

Multi-scale in-silico modelling to unveil structural requirements for DNA-PK inhibitors as radiosensitizers and MolSHAP based design of novel ligands

Soumya Mitra

Dr. B. C. Roy College of Pharmacy & Allied Health Sciences

Rakesh Kumar Dolai

Dr. B. C. Roy College of Pharmacy & Allied Health Sciences

Nilanjan Ghosh

Jadavpur University

Subhash C. Mandal

Jadavpur University

Amit Kumar Halder

amit.halder@bcrpc.org

Dr. B. C. Roy College of Pharmacy & Allied Health Sciences

Research Article

Keywords: Radiosensitizers, DNA-dependent Protein kinase (DNA-PK), QSAR, molecular dynamics simulation, MolSHAP

Posted Date: September 19th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7479073/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Computer-Aided Molecular Design on December 22nd, 2025. See the published version at <https://doi.org/10.1007/s10822-025-00733-x>.

Abstract

Radiosensitizers are agents that make tumour cells more sensitive to radiation therapy. One key mechanism involves inhibition of the DNA-dependent protein kinase (DNA-PK), an enzyme crucial for repairing DNA double-strand breaks in mammalian cells. Suppression of the DNA-PK enzyme compromises the double-strand break repairs to amplify the radiation induced toxicity among the tumour cells. In this study, 73 6-Anilino Imidazo[4,5-c]pyridin-2-one derivatives were curated as potent DNA-PK inhibitors and subjected them to 2D -and 3D-Quantitative Structure Activity Relationship (QSAR) analyses to explore their structural requirements. Apart from conventional methodology, we implemented newly developed MolSHAP analyses for R-group analyses. Significant information regarding structural requirements were retrieved from each of these cheminformatic analyses. Additionally, to understand the interaction between the ligands and the DNA-PK receptor, molecular dynamics (MD) simulation analysis of 100ns were carried out for the most and the least potent compounds among the dataset. The findings indicated H-bond and π - π interactions to be the key factors for binding interactions. Furthermore, novel ligands were designed through the MolSHAP tool and were validated through the chemometric model developed in this investigation. The designed compound exhibited favourable predicted activity and replicated key interaction profiles of the co-crystallized bound ligand in MD simulations. The investigation was carried out through open-access tools to safeguard reproducibility and accessibility among researchers.

1. INTRODUCTION

Cancer remains one of the most complicated diseases to appear during the course of history with no definite therapeutic strategies to succeed. With over 1.4 million new cases detected in India alone, therapies related to the management of cancer is highly dependent on radiotherapy. Radiotherapy (RT) has been one of the most commonly used approaches in the management of different types of cancers with around 52.6% of the total population of cancer patients undergoing RT at different stages of the disease [1, 2]. RT, in combination with chemotherapy, along with radiosensitizers has become the typical approach for treating head and neck small cell carcinoma [3]. While the strategy is beneficial in cases with detection of cancer, this therapeutic approach also presents with high occurrences of toxicity towards the unaffected cells and tissue [4]. To combat the adverse effects associated with radiotherapy, several investigations have been dedicated to enhance the efficacy of the RT with minimal exposure of the patients to the radiation. One such strategy involves the introduction of radiosensitizers which are chemical compounds capable of increasing the efficiency of the RT. Radiosensitizers are effectively compounds which make the tumor cells more susceptible towards RT. Radiotherapy deals with ionizing radiation which can damage the DNA in the cancerous cells in irreversible manner. RT produces free radical, especially hydroxyl radicals, which promote breaking of the DNA double strand. The oxygen then reacts with the hydroxyl free radical to give rise to peroxide radicals which causes a more permanent damage towards the DNA strands [5]. In case of tumors, they are often seen in regions with low oxygen content leading to hypoxic tumours. Having scarcity of oxygen often leads to resistance towards RT which is caused by the inability to form peroxide free radicals. To overcome the RT associated

resistance, investigators have resorted to developing compounds like nitroimidazole, which can act as substitute for oxygen in those low oxygen environments and cause irreversible DNA damage[6–9]. Another strategy to counter the RT resistance is by developing protein kinase inhibitors that primarily stops the hypoxic cells from recovering from the DNA damage. In recent years, more intensive investigations have been carried out which focussed on the pathways that lead to DNA damage repair through development of modern radiosensitizers [10, 11].

RT has been responsible for creating double strand breaks (DSB) while inducing cytotoxic effects on the cells. Resistance towards RT can be indicated by the ability of the DNA to repair itself; primarily by two pathways; which include non-homologous end joining (NHEJ) and homologous recombination (HR) [12]. However, the NHEJ plays the predominant role as it increases radiosensitivity of cells by mutating the NHEJ genes [11, 13]. Within the NHEJ pathway, the Ku heterodimers rapidly locate the DSB and attach itself to the DNA terminal to prevent any further nucleolytic damage [14]. Irrespective of the structural configuration of the DNA, the Ku dimers show high binding affinity to the damaged ends. Due to its ring-like shape (toroidal) it can easily attach itself to the DNA terminal and start recruiting DNA-PK catalytic subunits (DNA-PKcs) along with the nuclease Artemis. Artemis is responsible for trimming the damaged ends of the DNA and together with the DNA-PKcs, they form a DNA-PK holoenzyme. The association of the Ku with DNA-PKcs increases the stability of the Ku-DNA complex and promotes a conformational shift in DNA-PKcs which activates the kinase function through autophosphorylation at multiple regulatory sites. Once activated, DNA-PKcs induces phosphorylation at the C-terminal of the autoinhibitory segment of Artemis, enabling to perform the nuclease activity. Following end processing, the DNA-PKcs recruit gap-filling polymerases from the Pol X family, specifically DNA polymerase λ and DNA polymerase μ . These enzymes restore the missing nucleotides, following which the DNA ligase IV-XRCC4-XLF complex completes the ligation to facilitate DNA repair [15–22]. This loss of DNA-PK functions severely sensitises the DSBs [23, 24], that makes DNA-PK as major drug target [11, 25, 26]. Recently it was confirmed that the DNA-PK offers the higher radio sensitization than inhibitors of PARP-1 (poly (ADP-ribose) polymerase-1) in HNSCC (Head and neck squamous cell carcinoma) cells [27]. The limited resolution of DNA-PK structural data first made it difficult to find strong and specific DNA-PK inhibitors [28–31], and dependent on the homology model [32, 33] based on the phosphoinositide 3-kinase (PI3K) enzymes. Because of this, initial inhibitors like dactosilib, NU7441, KU0060648, and IC87361 show lack of precise selectivity to DNA-PK, they affected not only DNA-PK but also PI3K and other related enzymes in the PIKK (PI3K-related kinase) family [34–39]. This lack of selectivity led to unwanted adverse effect or toxicities in preclinical studies, limiting their therapeutic potency [40]. In recent time, more selective inhibitors to DNA-PK such as AZD7684, NU5455, Pepostertib (M3814), and 31t are identified and most of these are characterized at higher resolution in structural studies [41–47]. Two of them, Peposertib with RT as well as AZD7648 in combination with pegylated doxorubicin or Olaparib, have progressed to clinical trials. In the current work, we collected a data set containing 73 6-Anilino Imidazo [4,5-c] pyridin-2-one's derivatives that were recently reported as novel DNA-PK inhibitors and performed a range of in silico modelling analyses to understand the structural

requirements in order to ensure that more robust design strategy may be adopted with such information for obtaining DNA-PK inhibitors with greater potency [48].

2. MATERIALS AND METHODS

2.1 Dataset curation and preparation

In search of potent DNA-PK inhibitors, Hong et al. designed a series of novel 6Anilino Imidazo[4,5c]pyridin-2-one derivatives by deploying scaffold hopping techniques on an established PI3K and PIKK inhibitor Dactolisib [48]. The group developed new inhibitors and evaluated their inhibitory potential against the DNA-PK enzyme using the HotSpot assay platform. The nanomolar IC₅₀ values reported for the 73 compounds were subsequently transformed into pIC₅₀ ($= -\log_{10}(\text{IC}_{50}/10^9)$) values.

The SMILES notations of the dataset compounds are collected from the article of Hong et al [48]. The dataset was prepared by converting the curated SMILES into their canonical forms via RDKit (available at <https://www.rdkit.org/>) and were converted into .sdf files through Discovery Studio Visualizer tool (available at <https://discover.3ds.com/discovery-studio-visualizer-download>). This was followed by incorporation of the biological activity in terms of pIC₅₀ as the response variables required to develop QSAR models. Once the dataset was ready (Table S1 of the Supplementary information), the ligands of the dataset were subjected to structural standardization by the CDK standardizer. The standardization process entails adding explicit hydrogen atoms, performing aromatization, cleaning the structure in both 2D and 3D, neutralizing, and removing any salts.

2.2 2D-QSAR modelling

2.2.1 *Descriptor calculation*

For descriptor calculation, the open-access webserver OCHEM (<https://ochem.eu/login/show.do>) was utilized where 73 dataset ligands were fed to compute alvaDesc descriptors (<https://www.alvascience.com/alvadesc-descriptors/>) in .sdf format [49, 50]. The molecular structures of the 6anilino imidazo[4,5c]pyridin-2-one derivatives were carried out by the RDKit module of the webserver. The descriptor calculation took place in two steps, wherein the first step included computation of all the alvaDesc descriptors and the next step included computing only those alvaDesc descriptors which have better interpretability. The details of the interpretable descriptors have been provided in one of our previous works [51]. The descriptor files (.csv) were prepared by adding the identifier of dataset ligands and their respective pIC₅₀ values as response variables.

2.2.2 *Dataset division and model set-up*

As both the datasets with all AlvaDesc descriptors and interpretable descriptors were developed, they were segregated into training and test (or prediction) sets by using SFS-QSAR-tool_v2 (available at

https://github.com/ncordeirfcup/SFS-QSAR-tool_v2) [51]. The datasets were divided while deploying activity sorting 4, and the ratio between training and prediction set were kept at 4:1.

In regression analysis, feature selection helps in identifying the relevant features during model development [52, 53]. Two distinct feature selection modules were utilized for setting up the 2D-QSAR models; (i) sequential forward selection and (ii) genetic algorithm. The first method SFS works by iteratively adding the most significant feature that improves model performance until a stopping criterion is met. On the other hand, GA is a heuristic search inspired by natural selection [51]. Sequential forward selection (SFS) is a non-stochastic greedy algorithm which acts by selecting the optimum features in a dataset. It starts with no features and keeps adding features one by one until the termination criterion is met. The SFS-QSAR tool uses the Mlxtend module (available at <https://rasbt.github.io/mlxtend/>) for feature selection process. To prepare the data, variables with minimal variance were removed using a 0.0001 variance threshold where multicollinearity was managed with a correlation threshold of 0.99. To evaluate the predictive nature of the developed models, four scoring functions were deployed including the likes of R^2 (determination coefficient) which measures how much variance in the dependent parameter is explained by the model, *NMAE* (normalized mean absolute error) which measures the average of the absolute errors between the predicted value and actual values, *NMPD* (normalized mean Poisson deviance) which measures errors for count-based data assuming a Poisson distribution and finally *NMGD* (normalized mean gamma deviance) which measures error for positive continuous data assuming gamma distribution. Eight (8) different models with SFS-QSAR were generated, four each for 5-fold and 0-fold cross validation.

