

Automated *In Silico* identification of drug candidates for coronavirus through a novel programmatic tool and extensive computational (MD, DFT) studies of select drug candidates

Ben Geoffrey A S^{*a}, Rafal Madaj^b, Akhil Sanker^c Mario Sergio Valdés Tresanco^d, Host Antony David ^e, Gitanjali Roy^f, Rinnu Sarah Saji^a, Beutline Malgija^e, Abdulbasit Haliru Yakubu^g

^a University of Madras, Chepauk, Chennai 600 005, India

^b Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Poland

^c SRM University, Tamil Nadu 603203, India

^d University of Medellin, Medellin, Colombia

^e Madras Christian College, Chennai 600 059, India

^f Indiana University School of Medicine, Indianapolis, USA

^g University of Maiduguri, Maiduguri, Borno state, Nigeria.

*Corresponding author email : bengeof@gmail.com

Abstract

The work is composed of python based programmatic tool that automates the dry lab drug discovery workflow for coronavirus. Firstly, the python program is written to automate the process of data mining PubChem database to collect data required to perform a machine learning based AutoQSAR algorithm through which drug leads for coronavirus are generated. The data acquisition from PubChem was carried out through python web scrapping techniques. The workflow of the machine learning based AutoQSAR involves feature learning and descriptor selection, QSAR modelling, validation and prediction. The drug leads generated by the program are required to satisfy the Lipinski's drug likeness criteria as compounds that satisfy Lipinski's criteria are likely to be an orally active drug in humans. Drug leads generated by the program are fed as programmatic inputs to an *In Silico* modelling package to computer model the interaction of the compounds generated as drug leads and the coronaviral drug target identified with their PDB ID : 6Y84. The results are stored in the working folder of the user. The program also generates protein-ligand interaction profiling and stores the visualized images in the working folder of the user. Select drug leads were further studied extensively using Molecular Dynamics Simulations and best binders and their reactive profiles were analysed using Molecular Dynamics and Density Functional Theory calculations. Thus our programmatic tool ushers in a new age of automatic ease in drug identification for coronavirus.

1. Introduction

PubChem is a data repository of chemical compounds, their properties and biological activities [1] which can be programmatically accessed through web API packages such as PUB-REST and python web scrapping techniques [2,3]. Quantitative Structure-Activity Relationship(QSAR) studies are statistical based studies through which drug leads are generated which provide cost cutting advantages in testing and drug discovery for the pharmaceutical industry [4-7]. However the data set required to perform a QSAR study is curated by researchers before performing the statistically study. Our programmatic tool automates this process of data acquisition required to perform a QSAR study to generate drug leads for coronavirus through programmatic access of PubChem database and python web scrapping techniques [8,9]. The workflow of the QSAR study was also automated through a machine learning based AutoQSAR algorithm. The workflow of a machine learning based AutoQSAR algorithm involves feature learning and descriptor selection, QSAR modelling, validation and prediction [10-12]. The drug leads generated by the program are required to satisfy the Lipinski's drug likeness criteria as compounds that satisfy Lipinski's criteria are likely to be an orally active drug in humans [13]. Drug leads generated by the program are fed as programmatic inputs to an In Silico modelling package to computer model the interaction of the compounds generated as drug leads and the coronaviral drug target identified with PDB ID : 6Y84. The coronaviral drug targets identified from the literature was the main protease of SARS-CoV 2 (PDB ID : 6Y84) [14,15]. The results of the *In Silico* modelling are stored in the working folder of the user. The program also generates protein-ligand interaction profiling and stores the visualized images in the working folder of the user. Thus our programmatic tool ushers in the new age automatic ease in drug identification for coronavirus through a fully automated QSAR and an automated *In Silico* modelling of the drug leads generated by the AutoQSAR algorithm.

Our work is distinguished from previous attempts of virtual screening of large ligand libraries in way that we employ programmatic techniques as compared to other works that do not [16]. However as compared to recent data drive machine learning based approaches to drug discovery [17,18] that use pre-downloaded data sets we deploy a real time data mining which makes a case for a dynamic approach to drug lead generation and the results of the program are reflective of PubChem data library at the instant the program is run and thus approaches drug discovery from a dynamic approach in an age Big Data and constantly growing data libraries. We also add to the existing richness of the novelty of methods [18-20]

in data driven drug discovery in the following way. Our programmatic tool couples the drug leads generated by the AutoQSAR algorithm as programmatic inputs to an *In Silico* modelling package and programmatically profiles the protein-ligand interaction and stores the results in the working folder of the user. While adding to the existing richness in data driven machine learning based drug discovery methods, our work also adds new scientific findings to existing literature as the coronaviral drug target chosen for the study have been approached by data driven machine learning based methods.

2. Methods and Techniques

The workflow of the programmatic tool implementing programmatic data mining, AutoQSAR and automated *In Silico* modelling for identification of drugs against coronavirus is shown in Fig.1. The first process involved in programmatic workflow is the data mining of PubChem database to automate the process of data acquisition to implement a machine learning based AutoQSAR algorithm that automates the process of drug lead generation for coronavirus. The programmatic access to PubChem is accomplished through python commands [8, 9]. The program fetched activity data for reported for compounds on PubChem. At the instant we ran the program, data availability of PubChem provided us with activity data of 111 compounds reported to be active against coronavirus. The molecular descriptors for the active compounds was also programmatically fetched. The molecular descriptors include Molecular Weight, Heavy Atom Count, XLOGP, Complexity, Hydrogen bond Acceptor Count, Monoisotopic Mass, Rotatable Bond Count and Topological Surface Area. The next process in the workflow involves implementing the machine learning based AutoQSAR algorithm for drug lead generation. The drug leads are generated by the AutoQSAR algorithm through the workflow that involves feature learning, descriptor selection, QSAR modelling, validation and prediction [10-12]. QSAR models are usually linear or non-linear statistical correlation between the activity and molecular descriptors. While the total number of descriptors is 8, the program builds a QSAR model with every possible combination of descriptors by generating all possible combinations of descriptors where $n = 8$ and $r = 2, 3, 4, 5, 6, 7$ and nC_r in such a case gives a total of 256 combinations of descriptor selection for the QSAR model. The program builds a linear and non-linear regression based QSAR model with all 256 possible combinations of descriptors and selects the QSAR with highest R^2 value or R^2 value closest to 1. The highest R^2 value of $R^2 = 0.88$ was achieved for a non-linear regression based QSAR model which involved descriptors Molecular Weight, XlogP and Monoisotopic Mass.

Based on the validated QSAR model PubChem compound library is screened for drug lead generation based on the validated QSAR model. From the PubChem library the program identified 1355 compounds which are structurally associative to the 111 compounds that are active against coronavirus. The molecular descriptors of the 1355 compounds was collected by the program and the validated QSAR model was used to predict the activity against coronavirus for the 1355 compounds. The program prints out the top 50 compounds and identifies them as drug leads. The drug leads generated by the program were also required to satisfy the Lipinski's drug likeness criteria [13].

Running the program requires no more programming knowledge than running the python executable file in python 3 environment in Linux OS along with some python dependency packages installed such as:

pandas

biopandas

numpy

matplotlib

scikit-learn

seaborn

selenium (along with selenium's driver for firefox browser)

Other additional dependencies for automated *In Silico* modelling

openbabel 2.4.1

mgltools 1.5.4

autodock-vina 1.1.2-4

The program is hosted, maintained and supported at the GitHub repository link given below

<https://github.com/bengeof/Programmatic-tool-to-automate-the-drug-discovery-workflow-for-coronavirus>

