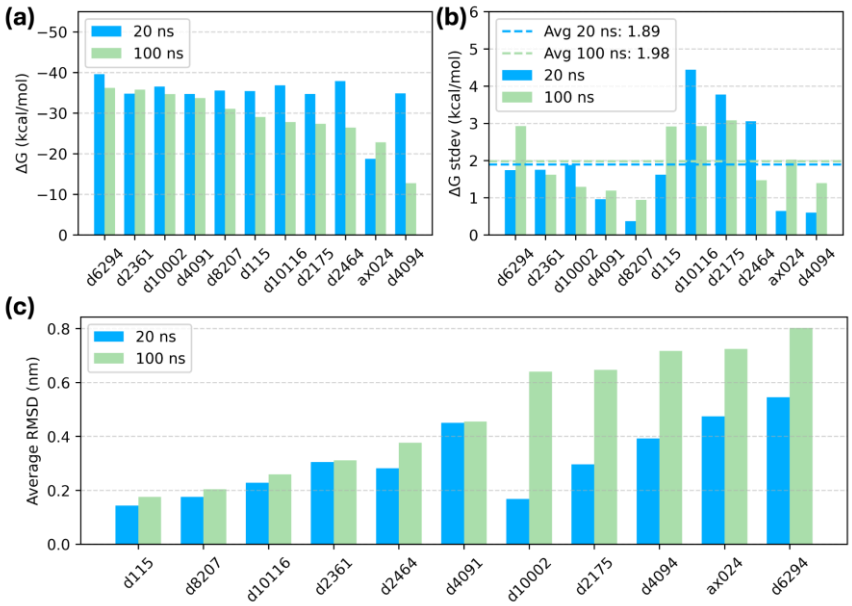

Figure 9. Comparison of ΔG ((a)), ΔG's standard deviation ((b)), and average RMSD ((c)) calculated for top molecules from 20 ns and in 100 ns MD simulations

I next sought to analyze the binding poses of the top molecules and understand the molecular forces that led to their stable binding to the Nck1 protein pocket. Among the three substituent positions, the R2 group shows the most interesting and influential-to-results features with all top candidates incorporating one of the three R2 motifs highlighted in Figure 10, which strongly suggests their central role in driving favorable binding interactions. By contrast, no clear preference is observed for the R3 group, particularly since the aniline-like moiety was excluded as a PAINS liability. The R1 groups, however, consistently feature a para-halogen atom on a six-membered aromatic ring or a halogen not adjacent to the core on a five-membered one, a recurring motif that is also present in the reference compound AX-024. Collectively, these structural patterns point to common binding mechanisms among the highest-ranking molecules.

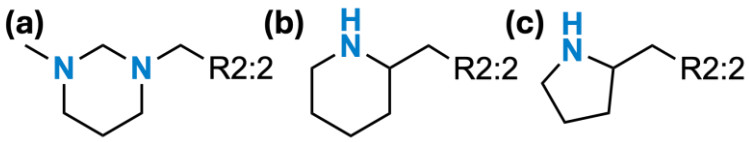

Figure 10. The three R2 groups observed in the top candidates

Candidates d6294, d2464, d10116, and d2175 adopt the R2 group shown in Figure 10(a) (molecules with this group may be referred to as Group A from now on). Looking at the diagram of the binding interactions of d6294 (Figure 11(a)), the SMI with the lowest ΔG scores among Group A SMIs and also all SMIs, it is observed that both nitrogen atoms in R2 are protonated and form two strong salt bridges with the negatively charged carboxylic acid group on E23. This results in substantially stronger binding within the pocket for Group A molecules compared to the single salt bridge observed in AX-024 (Figure 5(b)). Although AX-024 also engages in hydrogen bonding and cation-π interactions, the presence of an additional salt bridge can provide a stronger stabilizing force, as salt bridges involve direct electrostatic attraction between fully charged groups, whereas hydrogen bonds and cation-π interactions rely on partial or delocalized charges that typically contribute less to binding energy. The positioning of the two sp^3 nitrogens is also constrained: if they were separated by either two or zero carbon atoms on the ring, their hydrogens would point towards opposite directions, weakening or even eliminating one of the salt bridges and thereby resulting in weaker binding to the pocket. This is supported by the observations that adjacent sp^3 nitrogens in the R2 group show reduced basicity (Figure S4) and protonation and that no structure with sp3 nitrogen atoms separated by two carbon atoms showed up in the filtered molecules from the previous screen.