Unlike SFS, Genetic algorithm (GA) is a probabilistic optimization approach, which is inspired by the Darwinian evolution, employing processes analogous to inheritance, recombination, mutation and the survival of the fittest. Feature selection was performed through GeneticAlgorithm v.4.1_2 tool (available at http://teqip.jdvu.ac.in/QSAR_Tools/) [54]. As GA's stochastic approach can create variability in the selected features, the model generation was repeated for 25 times. Similar to SFS, a variance threshold of 0.0001 was maintained, but the collinearity threshold was relaxed at 0.975. The algorithm was executed with default parameters; mutation rate of 0.3, 100 evolutionary cycles and crossover rate at 1. To eliminate the risk of overfitting and retain the relevant features only, the maximum descriptor count was limited to 7 for both datasets across all feature selection runs. Quality of the generated 2D-QSAR models were determined on the basis of Q^2_{LOO} and R^2_{Predi} which estimates the internal and external predictivity of the models. The details of these parameters have been provided in the Supplementary information.

2.2.3 Applicability domain (AD) of the models

The AD is a theoretically determined descriptor space within which reliable and accurate predictions may be ascertained. The AD can help in estimating uncertainty, ensuring reliable predictions and also monitoring performance of the model. AD of the generated 2D-QSAR model was determined by Williams plot as well as by standardization approach. The Williams plot compares leverage values with

standardized coefficients, helping to detect both response and structural outliers. It envisages the impact of each data point's leverage against standardized residuals. When a data point's leverage value exceeds the threshold hat value (h^*), calculated as $3p/n$ (p : the number of descriptors + 1, and n : the number of data points in the training set), this datapoint is flagged as structural outlier. Response outliers are identified when any data point's standardized residual surpasses ± 2.5 . On the other hand, standardization approach is a relatively novel and more straightforward strategy for determining AD of QSAR models using basic theory of standardization[55, 56].

2.3 MolSHAP-based analysis

In the next phase of this investigation, a relatively novel and unexplored tool was deployed to understand the contributions of different fragments within a chemical structure for exerting binding potential. The MolSHAP is a relatively new tool, recently developed by Tian et al, to determine the contributions of diverse fragments of a compound. MolSHAP is a model-agonistic framework which interprets the quantitative structure activity relationship (QSAR) by focusing on *R*-group contributions within a molecule rather than fingerprints or SMILES as a whole [57]. MolSHAP works best with series of compounds with a common core which is then extracted by the decomposition of the structure. The decomposition function of the tool breaks the structures down into a core, common for all the compounds in the series, and its *R*-groups at different positions.

Once the compounds are decomposed, the next step involves *R*-group ranking where the contribution of each of the fragment is determined. The idea behind this step is to identify the fragments which contribute the most in exerting biological activity. For this, initially, two compounds are compared where only one *R*-group is different and the difference or similarity of the biological activity helps in evaluating the contribution of the fragment in that structure. This leads to ranking of the *R*-groups from the most important to least important to exhibit biological activity. The next stage in MolSHAP analysis involves training the compounds with different machine learning algorithms. In the current work, Gaussian Process Regressor was used as it was a part of MolSHAP software provided by Tian et al (<https://github.com/tiantz17/MolSHAP>) [57]. The major competence of MolSHAP lies in its ability to interpret the contribution of the diverse fragments in a structure while using a modified Shapley value-based interpretation. Unlike the traditional SHAP analysis, this tool does not remove the fragments from a structure, which may lead to the formation of an invalid or broken structure. On the contrary, the tool uses *R*-group masking task where the fragment in question is replaced by a background neutral fragment so as to retain the integrity of the structure. By systematically masking different *R*-groups and analyzing how predictions change, MolSHAP computes Shapley values that reflect each fragment's contribution to activity. These interpretations are evaluated using the *R*-group ranking task, which involves comparing the *R*-group rankings predicted by Shapley values to the actual rankings determined from experimentally observed or model-predicted differences in compound activity.

MolSHAP also provides the option of determining uncertainty estimation for models like Gaussian processes (GP) which helps in evaluating the models' quality. Determination of the model quality helps in eliminating models with low confidence level while retaining the ones with better model quality. Finally,

MolSHAP enables compound optimization by leveraging the calculated Shapley values. Starting from a lead compound, *R*-groups with low MolSHAP values are identified and substituted with fragments that have higher contributions. This guided *R*-group replacement strategy is iteratively applied to design new molecules that are predicted to have enhanced biological activity. The optimized molecules are screened and retained only if they show improved predicted potency, and their uncertainty is acceptably low.

For this particular study, MolSHAP tool was used (available at <https://github.com/tiantz17/MolSHAP>) where initially the dataset compounds were prepared by keeping three distinct column with ID, SMILES and Activity (response variable) as headers and subjected to *decompose.py* function of the module. This leads to generation of the core molecule and fragments of the dataset. Once the side chain decomposition was completed the *molshap.py* module of the tool was deployed to find the decomposition results, side chain contribution matrix, list of optimized compounds and a scattered plot of optimized compounds.

2.4 3D- QSAR modelling

2.4.1 Alignment of the dataset ligands

In this current work, we used rigid body alignment of the structures for the 3D-QSAR, with the help of open-source application called MinAlignConf (available at <https://github.com/ncordeirfcup/MinAlignConf>).

2.4.2 Model development

The 3D-QSAR relationship analysis was performed using Open3DQSAR, an open-source framework for modelling steric and electrostatic molecular interactions [58, 59]. The detailed methodology for the model development was reported previously [60, 61] and a brief description is provided in the Supplementary information mainly highlighting the dataset division and feature selection methodologies as well as statistical parameters. Finally, contour maps were visualized at partial least squared coefficient threshold of + 0.005 (green) and – 0.005 (yellow) for steric fields; and + 0.003 (blue) and – 0.003 (red) for electrostatic fields [62].

2.5 Molecular docking

For molecular docking studies, the DNA-PK protein was initially curated from RCSB Protein Data Bank (PDB ID: 7OTW) (available at <https://www.rcsb.org/structure/7OTW>). Docking analysis was carried out using Autodock Vina [63, 64]. The initial stage of molecular docking requires pre-processing of the protein structure which includes addition of all the missing hydrogen atoms, assigning Gasteiger partial atomic charges and subsequently converting the coordinate file into *pdbqt* format using AutoDock tools. Following the preparation of the protein structure, the 3D structures of the most active compound, the least active compound and the co-crystallized reference inhibitor AZD7648 were constructed. These ligands were then subjected to energy minimization and geometric optimization employing the MMFF94 force field to ensure accurate conformational stability prior to the docking simulations. The next step

includes defining a grid box around the active site of the protein with dimensions of X = 157.625, Y = 174.369 and Z = 218.693 with extension of 30Å. The resulting docking poses and binding affinities were subsequently utilized for molecular dynamics (MD) simulation analysis.

2.6 MD Simulation analysis

The protein-ligand complexes derived from the docking studies were utilized for MD simulation which was carried out with 100ns long run. The detailed methodology for the MD simulation study has been described in some of the previous investigations [51, 65]. Brief description is however provided in the Supplementary information to describe the set-up of MD simulation and its analyses.

3. RESULTS AND DISCUSSION

3.1 2D-QSAR modelling

As the strategy is outlined before, the 2D-QSAR models were developed with the intention of having better interpretable models without sacrificing the predictive quality of the models. To develop best linear model, two sets of alvaDesc descriptors sets were considered and these are (a) interpretable descriptors and (b) all descriptors [62] to check which set may provide the best solution. Naturally, the chance of getting a better QSAR model is always greater for the latter set since the former is just a subset of the latter. However, due to the fact the feature selection strategy is a complex method, it is always worthwhile to check which set produces the better result. If the second set is indeed found to generate the better results, it is required to check how much difference is found between the best models produced with by these sets. Evidently, a predictive model produced by the interpretable descriptors is capable of providing high statistical quality as well as mechanistic interpretability. Since, it is not wise to compromise the statistical results for sake of molecular interpretability, the current approach clarifies whether inclusion of all descriptors is indeed required.

In this particular investigation, two distinct approaches were conceived to develop the 2D-QSAR models which required utilization of SFS and GA algorithms. The linear regression models were developed from the statistics resulted from the training and test sets. To consider the best predictive model from this investigation the average of $Q^2_{LOO_i}$ an estimation of the internal predictivity, and $R^2_{Pred_i}$ estimating the external predictivity was considered. The findings from the 2D-QSAR analysis are listed in Table 1 [51].

Table 1
Synopsis of the statistical parameters from 2D-QSAR models.

Method	Score	Fold	Interpretable descriptors		All descriptors	
			Q^2_{LOO}	R^2_{Pred}	Q^2_{LOO}	R^2_{Pred}
SFS	R^2	0	0.642	ND	0.864	0.828
SFS	R^2	5	0.695	ND	0.825	ND
SFS	NMAE	0	0.664	ND	0.764	ND
SFS	NMAE	5	0.689	ND	0.522	ND
SFS	NMPD	0	0.664	ND	0.833	ND
SFS	NMPD	5	0.685	ND	0.795	ND
SFS	NMGD	0	0.678	ND	0.814	ND
SFS	NMGD	5	0.683	ND	0.795	ND
GA	NA	NA	0.762	0.208	0.827	ND

#ND: Not Determined

Since multiple models were generated here with different descriptor sets and model development strategy, we selected the most predictive model from maximum Q^2_{LOO} , which justified the internal predictivity of the model. Table 1 clearly demonstrates that the interpretable descriptors were inadequate to develop a predictive model as predictive as the models produced with all alvaDesc descriptors. The most predictive model was produced when the SFS-MLR model was constructed using R^2 as scoring function (without CV). This model (henceforth designated as Model 1) showed Q^2_{LOO} of 0.864 and R^2_{Pred} of 0.828. This SFS- MLR model is presented in Table 2 with 2D-QSAR equation and statistical parameters for the training and test sets. Overall, the model is generated with a satisfactory statistical predictivity and it is also evident from statistical parameters other than Q^2_{LOO} and R^2_{Pred} . Reports suggested that validation parameters based on rm^2 statistics are more robust than these two parameters [66]. The internal validation parameters rm^2_{Train} and the Δrm^2_{Train} values of the model were found as 0.814 and 0.059, respectively. At the same time, the rm^2_{Test} and Δrm^2_{Test} values are 0.769 and 0.120, respectively. The observed vs. predicted pIC₅₀ values of 2D-QSAR model is depicted in Fig. 1.

After that we checked the correlation matrix of this model (shown in Fig. 1) and it was observed that these descriptors are devoid of high inter-collinearity as the highest absolute correlation (Pearson r) between any two descriptor is 0.45 (i.e., < 0.95).

Y-randomization was performed by shuffling the pIC₅₀ values of the compounds over 1000 iterations, yielding a $^cR^2_p$ value of 0.837. The relatively high score confirms the statistical reliability of the model

and not generated by random chance.

Table 2
Detailed statistical parameters of 2D-QSAR model.