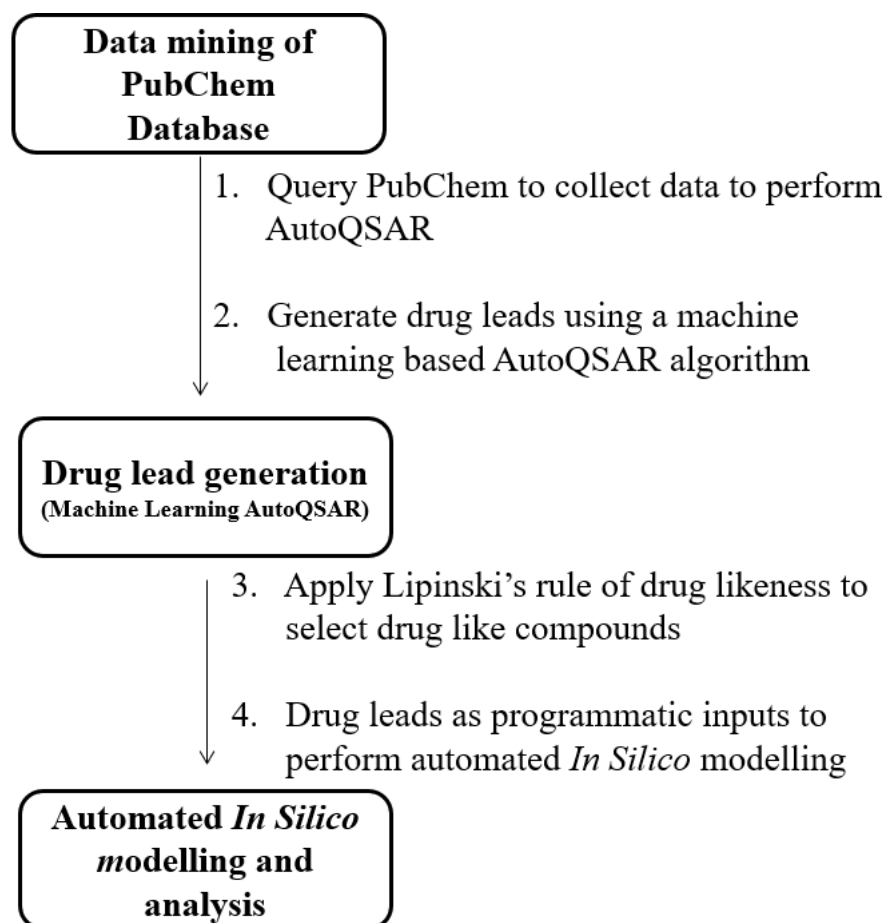


Fig. 1 – Workflow of the programmatic tool to automate the drug discovery for coronavirus

The running of the program requires a stable internet connection and the run time of the program is expected to be a few hours however it is expected to vary based on CPU and internet speed. The program prints out the PubChem CIDs of Top 50 compounds identified as drug leads for coronavirus drugs

The crystal structure of the SARS-CoV 2 main protease was downloaded from the RCSB-PDB database [21] with PDB ID : 6Y84. The ligand and the protein files were prepared for AutoDock process using AutoDockTools(ADT) scripts and the protein drug target files are to be kept in the working folder of the user and can be downloaded from the GitHub repository. The structure of the drug lead compounds generated by the program were programmatically downloaded from PubChem and programmatically prepared for molecular docking using AutoDockTools ligand preparation scripts. The virtual screening using AutoDock Vina

[22,23] was initiated programmatically through the program and the interaction between the drug targets proteins and lead drug compounds also automatically profiled and the visualized image of the protein-ligand interaction is saved in the working folder of the user by the program [24,25].

Molecular Dynamics study of select drug candidates

Select protein-ligand complexes associated with structurally diverse ligands having lowest binding energy obtained from AutoDock-Vina screening were selected for extensive molecular dynamics simulation studies and subsequently for molecular mechanics generalized-born surface area (MMGBSA) with pairwise decomposition calculations. The ligands were cut out from the complexes and optimized using AmberTools18 from Amber18 suite [26], followed by partial charges calculation according to AM1-BCC level of theory [27]. The topology and input coordinates were created using tleap. The protein was described using ff14SB force field, ligand using GAFF and water molecules by TIP3P. The system was placed inside truncated octahedral box with 16 Å boundary, solvated and neutralized. Minimization encompassed 20000 steps, 10000 steepest descent algorithm and the rest using conjugate gradients. The heating encompassed 50 to 300 K increase in the temperature and lasted 100 ps. Restraints on protein backbone of 4 kcal/mol*Å² were applied. The production run were performed under NPT ensemble for 20 ns, out of which 10 ns were truncated for equilibration purposes. The simulation runs were put under periodic boundary conditions, non-bonded interactions were evaluated with the Particle Mesh Ewald method with cut-off of 9 Å. Langevin thermostat and Monte Carlo Barostat were used for temperature and pressure maintenance [28]. All described simulations were done using pmemd module of AMBER 18, utilizing CPU for minimization and GPU for heating, equilibration and production. For each ligand, the simulations were repeated ten times. Subsequently, the trajectories were merged and clustered. The clusters encompassing ligand within the binding cavity were selected for MMGBSA calculations and then pairwise decomposition. The calculation were performed on the Intel® Core™ i9-9900KF CPU @ 3.60GHz x 16 with 32GB @ 2666MHz with GeForce RTX 2070 SUPER/PCIe/SSE2 on the Ubuntu 20.04 Focal Fossa.

DFT study of electrostatic topology and reactive sites of select compounds

Density Functional Theory is used as approximation techniques to solve the many-body Schrodinger equation system to study the electronic properties of atoms, molecules and compounds [29]. The energy functional which is composed of potential, kinetic and

exchange-correlation parts is minimized to produce the configuration of the system corresponding to minimum energy. The DFT problem was solved using the ORCA package at a level of theory which involved using the B3LYP exchange correlation function [30]. The multiwfn program was used to visualize the results and Electrostatic Potential(ESP), Electron Locationzation Potential (ELF), Localized Orbital Locator (LOL) plots were generated and used to study the electrostatic topology and reactive sites of select compounds identified as strong binders to coronaviral drug target [31-34].

3. Results and Discussion

The python program was run in Python 3 environment with the dependency packages mentioned in the methodology section. The program prints out the PubChem CIDs of the top 60 compounds identified as coronavirus drug leads by the program. This is done by the program through automated programmatic data mining of PubChem database to collect data required to perform a machine learning based AutoQSAR algorithm on the dataset to generate the drug leads for coronavirus. The drug leads generated by the program for the coronavirus are useful to screen PubChem database which is over a billion compounds and the generated drug leads are useful to further pursue *In Silico*, *In Vitro* and *In Vivo* testing and is expected to save computational and experimental testing costs for the pharmaceutical industry.

The drug leads generated by the program were required to satisfy the Lipinski's criteria of drug likeness as compounds that satisfy the Lipinski's criteria are likely to be orally active drug in humans. The structure of the compounds identified as drug leads were programmatically downloaded from PubChem by the program and they were fed as programmatic ligand input files after ligand preparation via ADT scripts to a *In Silico* modelling package used widely known as AutoDock-Vina. Therefore the study of the interaction of the drug lead compounds and the SARS-CoV 2 drug target (PDB ID : 6Y84) was automated through the programmatic inputs given to AutoDock-Vina in the program. The results of the virtual screening for the top 50 drug lead compounds are given in Table 1 and select images of interaction of the compounds generated as drug leads by the program and the coronaviral drug target, main protease (PDB ID : 6Y84) is shown in Fig 2, 3, 4 & 5.

Table 1 - Automated Virtual Screening for identifying drugs against coronavirus

PubChem CID	Drug Target	Binding Affinity (Kcal/mol)
137777	Main Protease of SARS-CoV 2 (PDB ID : 6Y84)	-5.5
901295		-6.3
1580642		-6.5
2196453		-6.0
3542734		-5.3
4363256		-5.9
5244119		-5.6
16075059		-6.1
16202740		-6.0
16203681		-5.0
16203682		-5.7
16203797		-7.0
16204319		-5.8
29559485		-6.0
44589253		-6.3
47391569		-5.8
50876798		-5.7
54178808		-5.1
61356255		-6.6
61356261		-6.3
61356831		-6.1
61359893		-5.7
61360057		-6.2
61727697		-5.9

61727698		-6.1
61743020		-5.9
62024579		-6.4
62024757		-6.4
64204896		-5.6
64205053		-5.0
64207301		-5.4
64207509		-4.6
65101793		-5.1
65102076		-5.2
65102365		-5.3
65237247		-5.0
67380717		-6.1
68862352		-6.0
70485909		-6.1
75268360		-5.7
80452260		-5.3
90975560		-5.6
91875621		-6.0
91879629		-6.2
104453473		-5.3
130552686		-5.1
130643811		-5.2
140338441		-5.9
142747432		-7.3
142747435		-7.1

