Dipole Moment, Solvation Energy, and Ovality Account for the Variations in the Biological Activity of HIV-1 Reverse Transcriptase Inhibitor Fragments

Derick Erl P. Sumalapao[1,2*], Jose Isagani B. Janairo[2] and Nina G. Gloriani[1]

[1]*Department of Medical Microbiology, College of Public Health, University of the Philippines Manila, Manila, Philippines.*
[2]*Department of Biology, College of Science, De La Salle University, Manila, Philippines.*

Authors' contributions

This work was carried out in collaboration between all authors. All authors read and approved the final manuscript.

Original Research Article

ABSTRACT

Objective: A computational approach was employed to determine the interaction of molecular descriptors and the biological activity of the different fragments of HIV-1 reverse transcriptase inhibitors (RTIs).
Methods: Using multiple linear regression analysis and leave-one-out validation method, a quantitative structure activity relationship (QSAR) model was developed to relate the biological activity (log IC_{50}) of the different fragment-sized compounds against HIV-1 RT(WT) DNA-dependent DNA polymerase and molecular descriptors of these compounds.
Results: QSAR model identified dipole moment, solvation energy, and ovality of fragment-sized compounds to confer reverse transcriptase inhibitory action. A highly significant correlation with log P, molecular weight, polarizability, molecular energy, zero-point energy, constant volume heat capacity at 298 K, and entropy was identified to account for the variations in the potency of RTIs. An increase in ovality, log P, and molecular weight of the fragment-sized compound renders a more

---

Corresponding author: E-mail: derick.sumalapao@dlsu.edu.ph;

active reverse transcriptase inhibition.

Conclusion: The quality of the established QSAR model has been validated and demonstrates its potential as a tool for computational design and synthesis of next generation RTIs.

1. INTRODUCTION

In 2016, individuals infected with human immunodeficiency virus (HIV) were approximately 36.7 million with around one million HIV-related mortality globally [1]. Considering that the number of new HIV infections continues to rise, it remains to be a major global public health issue. Given the magnitude of this pandemic, novel classes of antiretroviral drugs with mechanisms intended to treat and prevent HIV are needed [2]. In HIV replication, HIV-encoded reverse transcriptase (RT), the first successful enzyme target for HIV therapy [3], is responsible for the reverse transcription of viral RNA to proviral DNA [4]. Inhibition of the function of this enzyme, either genetic or chemical approach [5], can suppress viral replication. However, of the 26 anti-HIV drugs, only 13 can target HIV-1 RT [3].

In this study, fragments of HIV-1 reverse transcriptase (RT) inhibitors [2] along with their molecular descriptors accounting variations in their specific biological activity (log IC_{50}) were examined using quantitative structure activity relationship (QSAR) models. In drug design, QSAR is employed to relate the experimentally obtained biological activity of a compound with its molecular properties [6]. Findings of this study offer critical insight in the design of novel RTIs as therapeutic interventions for HIV. Moreover, the generated QSAR model unveils the relevant physicochemical properties of HIV-1 RT(WT) DNA-dependent DNA polymerase (DDDP) and the extent of the influence of a given molecular descriptor to effect inhibition of the reverse transcriptase activity (log IC_{50}) using multiple linear regression analysis and validated by leave-one-out technique.

2. MATERIALS AND METHODS

2.1 Molecular Descriptors and Biological Activity

Information on the biological activity (IC_{50}) of the 23 fragment-sized RTIs (Table 1) were obtained from previous study [2] with geometry optimized structures generated using Spartan'10 (Wavefunction, Inc.). Using density functional theory, the molecular descriptors of these fragment-sized compounds were calculated employing the B3LYP 6-31G* basis set. The calculated molecular descriptors were the molecular energy, highest occupied molecular orbital (HOMO) energy, lowest unoccupied molecular orbital (LUMO) energy, chemical potential, molecular hardness, molecular softness, electrophilicity, dipole moment, solvation energy, molecular weight, area, volume, polar surface area, ovality, log P, polarizability, zero-point energy, constant volume heat capacity at 298 K, enthalpy, entropy, and Gibb's free energy.

2.2 Model Construction

In this study, quantitative structure activity relationship (QSAR) model correlates the experimental biological activity (log IC_{50}) of fragment-sized RTIs with their structural and physicochemical properties using multiple linear regression analysis.