Equation	Training set	Test set
$pIC_{50} = -6.404(\pm 0.844) * VE1sign_B(m) - 2.456(\pm 0.595) * GATS8v + 3.866(\pm 0.729) * TDB03p + 1.815(\pm 0.28) * L3v - 0.333(\pm 0.066) * F02[C-N] + 0.257(\pm 0.051) * CATS3D_02_DL + 13.555(\pm 3.049) * WHALES20_IR + 5.46(\pm 1.756)$	$N_{tr} = 59, R^2 = 0.900,$	$N_{ts} = 14, R^2_{Pred}/Q^2_{F1}$
	$R^2_{adj} = 0.886,$	$= 0.828, Q^2_{F2} = 0.830,$
	$Q^2_{LOO} = 0.864,$	$Q^2_{F3} = 0.848,$
	$MAE = 0.029,$	$RMSEP = 0.420,$
	$MSE = 0.341$	$rm^2_{Test} = 0.769,$
	$rm^2_{Train} = 0.814,$	$\Delta rm^2_{Test} = 0.120$
	$\Delta rm^2_{Train} = 0.059$	

Moreover, it is absolutely important to check the applicability domain (AD) for the establishment of the 2D-QSAR model. The Williams plot was generated for the Model 1, which is presented in the Fig. 2. No structural or response outliers were found in this model. However, standardization approach detected one structural outlier in the training set and we retained this since no test set compound was identified as outlier.

Figure 2(B) provides an idea about the impact of each descriptor from the model M1. The model with the all descriptors shows insightful structural information of 6-Anilino Imidazo[4,5-c] pyridin-2-one derivatives. At the same time, out of 7 descriptors, two (*i.e.*, *WHALES20_IR* and *VE1sign_B(m)*) showed high significance and three (*i.e.*, *TDB03p*, *L3v* and *GATS8v*) demonstrated moderate significance. Two descriptors, namely *CATS3D_02_DL* and *F02[C-N]* depicted negligible significance. The descriptions of the descriptors of the model are outlined in Table 3. Noticeably, most of the descriptors of the model is topological in nature and less interpretable. We first describe the *WHALES20_IR* descriptor which is found to be the most influential descriptor from Model 1 and it was found to have positive correlation towards the response variable. Weighted Holistic Atom Localization and Entity Shape (*WHALES*) descriptors are kind of 3D topological descriptors, which capture molecular shape and partial charges concurrently. Here IR of the *WHALES20_IR* denotes isolation and remoteness, and 20 is ratio percentile. Isolation depends on the local electron density and nearby significant group; remoteness is influenced by

the topological distance in 3D structure. Molecules with different IR profiles may have different impact on the biological activity whereas 2D percentile value focuses on the less isolated regions, which is critical for the active site for the ligands. Overall, it can be concluded that the specific 3D shape of the compounds is important to determine the biological activity. The maximum value of this descriptor is found in **88**, which is also the most potent compound of the dataset with reported pIC_{50} value of 8.674. On the other hand, another compound **29** that showed the lowest value of *WHALES20_IR* (i.e., -0.135) was found to possess an pIC_{50} of 5.442. In Fig. 3, we demonstrated and compared the structures of the most potent compound **88** and least potent compound **41**, which has *WHALES20_IR* value of -0.108. In the same figure, similar comparisons were performed with second and third most influential descriptors - *VE1sign_B(m)* and *TDB03p*. The second most important descriptor is *VE1sign_B(m)*, with the negative correlation with the response variable. It is a 2D matrix-based descriptors that is weighted with atomic mass of the compounds. The third most important descriptor is *TDB03p*, under 3D autocorrelations category. This depends on the polarizability of the compounds, as it is positively correlated with the response variable the higher polarizability enhances the binding potential of the compounds which is shown in Fig. 3. Interestingly, the most potent compound **88** was found to have maximum value (i.e., 1.831) of this descriptor whereas the lowest descriptor value (i.e., 1.435) was found in a low active derivative **73** (pIC_{50} = 5.714). Two least potent derivatives **20** and **41** had *TDB03p* values of 1.486 and 1.479, respectively. In comparison to **88**, all these low active compounds contained more bulky non-polar residues such as cyclopentyl, that may be responsible for low activity towards PKI enzyme.

The fourth and the fifth most important descriptors are *GATS8v* and *L3v* respectively. Even though *GATS8v* belongs to the 2D autocorrelations category and *L3v* is a *WHIM* descriptor, both these are weighted by the van der Waals volume and both of them are found to be positively correlated with the response variable. Both descriptors bank on the van der Waals volume. Two descriptors with least significance are *F02[C-N]* and *CATS3D_02_DL*. The *F02[C-N]* belongs to 2D-atom pair descriptor which denotes the depicts the number of times the pair C and N appear within the molecular structure at topological distance of 2. Low value of this descriptor was found to be important. Finally, CATS descriptors are pharmacophore descriptors and *CATS3D_02_DL* quantifies the frequency of donor-lipophilic feature pairs at a defined 3D distance interval (bin 02). Positive correlation of this particular descriptor with the response variable signifies that the presence of this structural feature is beneficial to the binding potential of the compounds [67]. Detailed description of each molecular descriptor in the developed model has been provided in the Table S2 of Supplementary information.

3.2 MolSHAP analysis

2D-QSAR model provided satisfactory statistical results but provided us limited information about the contributions of various substitutions for increased or decreased potential towards DNA-PK. As it has been discussed earlier, MolSHAP is a comparatively novel strategy of QSAR model development that allows *R*-group analyses and may be considered as upgraded strategy of Free Wilson analysis. Here, the model generated with Gaussian Process Regressor (GPR) showed R^2 values of 0.987 and 0.844, respectively. However, it was necessary to more robust internal validation parameter and for this 5-fold

cross-validation was performed with the training set that resulted in R^2 value of 0.583. Therefore, the model depicted satisfactory statistical predictivity. First, the contributions of the R groups were analysed and the results are depicted in Fig. 4. In this figure the core common structure is shown and 10 most favourable and 10 most unfavourable fragments are provided.

The R -group analyses revealed that the R_6 and R_7 positions (both attached to imidazo[4,5-*c*]-pyridine-2-one moiety) are the most important regions for determining increased biological activity. The presence of 4-methoxyphenyl moiety at position R_6 of the common structure appears as the most contributing fragment for increasing biological activity and it is followed by 4-hydroxyphenyl and 4-(2-methyl-2-phenylpropanenitrile) moieties. Noticeably, the latter fragments are found in two most potent compounds; **87** and **88**, with pIC_{50} values of 8.357 and 8.674 respectively. Substitutions at positions R_1 , R_2 and R_4 also determine improved activity. Among this, the presence of methyl group at R_1 exhibits maximum impact with contribution of 0.83. Noticeably, the presence of isopropyl group at R_7 is responsible for lowering the biological potency. Noticeably, the presence of this moiety is found in **72** that has pIC_{50} value 5.975. However, replacement of this isopropyl moiety with methyl group gives rise to **53** with significant improvement of biological activity ($\text{pIC}_{50} = 8.134$). The MolSHAP analysis clearly indicates that the presence of methyl group at R_1 and R_4 is favourable whereas methyl group at R_5 is detrimental for increased potency. Similarly, influence of cyclopentyl was found to vary depending on its position. Substitution at R_6 was favourable whereas R_7 in unfavourable. Overall, the 2D-QSAR analysis highlights the favourable and unfavourable fragments but does not indicate the electrostatic and steric factors responsible for the increased or decreased potentials towards PKI enzyme. Therefore, we resorted to 3D-QSAR analyses to extract more information regarding these.

3.3 3D- QSAR modelling

As seen in Table 3, the better predictive quality was achieved through the deployment of FFD-SEL algorithm. The number of ligands considered to for training set was 59 while 14 datapoints were set aside for test set prediction. The Q^2_{LOO} (0.726), Q^2_{LTO} (0.760) and Q^2_{LMO} (0.736) values indicates that the model has strong internal predictivity, as well as external predictability ($R^2_{Pred} = 0.827$). Additionally, the 3D-QSAR model accurately captures the role of both steric as well as the electrostatic fields in model development. Figure 5 shows a visual depiction of the contour maps.

Table 3
Statistical parameters from 3D-QSAR analysis

Parameter	FFD-SEL	UVE-PLS
PC	4	4
N _{training}	59	59
F-test	134.456	125.997
R ² /SDEC	0.908/0.321	0.903/ 0.331
Q ² _{LOO} /SDEP	0.762/ 0.518	0.695/0.588
Q ² _{LTO} /SDEP	0.760/0.521	0.691/0.591
Q ² _{LMO} /SDEP	0.736/0.546	0.658/ 0.622
N _{test}	14	14
R ² _{Pred} /SDEP	0.827/0.446	0.786/0.495

The contributions of the steric and electrostatic fields were found as 58% and 42%, respectively indicating the joint responsibility of both components in shaping the model. There is substantial consistency between the MolSHAP analyses and 3D-QSAR modelling.

The presence of relatively bulky and electron rich 4- 2-methyl-2-phenylpropanenitrile moiety at R_6 position was inserted into a steric favourable as well as electronegative fields. At the same position (i.e., R_6), the lowest active **41** contained relatively less electron rich cyclopentyl ring that was inserted into an electronegative field. Being a bulky group, the cyclopentyl was found close to neither steric favourable nor steric unfavourable fields.

Compound **88**, the one with the highest binding affinity contains a bulky aromatic side chain that aligns well within the steric favourable region of the isocontour map. In addition, a methoxy and a methyl substituent are also positioned in sterically advantageous regions, prompting further enhancement in its binding potential. By contrast, compound **41**; the one with the weakest binding affinity, lacks any bulky substituents near the steric favourable region, explaining one of the causes of its weak binding potential. Analysis of the electrostatic isocontour maps also revealed that compound **88** places its electron rich aromatic ring in closer proximity to an electrostatically favourable region, supporting beneficial interactions. On the other hand, compound **41** contains a hydrophobic cyclopentyl moiety within the same electronegative favourable region. Since non-polar groups are unable to interact effectively in such regions, this unfavourable placement likely results in decreased binding affinity. Overall, these observations provide important insights into the structure-activity relationship of DNA-dependent protein kinase inhibitors.

3.4 Molecular dynamics simulation

The 2D-QSAR, MolSHAP and 3D-QSAR provided various information about the structural attributes for dataset compounds for enhanced potential regarding PKI. Nevertheless, it was necessary to understand the dynamic behaviours of the 6-Anilino Imidazo[4,5-c] pyridin-2-one's derivatives with PKI. In order to understand this, we selected the most active compound **88** and one of the least active compounds **41** for MD simulation analyses. For molecular docking the protein structure provided by Hong et al was used in the current work. Notably, they also provided the docked complexes of Dactolisib and another dataset compound **78**. In the current work, we used the AZD7648-PKI complex as a reference to compare the dynamic behaviours of **88** and **41**. Interestingly, when the molecular docking was performed with **88** and **41** using Autodock Vina, the best docked pose of **88** replicated the structure of both AZD7648 as well as **78** (both provided by Hong et al). However, the docked conformation of **41** failed to replicate the similar conformation that led us to perform forced alignment of **41** with the docked pose of **78** to check the reason for this (Fig. S1). For this alignment, the *.sdf* form of **78** was used as a template whereas the structure of **41** was provided as SMILES. Subsequently, the aligned structure of **41** was minimized by *OpenBabel* tool using MMFF94 forcefield. The minimized structure of **41** was found to have multiple clashes at the binding site (shown in Fig. S1) that led us to proceed with the docked conformation of both **88** and **41** for 100ns MD simulations. After the MD simulation, we performed trajectory analyses, MM-GBSA analyses and per-residue decomposition analyses. The MM-GBSA results that are provided in Table 4 match the differences in experimental biological activities of **88** and **41**.