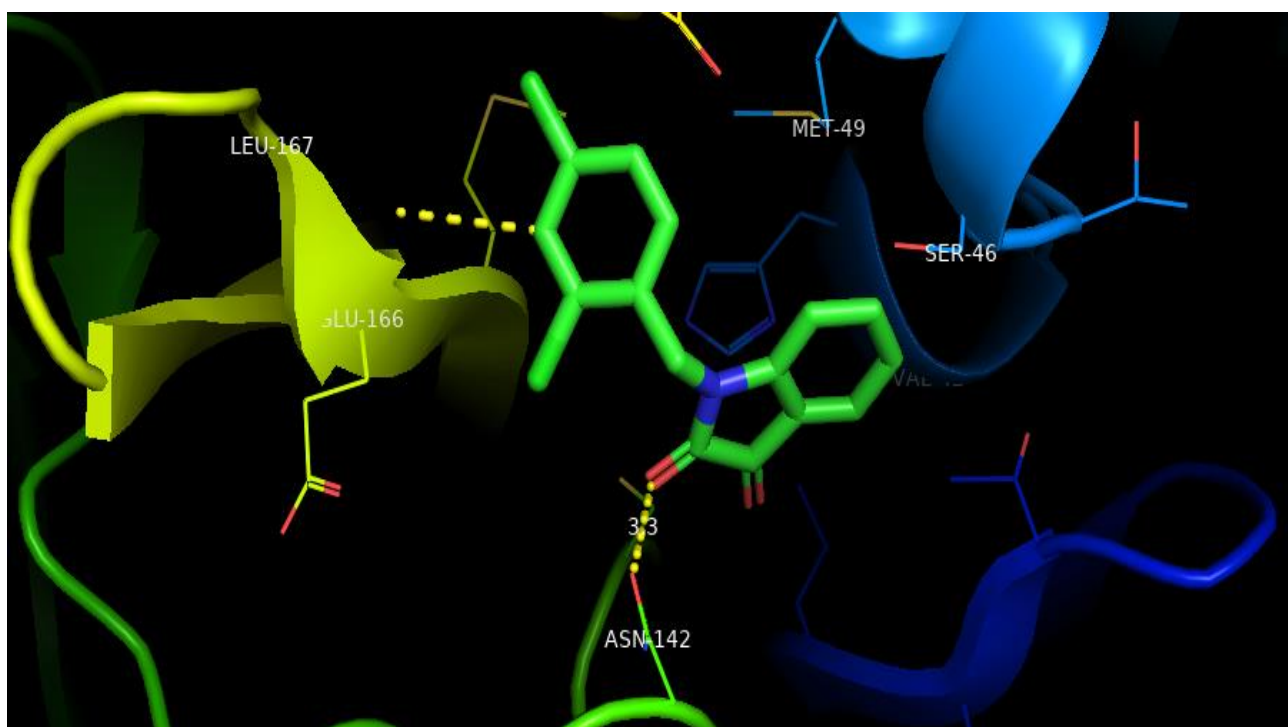


Fig 2 – Interaction of drug target and compound with PubChem CID 1580642

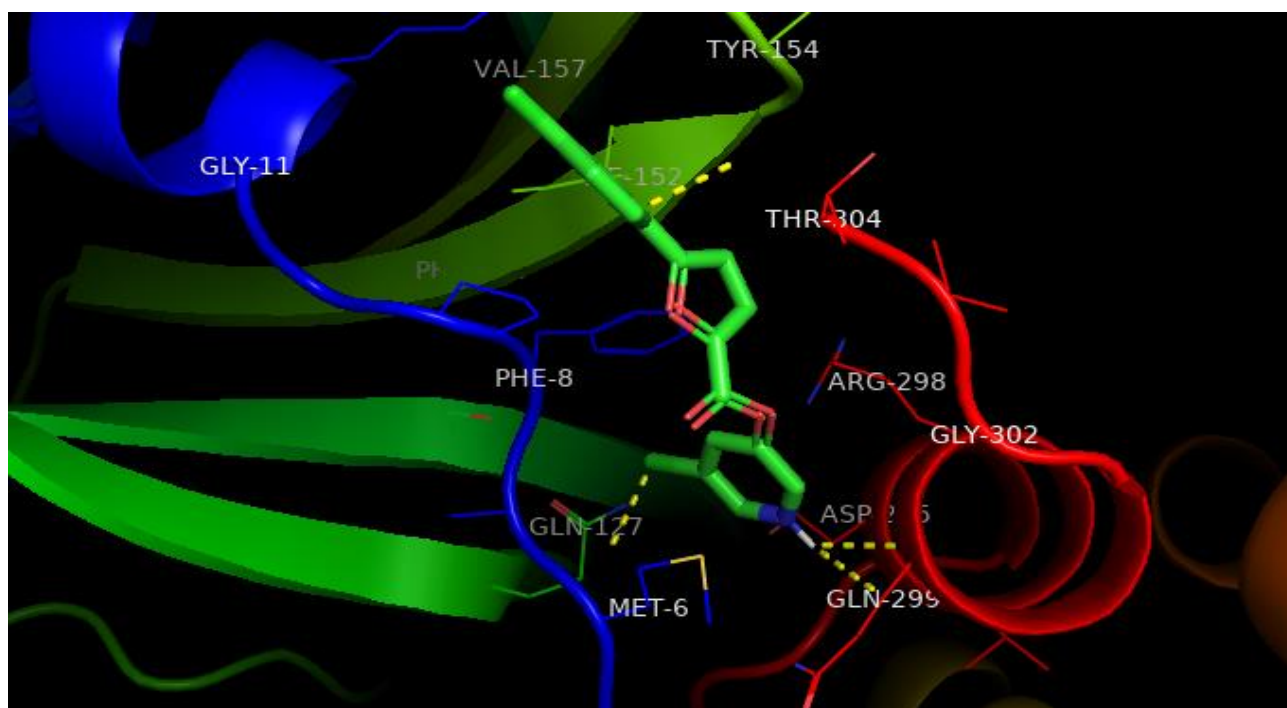


Fig 3 - Interaction of drug target and compound with PubChem CID 16203797

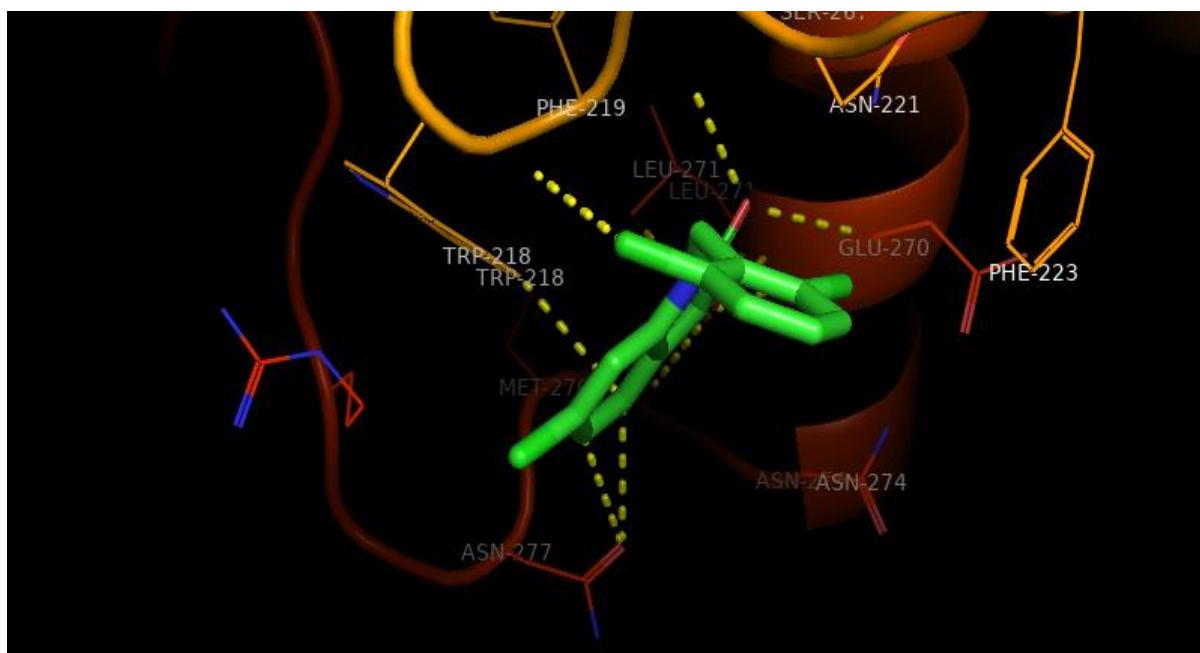


Fig 4 - Interaction of drug target and compound with PubChem CID 61356255

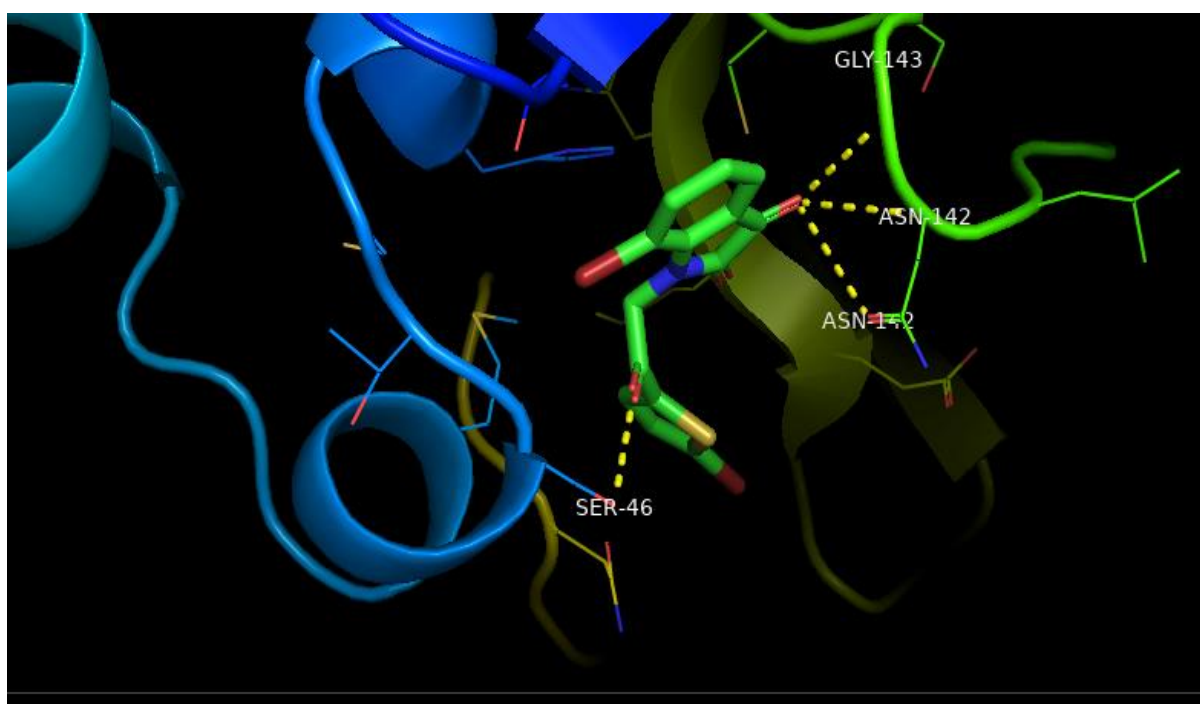


Fig 5 - Interaction of drug target and compound with PubChem CID 62024757

Select drug candidates were identified from the fast and automated virtual screening and were further studied extensively using computationally more expensive methods such as Molecular Dynamics Simulations and Density Functional Theory calculations. The Binding Free Energy change was calculated using molecular dynamics simulations and the interacting residues for the select drug candidate compounds are given in Table 2

Table 2 – MD

Method	Molecular docking	Molecular dynamics			
Feature	Lowest free energy of binding ΔG [kcal/mol]	% of time spent inside binding cavity	Total relative binding free energy $\Delta\Delta G$ (Generalized-Born) [kcal/mol]	Most important resids (more than -1 kcal/mol contribution)	Total relative binding free energy $\Delta\Delta G$ (Poisson-Boltzmann) [kcal/mol]
1580642	-6.3	83.00%	-18.95 ± 3.9	TRP218, PHE219, LEU220, ASN221, ARG222, GLU270, LEU271, ASN274	-3.64 ± 2.99
16203797	-7	92.00%	-16.6 ± -3.7	ASN151, ILE152, ASP153, TYR154, PHE294, VAL297, ARG298	0.33 ± 5.2
61356255	-6.6	88.00%	-16.9 ± -3.6	TRP218, PHE219, ASN221, ARG222, GLU270, LEU271, ASN274, ARG279	-2.33 ± 3.33

Electrostatic topology and reactive site study of the select drug candidates