Multiple linear regression analysis, a commonly employed method in QSAR, is simple and reproducible [7]. In this study, the *n* molecular descriptors (X_i) were related to the biological activity (Y) using a linear function $E(Y/X) = \alpha + \sum_1^n \beta_i X_i$ with determined parameters α and β. The regression model was obtained using backward elimination in V12.0 STATA®. To assess the quality of the obtained model, squared correlation coefficient (r^2) was calculated [8]. Moreover, with the molecular predictors included in the final model, using bivariate correlation studies, multicollinearity among these predictors was examined.

2.3 Model Validation

Using Leave-One-Out (LOO) technique [9], predictive ability of the obtained QSAR model was evaluated. In LOO cross-validation [10], the training set was basically modified by having a

Table 1. Chemical structures of fragments inhibiting HIV-1 RT(WT) DNA-dependent DNA polymerase activity [2]
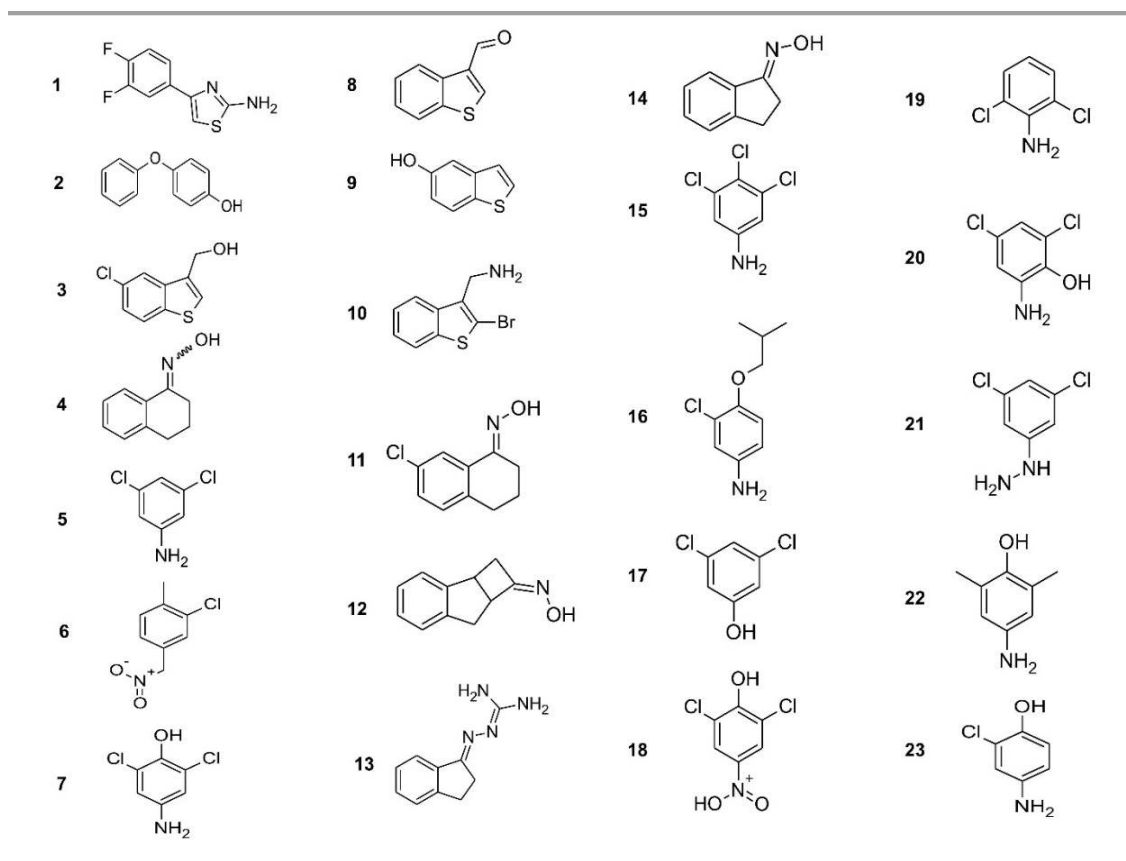
compound eliminated from the entire data set. Subsequently, with the remaining *n-1* compounds, a QSAR model was constructed and used to determine the biological activity of the deleted compound. These predicted response values were obtained for all the training set compounds and were used for the calculation of different internal validation parameters.