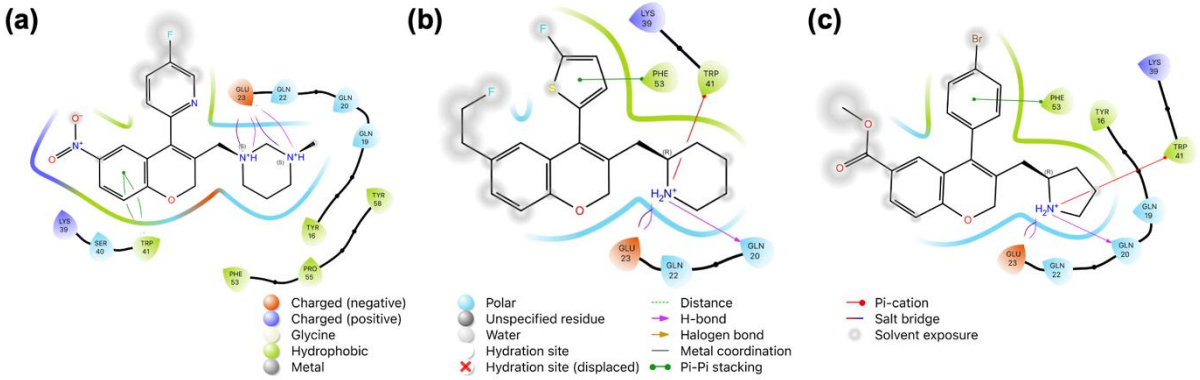

Figure 11. Diagram of binding of Nck1 with SMIs d6294 ((a)), d4091 ((b)), and d2361 ((c)), generated using Maestro Viewer[59]

Candidates d8027, d115, d4094, and d4091 share the R2 group shown in Figure 10(b) (Group B); meanwhile, d10002 and d2361 share the R2 group in shown Figure 10(c) (Group C). Both of these R2 groups feature a nitrogen at the neighboring position to the carbon connecting to the scaffold. In Figure 11(b) (d4091, lowest ΔG from 100 ns simulation in group B) and Figure 11(c) (d2361, lowest ΔG from 100 ns simulation in group C), I observe that due to this similarly positioned basic amine, both groups can feature on the two hydrogens on the nitrogen a salt bridge with E23, hydrogen bonding with amine on Q20, and cation-π interactions with W41's aromatic system (Figure 11(b)(c)). Comparing groups B and C, while both d10002 and d2361 feature all three different kinds of interactions (salt bridge, cation-π, hydrogen bonding), none of the Group B SMIs feature all three of them apart from d4091, possibly because the larger size of the six-membered piperidine ring (Figure 10(b)) compared with the five-membered pyrrolidine ring (Figure 10(c)) makes the former form stronger steric hindrance and blocks off certain amino acids from binding with the hydrogens on the protonated nitrogen. This also results in consistently better performance in ΔG scores in group C SMIs compared to group B in Figure 9(a). Nonetheless, most molecules in both Groups B and C outperform AX-024, in some cases by large margins, despite relying on similar types of intermolecular forces. Their advantage may be attributed to the presence of two hydrogens on the secondary amine, compared to only one in AX-024. The additional hydrogen may provide an extra hydrogen-bond donor to enable stronger or more versatile binding angles with other functional groups in the protein pocket. As a result, the intermolecular forces in Groups B and C reach binding energies that, while slightly weaker, approach those of the dual salt bridges observed in Group A.

3.4. Large Scale Analysis of Structure-Property Relationship

The relative stability of different structural regions within the top candidates can be assessed through the RMSF profiles of the ten SMIs (Figure S5). In general, all the ligands exhibit stable binding and low RMSF values of less than 2 Å for all heavy atoms; the core scaffold and the R1 substituents usually show the lowest fluctuations, indicating their relatively rigid conformation and stable binding throughout the simulations. This stability is consistent with their position in the Nck1 binding pocket, where they form nonchanging hydrophobic interactions and potential π-π stacking with residues. The R2 groups, however, display more heterogeneous behavior. In molecules such as d6294, d2464, d10116, d2175, and d4094, R2 atoms exhibit comparatively more pronounced flexibility, whereas in the remaining five candidates they are more restrained. These differences suggest that the R2 group plays a context-dependent role, either contributing to stable binding through specific contacts or remaining more mobile when solvent-exposed, a behavior that may also be influenced by local flexibility within the protein pocket.

The R3 groups show the highest RMSF values and greater fluctuations than either the core or other substituents. This is likely due to their partial or complete exposure to the solvent in most cases (Figure 11), which leaves them more susceptible to interactions with surrounding solvent molecules

of random orientation and collisions. Finally, a distinctive feature is observed in Group A molecules (Figure 10(a)), where the R2 groups are relatively unstable compared to Group B and C molecules, as shown by the slightly elevated RMSF values. This instability suggests that, although the dual amine R2 groups in Group A molecules form strong electrostatic salt bridges, their charged and solvent-exposed nature leaves them less structurally restrained, resulting in elevated flexibility that may influence overall binding dynamics.

