

# Computational development of aryl sulfone compounds as potential NNRTIs

Andy Zhang<sup>1,2</sup>, Rohan Adwankar<sup>\*1,2</sup>, Aarya Morgaonkar<sup>\*1,2</sup>, Dhyana Patel<sup>1,2</sup>, Maanibha Sengupta<sup>1,2</sup>, Justin Choi<sup>1</sup>

<sup>1</sup> Irvington High School, Fremont, California

<sup>2</sup> Irvington Research Club, Fremont, California

\*These authors contributed equally to this work.

## SUMMARY

Human immunodeficiency virus (HIV) attacks the host's immune system and affects millions of people globally. Due to the high mutation rate of HIV, it is of great importance that novel drugs are found to combat new strains. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are allosteric inhibitors that bind to the HIV reverse transcriptase and prevent replication. Indolyl aryl sulfones (IAS) and IAS derivatives have been found to be highly effective against mutant strains of HIV-1 reverse transcriptase. Here, we analyzed molecules designed using aryl sulfone scaffolds paired to cyclic compounds (that have previously been found to be beneficial groups for NNRTIs) as potential NNRTIs through the computational design and docking of 100 novel NNRTI candidates. Moreover, we explored the specific combinations of functional groups and aryl sulfones that resulted in the NNRTI candidates with the strongest binding affinity while testing all compounds for carcinogenicity. We hypothesized that the combination of an IAS scaffold and pyrimidine would produce the compounds with the best binding affinity. Our hypothesis was correct as the series of molecules with an IAS scaffold and pyrimidine exhibited the best average binding affinity. Additionally, this study found 32 molecules designed in this procedure with higher or equal binding affinities to the previously successful IAS derivative 5-bromo-3-[(3,5-dimethylphenyl)sulfonyl]indole-2-carboxamide when docked to HIV-1 reverse transcriptase. These findings contribute to the search for novel NNRTIs as well as expanding knowledge regarding the characteristics of the NNRTI pharmacophore.

## INTRODUCTION

Human immunodeficiency virus (HIV) affects millions of people around the world. In 2020, there were reportedly 37.6 million people who were living with HIV and an estimated 690,000 people died from acquired immunodeficiency syndrome (AIDS)-related illnesses (1). HIV infection can progress into AIDS if the number of CD4<sup>+</sup> T cells falls below 200 cells per cubic millimeter of blood, which is notably less than the average 500-1,600 CD4<sup>+</sup> T cells per cubic millimeter in an uninfected person's immune system (2). HIV can be transmitted through bodily fluids from an infected individual to a healthy individual (3). The virus attacks CD4<sup>+</sup> T cells, which

have a significant role in the body's immune response, which is needed for producing antibodies that reduce the effect of a pathogen. This is why HIV/AIDS is a global concern; it can severely deteriorate an individual's immune response, making them more susceptible to typically harmless infections (4).

HIV is a retrovirus that stores its genetic information using RNA instead of DNA. HIV is shaped like a sphere and the outer envelope of the virus has glycoproteins, which stick out like spikes, allowing the virus to lock into CD4 expressed on a subset of T cells and facilitating its envelope to fuse with the cell membrane of the CD4<sup>+</sup> cell. The capsid of the virus holds the genetic information as well as the reverse transcriptase (RT) enzyme. The capsid enters the cell once the envelope of the virus envelope fuses with the cell membrane. Once the capsid is in the cell, reverse transcriptase turns the HIV RNA into DNA (5). The HIV DNA then enters the CD4<sup>+</sup> T cell nucleus with the help of the integrase enzyme. Once the viral DNA is in a CD4<sup>+</sup> T cell nucleus, HIV can replicate by using the CD4<sup>+</sup> RNA polymerase (6).

Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are crucial to treating HIV/AIDS because they can inhibit the creation of HIV DNA by binding to the RT enzyme in its allosteric site (7). The allosteric site is located in the hydrophobic pocket of the HIV RT (8). The FDA-approved NNRTIs to combat HIV are Doravirine, Efavirenz, Etravirine, Nevirapine, and Rilpivirine. The RT enzyme of HIV causes it to constantly mutate, which can increase its resistance to NNRTIs (3). This means new inhibitors need continually be found to combat these new forms of HIV.

Due to the large costs and time commitments associated with synthesizing molecules, computer-aided drug discovery is an efficient method of finding and analyzing potential drug targets for synthesis (9). Computer-aided small molecule drug discovery generally consists of the design, pharmacological screening for toxicity, and docking of ligands (10). This procedure paired with the design of large libraries of compounds allowed for high-throughput virtual screening, which analyzes hundreds of potential drug candidates and allows chemists to find the best targets for synthesis (11). Multiple techniques are used in the design of molecules for compound libraries, including pharmacophore analysis and scaffold-based drug discovery. By examining the geometric configuration of features that are relevant to binding, a pharmacophore, or a theoretical model for understanding binding to the hydrophobic pocket within HIV-RT, can be developed (either by using computer assistance or simply considering important features in the design) that can be used to predict molecules that will also bind well (12). Scaffold-based drug discovery is the creation of novel compounds utilizing core scaffolds paired

with functional groups that previous research has postulated could be effective in binding to the target site (13). Virtual pharmacological screening for toxicity could save time and resources in comparison to *in vitro* toxicity assays, improving the speed of the drug discovery process (14). Ligand docking is performed by simulating the docking interactions *in silico* to make a prediction of the binding affinity the ligand would have to the target site (15). This helps predict which ligands would bind to the target the most effectively and reliably. Binding affinity predicts the strength of the bond between a ligand and an organic molecule, in this case, the RT enzyme of HIV.

Indolyl aryl sulfone (IAS) derivatives have shown large amounts of promise as NNRTIs (16). Specifically, Silvestri and Artico concluded that the IAS derivative 5-bromo-3-[(3,5-dimethylphenyl)sulfonyl]indole-2-carboxamide was very effective against HIV-1 mutants, which had NNRTI resistant mutations on the RT at amino acid positions 103 and 181 (17). This warrants further research into aryl sulfones as scaffolds for potential NNRTIs. We selected indolyl aryl sulfones, pyrrolyl aryl sulfones, phenyl vinyl sulfones, phenyl vinyl sulfonates, and benzyl vinyl sulfonates as scaffolds based on their success in prior studies (Figure 1) (17). Utilizing functional groups from previously successful molecules can increase the probability of success of the novel molecules (18). Upon examination of the FDA-approved NNRTIs Efavirenz, Etravirine, and Rilpivirine, we selected benzoxazine and pyrimidine as functional groups for this study (19). Second-generation NNRTIs are successful due in part to the larger aromatic systems, which improve the flexibility and hence the binding of the molecule, so we additionally selected styrene as a functional group (7).

