Pharmacophore-based screening and identification of molecular-level descriptors applied to non-nucleoside reverse transcriptase inhibitors (NNRTIs)

Bhavesh Ashok[1], Ayush Bajaj[2], Rohan Adwankar[3], Atri Surapaneni[2], Anvi Surapaneni[2], Allen Chen[4], Stephanie Sun[5], Kushal Chattopadhyay[4], Jeslyn Wu[4], Andrew Liang[6], Ayeeshi Poosarla[4], Karankumar Mageswaran[4], Isha Rao[7], Sania Kharshingkar[4], Sushruth Booma[4], *Robert Downing*[8], *Edward Njoo*[9]

[1]Amador Valley High School, Pleasanton, CA; [2]The Quarry Lane School, Dublin, CA; [3]Irvington High School, Fremont, CA; [4]Mission San Jose High School, Fremont, CA; [5]BASIS Independent Silicon Valley, San Jose, CA; [6]The College Preparatory School, Oakland, CA; [7]Dougherty Valley High School, San Ramon, CA; [8]Department of Computer Science & Engineering, Aspiring Scholars Directed Research Program, Fremont, CA; [9]Department of Chemistry, Biochemistry, & Physical Science, Aspiring Scholars Directed Research Program, Fremont, CA

ABSTRACT: Retroviruses such as HIV use the reverse transcriptase (RT) enzyme to transcribe their viral RNA into DNA, which is subsequently incorporated into the human host cell genome and expressed to replicate the virus. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are a class of antiretroviral drugs that have been used to treat human immunodeficiency virus (HIV) infections. In any small molecule bioactive compound, two important structural features that define the thermodynamics of ligand-receptor binding are the 3-dimensional orientation and positioning of hydrogen bond donor/acceptor sites and aromatic rings, which form hydrogen bonds and pi-stacking interactions, respectively with the target. Here, scripts to determine heteroatom and aromatic ring positions were written in the programming language Python, generating the Cartesian coordinates of these structural features in a 3-dimensional map. This allowed for the creation of a descriptor-driven platform that can identify and statistically analyze the structural correlation across a library of structurally diverse chemical entities that bind to a known biomolecular target, such as HIV-RT. Six FDA-approved NNRTIs and three non-FDA-approved NNRTIs were analyzed using PaDEL-Descriptor, a molecular characteristics quantification program, to determine and identify structural similarities among molecules that have displayed potent biological activity against HIV-RT.

INTRODUCTION

The ability of high throughput virtual screening (HTVS) to expedite the identification of hit compounds has assumed an increasingly influential role in the process of drug discovery and development (1). Due to the public availability of vast chemical libraries, thousands of molecules can be screened to determine those that can bind to the target site, and hence offer the possibility of being utilized as drugs. Within the binding pocket of the target site, hydrogen bond donors/acceptors and aromatic rings on ligands have been identified to facilitate hydrogen bonding and pi-stacking interactions, respectively, which allow for effective binding and interaction with the target receptor. However, these approaches often utilize computationally expensive methods, which greatly limits the extent of chemical space that can be sampled in any given HTVS campaign. Due to these limitations, advances in pharmacophore motif screening techniques are being made in order to act as a preliminary screening stage (2, 3). These screening techniques work by constraining the screening set to molecules that meet with identified structural motifs. By only selecting molecules that meet these criteria, the computational expense in conducting a HTVS can be greatly reduced.

Here, we envision and develop a model identifying and correlating the key pharmacophoric features of such molecules, which would be greatly advantageous in the drug discovery process. Moreover, we apply this to a library of non-nucleoside reverse transcriptase inhibitors (NNRTIs), which are small molecule antiretroviral compounds targeting the reverse transcriptase enzyme of the human immunodeficiency virus (HIV) and related retroviruses (**Figure 1**). We hypothesize that this methodology will

serve as a companion to HTVS in molecular classification and identification of molecular structure.

