

Article

Exploring the Prominent and Concealed Inhibitory Features for Cytoplasmic Isoforms of Hsp90 Using QSAR Analysis

Magdi E. A. Zaki^{1,*}, Sami A. Al-Hussain¹, Syed Nasir Abbas Bukhari², Vijay H. Masand^{3,*}, Mithilesh M. Rathore³, Sumer D. Thakur⁴ and Vaishali M. Patil⁵

¹ Department of Chemistry, Faculty of Science, Imam Mohammad Ibn Saud Islamic University, Riyadh 13318, Saudi Arabia; sahussain@imamu.edu.sa

² Department of Pharmaceutical Chemistry, College of Pharmacy, Jouf University, Al Jouf 72388, Saudi Arabia; sbukhari@ju.edu.sa

³ Department of Chemistry, Vidya Bharati Mahavidyalaya, Amravati 444 602, Maharashtra, India; mithsrathore@gmail.com

⁴ Department of Chemistry, RDIK and NKD College, Badnera-Amravati 444 701, Maharashtra, India; sdthakur11@gmail.com

⁵ Department of Pharmaceutical Chemistry, KIET School of Pharmacy, KIET Group of Institutions, Delhi-NCR, Ghaziabad 201206, Uttar Pradesh, India; vaishuwise@gmail.com

* Correspondence: mezaki@imamu.edu.sa (M.E.A.Z.); vijaymasand@gmail.com (V.H.M.)



Citation: Zaki, M.E.A.; Al-Hussain, S.A.; Bukhari, S.N.A.; Masand, V.H.; Rathore, M.M.; Thakur, S.D.; Patil, V.M. Exploring the Prominent and Concealed Inhibitory Features for Cytoplasmic Isoforms of Hsp90 Using QSAR Analysis. *Pharmaceuticals* **2022**, *15*, 303. <https://doi.org/10.3390/ph15030303>

Academic Editor: Osvaldo Andrade Santos-Filho

Received: 21 January 2022

Accepted: 23 February 2022

Published: 1 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Cancer is a major life-threatening disease with a high mortality rate in many countries. Even though different therapies and options are available, patients generally prefer chemotherapy. However, serious side effects of anti-cancer drugs compel us to search for a safer drug. To achieve this target, Hsp90 (heat shock protein 90), which is responsible for stabilization of many oncoproteins in cancer cells, is a promising target for developing an anti-cancer drug. The QSAR (Quantitative Structure–Activity Relationship) could be useful to identify crucial pharmacophoric features to develop a Hsp90 inhibitor. Therefore, in the present work, a larger dataset encompassing 1141 diverse compounds was used to develop a multi-linear QSAR model with a balance of acceptable predictive ability (Predictive QSAR) and mechanistic interpretation (Mechanistic QSAR). The new developed six-parameter model satisfies the recommended values for a good number of validation parameters such as $R2_{tr} = 0.78$, $Q2_{LMO} = 0.77$, $R2_{ex} = 0.78$, and $CCC_{ex} = 0.88$. The present analysis reveals that the Hsp90 inhibitory activity is correlated with different types of nitrogen atoms and other hidden structural features such as the presence of hydrophobic ring/aromatic carbon atoms within a specific distance from the center of mass of the molecule, etc. Thus, the model successfully identified a variety of reported as well as novel pharmacophoric features. The results of QSAR analysis are further vindicated by reported crystal structures of compounds with Hsp90.

Keywords: Hsp90; cancer; QSAR; machine learning; pharmacophores

1. Introduction

Cancer kills; therefore, medicinal chemists are continuously trying to develop therapeutic agents that could retard the growth of cancer cells. In cancer cells, a protein Hsp90 (heat shock protein 90, also known as HSPC) is overexpressed [1]. It is a highly conserved, non-fibrous, and chaperone protein with a key role in many cellular processes like proper folding of other proteins, apoptosis, cell cycle control, cell viability, and degradation and signaling events [1–6]. As the name indicates, Hsp (heat shock proteins) shield cells when stressed by higher temperatures. The number “90” comes from the fact that it weighs about 90 kDa. There are two isoforms of Hsp90: Hsp90 α (the inducible form) and Hsp90 β (the constitutive form), which are found in cytoplasm and share 85% sequence identity [1–6]. These two isoforms are like flexible biological catalysts and interact with a good number of newly synthesized proteins, such as Akt2, CDKs, PKC, MAP kinases, steroid receptors,

BCL-6, CAR, p53, Oct4, etc., to avoid their aggregation or mistakes in their folding [6]. Despite a crucial role, in cancer cells, these are responsible for the stabilization of a number of oncoproteins required for tumor growth, leading to their overexpression [1–6]. Consequently, Hsp90 is an attractive target for developing a drug for cancer.