A cross-validated r^2 (q^2) value was determined (Equation 1) to assess the statistical validity of the obtained model. These q^2 values were obtained from the model PRESS (prediction error sum of the squares) (Equation 2).

$$q^2 = 1 - \frac{\text{PRESS}}{\Sigma(Y_{obs\,(train)} - \bar{Y}_{train})^2} \tag{1}$$

$$\text{PRESS} = \Sigma(Y_{obs\,(train)} - Y_{pred\,(train)})^2 \tag{2}$$

Y_{obs} and Y_{pred} are the experimentally observed and LOO-predicted biological activity of the compound, respectively. $Y_{obs(train)}$ and $Y_{pred(train)}$ are the observed and predicted biological activity of the training set compounds, respectively, using the LOO technique. These two measures, r^2 and q^2, are useful determinants of model validity, which quantify the goodness-of-fit and the goodness of prediction [8] of the obtained QSAR models, and are considered satisfactory when $r^2 > 0.6$ and $q^2 > 0.5$ [11].

3. RESULTS AND DISCUSSION

Molecular descriptors of the fragment-sized compounds were defined using the density functional theory. Preliminary analysis identified and removed molecular descriptors that had invariant values in the 23 fragment-sized compounds, subsequently leaving 21 molecular descriptors. Considering that chemical potential, molecular hardness, molecular softness, and electrophilicity were derived from the HOMO and LUMO values of the molecule, then these four molecular descriptors were removed from the list

3

of 21 independent variables to avoid possible presence of multicollinearity. In this study, there were only 23 available fragment-sized compounds, and in model construction, the presence of one regressor variable in the model requires five samples [8,12] suggesting that the appropriate QSAR model should contain at most four descriptors.

When multiple linear regression analysis was performed on the 23 fragment-sized compounds using backward elimination method, four independent variables were included in the model (r^2 = 0.747, p = 0.004). The obtained QSAR model (Equation 3) has four molecular predictors and is sufficient to explain the variations in the biological activity of the fragment-sized compounds. These molecular descriptors explaining almost 75% of the variations in the biological activity (log IC_{50}) of the fragment-sized compounds were dipole moment (DIPOLE, p = 0.022), solvation energy (SOLVATION, p = 0.028), area-volume ratio (AVR, p = 0.009), and ovality (OVALITY, p = 0.032). The combination of dipole moment and solvation energy explains almost 50% of the fluctuations in the response variable, while the addition of AVR in the set of independent variables would account for 57% of the variations in the log IC_{50} values.

log IC_{50} = −11.757(±11.436) − 0.231(±0.090)*DIPOLE − 0.017(±0.007)*SOLVATION + 21.882(±7.160)*AVR − 17.648(±7.401)*OVALITY   (3)

Having identified the statistically significant molecular descriptors in the proposed QSAR model, assumptions of multiple linear regression analysis were assessed. These assumptions include linearity, independence, normality, equal variance, presence of outliers, and multicollinearity.

Linearity of the constructed regression function was examined using the scatter plot of the predicted values with the residuals (Fig. 1). Since the plot of the residuals does not exhibit any linear pattern, this suggests that there exists a linear relationship between the identified molecular descriptors and the biological activity of the fragment-sized compounds. For the test of independence of the error terms, the null hypothesis states that the data is random. Twelve observations had residual values below 0.036 (z = 0.22, p = 0.82), suggesting that the data is random.

In assessing normal distribution of the error terms, a scatter plot between residuals and the inverse normal values was generated (Fig. 2) which reflects agreement with normality. Moreover, Shapiro-Wilk test further confirmed the normal distribution of the residuals (p = 0.545).

For heteroskedasticity, Breusch-Pagan and Cook-Weisberg test was used (p = 0.829) suggesting that the error terms had a constant variance. Moreover, no outlier was identified in the data set as reflected in the generated boxplot of the residuals (Fig. 3). Lastly, the presence of multicollinearity was examined among the identified molecular descriptors using their respective variance inflation factor (VIF). If VIF exceeds 10, multicollinearity among independent variables is present. The VIF of DIPOLE, AVR, OVALITY, and SOLVATION is 1.55, 1.45, 1.19, and 1.06, respectively. There is no multicollinearity among the identified molecular descriptors since the VIFs were all less than 10.

