JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

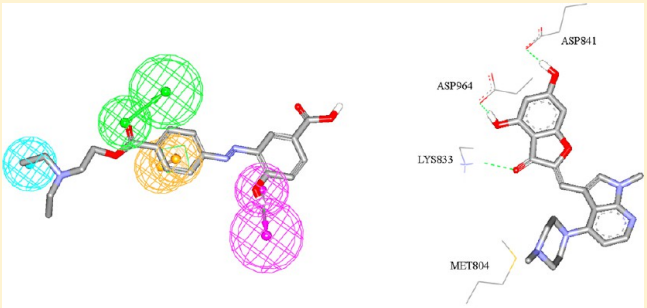
Elaborate Ligand-Based Modeling Coupled with Multiple Linear Regression and k Nearest Neighbor QSAR Analyses Unveiled New Nanomolar mTOR Inhibitors

Mohammad A. Khanfar and Mutasem O. Taha*

Drug Discovery Unit, Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman 11942, Jordan

Supporting Information

ABSTRACT: The mammalian target of rapamycin (mTOR) has an important role in cell growth, proliferation, and survival. mTOR is frequently hyperactivated in cancer, and therefore, it is a clinically validated target for cancer therapy. In this study, we combined exhaustive pharmacophore modeling and quantitative structure−activity relationship (QSAR) analysis to explore the structural requirements for potent mTOR inhibitors employing 210 known mTOR ligands. Genetic function algorithm (GFA) coupled with k nearest neighbor (kNN) and multiple linear regression (MLR) analyses were employed to build self-consistent and predictive QSAR models based on optimal combinations of pharmacophores and physicochemical descriptors. Successful pharmacophores were complemented with exclusion spheres to optimize their receiver operating characteristic curve (ROC) profiles. Optimal QSAR models and their associated pharmacophore hypotheses were validated by identification and experimental evaluation of several new promising mTOR inhibitory leads retrieved from the National Cancer Institute (NCI) structural database. The most potent hit illustrated an IC_{50} value of 48 nM.



1. INTRODUCTION

Mammalian target of rapamycin (mTOR) is a serine/threonine kinase and member of the PI3K-related kinase (PIKK) family.[1] It plays a central role in integrating signals from metabolism, energy homeostasis, cell cycle, and stress response.[1,2] Aberrant PI3K/mTOR activation is commonly observed in cancers.[3,4] mTOR plays an important role in supporting proliferation and cell survival of tumor under metabolic stress conditions.[3,4] Under hypoxic conditions, mTOR contributes to HIF-1α activation to support tumor cell survival.[5] Inhibition of mTOR leads to arrest of mitotic cells in G1 and may eventually result in cell death via apoptosis, possibly through downregulation of cyclin D1 translation.[6] Therefore, mTOR is a validated target for cancer treatment.[7]

In addition to cancer, mTOR is involved in other pathogenesis. It is hyperactivated in brains of Alzheimer's disease patients, and it appears to be accountable for the development of amyloid beta (Aβ) and tau proteins.[8,9] Furthermore, overstimulation of the mTOR pathway by excess food consumption may be a crucial factor underlying the diabetes.[10] mTOR hyperactivation during hyperfeeding leads to insulin desensitization. This results in reduced glucose uptake and glycogen synthesis in liver and muscle and increased gluconeogenesis and glucose release in liver. Mutually, these effects lead to worsening of the hyperglycemia and hyperinsulinemia.[10,11]

The pronounced current interest in developing new mTOR inhibitors as potential agents for treatment of cancer, Alzheimer's disease, and diabetes,[1−11] combined with the lack of crystallographic structure for mTOR kinase domain prompted us to explore the possibility of developing ligand-based 3D pharmacophores integrated within self-consistent QSAR models. The pharmacophore models can be used as 3D search queries to mine 3D libraries for new mTOR inhibitors, while the associated QSAR models can be used to predict the bioactivities of captured hits and therefore prioritize them for *in vitro* evaluation.

We previously reported the use of this innovative approach toward the discovery of new leads for glycogen synthase kinase 3β,[12] β-secretase,[13] CDK1,[14] β-D-galactosidase,[15] glycogen phosphorylase,[16] rho kinase,[17] inducible nitric oxide synthase (iNOS),[18] and Ca^{2+}/calmodulin-dependent protein kinase II.[19]

However, we herein present a new workflow that combines linear (MLR) and nonlinear (kNN) modeling approaches for better exploration of the bioactive chemical space of mTOR inhibitors. Figure 1 shows a schematic representation of the overall computational workflow of this novel approach. Interestingly, this workflow unveiled new pharmacophoric models that allowed us to better understand ligand binding into the mTOR binding site. Morover, the new binding models were used as three-dimensional search queries to discover new nanomolar bioactive hits.

Received: July 1, 2013

 © XXXX American Chemical Society

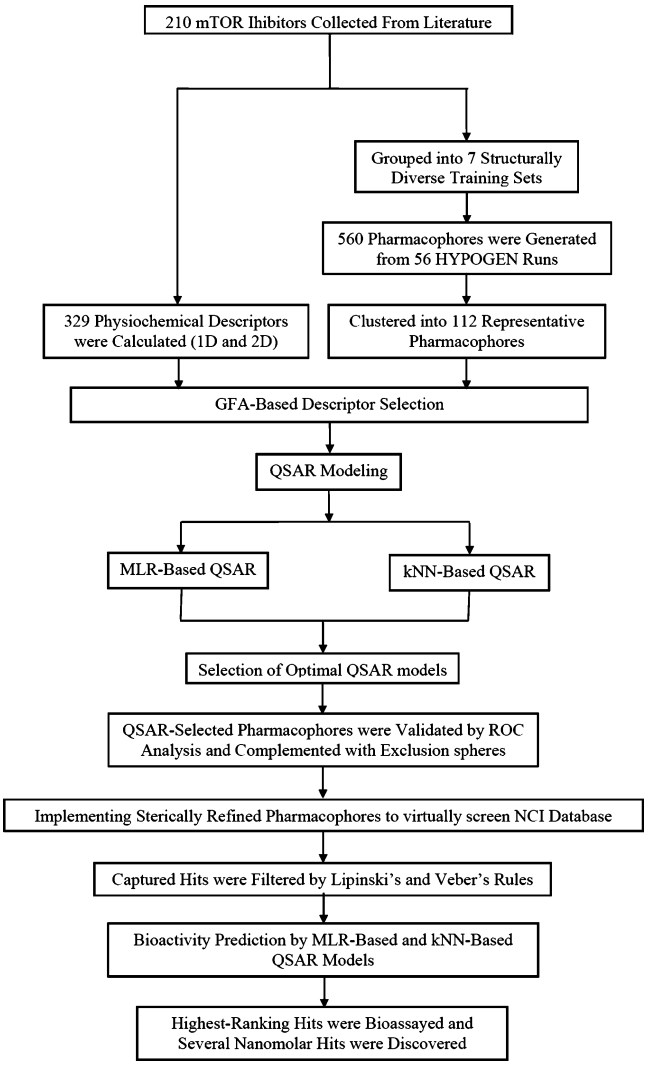


Figure 1. General computational workflow implemented herein for discovering novel mTOR inhibitors. Acronyms: GFA, genetic function approximation; MLR, multiple linear regression; kNN, k nearest neighbor; ROC, receiver operating characteristic.

We employed the HYPOGEN module from the CATALYST software package to construct numerous plausible binding hypotheses for mTOR inhibitors.[20−32] Subsequently, a genetic function algorithm (GFA) coupled with multiple linear regression (MLR) analysis or k nearest neighbor (kNN) analysis was employed to search for optimal QSAR models. Both approaches yielded QSAR models that combined high-quality binding pharmacophores with other physicochemical molecular descriptors capable of explaining bioactivity variation across a collection of diverse mTOR inhibitors.

The resulting pharmacophores were validated by evaluating their abilities to successfully classify a long list of compounds as actives or inactives, that is, by assessing their receiver-operating characteristic (ROC) curves. Subsequent decoration with steric exclusion spheres enhanced their ROC profiles.

The resulting sterically refined pharmacophores were used as 3D search queries to screen the National Cancer Institute (NCI) virtual molecular database for new mTOR inhibitors.

2. RESULTS AND DISCUSSION

CATALYST-HYPOGEN utilizes a collection of molecules with activities ranging over a number of orders of magnitude for automatic pharmacophore construction. HYPOGEN pharmacophores use the geometric localization of the chemical features present in the molecules to explain the variability of bioactivity. A total of 210 mTOR inhibitors (Figure A and Table A in Supporting Information) were used in this study to generate different binding pharmacophore hypotheses. The reader is advised to see sections SM-2 and SM-3 in Supporting Information for full description of HYPOGEN pharmacophore modeling algorithm.[33,36−39]

2.1. Exploration of mTOR Pharmacophoric Space. The literature was investigated to collect as many structurally diverse mTOR inhibitors as possible. The collected inhibitors were selected in such a way that they were assayed by the same procedure (**1−210**, see Figure A and Table A in Supporting Information).[20−35] Statistical consistency necessitates that QSAR and pharmacophore modeling are based on training compounds assayed by a single bioassay procedure.[12−19]

The pharmacophoric space of mTOR inhibitors was explored through 16 HYPOGEN automatic runs performed on seven carefully selected training subsets: **A**, **B**, **C**, **D**, **E**, **F**, and **G** (Table B in Supporting Information).[33,36−39] The training compounds were selected to guarantee wide structural diversity with bioactivities extended over more than 3.5 logarithmic cycles. To ensure sufficient molecular diversity within training subsets, member compounds were selected in such a way that each structural cluster of the collected compounds (Table A, Supporting Information) was sampled at least once in each training subset. Training subsets were selected in such a way that differences in mTOR inhibitory activities are primarily attributable to the presence or absence of pharmacophoric features (e.g., hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic (Hbic), or ring aromatic (Ring-Arom)) rather than steric shielding or bioactivity-enhancing or -reducing auxiliary groups (e.g., electron-donating or -withdrawing groups). A special emphasis was given to the structural diversity of the most-active compounds in each training subset (Table B in Supporting Information) because of their significant influence on the extent of the evaluated pharmacophoric space during the constructive phase of HYPOGEN algorithm.

HYPOGEN was instructed to explore only four- and five-featured pharmacophores and ignore models of lesser number of features (as shown in Table C in Supporting Information). The advantage of this restriction is to narrow the investigated pharmacophoric space while allowing good representation of the feature-rich nature of mTOR inhibitors.

Eventually, 560 pharmacophore models resulted from 56 automatic CATALYST-HYPOGEN runs, out of which 559 models illustrated confidence levels ≥90% (Fisher scrambling criteria, See section SM-3 in Supporting Information).[33,36−39] These successful models were clustered, and their best 112 representatives were used in subsequent QSAR modeling. Table D in Supporting Information shows the statistical criteria of the best representatives.

2.2. QSAR Modeling. Although pharmacophore models provide excellent insights into ligand−receptor recognition and binding phenomena, their predictive potential suffer from two important pitfalls, namely, (i) they fail to account for the steric constraints of binding pockets and (ii) they fail to explain

bioactivity enhancing or reducing effects associated with auxiliary groups (electron-donating and -withdrawing functionalities). Furthermore, our pharmacophore exploration yielded numerous high-quality models of comparable success criteria (Table D in Supporting Information), which renders selecting a particular binding hypothesis to explain bioactivity variations across all collected mTOR inhibitors rather daunting. Accordingly, we decided to implement QSAR as a competition platform to select the best possible combination of pharmacophores and other molecular descriptors collectively capable of explaining bioactivity variations across collected mTOR inhibitors.

We implemented GFA[40] as means for selecting different combinations of pharmacophores and molecular descriptors. However, we implemented two separate methodologies to evaluate the ability of the resulting descriptor and pharmacophore combinations in explaining bioactivity variations within mTOR inhibitors: (a) MLR analysis and (b) kNN regression.

MLR analysis assumes the existence of a linear correlation between molecular descriptors and corresponding bioactivities.[12−19] On the other hand, kNN is a nonlinear nonparametric method that predicts a ligand's bioactivity as distance weighted average of the bioactivities of its k nearest neighbors. The neighborhood is defined based on certain selected descriptors. The nearness is measured by an appropriate distance metric (e.g., a molecular similarity measure).[41,42]

The fit values obtained by mapping 112 representative hypotheses (generated from clustering of pharmacophore hypotheses) against collected inhibitors (**1−210**) were enrolled together with a selection of 2D descriptors as independent variables in GFA/MLR-based and GFA/kNN-based QSAR analyses.

2.2.1. Multiple Linear Regression-Based QSAR Modeling. Equation 1 shows the optimal GFA/MLR-based QSAR model. Figure 2 shows the corresponding scatter plot of experimental versus estimated bioactivities for the training and testing inhibitors.

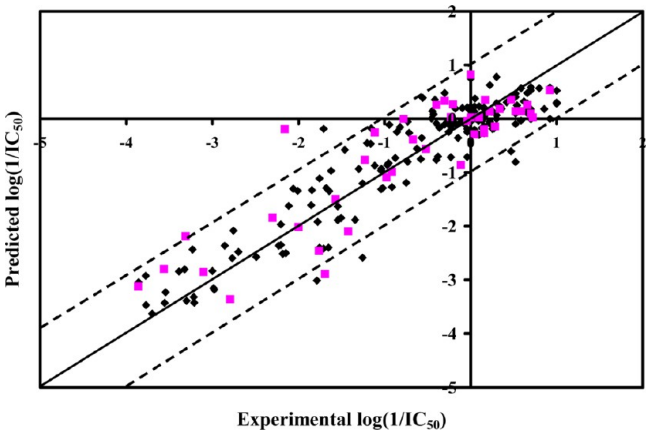


Figure 2. Experimental versus predicted bioactivities for the training compounds (black squares) and testing compounds (pink squares). Predicted bioactivities calculated from the best MLR-QSAR model eq 1. The solid line is the regression line for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate ±1.0 logarthmic error margins.

$$\log(1/\mathrm{IC}_{50}) = -6.11 + 0.27(\mathbf{SssCH}) + 0.045(\mathbf{AaN}) + 8.19(\mathbf{JursFNSA1}) + 0.11\mathbf{Hypo}(\mathbf{A\text{-}T7\text{-}8}) + 0.14\mathbf{Hypo}(\mathbf{E\text{-}T5\text{-}8}) + 0.12\mathbf{Hypo}(\mathbf{G\text{-}T2\text{-}1})$$
$$n = 168,\ r_{168}^2 = 0.86,\ F = 160.7,\ r_{\mathrm{LOO}}^2 = 0.84,\ r_{\mathrm{PRESS}(42)}^2 = 0.77 \quad (1)$$

where, n is the number of training compounds, r_{168}^2 is the correlation coefficient against 168 training compounds, r_{LOO}^2 is the leave-one-out cross-validation correlation coefficient, and r_{PRESS}^2 is the predictive r^2 determined for 42 randomly selected test compounds. **Hypo(A-T7-8)**, **Hypo(E-T5-8)**, and **Hypo(G-T2-1)** represent the fit values of the training compounds (as calculated from equation D in Supporting Information) against the corresponding pharmacophore models, as in Tables C and D in Supporting Information. Figures 3, 4 and 5 show the three models, while Table 1 shows the X, Y, and Z coordinates of the three pharmacophores.

The remaining descriptors are as follows: **SssCH** represents the count of trivalent CH fragments, while **AaN** represents the count of heterocyclic aromatic nitrogen atoms. **JursFNSA1** is a fractional negative charged partial surface area obtained by dividing the total charge weighted negative surface area by the total molecular solvent accessible surface area.

The statistical criteria of eq 1 have excellent predictive values. This model has excellent r_{LOO}^2 and r_{PRESS}^2 values against 42 compounds randomly selected from an external list.

The **JursFNSA1** descriptor has significant positive regression slope. This indicates that ligands with diffuse negative charges tend to have higher affinities to mTOR binding pocket. This is not unexpected since the putative binding pocket of mTOR includes six cationic amino acids, that is, Lys2166, Lys2187, Arg2251, Lys2171, Lys2256, and Lys2257 (the former three are intimately involved in ligand binding).[43−45] Accordingly, ligands with pronounced negatively charged centers tend to have higher binding affinities to mTOR binding pocket. This conclusion is further supported by the appearance of **AaN** combined with positive regression slope, which suggests that nitrogen heterocycles promote bioactivity. Heterocyclic nitrogens represent strongly electronegative centers capable of electrostatic and hydrogen-bonding interactions with cationic side chains of lysine and arginine residues.

Interestingly, the **SssCH** descriptor in eq 1 seems to correlate with the presence of 2,6-ethylene-bridged morpholine substituents in potent ligands, for example, compounds **1−17** (Figure A and Table A in Supporting Information). In contrast, this group is absent from the less active mTOR inhibitors, for example, compounds **108−124** (Figure A and Table A in Supporting Information). The ethylene bridge stacks at close proximity with a hydrophobic moiety within the binding pocket leading to the observed trend.[43]

The three binding models (**Hypo(A-T7-8)**, **Hypo(E-T5-8)**, and **Hypo(G-T2-1)**) in eq 1 suggest the existence of at least three binding modes assumed by inhibitors within the binding pocket of mTOR. They correspond nicely with binding interactions tying cocrystallized ligands within the highly mTOR homologous protein PI3K-γ (Figures 3, 4, and 5).