To better understand how chemical substituents shape binding affinity, I next aimed to recover correlations between ΔG values and the specific chemistries of the R groups. Figure 12 provides an overview of how different R groups influence the performance of the SMIs. In general, the average ΔG values associated with R1 and R3 groups show less variation than those for R2 groups. For example, the R2 group with Simplified Molecular Input Line Entry System (SMILES)[60–62] representation C1CNC(C[*:2])C1 (as in Figure 10(b)), which corresponds to the lowest (best) ΔG distribution among all R2 groups, achieves an average ΔG of -28.00 kcal/mol and a median of -30.91 kcal/mol. These values are substantially lower than the overall averages across all molecules (-19.17 kcal/mol and -17.23 kcal/mol, respectively). By contrast, the best-performing R1 and R3 groups only reach average/median ΔG values of -21.98/-19.25 kcal/mol and -20.64/-18.43 kcal/mol, respectively. This reinforces the observation that among the three substituent positions, the R2 group exerts the greatest influence on improving the predicted effectiveness of SMIs.

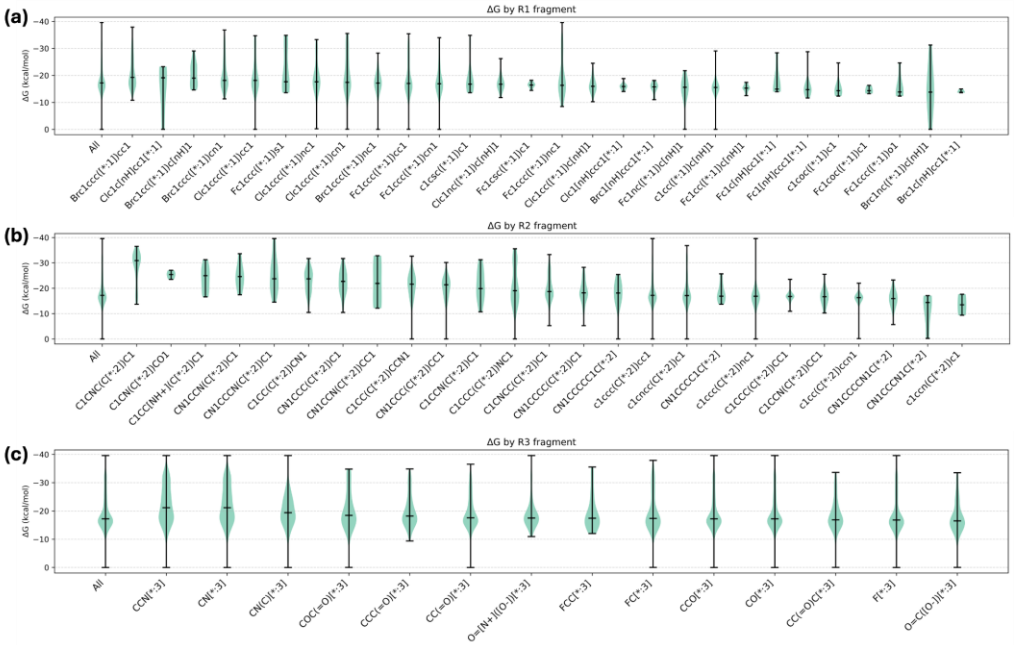

Figure 12. Violin plots showing the distribution of ΔG values of SMIs grouped according to the identity of R1 ((a)), R2 ((b)) and R3 ((c)) groups

Having established that the R2 group has the strongest influence on binding affinity, I next examined how specific physicochemical properties of R groups affect SMI performance (Figure 13). In R1, six-membered aromatic rings are slightly more favored over five-membered ones (Figure 13(a)), possibly due to their larger π systems and the increased distance they provide between the halogen substituent and the core. This trend is consistent with the pattern in Figure 13(b), where halogens positioned farther from the core are preferred over those directly adjacent to it. Both Figure 13(b) and Figure 13(c) further supports the observation that incorporating a halogen at R1 improves binding relative to unsubstituted rings. Among halogens, MD simulations indicate a preference for heavier atoms such as Br over Cl over F, which is likely due to their greater polarizability and ability to engage in hydrophobic interactions within the protein pocket. In contrast to R1 groups, R2 groups display a strong preference for five-membered rings over six-membered rings (Figure 13(d)) and it may arise from steric hindrance imposed by larger rings, as discussed previously for Groups B and C top candidates. Additional trends emerge when considering heteroatom content in the ring structure:

SMIs with two nitrogens in their R2 substituents consistently outperform those with a single nitrogen, which in turn show better performance than R2 groups lacking basic amines altogether (Figure 13(e)). Finally, R3 substituents show relatively minor effects on binding, with ΔG values being similar for groups containing oxygen versus fluorine (Figure 13(f)).