The negative coefficients of DIPOLE, SOLVATION, and OVALITY suggest that increasing their respective values would result to a lower log IC_{50} which improves potency of the RTIs, while the positive regression coefficient for AVR means higher log IC_{50} concentrations for fragment-sized compounds with larger AVR values. Moreover, to visualize the influence of other molecular descriptors generated in this study on the biological activity of the RTIs, these molecular descriptors were correlated with the relevant variables included in the QSAR model (Table 2).

It appears that there is some interrelatedness between the identified relevant regressors in the constructed QSAR model with the other molecular descriptors. Apparently, dipole moment is significantly directly correlated with the logarithm (base 10) of negative of molecular energy (LOG ENERGY) which suggests that the stability of the molecule influences the biological activity of the compound. Moreover, solvation energy (SOLVATION) is highly significantly directly correlated with log P (LOG P). However, solvation energy is highly significantly inversely correlated with highest occupied molecular orbital (HOMO) energy. This indicates a lower HOMO value corresponds to a lower log IC_{50} value. The positive coefficient of AVR, which is highly significantly inversely correlated with polarizability (POLAR), zero-point energy (ZPE), and constant volume heat capacity (CV)

suggests that more polarizable, higher ZPE, and higher CV variants of RTIs would have lower log IC_{50} values. Ovality of the molecule is highly significantly directly correlated with molecular weight (WEIGHT), CV, ZPE, and entropy (S). These results suggest that a more spherical fragment-sized compound with higher WEIGHT, CV, ZPE, and S values would be a more potent inhibitor of reverse transcriptase.

Moreover, a cross-validation was performed using the leave-one-out (LOO) method (Fig. 4) regarding the predictive ability of the obtained QSAR model. The generated QSAR model can

be utilized in designing next generation RTIs ($q^2 = 0.77$).

Given the findings of the present study, molecular descriptors of fragments of HIV-1 RT drugs can possibly lead to better elucidation of novel inhibitors of HIV-1 RT allosteric sites and enzymatic activity. The allosteric binding to HIV-1 RT [13] was identified to have a critical dependence on the conformational flexibility [14]. The presence of new allosteric pockets in the HIV-1 RT [15,16] can induce conformational changes in substrate binding site of reverse transcriptase [17] resulting to its reduced activity.
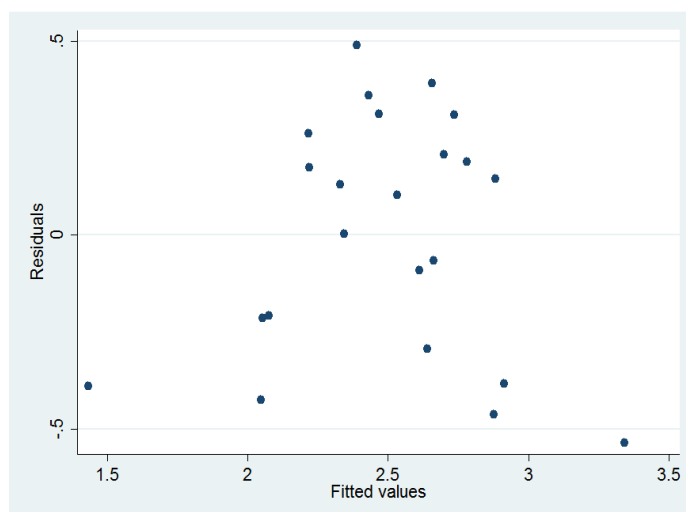

Fig. 1. Linearity assessment between residual and predicted values of the biological activity (log IC_{50}) of the 23 compounds
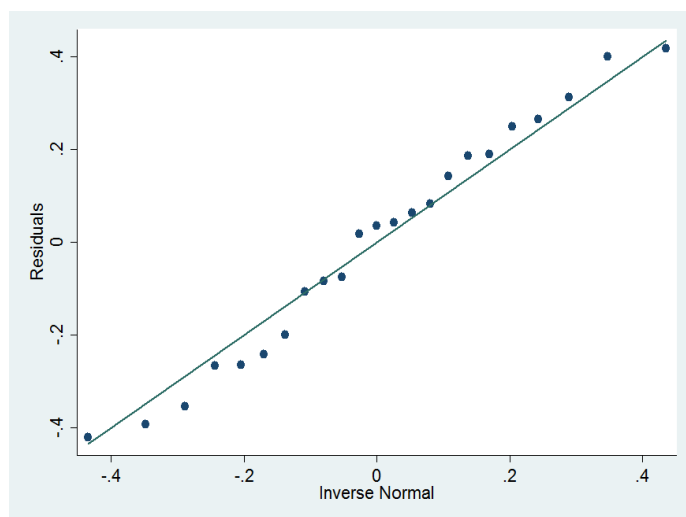

