

Development of Putative Isospecific Inhibitors for HDAC6 using Random Forest, QM-Polarized docking, Induced-fit docking, and Quantum mechanics

JOEL Ireoluwa Yinka¹, **ADIGUN** Temidayo Olamide¹, **BANKOLE** Olukayode Olusola^{1, 2}, **AJIBOLA** Ahmeedah Ololade¹, **OFENIFORO** Emmanuel Bankole¹, **AUTA** Faith Beyaan³, **OZOJIOFOR** Ugochukwu Okechukwu⁴, **REMI-ESAN** Ifelolu Adeseye², **AKANDE** Aminat Ifeoluwa¹

¹Department of Biochemistry, University of Ilorin, Ilorin Kwara state

Department of Science Laboratory Technology, School of Pure & Applied Sciences, Federal Polytechnic Ilaro, Ogun state

³Departement of Microbiology, Federal University Of Agriculture Makurdi, Benue state

⁴Department of Biotechnology, Nigerian Defence Academy, Kaduna, Kaduna State

Keywords: HDAC6; Qm-polarized docking; Induced-Fit docking; Machine learning; Quantum mechanics

Corresponding Author: **JOEL** Ireoluwa Yinka, Department of Biochemistry, University of Ilorin, Nigeria; joelireoluwa@gmail.com or 17-68ez012pg@students.unilorin.edu.ng

Abstract

Histone deacetylases have been recognized as a potential target for epigenetic aberrance reversal in the various strategies for cancer therapy, with HDAC6 implicated in various forms of tumor growth and cancers. Diverse inhibitors of HDAC6 has been developed, however, there is still the challenge of iso-specificity and toxicity. In this study, we trained a Random forest model on all HDAC6 inhibitors curated in the ChEMBL database (3,742). Upon rigorous validations the model had an 85% balanced accuracy and was used to screen the SCUBIDOO database; 7785 hit compounds resulted and were docked into HDAC6 CD2 active-site. The top two compounds having a benzimidazole moiety as its zinc-binding group had a binding affinity of -78.56kcal/mol and -78.21kcal/mol respectively. The compounds were subjected to exhaustive docking protocols (Qm-polarized docking and Induced-Fit docking) in other to elucidate a binding hypothesis and accurate binding affinity. Upon optimization, the compounds showed improved binding affinity (-81.42kcal/mol), putative specificity for HDAC6, and good ADMET properties. We have therefore developed a reliable model to screen for HDAC6 inhibitors and suggested a series of benzimidazole based inhibitors showing high binding affinity and putative specificity for HDAC6.

1.0. Introduction

Cancer is a complex disease showing multifaceted and multifactorial presentation, with a high rate of morbidities and mortalities second only to cardiovascular disease.^[1] There is growing clarity on the significant effect of epigenetic alterations in cancer initiation and development. The post-translational modifications involve reversible acetylation and deacetylation of histone^{[2][3]} by adding or removing acetyl groups to or from specific lysine residues respectively by histone acetylases and deacetylases enzymes at the histone tails.^{[4][5][6][7]} This increases or decreases ionic interaction respectively between the positively charged histones and negatively charged DNA to yield more or less compact chromatin structure,^[8] thereby regulating gene expression and the subsequent accessibility of cell growth transcription machinery as well as RNA polymerase enzyme accordingly for gene transcription repression or enhancement and microtubule stability regulation in response to both intrinsic and extrinsic stimuli.^{[9][8]}

The DNA histone deacetylation by Histone deacetylases results in tighter DNA-histone core wrapping, thereby leading to condensation of chromatin with its resultant gene expression silencing^[10] as well as gene transcription factors repression.^[11] This explains why over-expression of HDACs, resulting in abnormal levels of deacetylated histones and its associated normal gene transcription inhibition, have been linked to cancer cell proliferation,^{[12][13]} as different HDACs are over-expressed in many cancer cell lines and tissues, thereby indicating their crucial role in tumor development.^{[14][10]}

The human histone deacetylase 6 belongs to Class IIb of the family of zinc-dependent hydrolases in the eighteen currently identified isoforms of histone deacetylase enzymes based on their sequence homology, sub-cellular distribution and catalytic activity.^{[13][11]} It has two catalytic domains including a ubiquitin-binding domain at the C-terminal region and a dynein-binding domain.^{[15][16]} Histone deacetylase 6 has been shown to have a wide range of abnormal differential expressions in solid tumors and hematological malignancies including prostate cancer, lung cancer, liver cancer, ovarian cancer, renal cancer, thyroid cancer, breast cancer, urothelial cancers, gliomas, lymphomas, etc.^{[17][18][19][20]}, thereby rendering it a viable target for cancer therapy.^[21]

There are currently many HDAC inhibitors or pan-inhibitors that have shown certain anti-tumor activity under experimental and clinical trials,^{[22][23]} however non-specificity, low bioavailability and potential for drug-drug interactions (through cytochrome P450 inhibition) has been a major

challenge.^[1] The pharmacophoric features of HDAC inhibitors consist of a capping group (surface recognition moiety), a hydrophobic spacer (the linker), and a zinc-binding group (ZBG).^{[24][25]} The current ZBG ranges from hydroxamic acid, 2-aminoanilide, electrophilic ketones, to short-chain fatty acids.^[26] Bolden *et al.* (2006), Guo *et al.* (2012) and Lobera *et al.* (2018) suggested that changes in the cap group, the linker or the ZBG can individually provide selectivity for specific HDAC isoform inhibition^{[27][28]} as there is widespread speculation that development of isotype-selective inhibitors may contribute to therapeutic index improvement over non-selective ones.^[13]

In this study, we aimed to identify, optimize potential HDAC6 specific inhibitors, and determine the mechanism of inhibition and specificity. The workflow for this study includes:

- building a Random Forest Classifier (RFC) model which “learned” from all known HDAC6 inhibitors that have been deposited in the ChEMBL database from 2003 to date, and validated extensively;
- screening the SCUBIDOO database (a database of 21,000,000 compounds set up for fast and efficient lead identification) for hit compounds;
- subjecting the hit compounds to Induced-fit docking, QM-polarized docking, QM-MM geometric optimization, and Molecular dynamics (Normal mode analysis), in order to determine binding affinity, elucidate a binding hypothesis (with interest in the mechanism of specificity and inhibition) and stability of the protein-ligand complex;
- optimization of compounds (Bioisostere replacement protocols) for improved binding affinity and specificity;
- calculate electronic descriptors (using Quantum mechanics) to determine the mode of chemical reactivity (nucleophilic or electrophilic) and stability of the optimized compounds;
- predict ADMET properties.

2.0. Material and Methods

2.1. Hardware and software

All analysis was carried out using a Linux Ubuntu 18.04 distro system running on a 12 GB RAM, core i5, 4 Cores, 2.5GHz.

All python packages were run on Python 3.6 using Jupyter Lab 1.2.6. Python packages used include Scikit-learn (v0.22.2) for model training, Feature Selector for feature extractions and preparations, Pandas (v1.0.3) for data waggling, Matplotlib (v3.2.1) and Seaborn (v0.10.0) for data visualizations, Numpy (v1.18.1), SHAP and PDPbox for model interpretation, TPOT (v0.11.1) for AutoML analysis.

2.2. Data Extraction and Preparation

All inhibitors of HDAC6 present in the ChEMBL^[29] database were downloaded (3,744) and imported into a standalone MySQL database we created for analysis. All inhibitors with missing IC50 values were removed, thereafter inhibitor smiles and corresponding IC50 values were extracted into a CSV sheet.

Using the MOE descriptor calculator, 2D descriptors were calculated for the inhibitors. A python script implementing Feature selector module was written to remove inter-correlated descriptors (correlation threshold was set at 0.75). Using the same module, all descriptors that did not contribute to 0.95 cumulative importance were removed (feature selector uses XGBoost algorithm to determine the feature importance of the descriptors).

A python script was written to convert IC50 values to pIC50 values ($pIC50 = 9 - \log_{10}(IC50)$) (All IC50 were in nM). The pIC50 values were converted to categorical values of active (1) and non-active (0). The minimum pIC50 (2: the equivalent of 10000000 IC50 value) and maximum pIC50 value (12: the equivalent of 0.002nM IC50 value) were considered before setting activity threshold; this is to ensure balanced classes: Activity threshold was therefore set to 7 pIC50 and above (the equivalent of 37nM IC50 value). The data-set was finally divided into training, test, and external validation set.

2.3. Model building

A python script was written which implemented a Tree-based pipeline optimizer (TPOT).^[30] TPOT is an AutoML python module that searches for the best machine learning algorithm for our modeling task. It uses a genetic algorithm to search different tree-based algorithms (and corresponding best hyperparameter values), feature processing and selection algorithm, and outputs the pipeline with the highest accuracy. The following TPOT parameters were set: Generation: 100, Population Size: 100, Cross-validation: 10, and Early stopping: True. All other parameters were left at default values.

Random forest classifier (RFC) algorithm was the best performing algorithm with the following hyper-parameters: bootstrap: True, criterion: Gini, maximum features: 0.3, minimum samples leaf: 13, minimum samples split: 13, and number of estimators: 100.

Using Sci-kit learn RFC was fitted on the dataset and evaluated using Sci-kit learn classification evaluation metrics^[31] the following metric was calculated to evaluate the performance of the RFC model after 10 fold cross-validation: Recall (sensitivity), specificity, Precision, F1, Accuracy, Balanced accuracy, Error rate, and confusion matrix as described by Hossin *et al.*^[32]

In other to provide a suitable explanation for the RFC model we wrote a python script implementing SHAP and PDPbox module. SHAP plotted feature importance of the descriptors^[33], PDPbox plotted the partial dependence plot (PDP), and individual conditional exception (ICE) plot.^[34]

2.4. Database Screening

We used the SCUBIDOO database a freely available database of 21 million virtual compounds.^[35] The screening was in four steps:

- Predicting the activity of the M sample (~99,000 compounds) representation of SCUBIDOO with the RFC model
- Docking of predicted active compounds to HDAC6 active site and identification of part/fragment of the molecule making binding interaction of interest (chelating interaction)
- Searching the SUBIDOO database for compounds having the identified fragments
- Hits from the search are downloaded for further analysis

2.5. Structure-based drug discovery

2.5.1. Molecular docking

The crystal structure of HDAC6 in complex with TSA in the catalytic domain (CD) 2 (PDB: 5WGI) was downloaded from PDB. The protein was prepared using the Schrödinger protein preparation wizard^[36]; missing side chains and loops were filled with prime,^[37] waters beyond 5Å from het group was deleted and het states were generated using Epik^[38] (pH 7.0 +/- 2.0); all other parameters were left at default values.

The downloaded SCUBIDOO compounds were prepared using the Schrödinger Ligprep module; force field minimization using OPLS2005,^[39] het states generated using Epik (pH 7.0 +/- 2.0) and metal-binding states were added. All other parameters were left at default values.

Using the receptor grid generation module of Schrödinger the Grid file (active site coordinates and constraints) for the protein was extracted. The compounds were thereafter docked using Schrödinger's virtual screening workflow. Briefly, the workflow consists of filtering (drug-likeness criteria), ligand preparation (using Ligprep), docking (the docking phase involved using the three Glide docking protocols: HTVS, SP, XP) and postprocessing (rescoring) using prime molecular mechanics-generalized Born surface area (MM-GBSA)^[40]. All parameters were left at default values. The docking protocol was validated by redocking the co-crystallized ligand; the RMSD value was thereafter calculated.

2.5.2. Molecular Mechanics Generalized Born Surface Area

Molecular Mechanics Generalized Born Surface Area (MM-GBSA) evaluated the binding free energies (binding affinity) and minimized the docked protein-ligand complex.^[41] Using VSGB 2.0 implicit solvation model and OPLS-2005.^[42] The binding free energy was calculated using Eq.1

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} \quad (1)$$

Where G_{complex} , G_{protein} , and G_{ligand} represent the free binding energy of the protein-ligand complex, protein, and ligand respectively.

2.5.3. Induced-fit docking

Schrodinger induced-fit docking module^{[43][44]} was used to perform Induce-fit docking. All parameters were left at default values. Briefly, the ligand is docked into the active site of the protein using Glide protocol (SP)[45] with the protein active site residues held rigid, next the protein side chains or backbone are refined using prime refinement module,^[37] finally, the ligand is redocked into the refined protein conformation and induced fit docking score is calculated to rank the protein-ligand complex.

2.5.4. QM-polarized docking

QM-polarized docking (QPLD) was implemented using the Schrödinger QPLD module.^[46] Briefly, the ligands are docked into the active site of the protein using glide docking protocol^[45] and partial charges calculated by force field minimization, next the best-docked pose are selected and partial charges induced on the ligands by the active site residues are calculated using quantum mechanics this calculation is implemented using Schrödinger Qsite module,^[47] finally, the ligands are redocked using the QM derived partial charges and ranked using glide docking score.

2.5.5. QM/MM optimization

Using the Schrodinger Q-site module QM/MM optimization was carried out on docked protein-ligand complexes. Briefly, the ligand and active site residue (side-chain and backbone) involved in the interactions are treated as the Quantum mechanics (QM) region while the other protein complex is treated as molecular mechanics (MM) region. The Qm calculation was done using Density functional theory (DFT) with Becke's three-parameter exchange potential, Lee-Yang-Parr correlation functional (B3LYP) and basis set 631G** level,^[41] while MM region was treated using OPLS2005; minimization was done using Truncated Newton, maximum cycle:1000 with convergence criterion set to Energy Gradient.^{[41][47]} All other parameters were left at default.

2.5.6. Molecular dynamics

The stability of the docked protein-ligand complex investigated was using Normal mode analysis (NMA). The ease of deformability of the complex (eigenvalues), B-factor, variance, co-variance map, and the elastic network was calculated.^[48] Using iMOD server (<http://imods.chaconlab.org>) NMA calculations were implemented

2.5.7. Bioisostere replacement

In other optimized the lead compounds Schrodinger Bioisostere replacement module was used. The module replaced certain functional groups with the ligands while retaining biological activity or improving activity.

2.6. Electronic descriptor calculations

Quantum mechanics (QM) theories were used to determine Electronic descriptors of the ligands. Using Schrodinger Jaguar single energy point module,^[49] the ligands were optimized geometrically using hybrid Density flow theory (DFT) with Becke's three-parameter exchange potential, Lee-Yang-Parr correlation functional (B3LYP) and basis set 631G** level. The following descriptors were calculated: Highest Occupied Molecular Orbital (HOMO) energy, Lower Unoccupied Molecular Orbital (LUMO) energy, and Molecular Electrostatic Potential (MESP). From the HOMO and LUMO calculations, the following descriptors were further extrapolated:

- HOMO-LUMO gap= $E_{LUMO} - E_{HOMO}$;
- ionization energy (I) = $-E_{HOMO}$;
- electron affinity (A) = $-E_{LUMO}$;
- global hardness (η) = $(-E_{HOMO} + E_{LUMO})/2$
- chemical potential (μ) = $(E_{HOMO} + E_{LUMO})/2$ can be determined.
- Parr *et al.*^[50] proposed the global electrophilicity power of a ligand as $\omega = \mu/2\eta$.