Figure 3B shows how pharmacophore model **Hypo(A-T7-8)** maps a potent dual PIK3-γ/mTOR inhibitor compared with its cocrystallized structure within the binding pocket of PI3K-γ (Figure 3C). Mapping the urea hydrogens with HBD in

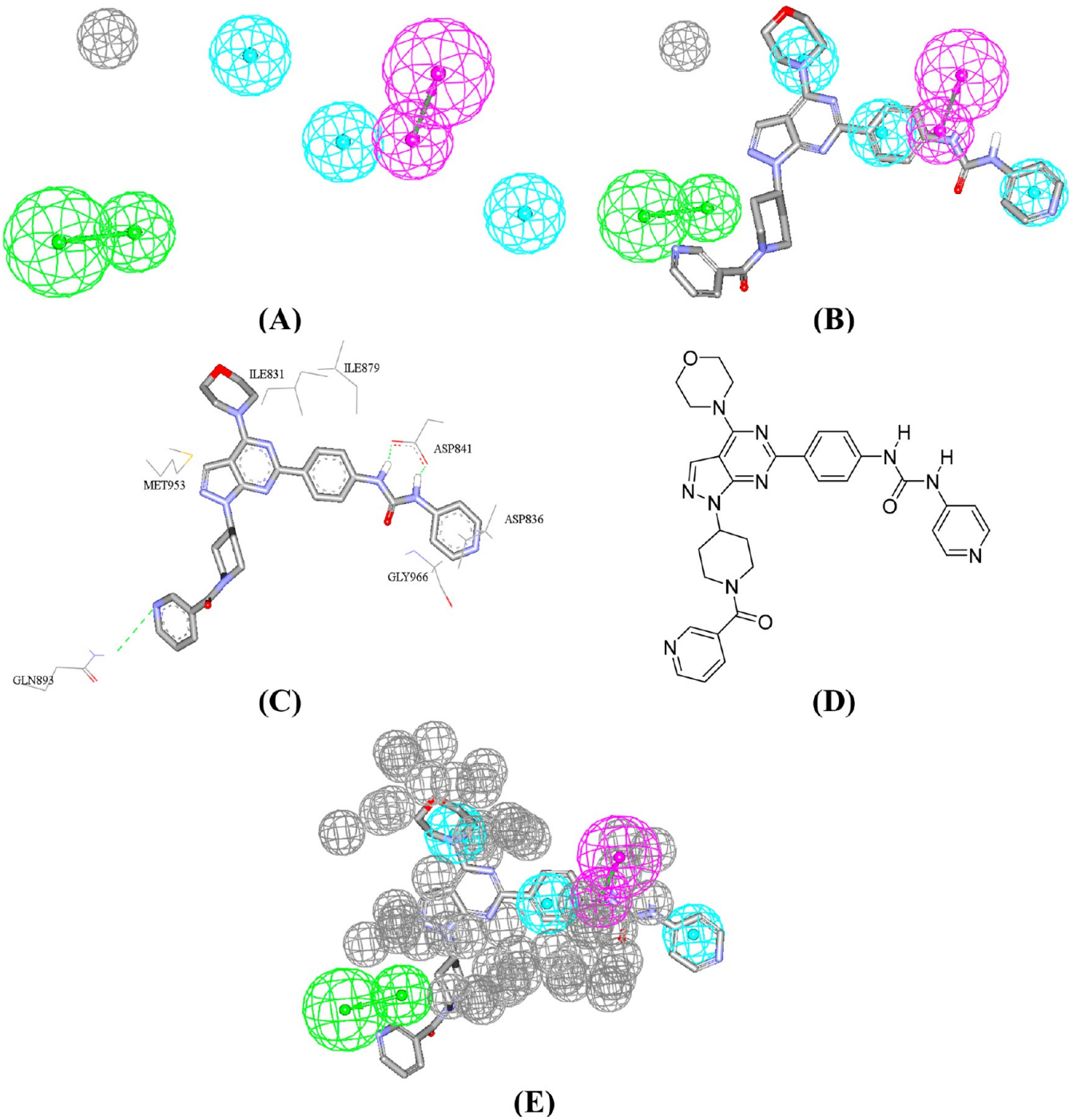


Figure 3. Hypo(A-T7-8). (A) Pharmacophoric features of the binding model: HBA as green vectored spheres, HBD as violet vectored spheres, and Hbic as blue spheres. (B) **Hypo(A-T7-8)** fitted against PI3K-γ cocrystallized ligand (pdb code 3IBE). (C) The key binding interactions of PI3K-γ cocrystallized ligand (pdb code 3IBE). (D) The chemical structures of the cocrystallized ligand. (E) HipHop-refined **Hypo(A-T7-8)** with exclusion volumes (gray spheres).

Hypo(A-T7-8) correlates with hydrogen-bonding interactions connecting the same urea hydrogens with the carboxylate of Asp841. Similarly, the less-than-optimal mapping of the meta-substituted pyridine nitrogen against HBA feature in **Hypo(A-T7-8)** corresponds to a stretched hydrogen-bonding interaction connecting the same atom with the amidic side chain of Gln893 (Figure 3C). Finally, mapping the terminal pyridine, benzene linker, and morpholine groups against three HBic features in **Hypo(A-T7-8)** agrees with hydrophobic interactions anchoring these groups with the hydrophobic side chains of Asp836, Ile879, Ile831, and Met953(Figure 3C).

Similarly, **Hypo(E-T5-8)** maps another ligand cocrystallized within PI3K-γ (Figure 4B). Mapping the catechol hydroxyls against two HBD features in **Hypo(E-T5-8)** correlates with hydrogen-bonding interactions connecting them to the carboxylates of Asp964 and Asp841 (Figure 4C), while mapping the chromone carbonyl against a HBA feature in **Hypo(E-T5-8)** seems to agree with a hydrogen-bonding interaction connecting this group with the peptidic NH of

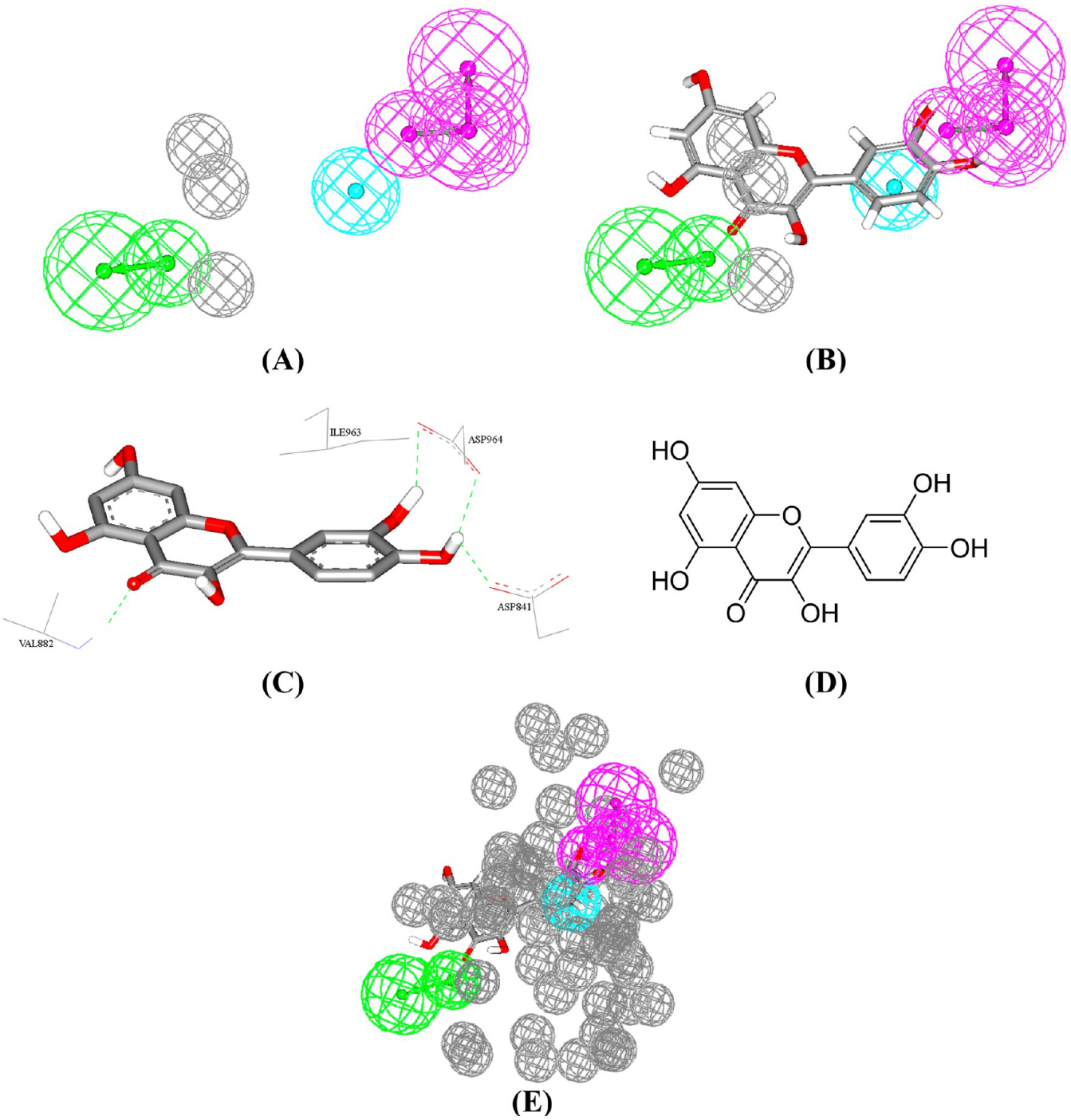


Figure 4. Hypo(E-T5-8). (A) Pharmacophoric features of the binding model: HBA as green vectored spheres, HBD as violet vectored spheres, and Hbic as blue spheres. (B) **Hypo(E-T5-8)** fitted against PI3K-γ cocrystallized ligand (pdb code 1E8W). (C) The key binding interactions of PI3K-γ cocrystallized ligand (pdb code 1E8W). (D) The chemical structures of the cocrystallized ligand. (E) HipHop-refined **Hypo(E-T5-8)** with exclusion volumes (gray spheres).

Val882 (Figure 4C). Finally, stacking the catechol aromatic ring against the hydrophobic side chain of Ile963 (Figure 4C) nicely agrees with mapping the same aromatic ring against a Hbic feature in **Hypo(E-T5-8)** (Figure 4B).

Finally, **Hypo(G-T2-1)** seems to encode for another separate binding mode by which ligands fit within mTOR (Figure 5). Figure 5B,C compares the way by which another dual PIK3-γ/mTOR inhibitor maps **Hypo(G-T2-1)** with the binding interactions tying the same ligand within the PI3K-γ cocrystallized complex. Mapping the ligand's sulfonamide NH_2 against a HBD feature in **Hypo(G-T2-1)** correlates with hydrogen bonding connecting the same NH_2 with the peptidic carbonyl oxygen of Ala885. Similarly, mapping the adjacent aromatic ring against a RingArom feature in **Hypo(G-T2-1)** agrees with π-stacking interactions anchoring this aromatic ring against the indole ring of Trp812. Likewise, the hydrogen-bonding interaction connecting the pyrimidine nitrogen atom with the peptidic NH of Val882 is encoded in **Hypo(G-T2-1)** by mapping the same heterocyclic nitrogen against a HBA feature. Finally, fitting the benzoyl ring of the cocrystallized ligand

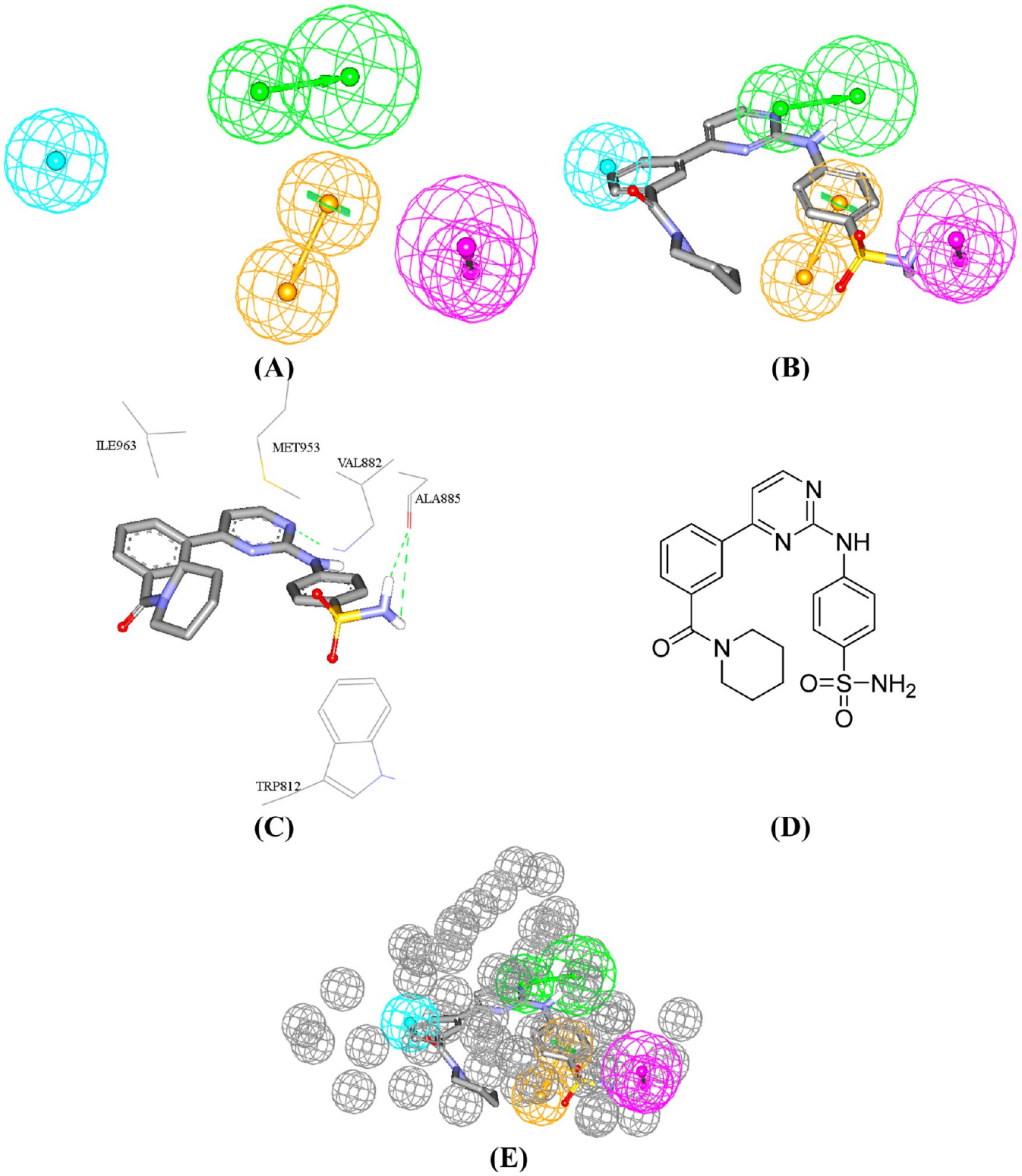


Figure 5. Hypo(G-T2-1). (A) Pharmacophoric features of the binding model: HBA as green vectored spheres, HBD as violet vectored spheres, Hbic as blue spheres, and RingArom as orange vectored spheres. (B) **Hypo(G-T2-1)** fitted against PI3K-γ cocrystallized ligand (pdb code 4FUl). (C) The key binding interactions of PI3K-γ cocrystallized ligand (pdb code 4FUl). (D) The chemical structures of the cocrystallized ligand. (E) HipHop-refined **Hypo(G-T2-1)** with exclusion volumes (gray spheres).

against a Hbic feature in **Hypo(G-T2-1)** corresponds to hydrophobic interactions linking this ring with the hydrophobic side chain of Ile963.

Accordingly, the three pharmacophores represent three corresponding binding modes assumed by different ligands within the binding pocket of mTOR. Needless to say that currently there is no available mTOR crystallographic structure in the protein databank, which prompted us to use the highly homologous PI3K-γ as an alternative crystallographic model for comparison.

