In silico quantitative structure-activity relationship analysis of a highly potent experimental HIV-1 protease inhibitor, GRL10413.

[1]Madhumita Aggunna and [1, 2]Ravikiran S. Yedidi*

[1]The Center for Advanced-Applied Biological Sciences & Entrepreneurship (TCABS-E), Visakhapatnam 530016, AP; [2]Department of Zoology, Andhra University, Visakhapatnam 530003, AP.
(*Correspondence to RSY: tcabse.india@gmail.com)

Keywords: HIV/AIDS, HIV protease, protease inhibitors, qsar, analogs, binding affinities.

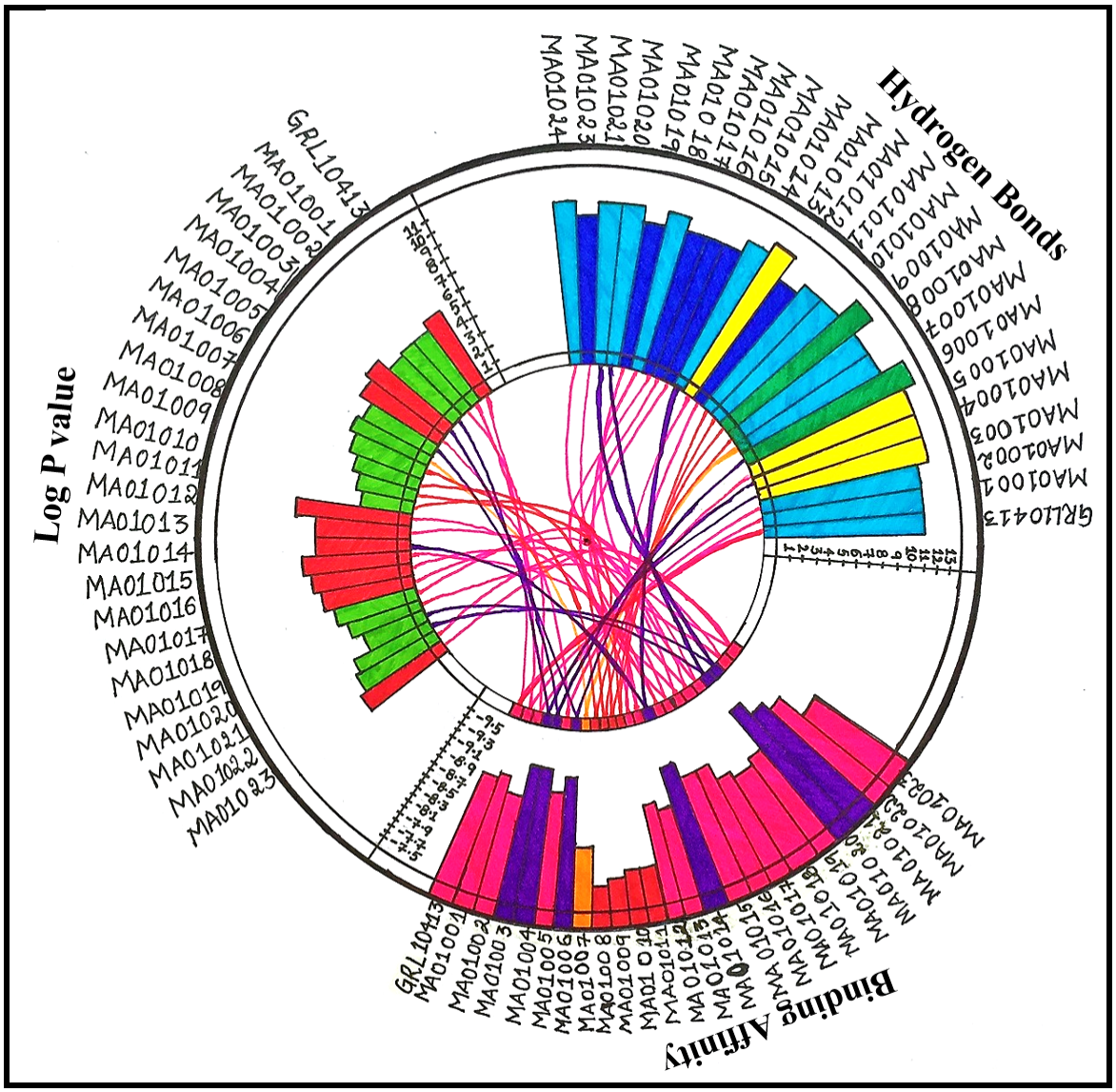

Figure 1. Hand drawn circular plot correlating the binding affinities of 23 analogs of GRL10413 with their corresponding number of hydrogen bonds and predicted LogP values analyzed in this study. Log-P value: (2.5 to 5.0) & (5.1 to 7.0); Hydrogen bonds: (10 H-bonds), (11 H-bonds), (12 H-bonds) & (13 H-bonds); Binding affinity (KCal/mol): (-7.5 KCal/mol to -7.9 KCal/mol), (-8.0 KCal/mol to -8.5 KCal/mol), (-8.6 KCal/mol to -9.0 KCal/mol) & (-9.1 KCal/mol to -9.5 KCal/mol).

Human Immunodeficiency Virus (HIV) has the highest mutation rate helping the viral early resistance to drugs that are currently in use. However, HIV-1 Protease has been a prime target for designing the drugs (protease inhibitors) due to the success rates of latest generation drugs such as darunavir. GRL10413 (an analog of darunavir) is a novel non-peptidic HIV-1 protease inhibitor that has shown highly potent antiviral activity against various multi-resistant HIV variants. In this study, GRL10413 was considered as the parent compound to create twenty three analogs whose molecular docking was performed using *Autodock tools* and *Autodock Vina* for *in silico* Quantitative Structure Activity Relationship (QSAR) evaluation in order to check whether the *in silico* binding affinity can further be enhanced significantly in the analogs. The hydrogen bonding profiles and predicted LogP values of these analogs were also evaluated along with their binding affinities. The overall highest increase in the binding affinity was <1.0 kcal/mol. In future, the analogs that showed better binding affinities should be evaluated against the multi-resistant variants of HIV-1.

