



International Journal of Pharmacy and Analytical Research (IJPAR)

IJPAR | Vol.14 | Issue 4 | Oct - Dec -2025

www.ijpar.com

ISSN: 2320-2831

DOI: <https://doi.org/10.61096/ijpar.v14.iss4.2025.1266-1279>

Review

QSAR Modeling of RET Kinase Inhibitors Based on Structure from 49 Various 5, 6-Fused Bicyclic Hetero aromatic Cores for Patent-Driven Verification.

Dr. Rahul B. Lovhare^{*1}, Dinesh N. Pawar², Jagruti R. Shewale³, Bhavana R. Vasave⁴, Shradha K. Patel⁵

^{*1}Professor, Department of Pharmaceutical Chemistry, Shatabdi Institute of Pharmacy, Bamdod, Nandurbar-425412 (MS), India.



^{2,4}Associate Professor, Department of Pharmaceutics, Shatabdi Institute of Pharmacy, Bamdod, Nandurbar-425412 (MS), India.

³Associate Professor, Department of Pharmaceutical Chemistry, Shatabdi Institute of Pharmacy, Bamdod, Nandurbar-425412 (MS), India.

⁵Lecturer of Pharmacy, Shatabdi Institute of Pharmacy, Bamdod, Nandurbar-425412 (MS), India

*Author for Correspondence: Rahul B. Lovhare

Email: pharmy786@gmail.com

	Abstract
Published on: 18.12. 2025	<p>The receptor tyrosine that is rearranged during transfection (RET) Kinase is essential for the growth of the enteric genitourinary tissues and the nervous system. 1 Research on cancer found that RET is a carcinogenic factor in nonsmall-cell lung cancer (NSCLC), 2 RET changes show a strong correlation with different types of cancer. 3 The turning on of down- stream pathways like MAPK and PI3K/AKT, which are found in the RET fusion protein tyrosine kinase, originates from Autophosphorylation and ligand-independent homodimerization brought about by the intrachromosomal rearrangement of RET. 3 The optimal QSAR model exhibited high predictive performance, with R² (of training data) and Q² (of test data) values of 0.801 and 0.794, respectively, effectively predicting known inhibitors. The optimal model was doubly verified by patent-filed RET inhibitors as the out-of-set data to demonstrate acceptable residual analysis results. Moreover, feature importance analysis of the QSAR model outlined the impact of substituent characteristics on the inhibitory activity within the 5,6- fused bicyclic heteroaromatic core structures. Furthermore, the relationship between structure and inhibitory activity was successfully applied to the RET screening of known clinical and nonclinical kinase inhibitors to afford accurate off-target prediction.</p>
Published by: Futuristic Publications	
<p>2025 All rights reserved.</p>  <p>Creative Commons Attribution 4.0 International License.</p>	
	<p>RET kinase inhibitors; QSAR modeling; 3D-QSAR; 5,6-fused bicyclic heteroaromatic cores; molecular docking; structure–activity relationship (SAR); kinase inhibitor design; virtual screening; patent validation; off-target prediction.</p>

1. INTRODUCTION

The rearranged during transfection (RET) receptor tyrosine kinase plays a crucial role in the development of the enteric nervous system and genitourinary tissues.¹ Cancer research identified RET as an oncogenic driver in non small-cell lung cancer (NSCLC),² RET alterations exhibit significant associations with various malignancies.³ The activation of down- stream pathways such as PI3K/AKT and MAPK, which are observed in the RET fusion protein tyrosine kinase, stems from Ligand independent homo dimerization and auto phosphorylation resulting from the intra chromosomal rearrangement of RET.³ Cancers related to RET alterations encompass lung adenocarcinoma, colon adenocarcinoma, medullary carcinoma of the thyroid gland, cutaneous melanoma, and melanoma; of the various alterations, RET point mutations and fusions are prevalent,⁴ both of which are closely linked to unfavorable prognoses. Notably, the occurrence of brain metastasis among RET-altered patients, RET(+) stands at 46%, markedly higher than that observed in RET(-) counterparts.⁵ Although rare, RET rearrangements delineate a distinct subtype of metastatic colorectal cancer characterized by poor prognosis under conventional therapies, warranting specialized management approaches.⁶ Following pivotal clinical trials, multikinase inhibitors (MKIs) and selective RET inhibitors have received FDA approval for the management of radioiodine-refractory differentiated thyroid cancer or metastatic medullary thyroid cancer (MTC).⁷ Notably, Se 1 percatinib (also known as Reteam or LOXO- 292) has been granted Accelerated Approval by the FDA.⁸ This approval applies to the use of the RET kinase inhibitor in patients with metastatic NSCLC harboring RET fusions, as well as those with MTC harboring RET mutations. Recent data demonstrate the efficacy of Se 1 percatinib over standard care therapy in cancer types characterized by recurrent RET alterations.⁹ Although RET inhibitors (Figure 1) have been continuously reported, quantitative structure–activity relationship (QSAR) models and other predictive models have not been sufficiently studied. To our best knowledge, reported 3D-QSAR models for RET kinase inhibitors used a focused data set sharing on common core structure (scaffold) such as indolin-2-one or pyrrole [2, 3-*d*] pyrimidine. However, the ideal.

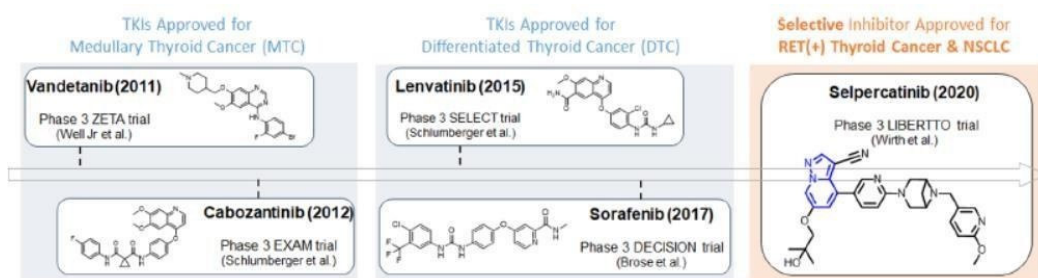


Figure 1. Summary of FDA-approved RET inhibitors. Abbreviation: TKI, tyrosine kinase inhibitor; DTC, radioiodine- refractory differentiated thyroid cancer; MTC, metastatic medullary thyroid cancer; NSCLC, non-small-cell lung cancer.