The majority of Hsp90 inhibitors occupy the ATP (adenosine tri-phosphate) pocket in the N-terminal domain of Hsp90, leading to limited ATPase activity [1–6]. At present, several natural and semi-synthetic Hsp90 inhibitors (see Figure 1) are in different stages of clinical trials for a variety of cancers [2,3,7–9].

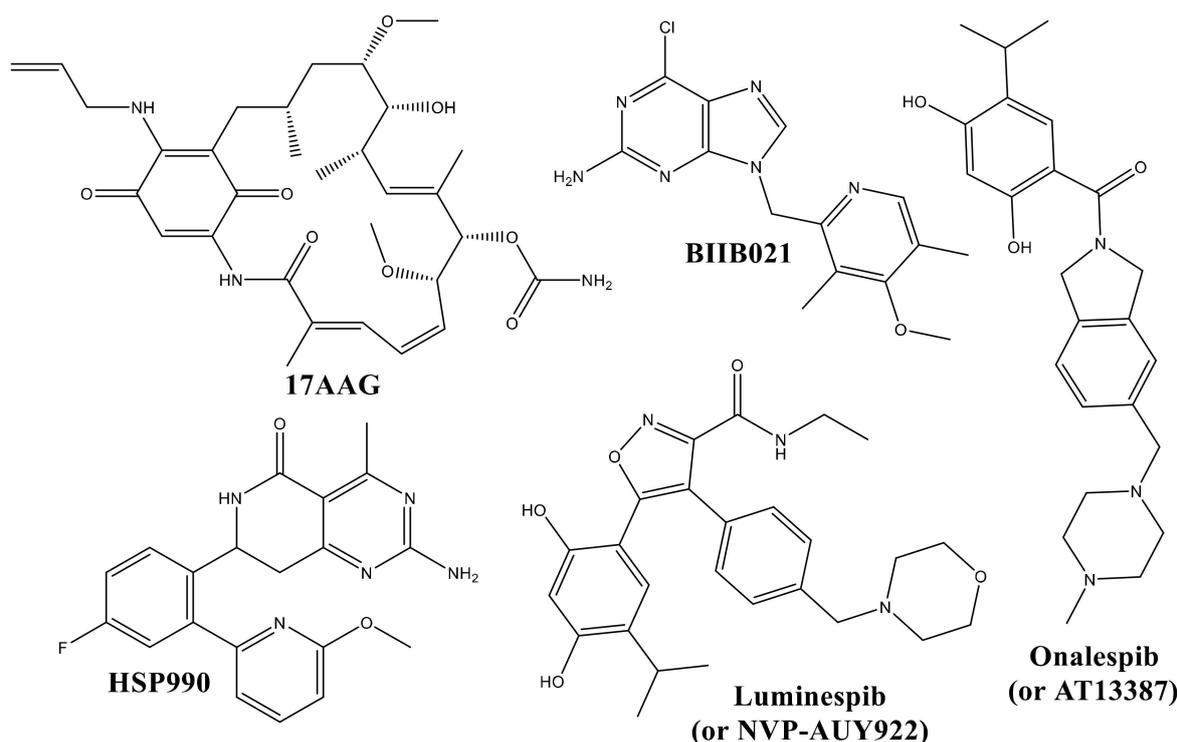


Figure 1. Different clinical trial candidates as inhibitors of Hsp90.

Unfortunately, several inhibitors have shown hepatotoxicity and ocular toxicity [2,10]; consequently, there is a need to modify them with retention of activity against Hsp90, which could be achieved on knowing the structural features responsible for their Hsp90 inhibitory activity. A simple, cost-effective, and faster yet effective strategy to know crucial pharmacophoric features is to use QSAR (Quantitative Structure–Activity Relationships), a successful, contemporary, and widely used branch of computer assisted-drug designing [11–16].

In QSAR analysis, generally, a good number of inhibitors are analyzed using a suitable technique like machine learning, deep learning, etc. There are two main advantages of using the QSAR approach [11,17,18]: (a) the analysis helps to identify the prominent structural features or patterns that influence the bioactivity profile of molecules (Mechanistic interpretation or Qualitative QSAR), and (b) the analysis could be used to predict the desired bioactivity of a molecule prior to its synthesis and lab testing (Predictive ability or Predictive QSAR). Therefore, many researchers prefer QSAR as a method of choice for drug/lead optimization. Nowadays, a QSAR analysis with a balance of mechanistic interpretation with predictive ability is highly preferred.