Human Immunodeficiency Virus (HIV) infect the immune cells (T_c cells, macrophages and dendritic cells) leading to lower immunity and decrease in cell-mediated immunity which leads to low immunity and quick infections and diseases [1, 2, 3, 6]. When HIV infection is left untreated it leads to the accumulation of opportunistic pathologies and results in Acquired Immunodeficiency Syndrome (AIDS). Approximately 37.7 million cases of HIV are reported in 2020 out of which 25.4 million cases are reported towards Africa. For about six lakhs eight thousand deaths are HIV-related infections and complications and approximately 1.5 million cases are reported to be HIV positive [1, 22]. Targeting the HIV infection for developing medications and treatments has become a challenge to scientists due high rate of mutations and drug resistance. Scientists believed that investigating the HIV viral replication would assist them in obtaining accurate specifics in designing the drugs for better treatment which might be the reason for choosing HIV-1 protease as their prime target for developing the protease inhibitors [15].

The HIV-1 protease is a retroviral aspartyl protease which is an enzyme that plays a significant role in HIV replication and survival. HIV remains uninfectious if the virus lacks efficient HIV protease [4, 5]. The HIV protease hydrolyzes the peptide bond and cleaves Gag-pol at nine sites which helps in the viral maturation that produces the functional proteins [15]. HIV is a virus with a higher rate of mutations that is challenging for treating the disease. HIV-1 protease is 22 kDa in its molecular weight which is a homodimer where each monomer consists of 99 amino acids [19]. The active site consists of a catalytic triad sequence of Asp-Thr-Gly (Asp25-Thr26-Gly27, which is an aspartyl group complex). The Asp25 of each monomer have to bind with the other monomer of the protease which implies that the HIV-1 protease can only function as a dimer. The Asp25 is protonated in one monomer while the other monomer consists of the deprotonated Asp25 due to the variation in the acid dissociation constant. The deprotonated Asp25 helps in the cleavage of peptide bonds between two amino acids which recurs back to its original deprotonated state [17, 20]. When the genetic material (RNA) of HIV is released into the cell, reverse transcriptase helps in converting the viral RNA into DNA which gets integrated into the host DNA with the help of an integrase enzyme. The HIV-1 protease is present in between the reverse transcriptase and the $p6^{pol}$. Reverse transcriptase is located towards the C-terminus of the protease and $p6^{pol}$ is located towards the N-terminus [18]. The reverse transcribed viral DNA either stays in the nucleus of the host cell in a dormant stage or it gets transcribed into mRNAs where the host cell