Table 4
MM-GBSA binding energies (in kcal/mol) from MD simulations

Complexes	ΔE_{vdW}	ΔE_{elec}	ΔG_{polar}	$\Delta G_{\text{nonpolar}}$	$T\Delta S$	$\Delta G_{\text{bind(T)}}$	pIC ₅₀
88	-54.58	-7.08	19.86	-6.37	-22.57	-25.6	8.674
41	-44.61	-6.32	16.42	-5.23	-29.64	-10.1	5.000
AZD7648	-47.66	-7.13	17.71	-4.95	-18.89	-23.14	-

More importantly, the best active **88** depicted significantly increased van der Waals interactions and considerably higher hydrogen bond interactions than low active **41**. Even though the nonpolar solvation energy of **88** was found to be more favourable, the polar solvation energy of the compound was less favourable for **88** that may be due to its larger molecular weight. The trajectory analyses, performed with **88**, **41** and **AZD7684** revealed sufficient dynamic stabilities of all these ligands during MD simulations. However, slightly increased fluctuations were noted for low active **41** (average RMSF = 1.66 Å) as compared to **88** (average RMSF = 1.55 Å). Furthermore, as shown in Fig. 6, **88** demonstrated higher number of hydrogen bond interactions in comparison with **41** further confirming that the former may produce increased electrostatic interactions with the receptor.

In Fig. 7, the last poses of these two compounds are depicted. However, to understand more specifically about the interactions between these ligands and DNA-PK, we performed the per residue decomposition

analyses with the amino acids located within 10Å radius of the ligand.

Noticeably, when we compare the per residue decomposition analyses results of **88** and **41**, the former depicted increased interactions with the receptor. In one hand, the electrostatic interactions of **88** with Lys3753 and Leu3806 were found to be significantly greater than **41** whereas van der Waals interaction of **88** with Trp3806 was found to be predominant in determining its higher binding potential. In this regard, we analysed the hydrogen bonds formed between these two ligands and the receptor. The trajectory analyses also revealed that the hydrogen bond interactions between pyridine nitrogen of **88** and Leu3806 may play the most significant role. Apart from that, the oxygen atom of the methoxy group of **88** forms hydrogen bond interactions with Lys3753. Compound **41** fails to show such significant interactions with the receptor and in this case, it is the Ser3731 that forms maximum hydrogen bond interaction with this compound. Noticeably, some π - π interactions were found to play pivotal role in determining the binding affinity. Interestingly, trajectory analyses revealed that almost 97% frames were identified in the last 50 ns MD simulation run where π - π interactions were found between heteroaromatic residues of **88** and Trp3805. However, only negligible number of frames were found where the aromatic rings of **41** engaged in π - π interactions with Trp3805 and Tyr3791. All these clearly indicate that it is the overall 3D topology of **41** that may hinder it to mimic the binding conformation adopted by **88** and other higher active compounds and it ultimately leads to lower activity towards DNA-PK.

3.5 MolSHAP based novel compound design

Previously, we discussed development predictive MolSHAP QSAR modelling to understand favourable and unfavourable fragments. However, the method is also used to design and propose new compounds. Here, total 239 compounds were optimized and ranked as per the predicted pIC_{50} values. We picked six structures with predicted pIC_{50} greater than **88** (the most active compound of the dataset) and their structures were found that are depicted in Fig. 8.

However, these MolSHAP predicted structures were screened with the 2D-QSAR model, which was so far proved to be the most predictive model, MS02 and MS06 were found as structural outliers by the standardization method. Significantly, the molecular docking performed with these six compounds also depicted that except these structural outliers, the best poses of remaining four MolSHAP predicted compounds superimposed well with the bound conformation of AZD or docked pose of **78** (Provided by Hong et al). Now, after screening with our 2D-QSAR model, these four compounds exhibited predicted pIC_{50} values as follows: MS01: 8.593, MS03: 8.397, MS04: 7.686 and MS05: 8.079. Therefore, as per the predictions provided by both MolSHAP and 2D-QSAR models, MS01 emerged as the most promising structure designed by MolSHAP. In order to ensure further about its reliability, we performed 100ns MD simulations with the docked conformation of it. This compound depicted $\Delta G_{\text{bind(T)}}$ of -20.43 kCal/mol, which is very close to AZD7684. The energy contributions are as follows: ΔE_{vdW} = -48.03 kCal/mol, ΔE_{elec} = -8.47 kCal/mol, ΔG_{polar} = 19.48 kCal/mol, $\Delta G_{\text{nonpolar}}$ = -5.46 kCal/mol and $T\Delta S$ = -22.04 kCal/mol. Noticeably, this compound has highly satisfactory polar interactions that is even greater than that

observed for **88**. Additionally, trajectory analyses (shown in Supplementary information) depicted that MS01 maintains conformational stability and forms 5–7 hydrogen bond interactions during MD run.

Analyses of the last pose of MS01 after MD simulation and the results of per residue decomposition analyses hint that MS01 may form stable hydrogen bond interactions with Leu3806 and Glu3804. Furthermore, hydrogen bond analyses indicated that its methoxy residue may also form stable hydrogen bond interactions with Asn3926. Similar to **88**, it also shows stable π - π interactions with Trp3805 as around 99% of frames in the last 50ns MD run depicted its existence and according to per residue decomposition analyses, van der Waals interactions with this amino acid residue may play predominant role in its binding with the receptor as shown in Fig. 9.

4. CONCLUSION

DNA-PK inhibitors have emerged as one of the leading candidates as radiosensitizers and the Hong et al designed a series of 6-Anilino Imidazo[4,5-c]pyridin-2-one derivatives through the much-coveted scaffold hopping strategy. This particular investigation, seeking to unravel the structural requirements of the DNA-PK inhibitors, constitutes of performing 2D-QSAR and 3D-QSAR analysis. Apart from the conventional QSAR analyses, the study also focuses on understanding the contributions of different structural fragments of the dataset ligands. This is followed by executing molecular dynamics simulation analysis for ligands with greatest and lowest biological activities. Finally, with the help of MolSHAP, new ligands were designed and the ligand with the best predicted activity was subjected to MD simulation analysis to reveal the various molecular interactions while binding with the receptor protein.

Firstly, the 2D-QSAR analysis retrieved satisfactory results with both internal and external predictivities of 0.886 and 0.828 respectively. Furthermore, the model also brought forward the most significant descriptors from the dataset. It was found out that topological descriptor (WHALES20_IR) leverages a crucial influence on the binding potential of the ligands. The 3D topology of the structures provides an important idea about the electron density of the various atoms in the structure which is relevant in exerting binding capabilities of the ligands. Subsequently, the presence of bulky steric groups may influence negatively to the pIC₅₀ of the ligands. Descriptors from the 2D-QSAR analyses strongly indicate the importance of partial charges and van der Waals volume, exerting key influence on the binding potential. This was further supported by the MolSHAP analysis which clearly states the contributions of different fragments of the compounds in the dataset. From the MolSHAP analysis, it was found out that 4-methoxyphenyl moiety had the highest positive contribution in showing greater binding potential. The greater electron density of the group holds a significant importance that is also backed up by the findings from the 2D-QSAR analyses. Similar effects can be observed for the hydroxyphenyl moiety which is a strong electron donating group, having higher density of electrons in the structure. This phenomenon was also observed with 2D-QSAR analyses which heavily relied on the contributions of 3D topology and presence of electron rich regions of the dataset ligands.

The findings from the 3D-QSAR analyses were found to be consistent with the information extracted from MolSHAP analysis which indicated the significance of electron rich 4-methoxyphenyl group in shaping the binding potential of the ligands. The 3D-QSAR isocontour maps also provides evidence of the electronegative fragment of the compound **88** being inserted into the electrostatic favourable region. The significance of the fragment was further consolidated by the fact that due to its bulky nature, it was also present in the steric favourable region of the isocontour map. On the other hand, for compounds with lesser binding affinity, the less electronegative fragment like cyclopentyl moiety was in close proximity to the electronegative region of the isocontour map. The consistency between the findings from the 2D, 3D and MolSHAP analyses encouraged us into looking at findings from the MD simulations. To understand the interactions between the ligands and receptor, 100ns MD simulation analyses were employed which gave a clear indication of the participating interactive force between them. The MM-GBSA analysis for both the highest and lowest active compounds along with the bound ligand AZD7684 provided insights on how electrostatic interactions might play an important role in determining the binding potential of the ligands. The trajectory analysis revealed that compound **88** participated in more hydrogen bond interactions than **41** which leads to its greater binding potential. The per-residue decomposition analysis provided evidence of electrostatic interactions between **88** and amino acid residues Lys3753 and Leu3806; while the van der Waals interactions with Trp3806 was the predominant factor contributing to the binding affinity of the ligand. These interactions were either absent or very weak in case of compound **41**; leading to its lower biological potential. Finally, it was revealed that π - π interaction between the heterocyclic structure of **88** and Trp3805 was captured in almost 97% frames of the last 50ns of the MD run which was scarce for compound **41**.

Upon understanding the ligand-receptor interaction, the novel designed molecule through MolSHAP was subjected to MD simulation analysis. The best predictive compound **MS01** was found to exert similar MM-GBSA analysis to that of AZD7684 and the polar interactions were found to be more potent than that of **88**. The per-residue analysis of the ligand and receptor revealed stable H-bond interactions similar to **88**. Furthermore, **MS01** also exhibited π - π interactions with Trp3805 which was found to be a crucial factor in exerting binding potential of the compounds. The findings from this investigation can be of enormous help in designing new DNA dependent protein kinase inhibitors as the structural requirements along with the relevant interactions with the receptor amino acids have been explored in details. Furthermore, the entire work is based on open-access tools that require no subscription cost and for this, the models are reproducible.

Declarations

CRediT authorship contribution statement

S. Mitra: Writing – original draft, Investigation, Formal analysis, Validation, Resources, Data curation, Software. **R. K. Dolai:** Investigation, Data curation, Writing – original draft, Software. **N. Ghosh:** Writing – review & editing, Formal analysis, Conceptualisation, Methodology, Funding acquisition, Investigation, Supervision. **S. C. Mandal:** Formal analysis, Validation, Resources. **A. K. Halder:** Writing – review &

editing, Conceptualisation, Visualisation, Validation, Supervision, Investigation, Formal analysis, Software.

Funding

This current work was funded by the Department of Science and Technology and Biotechnology, Govt. of West Bengal, India Vide Memo. 2027 (Sanc.)/STBT-11012 (19)/ 6/2023-ST SEC, dated 24-01-2024.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data is provided within the manuscript or supplementary information files

References

1. Delaney G, Jacob S, Featherstone C, Barton M (2005) The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer* 104:1129-37. <https://doi.org/10.1002/cncr.21324>.
2. Moding EJ, Kastan MB, Kirsch DG (2013) Strategies for optimizing the response of cancer and normal tissues to radiation. *Nat Rev Drug Discov* 12:526-42. <https://doi.org/10.1038/nrd4003>.
3. Pignon JP, Bourhis J, Domenge C, Designe L (2000) Chemotherapy added to locoregional treatment for head and neck squamous-cell carcinoma: three meta-analyses of updated individual data. MACH-NC Collaborative Group. *Meta-Analysis of Chemotherapy on Head and Neck Cancer*. *Lancet* 355:949-55.
4. Bentzen SM, Trotti A (2007) Evaluation of early and late toxicities in chemoradiation trials. *J Clin Oncol* 25:4096-103. <https://doi.org/10.1200/JCO.2007.13.3983>.
5. Ewing D (1998) The Oxygen Fixation Hypothesis: A Reevaluation. 21:355-61.
6. Jing X, Yang F, Shao C, Wei K, Xie M, Shen H, et al. (2019) Role of hypoxia in cancer therapy by regulating the tumor microenvironment. *Mol Cancer* 18:157. <https://doi.org/10.1186/s12943-019-1089-9>.
7. Wilson WR, Hay MP (2011) Targeting hypoxia in cancer therapy. *Nat Rev Cancer* 11:393-410. <https://doi.org/10.1038/nrc3064>.
8. Overgaard J (2007) Hypoxic radiosensitization: adored and ignored. *J Clin Oncol* 25:4066-74. <https://doi.org/10.1200/JCO.2007.12.7878>.
9. Liew LP, Shome A, Wong WW, Hong CR, Hicks KO, Jamieson SMF, et al. (2023) Design, Synthesis and Anticancer Evaluation of Nitroimidazole Radiosensitisers. *Molecules* 28.