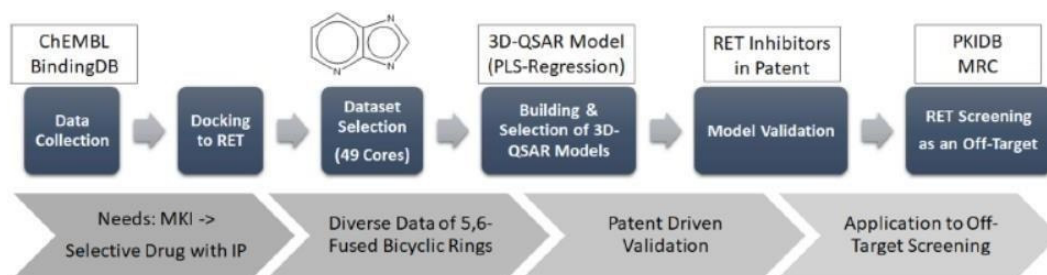


Figure 2. Overall workflow of this study

Design of kinase inhibitors necessitates the replacement and variation of both core structures and side chains to

optimize the synthetic method, drug efficacy, and ADMET. Heterocycle substitution, among several scaffold hopping strategies, offers a direct means of modifying the core structure of a known bioactive compound, which is especially beneficial for molecules posing synthetic hurdles. This method entails the placement of carbon atoms with heteroatoms within the core ring of drug molecules, with the goal of improving their physicochemical properties and pharmacokinetic profile. Moreover, recently reported considerable RET inhibitors commonly possess a 5, 6-fused bicyclic ring that interacts with the hinge region of RET kinase, and it is termed a hinge binder. Particularly, a 5, 6-fused bicyclic ring was more frequently used in RET-selective inhibitors than in multikinase inhibitors.

Therefore, we believe that researchers involved in the discovery of kinase inhibitors have a strong need to investigate the sophisticated SAR between these 5, 6-fused bicyclic derivatives. In other words, we are motivated to develop 3DQSAR models for explaining 5, 6-fused bicyclic hinge binders, which have increasingly entered nonclinical and clinical studies for RET (+) patients. For this purpose, herein, we collected RET inhibitors from public databases, classified them based on the ring system of their hinge binders, and compiled a comprehensive data set of 49 different 5, 6-fused bicyclic heteroaromatic rings with 952 inhibitors as a data set. In this study, we elucidate the subtle SAR difference among 5, 6-fused bicyclic derivatives through molecular docking simulations for the 3D alignment of the data set and 3D-QSAR modeling. Furthermore, the evaluation of the best QSAR model is presented to provide guidance on the rational design of selective and potent RET inhibitors. In particular, RET inhibitors filed in the patent were used as an out-of-set third data, which never include either training or test data. In parallel, the RET model was applied for RET screening as an off-target prediction of known kinase inhibitors. The workflow of this study is illustrated in Figure 2.

2. MATERIALS AND METHODS

2.1. Data Collection and Preparation.

Information on RET kinase inhibitor activity was retrieved from the ChEMBL (version 33), Binding DB, and Escape databases. Information from 4763 RET inhibitor patents and other sources was compiled and curated using the KNIME workflow. All of the preprocessing steps, such as target selection and bioactivity transformation, were performed using the KNIME analytical platform. The selection criteria included the target protein, assay details, and activity value type. The evaluation of activity data within the target category, conducted using wild-type, non cell-based assays, employed specific criteria to focus on data points related to target proteins, such as Kinesin-1 heavy chain/tyrosine-protein kinase receptor RET, proto-oncogene tyrosine-protein kinase receptor RET, and tyrosine-protein kinase receptor RET, where inhibitory concentration (IC₅₀) values were measured in nanomolar (nM) units. Compounds with activity annotations featuring blank fields or values containing “<” or “>” were excluded. In order to standardize the data and facilitate analysis, IC₅₀ values were converted into their negative logarithmic form using the formula $pIC_{50} = 9 - \log(IC_{50})$.

2.2. Molecular Docking Simulations.

Every ligand for molecular docking simulations was prepared using the Ligprep package from Schrodinger. 2D-to-3D conversion includes options for neutralizing compounds under the pH range of 7 \pm 2 and not generating tautomer's. The resulting structures of the 4763 compounds were docked to the RET protein. The crystal structure of the RET complex with inhibitors is available in the Protein Data Bank (PDB ID: 7DUA), and it was arranged via protein preparation (at pH 7 and the OPLS3e force field) in the Schrodinger Suite 2019-1. The protein preparation involves the process of adding hydrogen atoms.

Table 1. Filtering Criteria for Virtual Screening

	SlogP	TPSA	ExactMW	Num LipinskiHBA	Num LipinskiHBD	Num RotatableBonds	Num AmideBonds	Num HeteroAtoms	Num HeavyAtoms
lower bound	-1.4883	28.68	191.0695	2	1	0	0	2	14
upper bound	6.10482	170.23	550.2241	11	6	9	4	14	41

	NumRings	Num AromaticRings	Num AliphaticRings	Num Aromatic Heterocycles	Num Aliphatic Heterocycles
lower bound	2	2	0	1	0
upper bound	8	6	4	5	4

The criteria used for filtering were based on characteristics such as Slog P, TPSA, molecular

weight, number of Lipinski hydrogen-bond acceptors and donors, number of rotatable bonds, number of amide bonds, number of heteroatoms, number of rings, number of aromatic rings, number of aliphatic rings, number of aromatic heterocycles, and number of aliphatic heterocycles. (PROPKA) and removing unnecessary chains and water molecules, and an OPLS3e force field was applied to minimize the protein structure. The protein molecule underwent controlled minimization process to refine its geometry, achieving an RMSD of 0.3 Å. The grid was positioned at the core of the workspace ligand. The partial charge cutoff was set to 0.25, the van der Waals radius scaling factor was set to 1.0, and the docking box of size 20 Å × 20 Å × 20 Å was defined around the centroid of the ligand. Two spherical areas for the predicted binding site were determined through hydrogen-bonding interactions: (1) the oxygen of the carbonyl group of GLU805 as the H-bond donor and (2) the hydrogen of the amine group of ALA807 as the H-bond acceptor. Using Maestro's default docking settings, Glide was used for docking with default values of partial charge cutoff and van der Waal's radius scaling factor of 0.15 and 0.8, respectively. Ligand docking was run in the standard precision mode matching all of the grid constraints set previously, and the maximum number of poses per ligand was set to five. The docking model was validated by the RMSD calculation (0.0162 Å) between the X-ray ligand and redocked ligand and their pose comparison.

2.3. Substructure-Based Data-Set Selection.

The method for selecting the data set involved the similarity analysis of docking poses based on the literature. The docking pose that formed hydrogen bonds with the hinge residue ALA807 or GLU805 was selected. A list of 49 substructure filters involving fused bicyclic heteroaromatic rings, including pyrrole, pyrazine, pyrimidine, pyridine, and imidazole, was applied to remove molecules likely to be nonreactive.