2.7. ADMET property prediction

Using the Discovery studio ADMET property prediction module ADMET descriptors were calculated for the compounds. MOE (Molecular Operating Environment 2015) descriptor module was used to determine Linpinkis's drug-likeness and Opera lead-likeness of the compounds.

3.0. Results

3.1. Ligand-based drug discovery

3.1.1. Data extraction and preprocessing

A total of 3742 HDAC6 inhibitors developed from 2003 to date with IC₅₀ values ranging from 0.1 to 100000nM were downloaded from the ChEMBL database. The smiles of these inhibitors were extracted and a total of 206 2D descriptors were calculated using MOE software; the data was prepared and features extracted using a python script that implemented feature selector module (it removed inter-correlated, zero and low importance descriptors). Out of the 206 descriptors, 43 contributed to 0.95 cumulative importance and were selected for model building

3.1.2. Tree-Based Pipeline Optimization Analysis

Different machine learning algorithms have been developed, each with its array of hyper-parameters that should be tuned in order to achieve accurate models suited for different datasets. This tuning however requires expertise, experience, coupled with trial and error which is generally challenging and time-consuming. In order to solve this problem Auto-machine learning (AutoML) modules have been developed, these modules exhaustively search different algorithms and hyper-parameters that best suit the modeling task at hand. One of these AutoML modules is Tree-Based Pipeline Optimization (TPOT).^[30]

TPOT is an AutoML module that uses a genetic algorithm to search for tree-based algorithms, feature selection, and preprocessing algorithms with their corresponding hyper-parameter suitable for the provided data.^[50] TPOT classifier was therefore used to search for the best machine learning algorithm suited for our data (see methods). The analysis produced the Random Forest Algorithm with an accuracy of 0.83 after 100 generations of search, with hyper-parameter: criterion: Gini; maximum features: 0.3; minimum samples leaf: 13; minimum samples split: 13; number of estimators: 100.

3.1.3. Random Forest Classifier Model Building and Evaluation

The dataset was divided into 3 (training, test, and external validation set) with a ratio of 60:20:20 respectively, with the above hyper-parameters we fitted the Random Forest Classifier (RFC) on the training data set and evaluated its performance on the test and external data-set. Confusion matrix (Figure S1) and other classification matrix were used to evaluate the model performance (Table 1). The training set had a precision: 0.89; recall: 0.94; accuracy: 0.90; and balanced accuracy: 0.90 (Table 1). The test set had a precision: 0.78; recall: 0.90; accuracy: 0.87; and balanced accuracy: 0.86 (Table 1). The external validation set had a precision: 0.87; recall: 0.91; accuracy: 0.87; and balanced accuracy: 0.86 (Table 1). A Roc curve (True-positive rate Vs. False-positive rate) of the RFC model performance on its training, test, and external validation set was plotted (Figure 1) and AUC scores determined (Table 1; Figure 1). The receiver

operative curve (ROC) curve and Area under Curve (AUC) score evaluates how well a model picks a true positive ahead of a false positive in a pool of true positive and false positives.[51] The training set had an AUC of 0.91; test set an AUC of 0.83; external validation set an AUC of 0.86

Table 1: Model evaluation metrics

Metric	Training-set	Test-set	External-set
Precision	0.89	0.78	0.87
Recall	0.94	0.90	0.91
Accuracy	0.90	0.82	0.87
F1	0.91	0.84	0.89
Balanced Accuracy	0.90	0.83	0.86
ROC-AUC	0.91	0.83	0.86

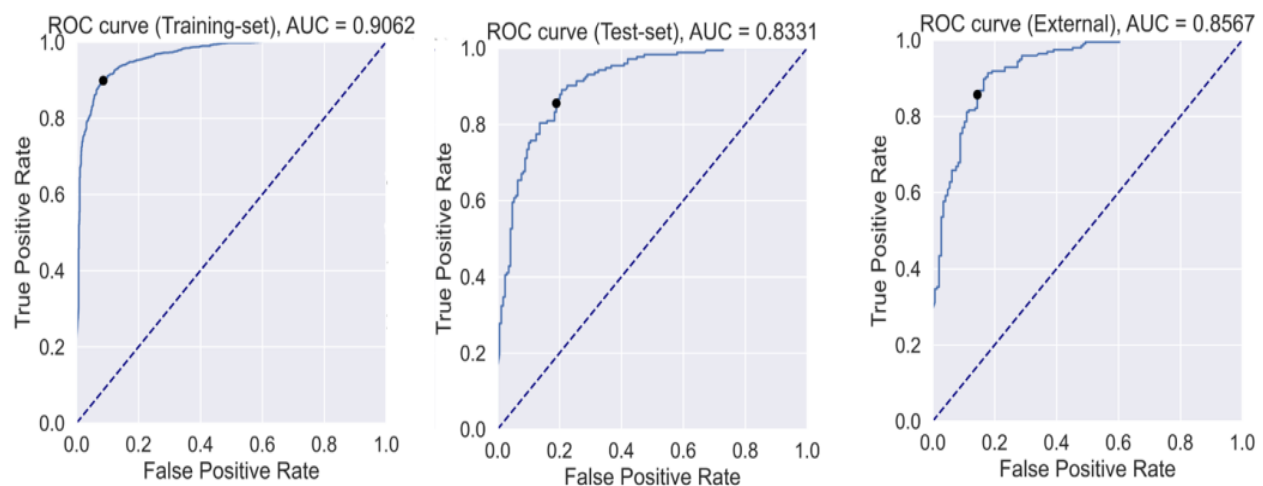


Figure 1: ROC-curve for Training, Test, and External validation data-set

3.1.4. Cross-Validation

Using Sci-Kit learn cross-val-predict function a 10 fold cross-validation was implemented on the datasets (Table 2). The training set had a precision: 0.83; recall: 0.39; accuracy: 0.83; and balanced accuracy: 0.83 (Table 2). The test set had a precision: 0.74; recall: 0.85; accuracy: 0.78; and balanced accuracy: 0.77 (Table 2). The external validation set had a precision: 0.79; recall: 0.90; accuracy: 0.80; and balanced accuracy: 0.78 (Table 2). The cross-validation Roc curve of the RFC model performance on its training, test, and external validation set was plotted (Figure S2) and AUC scores determined (Table 2). The Area under Curve (AUC) score was evaluated,

the training set had an AUC: 0.83; test set: 0.77; external validation set: 0.77. Based on these results, we held the model to be a robust and reliable model for screening.

Table 2: Cross-validation Evaluation Metrics

Metric	Training-set	Test-set	External-set
Precision	0.83	0.74	0.79
Recall	0.89	0.85	0.90
Accuracy	0.83	0.78	0.80
F1	0.86	0.78	0.83
Balanced Accuracy	0.83	0.77	0.78
ROC-AUC	0.83	0.77	0.77

3.1.5. Model Interpretation

Understanding how a model is using inputted descriptors to make predictions about the dependent variable (biological activity) is not only important in optimizing the model but also can give insights on dataset improvement (i.e. descriptors to improve for better activity).^[34] Using SHAP and PDPbox python module we plotted feature importance plot, partial dependence plot (PDP), Individual conditional expectation (ICE), and partial dependence interaction plots (see methods). SHAP module plotted the feature importance plot for the RFC model and showed the importance in this order GCUT_SLOGP_0, PEOE_VSA-1, SMR_VSA2, pKA, GCUT_SLOGP_2, etc. (Figure 2). It should be noted however that the contribution of these descriptors is towards class probability. Therefore, figure 2 shows the importance and contribution towards the probability of both classes modeled (class 1: active (blue), class 0: non-active (red))

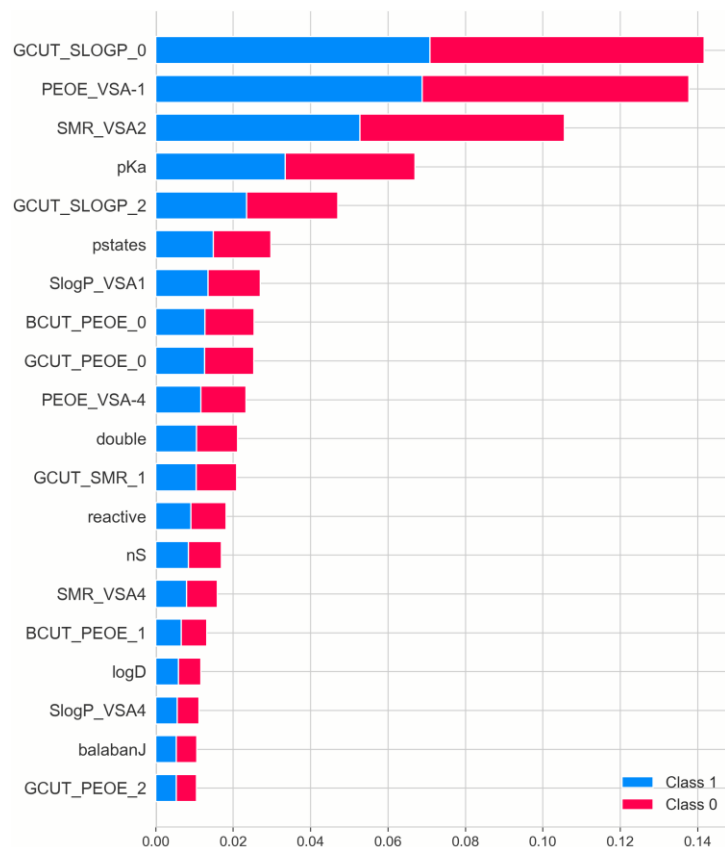


Figure 2: Random Forest Classifier feature importance

The feature importance plot is one important plot for understanding machine learning model predictions however; it does not show how varying values of individual descriptors affect the model predictions. The partial dependence plot (PDP) solves this problem by showing the marginal effect of varying values of a descriptor on model prediction.^[52] The PDP showed that increasing values of PEOE_VSA-1, SMR_VSA2, and GCUT_SLOGP_2 increased RFC model active class probability predictions (Figure S3). This is, however, not true for GCUT_SLOGP_0 and pKa descriptors, increasing values of GCUT_SLOG_0 resulted in a negative contribution towards the RFC model active class probability contribution. pKa descriptors contributed positively only within a narrow range (-10 to 12) (Figure S3a[i], S3d[i]).

Partial dependence plot goes a step forward in providing insight into the ‘black-box’ of the RFC model however, the PDP plot is an average of all data points, this hides possible variations. Individual Conditional Expectation (ICE) plot, therefore, disintegrates the PDP plot into their corresponding data points showing possible trend variation.^[52] Using the PDPbox module we plotted ICE plots for the top five descriptors (Figure 5). The ICE plot suggests that although most of the data point followed the PDP plot trend some however deviated, for example, the PDP plot indicated that increasing values of GCUT_LOGP_0 resulted in negative contribution towards active class probability prediction but the ICE plot has however shown that this is not true for all data points (this also applies to pKa) (Figure S4).

Going a step further in trying to make the model prediction process as clear as possible we investigated the partial dependence interaction matrix. This matrix shows how various values of descriptors when considered together affect the active class prediction probability (this, therefore, reveals optimum values for active class prediction probability); the matrix considered two features/descriptors at a time. Overall GCUT_LOGP_0 optimum values for various interaction consideration was -1.17, pKa optimum value was 8.8, PEOE_VSA-1 optimum value was 121.93, SMR_VSA2 optimum value was 71.78, and GCUT_SLOGP_2 optimum value was 0.33 (Figure 3).

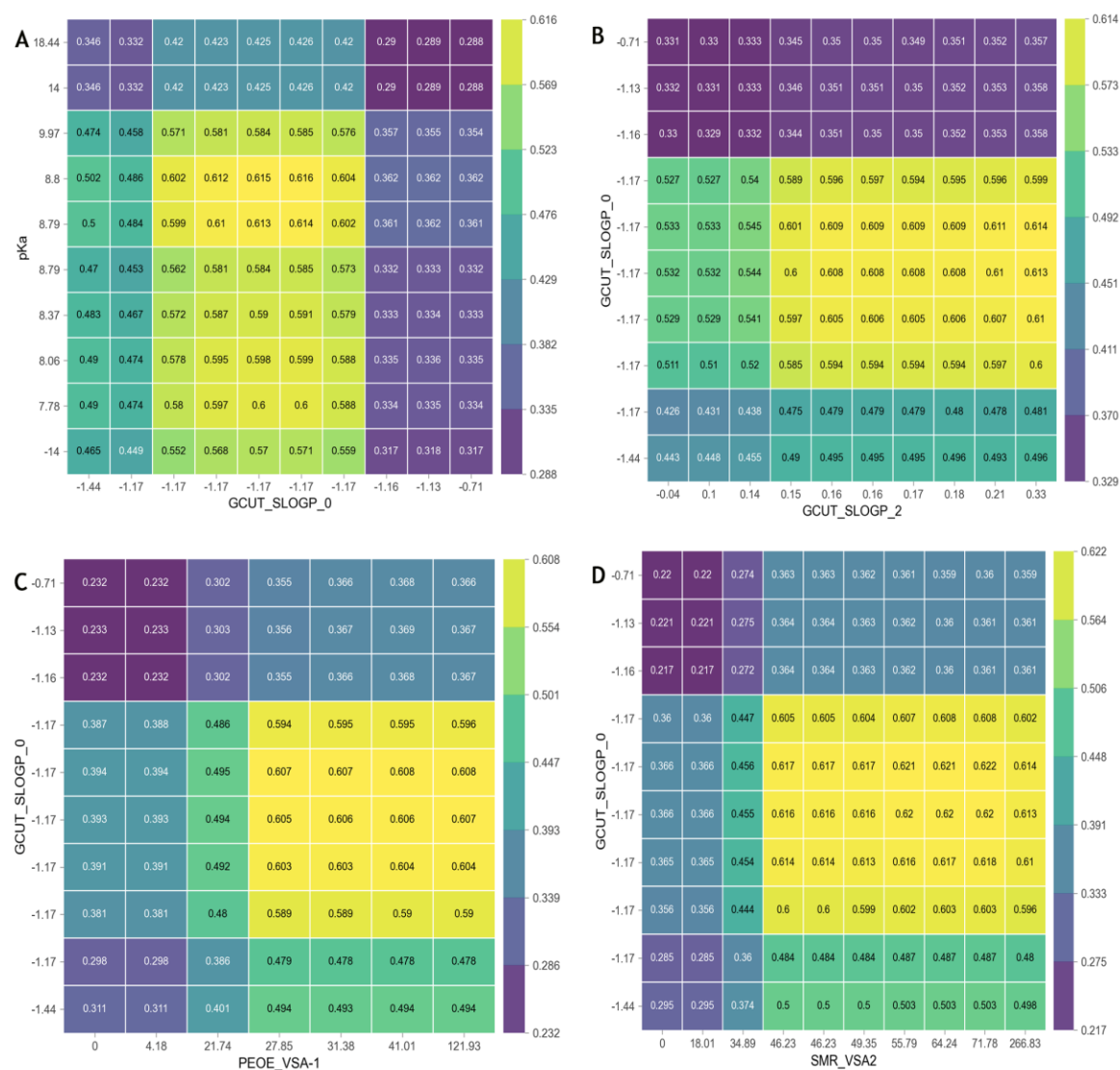


Figure 3: Partial dependence Interaction plot of a) pKa and GCUT_SLOGP_0 b) GCUT_SLOGP_0 and GCUT_SLOGP_2 c) GCUT_SLOGP_0 and PEOE_VSA-1 d) GCUT_SLOGP_0 and SMR_VSA2

The model interpretation analysis has shown that GCUT_SLOGP_0 is the most important descriptor for our RFC model with an optimum value of -1.17 and that increasing values of this descriptor resulted in a negative contribution towards active class prediction. This, therefore, suggests that an increase in GCUT_SLOGP_0 values might result in a reduction in the biological activity of HDAC6 inhibitors. pKa was the third important descriptor and followed a similar trend with GCUT_SLOGP_0 with an optimum value of 8.8, increasing biological activity through pKa is however within a narrow range after which biological activities drop. Other descriptors however increased biological activity as their values increased. It should be noted that this analysis is suggestive and specific only to this model, this, therefore, means that other machine learning algorithms might result in deferent results hence experimental verification is still needed (Table S1 describes the descriptors used in developing the model).