The literature survey reveals that QSAR analyses have been reported for Hsp90, but they are either based on a small dataset, lack general applicability, have poor predictive ability, are deficient of a mechanistic interpretation, or a combination of these factors, which limit their use [9,19–22]. Therefore, in the present work, we accomplished QSAR analysis for a larger and diverse dataset of Hsp90 inhibitors, and followed the OECD

(Organization for Economic Cooperation and Development) guidelines while developing a QSAR model to have a balance of mechanistic interpretation with predictive ability.

2. Results

The exhaustive and heuristic search resulted in the development of a six-descriptor-based QSAR model (see model-A), which was subjected to thorough statistical validation for internal and external validations.

Model-A: $\text{pIC}_{50} (M) = 3.903 (\pm 0.134) + 0.101 (\pm 0.013) \times \text{com_ringChyd_4A} + 0.433 (\pm 0.058) \times \text{faroCN2B} + 0.714 (\pm 0.214) \times \text{aroCminus_sumpc} + 0.065 (\pm 0.005) \times \text{aroC_aroN_5B} + 0.266 (\pm 0.048) \times \text{fringNsp3C5B} + 0.59 (\pm 0.082) \times \text{da_amdN_6B}$

Statistical validation of model-A:

$N_{\text{tr}} = 915$, $N_{\text{ext}} = 226$, $R^2_{\text{tr}} = 0.779$, $R^2_{\text{adj.}} = 0.777$, $R^2_{\text{tr}} - R^2_{\text{adj.}} = 0.002$, $\text{LOF} = 0.244$, $K_{\text{XX}} = 0.219$, $\Delta K = 0.122$, $\text{RMSE}_{\text{tr}} = 0.487$, $\text{MAE}_{\text{tr}} = 0.404$, $\text{RSS}_{\text{tr}} = 217.321$, $\text{CCC}_{\text{tr}} = 0.876$, $s = 0.489$, $F = 533.134$, $R^2_{\text{cv}} (Q^2_{\text{loo}}) = 0.775$, $R^2 - R^2_{\text{cv}} = 0.004$, $\text{RMSE}_{\text{cv}} = 0.491$, $\text{MAE}_{\text{cv}} = 0.407$, $\text{PRESS}_{\text{cv}} = 220.839$, $\text{CCC}_{\text{cv}} = 0.874$, $Q^2_{\text{LMO}} = 0.775$, $R^2_{\text{Yscr}} = 0.007$, $Q^2_{\text{Yscr}} = -0.009$, $\text{RMSE}_{\text{ex}} = 0.474$, $\text{MAE}_{\text{ex}} = 0.383$, $\text{PRESS}_{\text{ext}} = 50.675$, $R^2_{\text{ex}} = 0.779$, $Q^2 - F^1 = 0.778$, $Q^2 - F^2 = 0.778$, $Q^2 - F^3 = 0.791$, $\text{CCC}_{\text{ex}} = 0.876$, $R^2 - \text{ExPy} = 0.779$, $R'_o{}^2 = 0.727$, $k' = 0.989$, $1 - (R^2 / R'_o{}^2) = 0.066$, $r'^2 m = 0.602$, $R_o{}^2 = 0.779$, $k = 1.005$, $1 - (R^2 - \text{ExPy} / R_o{}^2) = 0$, $r^2 m = 0.766$

Different researchers have recommended the above statistical parameters to judge the robustness and external predictive ability of a QSAR model [11–16,23–31]. The formula to calculate them is available in the Supplementary Materials. It is clear that model-A fulfils the recommended threshold for many validation parameters and other criteria. A high value of different parameters like R^2_{tr} (coefficient of determination), $R^2_{\text{adj.}}$ (adjusted coefficient of determination), and R^2_{cv} (Q^2_{loo} , cross-validated coefficient of determination for leave-one-out), R^2_{ex} (external coefficient of determination), $Q^2 - F^n$, and CCC_{ex} (Concordance Correlation Coefficient), etc., and a low value of LOF (lack-of-fit), RMSE_{tr} (root mean square error), MAE_{tr} (mean absolute error), R^2_{Yscr} (R^2 for Y-scrambling), etc. along with the different graphs (see Figure 2) associated with the model indicate that the model possesses statistical robustness with excellent internal and external predictive ability as well as free from chance correlations. Additionally, the Williams plot specifies that the model is statistically acceptable (see Figure 2d). Therefore, it fulfils all the OECD recommended guidelines for creating a useful QSAR model.