The geometry of the select compounds was minimized at B3LYP level theory of DFT calculations carried out on ORCA. The minimized geometry of the select compounds is given in Fig. 6, 10, 14. The electrophilic and nucleophilic reactive sites of compounds can be identified using the Molecular Electrostatic Potential (MEP) map of the compound. The reactive site analysis through MEP is complemented by the study of ELF (electron localization function) and LOL (localized orbital locator) where the probability of finding an electron pair in regions of molecular space is studied using ELF and LOL studies. The white regions of the MEP Plot indicate electrophilic region and red regions indicate nucleophilic region. The value of ELF ranges from 0.0 to 1.0, where relatively large values in the interval 0.5 and 1.0 indicate regions containing bonding and nonbonding localized electrons, whereas smaller values (less than 0.5) describe regions where electron are expected to be delocalized. High values of ELF are seen around hydrogen atoms while around carbon, chlorine and nitrogen atoms, values of ELF are low. The LOL attains large values (greater than 0.5) in regions where the electron density is dominated by electron localization. A high localization of electrons due to the presence of a covalent bond (a lone pair of electrons) or a nuclear shell in that region is indicated by large value in that region [31-34]. Carbon atoms appear in the blue region and white regions around hydrogen atoms indicates that electron density exceeds the upper limit of colour scale (0.80). The blue regions around few carbon atoms show the delocalized electron cloud around it. Red colour around hydrogen and nitrogen atoms show the covalent regions. The MEP, ELF and LOL plots of the three select drug candidates are shown in Fig. 7,8,9,11,12,13,15,16 & 17 respectively