3.1.6. Database Screening

SUBIDOO a database with 21,000,000 virtual compounds/products^[35] was selected for screening, majorly for two reasons: the database has been set up in such a way as to reduce the computational cost of screening (see methods) and all compounds in the database come with detailed information on its synthesis and possible side reaction.

The RFC model was used to screen the M sample of the SCUBIDOO database for active compounds. Of the 99,977 compounds, only 167 compounds were classified as active; the 167 compounds were docked into the active site of HDAC6. The compound was filtered using the following criteria: docking score: > -8.0 kcal/mol and interactions with catalytic Zn^{2+} ion. Two compounds were selected having met these criteria (docking score: -9.216 and -9.034kcal/mol respectively; formation of salt bridges with catalytic Zn^{2+} ion (Figure 4)). The moiety of these compounds making interaction with the Zn^{2+} was the hydroypyrimidine functional group and thiadiazole functional group respectively. The SCUBIDOO database was, therefore, searched for compounds containing these functional groups and a total of 7785 compounds resulted from the search (Figure S4 shows the workflow for screening the database).

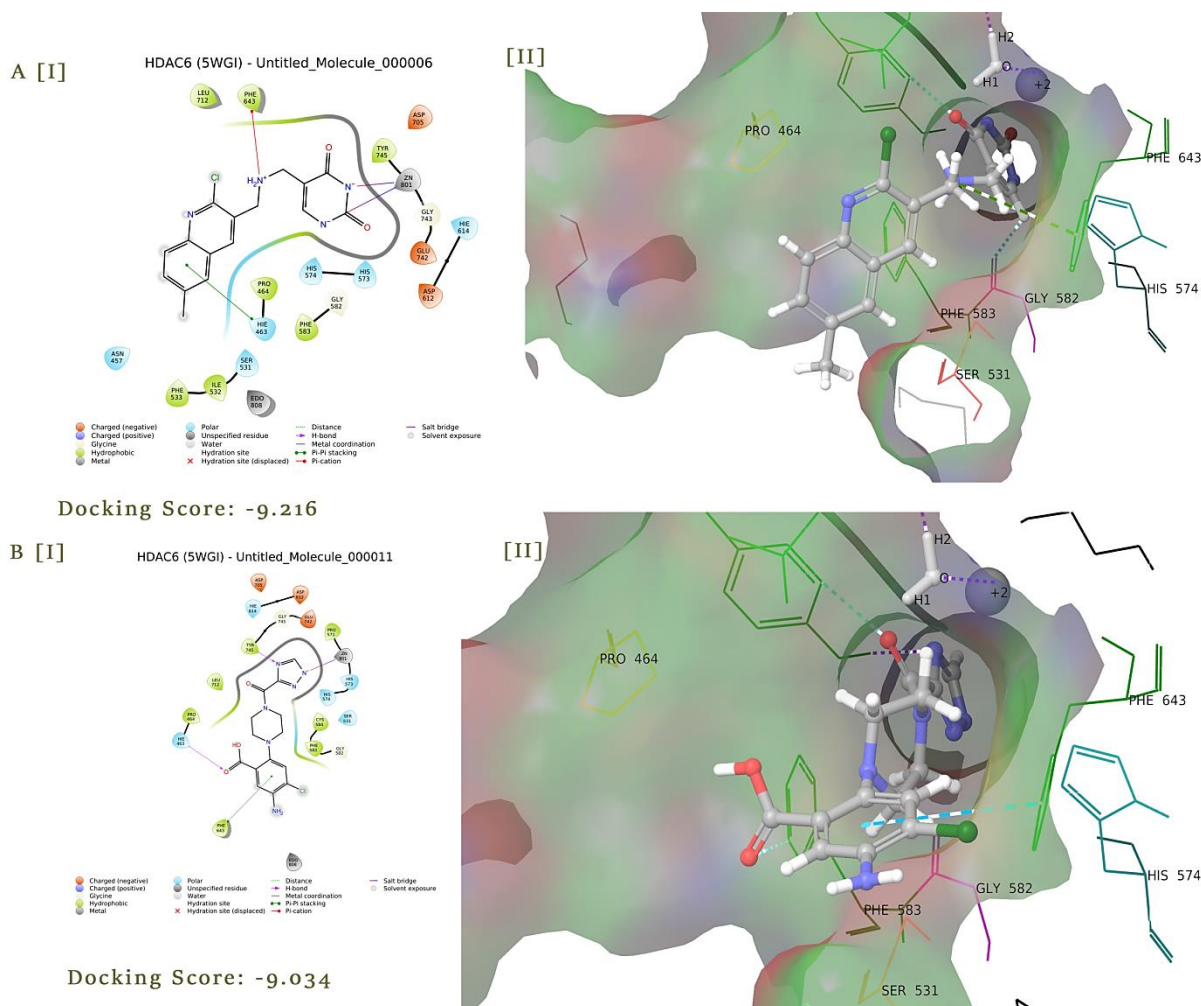


Figure 4: 2D and 3D interaction of top best two *M* sample compounds docked in HDAC6 (CD2) active site: a) compound 6 b) compound 11

3.2. Structure-based drug discovery

3.2.1. High throughput Virtual Screening

Using Schrodinger virtual screening workflow (see methods), the 7785 compounds downloaded from the previous section were docked into the catalytic domain 2 (CD2) of HDAC6. 24 compounds were returned with binding affinity ranging from -78 to -28kcal/mol (Table 2). Compounds 14660440 and 10651887 were the top compounds with the binding affinity of -78.56kcal/mol and -78.21kcal/mol and were thus selected for this study (Figure 5). The docking protocol was validated by redocking the co-crystallized ligand TSA. The redocked pose was superimposed on the crystallized structure; an RMSD score of 1.60 was obtained (Figure S6).

An inhibitor of HDAC6 which has shown very high selectivity (CAY10603)[52] and inhibitory activity for HDAC6 (IC50: 2 pM) was selected as the control for this study. When docked it had a binding affinity of -90.70kcal/mol (Table 3; Figure 6).

Table 3: Molecular docking, Induced fit docking, and Binding affinity

Compound ID	XP-Docking score (kcal/mol)	ΔG_{XP} (kcal/mol)
CAY 10603	-8.494	-90.70
14660440	-10.891	-78.56
10651887	-9.778	-78.21
20558972	-9.408	-69.83
5899732	-9.519	-69.75
2559755	-9.493	-68.47
10640503	-9.493	-68.47
8964976	-10.251	-66.42
5886114	-10.162	-62.07
14742603	-9.593	-61.29
5127713	-9.795	-59.05
19366888	-9.254	-57.35
3646127	-9.318	-56.07
1481504	-9.668	-55.84
686879	-9.429	-50.26
18504893	-9.63	-48.99
10688493	-9.118	-46.05
5814288	-9.664	-45.63
11634999	-10.176	-43.57
3284098	-9.191	-43.43
5127628	-9.604	-67.86
7227152	-10.053	-27.44
3789238	-9.544	-29.39

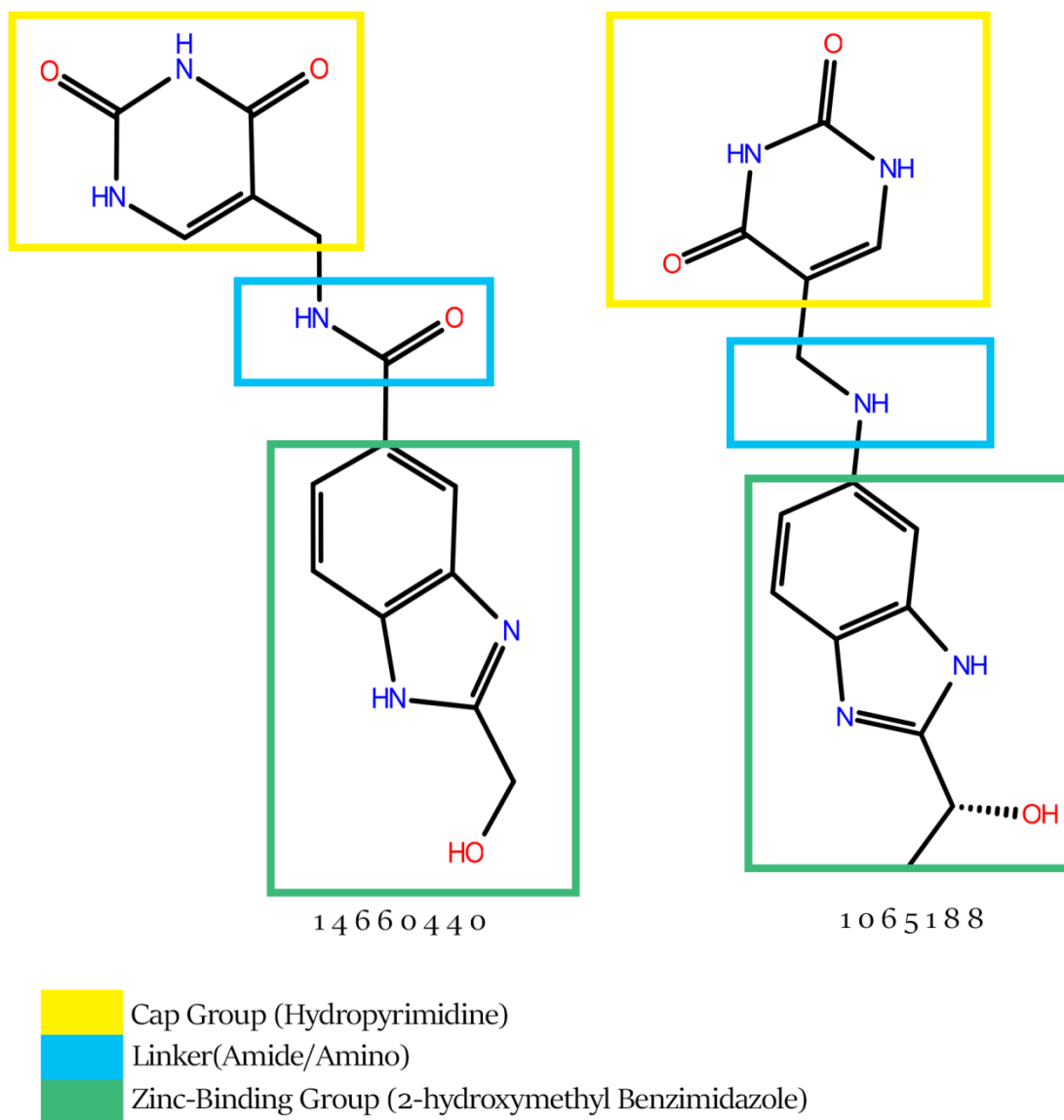


Figure 5: Compounds 14660440 and 1065188

Compound 14660440 formed a metal acceptor bond with the catalytic Zn^{2+} ion and Hydrogen Bond (H-bond) formation with HIS 573, HIS 574 using its hydroxymethyl R group. H-bond with catalytic water molecule via its amide linker and hydropyrimidine cap (Figure 6); HIS614 formed hydrophobic interactions (π - π bond) with the benzimidazole group. PHE 583, PHE643 also formed hydrophobic bonds (π - π , π -sigma) with the ligand; the hydropyrimidine cap formed H-bonds with PHE642 and ASN645 (Figure 6). Compound 10651887 formed metal acceptor bond with Zn^{2+} ion, H-bond with catalytic TRY745, H-bond with ASP612 and π - π interaction with HIS614 (both coordinators of the Zn^{2+} ion), π - π interactions with PHE583 and PHE643, H-bond with GLY582. The hydropyrimidine cap and amine linker formed H-bond with SER531 (Figure 6).

When comparing the binding poses of compound 14660440 and 10651887 with the CAY10603 (standard), the compounds had similar catalytic residue interactions. However, specificity differences are observed; interactions with PRO464 (H-bond, π -alkyl) and HIS462 (π -alkyl) (Figure 6). These two residues (PRO464 and HIS462) along with SER531 are specific only for HDAC6^[54]; interactions with these residues, therefore, suggest high specificity for HDAC6 which is consistent with experimental data which has shown CAY10603 to be highly specific for HDAC6^[31] and also suggest the reason for its high binding affinity (-90.70kcal/mol) observed.

Having identified these compounds, we decided to further subject the compounds to more rigorous and accurate docking protocols. This was in a bid to elucidate a binding hypothesis (i.e. mechanism of inhibition and specificity). We, therefore, subjected the compounds (14660440 and 10651887) to more accurate and exhaustive docking protocols (Induced fit docking and QM-polarized docking) and geometrically optimize the docked complex using QM/MM optimization while comparing with the standard.