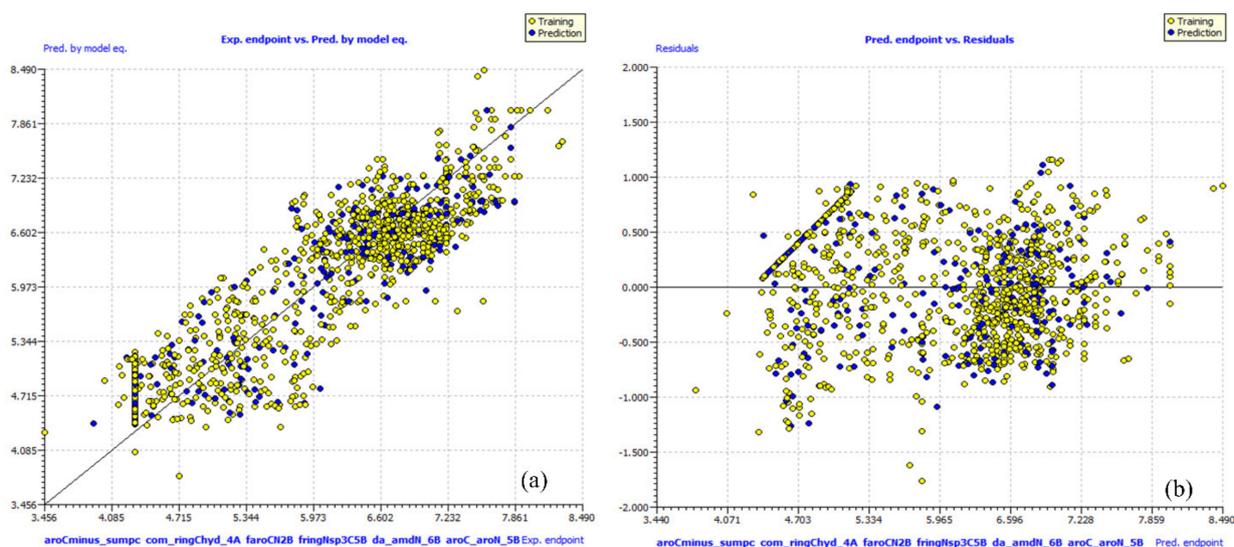


Figure 2. Cont.

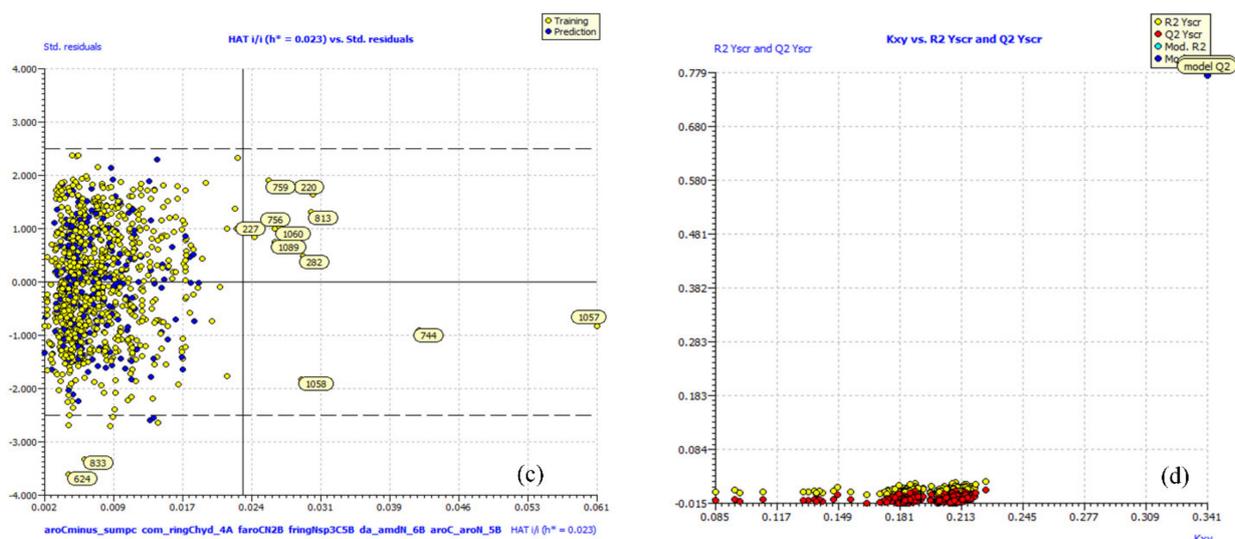


Figure 2. Different graphs associated with model-A: (a) experimental vs. predicted pIC_{50} (the solid line represents the regression line), (b) experimental vs. residuals, (c) Williams plot for applicability domain (the vertical solid line represents $h^* = 0.023$ and horizontal dashed lines represent the upper and lower boundaries for applicability domain), and (d) Y-randomization.