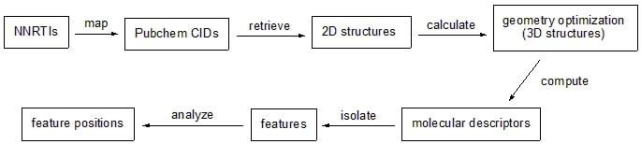


Figure 1 Workflow for the development of data and positional analyses on selected structural features

The human immunodeficiency virus (HIV) weakens the body's immune system by targeting CD4+ T lymphocytes. Progression of this disease leads to the onset of acquired immunodeficiency syndrome (AIDS), which is characterized by a low immune cell count, and consequently high vulnerability towards infectious diseases. Currently, there is no cure for HIV, so patients under antiretroviral therapy (ART) are treated with a drug regimen that targets different enzymes in the HIV life cycle (4). HIV belongs to the family Retroviridae, which use the host cell's functions and mechanisms to reproduce. The virus' RNA is converted into DNA by the RT enzyme, and the resulting complementary DNA (cDNA) is integrated into the host cell's genome (5). As such, the RT enzyme is a critical protein in HIV replication and is a common target for antiretroviral compounds, including NNRTIs. The first discovered strain of the HIV virus is classified as HIV-1 and makes up more than 95% of all worldwide HIV infections (6). NNRTIs inhibit HIV-1 RT by allosterically binding to a specific "pocket" on the enzyme, which contains a number of key hydrophobic amino acid residues and several aromatic side-chain residues that result in critical interactions with the bound NNRTI (7). The structural similarities among NNRTIs presents this class of molecules as a prime target for the analysis and quantification of common motifs. Through a comparison of a number of NNRTIs, it is clear that there are several recurring structural moieties including: aromatic rings to act as π-electron donors in aromatic ring stacking interactions, hydrocarbon chains to participate in hydrophobic interactions within the binding pocket, and functional groups with hydrogen bond donors and acceptors capable of forming strong hydrogen bonds (8). However, frequent mutations of the HIV-1 RT enzyme confer resistance towards current NNRTI drugs, necessitating the continued development of novel NNRTIs that will be more effective against these mutant strains.

First-generation NNRTIs—nevirapine (**Table 1.1**), delavirdine (**Table 1.2**), and efavirenz (**Table 1.3**)—share a "butterfly-like" geometry (9). Nevirapine, the first FDA-approved NNRTI, is a dipyridodiazepinone inhibitor with two conjugated pyridine systems that flank a central carbonyl; delavirdine also has a central carbonyl, flanked by an indole and a pyridyl piperazine ring system. The last of the first-generation NNRTIs, efavirenz, is a benzoxazinone and contains two key halogenated moieties—chlorobenzene and trifluoromethyl (10). Second-generation NNRTIs, etravirine and rilpivirine, which generally consist of larger aromatic systems, have greater torsional flexibility, resulting in greater potency towards mutant strains of HIV-1 (11). The most recent NNRTI approved by the FDA—doravirine (**Table 1.6**)—is a pyridinone inhibitor with even greater efficacy against common drug-resistant HIV-1 RTs. Furthermore, there are other NNRTIs from the following generation of compounds, all of which are or were investigational drugs. RDEA806 (**Table 1.9**) is part of the triazole class of NNRTIs, lersivirine (**Table 1.8**) is part of the pyrazole class, and fosdevirine (**Table 1.7**) is part of the phosphoindole class (12). Both lersivirine and fosdevirine continued up until Phase II clinical trials, resulting in a significant amount of data on its effectiveness (13, 14). The structures and FDA approval status of the nine NNRTIs are shown below. (**Table 1**)

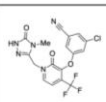
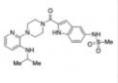
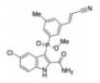
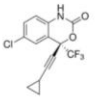
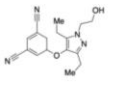
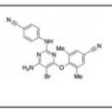
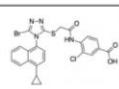
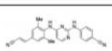
Name	Structure	FDA-Approval	Name	Structure	FDA-Approval
Nevirapine (1.1)		Yes (1996)	Doravirine (1.6)		Yes (2018)
Delavirdine (1.2)		Yes (2001)	Fosdevirine (1.7)		No
Efavirenz (1.3)		Yes (1998)	Lersivirine (1.8)		No
Etravirine (1.4)		Yes (2008)	RDEA806 (1.9)		No
Rilpivirine (1.5)		Yes (2011)			

Table 1 Attributes of the 9 NNRTIs chosen as data points for computational study.