2.4. Developing 3D-QSAR Models.

For QSAR modeling, PHASE implemented in Maestro 11.9 was used. While *receptor-guided alignment* of data sets is commonly employed in 3D-QSAR modeling, PHASE is specifically designed for *pharmacophore-based* 3D-QSAR modeling. Considerable QSAR studies using PHASE reported the general method consisting of (1) conformer generation with a reasonable number of conformers within available computing resources, (2) building pharmacophore hypothesis with different feature sets and number of points, (3) selection of the best pharmacophore model using scoring functions, and then (4) final building of the 3D-QSAR model using the best pharmacophore model. However, the large and heterogeneous data set in this study prevented implementation of the above-mentioned typical process, mainly owing to (1) constraints in conformer ensemble utilization with a large enough number of conformers (e.g., more than 952 inhibitors × 100 conformers) and (2) molecular superimposition inconsistencies, which arise either between substituents when bicyclic cores are fixed or between cores when substitution positions are fixed. Therefore, *receptor-guided alignment* was performed for building QSAR models using docking poses. Using the alignment from the best docking poses, various models were generated by varying the grid spacing (1.0, 1.5, and 2.0 Å), partial least-squares (PLS) factors (up to 10), and different data splitting methods, such as LMO (leave many out) or LOO (leave one out). The prepared data set was randomly divided into a training set and a test set at a 9:1 ratio. The PLS method was used for creating the QSAR model. For model generation, the $|t\text{-value}|$ was set below 2.0, and the grid spacing was 1.0 Å. Model validation was conducted using 10-fold cross-validation (10-CV), in which 306 compounds were excluded from the training set. The models were selected based on the criteria of low root-mean-square error (RMSE) and standard deviation (SD) values; RMSE below 0.8 and SD 0.7.

2.5. Out-of-Set Validation of the Optimal QSAR Model.

The out-of-set molecules consisted of 190 inhibitors and inhibitory activity (IC₅₀) against wild-type RET kinase, which were collected from the patent (US 10,807,986 B2). In Maestro, 2D ligands are neutralized and converted into 3D structures through ligand preparation without considering tautomeric. Ligand docking was performed using the extra precision mode without constraints, generating multiple docking poses for each molecule. Activity was predicted using QSAR, and the results were selected based on docking poses and the residual difference between the experimental activity and prediction.

2.6. Application of the Predictive QSAR Model into Virtual Screening.

The criteria used for filtering the kinase inhibitors, excluding RET inhibitors, were based on characteristics such as Slog P, TPSA, and molecular weight (Table 1). Public databases including PKIDB and MRC were used as screening libraries. Physicochemical data for the molecules used in 3D-QSAR modeling were revalidated using the KNIME workflow and subsequently employed for filtering purposes. A total of 459 compounds could be docked against the RET protein following the same configuration as described in Section 2.2. The 3D-QSAR model was then used as a three dimensional query to distinguish the hits with the highest predicted RET inhibitory activity. Two methods were employed to select the top 30 data sets including the structure information, docking score, actual activity (pIC₅₀), and predicted activity (pIC₅₀) most closely resembling a reference molecule based on their interactions. First, a specific

cutoff for the predicted activity was applied, followed by an evaluation of the interaction between docking poses and the target protein. Second, prioritization was based on rankings derived from the docking scores and predicted activity values.

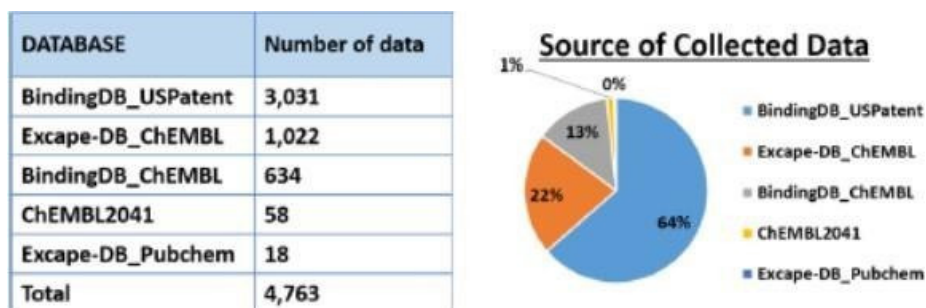


Figure 3. Composition of the initial collected data.

3. RESULTS

3.1. Activity Prediction of a Broad Range of 5, 6-Fused

Heteroaromatic Compounds. As mentioned in the histories of RET inhibitors above, the selectivity between RET and other receptor tyrosine kinases is important for clinical development, especially in terms of drug resistance. However, to our best knowledge, there are no reports providing the design logic for discriminating between MKIs and RET- selective inhibitors along with predicted potency. In this situation, the features of the RET-selective inhibitors can be a good clue to predict them. Notably, 5, 6 -fused heteroaromatic rings are frequently observed in both selective RET inhibitors and recently reported patents. Therefore, we expected that QSAR modeling with these compounds containing 5, 6 -fused heteroaromatic cores could come closer to the drug design logic (for selectivity and potency). When we collected RET inhibitors with enzymatic IC₅₀ values from five different data sources (Figure 3), a considerable number of compounds out of the 4763 inhibitors featured 5, 6-fused heteroaromatic rings in the substructure analysis using the KNIME workflow. Furthermore, while typical QSAR models use around 100 data sets with pIC₅₀ values in the range of about 3 (1000-fold gap between the best and worst data), we could acquire a data set that is ten times larger and has an activity range a thousand times wider than them (size: 4763; IC₅₀: <10 mM to >0.1 nM; pIC₅₀ range: >6) as shown in the Distribution of RET inhibitory activities as pIC₅₀ values (Figure 4). In sequence, we could generate 3D conformations of the 5, 6-fused heteroaromatic compounds and conduct their molecular docking simulations. The docking poses resulting from the simulations are approximate to bioactive conformation and also provide the superimposition (3D alignment) of the data set. After substructure filtering of the aligned conformers was conducted through the docking simulations, 49 different 5, 6-fused bicyclic heteroaromatic rings as substructures could be observed in the RET inhibitors (Supporting Information). Hierarchical clustering presented structural relevance between these 49 substructures, 41 and 10 representative structures were chosen as centroids from the 49 substructures (Figure 5). The substructure filtering process resulted in 952 ligands remaining, corresponding to the 3399 poses identified for the docking simulation. A total of 3399 pairs of structures and pIC₅₀ values were selected and used to construct the 3D-QSAR model. The finalized data set of 3399 conformers was randomly divided into the training set and test set at a 9:1 ratio. Then, the training set was fitted to a partial linear square (PLS) regression equation, which was generated based on the interaction features determined by the distance from each atom of the data set. The test set validated each PLS model. The top five models were selected based on the statistical metrics, particularly RMSE, which means the deviation between the actual and predicted pIC₅₀ values (Table 2). In other words, the RMSE denotes the root-mean-square error between the actual and predicted activities within a used data set and has the same unit with an SD of actual pIC₅₀ values. The optimal model presented an RMSE of 0.670 (log NM), an R² of 0.801, and a correlation coefficient of 0.891 between the experimental and predicted pIC₅₀. Hydrophobicity (H) is the most contributing feature in every model, followed by the electronic (E) feature and the hydrogen-bonding donor (D) feature in that order.