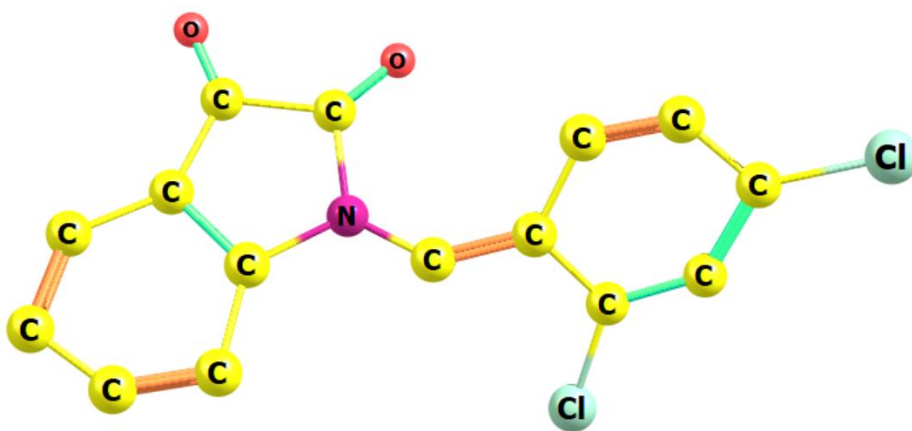


Fig. 6 – Optimized geometry of compound with PubChem CID 1580642

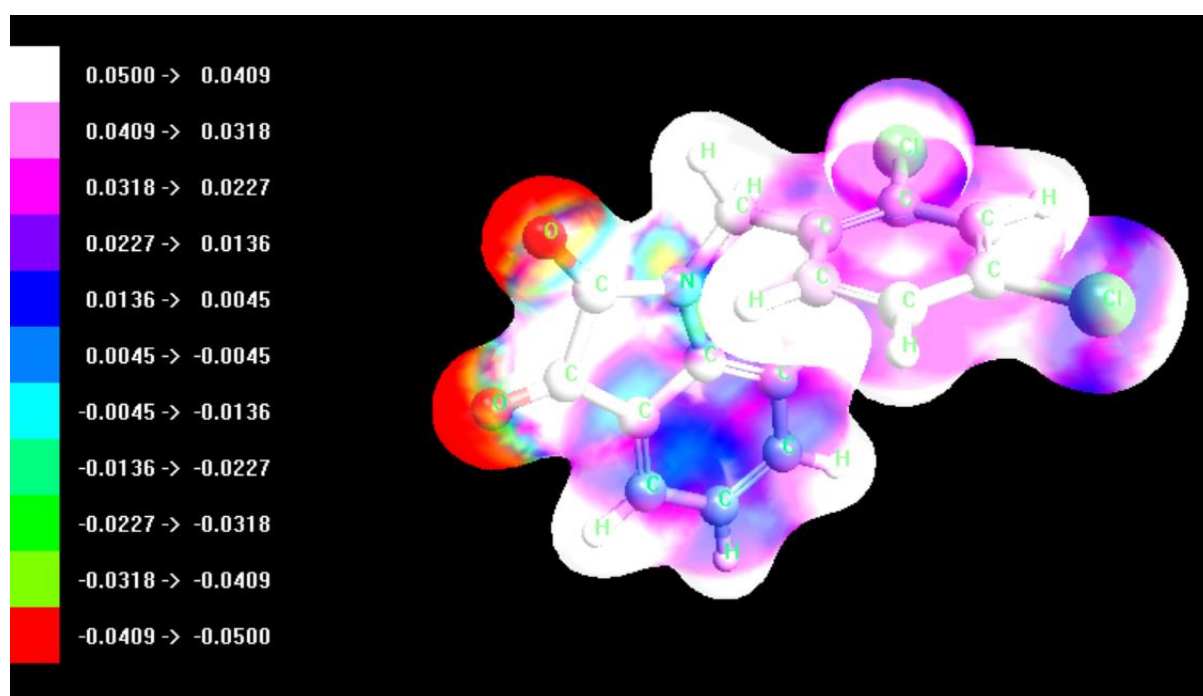


Fig.7 – MEP plot of compound with PubChem CID 1580642

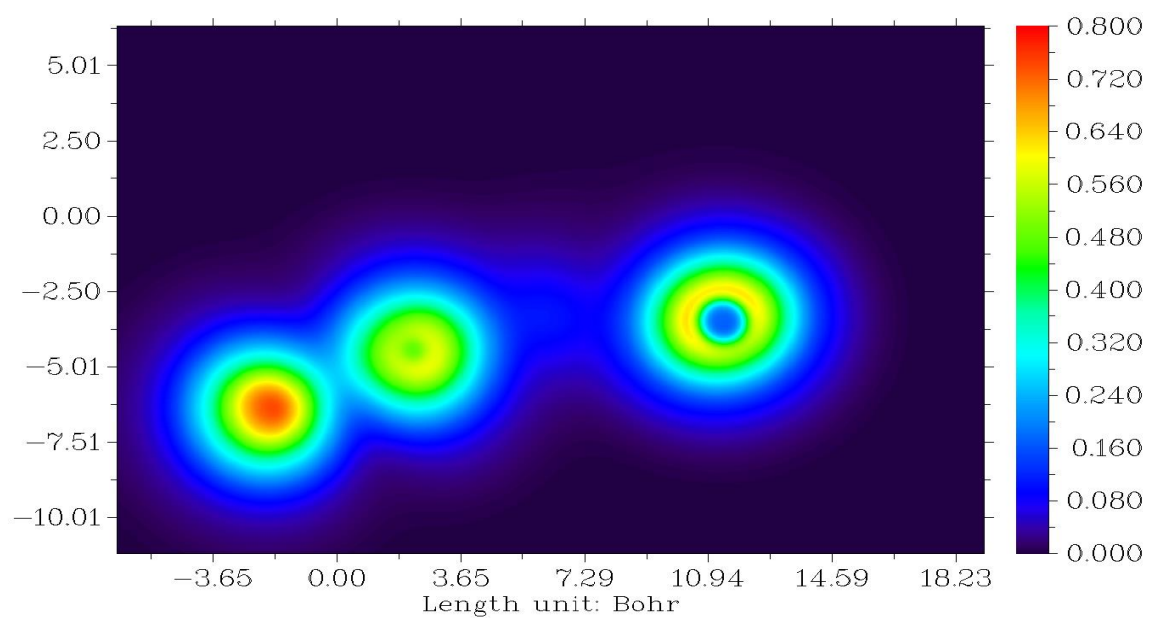


Fig.8 – LOL plot of compound with PubChem CID 1580642

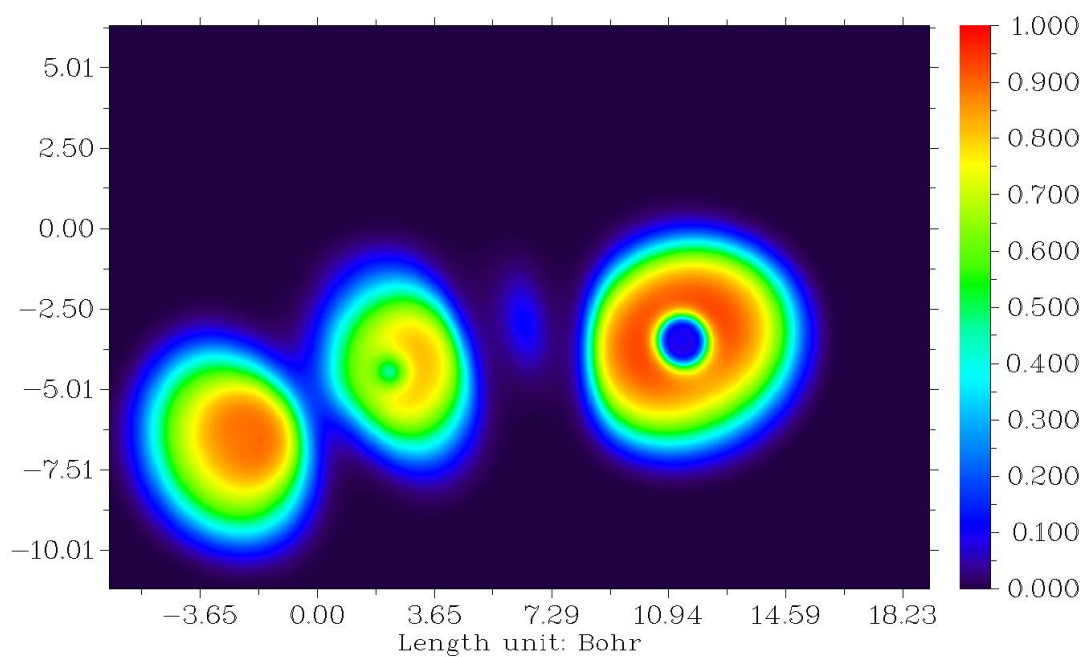


Fig. 9 – ELF plot of compound with PubChem CID 1580642

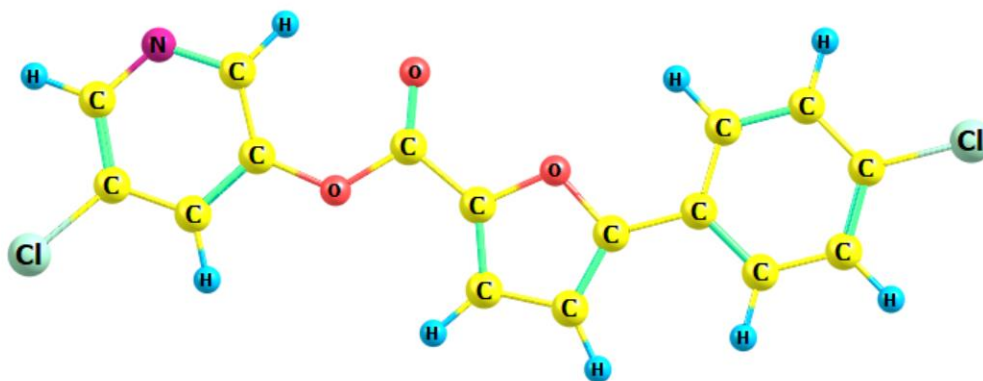


Fig. 10 - Optimized geometry of compound with PubChem CID 16203797

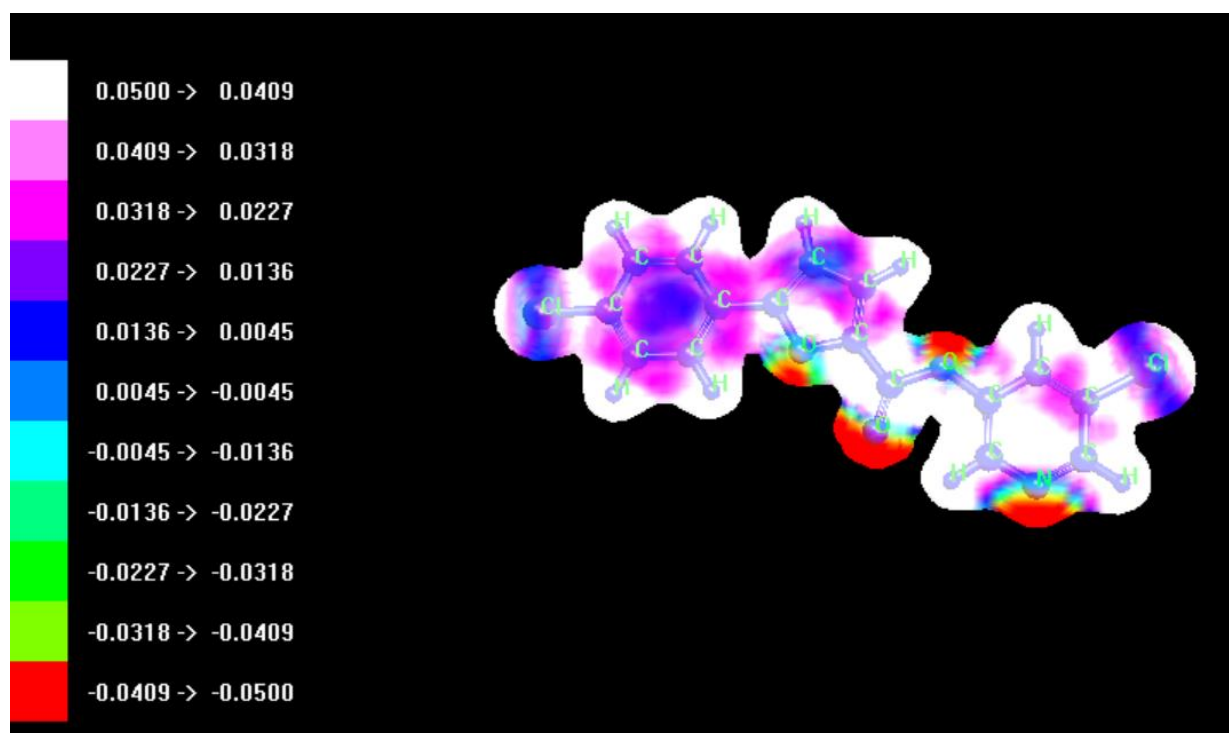


Fig.11 – MEP plot of compound with PubChem CID 16203797

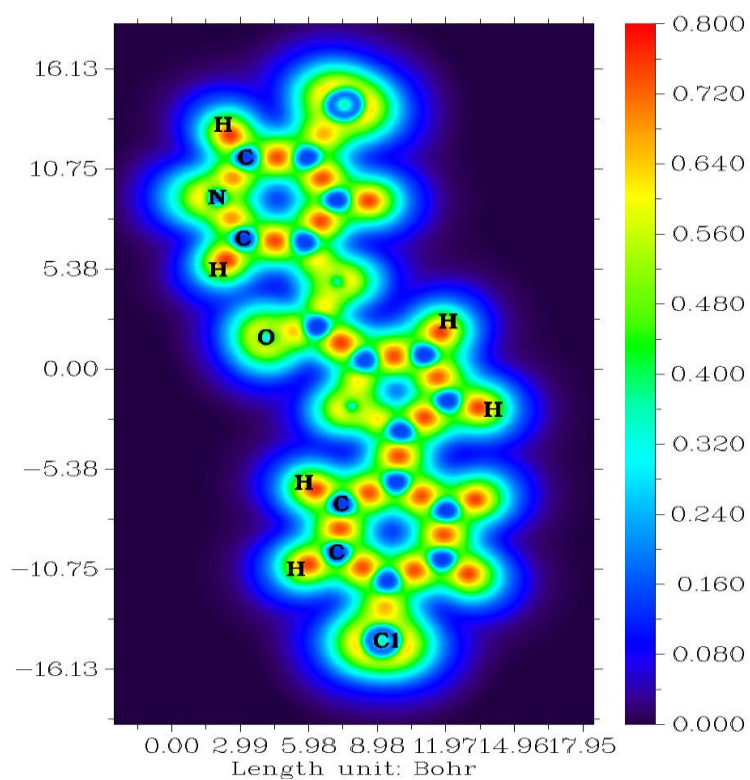


Fig.12 – LOL plot of compound with PubChem CID 16203797