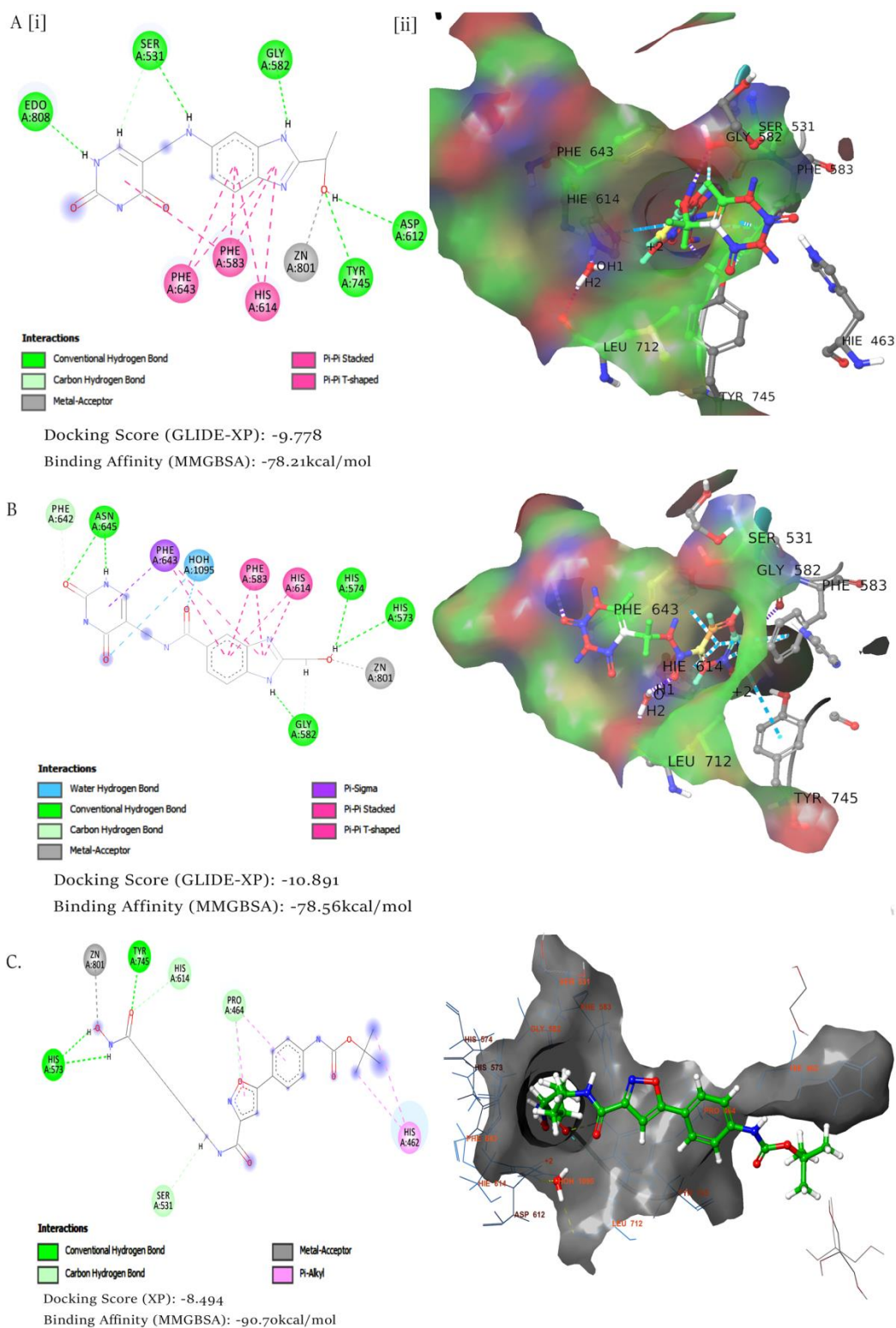


Figure 6: Molecular docking of compound a) 14660440 b) 10651887 c) CAY10603: i) 2D interactions ii) 3D interactions

3.2.2. QM-polarized docking

Accurate charges have been shown to improve docking accuracy,^[53] most docking protocols consider the charges on the ligand via force field minimization, however, charges induced on the ligand within the protein active site is not considered. QM-polarized docking (QPLD) however tries to improve docked ligand binding conformation by considering plausible charge polarization that might occur due to the active site residues.^[54] Using Schrödinger's QPLD module (see methods) QPLD was implemented on compound 14660440, 10651887, and CAY10603 to gain insight into the binding interactions of these compounds and calculate their corresponding binding affinity. The results showed CAY10603 to have the highest binding affinity of -98.55kcal/mol. Compound 10651997: -75.83kcal/mol and compound 14660440: -70.80kcal/mol (Table 4). New H-bond was observed between ASP612 and compound 10651887 other bonds remained when compared with the molecular docking. Compound 14660440 formed a new H-bond with SER531 however, this bond had a bond length of 2.78Å which was rather weak when compared to 2.49Å (Table 7) of compound 10651887 (Figure 7). CAY10603 formed H-bond with the catalytic water molecule, pi-alkyl bonds with LEU712, and alkyl bonds with LEU466 (Figure 7). The results showed that polarized charges induced by the active site residues on the ligands cause a reduction in the binding affinity of compound 10651887 and 14660440 for HDAC6.

Table 4: QPLD docking score and binding affinity

Compound ID	XP docking score(kcal/mol)	QPLD docking score (kcal/mol)	ΔG_{QPLD} (kcal/mol)
CAY 10603 (Standard)	-8.494	-8.810	-98.55
10651887	-9.778	-10.450	-75.83
14660440	-10.891	-10.422	-70.80

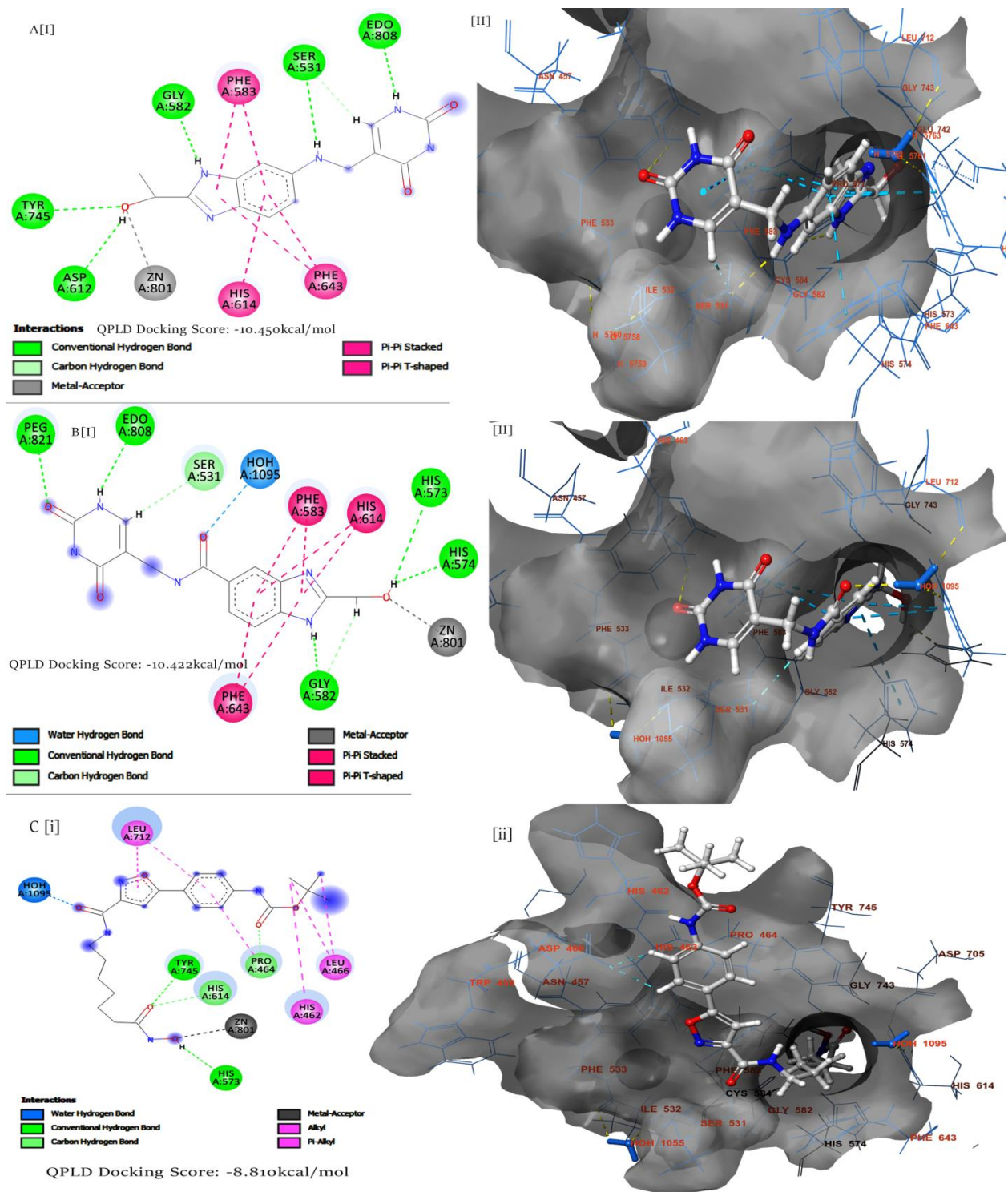


Figure 7: QM-polarized ligand docking of a) compound 10651887 b) compound 14660440 C) CAY10603: [I] 2D interactions [II] 3D interactions

3.2.3. Induced fit docking

Apart from accurate charges, conformational changes in the active site as a result of ligand binding needs to be considered before a reliable binding hypothesis can be elucidated. Induced fit docking (IFD) considers both the ligand and active site residues flexibility in its calculations, this, increases the accuracy of predicted binding interactions and binding affinity.^{[55][56]} We performed IFD (see methods) on compounds 14660440, 1065188, and CAY10603 and calculated binding affinity.

Compound 10651887 had the highest IFD score of -847.34; two new bonds were observed: H-bond with HIS 574 and HIS573, π -Donor H-bond with PRO464, and Carbon H-bond with HIS463 were also observed. However, H-bond formed previously with Asp612 was lost, all other bonds were retained (Figure 8). Binding affinity however increased from -78.21kcal/mol (XP) to -98.29kcal/mol (Table 5) becoming the ligand with the highest binding affinity. Compound 14660440 had an IFD score of -847.20; of note is the formation of a new H-bond was formed with TYR745 all other bonds were retained (Figure 8). The binding affinity also increased to -89.63kcal/mol (Table 5). CAY10603 formed new H-bonds with GLY743 and GLY582, π -sigma bond with PHE643, π - π interaction .with PHE583 and alkyl bonds with LEU466; H-bonds with TYR745, HIS573, and HIS614 were lost (Figure 8). The binding affinity of CAY10603 was however reduced from -90.70kcal/mol to -85.74kcal/mol (Table 5).

Table 5: IFD score and binding affinity

Compound ID	IFD score	ΔG_{IFD} (kcal/mol)
CAY 10603	-839.57	-85.74
14660440	-847.20	-89.63
10651887	-847.34	-98.29

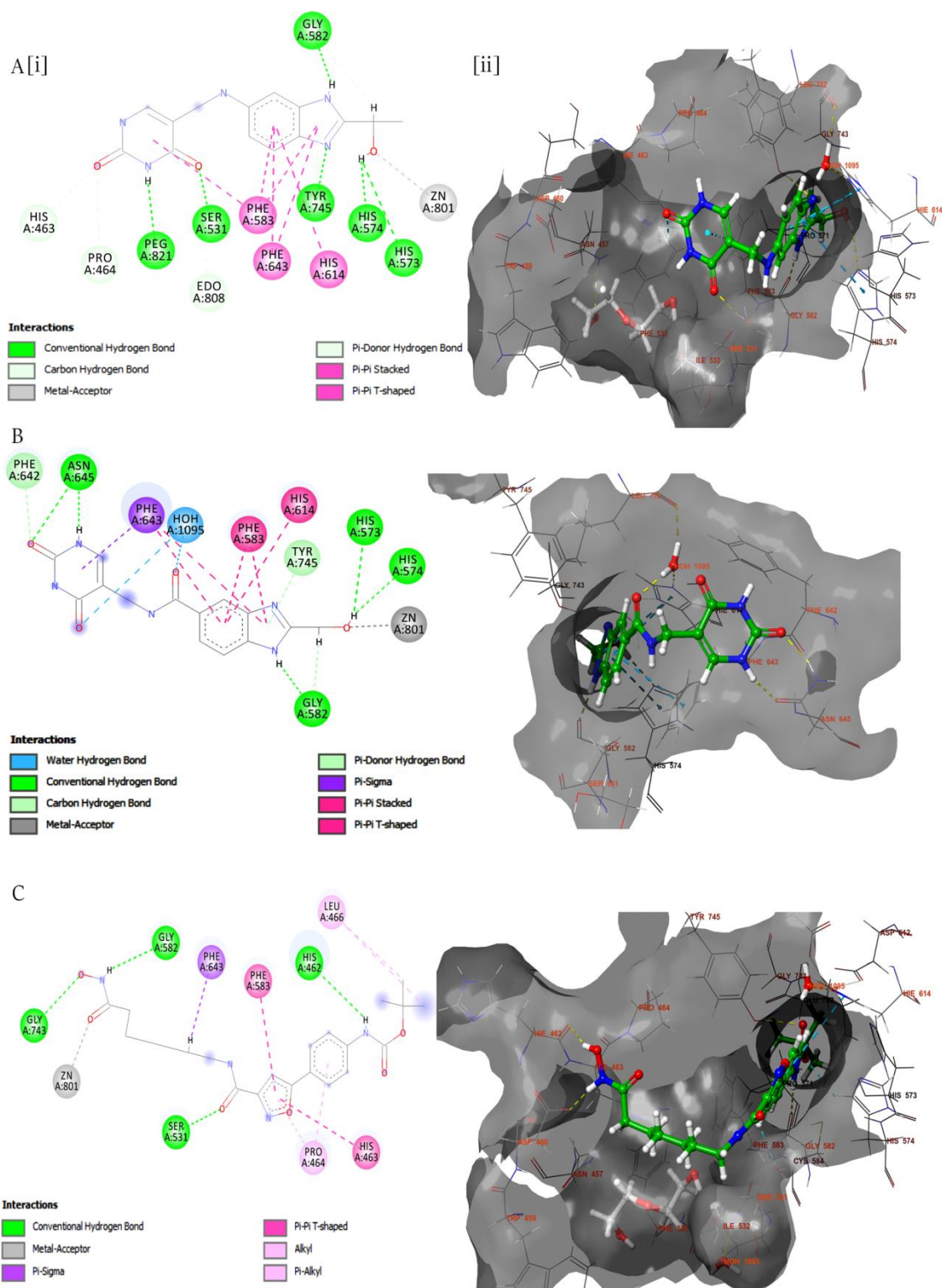


Figure 8: Induced fit docking of compound a) 10651887 b) 14660440 C) CAY10603: i) 2D interaction ii) 3D interaction

3.2.4. QM/MM optimization

The induced-fit protein-ligand complex was optimized using QM/MM geometric optimization in order to validate the binding interactions observed. In QM/MM optimization protocol, we treated the ligand and interacting side chains as the QM region while the whole protein was treated as the MM region (see methods).