We hypothesized that the combination of the IAS scaffold and pyrimidine would produce the set of molecules with the best average binding affinity to HIV-1 RT as compared to the array of other combinations of aryl sulfone scaffolds and functional groups. This was based on three pieces of evidence. Firstly, the scaffold results from Silvestri *et al.* concluded that the indolyl aryl sulfone derivatives worked best (17). Secondly, the aromatic properties of the pyrimidine ring may facilitate an effective binding. Third, pyrimidine is present in both Etravirine and Rilpivirine, while the other functional groups covered in this paper were only found in one or fewer FDA-approved NNRTIs. Our results are in line with our hypothesis since our third series, which was based on this combination, had the strongest average binding score.

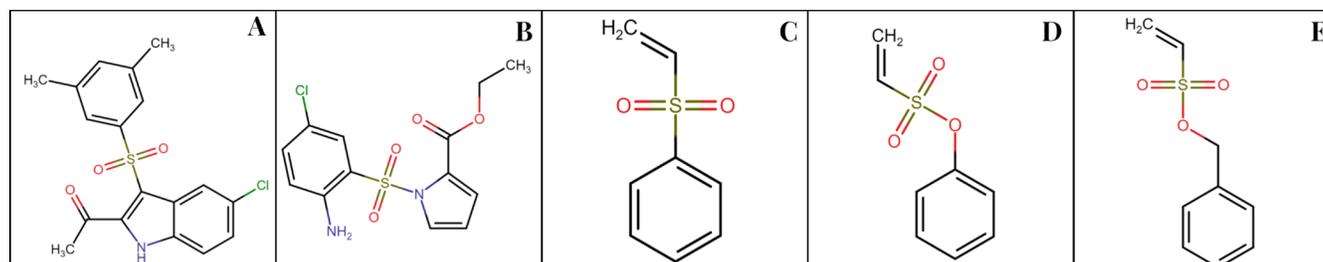
## RESULTS

We computationally designed, screened, optimized, and

docked one hundred novel ligands to the allosteric binding pocket of HIV-1 RT. Each molecule was formed using an aryl sulfone scaffold and a cyclic compound added as a functional group. We divided our research into 10 series, named alphabetically, each with a different variation to determine what combination would have the best binding scores, as we used functional groups and scaffolds based on our literature review (Table 1). To interpret the binding affinities of these novel ligands, the more negative the value, the more adept the ligand is at binding to the specific protein. As our control, the indolyl aryl sulfone derivative found in the Silvestri paper had a binding affinity of -7.8 kcal/mol; any molecule with an affinity better than that value was considered a good NNRTI candidate and selected for further analysis.

Series A is an indolyl aryl sulfone scaffold and a piperazine functional group, with an average binding affinity of -7.86 kcal/mol. Series B is a pyrrolyl aryl sulfone scaffold and a styrene functional group with an average binding affinity of -7.12 kcal/mol. Series C is an indolyl aryl sulfone scaffold and a pyrimidine functional group with an average binding affinity of -8.23 kcal/mol. Series D is a pyrrolyl aryl sulfone scaffold and a pyrimidine functional group, which has an average binding affinity of -7.22 kcal/mol. Series E was composed of an indolyl aryl sulfone with a benzoxazine group, which had an average binding affinity of -7.71 kcal/mol. Series F is a pyrrolyl aryl sulfone scaffold and a benzoxazine functional group with an average binding affinity of -6.95 kcal/mol. Series G is a phenyl vinyl sulfone scaffold with a styrene functional group which has an average binding affinity of -7.01 kcal/mol. Series H is a phenyl vinyl sulfonate scaffold with a pyrimidine functional group which has an average affinity of -6.08 kcal/mol. Series I is a Benzyl vinyl sulfonate with a benzoxazine functional group with an average of -6.39 kcal/mol. Series J is a IAS scaffold with a piperazine functional group with an average binding affinity of -8.15 kcal/mol.

The best performing series overall was Series C with an average binding affinity of -8.6 kcal/mol, as compared to the mean binding affinity of all the compounds, which was -7.3 kcal/mol. The aforementioned benchmark ligand 5-bromo-3-[(3,5-dimethylphenyl)sulfonyl]indole-2-carboxamide had a binding affinity of -7.8 kcal/mol when docked to HIV-1 RT using the same method as the rest of the screening. Comparing the binding scores of the aryl sulfone compounds to the benchmark, 32 of the 100 ligands performed better or equal in binding affinity (17). The most successful ligand generated was from Series A and is N-{5-chloro-3-[(3,5-dimethylphenyl)sulfonyl]-1H-indole-2-carbonyl}-1-[[{(2S)-piperazin-2-yl]carbamoyl]amino]formamide using an IAS scaffold and a piperazine



**Figure 1: Aryl sulfone scaffolds.** We selected the scaffolds (A) indolyl aryl sulfone, (B) pyrrolyl aryl sulfone, (C) phenyl vinyl sulfone, (D) phenyl vinyl sulfonate, and (E) benzyl vinyl sulfonate for this study. An indolyl aryl sulfone derivative had performed well against HIV-1 reverse transcriptase, warranting further research.