Table 1. mTOR Based Pharmacophore Models Selected by MLR-QSAR and kNN-QSAR Modeling

model	definitions		chemical features						
			HBA		HBD		Hbic	Hbic	Exv
Hypo(A-T7-8)[a]	weights		1.91		1.91		1.91	1.91	1.91
	tolerances		1.6	2.2	1.6	2.2	1.6	1.6	1.2
	coordinates	*X*	−4.82	−7.71	0.034	1.43	1.80	−1.16	−6.69
		Y	−1.03	−1.72	0.071	1.22	−4.04	−1.32	−3.14
		Z	−3.91	−4.37	5.84	−7.66	9.52	4.12	4.64

model	definitions		chemical features									
			HBA		HBD		HBD		Hbic	Exv1	Exv2	Exv3
Hypo(E-T5-8)[b]	weights		2.18		2.18		2.18		2.18	2.18	2.18	2.18
	tolerances		1.6	2.2	1.6	2.2	1.6	2.2	1.6	1.2	1.2	1.2
	coordinates	*X*	−2.61	−3.82	1.00	1.30	0.104	2.33	0.156	−3.22	−1.58	5.91
		Y	−3.29	−4.73	−1.43	0.116	−0.071	1.38	0.100	−5.36	−1.33	−3.83
		Z	−4.19	−6.52	7.70	−8.82	−5.91	6.88	−2.98	2.73	−5.36	−1.09

model	definitions		chemical features						
			HBA		HBD		RingArom		Hbic
Hypo(G-T2-1)[c]	weights		2.59		2.59		2.59		2.59
	tolerances		1.6	2.2	1.6	2.2	1.6	1.6	1.6
	coordinates	*X*	−0.69	1.62	1.06	0.45	1.11	4.04	−3.24
		Y	0.69	3.50	−2.43	0.37	0.90	−0.23	−4.36
		Z	0.84	0.035	8.51	10.1	3.33	3.37	−2.56

model	definitions		chemical features						
			HBA		HBD		HBD		Hbic
Hypo(E-T1-3)[d]	weights		2.18		2.18		2.18		
	tolerances		1.6	2.2	1.6	2.2	1.6	1.6	2.2
	coordinates	*X*	−0.44	0.67	7.67	8.72	5.54	−0.44	0.67
		Y	−0.80	−1.73	−1.52	1.25	−0.18	−0.80	−1.73
		Z	0.85	3.52	0.88	0.36	0.64	0.85	3.52

model	definitions		chemical features						
			HBA		HBD		RingArom		Hbic
Hypo(C-T2-9)[e]	weights		2.33		2.33		2.33		2.33
	tolerances		1.6	2.2	1.6	2.2	1.6	1.6	1.6
	coordinates	*X*	−0.42	1.46	7.89	9.16	3.33	2.98	−4.51
		Y	−0.20	−2.47	1.75	0.71	0.81	−1.09	4.65
		Z	−3.17	−4.24	−0.84	−3.35	−2.49	−0.06	−3.01

model	definitions		chemical features									
			HBA		HBD		RingArom		Hbic	EV1	EV2	EV3
Hypo(A-T6-8)[f]	weights		2.59		2.59		2.59		2.59	2.18	2.18	2.18
	tolerances		1.6	2.2	1.6	2.2	1.2	1.2	1.2	1.2	1.2	1.2
	coordinates	*X*	−4.34	−1.58	8.18	8.23	−1.34	−1.09	2.98	7.00	−1.20	−7.80
		Y	3.13	2.56	0.13	2.46	−1.3	0.63	0.34	5.12	1.89	−2.98
		Z	2.44	3.52	0.76	2.65	0.28	−2.00	0.00	6.13	6.27	0.40

[a]**Hypo(A-T7-8)** corresponds to the pharmacophore model generated by subset A (Table B), HYPOGEN run number 7 (Table C in the Supporting Information), eighth-ranked model. [b]**Hypo(E-T5-8)** corresponds to the pharmacophore model generated by subset E (Table B), HYPOGEN run number 5 (Table C in the Supporting Information), eighth-ranked model. [c]**Hypo(G-T2-1)** corresponds to the pharmacophore model generated by subset G (Table B), HYPOGEN run number 2 (Table C in the Supporting Information), first-ranked model. [d]**Hypo(E-T1-3)** corresponds to the pharmacophore model generated by subset E (Table B), HYPOGEN run number 1 (Table C in the Supporting Information), third-ranked model. [e]**Hypo(C-T2-9)** corresponds to the pharmacophore model generated by subset C (Table B), HYPOGEN run number 2 (Table C in the Supporting Information), ninth-ranked model. [f]**Hypo(A-T6-8)** corresponds to the pharmacophore model generated by subset A (Table B), HYPOGEN run number 6 (Table C in the Supporting Information), eighth-ranked model.

To further validate our QSAR-selected pharmacophores, we subjected them to ROC curve analysis to assess their abilities to selectively capture diverse mTOR inhibitors from a large list of decoys. The validity of a particular pharmacophore is indicated by the area under the curve (AUC) of the corresponding ROC curve, as well as the overall accuracy, specificity, true positive rate, and false negative rate of the pharmacophore (see **SM-4** in the Supporting Information). The ROC performances of the three QSAR-selected pharmacophores are excellent with ROC-AUC values of 0.935, 0.941, and 0.850 for **Hypo(A-T7-8)**, **Hypo(E-T5-8)**, and **Hypo(G-T2-1)**, respectively (Table 2 and Figure B in Supporting Information).

2.2.2. kNN-Based QSAR Modeling. By careful evaluation of different descriptors in QSAR model eq 1, we noticed that the three selected pharmacophores were moderately collinear with an average cross-correlation r^2 of 0.72. We believe this pitfall

Table 2. ROC[a] Performances of QSAR-Selected Pharmacophores and Their Sterically Refined Versions as 3D Search Queries

pharmacophore model	ROC[a]-AUC[b]	ACC[c]	SPC[d]	TPR[e]	FNR[f]
Hypo(A-T7-8)	0.935	0.968	0.979	0.625	0.021
Hypo(E-T5-8)	0.941	0.968	0.978	0.656	0.022
Hypo(G-T2-1)	0.850	0.968	0.972	0.843	0.028
sterically refined **Hypo(A-T7-8)**	0.999	0.968	0.982	0.531	0.0177
sterically refined **Hypo(E-T5-8)**	0.996	0.968	0.979	0.625	0.021
sterically refined **Hypo(G-T2-1)**	0.979	0.968	0.975	0.75	0.025
Hypo(A-T6-8)	0.911	0.968	0.972	0.843	0.028
Hypo(C-T2-9)	0.869	0.968	0.969	0.906	0.030
Hypo(E-T1-3)	0.909	0.968	0.979	0.626	0.021
sterically refined **Hypo(A-T6-8)**	0.996	0.968	0.978	0.656	0.021
sterically refined **Hypo(C-T2-9)**	0.983	0.968	0.978	0.688	0.022
sterically refined **Hypo(E-T1-3)**	0.988	0.968	0.982	0.531	0.0177

[a]ROC, receiver operating characteristic. [b]AUC, area under the curve. [c]ACC, overall accuracy. [d]SPC, overall specificity. [e]TPR, overall true positive rate. [f]FNR, overall false negative rate.

arises from the fact that MLR-based modeling assumes linear relationships between ligands' descriptors and bioactivities, thus forcing the GFA-based selection process to filter out any descriptors nonlinearly related to bioactivity. This limits the available pool of explanatory descriptors thus increasing the probability of selecting moderately collinear descriptors in the final QSAR model. Needless to say, collinear descriptors can significantly increase prediction errors in MLR-based QSAR models because they tend to reduce signal-to-noise ratio for successful selection of descriptors that best describe response among training lists.[46]

Accordingly, we decided to attempt QSAR modeling using a nonlinear modeling approach. We adopted kNN-based QSAR modeling for this purpose. The kNN-QSAR methodology relies on a distance learning approach such that the activity value of an unknown member is calculated from the activity values of a certain number (k) of nearest neighbors (kNNs) in the training set. The similarity is measured by a distance metric, and in the present study, the Euclidean distance is considered. We implemented the following kNN workflow: (1) calculate Euclidean distances between an unknown object (x) and all the objects in the training set with respect to certain descriptor(s) selected by GFA; (2) select k objects from the training set most similar to object x; (3) calculate the distance-weighted average bioactivities of k objects as predicted bioactivity of x; (4) correlate predicted bioactivities with experimental ones to determine the optimal k value and explanatory descriptors via leave-20% out cross-validation.[41,42]

Table 3 shows the selected descriptors, nearest neighbors, and statistical criteria of the top five kNN-based QSAR models. We selected model number 1 (Table 3) as the best representative for subsequent virtual screening and QSAR-based predictions because it exhibits excellent overall explanatory power with the least number of descriptors and nearest neighbors.[41,42] Interestingly, kNN-QSAR model 1 unveiled significantly different sets of explanatory descriptors compared with MLR-QSAR (eq 1).

Table 3. Optimal kNN-QSAR Models Including Their Corresponding Descriptors, Nearest Neighbors, and Statistical Criteria

model	selected descriptors	number of nearest neighbors	statistical criteria		
			r^{2a}	r_{LOO}^{2b}	$r_{L20\%Out}^{2c}$
1	**Hypo(E-T1-3)**[d]	3	0.97	0.80	0.88
	Hypo(C-T2-9)				
	Hypo(A-T6-8)				
	sssN[e]				
	dssC[f]				
	aaaC[g]				
	aaS[h]				
2	**Hypo(E-T1-3)**	3	0.97	0.82	0.89
	Hypo(C-T2-9)				
	Hypo(A-T6-8)				
	Hypo(A-T1−1)				
	sssN				
	dssC				
	aasC[i]				
	aaS				
3	**Hypo(G-T2-1)**	4	0.96	0.82	0.89
	Hypo(E-T2-6)				
	Hypo(C-T2-9)				
	Hypo(A-T1−9)				
	sssN				
	dssC				
	aasC				
	aaS				
4	**Hypo(G-T2−10)**	5	0.95	0.84	0.88
	Hypo(E-T7−9)				
	Hypo(C-T1−10)				
	Hypo(A-T6-8)				
	ssO[j]				
	sCH$_3$[k]				
	aaS				
	aaaC				
5	**Hypo(E-T7−9)**	7	0.94	0.82	0.88
	Hypo(C-T1−10)				
	Hypo(B-T8−2)				
	Hypo(A-T6-8)				
	Hypo(A-T1−1)				
	dssC				
	aaS				
	aaaC				
	sCH$_3$				

[a]Correlation coefficient between the predicted and experimental log(1/IC$_{50}$) values. [b]Leave one out cross correlation coefficient. [c]Leave 20% out cross correlation coefficient. [d]Pharmacophore names correspond to their training subsets (Table B), number of corresponding automatic HYPOGEN run (as in Table C in Supporting Information) and rank among other pharmacophores generated in that particular automatic run (Table D in the Supporting Information). For example, **Hypo(E-T1-3)** corresponds to the pharmacophore model generated by subset E (Table B), HYPOGEN run number 1 (Table D in the Supporting Information), third-ranked model. [e]Number of tertiary amine nitrogens. [f]Number of olefinic carbon atoms. [g]Number of quaternary aromatic carbon atoms. [h]Number of aromatic sulfur atoms. [i]Number of tertiary aromatic carbon atoms. [j]Number of etheric oxygens. [k]Number of methyl groups.

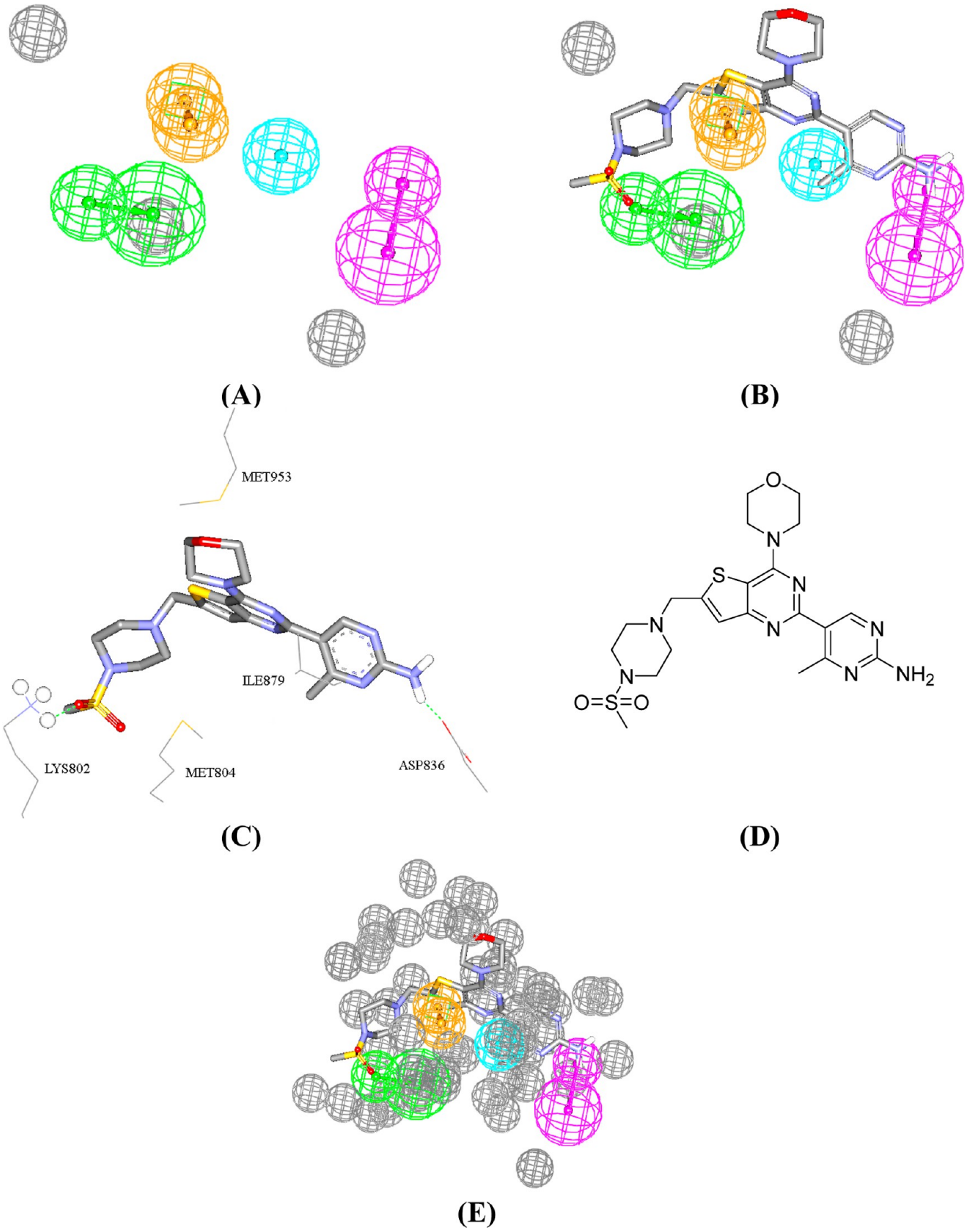


Figure 6. Hypo(A-T6-8). (A) Pharmacophoric features of the binding model: HBA as green vectored spheres, HBD as violet vectored spheres, and Hbic as blue spheres. (B) **Hypo(A-T6-8)** fitted against PI3K-γ cocrystallized ligand (pdb code 3L17). (C) The key binding interactions of PI3K-γ cocrystallized ligand (pdb code 3L17). (D) The chemical structures of the cocrystallized ligand. (E) HipHop-refined **Hypo(A-T6-8)** with exclusion volumes (gray spheres).