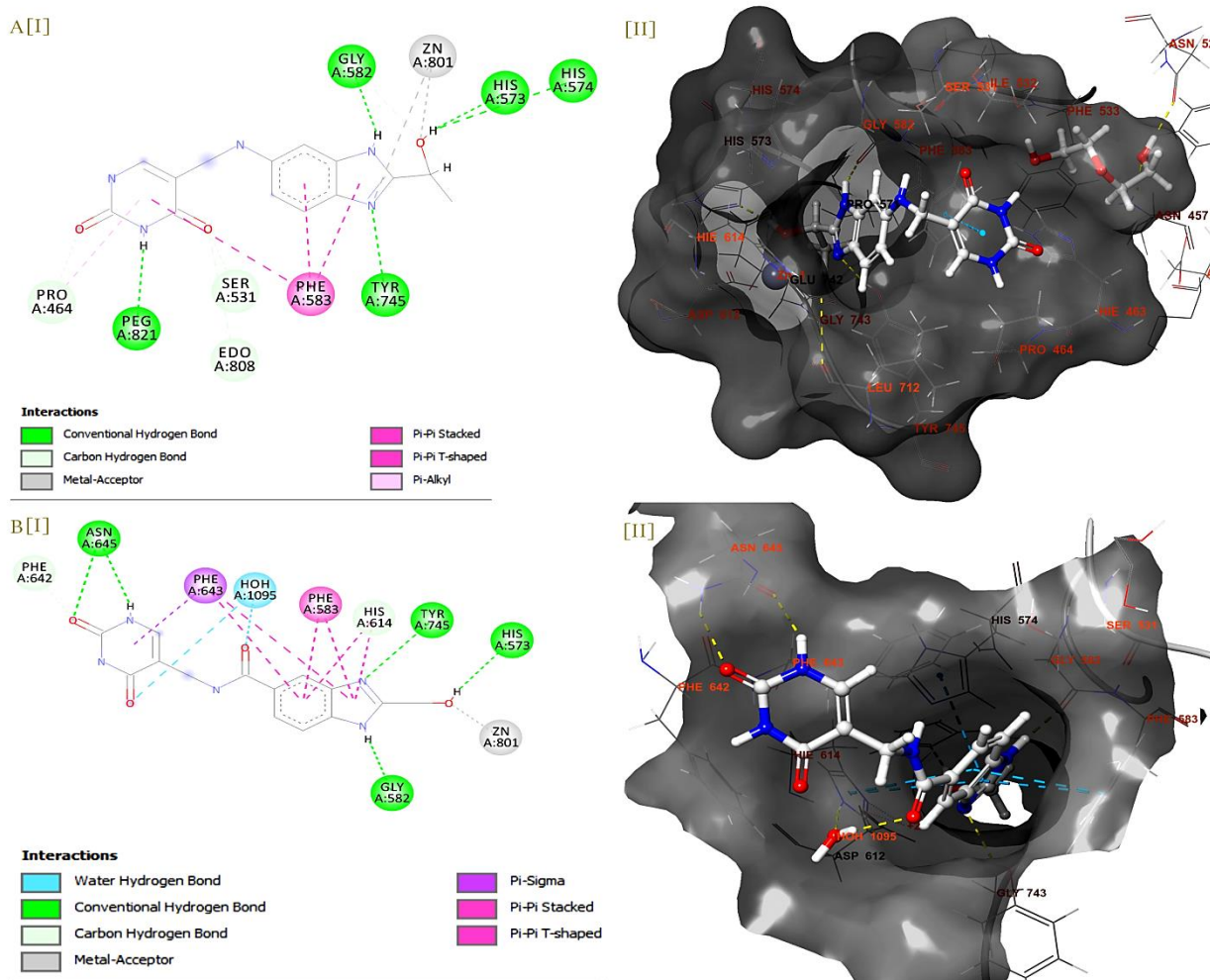


Figure 9: QM/MM optimization of protein-ligand docked complex: a) 10651997 b) 14660440

The QM/MM optimized pose was compared with the QPLD and IFD pose of the compounds. Compound 1466044 lost H-bonds formed with HIS 574 in the previous docking (IFD and QPLD) and reformed H-bonds with TYR745 which was missing in QPLD pose (Figure 9). Examining the bond length of the H-bond formed with HIS 573 and TYR 745 we observed stronger H-bonds upon optimization (Table 7); compound 10651997 retained H-bonds with HIS573, HIS574 which was absent in the QPLD pose (Figure 9). A new H-bond was formed with PRO464 which was not reported in any of the previous docking poses (IFD and QPLD).

The H-bond strength with HIS573 improved upon optimization, however, H-bond with HIS574 was not as stable as previous pose (increase in bond length) (Table 7). H-bond formed with SER531 was retained but not as strong as the previous docking pose (Table 7). H-bond with TYR745 improved when compared with the QPLD pose. The binding affinity of compound 10651887 was -79.62kcal/mol and compound 14660440 was -77.44kcal/mol (Table 6). Over the for binding affinities calculated (Molecular docking, induced-fit, Qm-polarized docking and QM/MM optimization) compound 10651887 had the highest binding affinity when compared to compound 14660440 (table 6)

Table 6: Binding affinity of all docking protocols

Compound ID	ΔG_{XP} (kcal/mol)	ΔG_{IFD} (kcal/mol)	ΔG_{QPLD} (kcal/mol)	$\Delta G_{QM/MM}$ (kcal/mol)
CAY 10603	-90.70	-85.74	-98.55	-
10651887	-78.21	-98.29	-75.83	-79.62
14660440	-78.56	-89.63	-70.80	-77.44

Table 7: Summary of binding interaction of compound 14660440, 10651887, and CAY10603 over various docking protocol and optimization.

Molecular Docking	Induced Fit Docking	QM-polarized Docking	QM/MM Optimization
Compound ID: 14660440			
ZN ²⁺ : Metal-Acceptor HIS ₅₇₃ : H-bond (2.59) HIS ₅₇₄ : H-bond (2.22) GLY ₅₈₂ : H-bond PHE ₅₈₃ : π - π , π - π HIS ₆₁₄ : π - π , π - π PHE ₆₄₂ : H-Bond PHE ₆₄₃ : π - π , π -sigma ASN ₆₄₅ : H-bond, H-bond TYR ₇₄₅ : π - π H ₂ O: H-bond, H-bond (1.90, 2.88)	ZN ²⁺ : Metal-Acceptor HIS ₅₇₃ : H-bond (2.45) HIS ₅₇₄ : H-bond (2.10) GLY ₅₈₂ : H-bond PHE ₅₈₃ : π - π , π - π , π - π HIS ₆₁₄ : π - π PHE ₆₄₃ : π -sigma, π - π , π - π , π - π ASN ₆₄₅ : H-bond, H-bond TYR ₇₄₅ : H-bond (2.49) H ₂ O: H-bond, H-bond (1.84, 2.73)	ZN ²⁺ : Metal-Acceptor SER ₅₃₁ : H-bond (2.78) HIS ₅₇₃ : H-bond (2.69) HIS ₅₇₄ : H-bond (1.95) GLY ₅₈₂ : H-bond PHE ₅₈₃ : π - π , π - π HIS ₆₁₄ : π - π , π - π PHE ₆₄₃ : π - π , π - π , π - π H ₂ O: H-bond (2.07)	ZN ²⁺ : Metal-Acceptor HIS ₅₇₃ : H-bond (1.93) GLY ₅₈₂ : H-bond PHE ₅₈₃ : π - π , π - π HIS ₆₁₄ : π - π , π - π PHE ₆₄₂ : H-bond PHE ₆₄₃ : π -sigma, π - π ASN ₆₄₅ : H-bond, H-bond TYR ₇₄₅ : H-bond (2.04) H ₂ O: H-bond (2.11), H-bond (3.01)
Compound ID: 10651887			
ZN ²⁺ : Metal-Acceptor SER ₅₃₁ : H-bond, H-bond (2.25, 2.52) GLY ₅₈₂ : H-bond PHE ₅₈₃ : π - π , π - π π - π ASP ₆₁₂ : H-bond (2.69) HIS ₆₁₄ : π - π , π - π PHE ₆₄₃ : π - π TYR ₇₄₅ : H-bond (2.32)	ZN ²⁺ : Metal-Acceptor HIS ₄₆₃ : H-bond (2.77) PRO ₄₆₄ : H-bond (2.284) SER ₅₃₁ : H-bond (1.90) GLY ₅₈₂ : H-bond, H-bond HIS ₅₇₃ : H-bond (2.33) HIS ₅₇₄ : H-bond (2.22) PHE ₅₈₃ : π - π , π - π , π - π HIS ₆₁₄ : π - π PHE ₆₄₃ : π - π , π - π TYR ₇₄₅ : H-bond (1.84)	ZN ²⁺ : Metal-Acceptor SER ₅₃₁ : H-bond, H-bond (2.32, 2.32) GLY ₅₈₂ : H-bond PHE ₅₈₃ : π - π , π - π ASP ₆₁₂ : H-bond (2.26) HIS ₆₁₄ : π - π PHE ₆₄₃ : π - π , π - π TYR ₇₄₅ : H-bond (2.32)	ZN ²⁺ : Metal-Acceptor, Metal-Acceptor SER ₅₃₁ : H-bond (2.71) GLY ₅₈₂ : H-bond, H-bond PHE ₅₈₃ : π - π , π - π , π - π HIS ₅₇₃ : H-bond (1.87) HIS ₅₇₄ : H-bond (2.94) PRO ₄₆₄ : π -alkyl, H-bond (5.43) TYR ₇₄₅ : H-bond (2.0)
Compound ID: CAY10603			
ZN ²⁺ : Metal-Acceptor HIS ₄₆₂ : π -alkyl, π -alkyl PRO ₄₆₄ : π -alkyl, π -alkyl, H-bond (3.07) SER ₅₃₁ : H-bond (2.70)	ZN ²⁺ : Metal-Acceptor HIS ₄₆₂ : H-bond (2.81) HIS ₄₆₃ : π - π LUE ₄₆₆ : Alkyl, Alkyl	ZN ²⁺ : Metal-Acceptor HIS ₄₆₂ : π -Alkyl PRO ₄₆₄ : π -Alkyl LEU ₄₆₆ : π -Alkyl, π -Alkyl	

HIS ₅₇₃ : H-bond, H-bond (1.95, 3.06)	SER ₅₃₁ : H-bond (1.86)	HIS ₅₇₃ : H-bond (1.86)
HIS ₆₁₄ : H-bond (3.05)	GLY ₅₈₂ : H-bond PHE ₅₈₃ : π - π GLY ₇₄₃ : H-bond PRO ₄₆₄ : π -alkyl, π -alkyl	HIS ₆₁₄ : H-bond (3.01) LEU ₇₁₂ : π -Alkyl, π -Alkyl TYR ₇₄₅ : H-bond (1.74)

3.3. Molecular Dynamics

Using Normal mode analysis (NMA) the stability of the induce-fit docked complex, and ease of protein structure deformability was investigated. Using the elastic network model (ENM)[57] method, the degree of deformability of each residue (figure 10i), B-factor obtained from PDB field and NMA mobility data (figure 10ii), eigenvalues representing ease of deformant (figure 10iii), covariance map (figure S7), and elastic network (figure S8) for the docked complex was calculated.

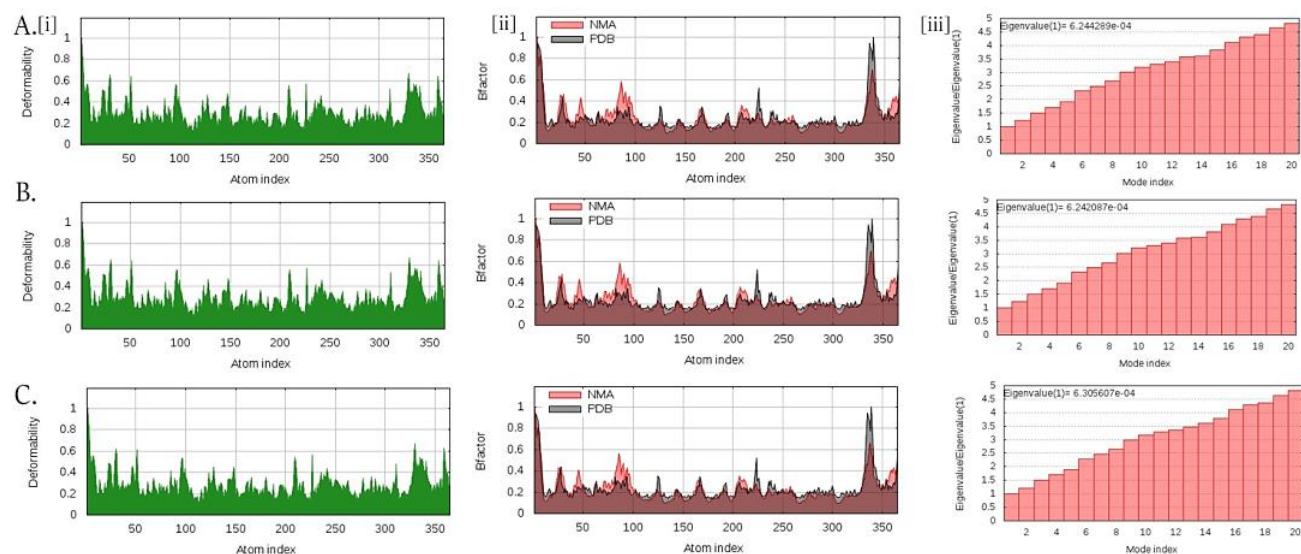


Figure 10: NMA analysis of a) 10651887 b) 14660440 C) CAY10603: i) deformability ii) B-factor iii) eigenvalues

The molecular dynamic simulation (NMA) of the complexes showed compound 14660440 to have an eigenvalue of 6.244289e-04, compound 10651887: 6.242087e-04, and compound CAY10603: 6.305607e-04. All the complexes showed similar eigenvalues however, compound 14660440 and 10651887 showed lower values.

3.4. Lead optimization

Bioisosteres are chemical moieties that are used to replace functional groups in compounds while retaining biological activity; bioisostere replacement is an established protocol for lead generation and optimization.^{[58][59]} Since benzimidazole has been suggested as a potential zinc-binding group^[60] and hydroypyrimidine as an established capping group,^[26] we, therefore, held

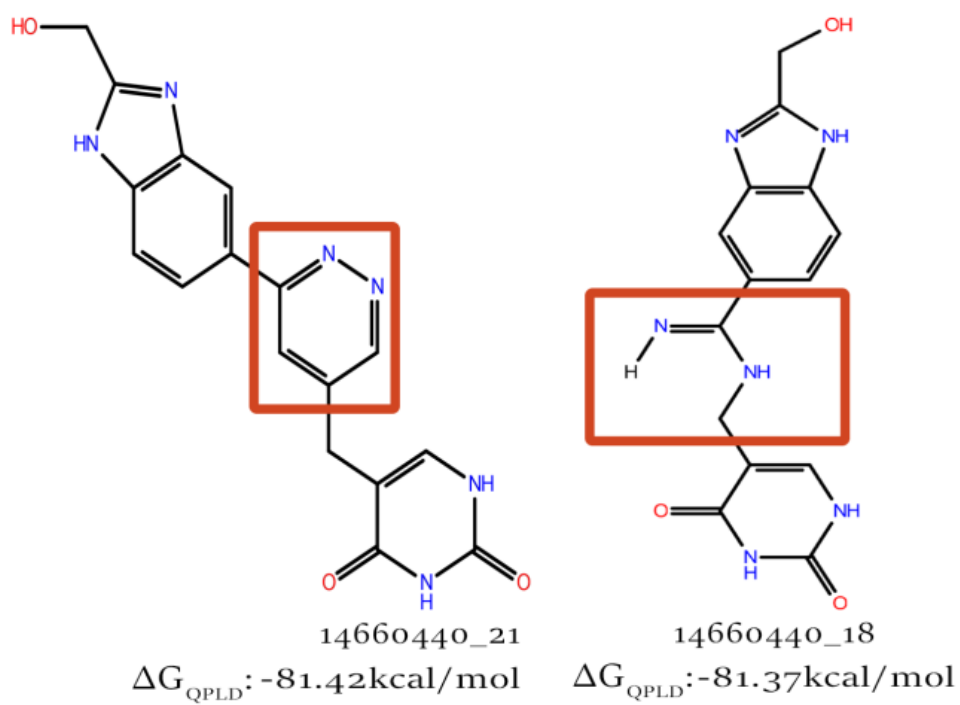
both regions immutable during the bioisostere replacement analysis; we sought to see if the binding affinity/binding affinity and specificity of the lead compounds would improve by substituting or modifying the linkers.