translates the mRNA into functional proteins like Gag-Pol polyproteins along with the mature HIV-1 protease [16, 17]. The mature HIV-1 protease can auto-synthesize the new mature HIV-1 proteases which helps the virus to replicate its genetic material inside the host cell leading to the production of a huge number of viruses [14, 15]. Also a mature HIV-1 protease is specific to nine sites of Gag-pol protein which gets cleaved and each of the subunits are processed to procure developed and functional proteins that are necessary in viral replication. The HIV-1 protease plays a very critical role in the viral survival and replication which is why it had high focus in targeting it for developing protease inhibitors. A weak or absence of active HIV-1 protease will lead to the non-infectious virions. Indinavir, Saquinavir, Ritonavir, Nelfinavir, Lopinavir, Amprenavir, Fosamprenavir, Atazanavir, Tipranavir, and Darunavir are the drugs that target the HIV-1 protease enzyme that are FDA approved till date.

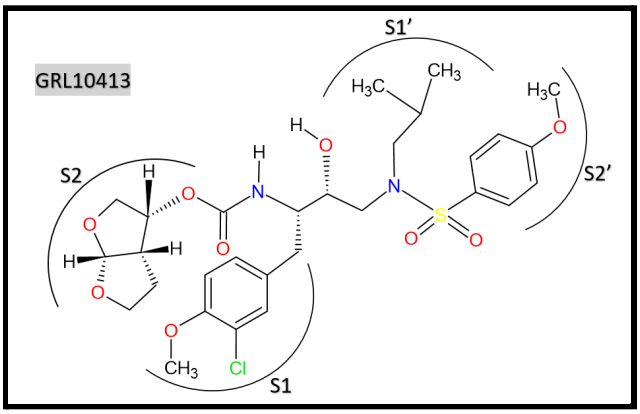

Figure 2. Two-dimensional structure of GRL10413. S2, S1, S1' and S2' represent the binding pockets from HIV-1 protease.

Though there are drugs available in the market for the treatment of HIV-AIDS it is not an easy task to cure the patients as the virus appears to attain resistance to the current existing inhibitors due its high rate of mutations in the genetic material. The combination of different drugs could be a better approach rather than using one drug at once [21]. The design of GRL-10413 molecule which is a novel non - peptidic HIV-1 protease inhibitor has efficient antiviral activity and is observed to show its activity on multidrug-resistant HIV variants. The rate of resistance toward GRL-10413 is reduced when compared to the Amprenavir (for upto the concentration of 5μM) toward the HIV-1_{NL4-3}, variants controlling the viral replication and shutting off the activity of infection. Darunavir has a similar structure to Ampiravir with an extra Tetrahydrofuran (THF) at the P2 position. The design of the nonpeptidic protease inhibitor GRL-10413 is the modification of the amine group of darunavir to the methoxy group at P2'. This modification has made GRL-10413 to show delayed development of antiviral resistance by constraining the viral replication and infectivity.