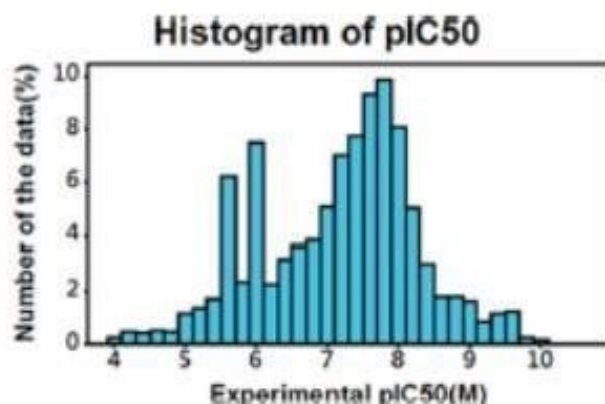


Figure 4. Distribution of the pIC50 value of collected RET inhibitors.

3.2. Evaluation of the Optimal QSAR Model.

This accurate result also encouraged us to characterize individual molecules in the out-of-set data. We selected the three most active and three most inactive molecules (Table 3). Very fortunately, the prediction of the subnanomolar ($pIC_{50} < 9$) active compounds was highly precise up to 0.03 pIC_{50} gap, demonstrating this model's capability for picomolar prediction. Similarly, millimolar activity (pIC_{50} close to 3–4) was also precisely predicted to demonstrate the predictable potency scope of the model. The reliable RMSE and R^2 encourage us to further evaluate the optimal model. Thus, we needed another out-of-set (third data set), neither training nor test data. For this purpose, we recruited the primary patent of TAS/HM06, a clinical phase II drug (US 10,807,986 B2). When we conducted docking simulations, 30 inhibitors out of 170 patent compounds satisfied the reliable docking poses with hydrogen bonding to ALA807 of the RET hinge region. The pIC_{50} values of the inhibitors were predicted using the optimal QSAR model to evaluate the practical utility of the model (see also the values in the Supporting Information). Clearly, their inhibitory activities presented a reasonable difference between actual and predicted activity values in the scatter plot of Figure 6. Moreover, we conducted residual analysis using the experimental and predicted values as shown in the following equation:

$$r_i = e_i / s(e_i) = e_i / \text{RSE} \sqrt{1 - h_{ii}}$$

e_i = the i th residual

RSE = the residual standard error of the model

h_{ii} = the leverage of the i th observation

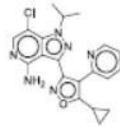
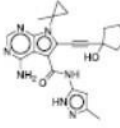
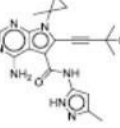
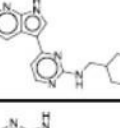
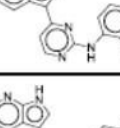
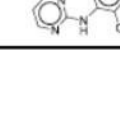
In general, even if a high correlation along with a high accuracy (high R^2) is observed between the predicted and actual values, it is difficult to trust the prediction results if the residuals follow any patterns. The analyzed residuals of Figure 7 did not show any pattern of consistent increase or decrease for every data point, thereby proving their reliability. Furthermore, the residual plot demonstrated rare outliers beyond the cutoff range (-2.0 - $+2.0$) of acceptable prediction.

3.3. Feature Analysis and 3D Visualization of RET–Inhibitor Interaction.

In machine learning-based prediction models, model interpretation, which explains how well the predictions are performed, is just as important as prediction accuracy. For this purpose, it is crucial to explain the important features involved in making predictions in a way that is understandable to researchers. Therefore, we tried to analyze the 3D molecular interaction features through the visualization of three representative inhibitors chosen from both the active ($pIC_{50} < 9$) and inactive (pIC_{50} close to 3–4) groups in Table 4. The feature distinctions between active and inactive compounds were analyzed using contour maps (a form of a 3D feature map) of the optimal QSAR model. The unit of the contour maps is a cube (1 Å³), which indicates whether the space is favorable or unfavorable areas for respective features such as H-bond acceptors and donors, hydrophobic groups, and electron-withdrawing (EW) elements. First of all, the critical hydrogen bond of RET inhibitors with the hinge region (Glu805 and Ala807) was perfectly matched between the 3D location of the H-donor feature with a positive coefficient (blue box) and the docking pose of the most potent compound (Figure 8a). Even if both the most potent and the least potent compounds commonly have hydrogen bonds (H-bond), the H-donor feature discriminated

between their subtle locations, leading to a mismatch between the H-donor feature and the H-bond in the

Table 3. Representative Compounds among the Data Set A The experimental activity and predicted activity are both expressed in the form of pIC50 values.

Data No.	Group	Structure	Experimental activity	Predicted activity
591	Active		10.770	9.224
377	Active		10.398	10.183
371	Active		10.155	10.126
44	Inactive		4.000	5.209
46	Inactive		4.000	4.840
83	Inactive		4.000	5.005

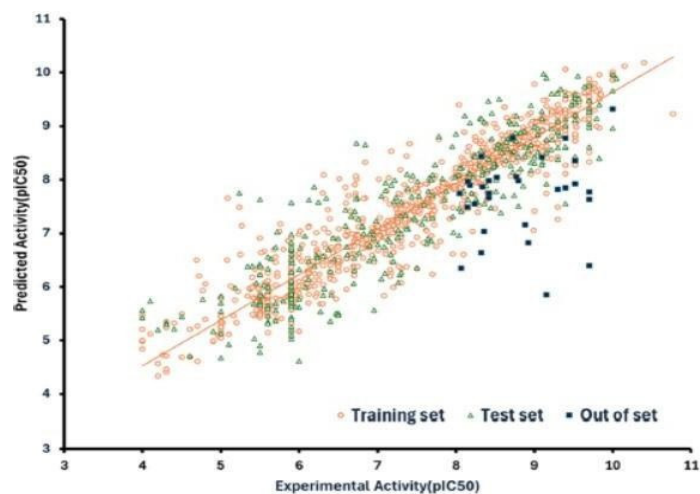


Figure 6. Scatter plot of the experimental activity and predicted activity of the training set, test set, and out-of-set molecules.