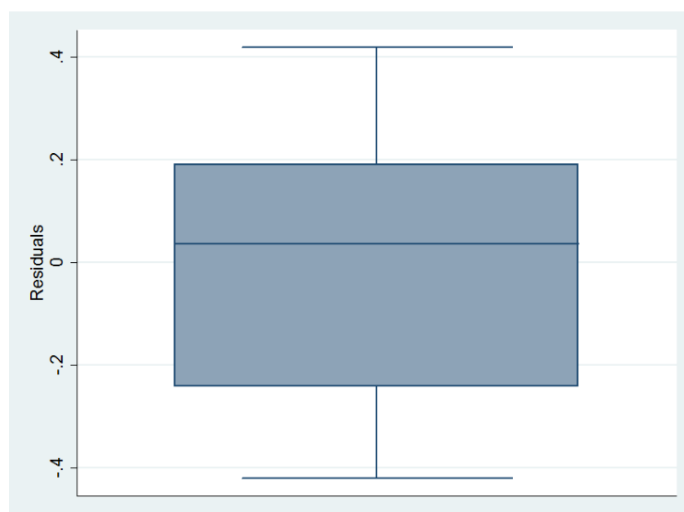**Fig. 2. Graphical representation of the normal distribution of the error terms**

Fig. 3. Box-plot of the residuals for outlier detection

Table 2. Correlation coefficients between the identified relevant regressors in the QSAR model and some of the molecular descriptors

	Dipole	Solvation	AVR	Ovality
LOG ENERGY	0.478*	0.072	0.405	0.067
HOMO	-0.152	-0.664**	0.011	0.233
WEIGHT	0.403	-0.076	-0.115	0.648**
LOG P	0.192	0.722**	0.029	0.196
POLAR	-0.069	-0.265	-0.608**	0.817**
ZPE	-0.349	-0.232	-0.638**	0.625**
CV	0.049	-0.274	-0.419*	0.904**
S	0.306	-0.205	-0.158	0.958**

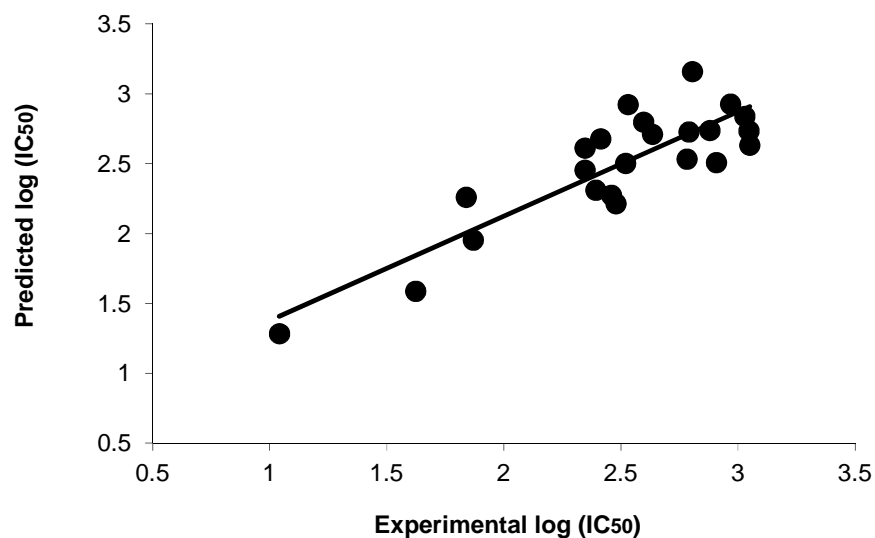*(0.05) and **(0.01) significance level, two-tailed*

Fig. 4. Experimental log IC_{50} versus calculated log IC_{50} values (Leave-One-Out cross-validation method, $q^2 = 0.77$, n = 23)