Quantitative Structure-Activity Relationship (QSAR) plays a significant role in understanding the interactions of the chemical compounds with the biological molecules which is based on the mathematical and statistical relations. The QSAR analysis takes an advantageous part in various ways by identifying the best analog with highest binding affinity through the computational analysis which helps in prediction of the chemical compounds performance, decreasing the costs of the wet lab experimentation and predicting the toxicity and other relevant requirements in identifying the best compound [8]. In this study, the analogs of the GRL-10413 are developed using the Autodock tools and Vina is used for obtaining the binding affinities with the HIV-1 protease for the analysis of the QSAR.

Materials & Methods:

Designing the analogs: Advanced Chemistry Development labs (ACD Labs)/Chemsketch is used to generate the 2D structures of all the 23 analogs of the GRL-10413. Chemsketch is the software that helps in modifying or creating the 2D structures of chemical compounds like small molecules to generate the log P, log D, pKa values etc [9]. Chemsketch has an option of

generating the molecular weight along with the analysis of the 3D view for each analog.

Preparation of the receptor: The 3D-structure of the PDB-ID: 5KAO was downloaded from the protein data bank (https://www.rcsb.org) and was analyzed using the PyMOL software. The protein-drug complex along with their interactions was saved for further analysis of the binding affinities analysis/the QSAR. The 5KAO is the 3D-structure of HIV-1 protease and the parent drug (GRL-10413) [7, 13]. The chain-A, chain-B and the 100th water molecule together are saved as a protein molecule that served as the docking receptor. The G43 chain-C is saved as an individual drug molecule. The parent drug molecule's chlorine atom has an alternate configuration at the P1 position which is modified by removing it. These structures of the protein and parent drug molecules were saved separately for the further conversion of the PDB files to PDBQT file formats for the *AutoDock Vina* to run and generate the binding affinities. The analogs of the GRL-10413 were generated using the PyMOL. Thus generated 23 analoges along with the output files created by the *Autodock Vina* were also evaluated against the receptor/protein molecule for finalizing the best binding pose of the analog.

Preparation of docking grid: AutoDock tools is maintained and developed by the Scripps Research Institute, Center for Computational Structure Biology [11]. *AutoDock tools* are the computational tools that help in molecular docking and analyzing the interactions of macromolecules and small molecules for identifying the conformations and best binding affinities [10]. The macromolecule is taken to set the grid (a virtual box of dimensions that is set to cover the active site of the macromolecule ((receptor), the area of the small molecule binding site) and the obtained points of the dimensions and the center is used to create a config text file that is given to the *Autodock Vina* to read the file for extracting the binding affinities of different poses of each analog of the small molecule (ligand) [12]. The version 1.5.7 of *Autodock tools* is used to set the protein structure molecule from the 5KAO is used to adjust the grid to 24, 20 and 20 for X-dimension, Y-dimension and Z-dimension respectively. The 3-dimensional center of the grid is taken as 4.994, -0.292 and -16.411 (x-center, y-center and z-center respectively) (Suppl Figure 1), which is saved in the file format of PDBQT. The rotatable and non-rotatable bonds of each of the analogs are adjusted and each analog is saved as individual files in the PDBQT file formats. The files are saved in the PDBQT formats for the *AutoDock Vina* to recognise and read the file for obtaining the binding affinities. *AutoDock Vina* produces the table of possible binding affinities by changing its modes of each analog (the binding pose of the drug molecule varies). The highest binding affinity is considered in each case. Analog MA01020 has the highest binding affinity of -9.3 KCal/mol while the parent drug (GRL-10413) has the binding affinity of -8.9 KCal/mol.

Molecular docking: The parent drug (GRL10413) is taken to obtain 23 analogs to evaluate if there are any possible analogue(s) with better characteristics than the parent drug molecule. All the 23 analogues that are listed in this study are the ones with deletion of possible chemical groups/classes (Suppl. Table 1). The parent drug (ligand) (GRL10413) is saved into a PDBQT format for evaluating the Binding affinities with the *Autodock Vina* by modifying the bond between the fourth carbon and the first nitrogen [S2 moiety: C4 & N1] and the bond between eighteenth carbon and sulphur [S2' moiety: C18 & S] are set to non-rotatable bonds which are represented in magenta color (Suppl. Figure 2) using *Autodock tools* software. After setting up the grid box at the active site of HIV-1 protease from the PDB ID: 5KAO structure (Suppl. Figure 1, refer to materials and methods: section- Autodock tools and Autodock Vina), the protein molecule (receptor: HIV-1