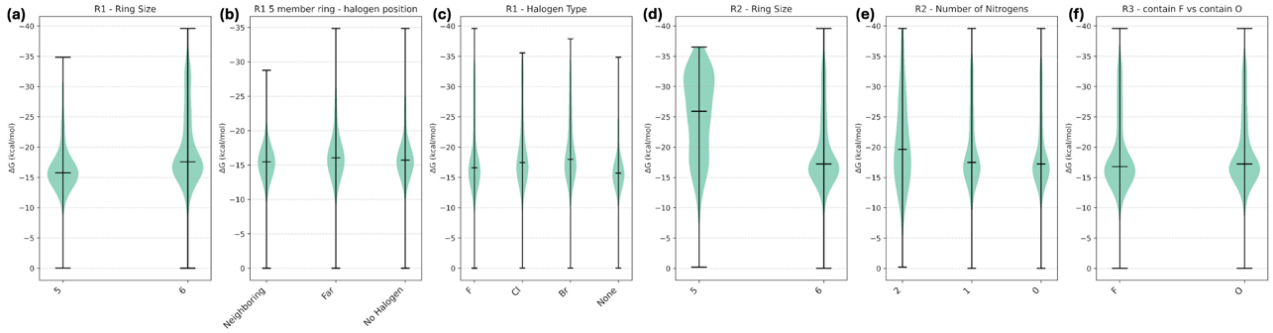

Figure 13. Violin plots of predicted binding free energies (ΔG) for SMIs grouped by R-group characteristics: (a) R1 ring size (five- vs six-membered), (b) R1 halogen position in five-membered rings (neighboring vs far vs none), (c) R1 halogen type (F, Cl, Br, none), (d) R2 ring size (five- vs six-membered), (e) number of nitrogen atoms in R2 (two, one, none), and (f) R3 substituents containing fluorine versus oxygen

4. Conclusion

In conclusion, this study successfully demonstrated the implementation of a multi-layer *in silico* workflow for the design, screening, and identification of SMIs with potential application to autoimmune diseases such as T1D. Starting from a library of 10,149 molecules generated through R-group enumeration of the experimentally validated SMI AX-024[30], I progressively narrowed down candidates through multiple filtering steps, including docking score evaluation, PAINS analysis, and MD-based binding free energy calculations. Most importantly, this screening workflow enabled the discovery of novel top-performing molecules with strong binding potential to Nck1 protein and the large-scale structural analysis for design rules for further drug design and optimization.

My analysis of both the top candidates and the full set of MD-simulated molecules provides valuable insight into the design and optimization of small molecules targeting Nck1. In particular, modifications to the R2 group introduced during the enumeration process emerged as a major driver of improved binding. Shifting the nitrogen atom to neighboring positions on the ring relative to its placement in AX-024, reducing the ring size to five members, and/or incorporating an additional nitrogen one atom apart were found to markedly enhance predicted binding affinity, as reflected by large improvements in binding free energy. Notably, candidates such as d6294, which incorporate such features, consistently outperformed AX-024 by significant margins in simulations. At the same time, the analysis also confirmed certain favorable structural characteristics of AX-024, such as the stabilizing contribution of its six-membered benzene ring. Together, these findings highlight both opportunities for structural innovation and the retained value of established design elements.

This work not only highlights a set of high-potential SMIs for synthesis and *in vitro* experimentation but also validates the multi-layered computational workflow featured as a broadly transferable strategy for drug design in other contexts. The top-performing candidates identified in this study stand out as promising leads for the development of next-generation therapeutics targeting autoimmune diseases related to the Nck1-CD3 signaling complex. Looking ahead, expanding the accessible chemical space through the generation of larger libraries, potentially accelerated by machine learning-guided molecular design[63], and incorporating enhanced sampling techniques[64] to obtain free energy estimates with higher accuracy, will enhance the precision of candidate selection. Together, these advances will not only enable the delivery of the most effective SMIs for *in vitro* validation but also pave the way for translation into *in vivo* studies, ultimately contributing to the development of more effective treatments for autoimmune and immune-related diseases.