A total of 24 bioisostere replacement for compounds 1466044 were generated, while no suitable replacement was found for compound 10651887. These newly generated compounds were subjected to Molecular docking and QM-polarized docking, corresponding binding affinity was calculated. Of the 24 compounds generated 7 had QPLD binding affinity ranging between -81 to -70kcal/mol (Table 8). The bioisostere replacement protocol generally modified the amide linker of compound 1466044 with pyridazine and imine functional groups. Compounds 14660440_21 which had the highest QPLD binding affinity of -81.42kcal/mol (Table 8) had its amide linker replaced with pyridazine functional group which resulted in the formation of a π - π bond with LEY712 (Figure 11, 12), while compound 14660440_18 had its amide linker modified to an imine functional group forming carboximidamide which resulted in the formation of pi-cation interaction with PHE 643 and H-bound with SER 531 (Figure 11, 12). These replacements increased the QPLD binding affinity of un-optimized 14660440 from -70.80 to -81.42kcal/mol (Table 8).

Table 8: Compound 14660440 Bioisostere replaced compounds: docking score, binding affinity, and electronic properties.

Compound ID	XP-docking Score (kcal/mol)	ΔG_{XP} (kcal/mol)	QPLD-docking score (kcal/mol)	ΔG_{QPLD} (kcal/mol)
CAY10603 (Standard)	-8.494	-90.70	-8.810	-98.55
14660440_21	-10.445	-85.45	-10.281	-81.42
14660440_18	-10.346	-80.43	-9.827	-81.37
14660440_13	-9.992	-80.35	-10.211	-67.92
14660440_19	-9.236	-79.74	-9.434	-51.19
14660440_24	-9.612	-79.55	-10.043	-73.46
14660440_3	-10.935	-78.84	-9.866	-67.85
14660440_14	-9.77	-78.46	-9.835	-71.39
14660440_23	-9.574	-78.43	-9.382	-78.96
14660440_16	-10.018	-78.32	-9.305	-68.45
14660440_5	-9.933	-78.19	-9.181	-71.42
14660440_11	-9.827	-78.12	-9.045	-70.98

The specificity for HDAC6 also improved; the pyridazine linker replacement in 14660440_21 resulted in the formation of bonds with residue-specific to HDAC6: pi-alkyl bond was formed with LEU712 and H-bond was formed with HIS463 (Figure 12). The amine ion linker in 14660440_18 resulted in the formation of H-bond with HIS463 and SER531 (Figure 12).



Bioisostere Replacement

Figure 11: Top optimized compounds: 14660440_21 and 14660440_18.

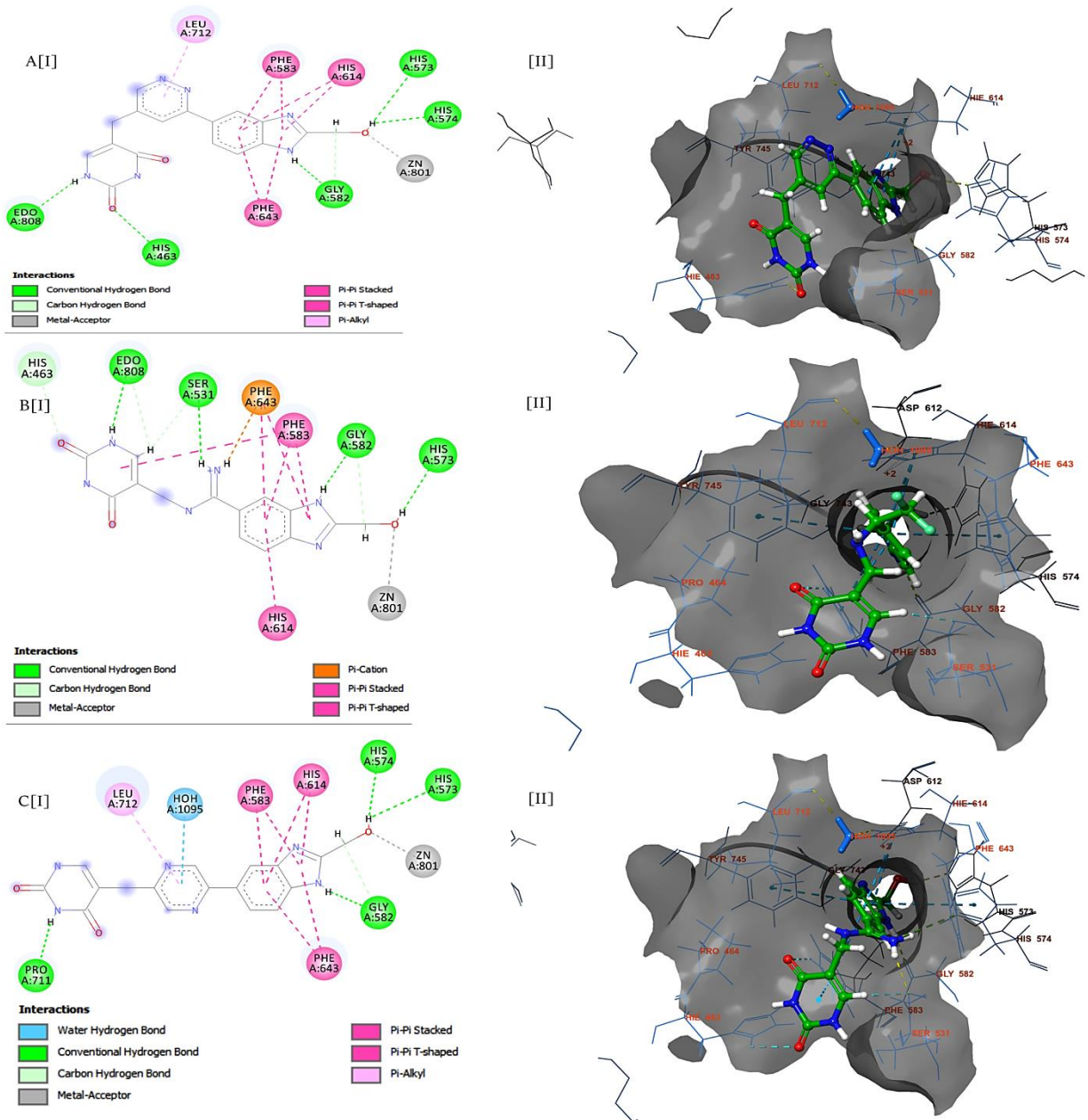


Figure 12: Compound 14660440 Bioisostere replacement compounds

3.5. Electronic Properties

Determining molecular orbitals and their properties help in understanding chemical reactivity of organic compounds; of these orbitals, Highest Occupied Molecular orbitals (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) are the most important.^[61] The HOMO are low energy regions of a compound that readily donates electrons (nucleophile) while LUMO are high energy regions in a compound which are deficient in electrons, hence, readily accept electrons (electrophile); the LUMO is, therefore, the most reactive part of the compound.

The HOMO and LUMO of compounds 14660440_21 and 14660440_18 are visualized in figure 13. The figure shows HOMO of 14660440_21 to be around atoms of the pyridazine linker and LUMO to be around atoms of the hydroxypyrimidine cap. HOMO of 14660440_18 is present around the benzimidazole zinc-binding pharmacophore; the LUMO is present around the imine linker. This, therefore, suggests the most reactive moiety of 14660440_21 is the hydroxypyrimidine cap while 14660440_18 is the imine linker (Figure 13).

From the calculated HOMO and LUMO energies different descriptors can be extrapolated; the HOMO-LUMO gap represents the difference between LUMO and HOMO energies, the gap indicates the ease by which electrons would move from HOMO to LUMO. A smaller gap indicates an easy movement of electrons which in turn indicates higher reactivity of the compounds. For both compounds 14660440_21 and 14660440_18 HOMO-LUMO gap was 0.16 (Table 9).

Other descriptors that can be extrapolated from HOMO and LUMO energies are ionization energy (I), electron affinity (A), global hardness (η), chemical potential (μ), and global electrophilicity power of a ligand (ω) (see methods).^[62] Extremely hard compounds are weak electron acceptors, while a compound with lower chemical potential indicates good electrophiles the global electrophilicity power descriptor indicates the general reactivity of compounds.^[62] Compound 14660440_18 had the highest global electrophilicity power of 0.12, electron affinity: 0.19, and chemical potential: -0.27 (Table 9) making it the most reactive compound.

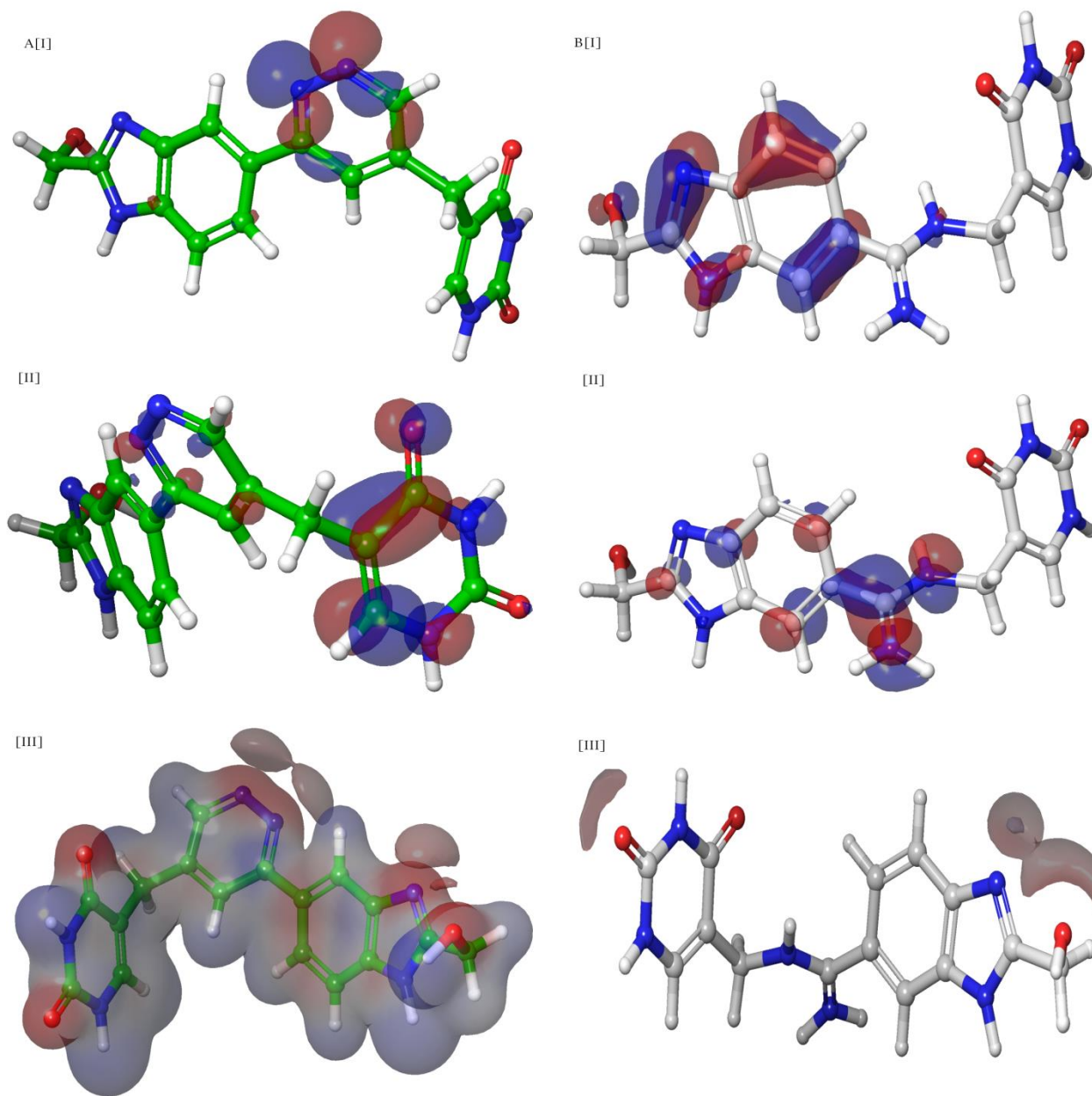


Figure 13: Electronic Properties of Compound a) 14660440_21 b) 14660440_18: i) HOMO ii) LUMO iii) MESP

Table 9: Electronic Descriptors of optimized 14660440 compounds

Compound ID	ΔG_{QPLD} (kcal/mol)	HOMO Energy (eV)	LUMO Energy (eV)	HOMO-LUMO GAP	Ionization energy (I)	Electron affinity (A)	Chemical potential (μ)	Global Hardness (η)	Global Electropositivity (ω)
14660440_21	-81.42	-0.22	-0.06	0.16	0.22	0.06	-0.14	0.08	0.12
14660440_18	-81.37	-0.35	-0.19	0.16	0.35	0.19	-0.27	0.08	0.25
14660440_13	-67.92	-0.23	-0.04	0.19	0.23	0.04	-0.14	0.10	0.10
14660440_19	-51.19	-0.14	-0.09	0.05	0.14	0.09	-0.12	0.03	0.24
14660440_24	-73.46	-0.23	-0.06	0.17	0.23	0.06	-0.15	0.09	0.13
14660440_3	-67.85	-0.34	-0.18	0.16	0.34	0.18	-0.18	0.08	0.20
14660440_14	-71.39	-0.22	-0.04	0.18	0.22	0.04	-0.13	0.09	0.09
14660440_23	-78.96	-0.22	-0.05	0.17	0.22	0.05	-0.14	0.09	0.11
14660440_16	-68.45	-0.21	-0.04	0.17	0.21	0.04	-0.13	0.09	0.09
14660440_5	-71.42	-0.23	-0.05	0.18	0.23	0.05	-0.14	0.09	0.11
14660440_11	-70.98	-0.21	-0.21	0	0.21	0.21	-0.21	0	0

3.5.1. Molecular electrostatic potential

Electrostatic potential explains the work done in moving a unit positive charge from infinity to a point, thus molecular electrostatic potential (MESP) is defined as the energy interaction between a positive charge and a molecule.^[62] MESP descriptor helps to identify regions in molecules that may be involved in electrophilic or nucleophilic attack thereby providing insights into reactivity of an organic compound.^[63] MESP provides a visual representation of positive and negative charge distribution in a molecule. Figure 13(a[iii], b[iii]) show MESP distribution of 14660440_21 and 14660440_18 respectively (Red: negatively charged, White: neutral, Blue: positively). Compound 14660440_21 was generally positive, with negative charges distributed over oxygen atoms on the hydroxymethyl R group, and nitrogen atoms on benzimidazole (Figure 13a[iii]). Compound 14660440_18 however was predicted a cation with

positive charges present over oxygen atoms on the hydropyrimidine cap and hydroxymethyl group on benzimidazole (Figure 13b[iii]).