3. Discussion

Mechanistic Interpretation of QSAR Model

A very crucial aspect of a useful QSAR analysis is to gain deep insight into the pharmacophore or structure-oriented linking of molecular descriptors [17,32]. This not only helps throughout the drug discovery process, but also expands the information and understanding of mechanistic aspects of different types of molecules. Though, in the present work, a specific molecular descriptor was used to equate the pIC_{50} values of different molecules, but an extending or reverse influence of unknown factors or other molecular descriptors, having a dominant effect in deciding the final pIC_{50} value of a molecule, cannot be ignored. To simplify, a single molecular descriptor (in turn structure feature) cannot decide the overall experimental pIC_{50} value of a molecule. In other words, the effective use of an appropriately validated QSAR model depends on the synchronous consideration of all constituent molecular descriptors. Interestingly, in model-A, all the molecular descriptors have positive coefficients, which indicates that increasing their value could result in a better Hsp90 inhibitory activity.

The descriptor **com_ringChyd_4A** represents the total number of hydrophobic ring carbons, having partial charge in the range ± 0.2 , within 4\AA from the com (center of mass) of the molecule. From this, it appears that mere total number of ring carbons is very important, but replacing **com_ringChyd_4A** with **nringC** (number of ring carbon atoms) or **naroC** (number of aromatic carbon atoms) significantly reduced the statistical performance of the model ($R^2 = 0.72$). To add further, **com_ringChyd_4A** has a positive correlation of $R = 0.488$ with pIC_{50} , whereas **nringC** and **naroC** have a correlation of $R = 0.461$ and 0.405 , respectively. **com_ringChyd_3A** and **com_ringChyd_5A** represent the total number of ring carbons, having partial charge in the range ± 0.2 , within 3\AA and 5\AA from the com (center of mass) of the molecule, respectively. Replacement of **com_ringChyd_4A** with **com_ringChyd_3A** or **com_ringChyd_5A** resulted in slightly reduced performance of the model with $R^2 = 0.75$ and 0.76 , respectively. This indicates that the optimum distance is 4\AA .

The importance of hydrophobic ring carbon atoms is supported by the X-ray-resolved structure of a good number of Hsp90 inhibitors because the active site of Hsp90 consists of lipophilic side chains of Leu48, Ile91, Val186, Leu315, Ile388, and Val391 [33,34], which favors the presence of hydrophobic moiety in the inhibitors. For example, a comparison of molecule 988 ($pIC_{50} = 6.009$, **com_ringChyd_4A** = 10) with 1007 ($pIC_{50} = 6.481$,

com_ringChyd_4A = 15) highlights the importance of **com_ringChyd_4A**. Another pair of molecules, viz. 794 and 814, also supports this observation. The molecular descriptor **com_ringChyd_4A** is depicted in Figure 3 for different molecules.