Series of Derivative	Name	Binding Affinity (kcal/mol)	Scaffold	Functional Group
A	(S)-5-chloro-3-((3,5-dimethylphenylthio)-N-(piperazin-2-yl carbamoyl)carbamoyl)-1H-indole-2-carboxamide	-9.4	Indoyl Aryl Sulfones	Piperazine
B	ethyl 5-chloro-8,8-dioxo-21-oxa-8 $\lambda^4$ -thia-1,9-diazapentacyclo[10.8.2.0 $^2$ .7.0 $^2$ .0 $^3$ .0 $^3$ .0 $^3$ ]docosane-2,4,6,10,12(22),13(18),14,16,19-nonaene-10-carboxylate	-7.7	Pyrrl Aryl Sulfones	Styrene
C	(3-(3,5-bis(pyrimidin-2-yl)methyl]benzenesulfonyl)-5-chloro-1H-indole-2-carbonyl)urea	-9.1	Indoyl Aryl Sulfones	Pyrimidine
D	1-(2-amino-5-chlorobenzenesulfonyl)-1H-pyrrole-2-carbonyl 2-oxo-2-(pyrimidin-5-yl)acetamido]acetate	-8.3	Pyrrl Aryl Sulfones	Pyrimidine
E	3-[3-(2H-1,2-benzoxazin-6-yl)-5-methylbenzenesulfonyl]-5-chloro-1H-indole-2-carboxamide	-8.4	Indoyl Aryl Sulfones	Benzoxazine
F	(2H-1,2-benzoxazin-8-yl)methyl 1-(2-amino-5-chlorobenzenesulfonyl)-1H-pyrrole-2-carboxylate	-7.4	Pyrrl Aryl Sulfones	Benzoxazine
G	2-(4-[(1Z,3E)-1,4-dihydroxy-4-[4-(2-oxoacetyl)phenyl]buta-1,3-diene-1-sulfonyl]phenyl)-2-oxoacetaldehyde	-7.5	Phenyl Vinyl Sulfones	Pyrimidine
H	(2R,3E)-8-(ethenesulfonyl)-3-[[quinazoline-8-sulfonyl]methylidene]-1,2,3,4-tetrahydroquinolin-2-amine	-7.9	Phenyl Vinyl Sulfones	Pyrimidine
I	2H,3H,10aH-naphtho[2,3-c][1,2]oxazin-8-yl ethene-1-sulfonate	-8.2	Benzyl Vinyl Sulfones	Benzoxazine
J	5-chloro-3-[[3-(2-oxo-3-(piperazin-2-yl)propyl)-5-(2,3,4-trioxobutyl)benzenesulfonyl]-1H-indole-2-carboxamide	-8.7	Indoyl Aryl Sulfones	Piperazine

**Table 1: Structure-Activity Relationship Table displaying best binding affinity per series.** Autodock Vina binding affinities of the best aryl sulfone compounds from each series screened against HIV-1 reverse transcriptase. Each series has 10 molecules.

zine functional group. It has a binding affinity of -9.4 kcal/mol.

**Table 2** shows key information from the pharmacological screening of these molecules. All of our molecules have been pharmacokinetically screened for carcinogenic constructs through the use of pkCSM software (20). This decreases the time needed for any research attempting to synthetically design the NNRTIs, since the risk of cancer has already been mitigated severely through determining Ames toxicity (a measure of the mutative properties of a chemical). The molecules that were found to be Ames toxic were not recorded as they were eliminated immediately. We also measured water solubility at 25°C, which is useful for understanding the intake method of any potential NNRTI. The chosen molecule from Series I had the best water solubility. Intestinal absorption is also a critical pharmacological property to understanding the efficacy of any potential drugs since low intestinal absorption requires higher intake amounts. The selected series B molecule had the most desired intestinal absorption, while the predicted percent absorption was significantly lower for the molecules in series D, G, and J. Overall, however, all of our molecules exceeded the desired intestinal absorption rate of 30%. Any value below 30% would decrease the efficacy of the drug significantly to the point of being impractical (20).

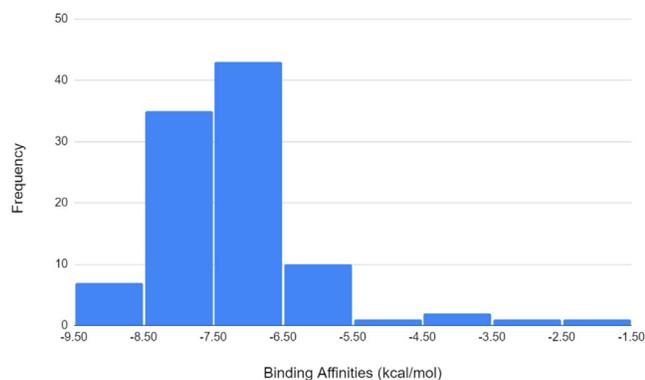
### Overall Statistical Analysis

The mean binding affinity was -7.3 kcal/mol, with the highest value being -1.8 kcal/mol and the lowest value being -9.4 kcal/mol (**Figure 2**).

There is a low degree of correlation between the molecular weight and the binding affinity, which is not statistically significant ( $R = -0.129$ ,  $p = 0.201$ , **Figure 3A**). Upon removing the outliers, there is still a low degree of correlation, but it is now statistically significant ( $R = -0.297$ ,  $p = 2.98 \times 10^{-3}$ ). The correlation is existent but should be viewed skeptically because the larger the molecule is, the greater chance that the groups

Series	Water Solubility (log mol/L)	Intestinal Absorption (% Absorbed)	Maximum Tolerated Dose (log mg/kg/day)	Oral Rat Acute Toxicity (LD50) (mol/kg)
A	-4.234	95.008	0.153	2.473
B	-4.221	96.49	0.246	2.655
C	-3.284	93.55	0.732	2.938
D	-3.486	64.893	0.361	1.835
E	-4.234	95.008	0.153	2.729
F	-3.886	89.298	0.06	2.268
G	-4.777	72.43	0.072	1.898
H	-4.294	77.14	0.583	2.193
I	-2.174	81.376	0.659	2.12
J	-4.02	69.642	0	2.46

**Table 2. Pharmacological table.** Key attributes from the ADMET Analysis of the best aryl sulfone compounds from each series. Water solubility measures the ability of a chemical to dissolve in water at 25 degrees Celsius. Intestinal absorption measures the amount of a drug that would be absorbed by the intestines if 100% was consumed. Max tolerated dose measures the amount a human could consume per day before serious health effects. The LD50 is the drug dosage that would kill 50% of all test subjects in a study. These values show the relative pharmacological stability of our ligands.