References

[1] Pisetsky, D.S. (2023) Pathogenesis of autoimmune disease. Nature Reviews Nephrology, 19(8), 509–524. https://doi.org/10.1038/s41581-023-00720-1

[2] Wang, L., Wang, F.-S. and Gershwin, M.E. (2015) Human autoimmune diseases: a comprehensive update. Journal of Internal Medicine, 278(4), 369–395. https://doi.org/10.1111/joim.12395

[3] Katsarou, A., Gudbjörnsdottir, S., Rawshani, A., Dabelea, D., Bonifacio, E., Anderson, B.J., Jacobsen, L.M., Schatz, D.A. and Lernmark, Å. (2017) Type 1 diabetes mellitus. Nature Reviews Disease Primers, 3(1), 17016. https://doi.org/10.1038/nrdp.2017.16

[4] Marrack, P., Kappler, J. and Kotzin, B.L. (2001) Autoimmune disease: why and where it occurs. Nature Medicine, 7(8), 899–905. https://doi.org/10.1038/90935

[5] Song, Y., Li, J. and Wu, Y. (2024) Evolving understanding of autoimmune mechanisms and new therapeutic strategies of autoimmune disorders. Signal Transduction and Targeted Therapy, 9(1), 263. https://doi.org/10.1038/s41392-024-01952-8

[6] Hu, H., Zhang, G., Chen, T., Liu, Y., Meng, L., Holmdahl, R., Dai, L. and Zhao, Y. (2025) Immunosenescence in autoimmune diseases. Autoimmunity Reviews, 24(6), 103805. https://doi.org/10.1016/j.autrev.2025.103805

[7] Pisetsky, D.S. (2023) Pathogenesis of autoimmune disease. Nature Reviews Nephrology, 19(8), 509–524. https://doi.org/10.1038/s41581-023-00720-1

[8] Steinman, L. (2004) Immune Therapy for Autoimmune Diseases. Science, 305(5681), 212–216. https://doi.org/10.1126/science.1099896

[9] Mukherjee, A.B. and Zhang, Z. (2011) Allergic Asthma: Influence of Genetic and Environmental Factors. Journal of Biological Chemistry, 286(38), 32883–32889. https://doi.org/10.1074/jbc.R110.197046

[10] Rolla, G. (2023) Why Current Therapy Does Not Cure Asthma. Is It Time to Move Towards a One Health Approach? Journal of Asthma and Allergy, 16(null), 933–936. https://doi.org/10.2147/JAA.S429646

[11] Chatenoud, L. and Bluestone, J.A. (2007) CD3-specific antibodies: a portal to the treatment of autoimmunity. Nature Reviews Immunology, 7(8), 622–632. https://doi.org/10.1038/nri2134

[12] Borroto, A., Reyes-Garau, D., Jiménez, M.A., Carrasco, E., Moreno, B., Martínez-Pasamar, S., Cortés, J.R., Perona, A., Abia, D., Blanco, S., Fuentes, M., Arellano, I., Lobo, J., Heidarieh, H., Rueda, J., Esteve, P., Cibrián, D., Martinez-Riaño, A., Mendoza, P., Prieto, C., Calleja, E., Oeste, C.L., Orfao, A., Fresno, M., Sánchez-Madrid, F., Alcamí, A., Bovolenta, P., Martín, P., Villoslada, P., Morreale, A., Messeguer, A. and Alarcon, B. (2016) First-in-class inhibitor of the T cell receptor for the treatment of autoimmune diseases. Science Translational Medicine, 8(370), 370ra184-370ra184. https://doi.org/10.1126/scitranslmed.aaf2140

[13] Atalla, L., Linker-Israeli, M., Steinman, L. and Rao, N.A. (1990) Inhibition of autoimmune uveitis by anti-CD4 antibody. Investigative Ophthalmology & Visual Science, 31(7), 1264–1270.

[14] Raphael, I., Joern, R.R. and Forsthuber, T.G. (2020) Memory CD4+ T Cells in Immunity and Autoimmune Diseases. Cells, 9(3). https://doi.org/10.3390/cells9030531

[15] Gravano, D.M. and Hoyer, K.K. (2013) Promotion and prevention of autoimmune disease by CD8+ T cells. Abul Abbas: An epitome of scholarship, 45, 68–79. https://doi.org/10.1016/j.jaut.2013.06.004

[16] Musette, P. and Bouaziz, J.D. (2018) B Cell Modulation Strategies in Autoimmune Diseases: New Concepts. Frontiers in Immunology, Volume 9-2018. https://www.frontiersin.org/journals/ immunology/ articles/10.3389/fimmu.2018.00622

[17] Lee, D.S.W., Rojas, O.L. and Gommerman, J.L. (2021) B cell depletion therapies in autoimmune disease: advances and mechanistic insights. Nature Reviews Drug Discovery, 20(3), 179–199. https://doi.org/10.1038/s41573-020-00092-2

[18] Serra, P. and Santamaria, P. (2019) Antigen-specific therapeutic approaches for autoimmunity. Nature Biotechnology, 37(3), 238–251. https://doi.org/10.1038/s41587-019-0015-4

[19] Peakman, M. and von Herrath, M. (2010) Antigen-Specific Immunotherapy for Type 1 Diabetes: Maximizing the Potential. Diabetes, 59(9), 2087–2093. https://doi.org/10.2337/db10-0630