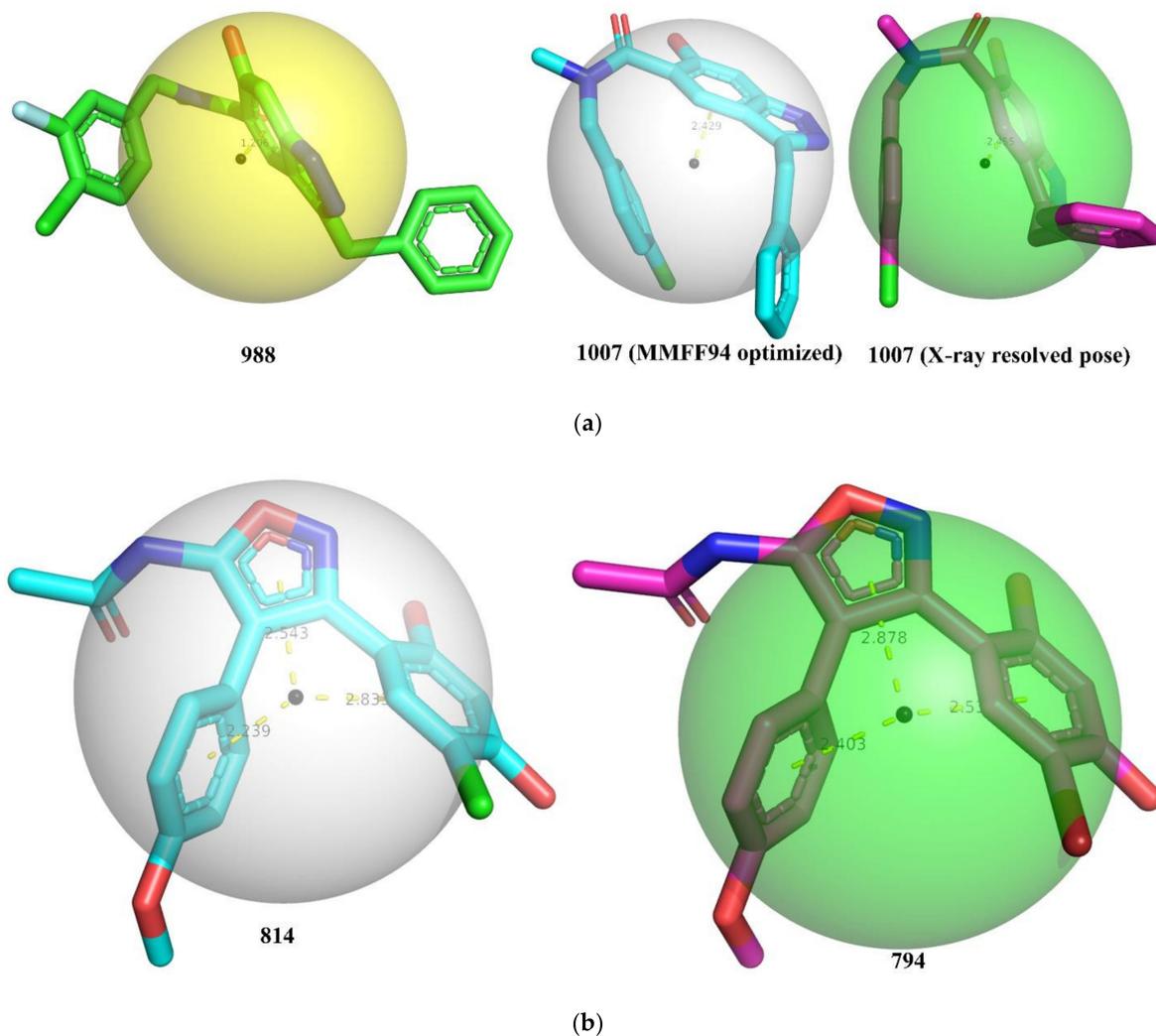


Figure 3. Depiction of **com_ringChyd_4A** using different molecules: (a) molecules 988, 1007 (MMFF94 optimized), and 1007 (X-ray resolved dock pose from pdb 6EY8); (b) molecules 794 and 814 (both X-ray-resolved poses from pdb 5XR9 and 4LWE, respectively). The small black sphere represents the com (center of mass) and the bigger transparent sphere represents the distance of 4Å from the center of mass. The dotted yellow line represents the distance (Å) of com from the centers of the different nearest rings.

From Figure 3, it is clear that the lowest energy conformer of molecule 988 has **com_ringChyd_4A** = 10 due to the closer presence of com (distance 1.206 Å) to the benzene ring of indazole ring. In case of molecule 1007 (MMFF94-optimized and X-ray-resolved pose from pdb 6EY8), the com is located slightly away from the benzene ring of Indazole ring at a distance > 2.40 due to specific conformation, thereby increasing the value of **com_ringChyd_4A** to 15. This could be a plausible reason for the difference in the bioactivity of these two compounds. Similarly, a better Hsp90 inhibitory activity of molecule 794 than 814 could be attributed to difference in their **com_ringChyd_4A** values.

Another molecular descriptor that has a positive effect on Hsp90 activity is **faroCN2B**, which signifies the presence of nitrogen exactly at two bonds from aromatic carbon atoms. If the same nitrogen atom is also present at two or less bonds from any other aromatic carbon atom, then it was excluded while calculating **faroCN2B**. This descriptor highlights the

importance of nitrogen atoms separated from aromatic ring (Benzene, etc.) by two bonds. As the majority of nitrogen atoms act as either an H-bond donor or acceptor; therefore, the presence of nitrogen atoms in the vicinity of aromatic rings could be useful in enhancing interactions with the polar residues of receptor (Hsp90). Additionally, the descriptor further points out the crucial role played by the aromatic rings undoubtedly due to their lipophilic nature. Taken together, the descriptor **faroCN2B** signifies the importance of two important structural features: aromatic rings and their vicinal nitrogen atoms.

This observation is confirmed when we compare the X-ray-resolved structures of molecule 727 ($pIC_{50} = 6.654$, **faroCN2B** = 1, pdb = 4O09) with 725 ($pIC_{50} = 7.137$, **faroCN2B** = 2, pdb = 4O05) depicted in Figure 4. The nitrogen atoms responsible for **faroCN2B** are highlighted by blue dotted circles. From Figure 4, it is clear that the aromatic ring B of both the molecules is responsible for hydrophobic interactions with the residue Met98. The nitrogen atom of ring A present in both the molecules is not only a constituent of **faroCN2B**, but also responsible for H-bonding with the residue Asp93. Thus, such a combination of aromatic carbons and nitrogen is highly beneficial to enhance the interactions with the receptor. In case of molecule 733, an additional nitrogen atom is present in ring E, which is a constituent of **faroCN2B**, and responsible for the H-bond interaction with the nearby water molecule. Thus, the present QSAR analysis revealed an important structural feature, which is also visible in X-ray-resolved structures of the same inhibitors with the same target enzyme Hsp90.