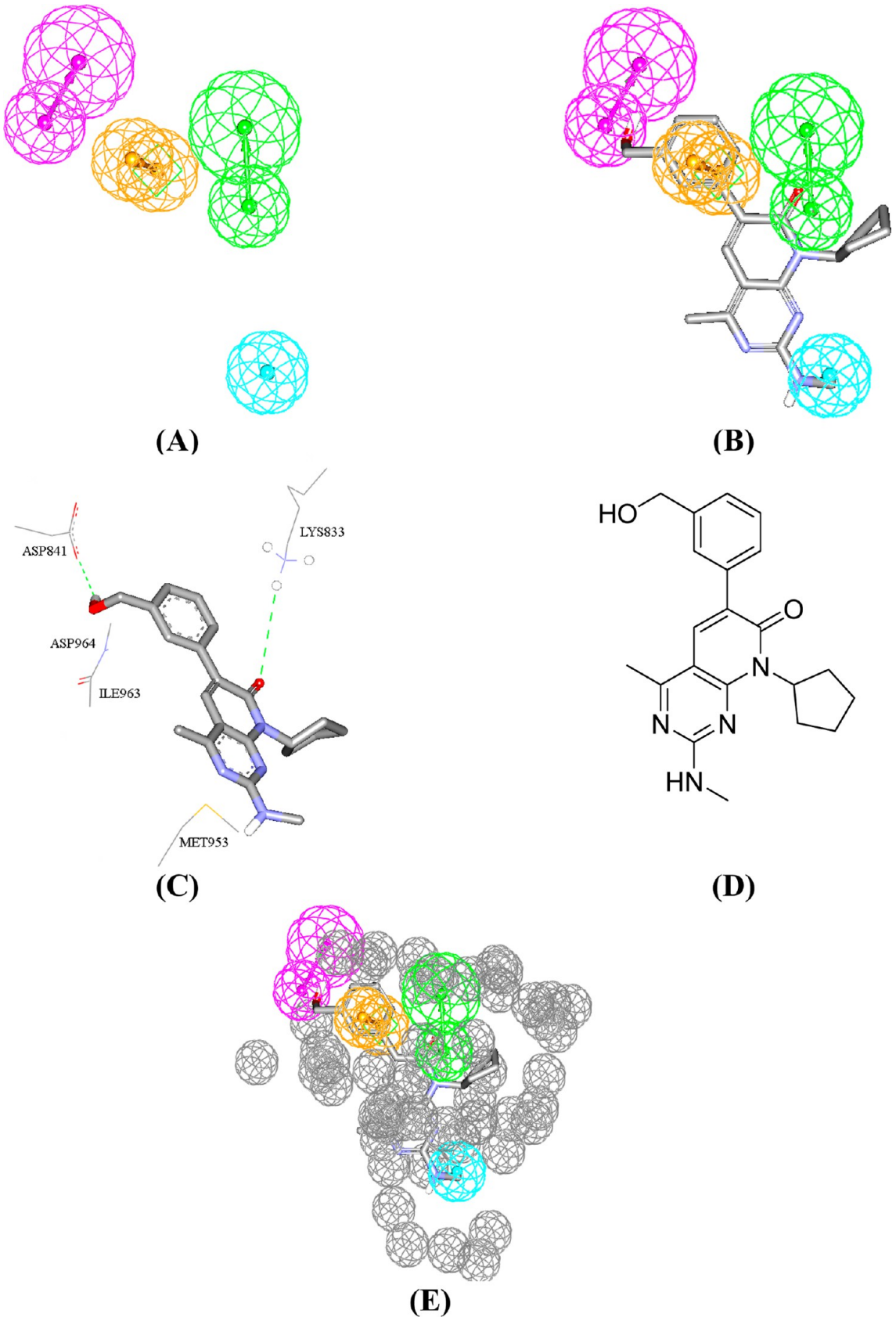


Figure 7. Hypo(C-T2-9). (A) Pharmacophoric features of the binding model: HBA as green vectored spheres, HBD as violet vectored spheres, and Hbic as blue spheres. (B) **Hypo(C-T2-9)** fitted against PI3K-γ cocrystallized ligand (pdb code 3ML8). (C) The key binding interactions of PI3K-γ cocrystallized ligand (pdb code 3ML8). (D) The chemical structures of the cocrystallized ligand. (E) HipHop-refined **Hypo(C-T2-9)** with exclusion volumes (gray spheres).

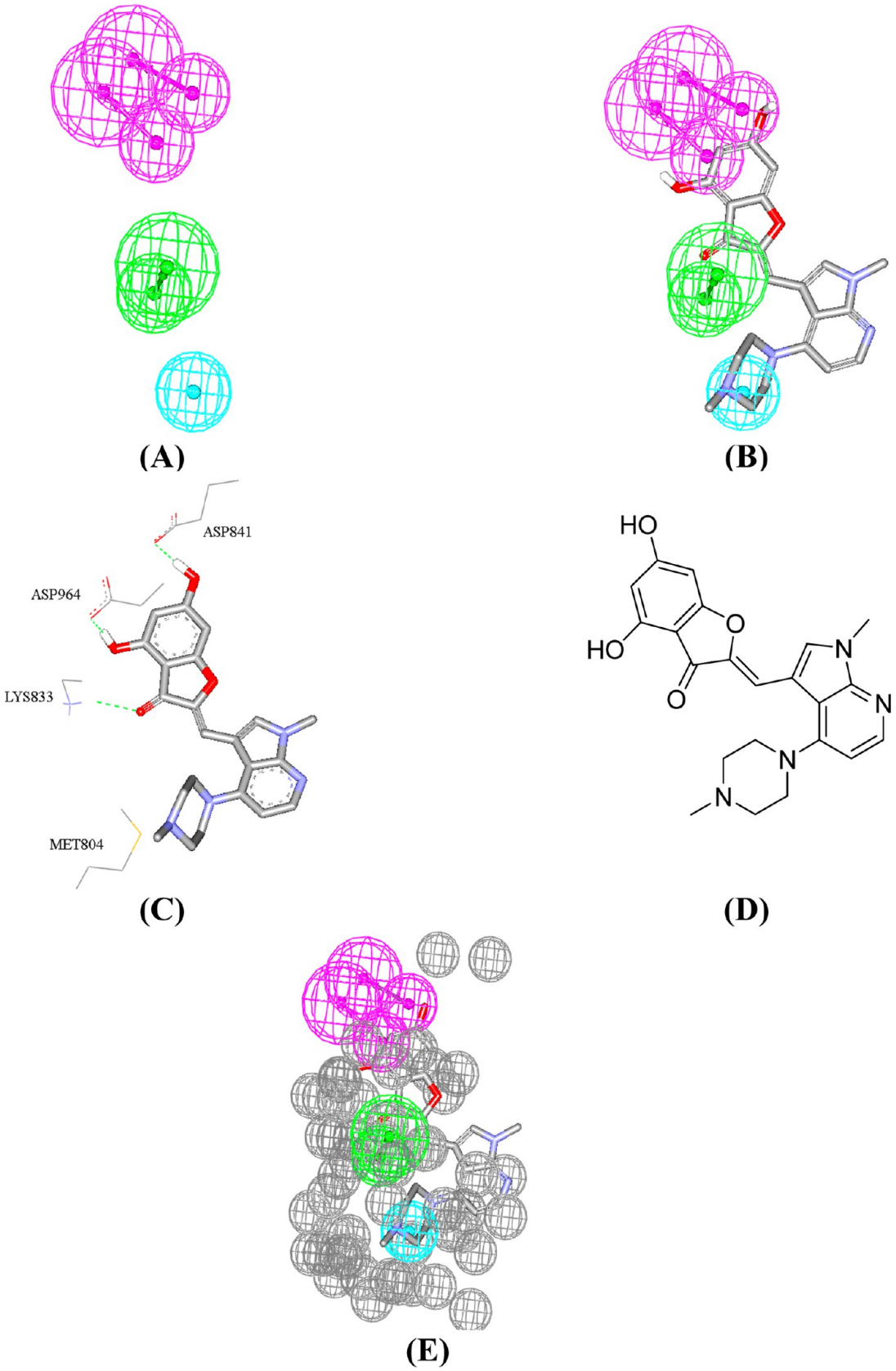


Figure 8. Hypo(E-T1-3). (A) Pharmacophoric features of the binding model: HBA as green vectored spheres, HBD as violet vectored spheres, and Hbic as blue spheres. (B) **Hypo(E-T1-3)** fitted against PI3K-γ cocrystallized ligand (pdb code 3LJ3). (C) The key binding interactions of PI3K-γ cocrystallized ligand (pdb code 3LJ3). (D) The chemical structures of the cocrystallized ligand. (E) HipHop-refined **Hypo(E-T1-3)** with exclusion volumes (gray spheres).

kNN-QSAR model 1 selected four one-dimensional descriptors encoding for the count of tertiary amines, aromatic sulfurs, and olefinic and aromatic carbon atoms. Moreover, it selected three pharmacophores as additional explanatory descriptors, namely, **Hypo(A-T6-8)**, **Hypo(E-T1-3)**, and **Hypo(C-T2-9)**. Interestingly, many of the descriptors in kNN-QSAR model 1 repeatedly emerged in other kNN-QSAR models, including the same three pharmacophore models, adding further weight to these descriptors.

The repeated appearance of the tertiary amine count descriptor in top ranking kNN-QSAR models is suggestive of a significant role played by amine moieties in ligand binding within mTOR probably through electrostatic attraction to acidic amino acid moieties in the binding pocket. The mTOR binding pocket contains two acidic amino acid moieties, Asp2357 and Asp2340.[43−45]

Similarly, the aromatic sulfur and olefinic and aromatic carbon atoms count descriptors probably encode for affinity interactions connecting different training ligands and hydrophobic moieties within the mTOR binding pocket. mTOR binding site contains several hydrophobic and aromatic moieties capable of π-stacking and hydrophobic interactions with various ligands, including, Ile2163, Ile2185, Ile2237, Ile2356, Met2345, Ala2248, and Tyr2225.[43−45]

The selection of three orthogonal pharmacophore binding models (average cross-correlation r^2 of 0.56) in the highest ranking kNN-QSAR model 1 (Table 3) further supports the notion of at least three binding modes assumed by inhibitors within the binding pocket of mTOR proposed by MLR-QSAR modeling mentioned in the previous section. Figures 6, 7, and 8 show **Hypo(A-T6-8)**, **Hypo(C-T2-9)**, and **Hypo(E-T1-3)** and how they map three cocrystallized ligands within the closely homologous protein PIK3-γ, while Table 1 shows the *X*, *Y*, and *Z* coordinates of the three pharmacophores. Interestingly, the three pharmacophores correspond nicely with binding interactions tying three cocrystallized ligands within the highly mTOR homologous protein analogue PI3K-γ.

Figure 6B shows how pharmacophore model **Hypo(A-T6-8)** maps a potent dual PIK3-γ and mTOR inhibitor compared with the corresponding PIK-3γ cocrystallized structure (Figure 6C). Mapping the terminal aminopyrimidine with HBD in **Hypo(A-T6-8)** correlates with hydrogen-bonding interactions connecting this amine with the carboxylate of Asp836. Similarly, mapping the sulfone oxygen within the cocrystallized ligand against a HBA feature in **Hypo(A-T6-8)** corresponds with a hydrogen-bonding interaction connecting the same atom with the ammonium Lys802. Moreover, mapping the ligand's methyl pyrimidine against a HBic feature in **Hypo(A-T6-8)** agrees with hydrophobic interactions anchoring this group with the hydrophobic side chain of Ile879. Finally, mapping the thiophene ring against a RingArom feature in **Hypo(A-T6-8)** agrees with sandwiching this ring between the sulfide moieties of Met804 and Met953.

A similar analogy can be seen upon comparing the cocrystallized pose of another potent and selective dual PIK3-γ and mTOR inhibitor with the way it maps **Hypo(C-T2-9)** (Figure 7B,C). Apparently, hydrogen-bonding interactions connecting the pyridone carbonyl and benzylic hydroxyl of the ligand to the ammonium and carboxylate side chains of Lys833 and Asp841, respectively, are represented by mapping the same carbonyl and hydroxyl groups against HBA and HBD features in **Hypo(C-T2-9)**, respectively. Similarly, mapping the terminal benzene ring against a RingArom feature in **Hypo(C-T2-9)** (Figure 7B) correlates with stacking interactions anchoring this ring against the peptidic amide joining Asp964 and Ile963. Finally, mapping the methyl of the methylaminopyrimidine against a HBic feature in **Hypo(C-T2-9)** (Figure 7B) corresponds to hydrophobic interactions tying this methyl with the sulfide side chain of Met953.

A similar comparison holds upon evaluating the bound pose of a third potent dual PIK3-γ/mTOR inhibitor within the PI3K-γ binding site with the way it fits **Hypo(E-T1-3)** (Figure 8B,C). Mapping the resorcinol hydroxyls against two HBD features in **Hypo(E-T1-3)** correlates with hydrogen bonding against the carboxylate side chains of Asp841 and Asp964. Comparably, hydrogen bonding connecting the benzofuranone carbonyl with the ammonium of Lys883 correlates with mapping the same carbonyl against a HBA in **Hypo(E-T1-3)** (Figure 8B,C). Finally, mapping the methylpipyrazine against a HBic feature in **Hypo(E-T1-3)** compares with hydrophobic interactions connecting the same group with the hydrophobic sulfide side chain of Met804 (Figure 8B,C).

To further validate our kNN-QSAR-selected pharmacophores, we subjected them to ROC curve analyses. The three pharmacophores showed excellent ROC performances with ROC-AUC values of 0.911, 0.869, and 0.909 for **Hypo(A-T6-8)**, **Hypo(C-T2-9)**, and **Hypo(E-T1-3)**, respectively (Table 2 and Figure C in Supporting Information).

2.3. In Silico Screening and Subsequent in Vitro Evaluation. Lack of steric constraints necessary to define the size of the binding pocket can render pharmacophore models rather promiscuous, that is, they can capture many false positive hits. Therefore, we decorated our pharmacophore models with appropriate exclusion spheres to resemble sterically inaccessible regions within mTOR's binding site. We employed the HipHop-REFINE module of CATALYST[47] for this purpose. A structurally diverse training subset was selected for HipHop-REFINE modeling (Table 4). The training compounds were selected in such a way that the bioactivities of inactive members are explained by steric clashes within the binding pocket (see section 4.1.7 and section **SM-5** in Supporting Information for more details). Figures 3E, 4E, 5E, 6E, 7E, and 8E show the sterically refined versions of the optimal pharmacophores, while Table 2 and Figures B and C in Supporting Information illustrate their corresponding ROC results. The sterically refined versions outperformed their unrefined counterparts, indicating significant improvements in their classification power upon addition of exclusion spheres.

We employed the sterically refined versions of optimal pharmacophores as 3D search queries to screen the NCI list of compounds (238 819 compounds) for new mTOR inhibitors. The captured hits were subsequently filtered by Lipinski's[48] and Veber's criteria.[49] Remaining hits were fitted against corresponding pharmacophores (fit values determined by eq D in Supporting Information) and their fit values were substituted in the MLR-based QSAR model (eq 1) or the kNN-based QSAR model (model 1 in Table 3) to determine their predicted bioactivities. Tables 5 and 6 and Figures 9 and 10 show the highest predicted hits, their QSAR-based predictions, and their experimental *in vitro* bioactivities.

Out of the 74 highest-ranking hits captured by the MLR-selected pharmacophores (eq 1), 53 were found to possess >50% anti-mTOR inhibitions at 10 μM prompting us to determine their anti-mTOR IC_{50} values (Table 5). Interestingly, 26 hits showed IC_{50} values within nanomolar range, while the rest were in the micromolar range. Figure D in Supporting

Table 4. The Training Compounds Used for Adding Excluded Spheres for All QSAR-Selected Pharmacophores (MLR- and kNN-Selected) Using HipHop-REFINE Module of CATALYST

compd[a]	IC_{50} (nM)	principal value	MaxOmitFeat[b]
40	0.1	2	0
58	0.1	2	0
59	0.1	2	0
29	0.2	2	0
55	0.2	2	0
64	0.2	2	0
20	0.22	2	0
47	0.3	2	0
37	0.6	2	0
43	0.9	2	0
31	500	0	2
108	950	0	2
121	970	0	2
201	1625	0	2
125	1650	0	2
82	3500	0	2
202	5000	0	2
126	7200	0	2
208	7300	0	2

[a]Compounds numbers are as in Figure A and Table A in Supporting Information. [b]MaxOmitFeat: maximum omitted features.

Information shows how MLR-selected pharmacophores map some of the most potent corresponding hits.

The dose−response curves of captured hits exhibit Hill slope values <1.0 and excellent correlation coefficients (Figure E in the Supporting Information), which strongly suggest the authenticity (i.e., nonpromiscuousity) of the inhibitors.[52] The NMR spectra and the exact mass of the most potent mTOR inhibitors are depicted in Figures G−L in Supporting Information.

On the other hand, out of the 27 highest-ranking hits captured by the kNN-based modeling strategy (Table 6), only 9 gave >50% inhibition at 10 μM. Upon further testing, five of them showed IC_{50} values within nanomolar range while the rest were within the micromolar range. Figure F in Supporting Information shows how kNN-selected pharmacophores map some of their potent hits.

These results suggest that MLR-based pharmacophores and QSAR exhibit superior success rate compared with their kNN-based counterparts in capturing potent hits. Still, the two methods seem to complement each other by capturing structurally distinct potent mTOR inhibitors.

In conclusion, combining linear and nonlinear modeling strategies provides better coverage of the conformational flexibility within mTOR's binding pocket and therefore better exploration of the bioactive chemical space of mTOR inhibitors. In fact, we believe this interesting combination of linear and nonlinear modeling methodologies can be implemented to effectively explore ligand chemical space of any other biological target.

Finally, to check the validity of our bioassay, we tested the procedure against a standard mTOR inhibitor (PF-04691502).[51] The measured IC_{50} value was found to be 76.8 nM, which is within reasonable range to the reported value (4 nM).[51]

2.4. Similarity Analysis between Training Compounds and Active Hits. Careful evaluation of Tables 5 and 6 shows discrepancies between experimental and QSAR-predicted bioactivities. We believe such prediction errors are due to the significant structural dissimilarity between training compounds and captured hits. Accordingly, in order to minimize the impact of any possible extrapolatory QSAR prediction errors on decisions regarding which hits merit subsequent in vitro testing,[50] we merely employed $\log(1/IC_{50})$ predictions to rank the corresponding hits and prioritize subsequent in vitro testing. Only the highest ranking hits were acquired for experimental validation.