3.6. ADMET Predictions

Having predicted the binding conformation, affinity, and mechanics of chemical reactivity, ADMET properties of the compounds were determined Using the discovery studio ADEMT descriptors calculator module (see methods). Lipinski drug-likeness and opea lead-likeness descriptors were calculated using the MOE descriptor calculator (see methods).

All compounds except compound CAY10603 (Standard) passed the Lipinski drug-likeness test.^[63] CAY10603 and 14660440_18 failed opea leadlike test^[64] (Table 10). Aques absorption prediction, predicts the solubility of the compounds in water, which in turn determine oral absorption. The aqueous absorption ranged from good to optimal (Table 10). Intestinal absorption predictions predict the concentration that would be absorbed into the bloodstream. The predictions showed that intestinal absorption of CAY10603, 14660440, 14660440_5, and 14660440_18 might low (Figure 14; Table 10). Only compound 14660440_14 was predicted to be able to penetrate the blood-brain barrier (Figure 14). All other compounds (except CAY10603, 14660440, 14660440_5, and 14660440_18) where within 99% confidence ellipses of intestinal absorption i.e. 99% intestinal absorption is likely (Figure 14).

Inhibition of cytochrome P450 2D6 (CYP2D6) by moieties of drugs is one of the major causes of drug-drug interaction^[65]. Therefore, ensuring that this does not occur is an important aspect of drug optimization and design. None of the compounds were predicted to inhibit CYP2D6 (Table 10). Plasma protein binding (PPB) of the drug were also predicted. Plasma protein binding prediction is of utmost importance as the binding of compounds to plasma protein shields the compounds from metabolism and reduces their pharmacological efficacy.^[66] The prediction predicts whether 90% (and above) concentration of the compounds would exhibit PPB. Only Compound CAY10603 was predicted to exhibit plasma protein binding (Table 10). Mutagenicity of the compounds was also predicted. Only compound CAY10603 was predicted to be mutagenic (Table 10).

Table 10: ADMET Descriptor predictions

Compound ID	ΔG_{QPLD}	Lipinski Druglike	Mutagenic	Oprea Leadlike	Aqueous Solubility	Cytochrome P450 2D6 (CYP2D6) Inhibition	Plasma Protein Binding	Human Intestinal Absorption
CAY10603	-90.6993	True	True	False	Good	False	True	Low
14660440	-70.801	True	False	True	Optimal	False	False	Low
10651887	-75.832	True	False	True	Optimal	False	False	High
14660440_21	-81.4237	True	False	True	Good	False	False	Moderate
14660440_13	-67.923	True	False	True	Good	False	False	High
14660440_18	-81.3747	True	False	False	Optimal	False	False	Low
14660440_24	-73.4641	True	False	True	Good	False	False	Moderate
14660440_16	-78.5321	True	False	True	Good	False	False	Moderate
14660440_3	-67.8505	True	False	True	Good	False	False	High
14660440_14	-71.3854	True	False	True	Good	False	False	High
14660440_19	-51.1944	True	False	True	Optimal	False	False	Moderate
14660440_23	-78.963	True	False	True	Good	False	False	Moderate
14660440_16	-68.4506	True	False	True	Good	False	False	Moderate
14660440_5	-71.4231	True	False	True	Good	False	False	Low

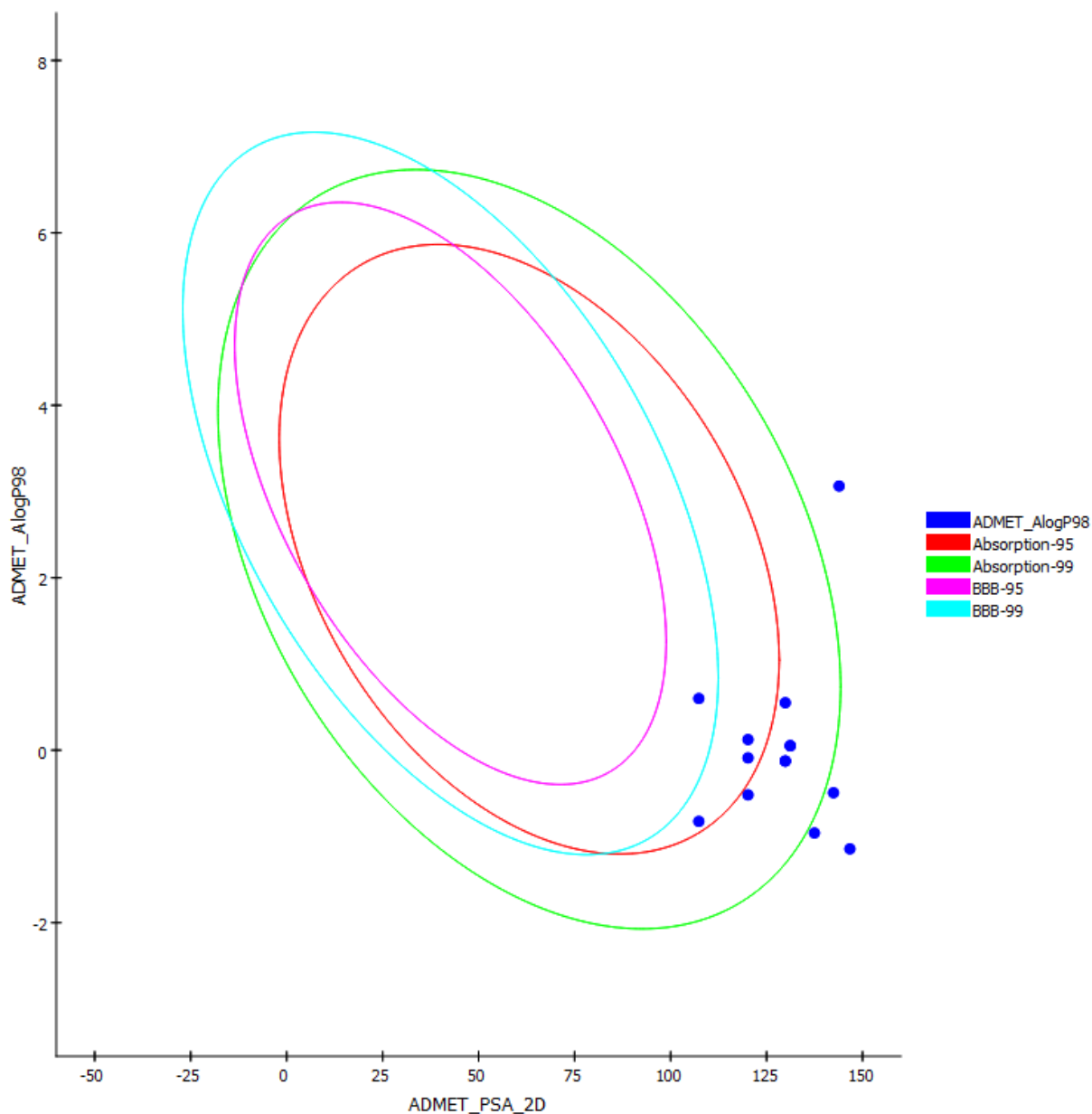


Figure 14: Human intestinal absorption and Blood-brain barrier penetration prediction plot: The ellipse describes the different regions in which well-absorbed compounds are expected. 95% of well-absorbed compounds are expected in the 95% ellipse and 99% of well-absorbed compounds are expected in the 99% ellipse.

4.0. Discussion

Developing specific inhibitors for HDAC6 is still a daunting task due to the highly conserved nature of HDAC family protein. Based on literatures,^{[26][67]} the formation of non-covalent bonds with the following residues is key in discussing the putative mechanism of inhibition and specificity for compounds in this study:

- Catalytic residues: TYR 745; HIS 573; HIS 574; ZN²⁺; ASP 612; HIS 614; Catalytic H₂O.
- Specificity residues: SER 531; HIS 462; HIS 463; PRO 464; LEU 712.

In this study, an RFC model was built on all know HDAC6 inhibitors (3742) deposited in ChEMBL (2003 to date). The model was validated on an external dataset using a 10 fold cross-validation, it had a balanced accuracy of 0.78, precision 0.79, recall 0.90, and AUC score of 0.768. The model was used to predict active compounds in the M sample representation of SCUBIDDOO; the predicted active compounds were docked and two compounds showing high docking score and fragment of the compounds having chelating interaction with ZN²⁺ were identified and used to search the SCUBIDDOO database. A total of 7785 compounds having this fragment were downloaded and virtually screened, 24 compounds resulted and compound 14660440 and 10651887 was selected because they had the highest docking binding affinity of -78.56kcal/mol and -78.21kcal respectively.

The two compounds were similar (RSMD: 0.61) in structure differing majorly in their linker moiety. They possessed a benzimidazole moiety with a 2-hydroxymethyl R group as their zinc-binding group, a hydroxypyrimidine as its capping group, an amide linker (14660440), and an amino linker (10651887). Recently Leandro *et al*^[60] (using Fragment-based drug discovery) suggested benzimidazole functional group as a potential zinc-binding group. With this as proof of concept, we investigated a potential binding hypothesis using Induced-fit (IFD), QM-polarized docking (QPLD), and QM/MM optimization. The hypothesis suggests a possible mechanism of inhibition and specificity of the compounds by identifying consistent interaction. The criteria for selection were these:

- A binding interaction must have been observed consistently in Molecular docking, QPLD, IFD, and QM-MM predicted binding pose;
- Any new interaction observed in the IFD docking pose and consistent in QM-MM optimized pose is also selected.

Based on these above criteria compound 10651887 with a binding affinity of -75.83kcal/mol (QPLD) might inhibit HDAC6 by interacting with: HIS 573, HIS574, TYR 745, and ZN²⁺, and show specificity by interacting with SER531 and PRO464. Compound 14660440 with a binding affinity of -70.80kcal/mol (QPLD), interactions with catalytic HIS573, HIS614, TYR745, H₂O, and ZN²⁺ might be its mechanism of inhibition. However, compound 14660440 did not interact with specificity residues. Considering the lack of specificity of compound 14660440 and the low

binding affinities when compared with the standard (QPLD: -98.55kcal/mol), we sought to optimize the compounds using bioisostere replacement. No suitable bioisostere was found for compound 10651887, however, 24 new compounds were generated for compound 14660440. After molecular docking and QPLD the top two compounds were 14660440_21 (QPLD: -81.42kcal/mol) and 14660440_18 (QPLD: -81.37kcal/mol) these two compounds showed a fold increase in binding affinity when compared with their parent compound (14660440: -79.80kcal/mol). Compound 14660440_21 had its amide linker replaced with a pyridazine functional group, this resulted in the formation of a π - π bond with LEY712 (specificity residue), and H-bond interactions with HIS463 (specificity residue) via its capping group (hydroypyrimidine). Compound 14660440_18 had its amide linker modified with an imine functional group forming carboximidamide which resulted in the formation of H-bond with SER 531, and H-bond with HIS463 via its capping group (hydroypyrimidine). Using Bioisostere replacement protocol we were successfully able to optimize and generate compounds with improved binding affinity and specificity.

Electronic descriptors were calculated to investigate the mode of the chemical reaction (nucleophilic or electrophilic), stability, and reactivity of the optimized compounds. Both compound 14660440_21 and 14660440_18 were shown to be very reactive and relatively stable (HOMO-LUMO gap: 0.16); compound 14660440_18 was predicted to be the most reactive (highest global electrophilicity power of 0.12). Based on HOMO-LUMO analysis the hydroypyrimidine cap of 14660440_21 was predicted to be likely involved in electrophilic substitution while for 14660440_18, the imine linker was predicted to be the electrophilic region. The MESP analysis showed the hydroypyrimidine cap, hydroxymethyl R group, and nitrogen atoms on benzimidazole to positively charge, suggesting that they might undergo electrophilic reactions. Compound 14660440_18 was predicted to be a cation suggesting that mode of reaction is likely to be purely electrophilic. ADMET predictions showed compound 14660440 and 10651887, with the optimized compounds to pass the drug-likeness test with no predicted mutagenicity, drug-drug interactions, and protein plasma binding. Oral absorption was predicted to be good and moderate for all the compounds, however, compound 14660440, 14660440_18, and 14660440_5 showed low intestinal absorption. Of note is that CAY10603 (Standard) was predicted to be a mutagenic and failed opera lead likeness test.

Having considered the binding interactions, electronic descriptors, and ADMET properties we suggest the following compounds as putative isospecific inhibitors of HDAC6: 10651887, 14660440_21, and 14660440_18. We consider these results to be predictive hence there is a need for experimental validations.

5.0. Conclusion

In this study, we have been able to build a random forest classifier model based on 3,742 inhibitors of HDAC6 developed from 2003 to date. The model had a 78% balanced accuracy after a 10 fold cross-validation on an external dataset; the model was used to screen the SCUBIDOO database. Two compounds were identified after virtual screening and were optimized for improved binding affinity and specificity. Electronic descriptors determined the mechanism of chemical reactivity and stability of the optimized compounds. Finally, ADMET properties were calculated with most of the compounds showing better drug properties than the standard.

Supplementary materials

All supplementary materials are available in MendelyData at DOI: 10.17632/775s3xrhrk.1

- Jupyter notebooks and Python scripts
- Supplementary Tables and Figures
- SMILES of lead compounds and Bioisostere replacement optimized compounds
- Complete Data (Raw and Processed)

Acknowledgment

We acknowledge the expert opinion of Idris Oluwaseun Mukhtar, George Oche Ambrose, and Alakanse Suleiman Oluwaseun.