The basis of an NNRTI's inhibitory activity directly correlates with the energetics of the protein-ligand interactions between the inhibitor and the RT enzyme. Our work focuses on two interactions: hydrogen bonding and pi stacking interactions. Hydrogen bonds are a predominant intermolecular force in ligand-protein interactions, and this is also described in Lipinski's Rule of 5. Lipinski's rule outlines that an active small molecule drug usually contains more than 5 hydrogen bond donors and 10 hydrogen bond acceptors, emphasizing the importance of a small molecule drug having the ability to form hydrogen bonds within the binding site of the intended target (15). One result of hydrogen bond formation in the ligand complex is the relative loss of flexibility of the rest of the ligand, which implies that the availability of hydrogen bonding interactions greatly affects the conformation of the ligand (16). Also, because the NNRTI binding pocket in HIV-1 RT contains many residues with aromatic rings, pi stacking interactions have the potential to create more favorable

thermodynamic energetics to promote the binding of NNRTIs to HIV-1 RT (**Figure 2**) (17).

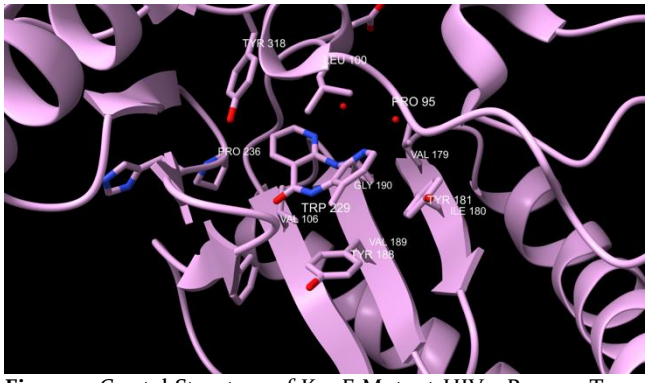


Figure 2 Crystal Structure of K101E Mutant HIV-1 Reverse Transcriptase in Complex with Nevirapine with aromatic and hydrophobic residues labeled (PDB: 2HND)

A search for hydrogen bond donor and acceptor positions as well as aromatic ring positions within each NNRTI therefore allows for the determination of spatial correlations between these motifs, providing insight into a more efficient method of screening potential NNRTI compounds.

RESULTS

PaDEL-Descriptor (PaDEL), is a software capable of generating 1,875 descriptors and 12 fingerprints for molecules, including descriptors corresponding to the hydrogen bond donor and acceptor count, and aromatic ring count (18). The 9 NNRTIs were screened in PaDEL. The resultant CSV file from PaDEL was analyzed in Python to derive the means and standard deviations of the 76 descriptors relating to the number of hydrogen-bond donors and acceptors and aromatic rings. We calculated the means in order to average all of these descriptors and the standard deviations to analyze the variance of the distribution. These standard deviations were used to calculate descriptors' correlations, where descriptors with a standard deviation of 1 or below were declared to be correlated with the structure of an NNRTI (**Table 2**).

Descriptor Name	Standard Deviation	Mean
nHBAcc2	0	6
nF10HeteroRing	0.333333	0.111111
nT10HeteroRing	0.333333	0.111111
n7Ring	0.333333	0.111111
nFG12Ring	0.333333	0.111111
n7HeteroRing	0.333333	0.111111
nFG12HeteroRing	0.333333	0.111111
nT7HeteroRing	0.333333	0.111111
nTG12HeteroRing	0.333333	0.111111
nF9Ring	0.440959	0.222222
nF10Ring	0.440959	0.222222
nF9HeteroRing	0.440959	0.222222
nT9HeteroRing	0.440959	0.222222
n3Ring	0.5	0.333333
n5Ring	0.527046	0.555556
n5HeteroRing	0.527046	0.555556
nT5HeteroRing	0.527046	0.555556
nF11Ring	0.666667	0.222222
nF11HeteroRing	0.666667	0.222222
nT11HeteroRing	0.666667	0.222222
n6Ring	0.707107	2.333333
nHBAcc3	0.707107	5.5
n6HeteroRing	0.781736	0.888889
nT6HeteroRing	0.781736	0.888889
nRing	0.866025	3.333333
nHeteroRing	0.881917	1.555556
nFRing	0.971825	0.777778