**Figure 2: Binding affinity distribution.** Distribution of binding affinities of 100 novel aryl sulfone compounds as calculated by docking them to HIV-1 Reverse Transcriptase (RT) using AutoDock Vina. The distribution is concentrated in the region from -8.5 kcal/mol through -6.5 kcal/mol and shows that 32 molecules had performed better than the control, the indoyl aryl sulfone (-7.8 kcal/mol).

necessary for a strong bond would be present as opposed to the binding affinity improving solely because of the weight. A larger molecule has a greater probability of having these groups, but it is not guaranteed to have them. This analysis indicates that despite the correlation, it is not necessarily beneficial to take into account the weight in the design of aryl sulfone NNRTIs. Additionally, in line with Lipinski's Rule of Five, molecules with a weight higher than 500 Daltons are generally considered pharmacologically unsuitable (21). As such, any molecules exceeding this weight are considered not viable for oral consumption.

We observed a low degree of correlation between the logP and the binding affinity, which is statistically nonsignificant ( $R = -0.166$ ,  $p = 9.82 \times 10^{-2}$ , **Figure 3B**). After removing the outliers, there remains a low degree of correlation which is not statistically significant ( $R = -0.166$ ,  $p = 0.101$ ). LogP refers to the partition coefficient and measures the hydrophobicity or hydrophilicity of a molecule. Though a broad correlation cannot be applied due to being statistically nonsignificant, it is notable that the majority of the data is in the positive range from  $\log P = 1-5$ . This indicates that most of our calculated aryl sulfone NNRTI candidates have a higher affinity for the lipid phase, making them lipophilic. Additionally, Lipinski's rule is relevant again, as a logP above 5 would cause a molecule to be not viable to be consumed orally (21). Nevertheless, very few ligands had a coefficient larger than five, and this criterion only disqualified six molecules.

There exists a low degree of correlation between the number of rotatable bonds and the binding affinity, which is not statistically significant ( $R = 0.119$ ,  $p = 0.239$ , **Figure 3C**). After removing the outliers, the correlation drastically drops and is certainly statistically nonsignificant ( $R = 3.60 \times 10^{-4}$ ,  $p = 0.997$ ). Overall, this reveals that there is no statistically significant correlation between rotatable bonds and binding affinity. This means that the number of rotatable bonds should not be strongly considered in the design of aryl sulfone NNRTIs.

We noticed a low degree of correlation between the number of hydrogen bonds and the binding affinity, which is statistically nonsignificant ( $R = -0.272$ ,  $p = 6.19 \times 10^{-3}$ , **Figure 3D**). Following the removal of the outliers, there is a low degree of correlation that is statistically significant ( $R = -0.306$ ,  $p =$

$2.21 \times 10^{-3}$ ). This means that it is important to optimize for more hydrogen bond donors when designing aryl sulfones as NNRTIs.

We found a moderate degree of correlation between the number of aromatic rings and the binding affinity which is not statistically significant ( $R = -0.378$ ,  $p = 1.03 \times 10^{-4}$ , **Figure 3E**). After removing the outliers, the correlation improves ( $R = -0.453$ ,  $p = 2.74 \times 10^{-6}$ ). Hence, aromatic rings have a beneficial impact and should be strongly considered when designing aryl sulfones as NNRTIs.

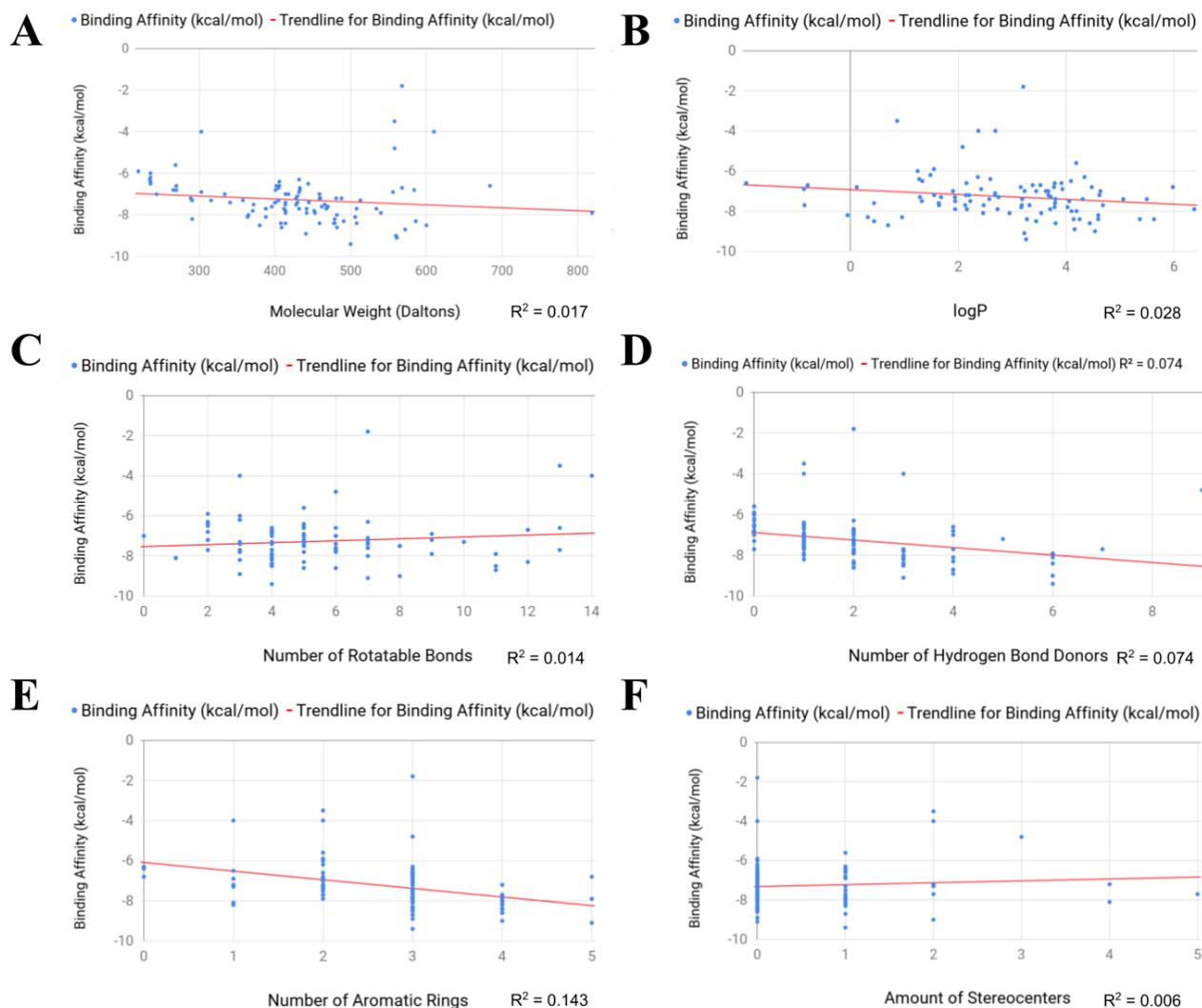
There is a very weak correlation between the number of atomic stereocenters and the binding affinity of the ligands, which is statistically nonsignificant ( $R = 7.89 \times 10^{-2}$ ,  $p = 0.435$ , **Figure 3F**). After removing the outliers, the correlation is still weak and not statistically significant ( $R = 6.01 \times 10^{-2}$ ,  $p = 0.557$ ). However, this experiment is inconclusive, because the data points are too heavily clustered on zero and one atomic stereocenters. Ninety percent of the ligands have either zero or one atomic stereocenter, and 71% have zero atomic stereocenters. This indicates that most aryl sulfone compounds designed in this procedure have a very small number of stereocenters.