protease) is also saved in the PDBQT format. The same procedure of setting up the non-rotatable bonds is followed for all the 23 analogs. Each of the PDBQT files of the analogs are used to compile in the configuration files that helps *Autodock Vina* to read the files and compile which calculates the binding affinities. The command " ***--config conf.txt --log log.txt*** " helps the *Autodock Vina* to read the configuration file and write the log file (which is also a text file that gets automatically created after *Vina* finishes its run). *Autodock Vina* creates a possible number of poses of each analog, based on the evaluation of each binding pose with respect to the HIV-1 protease (receptor/protein); the best one with the best binding pose is considered for further calculations of log-p values. The output files of each analog with the possible binding poses were evaluated using the PyMOL software and all the analogs are found with the better in the first binding pose. The molecular weights and the Log-P values of all the 23 analogs are calculated using the Chemsketcher software.

Results and Discussion:

In silico QSAR of the analogs: As shown in Figure 3, the minimum binding affinity is -7.7 kcal/mol. in analog MA01008 and the analog MA01020 has the maximum binding affinity of -9.3 kcal/mol. There are 12 analogs (Analog IDs: MA01008, MA01009, MA01010, MA01011, MA01007, MA01013, MA01018, MA01023, MA01012, MA01017, MA01019, MA01002) with the binding affinities below the parent drug (-8.9 KCal/mol) while there are 11 analogs (Analog IDs: MA01005, MA01015, MA01016, MA01001, MA01022, MA01006, MA01014, MA01021, MA01003, MA01004, MA01020) with relatively higher binding affinity values.

Hydrogen Bonds profiles of the analogs: The hydrogen bonds of all the 11 analogs with high binding affinities are analogized with the parent drug (GRL10413) present in the panel L (see Figure 4). The parent drug binds to HIV-1 protease with 11 hydrogen bonds. The evaluation states that 4 analogs (Panels A, F, I & J; see Figure 4) among 11 analogs have the same number of hydrogen bonds as the parent drug molecule.