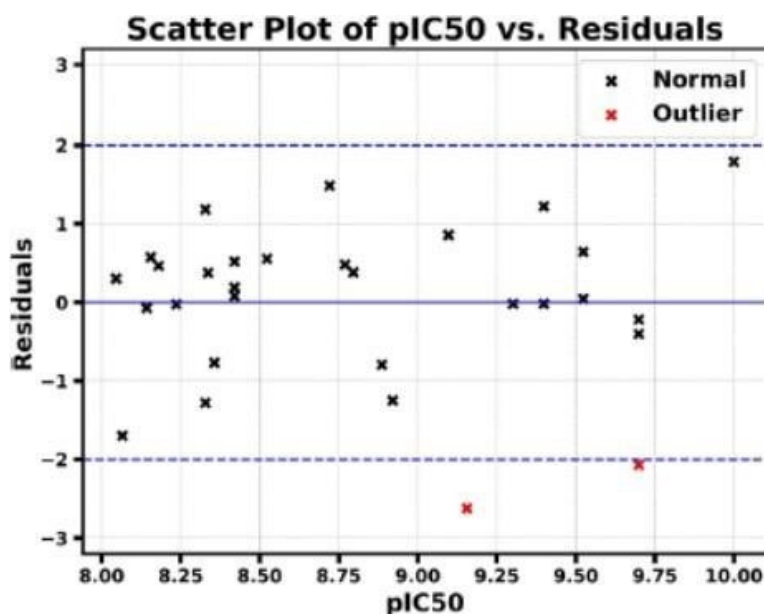
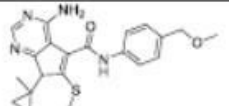
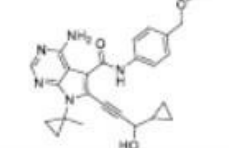
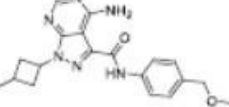
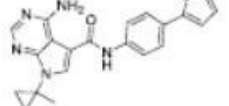


Figure 7. Residual analysis of patent-filed RET inhibitors as out-of-set data. 28 out-of-set molecules showed standardized residuals within the range of -2.0 - $+2.0$ (“normal”)

least potent compound (Figure 8b). Meanwhile, the difference in the Honor feature was transformed into the design logic. The presence of a H- donor at the C4 position, such as an amino group, which forms a H-bond with Glu805, is critical for the design. The presence of the H-donor feature, highlighted by an amino group at the C4 position, characterized the active molecules (molecules 591, 377, and 371). Despite the presence of aromatic amino groups in the inactive compound (44, 46, and 83), they did not align with the H-donor feature at the same location (Figure 8b). Moreover, the H-donor feature with a negative coefficient (red box) was near the aromatic amino group. This subtle difference in the location of the Honor suggests a conducive principle in the design of highly active RET inhibitors. In feature analysis, the functional group at the C3 position was another important feature to discriminate between active and inactive compounds. There are several H-donor cubes near the C3 position. Notably, the N-substituted amide group at the C3 position in the active molecules was surrounded by several blue boxes of the positive coefficient (Figure 8a). Second, we analyzed the 3D location of the hydrophobic feature, which is the most contributing feature in the QSAR model. As is well-known in data science, contour map visualization relies on coefficients, depicting how features affect activity in the QSAR model. In other words, the chosen coefficient cutoff determines which cube boxes are visualized on a contour map. Surely, the hydrophobic feature cube boxes were more dispersed than those of the H-donor. As shown in Figure 9, numerous cube boxes are present at the same coefficient level in the contour map for the hydrophobic feature. The clustered hydrophobic favored boxes (green boxes) led us to assign them as N1, C2, C3, and C4 substituents. The alkenyl group at the C2 position and the *N* alkyl group at the N1 position matched well with the location of hydrophobic boxes (Figure 9a). This implies that the incorporation of a hydrophobic group at the N1 and C2 positions is recommended in the RET inhibitor design. Clearly, while the most inactive molecules lacked substituents to the N1 or C2 positions (Figure 9b), the most potent RET inhibitors have isopropyl (molecules 591 and 371) or 1- Methyl cyclopropyl (molecule 377) a long with the alkenyl group at the C2 position (Figure 9a). Additionally, incorporating an amino group at the C4 position is recommended for highly potent RET inhibitors in terms of hydrophobic feature (negative coefficient) as well as the above H-donor (positive coefficient). Meanwhile, the C3 position was occupied by a mix of green and purple boxes more than any other position in Figure 9. In other words, the C3 position tended to favor hydrophilic substituents in some specific locations. However, although both the active and inactive molecules commonly possessed a hydrophilic group at the β - position of C3 carbon, there was a clear difference in how the 3D location was occupied by rigid amides (of active molecules) and rotatable aromatic amines (of inactive molecules) respectively. Finally, the 3D location of the EW feature was analyzed. Rather than those of other features, the EW location was more shared between active and inactive molecules (Figure 10). Particularly, considerable green cube boxes were commonly occupied by both the active and inactive molecules. It indicates that the EW feature in the hinge binder region has a

weak ability to discriminate the RET inhibitory activity. Meanwhile, C2 and C3 positions presented different aspects. When we integrate every feature, any small substituent or functional group can possess EW and hydrophilic features. Therefore, we derived another design logic from the feature analysis: the introduction of a hydrophilic substituent that attracts electrons at the C3 position becomes necessary when these two distinct features are considered together.

Table 4. Analysis of Best-Performing and Outlier Compounds in Out-of-Set Patent-Filed Data

Data No. in Patent	Structure	Docking Score	Experimental Activity	Predicted Activity	Standardized Residuals	Group
133		-5.807	9.155	5.855	-2.622	Outlier
104		-8.547	9.699	6.402	-2.070	Outlier
14		-5.768	8.357	7.038	1.319	Normal
106		-6.825	8.420	7.671	0.749	Normal

Docking score, experimental activity, and predicted activity from the atom-based QSAR model and the standardized residuals of four ligands from the model validation. The ligand numbering corresponds to that in the data source. Ligands 133 and 104, exhibiting standardized residuals below -2.000, were classified as “outliers.” Ligands 14 and 106, categorized as “normal,” were further analyzed for comparative purposes.

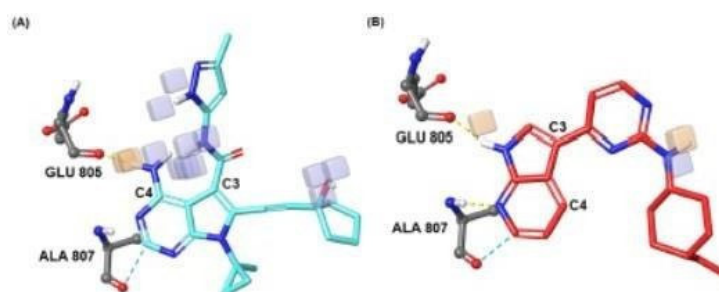


Figure 8. 3D feature map of hydrogen-bond interactions in the optimal QSAR model. (A) Representative active compound 377 (cyan) and (B) inactive compound 44 (red). Legend: positive coefficient (favored for H-donor), light-blue cube; negative coefficient (disfavored for H-donor): light-orange cube; hydrogen bond, yellow dashed line; aromatic hydrogen bond, blue dashed line.