Table 2 Descriptor distributions where $\sigma < 1$ and $\mu \neq 0$

Standard deviations closer to zero indicate the distribution of that descriptor has low variance, or a higher degree of correlation among NNRTIs at that characteristic than descriptors with standard deviations that are farther from zero. Descriptors with z-scores with a magnitude of less than 1 were chosen in order to solely select descriptors which had high levels of correlation between the NNRTIs. Descriptors with a mean of zero and a standard deviation of zero were disregarded in statistical analysis as they indicated that the value of the descriptor was 0 among all of the NNRTIs, meaning that specific structure did not exist in any of the screened NNRTIs. However, descriptors with a mean of 0 and standard deviation of 0 were still pertinent to the classification purpose of the platform, as they imply that the characteristic is not expected to be present in an NNRTI compound, and thus serve as an additional filter for the classification of the compound. As shown in Table 2, the descriptor that had the highest similarity among all 9 of the screened NNRTIs was nHBAcc2, which has a mean of 6 and a standard deviation of 0. The descriptor nHBAcc2 corresponds to the number of hydrogen bond acceptors, and the data analysis indicates that all nine NNRTIs have 6 hydrogen bond acceptors. The high degree of similarity in the values of

each NNRTI at this descriptor validates the key roles that hydrogen bond interactions play in ligand binding.

Similarly, the trends of low standard deviation continued in descriptors associated with the number of rings in each NNRTI. Notable descriptors include nHeteroRing, the number of rings containing heteroatoms, and n6HeteroRing, the number of 6-membered rings containing heteroatoms. A lower standard deviation was observed in the distribution of n6HeteroRing than nHeteroRing across the screened molecules, indicating that the presence of 6-membered aromatic ring systems was also salient to the structure and function of NNRTIs, while other aromatic systems were observed less frequently across all nine of the screened NNRTIs.

In determining the interactions made between the NNRTI and the RT binding pocket, the positions of the aromatic rings, hydrogen bond donors, and hydrogen bond acceptors must be considered. As such, positional analysis was used to determine common locations for these components among the nine screened NNRTIs. To perform positional analysis of the aromatic rings and hydrogen bond donors and acceptors on the 9 NNRTIs studied, Python scripts were written to determine the quantity and specific locations of these molecular features (**Table 3**). After applying our identification algorithms to density functional theory (DFT) geometry-optimized models of the aforementioned NNRTIs, it was determined from the output files that the algorithms were functional and that the resulting .xyz files mapped to the exact coordinates of the hydrogen bond donors/acceptors and aromatic rings of all 9 NNRTIs. An example of this is shown in Figure 3, where Doravirine (**Figure 3A**) is used as input for the identification scripts, and the resulting output files successfully isolate the hydrogen bond donors and acceptors (**Figure 3B**) as well as the aromatic rings (**Figure 3C**). Due to a lack of normalized inputs, the determined coordinate locations of the NNRTI constituents were pertinent only to the structure of the NNRTI they belonged to. However, the analysis revealed a clustering of the aromatic rings and heteroatoms within each the NNRTIs. This indicates that the efficacy of the NNRTI's binding ability to the RT enzyme's binding pocket is reliant on structures that have many aromatic systems and heteroatoms close together.