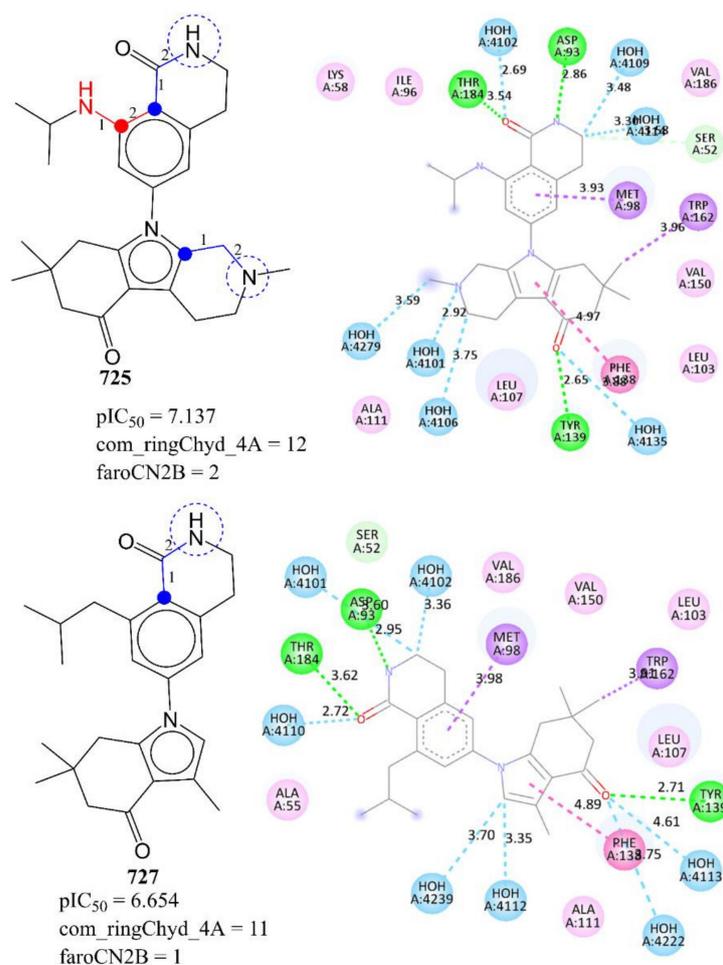


Figure 4. Depiction of **faroCN2B** using representative examples only.

A comparison of the following pairs of molecules further vindicates the importance of **faroCN2B** in determining the bioactivity: 213 ($pIC_{50} = 6.523$, **faroCN2B** = 2) with

212 ($pIC_{50} = 6.469$, **faroCN2B** = 1) and 758 ($pIC_{50} = 7.444$, **faroCN2B** = 2) with 759 ($pIC_{50} = 7.569$, **faroCN2B** = 3).

The importance of aromatic carbon atoms is further emphasized with the presence of **aroCminus_sumpc** as a constituent variable of model-A. The molecular descriptor **aroCminus_sumpc** represents the sum of partial charges on negatively charged aromatic carbon atoms. The positive coefficient for **aroCminus_sumpc** indicates that the higher the value of this descriptor, the better the activity profile. The sum of partial charges on negatively charged aromatic carbon atoms will always be negative; therefore, in reality, this descriptor actually decreases the pIC_{50} value. Further, the replacement of **aroCminus_sumpc** by **aroCplus_sumpc** (sum of partial charges on positively charged aromatic carbon atoms) led to a model with almost identical statistical performance ($R^2_{tr} = 0.772$, $Q^2_{LMO} = 0.767$, $R^2_{ex} = 0.78$, $CCC_{ex} = 0.876$). In fact, **aroCplus_sumpc** has a better correlation ($R = 0.33$) with pIC_{50} than **aroCminus_sumpc** ($R = 0.10$). From this it is clear that, if aromatic carbons are positively charged than the molecule possesses better Hsp90 inhibitory activity. Therefore, the best strategy is to attach atoms or groups that enhance lipophilic and mild polar interactions with the receptor (for example -Cl, etc.) to the aromatic carbon atoms. In short, substituted aromatic rings are preferable for better activity. This observation is supported by comparing following pairs of molecules: 2 with 3, 1054 with 1059, and 214 with 212.