In order to establish the structural dissimilarity between training compounds and captured active hits, we employed three library comparison methods implemented in Discovery Studio 2.5 to assess structural similarity and diversity between training compounds and captured active hits, namely, Murcko assemblies, Bayesian model, and global fingerprints.

In Murcko assemblies, the algorithm breaks the ligands of each library into unique occurrences of molecular rings, ring assemblies, bridge assemblies, chains, Murcko assemblies, or any combination of these. Murcko assemblies are contiguous ring systems plus chains that link two or more rings.[53] The two libraries are compared using a Tanimoto similarity of the assemblies based on the fragments that are common and unique to each library.[38]

On the other hand, in the Bayesian model approach, two Bayesian models were built, one to learn library A and one to learn library B. Finally, it scores all ligands using both models. A distance is computed as eq 2:

$$\text{distance} = \text{scoreAA} + \text{scoreBB} + \text{scoreAB} + \text{scoreBA} \quad (2)$$

where scoreAA is the average score of library A molecules scored by the Bayesian model that learned library A molecules, while scoreBB is the average score of library B molecules scored by the Bayesian model that learned library B. ScoreAB and scoreBA are the average scores of libraries A and B molecules scored by the Bayesian models that learned libraries B and A, respectively. The higher the distance, the more dissimilar the libraries are.[38]

Finally, the global fingerprint comparison algorithm generates a global fingerprint for all ligands in the training list and all ligands in the hits list and then computes a Tanimoto similarity coefficient between the two libraries.[38]

The three methods suggest minimal structural similarity between known mTOR inhibitors and captured hits (Tables 7 and 8), which probably explains the inconsistencies between experimental anti-TOR and QSAR predicted bioactivities in both MLR- and kNN-based QSAR models.

3. CONCLUSION

mTOR is currently considered a validated target for cancer therapy. The pharmacophoric space of mTOR inhibitors was explored via seven diverse training sets of compounds. Subsequently, GFA and MLR analysis was employed to access an optimal linear QSAR model. Moreover, we implemented a GFA-driven kNN-based modeling to access an optimal nonlinear QSAR model. Both approaches culminated in identification of several binding modes accessible to ligands within the mTOR binding site. The resulting QSAR models and associated pharmacophores were validated by the identification of 62 potent mTOR inhibitors retrieved from

Table 5. The Captured Hit Molecules with Their Fit Values, Their Corresponding MLR-QSAR Estimates from Eq 1 and Their *in Vitro* Bioactivities

hits		fit values against[b]			*in vitro* anti-mTOR activity[c]		
tested hits[a]	hit name	Hypo(A-T7-8)	Hypo(E-T5-8)	Hypo(G-T2-1)	predicted IC_{50} (nM)	% inhibition at 10 μM	experimental IC_{50} (nM)
211	NCI0032457	0	0	9.82	315.5	100	48.1
212	NCI0162404	0	5.46	8.83	18.5	100	93.9
213	NCI0328098	0	5.70	9.32	4.2	95	161.6
214	NCI0123517	6.81	8.10	0	0.38	100	162.5
215	NCI0348965	0	0	9.55	56.8	100	163.2
216	NCI0294133	5.62	0	9.69	16.1	84	186.8
217	NCI0031278	0	0	10.09	34.9	96	236.5
218	NCI0145408	0	4.18	9.08	16.2	91	245.2
219	NCI0019802	0	4.77	7.50	36.1	100	304.8
220	NCI0288051	0	0	9.92	11.4	100	325.6
221	NCI0045940	7.56	8.21	8.19	13.0	77	339.1
222	NCI0213858	0	5.02	9.55	1.1	100	350.1
223	NCI0291571	0	1.38	9.91	96.1	84	388.2
224	NCI0305180	4.72	7.35	5.90	4.8	95	393.4
225	NCI0205578	6.37	7.67	2.02	42.1	90	402.2
226	NCI0332542	0	5.97	7.75	5.8	88	456
227	NCI0245021	5.20	0	9.30	18.3	92	473.2
228	NCI0205709	0	4.93	9.00	3.45	100	489
229	NCI0137218	7.27	7.99	0	0.52	100	555.8
230	NCI0132098	6.46	8.06	0	0.68	89	689.8
231	NCI0602671	0	6.16	6.81	1.55	85	690.7
232	NCI0045942	4.44	7.81	5.71	25.3	95	705.1
233	NCI0040052	0	1.66	9.81	779.5	88	712.3
234	NCI0066756	0	4.24	8.02	153.0	86	756.4
235	NCI0114564	0	0	9.96	25.5	75	944.3
236	NCI0666767	0	4.08	7.47	10.6	76	986.6
237	NCI0251741	0.068	2.15	9.90	2.03	72	1108
238	NCI0114368	6.26	0	7.85	2.11	85	1202
239	NCI0114442	0	6.74	7.04	9.23	70	1302
240	NCI0084126	0	6.08	8.75	3.56	74	1450
241	NCI0102809	1.26	6.91	9.05	0.67	74	1552
242	NCI0608955	0	5.19	5.60	3.43	66	1563
243	NCI0329251	0	0.011	9.77	2.27	71	1727
244	NCI0205838	0	4.72	9.05	27.1	74	1790
245	NCI0143140	0	0	9.68	31.1	94	1930
246	NCI0185056	0	4.35	9.40	2.07	71	1951
247	NCI0215722	0	4.93	9.38	3.79	62	2059
248	NCI0319992	3.95	7.78	7.65	9.54	62	2315
249	NCI0379471	0.86	7.35	8.25	16.06	69	3327
250	NCI0270062	0	0	10.07	21.12	57	3580
251	NCI0034845	0	0	10.07	1.95	58	3672
252	NCI0118984	0	0	9.38	278.3	59	4126
253	NCI0145409	6.87	0	9.71	4.79	56	4492
254	NCI0665514	0	4.31	7.93	0.97	50	4532
255	NCI0185054	0	2.33	9.44	2.79	66	4596
256	NCI0185055	0	2.90	9.41	0.364	57	4777
257	NCI0366659	0	1.74	9.93	12.06	52	6558
258	NCI0211827	0.79	3.87	9.06	1.158	55	8117
259	NCI0291572	0.002	2.10	9.91	57.5	53	8439
260	NCI0212418	0	6.91	8.98	0.96	52	8774
261	NCI0013793	0	6.20	8.06	105.0	52	9204
262	NCI0134150	0	0	9.93	2.96	52	9271
263	NCI0133679	6.73	7.54	0	0.852	50	9853
264	NCI0665512	0	0	0	22.12	49	>10000
265	NCI0246978	0	0	0	182.6	46	>10000
266	NCI0204099	0	0	0	10369.5	42	>10000
267	NCI0185057	0	1.89	9.57	10.975	41	>10000
268	NCI0294402	5.793	0	9.72	36.8	38	>10000

Table 5. continued

hits		fit values against[b]			*in vitro* anti-mTOR activity[c]		
tested hits[a]	hit name	**Hypo(A-T7-8)**	**Hypo(E-T5-8)**	**Hypo(G-T2-1)**	predicted IC_{50} (nM)	% inhibition at 10 μM	experimental IC_{50} (nM)
269	NCI0319041	0.99	7.00	8.04	459.5	36	>10000
270	NCI0185052	0	1.70	9.17	49.8	36	>10000
271	NCI0290649	0	6.08	8.28	29.14	35	>10000
272	NCI0602692	8.68	7.71	9.94	2.44	28	>10000
273	NCI0114361	0	0	9.96	14.60	28	>10000
274	NCI0329253	0	0.91	9.89	7.09	28	>10000
275	NCI0147886	0	7.24	8.36	5.99	25	>10000
276	NCI0328131	0	3.53	8.34	29.95	25	>10000
277	NCI0162537	0	7.94	9.22	0.063	25	>10000
278	NCI0031279	0	0	9.51	119.4	22	>10000
279	NCI0204174	4.35	0	6.77	4.20	20	>10000
280	NCI0608329	0	7.57	5.31	101.0	12	>10000
281	NCI0291569	0.16	1.56	9.89	2.68	11	>10000
282	NCI0045941	3.78	7.65	8.11	21.46	10	>10000
283	NCI0608953	0	5.02	5.61	5.24	9	>10000
284	NCI0609070	7.29	5.67	10.13	0.78	6	>10000
PF-04691502[d]						100	76.8

[a]Compound numbers as in Figure 9. [b]Best-fit values against each binding hypothesis calculated by eq D in Supporting Information. [c]Bioactivity values are the average of at least duplicate measurements. [d]PF-04691502 is the standard positive control applied in mTOR inhibitory assay. The reported IC_{50} of PF-04691502 is 4 nM.[51]

Table 6. The Captured Hit Molecules with Their Fit Values, Their Corresponding kNN-QSAR Estimates and Their *in Vitro* Bioactivities

hits		fit values against[b]			*in vitro* anti-mTOR activity[c]		
tested hits[a]	hit name	**Hypo(A-T6-8)**	**Hypo(C-T2-9)**	**Hypo(E-T1-3)**	predicted IC_{50} (nM)	% inhibition at 10.0 μM	experimental IC_{50} (nM)
285	NCI0659390	6.93	7.27	5.64	1.67	100	211
286	NCI0309121	5.01	6.43	7.02	42.82	100	310
287	NCI0603664	6.08	5.89	6.31	2.66	86	815
288	NCI0359466	7.14	7.00	6.72	1.67	95	879
289	NCI0134179	6.52	7.56	6.08	1.67	58	912
290	NCI0153166	4.87	5.94	6.04	42.82	81	1120
291	NCI0353681	6.76	6.27	6.63	2.66	77	2489
292	NCI0067736	6.14	6.58	7.64	3.25	60.5	4800
293	NCI0215649	4.78	5.84	6.80	42.82	71	5711
294	NCI0117269	3.45	7.98	8.30	0.49	47	>10000
295	NCI0109161	6.42	7.54	8.29	0.49	27	>10000
296	NCI0221018	4.62	5.68	6.95	15.96	44	>10000
297	NCI0403440	6.46	8.22	7.96	0.49	22	>10000
298	NCI0375162	8.29	4.68	4.48	6.89	8	>10000
299	NCI0366657	0.65	9.20	8.45	14.58	36	>10000
300	NCI0680410	4.92	6.98	7.08	27.01	29	>10000
301	NCI0672070	1.74	9.24	8.47	4.05	27	>10000
302	NCI0667562	5.78	7.55	8.35	0.49	48	>10000
303	NCI0667561	5.78	7.55	8.35	0.49	42	>10000
304	NCI0063688	6.38	6.43	7.45	2.66	18	>10000
305	NCI0062766	6.33	6.53	6.51	2.66	12	>10000
306	NCI0052105	5.76	8.19	8.07	0.49	7	>10000
307	NCI0038278	7.15	8.53	2.37	9.45	33	>10000
308	NCI0012749	5.04	6.19	7.01	42.82	41	>10000
309	NCI0337610	2.34	9.22	8.45	4.05	16	>10000
310	NCI0332448	1.85	7.60	8.10	13.57	29	>10000
311	NCI0120183	5.92	6.32	5.57	2.60	12	>10000
PF-04691502[d]						100	76.8

[a]Compound numbers as in Figure 10. [b]Best-fit values against each binding hypothesis calculated by eq D in Supporting Information. [c]Bioactivity values are the average of at least duplicate measurements. [d]PF-04691502 is the standard positive control applied in mTOR inhibitory assay. The reported IC_{50} of PF-04691502 is 4 nM.[51]