<https://doi.org/10.3390/molecules28114457>.

10. Groelly FJ, Fawkes M, Dagg RA, Blackford AN, Tarsounas M (2023) Targeting DNA damage response pathways in cancer. *Nat Rev Cancer* 23:78-94. <https://doi.org/10.1038/s41568-022-00535-5>.
11. O'Connor MJ (2015) Targeting the DNA Damage Response in Cancer. *Mol Cell* 60:547-60. <https://doi.org/10.1016/j.molcel.2015.10.040>.
12. Paques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 63:349-404. <https://doi.org/10.1128/MMBR.63.2.349-404.1999>.
13. Mahaney BL, Meek K, Lees-Miller SP (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 417:639-50. <https://doi.org/10.1042/BJ20080413>.
14. Scully R, Panday A, Elango R, Willis NA (2019) DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* 20:698-714. <https://doi.org/10.1038/s41580-019-0152-0>.
15. Shao Z, Davis AJ, Fattah KR, So S, Sun J, Lee KJ, et al. (2012) Persistently bound Ku at DNA ends attenuates DNA end resection and homologous recombination. *DNA Repair (Amst)* 11:310-6. <https://doi.org/10.1016/j.dnarep.2011.12.007>.
16. Downs JA, Jackson SP (2004) A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol* 5:367-78. <https://doi.org/10.1038/nrm1367>.
17. Mimori T, Hardin JA (1986) Mechanism of interaction between Ku protein and DNA. *J Biol Chem* 261:10375-9.
18. Hartley KO, Gell D, Smith GC, Zhang H, Divecha N, Connelly MA, et al. (1995) DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell* 82:849-56. [https://doi.org/10.1016/0092-8674\(95\)90482-4](https://doi.org/10.1016/0092-8674(95)90482-4).
19. Ding Q, Reddy YV, Wang W, Woods T, Douglas P, Ramsden DA, et al. (2003) Autophosphorylation of the catalytic subunit of the DNA-dependent protein kinase is required for efficient end processing during DNA double-strand break repair. *Mol Cell Biol* 23:5836-48. <https://doi.org/10.1128/MCB.23.16.5836-5848.2003>.
20. Mahajan KN, Nick McElhinny SA, Mitchell BS, Ramsden DA (2002) Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair. *Mol Cell Biol* 22:5194-202. <https://doi.org/10.1128/MCB.22.14.5194-5202.2002>.
21. Niewolik D, Pannicke U, Lu H, Ma Y, Wang LC, Kulesza P, et al. (2006) DNA-PKcs dependence of Artemis endonucleolytic activity, differences between hairpins and 5' or 3' overhangs. *J Biol Chem* 281:33900-9. <https://doi.org/10.1074/jbc.M606023200>.
22. Niewolik D, Peter I, Butscher C, Schwarz K (2017) Autoinhibition of the Nuclease ARTEMIS Is Mediated by a Physical Interaction between Its Catalytic and C-terminal Domains. *J Biol Chem* 292:3351-65. <https://doi.org/10.1074/jbc.M116.770461>.

23. Fulop GM, Phillips RA (1990) The scid mutation in mice causes a general defect in DNA repair. *Nature* 347:479-82. <https://doi.org/10.1038/347479a0>.
24. Biedermann KA, Sun JR, Giaccia AJ, Tosto LM, Brown JM (1991) scid mutation in mice confers hypersensitivity to ionizing radiation and a deficiency in DNA double-strand break repair. *Proc Natl Acad Sci U S A* 88:1394-7. <https://doi.org/10.1073/pnas.88.4.1394>.
25. Curtin NJ (2012) DNA repair dysregulation from cancer driver to therapeutic target. *Nat Rev Cancer* 12:801-17. <https://doi.org/10.1038/nrc3399>.
26. Davidson D, Amrein L, Panasci L, Aloyz R (2013) Small Molecules, Inhibitors of DNA-PK, Targeting DNA Repair, and Beyond. *Front Pharmacol* 4:5. <https://doi.org/10.3389/fphar.2013.00005>.
27. Lee TW, Wong WW, Dickson BD, Lipert B, Cheng GJ, Hunter FW, et al. (2019) Radiosensitization of head and neck squamous cell carcinoma lines by DNA-PK inhibitors is more effective than PARP-1 inhibition and is enhanced by SLFN11 and hypoxia. *Int J Radiat Biol* 95:1597-612. <https://doi.org/10.1080/09553002.2019.1664787>.
28. Sibanda BL, Chirgadze DY, Blundell TL (2010) Crystal structure of DNA-PKcs reveals a large opening cradle comprised of HEAT repeats. *Nature* 463:118-21. <https://doi.org/10.1038/nature08648>.
29. Sibanda BL, Chirgadze DY, Ascher DB, Blundell TL (2017) DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair. *Science* 355:520-4. <https://doi.org/10.1126/science.aak9654>.
30. Yin X, Liu M, Tian Y, Wang J, Xu Y (2017) Cryo-EM structure of human DNA-PK holoenzyme. *Cell Res* 27:1341-50. <https://doi.org/10.1038/cr.2017.110>.
31. Cao R, Zeng H, Zhang H (2009) 3D-QSAR studies on a series of inhibitors docked into a new homology model of the DNA-PK receptor. *Curr Pharm Des* 15:3796-825. <https://doi.org/10.2174/138161209789649484>.
32. Ihmaid S, Al-Rawi J, Bradley C, Angove MJ, Robertson MN, Clark RL (2011) Synthesis, structural elucidation, DNA-PK inhibition, homology modelling and anti-platelet activity of morpholino-substituted-1,3-naphth-oxazines. *Bioorg Med Chem* 19:3983-94. <https://doi.org/10.1016/j.bmc.2011.05.032>.
33. Clapham KM, Rennison T, Jones G, Craven F, Bardos J, Golding BT, et al. (2012) Potent enantioselective inhibition of DNA-dependent protein kinase (DNA-PK) by atropisomeric chromenone derivatives. *Org Biomol Chem* 10:6747-57. <https://doi.org/10.1039/c2ob26035b>.
34. Maira SM, Stauffer F, Brueggen J, Furet P, Schnell C, Fritsch C, et al. (2008) Identification and characterization of NVP-BEZ235, a new orally available dual phosphatidylinositol 3-kinase/mammalian target of rapamycin inhibitor with potent in vivo antitumor activity. *Mol Cancer Ther* 7:1851-63. <https://doi.org/10.1158/1535-7163.MCT-08-0017>.
35. Mukherjee B, Tomimatsu N, Amancherla K, Camacho CV, Pichamoorthy N, Burma S (2012) The dual PI3K/mTOR inhibitor NVP-BEZ235 is a potent inhibitor of ATM- and DNA-PKcs-mediated DNA damage responses. *Neoplasia* 14:34-43. <https://doi.org/10.1593/neo.111512>.

36. Hardcastle IR, Cockcroft X, Curtin NJ, El-Murr MD, Leahy JJ, Stockley M, et al. (2005) Discovery of potent chromen-4-one inhibitors of the DNA-dependent protein kinase (DNA-PK) using a small-molecule library approach. *J Med Chem* 48:7829-46. <https://doi.org/10.1021/jm050444b>.
37. Zhao Y, Thomas HD, Batey MA, Cowell IG, Richardson CJ, Griffin RJ, et al. (2006) Preclinical evaluation of a potent novel DNA-dependent protein kinase inhibitor NU7441. *Cancer Res* 66:5354-62. <https://doi.org/10.1158/0008-5472.CAN-05-4275>.
38. Cano C, Saravanan K, Bailey C, Bardos J, Curtin NJ, Frigerio M, et al. (2013) 1-substituted (Dibenzo[b,d]thiophen-4-yl)-2-morpholino-4H-chromen-4-ones endowed with dual DNA-PK/PI3-K inhibitory activity. *J Med Chem* 56:6386-401. <https://doi.org/10.1021/jm400915j>.
39. Munck JM, Batey MA, Zhao Y, Jenkins H, Richardson CJ, Cano C, et al. (2012) Chemosensitization of cancer cells by KU-0060648, a dual inhibitor of DNA-PK and PI-3K. *Mol Cancer Ther* 11:1789-98. <https://doi.org/10.1158/1535-7163.MCT-11-0535>.
40. Shinohara ET, Geng L, Tan J, Chen H, Shir Y, Edwards E, et al. (2005) DNA-dependent protein kinase is a molecular target for the development of noncytotoxic radiation-sensitizing drugs. *Cancer Res* 65:4987-92. <https://doi.org/10.1158/0008-5472.CAN-04-4250>.
41. Fok JHL, Ramos-Montoya A, Vazquez-Chantada M, Wijnhoven PWG, Follia V, James N, et al. (2019) AZD7648 is a potent and selective DNA-PK inhibitor that enhances radiation, chemotherapy and olaparib activity. *Nat Commun* 10:5065. <https://doi.org/10.1038/s41467-019-12836-9>.
42. Goldberg FW, Finlay MRV, Ting AKT, Beattie D, Lamont GM, Fallan C, et al. (2020) The Discovery of 7-Methyl-2-[(7-methyl[1,2,4]triazolo[1,5-a]pyridin-6-yl)amino]-9-(tetrahydro-2H-pyran-4-yl)-7,9-dihydro-8H-purin-8-one (AZD7648), a Potent and Selective DNA-Dependent Protein Kinase (DNA-PK) Inhibitor. *J Med Chem* 63:3461-71. <https://doi.org/10.1021/acs.jmedchem.9b01684>.
43. Willoughby CE, Jiang Y, Thomas HD, Willmore E, Kyle S, Wittner A, et al. (2020) Selective DNA-PKcs inhibition extends the therapeutic index of localized radiotherapy and chemotherapy. *J Clin Invest* 130:258-71. <https://doi.org/10.1172/JCI127483>.
44. Zenke FT, Zimmermann A, Sirrenberg C, Dahmen H, Kirkin V, Pehl U, et al. (2020) Pharmacologic Inhibitor of DNA-PK, M3814, Potentiates Radiotherapy and Regresses Human Tumors in Mouse Models. *Mol Cancer Ther* 19:1091-101. <https://doi.org/10.1158/1535-7163.MCT-19-0734>.
45. Chaplin AK, Hardwick SW, Liang S, Kefala Stavridi A, Hnizda A, Cooper LR, et al. (2021) Dimers of DNA-PK create a stage for DNA double-strand break repair. *Nat Struct Mol Biol* 28:13-9. <https://doi.org/10.1038/s41594-020-00517-x>.
46. Liang S, Blundell TL (2023) Human DNA-dependent protein kinase activation mechanism. *Nat Struct Mol Biol* 30:140-7. <https://doi.org/10.1038/s41594-022-00881-w>.
47. Liang S, Thomas SE, Chaplin AK, Hardwick SW, Chirgadze DY, Blundell TL (2022) Structural insights into inhibitor regulation of the DNA repair protein DNA-PKcs. *Nature* 601:643-8. <https://doi.org/10.1038/s41586-021-04274-9>.
48. Hong CR, Liew LP, Wong WW, Dickson BD, Cheng G, Shome A, et al. (2024) Identification of 6-Anilino Imidazo[4,5-c]pyridin-2-ones as Selective DNA-Dependent Protein Kinase Inhibitors and Their