[20] Serr, I., Drost, F., Schubert, B. and Daniel, C. (2021) Antigen-Specific Treg Therapy in Type 1 Diabetes – Challenges and Opportunities. Frontiers in Immunology, Volume 12-2021. https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2021.712870

[21] Warshauer, J.T., Bluestone, J.A. and Anderson, M.S. (2020) New Frontiers in the Treatment of Type 1 Diabetes. Cell Metabolism, 31(1), 46–61. https://doi.org/10.1016/j.cmet.2019.11.017

[22] Ngoenkam, J., Schamel, W.W. and Pongcharoen, S. (2018) Selected signalling proteins recruited to the T-cell receptor–CD3 complex. Immunology, 153(1), 42–50. https://doi.org/10.1111/imm.12809

[23] Straus, D.B., Chan, A.C., Patai, B. and Weiss, A. (1996) SH2 Domain Function Is Essential for the Role of the Lck Tyrosine Kinase in T Cell Receptor Signal Transduction (*). Journal of Biological Chemistry, 271(17), 9976–9981. https://doi.org/10.1074/jbc.271.17.9976

[24] Isakov, N., Wange, R.L., Burgess, W.H., Watts, J.D., Aebersold, R. and Samelson, L.E. (1995) ZAP-70 binding specificity to T cell receptor tyrosine-based activation motifs: the tandem SH2 domains of ZAP-70 bind distinct tyrosine-based activation motifs with varying affinity. Journal of Experimental Medicine, 181(1), 375–380. https://doi.org/10.1084/jem.181.1.375

[25] Paensuwan, P., Hartl, F.A., Yousefi, O.S., Ngoenkam, J., Wipa, P., Beck-Garcia, E., Dopfer, E.P., Khamsri, B., Sanguansermsri, D., Minguet, S., Schamel, W.W. and Pongcharoen, S. (2016) Nck Binds to the T Cell Antigen Receptor Using Its SH3.1 and SH2 Domains in a Cooperative Manner, Promoting TCR Functioning. The Journal of Immunology, 196(1), 448–458. https://doi.org/10.4049/jimmunol.1500958

[26] Paensuwan, P., Hartl, F.A., Yousefi, O.S., Ngoenkam, J., Wipa, P., Beck-Garcia, E., Dopfer, E.P., Khamsri, B., Sanguansermsri, D., Minguet, S., Schamel, W.W. and Pongcharoen, S. (2016) Correction: Nck Binds to the T Cell Antigen Receptor Using Its SH3.1 and SH2 Domains in a Cooperative Manner, Promoting TCR Functioning. The Journal of Immunology, 196(11), 4833–4833. https://doi.org/10.4049/jimmunol.1600551

[27] Ngoenkam, J., Paensuwan, P., Preechanukul, K., Khamsri, B., Yiemwattana, I., Beck-García, E., Minguet, S., Schamel, W.W. and Pongcharoen, S. (2014) Non-overlapping functions of Nck1 and Nck2 adaptor proteins in T cell activation. Cell Communication and Signaling, 12(1), 21. https://doi.org/10.1186/1478-811X-12-21

[28] Ngoenkam, J., Paensuwan, P., Preechanukul, K., Khamsri, B., Yiemwattana, I., Beck-García, E., Minguet, S., Schamel, W.W. and Pongcharoen, S. (2014) Non-overlapping functions of Nck1 and Nck2 adaptor proteins in T cell activation. Cell Communication and Signaling, 12(1), 21. https://doi.org/10.1186/1478-811X-12-21

[29] Richter, K., Rufer, A.C., Muller, M., Burger, D., Casagrande, F., Grossenbacher, T., Huber, S., Hug, M.N., Koldewey, P., D'Osualdo, A., Schlatter, D., Stoll, T. and Rudolph, M.G. (2020) Small molecule AX-024 reduces T cell proliferation independently of CD3ϵ/Nck1 interaction, which is governed by a domain swap in the Nck1-SH3.1 domain. Journal of Biological Chemistry, 295(23), 7849–7864. https://doi.org/10.1074/jbc.RA120.012788

[30] Alarcon, B. and Borroto, A. (2020) Small molecule AX-024 targets T cell receptor signaling by disrupting CD3e-Nck interaction. Journal of Biological Chemistry, 295(29), 10076. https://doi.org/10.1074/jbc.l120.014338

[31] Richter, K., Rufer, A.C., Muller, M., Burger, D., Casagrande, F., Grossenbacher, T., Huber, S., Hug, M.N., Koldewey, P., D'Osualdo, A., Schlatter, D., Stoll, T. and Rudolph, M.G. (2020) Reply to Alarcon and Borroto: Small molecule AX-024 reduces T cell proliferation independently of CD3ε-Nck1 interaction at SH3.1. The Journal of biological chemistry, 295(29), 10077. https://doi.org/10.1074/jbc.RL120.014441