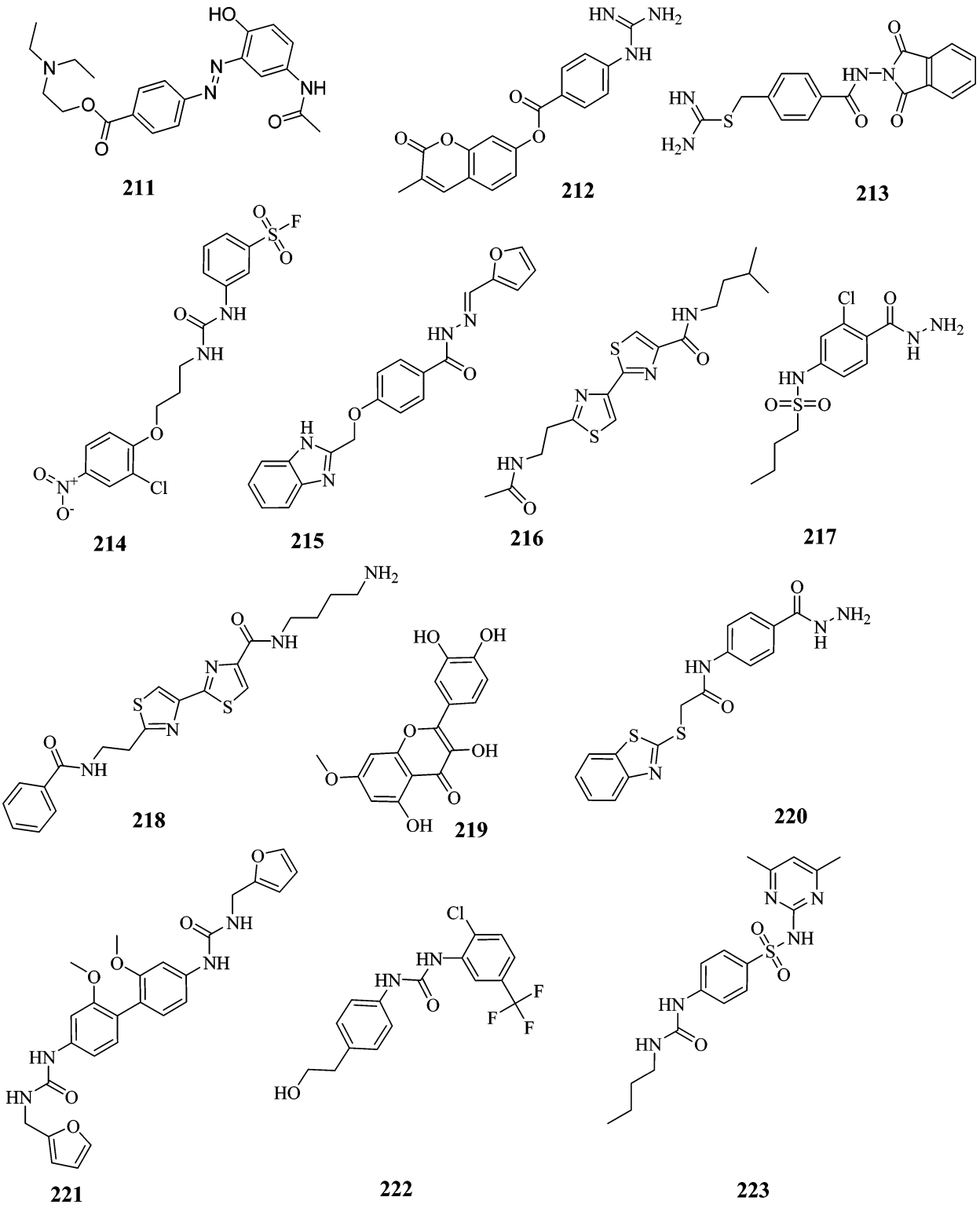


Figure 9. continued

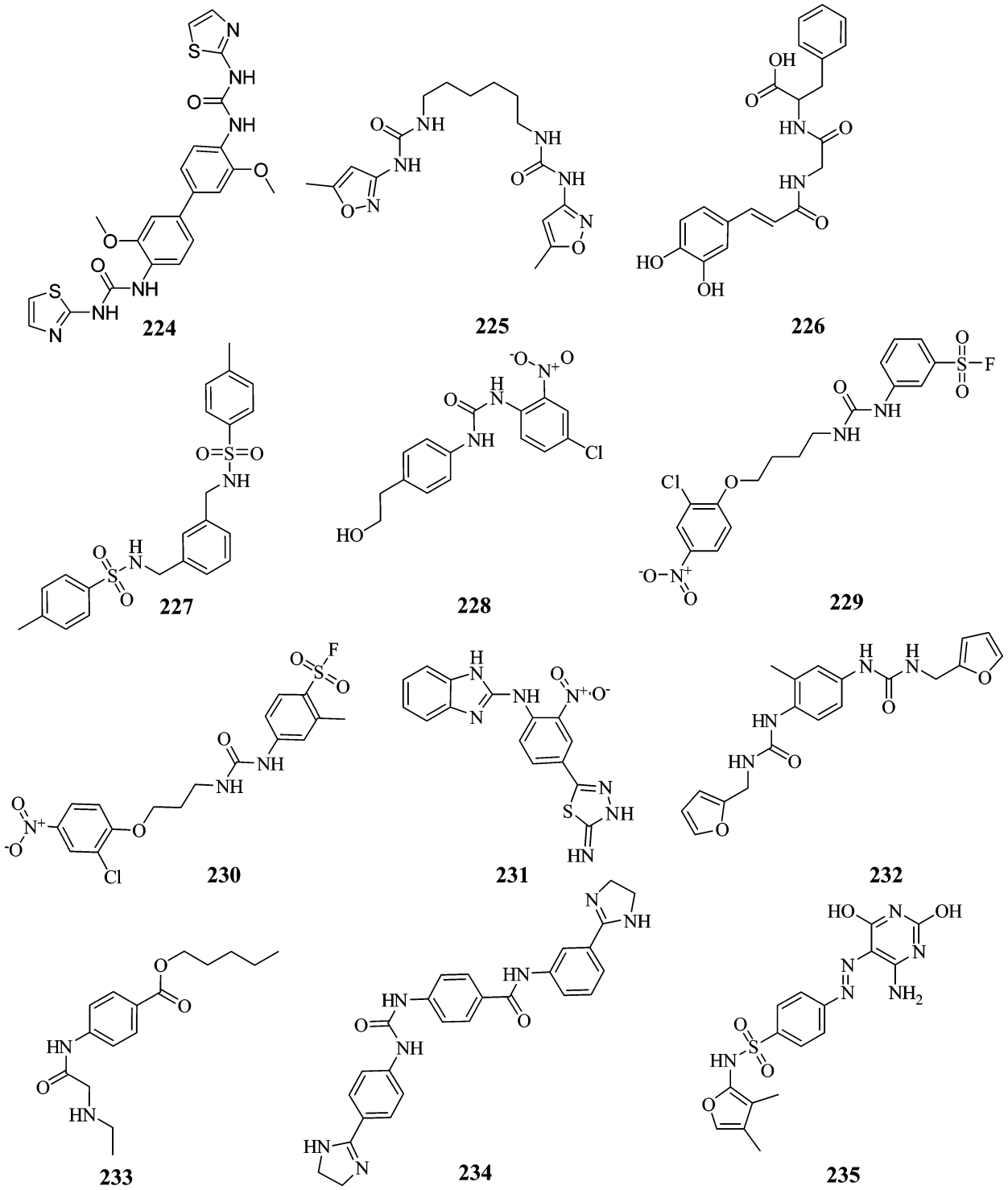


Figure 9. continued

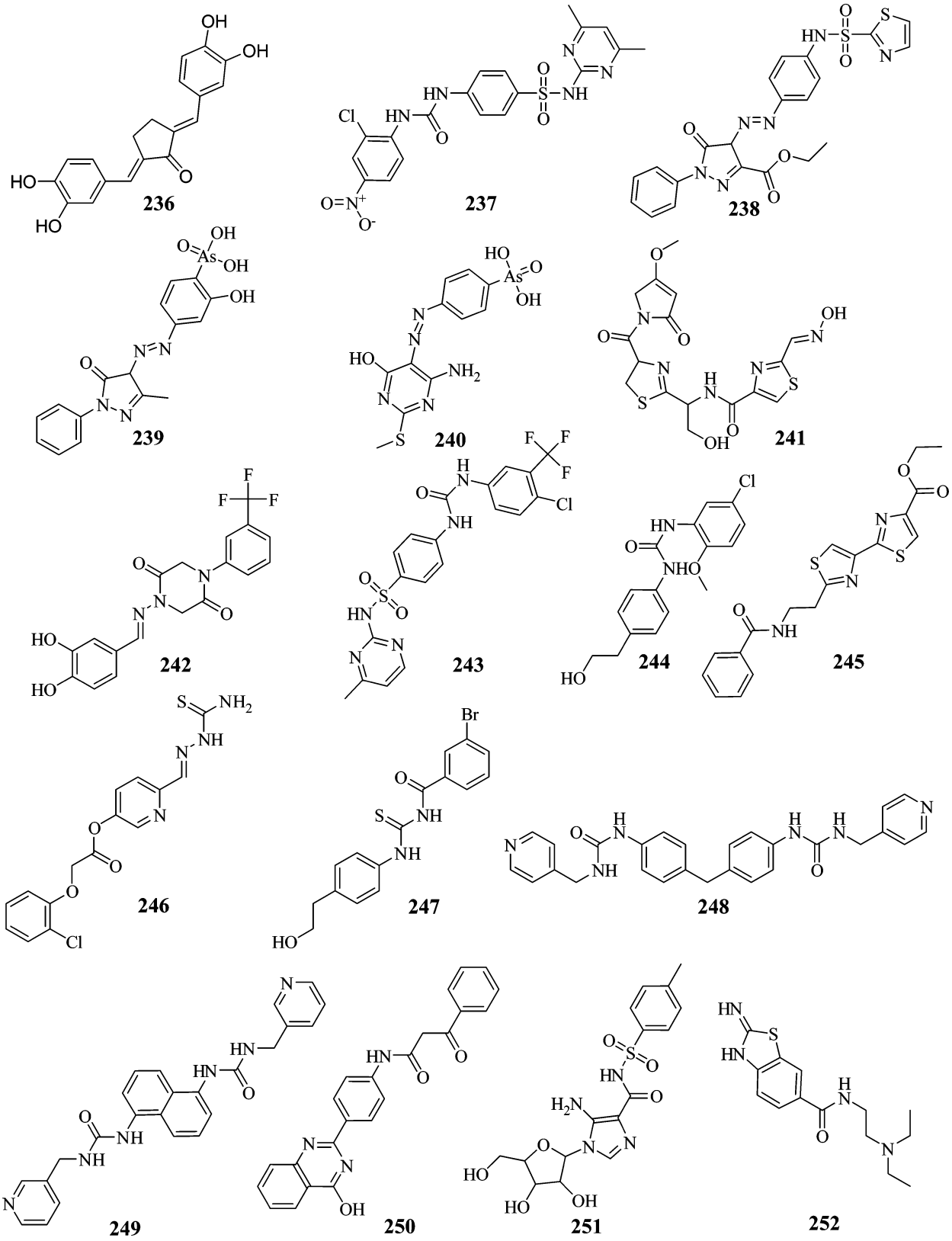


Figure 9. continued

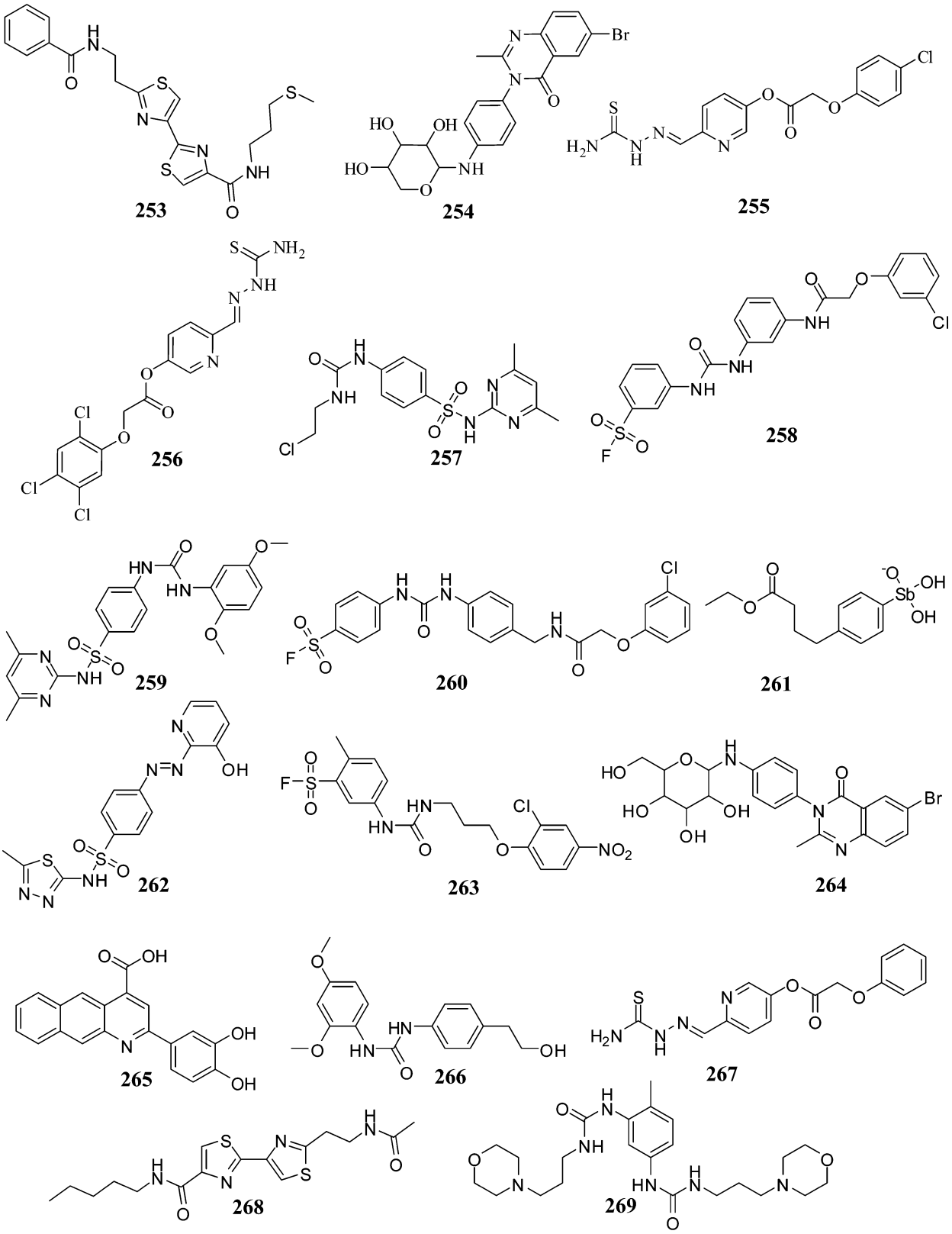


Figure 9. continued

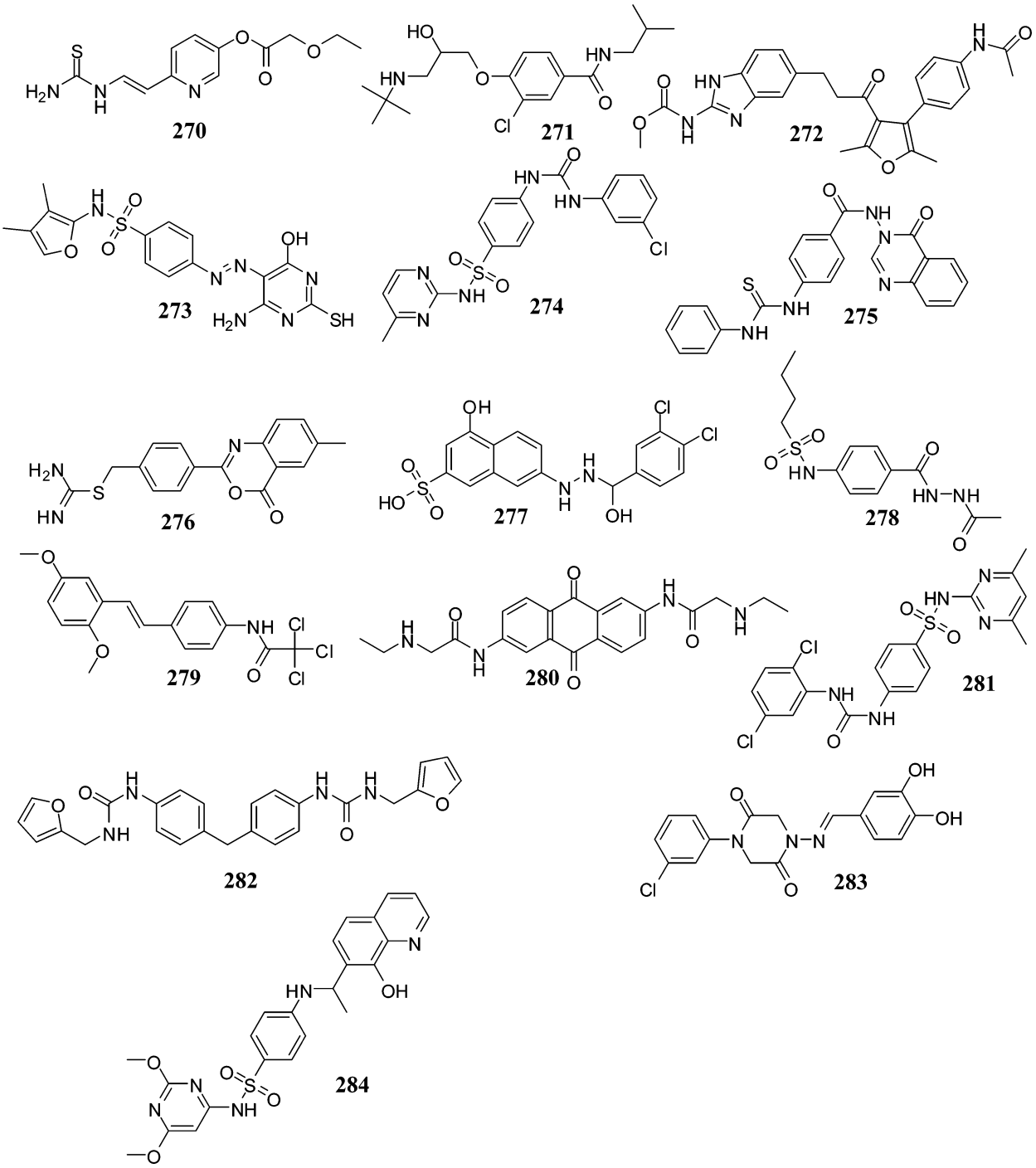


Figure 9. The chemical structures of the tested highest-ranking mTOR hits predicted by the MLR-QSAR model (eq 1) and associated pharmacophores.

the NCI structural database. The most potent hit illustrated an anti-mTOR IC_{50} value of 48 nM.

4. MATERIALS AND METHODS

4.1. Molecular Modeling. The HYPOGEN module from the CATALYST software package was employed to construct numerous plausible binding hypotheses for mTOR inhibitors.[20−32] The conformational space of each inhibitor (**1−210**, Figure A and Table A in Supporting Information) was explored adopting the "CAESAR" option within CATALYST.[35,36] Detailed experimental and theoretical explanations of pharmacophore modeling and conformational analysis are provided in the Supporting Information (section SM-1, SM2, and SM-3).

4.1.1. Data Set. The structures of 210 mTOR inhibitors (Figure A and Table A in Supporting Information) were collected from articles published by a single research group,[20−32] which strongly supports the notion that their *in vitro* bioactivities were determined by a single assay procedure. The bioactivities were expressed as the concentration of the test compound that inhibited the activity of mTOR by 50% (IC_{50}). The logarithm of measured IC_{50} (nM) values was used in