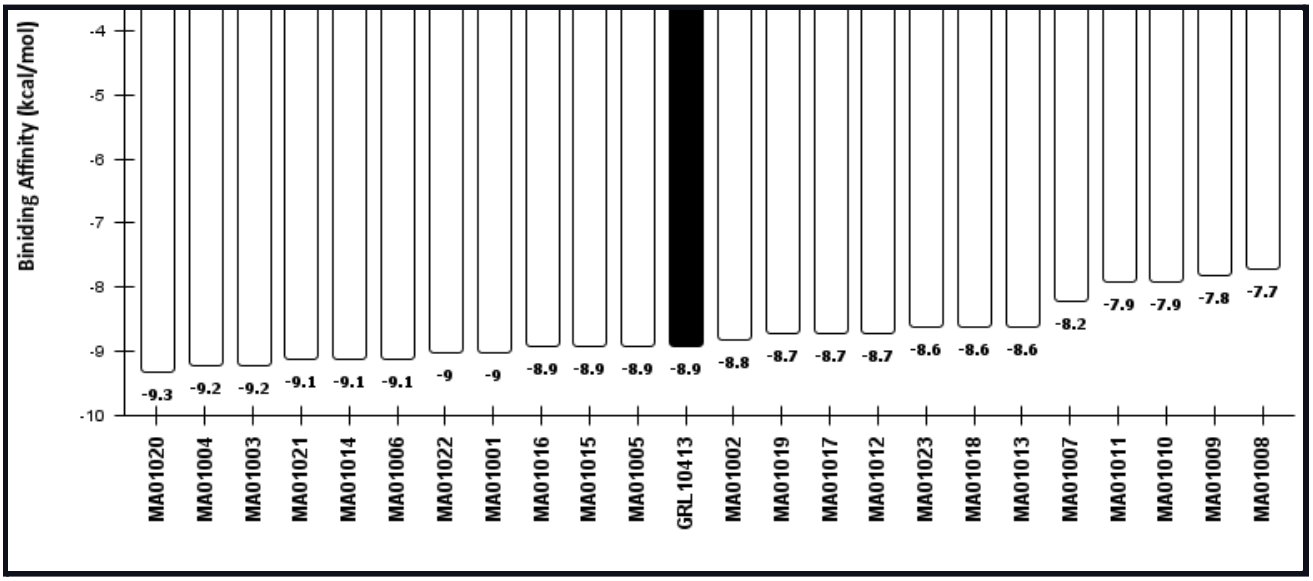

Figure 3. Binding affinities (kcal/mol.) of parent and analogs.