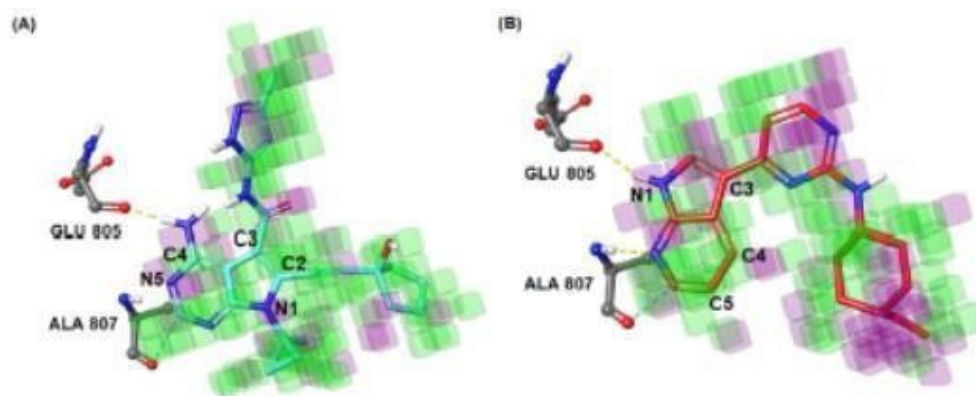


Figure 9. 3D feature map of hydrophobic/nonpolar interaction in the optimal QSAR model. (A) Representative active compound 377 (cyan) and (B) inactive compound 44 (red). Legend: positive coefficient (favored for hydrophobic), green cube; negative coefficient (disfavored for hydrophobic), purple cube; hydrogen bond, yellow dashed line; aromatic hydrogen bond, blue dashed line.

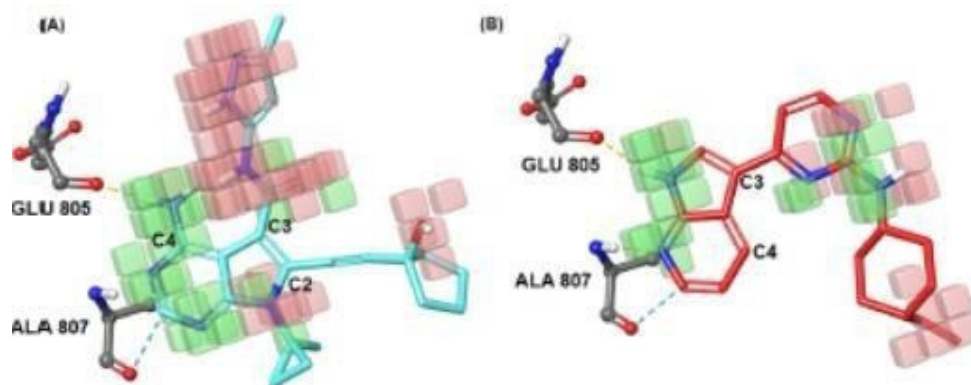


Figure 10. 3D feature map of EW interaction in the optimal QSAR model. (A) Representative active compound 377 (cyan) and (B) inactive compound 44 (red). Legend: positive coefficient (favored for EW), pale-red cube; negative coefficient (disfavored for EW), green cube; hydrogen bond, yellow dashed line; aromatic hydrogen bond, blue dashed line.

4. DISCUSSION

4.1. 3D-QSAR Model for RET Inhibitors Elucidates the Interaction Mechanisms and Logic of Drug Design in the Out-of-Set Data.

In general, an ideal QSAR model corresponds to the non covalent bonding interactions within the respective ligand-target docking complex. This means that the features in QSAR and the interactions observed in the docking complex are closely aligned with each other. Therefore, we also tried to match the RET docking poses with our optimal QSAR model. For this purpose, we examined docking poses of patent-filed RET inhibitors as out-of-set. The docking poses of both the outlier and normal groups commonly exhibited proximity to the hinge region (ALA807) and the nearby ligand-protein interactions. The reverse orientation of the core structure (5, 6 -fused bicyclic hetero aromatic ring) in compounds 104 and 133 (outlier group) compared to compounds 14 and 106 (normal group) significantly influenced the positioning of substituents at C2, C3, C4, and N1 (Figure 11).

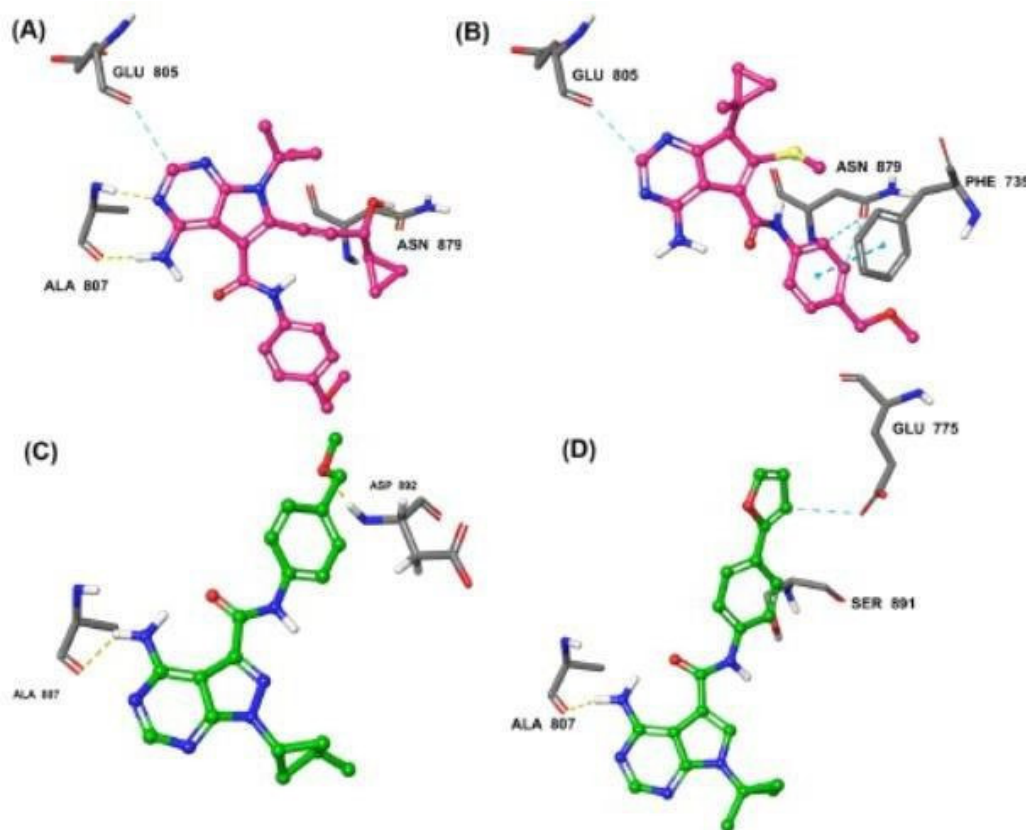


Figure 11 Rison of out-of-set molecules from patent-filed RET inhibitors. Outlier groue molecules include (A) compound 104p and (B) compound 133p and normal group molecules include (C) compound 14p and (D) compound 106p.