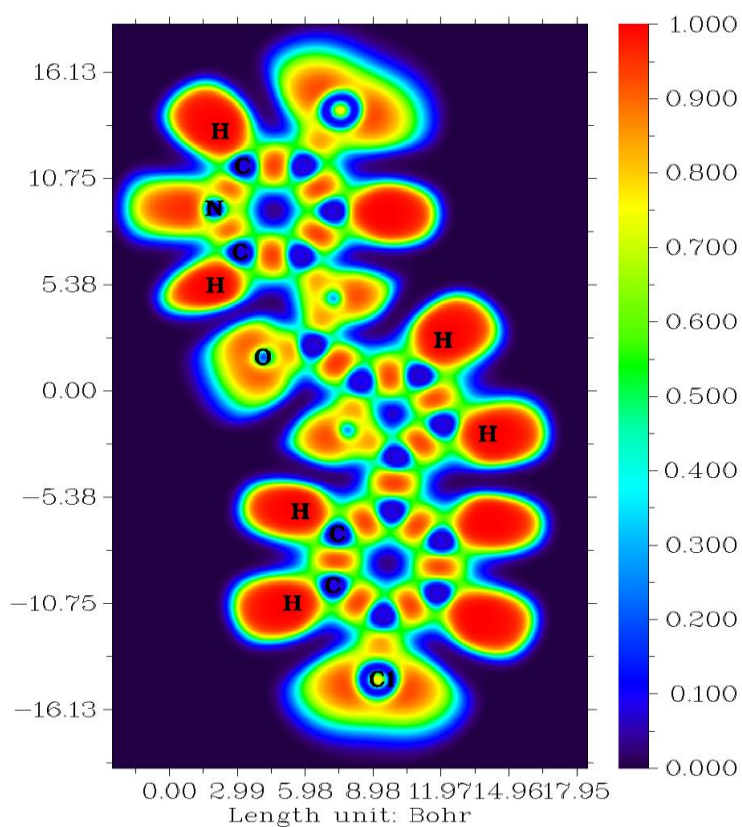


Fig. 13 – ELF plot of compound with PubChem CID 16203797

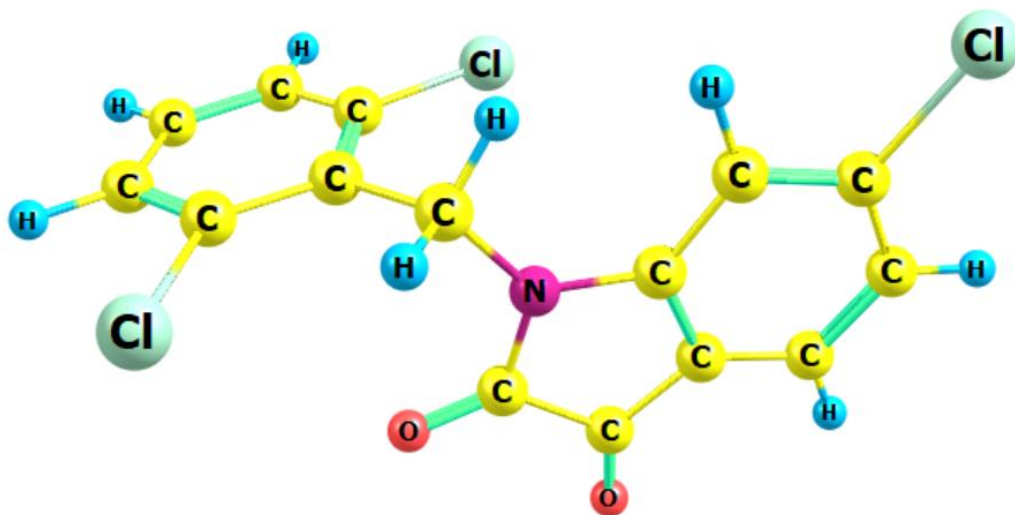


Fig. 14 - Optimized geometry of compound with PubChem CID 16203797

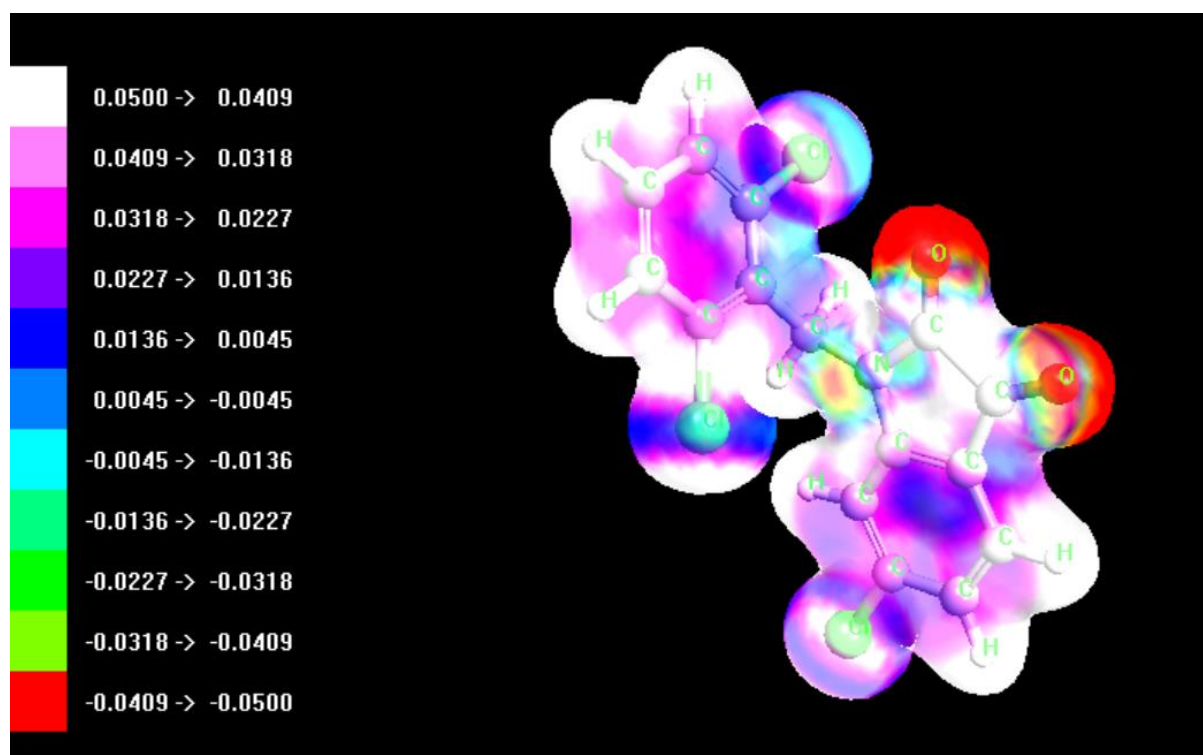


Fig.15 – MEP plot of compound with PubChem CID 16203797

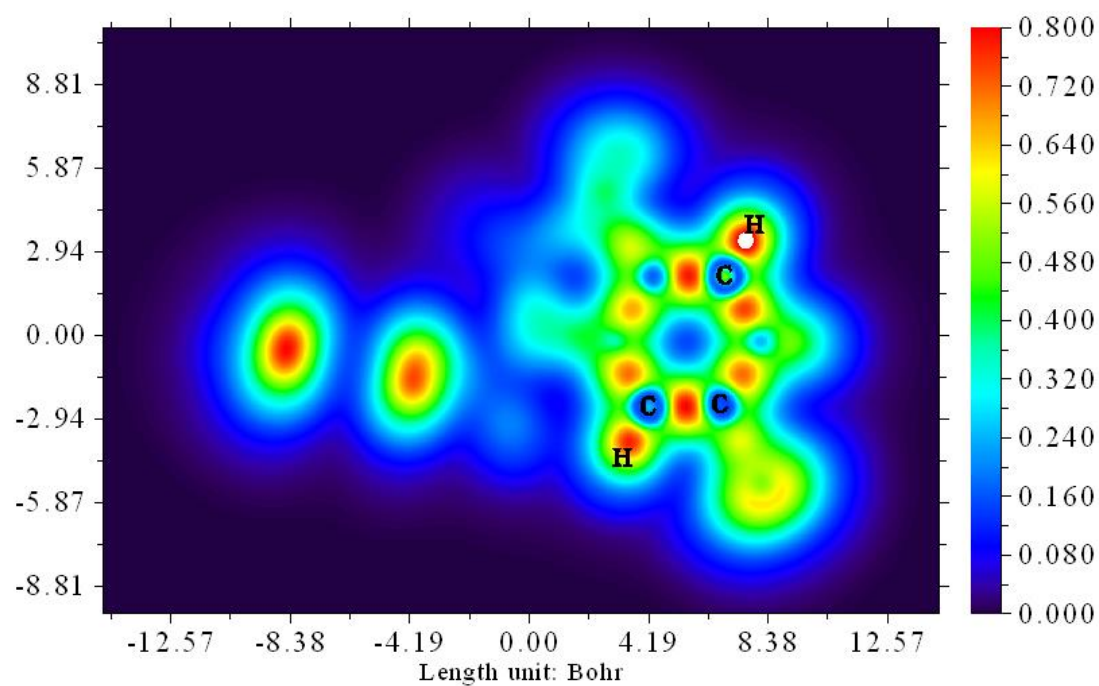


Fig.16 – LOL plot of compound with PubChem CID 16203797

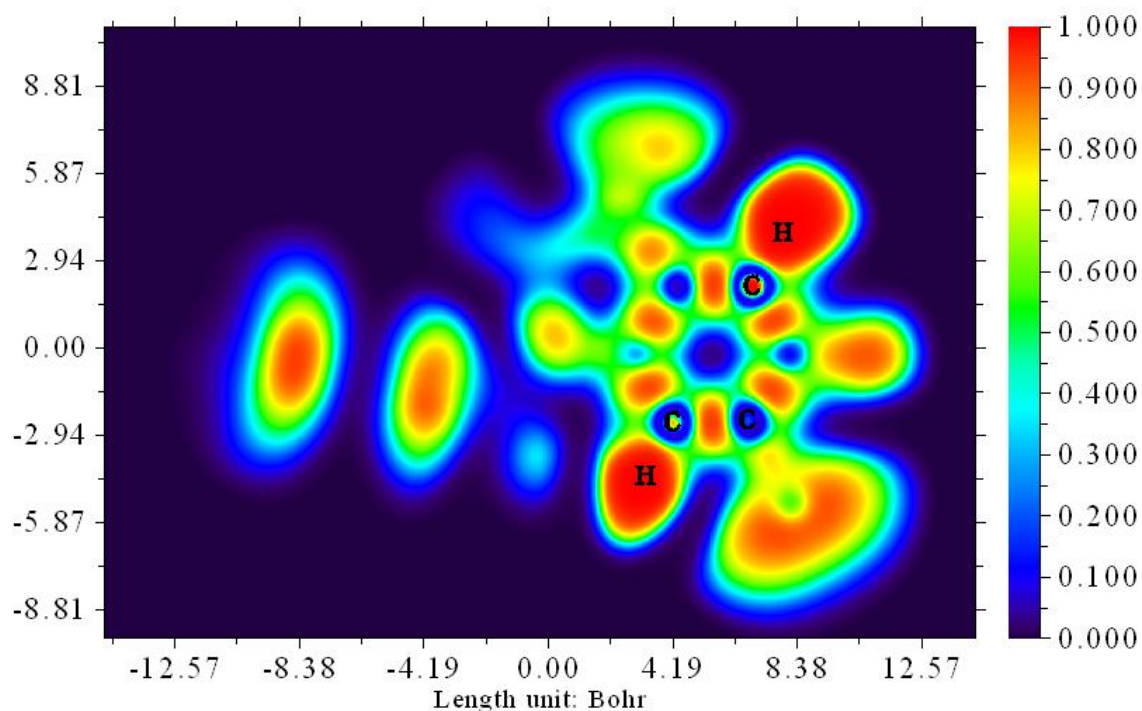


Fig. 17 – ELF plot of compound with PubChem CID 16203797

Therefore the usefulness of the programmatic tool is demonstrated in reducing the computational complexity of virtual screening for identifying drugs against coronavirus achieved through programmatic automation of data mining of PubChem to collect data required to implement a machine learning based AutoQSAR algorithm for automatic drug lead generation for coronavirus. The program requires that the drug leads generated by the program are required to satisfy the Lipinski's drug-likeness criteria. Further, the program automates the *In Silico* modelling of the interaction of the compounds generated as drug leads and the drug target of coronavirus and stores the results in the working folder of the user. Thus, the program helps achieve a completely 'hands off' automation in identifying drugs candidates against coronavirus which further has to be examined for drug potential through experimental testing such as *In Vitro* and *In Vivo* testing.