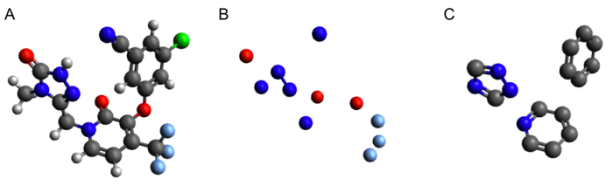


Figure 3 Results from hydrogen bond donor/acceptor and aromatic ring detection scripts (A) Geometry optimized structure of Doravirine (B) All hydrogen bond donors and acceptors present in Doravirine. (C) All aromatic rings present in Doravirine.

Name	Total # of Heteroatoms	# of H-Bond Donors/Acceptors	Total # of Rings	# of Aromatic Rings
Nevirapine	5	5	3	2
Delavirdine	10	9	4	3
Efavirenz	7	3	3	1
Etravirine	8	7	3	3
Rilpivirine	6	6	3	3
Doravirine	12	8	3	3
Fosdevirine	8	6	3	3
Lersivirine	6	6	2	1
RDEA806	10	7	4	3

Table 3 Heteroatoms and aromatic ring information of the 9 screened NNRTIs procured by the Python scripts.

DISCUSSION

The identification of highly correlated descriptors for HIV-1 RT provides valuable insight into future applications of high-throughput molecular screening. Specifically, these methods provide an efficient workflow for open source bioinformatic analysis of small molecules. Based on the descriptors that have a non-zero mean and low standard deviation, we can conclude the following: the structure of molecules belonging to the NNRTI class of drugs share commonalities in the number of aromatic atoms, aromatic bonds, hydrogen bond acceptors, and aromatic rings. Specifically, there is a greater presence of hydrogen bond acceptors in the NNRTIs studied, indicated by the highest mean values of 6 and 5.5 in the nHBAcc2 and nHBAcc3 descriptors respectively. This indicates that hydrogen bond acceptors in particular are mainly responsible for the formation of specific intermolecular hydrogen bonds between the NNRTI and RT binding pocket. Furthermore, among the mean values of descriptors for aromatic rings, the highest value of 2.333333 is associated with the number of 6-membered rings (n6Ring), indicating a strong presence of these substructures in all 9 NNRTIs. It is clear that for all compounds studied, the majority of molecular rings present in each molecule are aromatic and contain 6 atomic constituents (**Table 1**). It is evident that 6-membered aromatic rings (including heterocyclic structures) play a significant role in determining the specific conformation of the overall NNRTI when bound to the RT enzyme, and therefore contributes to the compound's overall success.

A limitation of the method described is the exclusion of rotatable bonds and/or conformational flexibility in the ligand. As a result, all molecules screened with this algorithm were represented as rigid structures, and only the relative positions of the heteroatoms and aromatic rings on the geometrically-optimized NNRTIs were determined instead of sampling all possible conformers of each compound. In traditional docking algorithms, bonds defined as rotatable are able to create additional conformers within the HIV-1 RT binding pocket.

While the results presented here were specific to small molecule antivirals targeting HIV-RT, it is conceivable that such an approach can be applied to any biomolecular target with a set of known ligands that bind to a well-defined binding pocket. In addition, this method of screening allows for the identification of descriptive variables in motif identification for machine learning approaches. These methods can potentially have broad applications in any campaign involving screening for molecules with common moieties. The methodology described might serve as an effective supplement to an HTVS campaign in filtering large libraries of molecules into classes that contain the desired pharmacophoric features identified in smaller libraries of lead structures.

MATERIALS AND METHODS

Quantified Characteristic Analysis PaDEL-Descriptor (PaDEL), an open-source software capable of calculating molecular descriptors and fingerprints, was identified to be a suitable software to quantify complex chemical characteristics of the NNRTI molecule. PaDEL is capable of generating 1,875 descriptors and 12 fingerprints for molecules, including descriptors corresponding to the hydrogen bond donor count, hydrogen bond acceptor count, and aromatic ring count (**Table 4**). Molecules can be analyzed in batch and the results are presented in a CSV format that can be parsed by scripts. The simplified molecular-input line-entry system (SMILES) code was obtained for the nine NNRTIs chosen as data points from PubChem (19). These codes were then submitted to the PaDEL-Descriptor interface to obtain the resultant CSV files. In PaDEL, all NNRTIs went through geometry optimization using the Merck Molecular Force Field (MMFF94) to 10,000 steps.