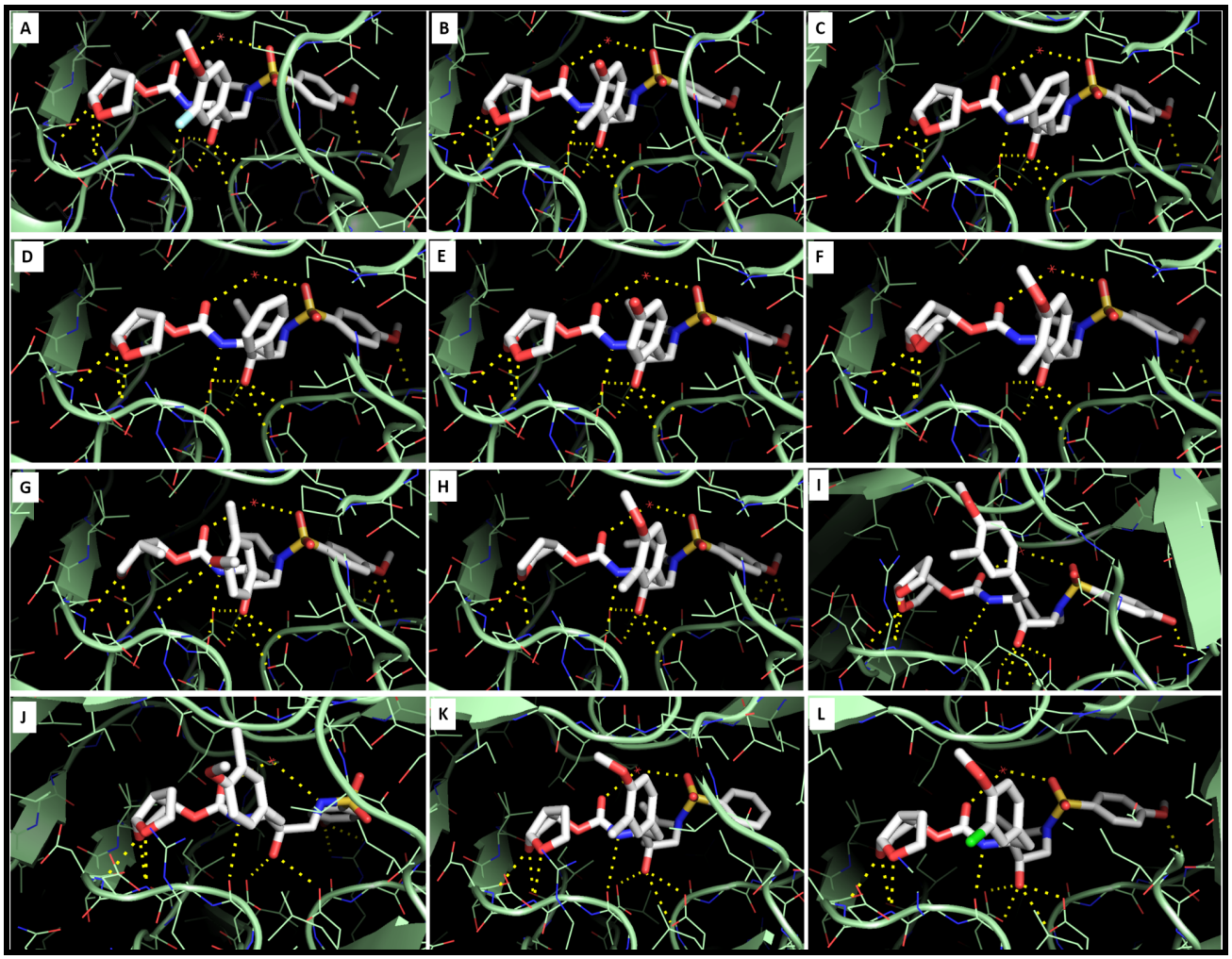

Figure 4. Hydrogen bonding profiles of the analogs.

There are 4 analogs (Panels B, C, D & E; see Figure 4) with higher number of hydrogen bonds. There are 3 analogs (Panels G, H & K) with lower number of hydrogen bonds in comparison with the parent drug (Panel L) (Supp. Table 2). Out of the 11 analogs; 5 analogs (MA01001, MA01003, MA01004, MA01005 & MA01006) are of P1 moiety modifications, 3 analogs (MA01014, MA01015 & MA01016) are modified P2 moiety, 1 analog (MA01020) is a modified P1' moiety and 2 analogs (MA01021, MA01022) of modified P2' moiety.

Prediction of best analogs of GRL10413 using the circular plot: The Binding Affinities, hydrogen bonds and Log P values of all the 23 analogs of GRL10413 are plotted in a circular plot (Figure 1) to choose the best analogs. By viewing the correlations between all three parameters, one can make better conclusions. The circular plot contains three sectors of Binding Affinity, Hydrogen bonds and Log-P values. The sectors of hydrogen bonds and Log-P values have positive scale and hence the peaks are projecting towards the outer circle originating from the inner circle. The sector of Binding Affinity is a negative scale hence the peaks are projected inside i.e., towards the inner circle. The sector of Log-P values contains two colors: red (ranging from 2.5 to 5.0) and bright green (ranging from 5.1 to 7.0). The sector of hydrogen bonds consists of colors: Dark blue (for 10 H-bonds), light blue (for 11 H-bonds), yellow (for 12 H-bonds) & Dark green (for 13

H-bonds). The Binding affinity (kcal / mol) sector comprises of colors: Red (ranging from -7.5 KCal/mol to -7.9 KCal/mol), Orange (ranging from -8.0 KCal/mol to -8.5 KCal/mol), Dark pink (ranging from -8.6 KCal/mol to -9.0 KCal/mol) & Dark purple (ranging from -9.1 KCal/mol to -9.5 KCal/mol). The Inner circle consists of the correlations of the Binding Affinities of each analog towards the H-bonds and Log-P values. The correlations concluded that there are 6 analogs: MA01003, MA01004, MA01006, MA01014, MA01020 & MA01021; with high binding affinities (ranging from 9.1 KCal/mol to 9.5 KCal/mol). Analogs MA01003 (12 H-bonds), MA01006 (13 H-bonds), MA01020 (11 H-bonds) & MA01021 (11 H-bonds) consist of Log-P values below 5.0; while the other 2 analogs, MA01004 (12 H-bonds) & MA01014 (11 H-bonds) contains Log-P values above 5.1. Analogs with the Log-P value ranging from -0.4 to 5 are considered according to Lipinski's Rule of 5. This filters out 4 analogs (MA01003, MA01006, MA01020 & MA01021) as the best analogs out of all the 23 analogs.