### Analysis of Scaffolds and Functional Groups

We calculated the mean binding affinities of each scaffold and functional group using the means of the respective series that contain them. The most successful scaffold was the indoyl aryl sulfone scaffold whereas piperazine was the best functional group by a large margin. The indoyl aryl sulfone scaffold was consistently effective in producing molecules that had an above-average binding affinity when paired with pyrimidine, styrene, and piperazine. The second-best scaffold was pyrrol aryl sulfones, proving that indoyl aryl sulfones are a superior scaffold by a wide margin, and should be further pursued in future research. The cyclic ring with the highest average binding affinity was piperazine. However, our research showed that Series C based on the indoyl aryl sulfone scaffold combined with the pyrimidine functional group had a lower binding score on average than series J, which combined the indoyl aryl sulfones with piperazine, by an average binding affinity of -0.3 kcal/mol.

### Chimera Binding Analysis

We analyzed the three best performing molecules, denoted A9, C6, and J10 from the A, C, and J series, respectively, as well as the previously suggested IAS derivative 5-bromo-3-[(3,5-dimethylphenyl)sulfonyl]indole-2-carboxamide by visualizing the bonds in Chimera on both the A and B strands of the HIV1-RT protein (**Figure 4**). A8 was the highest performing molecule (based solely on binding affinity) we created during our research, which had a single hydrogen and several contact bonds within the HIV binding pocket. The only hydrogen bond present was between tyrosine 181.A (the A portion indicating the strand) at 2.37 Å and the charged hydrogen in piperazine. Other significant bonds were a nonpolar bond between proline 25.B at 3.749 Å to chlorine in chlorobenzene and glutamine 91.A, which had two polar interactions at 3.1 Å to a nitrogen in central part of scaffold and 3.423 Å to carbon adjacent to nitrogen. There was also a hydrophobic interaction between leucine 26.B at 2.938 Å and a benzene ring methyl group. The other methyl group bonded to both valine 381.A at 2.992 Å and asparagine 137.B at 2.928 Å at C22. The final



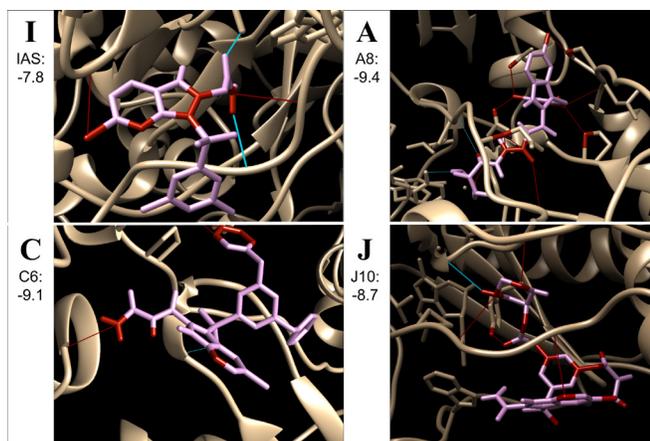
**Figure 3: Chemical property correlation with binding affinity.** (A) The relationship between molecular weight (Daltons) and binding affinity (kcal/mol) of the 100 novel aryl sulfone compounds to HIV-1 RT ( $R = -0.129$ ,  $p = 0.201$ ,  $R^2 = 0.017$ ). (B) The relationship between logP and binding affinity (kcal/mol) of the 100 novel aryl sulfone compounds to HIV-1 RT ( $R = -0.166$ ,  $p = 9.82 \times 10^{-2}$ ,  $R^2 = 0.028$ ). (C) The relationship between the number of rotatable bonds and binding affinity (kcal/mol) of the aryl sulfone compounds to HIV-1 RT ( $R = 0.119$ ,  $p = 0.239$ ,  $R^2 = 0.014$ ). (D) The relationship between the number of hydrogen bond donors and binding affinity (kcal/mol) of the aryl sulfone compounds to HIV-1 RT ( $R = -0.272$ ,  $p = 6.19 \times 10^{-3}$ ,  $R^2 = 0.074$ ). (E) The relationship between the number of aromatic rings and binding affinity (kcal/mol) of the aryl sulfone compounds to HIV-1 RT ( $R = -0.378$ ,  $p = 1.03 \times 10^{-4}$ ,  $R^2 = 0.143$ ). (F) The relationship between the number of atomic stereocenters and binding affinity (kcal/mol) of the aryl sulfone compounds to HIV-1 RT ( $R = 7.89 \times 10^{-2}$ ,  $p = 0.435$ ,  $R^2 = 0.006$ ).

two contact bonds for A9 were with glycine 93.A at 2.083 Å to a methylformamide group and asparagine 137.B at 2.982 Å to a benzene ring.

C6 was the second-best performing derivative created during our research. This molecule also had several key interactions including two hydrogen bonds, which may have contributed to the binding affinity. It had a bond with tyrosine 319.A at 2.638 Å from the hydrogen end of the amine group, and a bond between glutamic acid 28.B at the other hydrogen in the same group. There was also a bond from the nitrogen in the amine group to tryptophan 383.A at 2.925 Å. The glutamic acid 28.B amino acid also bonded to a hydrogen on the amine attached to the backbone of the drug at 2.044 Å. This led to the formyl urea group in C6 being one of the functional groups with the most interactions in the entire molecule. The

final significant interaction was between glutamic acid 138.B at 3.519 Å to the outermost carbon in the top pyrimidine group. The most significant hydrogen bond was between the oxygen on the oxidane-SO-thioperoxol group and isoleucine at 3.236 Å.