Descriptor Type	Descriptor Names
Aromatic Atoms Count	naAromAtom
Aromatic Bonds Count	nAromBond
Hydrogen Bond Acceptor Count	nHBAcc, nHBAcc2, nHBAcc3, nHBAcc_Lipinski
Hydrogen Bond Donor Count	nHBDon, nHBDon_Lipinski
Ring Count	nRing, n3Ring, n4Ring, n5Ring, n6Ring, n7Ring, n8Ring, n9Ring, n10Ring, n11Ring, n12Ring, nG12Ring, nFRing, nF4Ring, nF5Ring, nF6Ring, nF7Ring, nF8Ring, nF9Ring, nF10Ring, nF11Ring, nF12Ring, nFG12Ring, nHeteroRing, n3HeteroRing, n4HeteroRing, n5HeteroRing, n6HeteroRing, n7HeteroRing, n8HeteroRing, n9HeteroRing, n10HeteroRing, n11HeteroRing, n12HeteroRing, nG12HeteroRing, nFHeteroRing, nF4HeteroRing, nF5HeteroRing, nF6HeteroRing, nF7HeteroRing, nF8HeteroRing, nF9HeteroRing, nF10HeteroRing, nF11HeteroRing, nF12HeteroRing, nFG12HeteroRing, nTHeteroRing, nT4HeteroRing, nT5HeteroRing, nT6HeteroRing, nT7HeteroRing, nT8HeteroRing, nT9HeteroRing, nT10HeteroRing, nT11HeteroRing, nT12HeteroRing, nTG12HeteroRing

Table 4 Descriptor names in correspondence to their descriptor type

Data Analysis The Python package pandas was used to create a script to parse the CSV file outputted from PaDEL (20-22). The CSV file generated from PaDEL was a matrix with columns that consisted of descriptors and rows for each of the nine NNRTIs and their values for each descriptor. The CSV was then imported into python and the descriptors shown in Table 4 were isolated. Following this, the means and standard deviations for each descriptor were calculated, and data was ordered by the standard deviation in ascending order, while attributes with values of 0 for both the mean and standard deviation were omitted. Finally this dataframe was outputted as a CSV.

Position Analysis Determination of precise locations of hydrogen bond donors and aromatic rings in a three-dimensional environment was carried out through the creation of Python scripts to parse coordinates of atoms in an .xyz file of each geometry-optimized NNRTI. RDKit, an external cheminformatics Python library, was used to search for and identify individual atomic constituents of any aromatic rings present on the molecule being analyzed (23). RDKit contains various functions able to identify properties associated with molecular compounds, and was deemed an appropriate library to integrate into the aromatic ring detection script due to flexibility in its ability to identify aromatic bonds rather than entire aromatic systems. As such, the user-defined model for aromaticity used to identify these aromatic bonds was an intramolecular bond with a bond order of 1.5. The algorithm would isolate all rings from the NNRTI and determine the aromaticity of each ring through confirmation of all intramolecular bonds within the ring having a bond order of 1.5 - indicative of a fully conjugated ring system.

REFERENCES

(1) Shoichet, Brian K. "Virtual Screening of Chemical Libraries." Nature, vol. 432, no. 7019, 2004, pp. 862–865., doi:10.1038/nature03197.

(2) Singh, Nidhi, et al. "Identification of Novel Inhibitors of Mycobacterium Tuberculosis PknG Using Pharmacophore Based Virtual Screening, Docking, Molecular Dynamics Simulation, and Their Biological Evaluation." Journal of Chemical Information and Modeling, vol. 55, no. 6, 2015, pp. 1120–1129., doi:10.1021/acs.jcim.5b00150.

(3) Kaserer, Teresa, et al. "Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases." Molecules, vol. 20, no. 12, 2015, pp. 22799–22832., doi:10.3390/molecules201219880.

(4) Published: Sep 09, 2019. "The Global HIV/AIDS Epidemic." KFF, 30 Sept. 2019, www.kff.org/global-health-policy/fact-sheet/the-global-hivaids-epidemic/.