[32] Juraske, C., Wipa, P., Morath, A., Hidalgo, J.V., Hartl, F.A., Raute, K., Oberg, H.-H., Wesch, D., Fisch, P., Minguet, S., Pongcharoen, S. and Schamel, W.W. (2018) Anti-CD3 Fab Fragments Enhance Tumor Killing by Human γδ T Cells Independent of Nck Recruitment to the γδ T Cell Antigen Receptor. Frontiers in Immunology, Volume 9-2018. https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2018.01579

[33] Ramírez, Y., Laguna, M.F. and Holgado, M. (2022) New Label-Free Biosensing for the Evaluation of the AX-024 Inhibitor: Case Study for the Development of New Drugs in Autoimmune Diseases. Sensors, 22(3). https://doi.org/10.3390/s22031218

[34] Sadybekov, A.V. and Katritch, V. (2023) Computational approaches streamlining drug discovery. Nature, 616(7958), 673–685. https://doi.org/10.1038/s41586-023-05905-z

[35] Durrant, J.D. and McCammon, J.A. (2011) Molecular dynamics simulations and drug discovery. BMC Biology, 9(1), 71. https://doi.org/10.1186/1741-7007-9-71

[36] Ghattas, M.A., Al Rawashdeh, S., Atatreh, N. and Bryce, R.A. (2020) How Do Small Molecule Aggregates Inhibit Enzyme Activity? A Molecular Dynamics Study. Journal of Chemical Information and Modeling, 60(8), 3901–3909. https://doi.org/10.1021/acs.jcim.0c00540

[37] Santos, L.H.S., Ferreira, R.S. and Caffarena, E.R. (2019) Integrating Molecular Docking and Molecular Dynamics Simulations. In W.F. de Azevedo Jr., ed. Docking Screens for Drug Discovery. New York, NY: Springer New York. 13–34. https://doi.org/10.1007/978-1-4939-9752-7_2

[38] Guan, S., Wang, T., Kuai, Z., Qian, M., Tian, X., Zhang, X., Yu, Y., Wang, S., Zhang, H., Li, H., Kong, W. and Shan, Y. (2017) Exploration of binding and inhibition mechanism of a small molecule inhibitor of influenza virus H1N1 hemagglutinin by molecular dynamics simulation. Scientific Reports, 7(1), 3786. https://doi.org/10.1038/s41598-017-03719-4

[39] Alonso, H., Bliznyuk, A.A. and Gready, J.E. (2006) Combining docking and molecular dynamic simulations in drug design. Medicinal Research Reviews, 26(5), 531–568. https://doi.org/10.1002/med.20067

[40] Vidal-Limon, A., Aguilar-Toalá, J.E. and Liceaga, A.M. (2022) Integration of Molecular Docking Analysis and Molecular Dynamics Simulations for Studying Food Proteins and Bioactive Peptides. Journal of Agricultural and Food Chemistry, 70(4), 934–943. https://doi.org/10.1021/acs.jafc.1c06110

[41] Pathak, R.K., Kim, W.-I. and Kim, J.-M. (2023) Targeting the PEDV 3CL protease for identification of small molecule inhibitors: an insight from virtual screening, ADMET prediction, molecular dynamics, free energy landscape, and binding energy calculations. Journal of Biological Engineering, 17(1), 29. https://doi.org/10.1186/s13036-023-00342-y

[42] Landrum, G., Tosco, P., Kelley, B., Rodriguez, R., Cosgrove, D., Vianello, R., sriniker, Gedeck, P., Jones, G., Kawashima, E., NadineSchneider, Nealschneider, D., Dalke, A., tadhurst-cdd, Swain, M., Cole, B., Turk, S., Savelev, A., Vaucher, A., Wójcikowski, M., Take, I., Faara, H., Scalfani, V.F., Walker, R., Probst, D., Ujihara, K., Maeder, N., Monat, J., Lehtivarjo, J. and godin, guillaume (2025) rdkit/rdkit: 2025_03_5 (Q1 2025) Release. Zenodo. https://doi.org/10.5281/zenodo.16439048

[43] Baell, J.B. and Holloway, G.A. (2010) New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. Journal of Medicinal Chemistry, 53(7), 2719–2740. https://doi.org/10.1021/jm901137j

[44] Baell, J. and Walters, M.A. (2014) Chemistry: Chemical con artists foil drug discovery. Nature, 513(7519), 481–483. https://doi.org/10.1038/513481a

[45] Trott, O. and Olson, A.J. (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry, 31(2), 455–461. https://doi.org/10.1002/jcc.21334

[46] Eberhardt, J., Santos-Martins, D., Tillack, A.F. and Forli, S. (2021) AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. Journal of Chemical Information and Modeling, 61(8), 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203