- Application as Radiosensitizers. *J Med Chem* 67:12366-85.
<https://doi.org/10.1021/acs.jmedchem.4c01120>.
49. Sushko I, Novotarskyi S, Korner R, Pandey AK, Rupp M, Teetz W, et al. (2011) Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *J Comput Aided Mol Des* 25:533-54. <https://doi.org/10.1007/s10822-011-9440-2>.
50. Mauri A. alvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints. In: Roy K, editor. *Ecotoxicological QSARs*. New York, NY: Springer US; 2020. p. 801-20.
51. Mitra S, Halder AK, Koley A, Ghosh N, Panda P, Mandal SC, et al. (2024) Unveiling structural determinants for FXR antagonism in 1,3,4-trisubstituted-Pyrazol amide derivatives: A multi-scale in silico modelling approach. *Comput Biol Med* 180:108991.
<https://doi.org/10.1016/j.compbiomed.2024.108991>.
52. Danishuddin, Khan AU (2016) Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discov Today* 21:1291-302. <https://doi.org/10.1016/j.drudis.2016.06.013>.
53. Lesani S, Tavalla M, Eslami G, Boozhmehrani MJ (2025) Quantitative Structure-Activity Relationship Modeling and Molecular Docking Studies of TgCDPK1 Inhibitors in *Toxoplasma gondii*. *Microbiologyopen* 14:e70039. <https://doi.org/10.1002/mbo3.70039>.
54. Ambure P, Aher RB, Gajewicz A, Puzyn T, Roy K (2015) "NanoBRIDGES" software: Open access tools to perform QSAR and nano-QSAR modeling. *Chemometrics and Intelligent Laboratory Systems* 147:1-13. <https://doi.org/https://doi.org/10.1016/j.chemolab.2015.07.007>.
55. Gramatica P (2007) Principles of QSAR models validation: internal and external. 26:694-701.
<https://doi.org/https://doi.org/10.1002/qsar.200610151>.
56. Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems* 145:22-9.
<https://doi.org/https://doi.org/10.1016/j.chemolab.2015.04.013>.
57. Tian T, Li S, Fang M, Zhao D, Zeng J (2024) MolSHAP: Interpreting Quantitative Structure-Activity Relationships Using Shapley Values of R-Groups. *J Chem Inf Model* 64:2236-49.
<https://doi.org/10.1021/acs.jcim.3c00465>.
58. Tosco P, Balle T (2011) Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. *J Mol Model* 17:201-8.
<https://doi.org/10.1007/s00894-010-0684-x>.
59. Tosco P, Balle T, Shiri F (2011) Open3DALIGN: an open-source software aimed at unsupervised ligand alignment. *J Comput Aided Mol Des* 25:777-83. <https://doi.org/10.1007/s10822-011-9462-9>.
60. Mondal I, Halder AK, Pattanayak N, Mandal SK, Cordeiro M (2024) Shaping the Future of Obesity Treatment: In Silico Multi-Modeling of IP6K1 Inhibitors for Obesity and Metabolic Dysfunction. *Pharmaceuticals (Basel)* 17. <https://doi.org/10.3390/ph17020263>.
61. Mitra S, Chatterjee S, Bose S, Panda P, Basak S, Ghosh N, et al. (2024) Finding structural requirements of structurally diverse alpha-glucosidase and alpha-amylase inhibitors through

validated and predictive 2D-QSAR and 3D-QSAR analyses. J Mol Graph Model 126:108640. <https://doi.org/10.1016/j.jmgm.2023.108640>.

62. Ghosh A, Panda P, Halder AK, Cordeiro M (2022) In silico characterization of aryl benzoyl hydrazide derivatives as potential inhibitors of RdRp enzyme of H5N1 influenza virus. Front Pharmacol 13:1004255. <https://doi.org/10.3389/fphar.2022.1004255>.
63. Eberhardt J, Santos-Martins D, Tillack AF, Forli S (2021) AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model 61:3891-8. <https://doi.org/10.1021/acs.jcim.1c00203>.
64. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455-61. <https://doi.org/10.1002/jcc.21334>.
65. Mitra S, Halder AK, Ghosh N, Mandal SC, Cordeiro M (2023) Multi-model in silico characterization of 3-benzamidobenzoic acid derivatives as partial agonists of Farnesoid X receptor in the management of NAFLD. Comput Biol Med 157:106789. <https://doi.org/10.1016/j.combiomed.2023.106789>.
66. Roy K, Kar S (2014) The rm2 metrics and regression through origin approach: reliable and useful validation tools for predictive QSAR models (Commentary on 'Is regression through origin useful in external validation of QSAR models?'). Eur J Pharm Sci 62:111-4. <https://doi.org/10.1016/j.ejps.2014.05.019>.
67. Todeschini R, Consonni V (2008) Handbook of molecular descriptors. John Wiley & Sons, Place.

Figures

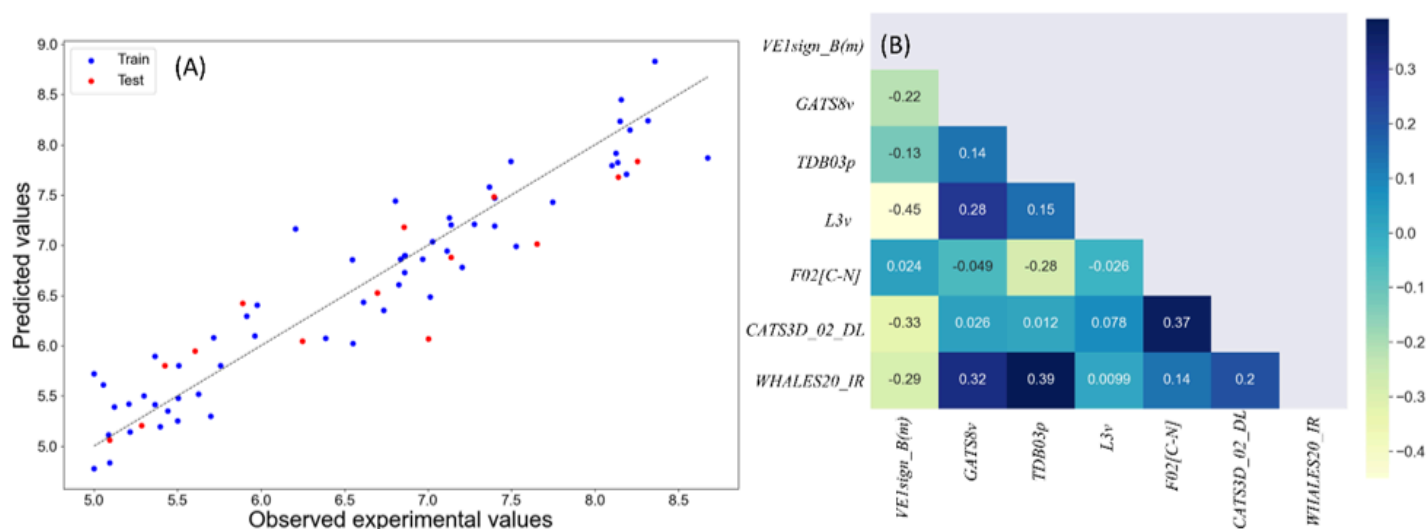


Figure 1

(A) Observed vs predicted pIC_{50} values and (B) correlation matrix of 2D-QSAR model.

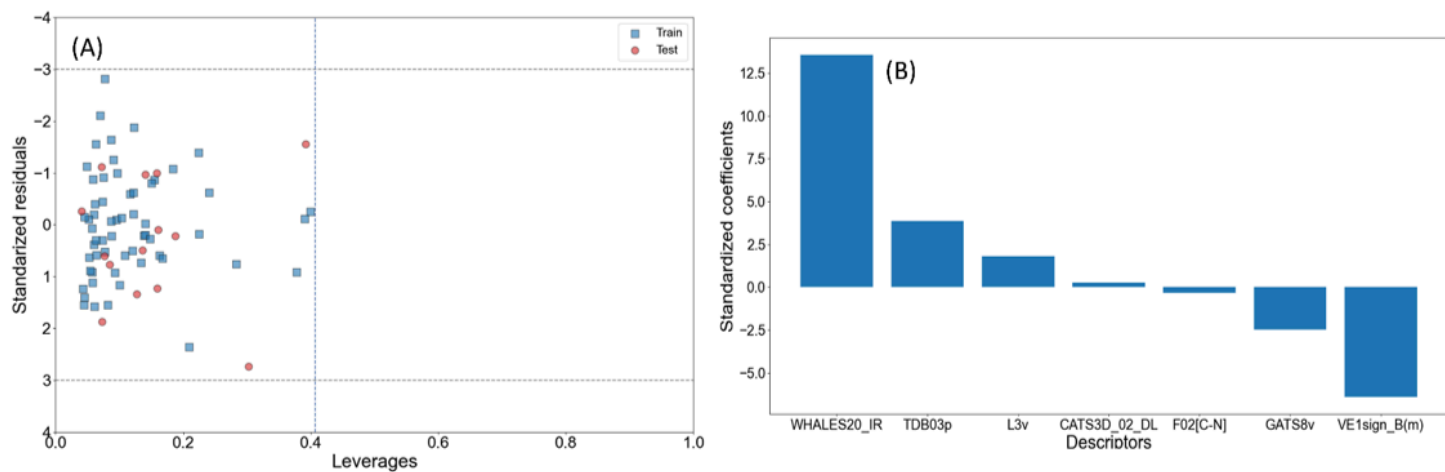


Figure 2

(A) Williams plot for model M1 and (B) relative significance of the descriptors of model M1 with respect to standardized coefficient values.