The J10 ligand had several contact bonds and two hydrogen bonds, which still did not increase its binding affinity beyond A9 or C6. One notable interaction was a nonpolar bond between valine 90.A and the chlorobenzene ring at 2.962 Å. There were also two significant interactions from the piperazine amine group, one from a hydrogen at 2.079 Å to asparagine 137.B O, and one from a nitrogen in the same amine to glycine 93.A at 2.951 Å which was nonpolar. Another interaction present was between a prop-2-one oxygen and 2.814 Å to tyrosine 181.A. The final significant non-hydrogen



**Figure 4: Protein-ligand complex visualizations.** Chimera visualization of the three best performing molecules (A8 with an indolyl aryl sulfone scaffold and piperazine functional group, C6 with an indolyl aryl sulfone scaffold and pyrimidine functional group, J10 with an indolyl aryl sulfone scaffold and piperazine functional group) and the IAS derivative to which they were compared. Protein docked against was 1REV.pdb.

interaction for this molecule was between tyrosine 115.A and the oxygen in the 2,3-dioxobutanal group at 2.825 Å. The first significant hydrogen bond was between tyrosine 115.A and an oxygen atom in the same dioxobutanal group as above. The second hydrogen bond was between tyrosine 183.A and the sulfur oxide group at 3.065 Å.

These polar and nonpolar bonds were compared to the IAS derivative 5-bromo-3-[(3,5-dimethylphenyl)sulfonyl]indole-2-carboxamide interactions with the HIV protein. The bromine group in the derivative had three separate interactions with the protein: with tyrosine 115.A at 3.74 Å, with proline 157.A at 3.738 Å, and with methionine 184.A at 4.01 Å. Despite these bonds being very weak, these bond lengths give bromine a specific area to attach to within the binding pocket. Another area on the derivative with multiple bonds was the amino group on the amide. The nitrogen in the group interacts hydrophobically with glycine 93.A at 2.815 Å, and the hydrogen interacts with glutamine 91.A at 2.123 Å. There are also hydrophobic interactions between two carbons in the pyrimidine interact with valine 90.A at 3.284 Å and 3.3 Å. The final significant non-hydrogen bond was a hydrophobic interaction between the carbon on the bromobenzene ring and methionine 184.A at 3.404 Å. The only significant hydrogen bond was between tyrosine 181.A and the oxygen on the amide group at 2.892 Å. These results confirmed that hydrophobic groups had a higher probability of binding to the molecule, as there were very few strong polar interactions.

## DISCUSSION

In the screening for potential NNRTIs built on the aryl sulfone scaffold using cyclic compounds, we found 32 molecules that bind to HIV-1 RT as well or better than the previously published IAS derivative 5-bromo-3-[(3,5-dimethylphenyl)sulfonyl]indole-2-carboxamide (16). Furthermore, by dividing our molecules into 10 series, we were able to discover general trends among different scaffolds and functional groups. This affirms our initial hypothesis that the combination of the indolyl aryl sulfone scaffold and pyrimidine would result in the series with

the best binding affinity due to the prominence of pyrimidine in FDA-approved NNRTIs and the prior success of IAS derivatives (17). Notably, the indolyl aryl sulfone combination with piperazine did not create the series with the highest binding affinity despite them both individually having the highest binding affinities on average. Upon analyzing the molecules present in the series, we postulated that the potential incompatibility between the molecules could be attributed to the number of hydrogen bonds within the piperazine-containing molecules, which may interfere with the indolyl aryl sulfone interactions. Further binding pocket analysis should be conducted to better understand the interference in the binding of piperazine and the indolyl aryl sulfone scaffold to HIV-1 RT. We were also able to identify non-effective scaffolds, such as phenyl vinyl sulfonate and benzyl vinyl sulfonate that exhibited significantly poorer binding affinity.

Our statistical analysis of various chemical properties provided multiple important discoveries regarding the properties of the aryl sulfone NNRTI pharmacophore. In the design of aryl sulfone NNRTIs, an increased weight improves the probability of an increased binding affinity but should not be emphasized in the design of aryl sulfone NNRTIs since this correlation is likely due to the probability of an ideal functional group improving with larger molecules. Additionally, most aryl sulfone NNRTIs are lipophilic with logPs between 1-5, which signals a good permeability and hence a better potency. The FDA-approved NNRTIs Efavirenz, Etravirine, Nevirapine, Delavirdine, and Rilpivirine have logPs of 4, 4.5, 2, 2.4, 4.5, respectively, as calculated by PubChem (18,19,22–24). These values are in this lipophilic range, which is a positive sign for their potential efficacy as NNRTIs. Our research also found that in the design of aryl sulfone NNRTIs, rotatable bonds should not be strongly considered, while it is very beneficial to increase the amount of hydrogen bond donors. Aromatic rings are also very beneficial for good binding scores in NNRTIs. Additionally, the majority of the ligands found here have zero stereocenters, which is beneficial from a synthesis perspective for reducing the potential of diastereomers and enantiomers with varying binding affinities decreasing the overall potency of the drug or increasing the expense of synthesis.

A few key binding residues on the HIV protein were tyrosine 181.A, asparagine 137.B, and tyrosine 115.A. Tyrosine 181.A was present in two hydrogen bonds for A9 and the derivative, and a contact bond for C6. Its high affinity to have hydrogen bonds and its ability to have contact bonds make Tyrosine 181.A an area of interest for future NNRTI research. Asparagine 137.B was able to bind to a methyl group and a hydrogen on an amine group, meaning that it can make contact with highly polar and hydrophobic substances as well. Tyrosine 115.A was also able to make hydrogen and contact bonds, similar to Tyrosine 181.A. These three amino acids were all successful because of their propensity for binding to both nonpolar and polar amino acid residues, making them all important for any NNRTI research regardless of the functional groups used.