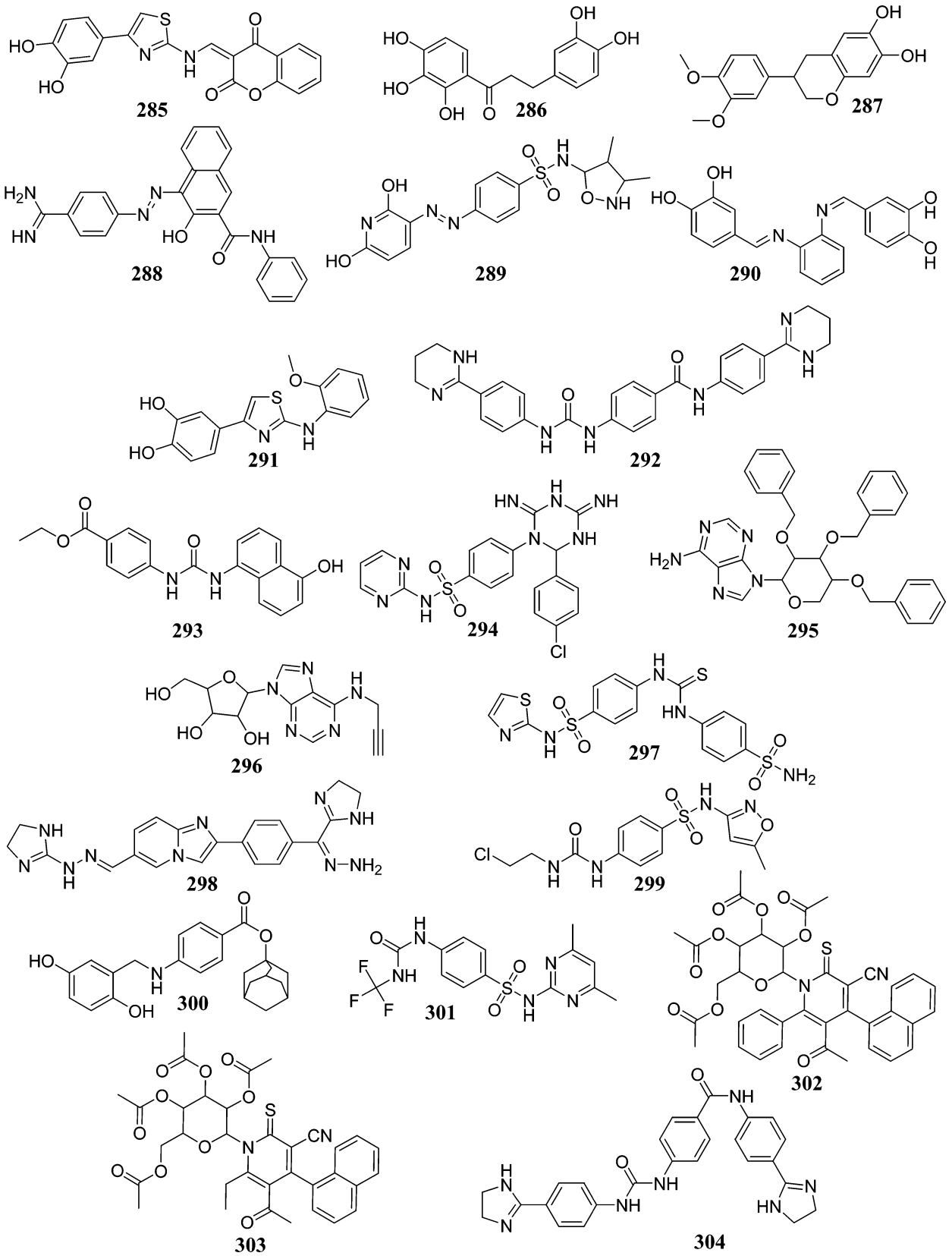


Figure 10. continued

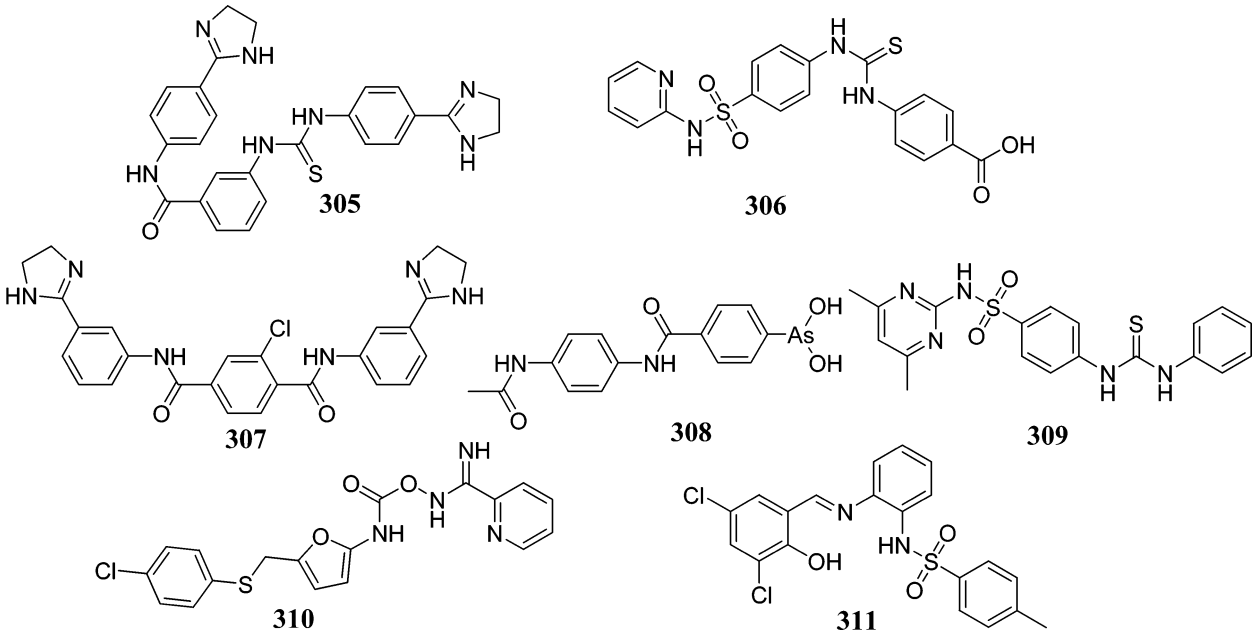


Figure 10. The chemical structures of the tested highest-ranking mTOR hits predicted by the kNN-QSAR model 1 and associated pharmacophores.

Table 7. Results of Similarity Analysis between Training Compounds and Active Hits Captured by MLR-QSAR-Selected Pharmacophores (Hypo(A-T7-8), Hypo(E-T5-8), and Hypo(G-T2-1))

Murcko assemblies		Bayesian model		global fingerprints[a]	
number of total assemblies	178	average LibA score of library A ligands	31.33	number of total global fingerprint bits	2088
number of common assemblies	6	average LibB score of library A ligands	−37.02	number of common global fingerprint bits	223
number of assemblies only in library A[b]	46	average LibA score of library B ligands	−93.33	number of global fingerprint bits only in library A	1025
number of assemblies only in library B[c]	126	average LibB score of library B ligands	8.61	number of global fingerprint bits only in library B	840
similarity score between the two libraries	0.034	Bayesian distance between the two libraries	170.29	similarity score between the two libraries	0.107

[a]Done by implementing the fingerprint descriptor FCFC_6, which correspond to functional-class extended-connectivity fingerprint count up to diameter 6. [b]Library A list includes all training and testing compounds employed in pharmacophore and QSAR modeling (**1−210**, Figure A and Table A in Supporting Information). [c]Library B includes hits captured by MLR-QSAR modeling (**211−284**, Figure 9 and Table 5).

Table 8. Results of Similarity Analysis between Training Compounds and Active Hits Active Hits Captured by kNN-QSAR-Selected Pharmacophores (Hypo(A-T6-8), Hypo(C-T2-9), and Hypo(E-T1-3))

Murcko assemblies		Bayesian model		global fingerprints[a]	
number of total assemblies	152	average LibA score of library A ligands	28.89	number of total global fingerprint bits	1558
number of common assemblies	3	average LibB score of library A ligands	−22.35	number of common global fingerprint bits	148
number of assemblies only in library A[b]	20	average LibA score of library B ligands	−67.33	number of global fingerprint bits only in library A	495
number of assemblies only in library B[c]	129	average LibB score of library B ligands	2.75	number of global fingerprint bits only in library B	915
similarity score between the two libraries	0.0197	Bayesian distance between the two libraries	121.33	similarity score between the two libraries	0.095

[a]Done by implementing the fingerprint descriptor FCFC_6, which correspond to functional-class extended-connectivity fingerprint count up to diameter 6. [b]Library A list includes all training and testing compounds employed in pharmacophore and QSAR modeling (**1−210**, Figure A and Table A in Supporting Information). [c]Library B includes hits captured by kNN-QSAR modeling (**285−311**, Figure 10 and Table 6).

pharmacophore modeling and QSAR analysis, thus correlating the data linearly to the free energy change.

The chemical structures of the inhibitors were converted into corresponding standard 3D structures and energy minimized to the closest local minimum using the molecular mechanics CHARMm force field. The resulting 3D structures were utilized as starting conformers for conformational analysis for pharmacophore modeling.

4.1.2. Pharmacophoric Hypotheses Generation. Seven structurally diverse training subsets (Table B in Supporting Information) were carefully selected from the collected compounds for pharmacophore modeling. Each training subset was utilized to conduct eight modeling runs to explore the pharmacophoric space of mTOR inhibitors. Different hypotheses were generated by altering the interfeature spacing and the number of allowed features in the resulting pharmacophores (see Table C in Supporting Information).

Eventually, pharmacophore exploration (eight automatic runs, Tables C and D in Supporting Information) culminated in 560 pharmacophore models of variable qualities (See SM-2 in Supporting Information for details about CATALYST pharmacophore generation algorithm).[33,37]

4.1.3. Assessment of the Generated Hypotheses. When generating hypotheses, CATALYST attempts to minimize a cost function consisting of three terms: weight cost, error cost, and configuration cost.[36,38,39] A total of 559 pharmacophores, out of 560 generated models, were found to possess Fisher confidence values ≥90% (see section SM-3 in Supporting Information). Tables C and D in Supporting Information show the success criteria of representative pharmacophores from each run. Detailed theoretical explanations of CATALYST's assessment of binding hypotheses are provided in SM-3 in the Supporting Information.

4.1.4. Clustering of the Generated Pharmacophore Hypotheses. The successful models (559) were clustered into 112 groups utilizing the hierarchical average linkage method available in CATALYST. Therefore, closely related pharmacophores were grouped in five-membered clusters. Subsequently, the highest-ranking representatives, as judged based on their fit-to-bioactivity correlation r^2-values (calculated against collected compounds **1−210**), were selected to represent their corresponding clusters in subsequent QSAR modeling (Table D in Supporting Information).

4.1.5. Genetic Function Algorithm-Based QSAR Modeling. GFA techniques rely on the evolutionary operations of "crossover and mutation" to select optimal combination of descriptors capable of explaining bioactivity variation among training compounds. GFA operates through a cycle of four stages: (i) encoding mechanism; (ii) definition of a fitness function; (iii) creating a population of chromosomes; (iv) genetic manipulation of chromosomes.[40] We implemented a gene-based encoding system. In this scheme, the possible models (chromosomes) differ from one another by the set of independent variables (descriptors) that comprise each model. If the general number of independent variables (descriptors) is equal to P (in this particular case, P = 431 variables corresponding to 112 pharmacophore fit values and 319 calculated descriptors, see below), the chromosome corresponding to any model consists of a string of P binary digits (bits) called "genes". Each value in the string represents an independent variable (0 = absent, 1 = present). Each chromosome is associated with a fitness value that reflects how good it is compared with other solutions. The following are important control parameters used in the GFA-based selection of optimal descriptors:

- Creating an initial population: The user must specify a number of initial random chromosomes.
- Mating population: Mating is an operation during which two parent chromosomes are combined to generate new solutions (offspring).
- Mutation operator: This operator modifies any single chromosome with a given probability, which can take values between 0.0 and 1.0. A mutation operator changes one or more bits in the chromosome to its complement.
- Maximum number of generations: This is needed for an exit from a basic cycle and completion of the algorithm.[40]

The independent descriptors were generated as follows: The chemical structures of the inhibitors were imported into Discovery Studio (version 2.55) as standard 3D single conformer representations in SD format. Subsequently, different descriptor groups were calculated for each compound employing the C2.DESCRIPTOR module within Discovery Studio. The calculated descriptors were 319 properties that included various simple and valence connectivity indices, electrotopological state indices, and other molecular descriptors (e.g., logarithm of partition coefficient, polarizability, dipole moment, molecular volume, molecular weight, molecular surface area, energies of the lowest and highest occupied molecular orbitals, etc.).[38] Furthermore, the training compounds were fitted (using the Best-fit option in CATALYST) against the representative pharmacophores (112 models, Table D in Supporting Information), and their fit values were added as additional descriptors. The fit value for any compound is obtained automatically via eq D, Supporting Information.[36]

4.1.5.1. MLR-Based Selection of Descriptors. GFA was employed to search for the best possible QSAR regression equation capable of correlating the variations in biological activities of the training compounds with variations in the generated descriptors, that is, MLR modeling. The fitness function employed herein is based on Friedman's "lack-of-fit" (LOF). The following GFA parameters were employed: explore linear, quadratic, and spline equations at mating and mutation probabilities of 50%; population size = 500; number of genetic iterations (generations) = 10000; LOF smoothness parameter = 0.5. However, to determine the optimal number of explanatory terms (QSAR descriptors), it was decided to scan and evaluate all possible QSAR models resulting from 4 to 10 explanatory terms.

All QSAR models were validated employing leave one-out cross-validation (r^2_{LOO}), and predictive r^2 (r^2_{PRESS}) calculated from the randomly selected external test subset (see selection criteria mentioned earlier).

Predictive r^2_{PRESS} is defined as

$$r_{PRESS}{}^2 = (SD - PRESS)/SD \quad (3)$$

where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of the training set molecules and PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set.

A subset of 168 compounds from the total list of inhibitors (**1−210**) was utilized as a training set for QSAR modeling. However, since it is essential to assess the predictive power of the resulting QSAR models on an external set of inhibitors, the remaining 42 molecules (ca. 20% of the data set) were employed as an external test subset for validating the QSAR models. (Figure A and Table A in Supporting Information). The test molecules were selected as follows: the collected inhibitors (**1−210**, Figure A and Table A in Supporting Information) were ranked according to their IC_{50} values, and then every fifth compound was selected for the test set starting from the high-potency end. In this way, the test molecules represent a range of biological activities similar to that of the training set.

4.1.5.2. KNN-Based Descriptor Selection. The kNN-QSAR methodology relies on a distance learning approach such that the activity value of an unknown member is calculated from the activity values of certain number (k) of nearest neighbors (kNNs) in the training set. The similarity is measured by a distance metric and in the present study the Euclidean distance is considered. The standard kNN method is implemented through the following workflow: (i) calculate distances between

an unknown object (e.g., x) and all the objects in the training set; (ii) select k objects from the training set most similar to object x, according to the calculated distances; (iii) calculate the activity value of object x as a weighted average of the activities of its kNNs. The best k value has been found empirically to lie between 1 and 5.[41,42] In our kNN approach, 20% of the observations are left out of the training set, and their activities are predicted as the weighted average. The process is repeated over five cycles such that in each cycle the selected testing set is different from those for the other cycles. The predicted activity value of each compound is calculated as weighted average of its nearest neighbors using the following formula:

$$\overline{y_x} = \frac{\sum_{k\text{-nearest neighbors}} y_i d_i}{\sum_{k\text{-nearest neighbors}} d_i} \tag{4}$$

where $\overline{y_x}$ is the predicted activity of compound x, y_i represent the activities of the nearest k-neighbors, and d_i is the Euclidean distance of the compound from its kNNs. The leave 20%-out cross-validated coefficient is calculated using the formula

$$r_{\mathrm{L20\%O}}{}^2 = 1 - \frac{\sum_{x=1}^{\text{training set}} (y_x - \overline{y_x})^2}{\sum_{x=1}^{\text{training set}} (y_x - y_{\mathrm{avg.tr}})^2} \tag{5}$$

where y_x is the experimental bioactivity of compound x and $y_{\mathrm{avg.tr}}$ is the average bioactivity of training compounds (i.e., after excluding the testing set).

GFA was employed to search for the best possible combination of descriptors capable of explaining variation in biological activities of training compounds via reasonable kNN model. The fitness function employed herein is $r^2_{\mathrm{L20\%O}}$. The following GFA parameters were employed: explore a combination of 1−10 descriptors using Gaussian-based random mutation and a mating probability of 80%; population size = 100; number of genetic iterations (generations) = 200.

4.1.6. ROC Curve Analysis. Successful GFA-MLR or GFA-kNN selected pharmacophore models were validated by assessing their abilities to selectively capture diverse mTOR inhibitors from a large list of decoys employing ROC analysis as described by Verdonk and co-workers.[54−56] For each active compound in the testing set, an average of 41 decoys were randomly chosen from the ZINC database.[57] See section SM-4 in the Supporting Information for detailed experimental and theoretical explanations of ROC analysis.

4.1.7. Addition of Exclusion Volumes. To account for the steric constraints of the binding pocket and to optimize the ROC curves of our QSAR-selected pharmacophores, it was decided to add exclusion volumes to the successful GFA-MLR or GFA-kNN selected pharmacophore models employing the HipHop-REFINE module of CATALYST. HipHop-REFINE uses inactive training compounds to add exclusion spheres to resemble the steric constraints of the binding pocket. It identifies spaces occupied by the conformations of inactive compounds and free from active ones. These regions are then filled with excluded volumes.[47] More details are provided in the Supporting Information (section SM-5).

4.2. Bioassay of Captured Hits. Briefly, recombinant mTOR was purchased from Invitrogen (Carlsbad, CA). The mTOR kinase assays were carried out with the Invitrogen Z′-LYTE kinase assay kit - Ser/Thr 11 peptide. The assay was optimized for use with mTOR as described in the Invitrogen protocol. The mTOR concentration was optimized to obtain the desired percent phosphorylation with an acceptable Z′-factor value, which indicates the quality of an assay; Z′-factor values of 0.5 or greater classify an assay as excellent. A Z′-factor value of 0.74 was obtained at final kinase and ATP concentrations of 14 nM and 100 μM, respectively. Tested concentrations ranged from 10 nM to 10 μM distributed log-linearly across the concentration range, and at least two data points from each concentration were collected. The IC_{50} value for each experiment was obtained using nonlinear regression of the log(concentration) versus percent inhibition values (GraphPad Prism 5.0). The assay conditions were validated by running positive (PF-04691502) and negative (provided in Z′-LYTE Kinase Assay kit) controls.

■ ASSOCIATED CONTENT

Supporting Information

The detailed theoretical and experimental procedures of pharmacophoric and QSAR modeling and analytical data of active hits discovered in this study. This information is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author

*Telephone: 00962 6 5355000, ext. 23305. Fax: 00962 6 5339649. E-mail: mutasem@ju.edu.jo.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for their generous funds. The authors are also thankful to the National Cancer Institute for freely providing NCI hits.