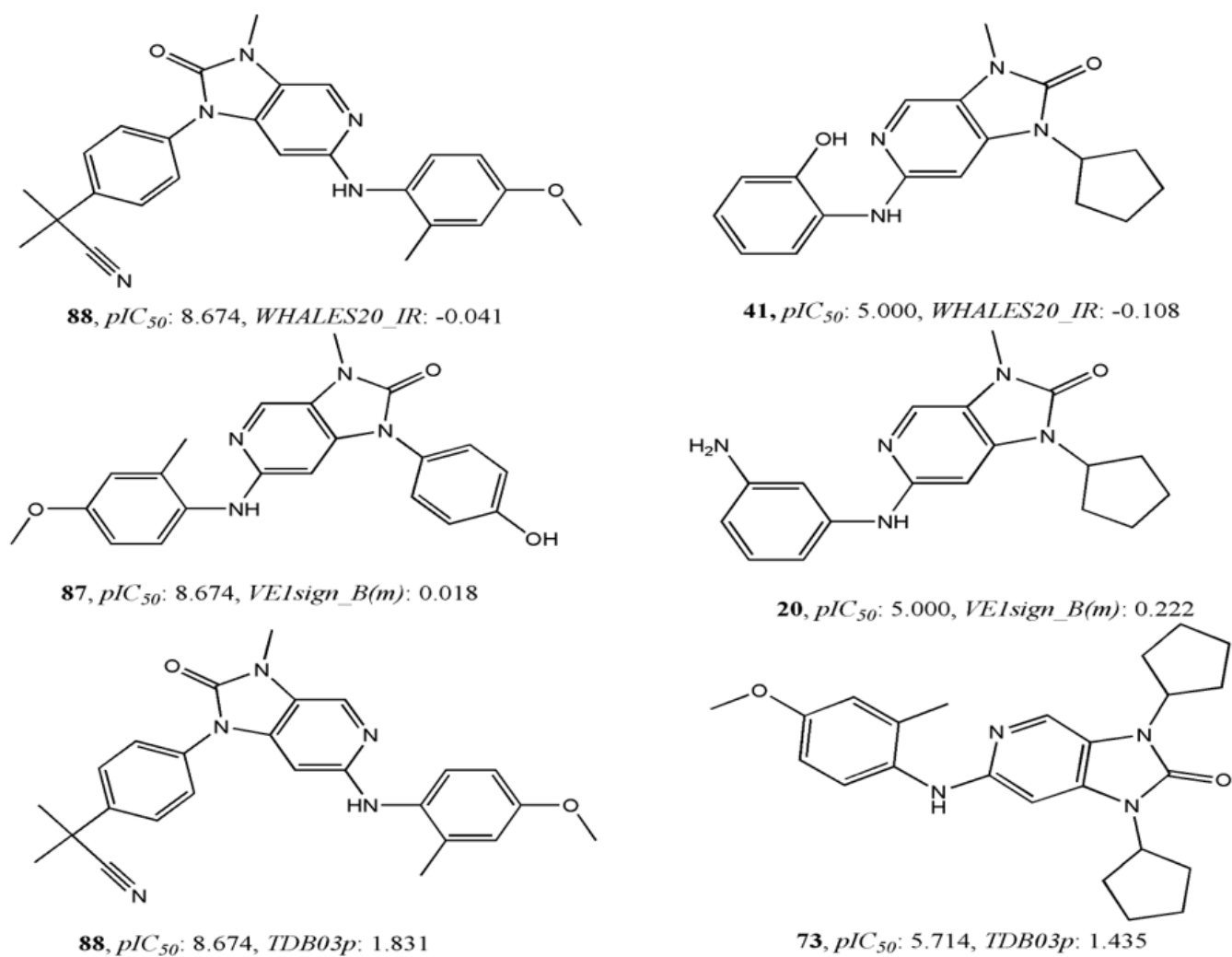


Figure 3

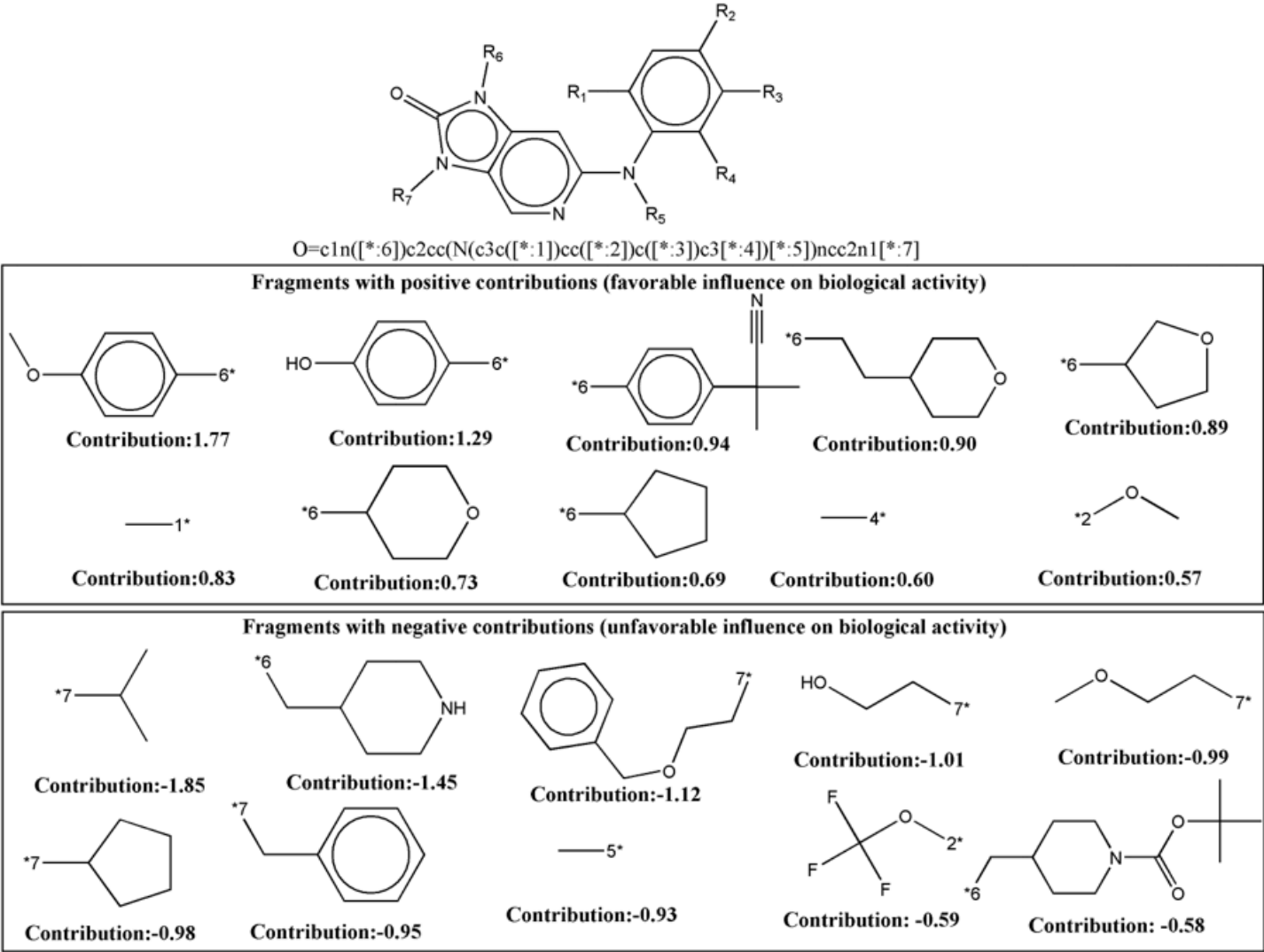


Figure 4

R-group analyses from MolSHAP

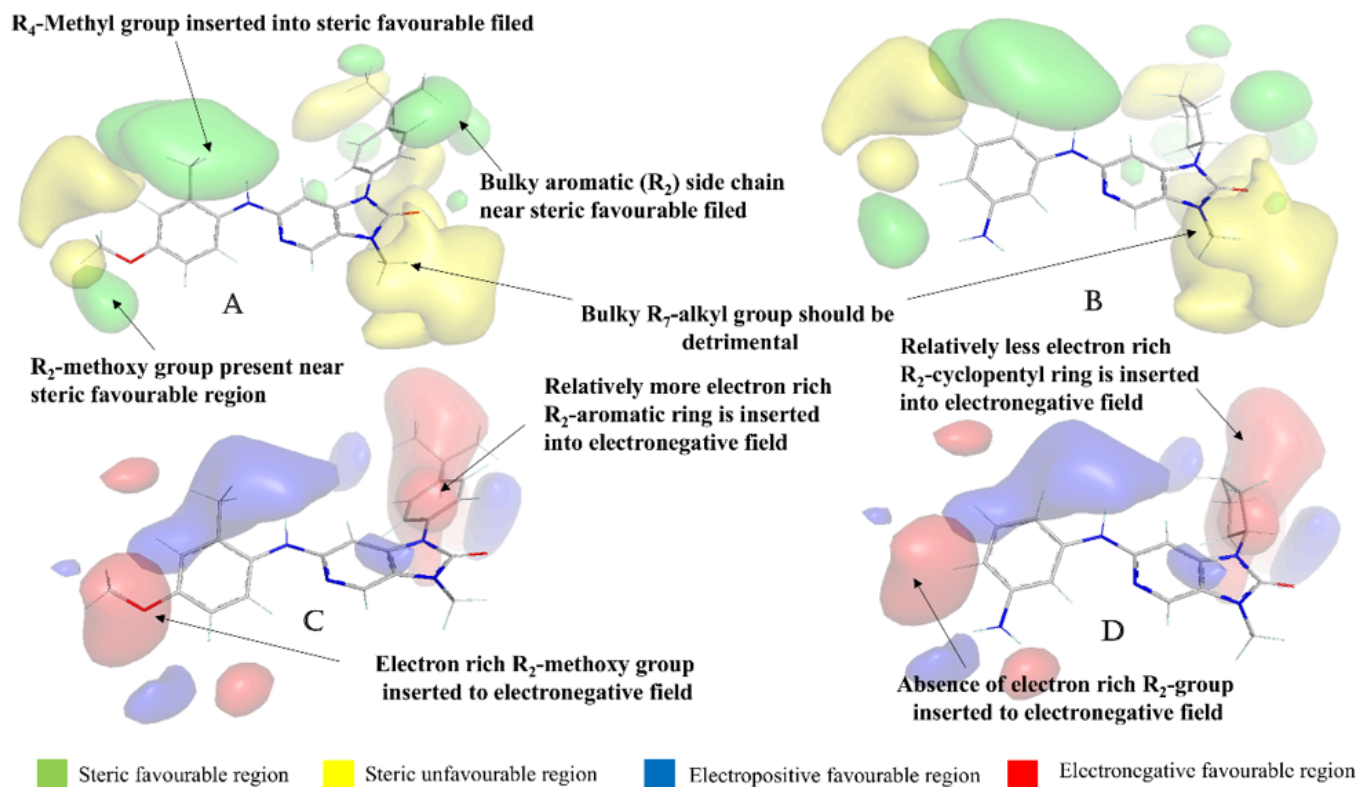


Figure 5

Isocontour maps from 3D-QSAR modelling analyses of compounds 88 and 41 with respect to steric and electrostatic regions

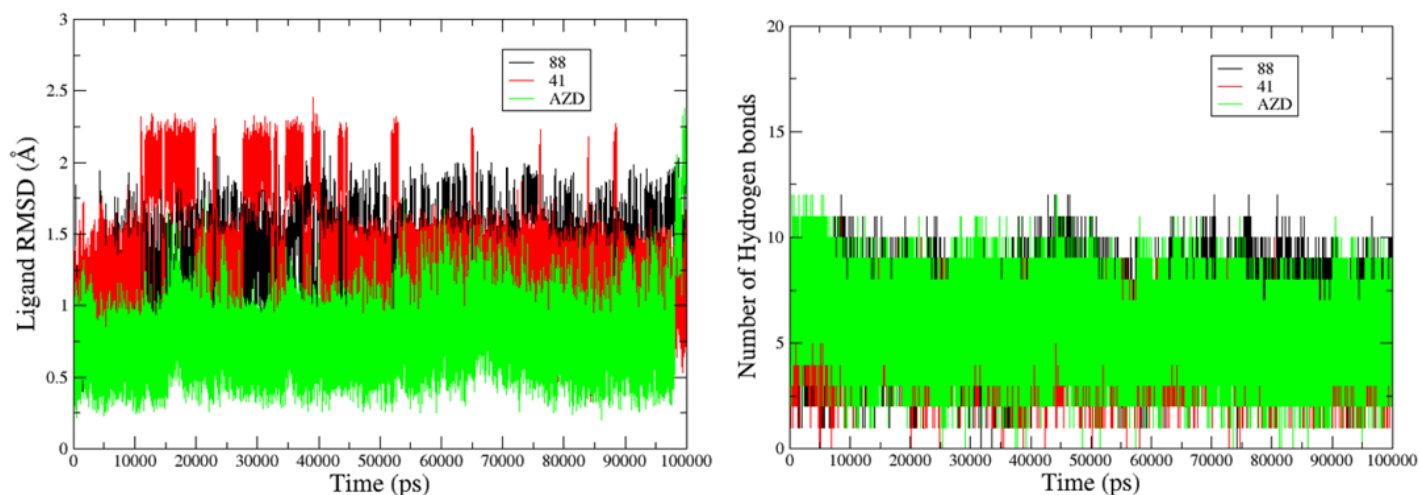


Figure 6

Plots obtained from trajectory analyses (A) ligand RMSD and (B) number of hydrogen bonds formed between amino acids and ligands