This research offers three significant contributions. The first is identifying 32 novel compounds as potential NNRTI candidates. The second is contributing to the pharmacophore knowledge of NNRTIs by contrasting the efficacies of various combinations of scaffolds and functional groups in binding to RT. The third is contributing detailed knowledge on which amino acids are more likely to form bonds with functional groups,

thereby helping future researchers understand the nature of the HIV-1 RT and which interactions will be successful for inhibition.

We plan to further study the aryl sulfone compounds generated in this research by docking them against mutant HIV-1 reverse transcriptase enzymes. Of specific interest would be the mutations at positions 103 and 181 to test if the novel aryl sulfones maintain their capability in different HIV strains as seen in 5-bromo-3-[(3,5-dimethylphenyl)sulfonyl]indole-2-carboxamide (16). This study's screening also warrants further analysis into the potential for the incorporation of other functional groups and more ring structures using the aryl sulfone scaffold such as further functionalization of the phenyl ring.

## METHODS

### Modelling, Screening and DFT

The aryl sulfone library was created and modeled using application version 1.2.0 of the molecular modeling program Avogadro, and version 21.14.0 of Marvin Sketch, by designing 100 molecules that contain their respective aryl sulfone and cyclic compounds in accordance with their series (25, 26).

Every ligand design was screened for Ames toxicity utilizing the pkCSM (20) online service. Only ligands that were not Ames toxic (the 100 mentioned throughout the paper) were docked and considered in this study. Pharmacological analysis was conducted on the most negative binding affinity aryl sulfone compounds from each series to assess the key pharmacokinetic properties of water solubility, intestinal absorption, Ames toxicity, maximum tolerated dose, and oral rat acute toxicity. The LD50 helps determine the potential for short-term poisoning and is measured by the amount of dose needed to kill fifty percent of rodent test subjects.

Avogadro's built-in ORCA (26) extension was used to generate the ORCA input files to perform geometry optimization on the structures using DFT in ORCA version 4.2.1 using the def2-SVP basis set and the BP86 functional.

### Docking and Visualization

AutoDock Vina (28) and Chimera, version 1.15 (29), were used to dock the optimized structure to the allosteric binding pocket of HIV-1 reverse transcriptase (30). The built-in Chimera AutoDock Vina tool was used to submit commands to AutoDock Vina (version 1.1.2) (24). For the receptor search volume, the center was set to [6,-34,23] with the size parameters of [30,25,31]. The default receptor, ligand options, and advanced options were kept according to the default settings. The most negative conformer binding affinity was recorded.

Chimera was used to visualize the bond of the conformer bound to the allosteric site. The residues that had a van der Waals radii overlap greater than or equal to 0.1 were named and colored in red. Any contacts due to hydrogen bonds were labeled in blue. Chimera was used to visualize hydrogen bonds, contacts, hydrophobic and polar pockets within the HIV reverse transcriptase receptor, and the identity of critical amino acids.

### Chemical Property Analysis

Following the docking of the ligands, the molecular weight, Wildman-Crippen logP value, number of rotatable bonds, number of hydrogen bond donors, and the number of atomic stereocenters were calculated using the RDKit Python package (version 2021.03.1) (31). Microsoft Excel (version 16.43)

and Google Sheets were used to visualize the relationship between these descriptors and the docking scores. Additionally, utilizing RDKit, a cheminformatics software, the relationships between various chemical properties were calculated. The Pearson product-moment correlation coefficient ( $R$ ) was calculated using the Google Sheets correlation function, comparing the chemical properties to the binding affinity. The coefficient of determination of the trendline was calculated using the Google Sheets line of best fit ( $R^2$ ). To calculate the p-value, the t-statistic and degrees of freedom (df) were calculated. The t-statistic, with  $N$  being the sample size and  $R$  being the correlation is calculated in google sheets using the function:

$$t = (R \cdot \sqrt{N - 2}) / (\sqrt{1 - R^2})$$

The degrees of freedom is calculated as  $N-1$  (33). Finally, the p-value is calculated using the function: =TDIST(t,df,2) (34). To determine the alpha, the Bonferroni Correction is applied to the initial alpha of 0.05. There are 12 comparisons so the significance is defined as  $p \leq 4.17 \times 10^{-3}$ . For each comparison, the  $R$  and p-values are calculated with the full set of molecules, and then again with outliers omitted. For outlier detection, the Z-scores of the binding affinities of the molecules were calculated and the ligands with Z-scores greater than 3 or less than -3 were omitted. This resulted in 2 ligands being excluded from the analysis.

## ACKNOWLEDGMENTS

We would like to thank the Irvington Research Club for all of their guidance, resources, and support for the creation of this research project and manuscript.

**Received:** October 19, 2021

**Accepted:** March 03, 2022

**Published:** October 12, 2022

## REFERENCES

1. "Global HIV & AIDS Statistics - Fact Sheet." *UNAIDS*, www.unaids.org/en/resources/fact-sheet.
2. "What Are HIV and AIDS?" *HIV.gov*, 8 Apr. 2021, www.hiv.gov/hiv-basics/overview/about-hiv-and-aids/what-are-hiv-and-aids.
3. "Sexual Health & Reproductive System." *Rady Children's Hospital-San Diego*, www.rchsd.org/health-articles/how-do-people-get-aids/.
4. Vidya Vijayan, K. K., *et al.* "Pathophysiology of CD4+ T-Cell Depletion in HIV-1 and HIV-2 Infections." *Frontiers in Immunology*, vol. 8, no. 580, International Union of Immunological Societies, May 2017, doi:10.3389/fimmu.2017.00580.
5. "List of NNRTIs." *Drugs.com*, www.drugs.com/drug-class/nnrtis.html#:~:text=Non%2Dnucleoside%20reverse%20transcriptase%20inhibitors,immune%20deficiency%20syndrome%20(AIDS).
6. "The Science of HIV and AIDS - Overview." *Avert*, 10 Oct. 2019, www.avert.org/professionals/hiv-science/overview.
7. Surapaneni, Anvi *et al.* "A Review of Non-Nucleoside Reverse Transcriptase Inhibitors' Structure and Activity with HIV-1 Reverse Transcriptase." *Young Scientists Journal*, 15 Aug. 2020, ysjournal.com/a-review-of-non-nucleosi-