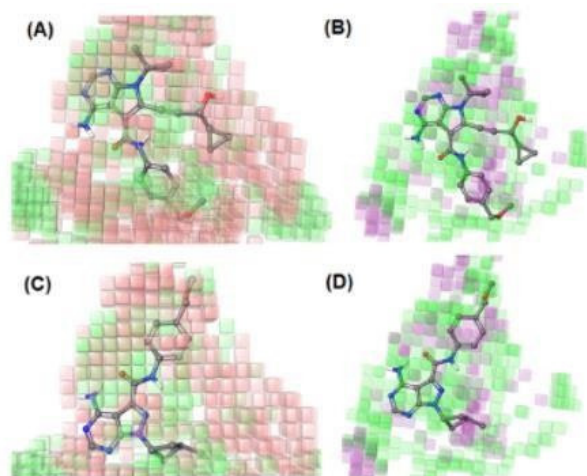


Figure 12. Whole feature analysis of representative out-of-set from patent-filed RET inhibitors. (A) EW interaction feature of compound 104p, (B) hydrophobic/nonpolar interaction feature of compound 104p, (C) EW interaction feature of compound 14p, and (D) hydrophobic/nonpolar interaction feature of compound 14p. Legend: positive coefficient, pale- red cube (for EW) and green cube (for hydrophobic); negative: green cube (for EW) and purple cube (for hydrophobic).

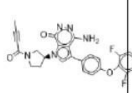
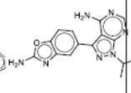
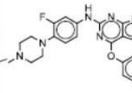
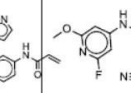
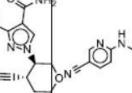
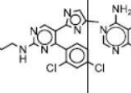
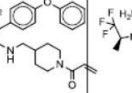
Therefore, 3D-QSAR models, which rely on the alignment of the inhibitors from docking results, can explain the conflicting behaviors of QSAR feature cubes surrounding active and inactive compounds

despite identical substituents. While key interactions with ALA807 at the hinge are common in both groups, our optimal QSAR model accounts for the variation in residual size between the experimental and predicted activities. This variance is primarily due to structural differences, particularly the introduction of substituents at the C2 position in the outlier group (Figure 12). For instance, in the contour map of compound 104, the alkenyl group was matched with the negative EWG feature (electron donating) and the hydroxyl group was well matched with the positive EWG features. In addition to the structural disparities at the C2 position, substituents can be compared. Notably, the presence of an isopropyl group at N1 in compound 104 was regarded as unfavorable (depicted by the negative hydrophobic), in contrast to its favorable depiction in compound 14 (Figure 12). The overall view in Figure 13 simplifies the SAR for designing RET inhibitors based on the comparative analysis of every contour map. The N-substituted amide at the C3 position was commonly observed in active compounds, and hence we named the C3 substituent as a *linker*, such as reported in another study.

4.2. The QSAR Model Applied to the Off-Target Screening of Clinical and Nonclinical Kinase Inhibitors.

The reliable evaluation and interpretation of the optimal QSAR model encouraged us to apply the QSAR model to an off-target screening for RET alteration in known clinical and nonclinical kinase inhibitors. Among 633 drug candidates collected from PKIDB and MRC, 459 kinase inhibitors were docked to RET kinase to produce 3D conformation for QSAR prediction. After achieving both docking scores and predicted pIC50 values, the top 30 inhibitors were selected under two different ranking methods. The first method involved assigning ranks to the docking scores and the predicted pIC50 values separately, with the top 30 selected based on the lowest sum of these ranks. The second method used the cutoff (>7.5) of the predicted pIC50 along with the evaluation of the interaction between docking poses and the target protein. Surprisingly, seven compounds in the chosen top 30 list were acknowledged for their inhibitory efficacy against RET in the literature (Table 5). In the first announced patent, Edralbrutinib, Evobrutinib and Pirtobrutinib are patented as BTK inhibitors; Sapanisertib functions as an mTOR inhibitor; Abivertinib serves as an EGFR inhibitor; Atinvcitinib acts as a JAK inhibitor; and Laduviglusib (also known as CHIR-99021) is an inhibitor of GSK3 β signaling. Despite the absence of a core structure commonly found in active molecules, according to the QSAR model, the molecule contained an imidazole moiety. In the docking pose, the imidazole group was positioned in proximity to GLU805 and ALA807, forming a hydrogen bond, as evidenced by the light blue cubes near the nitrogen atom in the imidazole ring, indicating the presence of a favorable hydrogen-bond donor feature. The predicted value can also be rationalized by the QSAR cubes near other substituents, such as amino and cyan groups. The presence of an amino group, including N24 in Laduviglusib, is favorable, as indicated by the light-blue cube representing hydrogen-bond donor features. Furthermore, the addition of a cyano group to the pyridine ring improved activity because of its EW features, as shown by the pale-red cube in the QSAR model.

Table 5. RET Off-Target Screening of Known Clinical and Nonclinical Kinase Inhibitors

	Edralbrutinib	Sapanisertib	Abivertinib	Atinvcitinib	Laduviglusib	Evobrutinib	Pirtobrutinib
Structure							
Predicted activity(pIC50)	7.915	7.784	7.822	7.924	8.225	7.611	7.306
LEU 730	Aromatic-HB(A)		Aromatic-HB(A)				
LYS 758		cation- π		HB(A)			
GLU 775		HB(D)			Aromatic-HB(A)		
GLU 805	HB(D)	HB(D)	HB(D)	HB(D)	HB(D)	HB(D)	
TYR 806				Aromatic-HB(D)			Aromatic-HB(D)
ALA 807	HB(D) HB(A)	HB(A) Aromatic-HB(A)	HB(D) HB(A) Aromatic-HB(A)	HB(A)	HB(D) Aromatic-HB(A)	HB(A) Aromatic-HB(A)	HB(A)
SER 811							HB(A)
SER 891		Aromatic-HB(A)					
ASP 892	HB(A)	HB(A)		HB(A)			

5. CONCLUSION

To sum up, the main goal of this research was to analyze the connection between ligand structures and 3D-QSAR modeling of inhibitory activity. The ideal RET kinase inhibitors were used to create the 3D-QSAR model. with 49 distinct substructures to show a high predictive performance (Q2 value of 0.801 and R2 value of 0.794). Additionally, the patent-based out-of- set validation molecules demonstrated the usefulness of identifying active molecules. Additionally, the high-resolution 3D spatial feature analysis illuminated the distinctive substituent patterns of the 5, 6-fused heteroaromatic bicyclic core structures. At last, the model was used to assess the potential of screening kinase inhibitors applications. This study will be expanded in the future by confirming results from experiments and evaluating the medication develop ability of the compounds found.