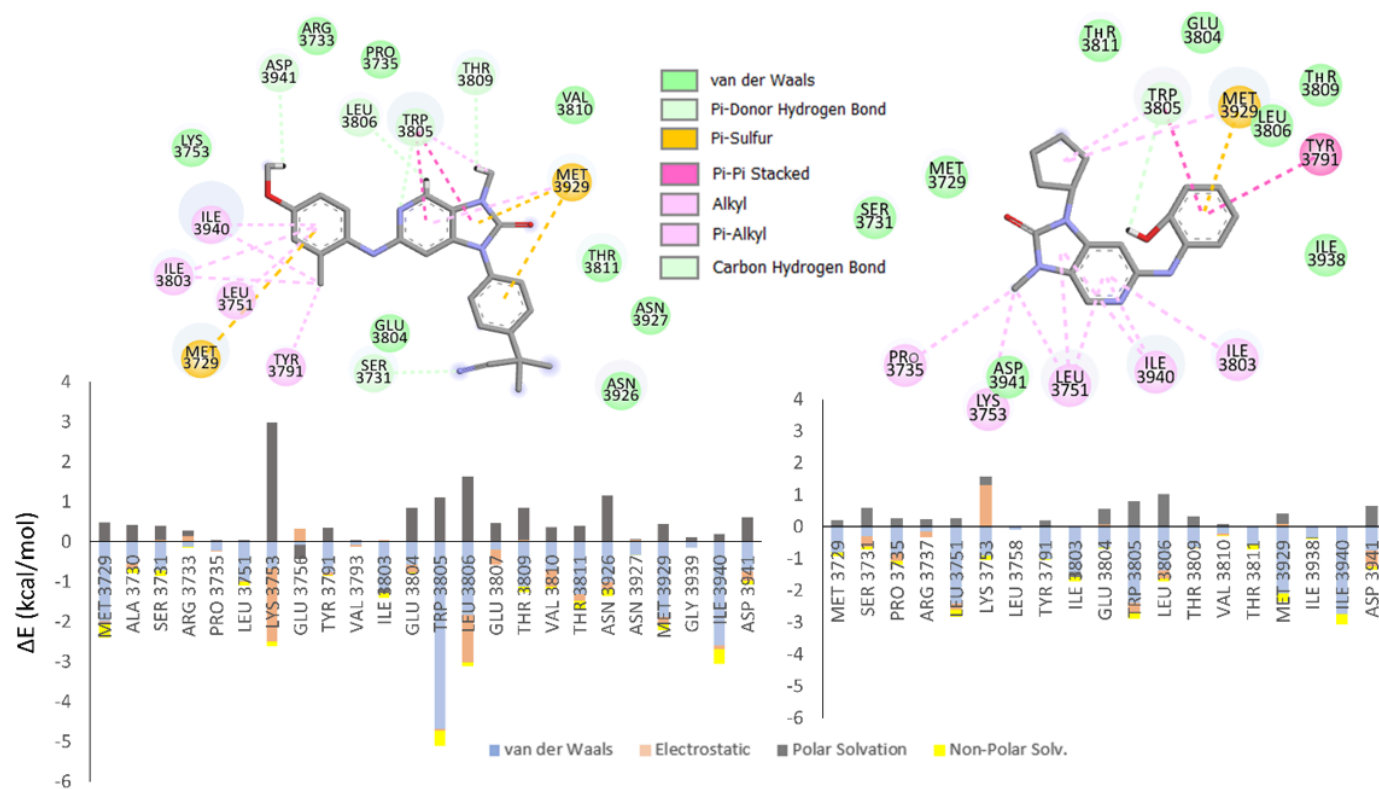


Figure 7

(A) The receptor-ligand interactions for the 88-complex (left) and the 41-complex (right) extracted from the last frame of the MD simulations. (B) Per-residue decomposition analyses of compounds 88 (left) and 41 (right)

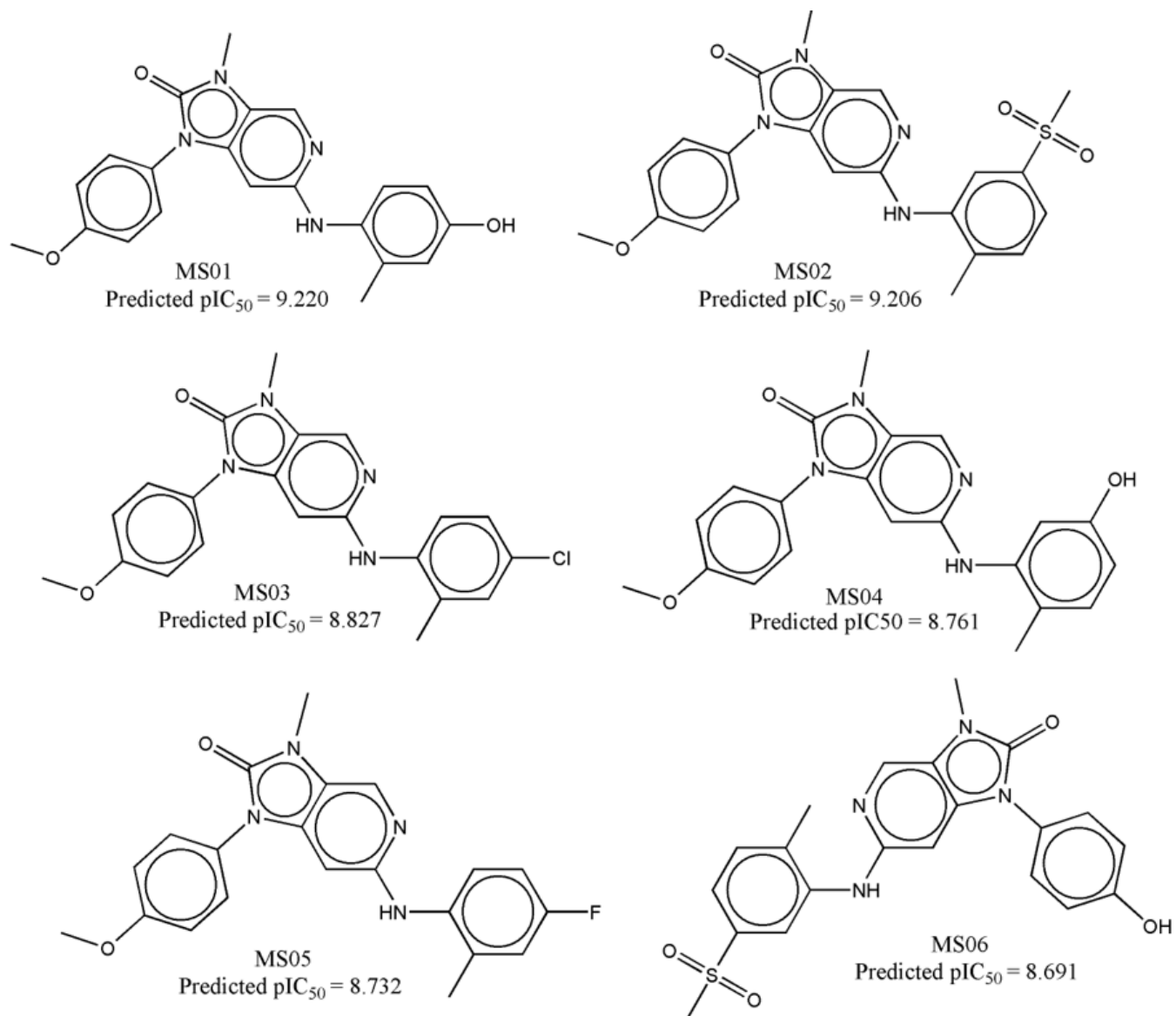


Figure 8

Novel ligands designed by MolSHAP

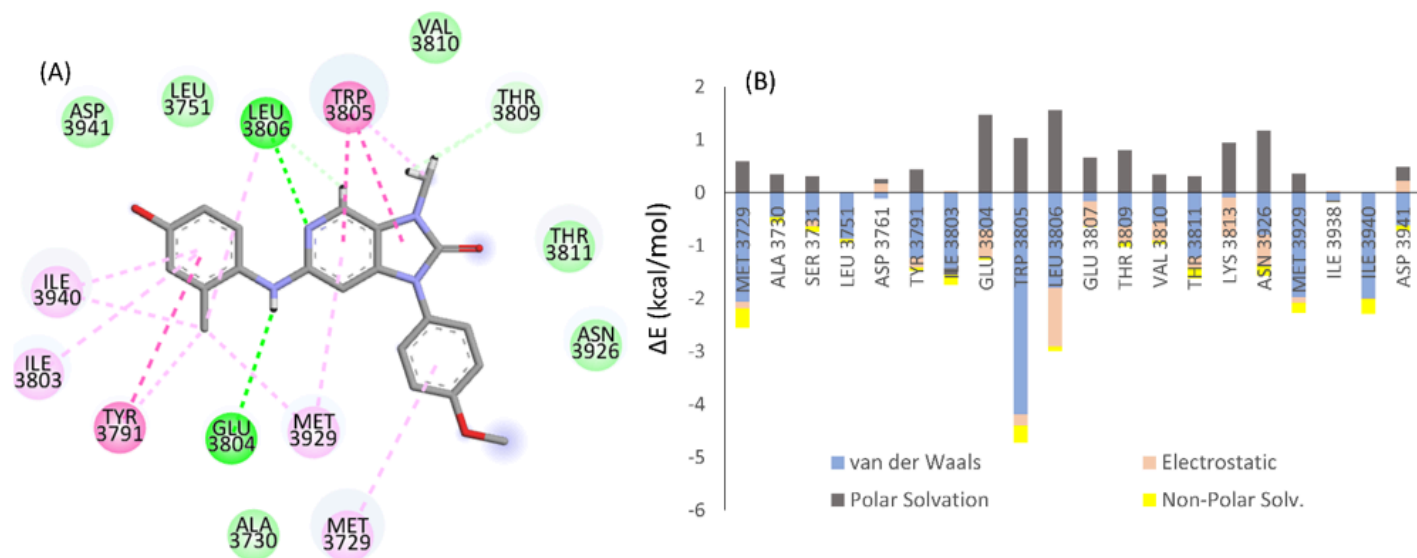


Figure 9

(A) Receptor ligand interaction between MS01 and receptor; (B) Per-residue decomposition analyses of MS01

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinformation.docx](#)