- de-reverse-transcriptase-inhibitors-structure-and-activity-with-hiv-1-reverse-transcriptase/.
8. Mackie, Nicola. "Resistance to Non-Nucleoside Reverse Transcriptase Inhibitors." *Antiretroviral Resistance in Clinical Practice*, U.S. National Library of Medicine, 1 Jan. 1970,
  9. Agostino, Bruno, et al. "The in Silico Drug DISCOVERY Toolbox: Applications in Lead Discovery and Optimization." *Current Medicinal Chemistry*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/29110597/.
  10. Sliwoski, Gregory, et al. "Computational Methods in Drug Discovery." *Pharmacological Reviews*, The American Society for Pharmacology and Experimental Therapeutics, 31 Dec. 2013, doi: 10.1124/pr.112.007336
  11. Dhasmana, Anupam, et al. "High-Throughput Virtual Screening (HTVS) of Natural Compounds and Exploration of Their BIOMOLECULAR Mechanisms: An In Silico APPROACH." *New Look to Phytomedicine*, Academic Press, 26 Oct. 2018, www.sciencedirect.com/science/article/pii/B9780128146194000203.
  12. Qingzhi, Gao, et al. "Pharmacophore Based Drug Design Approach as a Practical Process in Drug Discovery." *Current Computer-Aided Drug Design*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/20370694/.
  13. "Deep Scaffold: A Comprehensive Tool for Scaffold-Based De NOVO Drug Discovery Using Deep Learning." *Journal of Chemical Information and Modeling*, vol. 60, 2020, pp. 77-91. pubs.acs.org/doi/10.1021/acs.jcim.9b00727.
  14. Dhasmana, Anupam, et al. "High-Throughput Virtual Screening (HTVS) of Natural Compounds and Exploration of Their Biomolecular Mechanisms: An in Silico Approach." *New Look to Phytomedicine*, Academic Press, 26 Oct. 2018, www.sciencedirect.com/science/article/pii/B9780128146194000203.
  15. Meng, Xuan-Yu, et al. "Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery." *Current Computer-Aided Drug Design*, U.S. National Library of Medicine, June 2011, www.ncbi.nlm.nih.gov/pmc/articles/PMC3151162/#:~:text=Essentially%2C%20the%20aim%20of%20molecular,complex%20structure%20using%20computation%20methods.
  16. Famigliani, Valeria and Romano Silvestri. "Indolylarylsulfones, a Fascinating Story of Highly Potent Human Immunodeficiency Virus Type 1 Non-Nucleoside Reverse Transcriptase Inhibitors." *Antiviral Chemistry & Chemotherapy*, vol. 26, 2018. doi:10.1177/2040206617753443.
  17. Romano, Silvestri and Marino, Artico. "Indolyl Aryl Sulfones (IASs): Development of Highly Potent NNRTIs Active against Wt-HIV-1 and Clinically Relevant Drug Resistant Mutants." *Current Pharmaceutical Design*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/16305512/.
  18. "Efavirenz." *National Center for Biotechnology Information. PubChem Compound Database*, U.S. National Library of Medicine, pubchem.ncbi.nlm.nih.gov/compound/64139#section=Depositor-Supplied-Synonyms.
  19. "Etravirine." *National Center for Biotechnology Information. PubChem Compound Database*, U.S. National Library of Medicine, pubchem.ncbi.nlm.nih.gov/compound/193962#section=Depositor-Supplied-Synonyms.
  20. Pires, Douglas E V, et al. "PkcsM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures." *Journal of Medicinal Chemistry*, American Chemical Society, 14 May 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC4434528/.
  21. "Lipinski's Rule of Five." *Lipinski's Rule of Five - an Overview | ScienceDirect Topics*, www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/lipinskis-rule-of-five#:~:text=Traditionally%2C%20therapeutics%20have%20been%20small,P%20not%20greater%20than%205).
  22. "Nevirapine." *National Center for Biotechnology Information. PubChem Compound Database*, U.S. National Library of Medicine, pubchem.ncbi.nlm.nih.gov/compound/4463.
  23. "Delavirdine." *National Center for Biotechnology Information. PubChem Compound Database*, U.S. National Library of Medicine, pubchem.ncbi.nlm.nih.gov/compound/5625#section=Depositor-Supplied-Synonyms.
  24. "Rilpivirine." *National Center for Biotechnology Information. PubChem Compound Database*, U.S. National Library of Medicine, pubchem.ncbi.nlm.nih.gov/compound/6451164.
  25. Hanwell, Marcus D., et al. "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform." *Journal of Cheminformatics*, vol. 4, no. 17, 2012, doi:10.1186/1758-2946-4-17.
  26. ChemAxon, version 21.11.0, *Marvin JS*. chemaxon.com/products/marvin-js/download
  27. Neese, Frank. "The ORCA Program System." *WIREs Computational Molecular Science*, vol. 2, no. 1, 2011, pp. 73-78. doi:10.1002/wcms.81.
  28. Trott, Oleg, and Arthur J. Olson. "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading." *Journal of Computational Chemistry*, vol. 31, no. 2, 2010, pp. 455-61. doi:10.1002/jcc.21334
  29. Pettersen, Eric F, et al. "UCSF Chimera—a visualization system for exploratory research and analysis." *Journal of Computational Chemistry*, vol. 25, no. 13, 2004, pp. 1605-12. doi:10.1002/jcc.20084
  30. RCSB Protein Data Bank. "1REV: HIV-1 REVERSE TRANSCRIPTASE." *RCSB PDB*, www.rcsb.org/structure/1rev.
  31. Landrum, Greg. "Open-Source Cheminformatics Software." *RDKit*, www.rdkit.org/.
  32. "Hypothesis Test for the Population Correlation Coefficient." *Penn State Eberly College of Science: STAT 501*, online.stat.psu.edu/stat501/lesson/1/1.9. Accessed 3 Aug. 2022.
  33. Calder, Kate. "Statistics 528 - Lecture 23." *Ohio State College of Arts and Sciences*, asc.ohio-state.edu/calder.13/stat528/Lectures/lecture23\_2slides.PDF. Accessed 3 Aug. 2022.
  34. "TDIST and TINV." *California State University Bakersfield*, csu.edu/~ychoi2/MGMT\_BA%20301/Ch08/TDIST%20and%20TINV.doc. Accessed 3 Aug. 2022.

**Copyright:** © 2022 Zhang, Adwankar, Morgaonkar, Patel, Sengupta, Choi. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>). 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.