ACKNOWLEDGMENTS

This study was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (No. 2022R1A2C2025810). This research was also supported by a grant of the Korea Machine Learning Ledger Orchestration for Drug Discovery Project (KMELLODDY), funded by the Ministry of Health & Welfare and Ministry of Science and ICT, Republic of Korea (No.: RS 2025-12345678). S.J. appreciates Gachon University's Scholarship Program.

REFERENCES

1. Vargas-Leal, V.; Bruno, R.; Derfuss, T.; Krumbholz, M.; Hohlfeld, R.; Meinl, E. Expression and function of glial cell line derived neurotrophic factor family ligands and their receptors on human immune cells. *J. Immunol.* 2005, *175* (4), 2301–2308.
2. Lipson, D.; Capelletti, M.; Yelensky, R.; Otto, G.; Parker, A.; Jarosz, M.; Curran, J. A.; Balasubramanian, S.; Bloom, T.; Brennan, K. W. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nature medicine* 2012, *18* (3), 382–384.
3. Novello, S.; Califano, R.; Reinmuth, N.; Tamma, A.; Puri, T. RET Fusion-Positive Non- small Cell Lung Cancer: The Evolving Treatment Landscape. *Oncologist* 2023, *28* (5), 402–413.
4. Consortium, A. P. G.; Consortium, A. P. G.; Andre, F.; Arnedos, M.; Baras, A. S.; Baselga, J.; Bedard, P. L.; Berger, M. F.; Bierkens, M.; Calvo, F. AACR Project GENIE: powering precision medicine through an international consortium. *Cancer Discovery* 2017, *7* (8), 818–831.
5. Drilon, A.; Lin, J. J.; Filleron, T.; Ni, A.; Milia, J.; Bergagnini, I.; Hatzoglou, V.; Velcheti, V.; Offin, M.; Li, B. Frequency of brainmetastases and multikinase inhibitor outcomes in patients with RET–rearranged lung cancers. *J. Thorac. Oncol.* 2018, *13* (10), 1595–1601.
6. Pietrantonio, F.; Di Nicolantonio, F.; Schrock, A.; Lee, J.; Morano, F.; Fuca, G.; Nikolinakos, P.; Drilon, A.; Hechtman, J.; Christiansen, J. RET fusions in a small subset of advanced colorectal cancers at risk of being neglected. *Ann. Oncol.* 2018, *29* (6), 1394–1401.
7. Huang, L.; Jiang, S.; Shi, Y. Tyrosine kinase inhibitors for solid tumors in the past 20 years (2001–2020). *J. Hematol. Oncol.* 2020, *13*, 143.
8. Duke, E. S.; Bradford, D.; Marcovitz, M.; Amatya, A. K.; Mishra- Kalyani, P. S.; Nguyen, E.; Price, L. S.; Zirkelbach, J. F.; Li, Y.; Bi, Y. FDA Approval Summary: Selpercatinib for the treatment of advanced RET fusion-positive solid tumors. *Clin. Cancer Res.* 2023, *29*, 3573.
9. Romero, D. Activity of selpercatinib confirmed in phase III trials. *Nature Reviews Clinical Oncology* 2024, *21* (1), 5–5.
10. Mologni, L.; Rostagno, R.; Brussolo, S.; Knowles, P. P.; Kjaer, S.; Murray-Rust, J.; Rosso, E.; Zambon, A.; Scapozza, L.; McDonald, N. Q. Synthesis, structure–activity relationship and crystallographic studies of 3-substituted indolin-2-one RET inhibitors. *Bioorg. Med. Chem.* 2010, *18* (4), 1482–1496.
11. Bhattacharya, S.; Asati, V.; Ali, A.; Ali, A.; Gupta, G. In-silico studies for the development of novel RET inhibitors for cancer treatment. *J. Mol. Struct.* 2022, *1251*, No. 132040.
12. Hu, Y.; Stumpfe, D.; Bajorath, J. Recent advances in scaffold hopping: miniperspective. *Journal of medicinal chemistry* 2017, *60* (4), 1238–1246.
13. Acharya, A.; Yadav, M.; Nagpure, M.; Kumarsan, S.; Guchhait, S. K. Molecular medicinal insights into scaffold hopping-based drug discovery success. *Drug Discov. Today* 2023, *29*, No. 103845.
14. Mathison, C. J.; Chianelli, D.; Rucker, P. V.; Nelson, J.; Roland, J.; Huang, Z.; Yang, Y.; Jiang, J.; Xie, Y. F.; Epple, R. Efficacy and tolerability of pyrazolo [1,5-a] pyrimidine RET kinase inhibitors for the treatment of lung adenocarcinoma. *ACS Med. Chem. Lett.* 2020, *11* (4), 558–565.
15. Lee, S.-H.; Lee, J.-K.; Ahn, M.-J.; Kim, D.-W.; Sun, J.-M.; Keam, B.; Kim, T.; Heo, D.; Ahn, J.; Choi, Y.-L. Vandetanib in pretreated patients with advanced non-small cell lung cancer- harboring RET rearrangement: a phase II clinical trial. *Ann. Oncol.* 2017, *28* (2), 292–297.

16. ChEMBL2023.
17. BindingDB2023.
18. Sun, J.; Jeliaskova, N.; Chupakhin, V.; Golib-Dzib, J.-F.; Engkvist, O.; Carlsson, L.; Wegner, J.; Ceulemans, H.; Georgiev, I.; Jeliaskov, V. ExCAPE-DB: an integrated large scale dataset facilitating Big Data analysis in chemogenomics. *J. Cheminf.* 2017, 9, 17.
19. Lee, J.; Kumar, S.; Lee, S. Y.; Park, S. J.; Kim, M. H. Development of Predictive Models for Identifying Potential S100A9 Inhibitors Based on Machine Learning Methods. *Front Chem.* 2019, 7, 779.
20. Arabie, P.; Baier, N. D.; Critchley, C. F.; Keynes, M. *Studies in Classification, Data Analysis, and Knowledge Organization*; Springer, 2006.
21. *Schrödinger Release 2019-1: LigPrep*; Schrödinger, LLC: New York, NY, 2019
22. Sastry, G. M.; 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 (3), 221–234.
23. Miyazaki, I.; Odintsov, I.; Ishida, K.; Lui, A. J.; Kato, M.; Suzuki, T.; Zhang, T.; Wakayama, K.; Kurth, R. I.; Cheng, R. Vepafestininib is a pharmacologically advanced RET - selective inhibitor with high CNS penetration and inhibitory activity against RET solvent front mutations. *Nature Cancer* 2023, 4 (9), 1345–1361.
24. *Schrödinger Release 2019-1: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2019; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2019.*