References

1. https://www.hiv.gov/hiv-basics/overview/data-and-trends/global-statistics
2. https://www.hiv.gov/hiv-basics/overview/about-hiv-and-aids/what-are-hiv-and-aids
3. Cunningham AL, Donaghy H, Harman AN, Kim M, Turville SG (August 2010). "Manipulation of dendritic cell function by viruses". *Current Opinion in Microbiology*. **13** (4): 524–9.
4. Kräusslich HG, Ingraham RH, Skoog MT, Wimmer E, Pallai PV, Carter CA (February 1989). "Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides". *Proceedings of the National Academy of Sciences of the United States of America*. **86** (3): 807–11.
5. Kohl NE, Emini EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM, Sigal IS (July 1988). "Active human immunodeficiency virus protease is required for viral infectivity". *Proceedings of the National Academy of Sciences of the United States of America*. **85** (13): 4686–90.
6. Marmor M, Hertzmark K, Thomas SM, Halkitis PN, Vogler M. Resistance to HIV infection. *J Urban Health*. 2006;83(1):5-17.
7. ASM Journals Antimicrobial Agents and Chemotherapy, Vol. 60, No. 12 A Modified P1 Moiety Enhances In Vitro Antiviral Activity against Various Multidrug-Resistant HIV-1 Variants and In Vitro Central Nervous System Penetration Properties of a Novel Nonpeptidic Protease Inhibitor, GRL-10413.
8. https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/quantitative-structure-activity-relationship#:~:text=Quantitative%20structure%20activity%20relationship%20(QSAR,the%20substance%20will%20be%20modified.
9. https://www.technologynetworks.com/drug-discovery/product-news/new-freeware-offering-makes-acdlabs-lipophilicity-predictions-available-to-all-220675
10. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455-461. doi:10.1002/jcc.21334
11. https://ccsb.scripps.edu/ "The Center for Computational Structural Biology". *The Center for Computational Structural Biology*. 2020-05-15. Retrieved 2020-05-15.
12. http://www.bch.cuhk.edu.hk/kbwong/teaching/autodock/autodock.html
13. https://www.rcsb.org/structure/5KAO
14. Huang L, Chen C (July 2013). "Understanding HIV-1 protease autoprocessing for novel therapeutic development". *Future Medicinal Chemistry*. **5** (11): 1215–29.

15. Lv Z, Chu Y, Wang Y (April 2015). "HIV protease inhibitors: a review of molecular selectivity and toxicity". *HIV/AIDS: Research and Palliative Care*. **7**: 95–104.
16. Pettit SC, Everitt LE, Choudhury S, Dunn BM, Kaplan AH (August 2004). "Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism". *Journal of Virology*. **78** (16): 8477–85.
17. Brik A, Wong CH (January 2003). "HIV-1 protease: mechanism and drug discovery". *Organic & Biomolecular Chemistry*. **1** (1): 5–14.
18. Louis JM, Clore GM, Gronenborn AM (September 1999). "Autoprocessing of HIV-1 protease is tightly coupled to protein folding". *Nature Structural Biology*. **6** (9): 868–75.
19. Davies DR (1990). "The structure and function of the aspartic proteinases". *Annual Review of Biophysics and Biophysical Chemistry*. **19** (1): 189–215.
20. Liu H, Müller-Plathe F, van Gunsteren WF (August 1996). "A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease". *Journal of Molecular Biology*. **261** (3): 454–69.
21. Moore JP, Stevenson M (October 2000). "New targets for inhibitors of HIV-1 replication". *Nature Reviews. Molecular Cell Biology*. **1** (1): 40–9.
22. https://www.who.int/news-room/fact-sheets/detail/hiv-aids#:~:text=There%20were%20an%20estimated%2037.7,2.0%20million%5D%20people%20acquired%20HIV.

How to cite this article?
Aggunna and Yedidi (2022). *TCABSE-J*, Vol. 1, Issue 3:10-17. Epub: Apr2nd, 2022.

Acknowledgements: The authors thank T*yi*DE-Toronto, Canada for helping to write this manuscript.

Funding: The authors thank TCABS-E, Rajahmundry, India and T*yi*DE-Toronto, Canada for financial support.

Conflict of interest: This research article is an ongoing project currently at TCABS-E, Rajahmundry, India. The authors invite collaborations without any conflict of interest.