■ ABBREVIATIONS USED

GFA, genetic function algorithm; Hbic, hydrophobic; kNN, k nearest neighbor; LOF, lack-of-fit; MLR, multiple linear regression; mTOR, mammalian target of rapamycin; NCI, National Cancer Institute; RingArom, ring aromatic; ROC, receiver operating characteristic

■ REFERENCES

(1) Kim, D. H.; Sarbassov, D. D.; Ali, S. M.; King, J. E.; Latek, R. R.; Erdjument-Bromage, H.; Tempst, P.; Sabatini, D. M. mTOR Interacts with Raptor to Form a Nutrient-Sensitive Complex That Signals to the Cell Growth Machinery. *Cell* **2002**, *110*, 163−175.

(2) Hay, N.; Sonenberg, N. Upstream and Downstream of mTOR. *Genes Dev.* **2004**, *18*, 1926−1945.

(3) Meric-Bernstam, F.; Gonzalez-Angulo, A. M. Targeting the mTOR Signaling Network for Cancer Therapy. *J. Clin. Oncol.* **2009**, *27*, 2278−2287.

(4) Engelman, J. A. Targeting PI3K Signalling in Cancer: Opportunities, Challenges and Limitations. *Nat. Rev. Cancer* **2009**, *9*, 550−562.

(5) Land, S. C.; Tee, A. R. Hypoxia-Inducible Factor 1alpha is Regulated by the Mammalian Target of Rapamycin (mTOR) Via an mTOR Signaling Motif. *J. Biol. Chem.* **2007**, *282*, 20534−20543.

(6) Gao, N.; Flynn, D. C.; Zhang, Z.; Zhong, X. S.; Walker, V.; Liu, K. J.; Shi, X.; Jiang, B. H. G1 Cell Cycle Progression and the Expression of G1 Cyclins are Regulated by PI3K/AKT/mTOR/p70S6K1Signaling in Human Ovarian Cancer Cells. *Am. J. Physiol. Cell Physiol.* **2004**, *287*, C281−291.

(7) Don, A. S.; Zheng, X. F. Recent Clinical Trials of mTOR-Targeted Cancer Therapies. *Rev. Recent Clin. Trials* **2011**, *6*, 24−35.

(8) Chano, T.; Okabe, H.; Hulette, C. M. RB1CC1 Insufficiency Causes Neuronal Atrophy through mTOR Signaling Alteration and Involved in the Pathology of Alzheimer's Diseases. *Brain Res.* **2007**, *1168*, 97−105.

(9) Caccamo, A.; Majumder, S.; Richardson, A.; Strong, R.; Oddo, S. Molecular Interplay between Mammalian Target of Rapamycin (mTOR), Amyloid-Beta, and Tau: Effects on Cognitive Impairments. *J. Biol. Chem.* **2010**, *285*, 13107−13120.

(10) Zoncu, R.; Efeyan, A.; Sabatini, D. M. mTOR: From Growth Signal Integration to Cancer, Diabetes and Ageing. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 21−35.

(11) Di Paolo, S.; Teutonico, A.; Leogrande, D.; Capobianco, C.; Schena, P. F. Chronic Inhibition of Mammalian Target of Rapamycin Signaling Downregulates Insulin Receptor Substrates 1 and 2 and AKT Activation: A Crossroad between Cancer and Diabetes? *J. Am. Soc. Nephrol.* **2006**, *17*, 2236−2244.

(12) Taha, M. O.; Bustanji, Y.; Al-Ghussein, M. A.; Mohammad, M.; Zalloum, H.; Al-Masri, I. M.; Atallah, N. Pharmacophore Modeling, Quantitative Structure-Activity Relationship Analysis, and in Silico Screening Reveal Potent Glycogen Synthase Kinase-3beta Inhibitory Activities for Cimetidine, Hydroxychloroquine, and Gemifloxacin. *J. Med. Chem.* **2008**, *51*, 2062−2077.

(13) Al-Nadaf, A.; Abu Sheikha, G.; Taha, M. O. Elaborate Ligand-Based Pharmacophore Exploration and QSAR Analysis Guide the Synthesis of Novel Pyridinium-Based Potent Beta-Secretase Inhibitory Leads. *Bioorg. Med. Chem.* **2010**, *18*, 3088−3115.

(14) Al-Sha'er, M. A.; Taha, M. O. Discovery of Novel CDK1 Inhibitors by Combining Pharmacophore Modeling, QSAR Analysis and in Silico Screening Followed by in Vitro Bioassay. *Eur. J. Med. Chem.* **2010**, *45*, 4316−4330.

(15) Abdula, A. M.; Khalaf, R. A.; Mubarak, M. S.; Taha, M. O. Discovery of New Beta-D-Galactosidase Inhibitors via Pharmacophore Modeling and QSAR Analysis Followed by in silico Screening. *J .Comput. Chem.* **2011**, *32*, 463−482.

(16) Habash, M.; Taha, M. O. Ligand-Based Modelling Followed by Synthetic Exploration Unveil Novel Glycogen Phosphorylase Inhibitory Leads. *Bioorg. Med. Chem.* **2011**, *19*, 4746−4771.

(17) Shahin, R.; Alqtaishat, S.; Taha, M. O. Elaborate Ligand-Based Modeling Reveal New Submicromolar Rho Kinase Inhibitors. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 249−266.

(18) Suaifan, G. A.; Shehadehh, M.; Al-Ijel, H.; Taha, M. O. Extensive Ligand-Based Modeling and in Silico Screening Reveal Nanomolar Inducible Nitric Oxide Synthase (iNOS) Inhibitors. *J. Mol. Graphics Modell.* **2012**, *37*, 1−26.

(19) Shahin, R.; Taha, M. O. Elaborate Ligand-Based Modeling and Subsequent Synthetic Exploration Unveil New Nanomolar Ca2+/Calmodulin-Dependent Protein Kinase II Inhibitory Leads. *Bioorg. Med. Chem.* **2012**, *20*, 377−400.

(20) Zask, A.; Verheijen, J. C.; Richard, D. J.; Kaplan, J.; Curran, K.; Toral-Barza, L.; Lucas, J.; Hollander, I.; Yu, K. Discovery of 2-Ureidophenyltriazines Bearing Bridged Morpholines as Potent and Selective ATP-Competitive mTOR Inhibitors. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2644−2647.

(21) Curran, K. J.; Verheijen, J. C.; Kaplan, J.; Richard, D. J.; Toral-Barza, L.; Hollander, I.; Lucas, J.; Ayral-Kaloustian, S.; Yu, K.; Zask, A. Pyrazolopyrimidines as Highly Potent and Selective, ATP-Competitive Inhibitors of the Mammalian Target of Rapamycin (mTOR): Optimization of the 1-Substituent. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 1440−1444.

(22) Kaplan, J.; Verheijen, J. C.; Brooijmans, N.; Toral-Barza, L.; Hollander, I.; Yu, K.; Zask, A. Discovery of 3,6-Dihydro-2H-Pyran as a Morpholine Replacement in 6-Aryl-1H-Pyrazolo[3,4-D]Pyrimidines and 2-Arylthieno[3,2-d]Pyrimidines: ATP-Competitive Inhibitors of the Mammalian Target of Rapamycin (mTOR). *Bioorg. Med. Chem. Lett.* **2010**, *20*, 640−643.

(23) Richard, D. J.; Verheijen, J. C.; Curran, K.; Kaplan, J.; Toral-Barza, L.; Hollander, I.; Lucas, J.; Yu, K.; Zask, A. Incorporation of Water-Solubilizing Groups in Pyrazolopyrimidine mTOR Inhibitors: Discovery of Highly Potent and Selective Analogs with Improved Human Microsomal Stability. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6830−6835.

(24) Richard, D. J.; Verheijen, J. C.; Yu, K.; Zask, A. Triazines Incorporating (R)-3-Methylmorpholine Are Potent Inhibitors of the Mammalian Target of Rapamycin (mTOR) with Selectivity over PI3Kalpha. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2654−2657.

(25) Tsou, H. R.; MacEwan, G.; Birnberg, G.; Grosu, G.; Bursavich, M. G.; Bard, J.; Brooijmans, N.; Toral-Barza, L.; Hollander, I.; Mansour, T. S.; Ayral-Kaloustian, S.; Yu, K. Discovery and Optimization of 2-(4-Substituted-pyrrolo[2,3-b]Pyridin-3-yl)-Methylene-4-Hydroxybenzofuran-3(2H)-ones as Potent and Selective ATP-Competitive Inhibitors of the Mammalian Target of Rapamycin (mTOR). *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2321−2325.

(26) Tsou, H. R.; MacEwan, G.; Birnberg, G.; Zhang, N.; Brooijmans, N.; Toral-Barza, L.; Hollander, I.; Ayral-Kaloustian, S.; Yu, K. 4-Substituted-7-Azaindoles Bearing a Ureidobenzofuranone Moiety as Potent and Selective, ATP-Competitive Inhibitors of the Mammalian Target of Rapamycin (mTOR). *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2259−2263.

(27) Venkatesan, A. M.; Dehnhardt, C. M.; Chen, Z.; Santos, E. D.; Dos Santos, O.; Bursavich, M.; Gilbert, A. M.; Ellingboe, J. W.; Ayral-Kaloustian, S.; Khafizova, G.; Brooijmans, N.; Mallon, R.; Hollander, I.; Feldberg, L.; Lucas, J.; Yu, K.; Gibbons, J.; Abraham, R.; Mansour, T. S. Novel Imidazolopyrimidines as Dual PI3-Kinase/mTOR Inhibitors. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 653−656.

(28) Verheijen, J. C.; Richard, D. J.; Curran, K.; Kaplan, J.; Yu, K.; Zask, A. 2-Arylureidophenyl-4-(3-oxa-8-azabicyclo[3.2.1]Octan-8-yl)-Triazines as Highly Potent and Selective ATP Competitive mTOR Inhibitors: Optimization of Human Microsomal Stability. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2648−2653.

(29) Verheijen, J. C.; Yu, K.; Toral-Barza, L.; Hollander, I.; Zask, A. Discovery of 2-Arylthieno[3,2-d]Pyrimidines Containing 8-oxa-3-azabi-Cyclo[3.2.1]Octane in the 4-Position as Potent Inhibitors of mTOR with Selectivity over PI3K. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 375−379.

(30) Zhang, N.; Ayral-Kaloustian, S.; Anderson, J. T.; Nguyen, T.; Das, S.; Venkatesan, A. M.; Brooijmans, N.; Lucas, J.; Yu, K.; Hollander, I.; Mallon, R. 5-Ureidobenzofuranone Indoles as Potent and Efficacious Inhibitors of PI3 Kinase-Alpha and mTOR for the Treatment of Breast Cancer. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3526−3529.

(31) Chen, Z.; Venkatesan, A. M.; Dehnhardt, C. M.; Ayral-Kaloustian, S.; Brooijmans, N.; Mallon, R.; Feldberg, L.; Hollander, I.; Lucas, J.; Yu, K.; Kong, F.; Mansour, T. S. Synthesis and SAR of Novel 4-Morpholinopyrrolopyrimidine Derivatives as Potent Phosphatidylinositol 3-Kinase Inhibitors. *J. Med. Chem.* **2010**, *53*, 3169−3182.

(32) Nowak, P.; Cole, D. C.; Brooijmans, N.; Bursavich, M. G.; Curran, K. J.; Ellingboe, J. W.; Gibbons, J. J.; Hollander, I.; Hu, Y.; Kaplan, J.; Malwitz, D. J.; Toral-Barza, L.; Verheijen, J. C.; Zask, A.; Zhang, W. G.; Yu, K. Discovery of Potent and Selective Inhibitors of the Mammalian Target of Rapamycin (mTOR) Kinase. *J. Med. Chem.* **2009**, *52*, 7081−7089.

(33) Li, H.; Sutter, J.; Hoffmann, R. HypoGen: An Automated System for Generating 3D Predictive Pharmacophore Models. In *Pharmacophore Perception, Development, and Use in Drug Design*; Gűner, O. F., Ed.; International University Line, San Diego, CA, 2000; pp 171−190.

(34) Smellie, A.; Teig, S. L.; Towbin, P. Poling - Promoting Conformational Variation. *J. Comput. Chem.* **1995**, *16*, 171−187.

(35) Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of Conformational Coverage 0.2. Application of Conformational Models. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 295−304.

(36) Catalyst 4.11 User Guide, Accelrys Software Inc., San Diego, CA, 2005.

(37) Guner, O.; Clement, O.; Kurogi, Y. Pharmacophore Modeling and Three Dimensional Database Searching for Drug Design Using Catalyst: Recent Advances. *Curr. Med. Chem.* **2004**, *11*, 2991−3005.

(38) Discovery Studio 2.5.5 User Guide, Accelrys Inc.: San Diego, CA, 2010.

(39) 2010. Sutter, J.; Güner, O.; Hoffmann, R.; Li, H.; Waldman, M. Effect of Variable Weights and Tolerances on Predictive Model Generation. In *Pharmacophore Perception, Development, and Use in Drug Design*; Gűner, O. F., Ed.; International University Line, San Diego, CA, 2000; pp 499−512.

(40) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.

(41) Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics*; Wiley: New York, 1986.

(42) Zheng, W.; Tropsha, A. Novel Variable Selection Quantitative Structure–Property Relationship Approach Based on the *k*-Nearest-Neighbor Principle. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185−194.

(43) Zask, A.; Kaplan, J.; Verheijen, J. C.; Richard, D. J.; Curran, K.; Brooijmans, N.; Bennett, E. M.; Toral-Barza, L.; Hollander, I.; Ayral-Kaloustian, S.; Yu, K. Morpholine Derivatives Greatly Enhance the Selectivity of Mammalian Target of Rapamycin (mTOR) Inhibitors. *J. Med. Chem.* **2009**, *52*, 7942−7945.

(44) Cohen, F.; Bergeron, P.; Blackwood, E.; Bowman, K. K.; Chen, H.; Dipasquale, A. G.; Epler, J. A.; Koehler, M. F.; Lau, K.; Lewis, C.; Liu, L.; Ly, C. Q.; Malek, S.; Nonomiya, J.; Ortwine, D. F.; Pei, Z.; Robarge, K. D.; Sideris, S.; Trinh, L.; Truong, T.; Wu, J.; Zhao, X.; Lyssikatos, J. P. Potent, Selective, and Orally Bioavailable Inhibitors of Mammalian Target of Rapamycin (mTOR) Kinase Based on a Quaternary Substituted Dihydrofuropyrimidine. *J. Med. Chem.* **2011**, *54*, 3426−3435.

(45) Khanfar, M. A.; Abukhader, M. M.; Alqtaishat, S.; Taha, M. O. Pharmacophore Modeling, Homology Modeling, and in Silico Screening Reveal Mammalian Target of Rapamycin Inhibitory Activities for Sotalol, Glyburide, Metipranolol, Sulfamethizole, Glipizide, and Pioglitazone. *J. Mol. Graphics Modell.* **2013**, *42*, 39−49.

(46) Farrar, D. E.; Glauber, R. R. *Multicollinearity in Regression Analysis; the Problem Revisited*. BiblioBazaar: Charleston, SC, 2011.

(47) Clement, O. O.; Mehl, A. T. HipHop: Pharmacophores based on multiple common-feature alignment. In *Pharmacophore Perception, Development, and Use in Drug Design*; Gűner, O. F., Ed.; International University Line, San Diego, CA, 2000; pp 69−84.

(48) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3−26.

(49) Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615−2623.

(50) Cronin, M. T. D.; Schultz, T. W. Pitfalls in Qsar. *J. Mol. Struc. (Theochem)* **2003**, *622*, 39−51.

(51) Yuan, J.; Mehta, P. P.; Yin, M. J.; Sun, S. X.; Zou, A. H.; Chen, J.; Rafidi, K.; Feng, Z.; Nickel, J.; Engebretsen, J.; Hallin, J.; Blasina, A.; Zhang, E.; Nguyen, L.; Sun, M. H.; Vogt, P. K.; McHarg, A.; Cheng, H. M.; Christensen, J. G.; Kan, J. L. C.; Bagrodia, S. PF-04691502, a Potent and Selective Oral Inhibitor of PI3K and mTOR Kinases with Antitumor Activity. *Mol. Cancer Ther.* **2011**, *10*, 2189−2199.

(52) Shoichet, B. K. Interpreting Steep Dose-Response Curves in Early Inhibitor Discovery. *J. Med. Chem.* **2006**, *49*, 7274−7277.

(53) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.

(54) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the Performance of 3d Virtual Screening Protocols: Rmsd Comparisons, Enrichment Assessments, and Decoy Selection–What Can We Learn from Earlier Mistakes? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 213−228.

(55) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual Screening Using Protein-Ligand Docking: Avoiding Artificial Enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.

(56) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual Screening Workflow Development Guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *J. Med. Chem.* **2005**, *48*, 2534−2547.

(57) Irwin, J. J.; Shoichet, B. K. Zinc–a Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−182.