4. Conclusion and future scope

Thus the presented work is an attempt to automate the dry lab drug discovery workflow of drug discovery for coronavirus by a python program automate the process of data mining PubChem database to collect data required to perform a machine learning based AutoQSAR algorithm through which drug leads for coronavirus are generated. The data acquisition from PubChem was carried out through python web scrapping techniques. The workflow of a machine learning based AutoQSAR involves feature learning and descriptor selection, QSAR modelling, validation and prediction. The drug leads generated by the program are required to satisfy the Lipinski's drug likeness criteria. Drug leads generated by the program are fed as programmatic inputs to an In Silico modelling package to computer model the interaction of the compounds generated as drug leads and coronaviral drug target, main protease of SARS-CoV 2 identified with PDB ID : 6Y84. The results are stored in the working folder of the user. The program also generates protein-ligand interaction profiling and stores the visualized images in the working folder of the user. Thus our programmatic tool ushers in the new age automatic ease in drug candidate identification for coronavirus through a fully automated QSAR and an automated *In Silico* modelling of the drug leads generated by the AutoQSAR algorithm. The program reduces the computational complexity of virtual screening and helps identify best drug candidates to employ computationally more expensive methods such as Molecular Dynamics and DFT studies to study binding and reactive profiles of select drug candidates. While the compounds identified through the automated workflow must be test experimentally by the experimental research community for their drug potential against coronavirus, there is still a lot of scope to make the automation algorithm more self-aware of technical nuances which will help increase its accuracy in drug identification which we bring to the attention of the computational research community for their scholarly attention and efforts on the same.

References

1. Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Bryant, S. H. (2015). PubChem Substance and Compound databases. *Nucleic Acids Research*, 44(D1). doi:10.1093/nar/gkv951
2. Kim, S., Thiessen, P. A., Bolton, E. E., & Bryant, S. H. (2015). PUG-SOAP and PUG-REST: Web services for programmatic access to chemical information in PubChem. *Nucleic Acids Research*, 43(W1). doi:10.1093/nar/gkv396
3. Swain, M. (2014). PubChemPy: A way to interact with PubChem in Python.
4. Bhardwaj, V. (2014). Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. *Medicinal Chemistry Research*, 23(12), 4991-5007. doi:10.1007/s00044-014-1072-3
5. Eriksson, L., & Johansson, E. (1996). Multivariate design and modeling in QSAR. *Chemometrics and Intelligent Laboratory Systems*, 34(1), 1-19. doi:10.1016/0169-7439(96)00023-8.
6. Bajot, F. (2009). The Use of Qsar and Computational Methods in Drug Design. *Challenges and Advances in Computational Chemistry and Physics Recent Advances in QSAR Studies*, 261-282. doi:10.1007/978-1-4020-9783-6_9
7. Sippl, W. (2009). 3D-QSAR – Applications, Recent Advances, and Limitations. *Challenges and Advances in Computational Chemistry and Physics Recent Advances in QSAR Studies*, 103-125. doi:10.1007/978-1-4020-9783-6_4.
8. Mitchell, R. (2018). *Web Scraping with Python: Collecting More Data from the Modern Web*. O'Reilly Media, Incorporated.
9. Vanden Broucke, S., & Baesens, B. (2018). *Practical Web scraping for data science* (pp. 3-5). New York, NY: Apress.
10. Ton, A. T., Gentile, F., Hsing, M., Ban, F., & Cherkasov, A. (2020). Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Molecular informatics*.
11. Dixon, S. L., Duan, J., Smith, E., Von Bargen, C. D., Sherman, W., & Repasky, M. P. (2016). AutoQSAR: an automated machine learning tool for best-practice quantitative structure–activity relationship modeling. *Future medicinal chemistry*, 8(15), 1825-1839.
12. Kim, S., & Cho, K. H. (2019). PyQSAR: A Fast QSAR Modeling Platform Using Machine Learning and Jupyter Notebook. *Bulletin of the Korean Chemical Society*, 40(1), 39-44.

13. Lipinski, C. A. (2004). Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technologies*, 1(4), 337-341.
14. Kumar, V., Jung, Y. S., & Liang, P. H. (2013). Anti-SARS coronavirus agents: a patent review (2008–present). *Expert opinion on therapeutic patents*, 23(10), 1337-1348.
15. Nascimento Junior, J. A. C., Santos, A. M., Quintans-Júnior, L. J., Walker, C. I. B., Borges, L. P., & Serafini, M. R. (2020). SARS, MERS and SARS-CoV-2 (COVID-19) treatment: a patent review. *Expert Opinion on Therapeutic Patents*, (just-accepted).
16. Fischer, A., Sellner, M., Neranjan, S., Smieško, M., & Lill, M. A. (2020). Potential Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 606 Million Compounds. *International Journal of Molecular Sciences*, 21(10), 3626.
17. Rifaioğlu, A., Sinoplu, E., Atalay, V., Martin, M., Cetin-Atalay, R., & Dogan, T. (2020). DEEPScreen: High Performance Drug-Target Interaction Prediction with Convolutional Neural Networks Using 2-D Structural Compound Representations. *Chemical Science*.
18. Gentile, F., Agrawal, V., Hsing, M., Ton, A. T., Ban, F., Norinder, U., ... & Cherkasov, A. (2020). Deep Docking: A Deep Learning Platform for Augmentation of Structure Based Drug Discovery. *ACS Central Science*.
19. Liao, Z., You, R., Huang, X., Yao, X., Huang, T., & Zhu, S. (2019, November). DeepDock: Enhancing Ligand-protein Interaction Prediction by a Combination of Ligand and Structure Information. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 311-317). IEEE.
20. Rodgers, S. L., Davis, A. M., Tomkinson, N. P., & van de Waterbeemd, H. (2011). Predictivity of simulated ADME AutoQSAR models over time. *Molecular informatics*, 30(2-3), 256-266.
21. Deshpande, N., Address, K. J., Bluhm, W. F., Merino-Ott, J. C., Townsend-Merino, W., Zhang, Q., ... & Kramer Green, R. (2005). The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic acids research*, 33(suppl_1), D233-D237.
22. Jaghoori, M. M., Bleijlevens, B., & Olabarriaga, S. D. (2016). 1001 Ways to run AutoDock Vina for virtual screening. *Journal of computer-aided molecular design*, 30(3), 237-249.

23. Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J. (2016). Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature protocols*, 11(5), 905.
24. Tibaut, T., Borišek, J., Novič, M., & Turk, D. (2016). Comparison of in silico tools for binding site prediction applied for structure-based design of autolysin inhibitors. *SAR and QSAR in Environmental Research*, 27(7), 573-587.
25. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., & Schroeder, M. (2015). PLIP: fully automated protein–ligand interaction profiler. *Nucleic acids research*, 43(W1), W443-W447.
26. Case, D., et al. "AMBER 18; 2018." *University of California, San Francisco*.
27. Jakalian, Araz, et al. "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method." *Journal of computational chemistry* 21.2 (2000): 132-146.
28. Hernández, Eduardo R. "Molecular Dynamics: from basic techniques to applications (A Molecular Dynamics Primer)." *AIP Conference Proceedings*. Vol. 1077. No. 1. American Institute of Physics, 2008.
29. Parr, Robert G. "Density functional theory of atoms and molecules." *Horizons of quantum chemistry*. Springer, Dordrecht, 1980. 5-15.
30. Neese, Frank. "ORCA—An ab initio DFT and Semi-empirical Electronic Structure Package, version 2.9. 1." *Max Planck Institute for Bioinorganic Chemistry* (2012).
31. Lu, Tian, and Feiwu Chen. "Multiwfn: a multifunctional wavefunction analyzer." *Journal of computational chemistry* 33.5 (2012): 580-592.
32. Prasana, Johanan Christian, S. Muthu, and Christina Susan Abraham. "Molecular docking studies, charge transfer excitation and wave function analyses (ESP, ELF, LOL) on valacyclovir: A potential antiviral drug." *Computational biology and chemistry* 78 (2019): 9-17.
33. Muthu, S., et al. "Spectroscopic (FT-IR, FT-Raman) investigation, topology (ESP, ELF, LOL) analyses, charge transfer excitation and molecular docking (dengue, HCV) studies on ribavirin." *Chemical Data Collections* 17 (2018): 236-250.
34. Khemalapure, Seema S., et al. "Spectroscopic (FT-IR, FT-Raman, NMR and UV-Vis), ELF, LOL, NBO, and Fukui function investigations on (5-bromo-benzofuran-3-yl)-acetic acid hydrazide (5BBAH): Experimental and theoretical approach." *Journal of Molecular Structure* 1196 (2019): 280-290.