Reference

- [1] Wu D, Huang Q, Zhang Y, Zhang Q, Liu Q, Gao J, et al. Screening of selective histone deacetylase inhibitors by proteochemometric modeling. *BMC Bioinformatics* 2012;13. <https://doi.org/10.1186/1471-2105-13-212>.
- [2] Januar V, Saffery R, Ryan J. Epigenetics and depressive disorders: A review of current progress and future directions. *Int J Epidemiol* 2015;44:1364–87. <https://doi.org/10.1093/ije/dyu273>.
- [3] Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res* 2011;21:381–95. <https://doi.org/10.1038/cr.2011.22>.
- [4] Gryder BE, Sodji QH, Oyelere AK. Targeted cancer therapy: Giving histone deacetylase inhibitors all they need to succeed. *Future Med Chem* 2012;4:505–24. <https://doi.org/10.4155/fmc.12.3>.
- [5] Hubbert C, Guardiola A, Shao R, Kawaguchi Y, Ito A, Nixon A, et al. HDAC6 is a microtubule-associated deacetylase. *Nature* 2002;417:455–8. <https://doi.org/10.1038/417455a>.
- [6] Stimson L, La Thangue NB. Biomarkers for predicting clinical responses to HDAC inhibitors. *Cancer Lett* 2009;280:177–83. <https://doi.org/10.1016/j.canlet.2009.03.016>.
- [7] Anne M, Sammartino D, Barginear MF, Budman D. Profile of panobinostat and its potential for treatment in solid tumors: An update. *Onco Targets Ther* 2013;6:1613–24. <https://doi.org/10.2147/OTT.S30773>.
- [8] Chen J, Zhang J, Shaik NF, Yi B, Wei X, Yang XF, et al. The histone deacetylase inhibitor tubacin mitigates endothelial dysfunction by up-regulating the expression of endothelial nitric oxide synthase. *J Biol Chem* 2019;294:19565–76. <https://doi.org/10.1074/jbc.RA119.011317>.
- [9] Zhao Z, Shilatifard A. Epigenetic modifications of histones in cancer. *Genome Biol* 2019;20:1–16. <https://doi.org/10.1186/s13059-019-1870-5>.
- [10] Zang LL, Wang XJ, Li XB, Wang SQ, Xu WR, Xie X Bin, et al. SAHA-based novel HDAC inhibitor design by core hopping method. *J Mol Graph Model* 2014;54:10–8. <https://doi.org/10.1016/j.jmgm.2014.08.005>.
- [11] Bora-Tatar G, Dayangaç-Erden D, Demir AS, Dalkara S, Yelekçi K, Erdem-Yurter H. Molecular modifications on carboxylic acid derivatives as potent histone deacetylase inhibitors: Activity and docking studies. *Bioorganic Med Chem* 2009;17:5219–28. <https://doi.org/10.1016/j.bmc.2009.05.042>.
- [12] Grozinger CM, Schreiber SL. Deacetylase enzymes: Biological functions and the use of small-molecule inhibitors. *Chem Biol* 2002;9:3–16. [https://doi.org/10.1016/S1074-5521\(02\)00092-3](https://doi.org/10.1016/S1074-5521(02)00092-3).
- [13] Karagiannis TC, El-Osta A. Will broad-spectrum histone deacetylase inhibitors be superseded by more specific compounds? *Leukemia* 2007;21:61–5.

<https://doi.org/10.1038/sj.leu.2404464>.

- [14] Zhang L, Han Y, Jiang Q, Wang C, Chen X, Li X, et al. Trend of Histone Deacetylase Inhibitors in Cancer Therapy: Isoform Selectivity or Multitargeted Strategy. *Med Res Rev* 2014;1–21. <https://doi.org/10.1002/med.21320>.
- [15] Li Y, Shin D, Kwon SH. Histone deacetylase 6 plays a role as a distinct regulator of diverse cellular processes. *FEBS J* 2013;280:775–93. <https://doi.org/10.1111/febs.12079>.
- [16] Namdar M, Perez G, Ngo L, Marks PA. Selective inhibition of histone deacetylase 6 (HDAC6) induces DNA damage and sensitizes transformed cells to anticancer agents. *Proc Natl Acad Sci U S A* 2010;107:20003–8. <https://doi.org/10.1073/pnas.1013754107>.
- [17] Sakamoto KM, Aldana-Masangkay GI. The role of HDAC6 in cancer. *J Biomed Biotechnol* 2011;2011. <https://doi.org/10.1155/2011/875824>.
- [18] Amin SA, Adhikari N, Jha T. Structure-activity relationships of HDAC8 inhibitors: Non-hydroxamates as anticancer agents. *Pharmacol Res* 2018;131:128–42. <https://doi.org/10.1016/j.phrs.2018.03.001>.
- [19] Wang XX, Wan RZ, Liu ZP. Recent advances in the discovery of potent and selective HDAC6 inhibitors. *Eur J Med Chem* 2018;143:1406–18. <https://doi.org/10.1016/j.ejmech.2017.10.040>.
- [20] Yang P. HDAC6: Physiological function and its selective inhibitors for cancer treatment. *Drug Discov Ther* 2013;7:233–42. <https://doi.org/10.5582/ddt.2013.v7.6.233>.
- [21] Li T, Zhang C, Hassan S, Liu X, Song F, Chen K, et al. Histone deacetylase 6 in cancer. *J Hematol Oncol* 2018;11:1–10. <https://doi.org/10.1186/s13045-018-0654-9>.
- [22] Minucci S, Pelicci PG. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat Rev Cancer* 2006;6:38–51. <https://doi.org/10.1038/nrc1779>.
- [23] Ojha R, Huang HL, HuangFu WC, Wu YW, Nepali K, Lai MJ, et al. 1-Aroylindoline-hydroxamic acids as anticancer agents, inhibitors of HSP90 and HDAC. *Eur J Med Chem* 2018;150:667–77. <https://doi.org/10.1016/j.ejmech.2018.03.006>.
- [24] Hsu KC, Liu CY, Lin TE, Hsieh JH, Sung TY, Tseng HJ, et al. Novel Class IIa-Selective Histone Deacetylase Inhibitors Discovered Using an in Silico Virtual Screening Approach. *Sci Rep* 2017;7:1–13. <https://doi.org/10.1038/s41598-017-03417-1>.
- [25] Kang H, Sheng Z, Zhu R, Huang Q, Liu Q, Cao Z. Virtual drug screen schema based on multiview similarity integration and ranking aggregation. *J Chem Inf Model* 2012;52:834–43. <https://doi.org/10.1021/ci200481c>.
- [26] Manal M, Chandrasekar MJN, Gomathi Priya J, Nanjan MJ. Inhibitors of histone deacetylase as antitumor agents: A critical review. *Bioorg Chem* 2016;67:18–42. <https://doi.org/10.1016/j.bioorg.2016.05.005>.

- [27] Bolden JE, Peart MJ, Johnstone RW. Anticancer activities of histone deacetylase inhibitors. *Nat Rev Drug Discov* 2006;5:769–84. <https://doi.org/10.1038/nrd2133>.
- [28] Lobera M, Madauss KP, Pohlhaus DT, Wright QG, Trocha M, Schmidt DR, et al. Selective class IIa histone deacetylase inhibition via a nonchelating zinc-binding group. *Nat Chem Biol* 2013;9:319–25. <https://doi.org/10.1038/nchembio.1223>.
- [29] Gaulton A, Hersey A, Nowotka ML, Patricia Bento A, Chambers J, Mendez D, et al. The ChEMBL database in 2017. *Nucleic Acids Res* 2017;45:D945–54. <https://doi.org/10.1093/nar/gkw1074>.
- [30] Olson RS, Moore JH. TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning 2019:151–60. https://doi.org/10.1007/978-3-030-05318-5_8.
- [31] Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. *GetMobile Mob Comput Commun* 2015;19:29–33. <https://doi.org/10.1145/2786984.2786995>.
- [32] M H, M.N S. A Review on Evaluation Metrics for Data Classification Evaluations. *Int J Data Min Knowl Manag Process* 2015;5:01–11. <https://doi.org/10.5121/ijdkp.2015.5201>.
- [33] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2:56–67. <https://doi.org/10.1038/s42256-019-0138-9>.
- [34] Molnar C. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable.* Book 2019:247.
- [35] Chevillard F, Kolb P. SCUBIDOO: A Large yet Screenable and Easily Searchable Database of Computationally Created Chemical Compounds Optimized toward High Likelihood of Synthetic Tractability. *J Chem Inf Model* 2015;55:1824–35. <https://doi.org/10.1021/acs.jcim.5b00203>.
- [36] Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 2013;27:221–34. <https://doi.org/10.1007/s10822-013-9644-8>.
- [37] Jacobson MP, Pincus DL, Rapp CS, Day TJF, Honig B, Shaw DE, et al. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins Struct Funct Genet* 2004;55:351–67. <https://doi.org/10.1002/prot.10613>.
- [38] Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: A software program for pKa prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des* 2007;21:681–91. <https://doi.org/10.1007/s10822-007-9133-z>.
- [39] Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J Chem Theory Comput* 2016;12:281–96. <https://doi.org/10.1021/acs.jctc.5b00864>.

- [40] Aly AA, Ibrahim MAA, El-Sheref EM, Hassan AMA, Brown AB. Prospective new amidinothiazoles as leukotriene B₄ inhibitors. *J Mol Struct* 2019;1175:414–27. <https://doi.org/10.1016/j.molstruc.2018.07.085>.
- [41] Singh N, Villoutreix BO, Ecker GF. Rigorous sampling of docking poses unveils binding hypothesis for the halogenated ligands of L-type Amino acid Transporter 1 (LAT1). *Sci Rep* 2019;9:1–20. <https://doi.org/10.1038/s41598-019-51455-8>.
- [42] Lyne PD, Lamb ML, Saeh JC. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J Med Chem* 2006;49:4805–8. <https://doi.org/10.1021/jm060522a>.
- [43] Farid R, Day T, Friesner RA, Pearlstein RA. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorganic Med Chem* 2006;14:3160–73. <https://doi.org/10.1016/j.bmc.2005.12.032>.
- [44] Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem* 2006;49:534–53. <https://doi.org/10.1021/jm050540c>.
- [45] Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J Med Chem* 2004;47:1750–9. <https://doi.org/10.1021/jm030644s>.
- [46] Cho AE, Guallar V, Berne BJ, Friesner R. Importance of accurate charges in molecular docking: Quantum Mechanical/Molecular Mechanical (QM/MM) approach. *J Comput Chem* 2005;26:915–31. <https://doi.org/10.1002/jcc.20222>.
- [47] Murphy RB, Philipp DM, Friesner RA. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J Comput Chem* 2000;21:1442–57. [https://doi.org/10.1002/1096-987X\(200012\)21:163.O.CO;2-O](https://doi.org/10.1002/1096-987X(200012)21:163.O.CO;2-O).
- [48] López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P. IMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Res* 2014;42:271–6. <https://doi.org/10.1093/nar/gku339>.
- [49] Bochevarov AD, Harder E, Hughes TF, Greenwood JR, Braden DA, Philipp DM, et al. Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *Int J Quantum Chem* 2013, 2013:2110–2142. <https://doi.org/10.1002/qua.24481>.
- [50] Parr RG, Szentpály L V., Liu S. Electrophilicity index. *J Am Chem Soc* 1999;121:1922–4. <https://doi.org/10.1021/ja983494x>.
- [51] Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10:1–21. <https://doi.org/10.1371/journal.pone.0118432>.
- [52] Kozikowski AP, Tapadar S, Luchini DN, Ki HK, Billadeau DD. Use of the Nitrile Oxide

- Cycloaddition (NOC) reaction for molecular probe generation: A new class of enzyme selective histone deacetylase inhibitors (HDACIs) showing picomolar activity at HDAC6. *J Med Chem* 2008;51:4370–3. <https://doi.org/10.1021/jm8002894>.
- [53] Tsai K, Wang S, Hsiao N, Li M, Wang B. Bioorganic & Medicinal Chemistry Letters The effect of different electrostatic potentials on docking accuracy: A case study using DOCK5. *Bioorg Med Chem Lett* 2008;18:3509–12. <https://doi.org/10.1016/j.bmcl.2008.05.026>.
- [54] Kombo DC, Grinevich VP, Hauser TA, Sidach S, Bencherif M. QM-polarized ligand docking accurately predicts the trend in binding affinity of a series of arylmethylen quinuclidine-like derivatives at the $\alpha 4\beta 2$ and $\alpha 3\beta 4$ nicotinic acetylcholine receptors (nAChRs). *Bioorganic Med Chem Lett* 2013;23:4842–7. <https://doi.org/10.1016/j.bmcl.2013.06.094>.
- [55] A. Sotriffer C. Accounting for Induced-Fit Effects in Docking: What is Possible and What is Not? *Curr Top Med Chem* 2012;11:179–91. <https://doi.org/10.2174/156802611794863544>.
- [56] Barreca ML, Iraci N, De Luca L, Chimirri A. Induced-fit docking approach provides insight into the binding mode and mechanism of action of HIV-1 integrase inhibitors. *ChemMedChem* 2009;4:1446–56. <https://doi.org/10.1002/cmdc.200900166>.
- [57] Bauer JA, Pavlović J, Bauerová-Hlinková V. Normal mode analysis as a routine part of a structural investigation. *Molecules* 2019;24. <https://doi.org/10.3390/molecules24183293>.
- [58] Yang Z, Li X, Ma H, Zheng J, Zhen X, Zhang X. Replacement of amide with bioisosteres led to a new series of potent adenosine A2A receptor antagonists. *Bioorganic Med Chem Lett* 2014;24:152–5. <https://doi.org/10.1016/j.bmcl.2013.11.051>.
- [59] Langdon SR, Ertl P, Brown N. Bioisosteric replacement and scaffold hopping in lead generation and optimization. *Mol Inform* 2010;29:366–85. <https://doi.org/10.1002/minf.201000019>.
- [60] Alves ALA, Dusan R, Nemanja D, Thomas K, Katarina N. Structure-Based Design of Selective Histone Deacetylase 6 Zinc Binding Groups Leandro. *J Biomol Struct Dyn* 2019;0:000. <https://doi.org/10.1080/07391102.2019.1652687>.
- [61] Prashanth J, Ramesh G, Laxman Naik J, Kishan Ojha J, Venkatram Reddy B. Molecular geometry, NBO analysis, Hyperpolarizability and HOMO-LUMO energies of 2-azido-1-phenylethanone using Quantum chemical calculations. *Mater. Today Proc.*, vol. 3, Elsevier Ltd; 2016, p. 3761–9. <https://doi.org/10.1016/j.matpr.2016.11.025>.
- [62] Gupta VP. Principles and Applications of Quantum Chemistry. n.d.
- [63] Lipinski CA. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov Today Technol* 2004;1:337–41. <https://doi.org/10.1016/j.ddtec.2004.11.007>.
- [64] Oprea TI, Gottfries J. Chemography: The art of navigating in chemical space. *J Comb Chem* 2001;3:157–66. <https://doi.org/10.1021/cc0000388>.

- [65] Arimoto R. Computational Models for Predicting Interactions with Cytochrome p450 Enzyme. *Curr Top Med Chem* 2006;6:1609–18. <https://doi.org/10.2174/156802606778108951>.
- [66] Smith DA, Di L, Kerns EH. The effect of plasma protein binding on in vivo efficacy: Misconceptions in drug discovery. *Nat Rev Drug Discov* 2010;9:929–39. <https://doi.org/10.1038/nrd3287>.
- [67] Porter NJ. Selective Inhibition Of Histone Deacetylases. University of Pennsylvania, 2019.