[47] Forli, S., Huey, R., Pique, M.E., Sanner, M.F., Goodsell, D.S. and Olson, A.J. (2016) Computational protein–ligand docking and virtual drug screening with the AutoDock suite. Nature Protocols, 11(5), 905–919. https://doi.org/10.1038/nprot.2016.051

[48] Abraham, M., Alekseenko, A., Basov, V., Bergh, C., Briand, E., Brown, A., Doijade, M., Fiorin, G., Fleischmann, S., Gorelov, S., Gouaillardet, G., Gray, A., Irrgang, M.E., Jalalypour, F., Jordan, J., Kutzner, C., Lemkul, J.A., Lundborg, M., Merz, P., Miletic, V., Morozov, D., Nabet, J., Pall, S., Pasquadibisceglie, A., Pellegrino, M., Santuz, H., Schulz, R., Shugaeva, T., Shvetsov, A., Villa, A., Wingbermuehle, S., Hess, B. and Lindahl, E. (2024) GROMACS 2024.4 Manual. Zenodo. https://doi.org/10.5281/zenodo.14016613

[49] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics, 65(3), 712–725. https://doi.org/10.1002/prot.21123

[50] Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A. (2004) Development and testing of a general amber force field. Journal of Computational Chemistry, 25(9), 1157–1174. https://doi.org/ https://doi.org/10.1002/jcc.20035

[51] Casoria, M., Macchiagodena, M., Rovero, P., Andreini, C., Papini, A.M., Cardini, G. and Pagliai, M. (2024) Upgrading of the general AMBER force field 2 for fluorinated alcohol biosolvents: A validation for water solutions and melittin solvation. Journal of Peptide Science, 30(2), e3543. https://doi.org/ https://doi.org/10.1002/psc.3543

[52] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics, 79(2), 926–935. https://doi.org/10.1063/1.445869

[53] Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. The Journal of Chemical Physics, 98(12), 10089–10092. https://doi.org/10.1063/ 1.464397

[54] Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry, 18(12), 1463–1472. https://doi.org/ 10.1002/ (SICI)1096-987X(199709)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H

[55] Miller, B.R. 3rd, McGee, T.D.J., Swails, J.M., Homeyer, N., Gohlke, H. and Roitberg, A.E. (2012) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. Journal of chemical theory and computation, 8(9), 3314–3321. https://doi.org/10.1021/ct300418h

[56] Valdés-Tresanco, M.S., Valdés-Tresanco, M.E., Valiente, P.A. and Moreno, E. (2021) gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. Journal of Chemical Theory and Computation, 17(10), 6281–6291. https://doi.org/10.1021/acs.jctc.1c00645

[57] Gowers, R.J., Linke, M., Barnoud, J., Reddy, T.J.E., Melo, M.N., Seyler, S.L., Domański, J., Dotson, D.L., Buchoux, S., Kenney, I.M. and Beckstein, O. (2016) MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In S. Benthall and S. Rostrup, eds. Proceedings of the 15th Python in Science Conference. 98–105. https://doi.org/10.25080/Majora-629e541a-00e

[58] Michaud-Agrawal, N., Denning, E.J., Woolf, T.B. and Beckstein, O. (2011) MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. Journal of Computational Chemistry, 32(10), 2319–2327. https://doi.org/10.1002/jcc.21787

[59] Schrödinger Release 2025-3: Maestro (2025) New York, NY: Schrödinger, LLC.

[60] Weininger, D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of Chemical Information and Computer Sciences, 28(1), 31–36. https://doi.org/10.1021/ci00057a005

[61] Weininger, D., Weininger, A. and Weininger, J.L. (1989) SMILES. 2. Algorithm for generation of unique SMILES notation. Journal of Chemical Information and Computer Sciences, 29(2), 97–101. https://doi.org/10.1021/ci00062a008

[62] Weininger, D. (1990) SMILES. 3. DEPICT. Graphical depiction of chemical structures. Journal of Chemical Information and Computer Sciences, 30(3), 237–243. https://doi.org/10.1021/ci00067a005

[63] Xie, S., Zhu, H. and Huang, N. (2025) AI-Designed Molecules in Drug Discovery, Structural Novelty Evaluation, and Implications. Journal of Chemical Information and Modeling, 65(17), 8924–8933. https://doi.org/10.1021/acs.jcim.5c00921

[64] Bertazzo, M., Gobbo, D., Decherchi, S. and Cavalli, A. (2021) Machine Learning and Enhanced Sampling Simulations for Computing the Potential of Mean Force and Standard Binding Free Energy. Journal of Chemical Theory and Computation, 17(8), 5287–5300. https://doi.org/10.1021/acs.jctc.1c00177