(5) "The HIV Life Cycle Understanding HIV/AIDS." National Institutes of Health, U.S. Department of Health and Human Services, 1 July 2019, aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/19/73/the-hiv-life-cycle.

(6) "HIV Strains and Types." Avert, 26 Feb. 2019, www.avert.org/professionals/hiv-science/types-strains.

(7) Seckler, James M., et al. "Allosteric Suppression of HIV-1 Reverse Transcriptase Structural Dynamics upon Inhibitor Bind-

ing." Biophysical Journal, vol. 100, no. 1, 2011, pp. 144–153., doi:10.1016/j.bpj.2010.11.004.
(8) Sluis-Cremer, Nicolas, et al. "Conformational Changes in HIV-1 Reverse Transcriptase Induced by Nonnucleoside Reverse Transcriptase Inhibitor Binding." Current HIV Research, vol. 2, no. 4, 2004, pp. 323–332., doi:10.2174/1570162043351093.
(9) Chen, Xuwang, et al. "Recent Advances in DAPYs and Related Analogues as HIV-1 NNRTIs." Current Medicinal Chemistry, vol. 18, no. 3, 2011, pp. 359–376., doi:10.2174/092986711794839142.
(10) Béthune, Marie-Pierre De. "Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs), Their Discovery, Development, and Use in the Treatment of HIV-1 Infection: A Review of the Last 20 Years (1989–2009)." Antiviral Research, vol. 85, no. 1, 2010, pp. 75–90., doi:10.1016/j.antiviral.2009.09.008.
(11) Das, Kalyan, et al. "Crystallography and the Design of Anti-AIDS Drugs: Conformational Flexibility and Positional Adaptability Are Important in the Design of Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors." Progress in Biophysics and Molecular Biology, vol. 88, no. 2, 2005, pp. 209–231., doi:10.1016/j.pbiomolbio.2004.07.001.
(12) Behja, Wollela, and Mudin Jemal. "Anti-HIV Drug Discovery, Development and Synthesis of Delavirdine: Review Article." International Research Journal of Pure and Applied Chemistry, 2019, pp. 1–16., doi:10.9734/irjpac/2019/v20i330137.
(13) "Lersivirine Clinical Trials, Side Effects." National Institutes of Health, U.S. Department of Health and Human Services, aidsinfo.nih.gov/drugs/510/lersivirine/0/patient.
(14) De Clercq, Erik. "Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs): Past, Present, and Future." Chemistry & Biodiversity, vol. 1, no. 1, 2004, pp. 44–64., doi:10.1002/cbdv.200490012.
(15) Lipinski, Christopher A. "Lead- and Drug-like Compounds: the Rule-of-Five Revolution." Drug Discovery Today: Technologies, vol. 1, no. 4, 2004, pp. 337–341., doi:10.1016/j.ddtec.2004.11.007.
(16) Kubinyi, Hugo. 3D QSAR in Drug Design. Kluwer/Escom, 2000.
(17) Ren, Jingshan et al. "Structural insights into mechanisms of non-nucleoside drug resistance for HIV-1 reverse transcriptases mutated at codons 101 or 138." The FEBS journal vol. 273,16 (2006): 3850-60. doi:10.1111/j.1742-4658.2006.05392.x
(18) Yap, Chun Wei. "PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints." Journal of Computational Chemistry, vol. 32, no. 7, 2010, pp. 1466–1474., doi:10.1002/jcc.21707.
(19) Kim, Sunghwan, et al. "PubChem 2019 Update: Improved Access to Chemical Data." Nucleic Acids Research, vol. 47, no. D1, 2018, doi:10.1093/nar/gky1033.
(20) "G. van Rossum, Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995."
(21) Mckinney, Wes. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference.
(22) The pandas development team. (2020, March 18). pandas-dev/pandas: Pandas 1.0.3 (Version v1.0.3). Zenodo. http://doi.org/10.5281/zenodo.3715232.
(23) RDKit: "Open-Source Cheminformatics Software." RDKit, www.rdkit.org/.