4. CONCLUSION

Computational approach such as quantitative structure activity relationship studies identified the activity and designed the structure of new compounds. In this study, a QSAR model was constructed to identify molecular descriptors of HIV-1 RT(WT) DNA-dependent DNA polymerase influencing their biological activity. Using multiple linear regression analysis on a set of 23 fragments derived from known reverse transcriptase inhibitors, four molecular descriptors were identified to confer reverse transcriptase inhibitory action. These molecular descriptors were dipole moment, solvation energy, ovality, and area to volume ratio of RTIs. A highly significant correlation with log P, molecular weight, polarizability, molecular energy, zero-point energy, constant volume heat capacity, and entropy was identified to account for the variations in the potency of RTIs. Increase in ovality, log P, and molecular weight of the fragment-sized compound render a more active reverse transcriptase inhibition. The established QSAR model has been validated and demonstrates its potential as a tool for computational design and synthesis of next generations RTIs.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. World Health Organization. HIV/AIDS Fact Sheet N_360; 2017.
   Available:http://www.who.int/mediacentre/factsheets/fs360/en/
   (Accessed 15 December 2017)
2. La J, Latham CF, Tinetti RN, Johnson A, Tyssen D, Huberd KD, Sluis-Cremer N, Simpson JS, Headey SJ, Chalmers DK, Tachedjian G. Identification of mechanistically distinct inhibitors of HIV-1 reverse transcriptase through fragment screening. PNAS. 2015;112(22):6979–6984.
3. Esposito F, Corona A, Tramontano E. HIV-1 reverse transcriptase still remains a new drug target: Structure, function, classical inhibitors, and new inhibitors with innovative mechanisms of actions. Mol Biol Int. 2012;2012:586401.
4. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamart S, Greust J, Dausgeut C, Axler-Bin C, Vézinit-Brum F, Rouzioux C, Rosenbaum W, Montagnier L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immunodeficiency syndrome (AIDS). Science. 1983;220:868–871.
5. Schinazi RF, Larder BA, Mellors JW. Mutations in retroviral genes associated with drug resistance: 1999–2000 update. Int. Antivir. News. 1999;7:46–69.
6. Warr WA. Some Trends in Chem(o)informatics. In: Chemoinformatics and computational chemical biology, Bajorath J. Ed., Humana Press, New York. 2011;1–38.
7. Snedecor GW, Cochran WG. Statistical methods. Oxford and IBH. New Delhi, India; 1967.
8. Leach AR, Gillet VJ. Computational models. In: An Introduction to Cheminformatics, Revised Edition, Springer, The Netherlands. 2007;75–97.
9. Gong G. Cross-validation, the Jacknife and the Bootstrap: Excess error estimation in forward logistic regression. J. Am. Stat. Assoc. 1986;81(393):108–113.
10. Roy K, Mitra I. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. Comb Chem High Throughput Screen. 2011;14:450-474.
11. Golbraikh A, Tropsha A. Beware of q^2! J Mol Graph Model. 2002;20:269-276.
12. Tabachnick BG, Fidell LS. Multiple Regression. In: Using Multivariate Statistics. 5th Edition, London: Pearson/Allyn & Bacon. 2007;117–194.
13. Kohlstaedt L, Wang J, Friedman J, Rice P, Steitz T. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science. 1992;256:1783-1790.
14. Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG. Collective motions in HIV-1 reverse transcriptase: Examination of flexibility and enzyme function. J Mol Biol. 1999;285(3):1023–1037.
15. Bauman JD, Patel D, Dharia C, Fromer MW, Ahmed S, Frenkel Y, Vijayan RSK, Eck JT, Ho WC, Das K, Shatkin AJ, Arnold E. Detecting allosteric sites of HIV-1 reverse transcriptase by X-ray crystallographic fragment screening. J Med Chem. 2013;56(7):2738–2746.

16. Freisz S, Bec G, Radi M, Wolff P, Crespan E, Angeli L, Dumas P, Maga G, Botta M, Ennifar E. Crystal structure of HIV-1 reverse transcriptase bound to a nonnucleoside inhibitor with a novel mechanism of action. Angew Chem Int Ed Engl. 2010;49(10):1805–1808.

17. Peletskaya EN, Kogon AA, Tuske S, Arnold E, Hughes SH. Nonnucleoside inhibitor binding affects the interactions of the fingers subdomain of human immunodeficiency virus type 1 reverse transcriptase with DNA. J. Virol. 2004;78:3387-3397.