aroC_aroN_5B, which represents the total number of aromatic carbon atoms within five bonds from aromatic nitrogen atoms, again points out the key role played by aromatic carbon atoms in deciding Hsp90 inhibitory activity. It also underlines the usefulness of aromatic nitrogen atoms. This descriptor has a positive correlation with pIC_{50} with $R = 0.63$. Therefore, an increase in number of aromatic carbon atoms within five bonds from aromatic nitrogen atoms leads to better Hsp90 inhibitory activity. The following pairs of the molecules support this observation: 888 ($pIC_{50} = 7.523$, **aroC_aroN_5B** = 22) with 887 ($pIC_{50} = 6.046$, **aroC_aroN_5B** = 20) and 107 ($pIC_{50} = 5.953$, **aroC_aroN_5B** = 13) with 108 ($pIC_{50} = 4.874$, **aroC_aroN_5B** = 10), to mention a few. Further, the 50 most active molecules possess relatively higher value of **aroC_aroN_5B** (range 8–17) than the 50 least active molecules (range 0–8).

fringNsp3C5B stands for the number of sp^3 -hybridized carbon atoms exactly at five bonds from the ring nitrogen atom. If the same sp^3 -hybridized carbon atom is also present at four or less bonds from any other ring nitrogen atom, then it was excluded while calculating **fringNsp3C5B**. It is interesting to note that the 50 most active molecules, except molecule 618, possess at least one or more of such a combination of carbon and ring nitrogen, whereas the 50 least active molecules either lack it or have **fringNsp3C5B** = 1. In the majority of compounds, the sp^3 -hybridized carbon atoms are present either as a linker between two rings or as a substituent, which therefore enhances conformational flexibility of the molecule to adopt a bioactive conformer or lipophilic characters of the molecule. A comparison of 895 ($pIC_{50} = 7.071$, **fringNsp3C5B** = 2) with 896 ($pIC_{50} = 6.777$, **fringNsp3C5B** = 1), 859 ($pIC_{50} = 7.237$, **fringNsp3C5B** = 2) with 896 ($pIC_{50} = 7.071$, **fringNsp3C5B** = 1), 326 ($pIC_{50} = 6.921$, **fringNsp3C5B** = 1) with 328 ($pIC_{50} = 7.046$, **fringNsp3C5B** = 2), and 412 ($pIC_{50} = 7.155$, **fringNsp3C5B** = 1) with 411 ($pIC_{50} = 6.959$, **fringNsp3C5B** = 0) and 410 ($pIC_{50} = 6.854$, **fringNsp3C5B** = 0) confirms the importance of **fringNsp3C5B** in deciding the activity.

A molecular descriptor that identifies the relation of total number amide nitrogen atoms within six bonds from the H-bond donor and acceptor atoms is **da_amdN_6B**. In the majority of compounds in the present dataset, the amide group is present as a substituent on aromatic ring or as a linker between two rings. The descriptor **da_amdN_6B** suggests the significance of amide group and its correlation with the H-bond donor and acceptor atoms. This observation is confirmed on comparing molecule **A** with molecules **B** and **C** (see Figure 5).

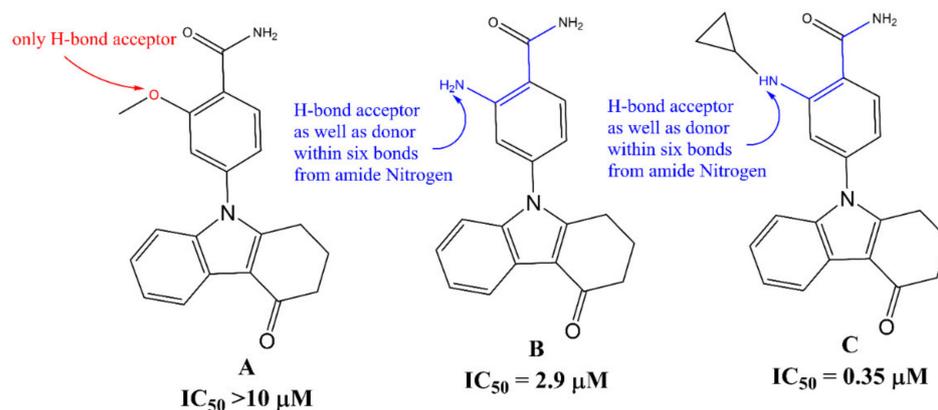


Figure 5. Pictorial representation of **da_amdN_6B** using representative examples only.

A good number of researchers have also pointed out that the amide group is crucial for Hsp90 inhibitors to establish H-bonding with residues of the active site (see pdb 4AWO). For example, Zhao et al. [4] pointed out that the distance between the nitrogen atoms on the piperidine ring and the amide are important for Hsp90 inhibition. Similarly, Baruchello and co-workers [35] studied a library of 3,4-isoxazole diamides for Hsp90 binding and found that a substantial reduction in Hsp90 binding affinity when the amide was replaced with substituted amines. In addition, a H-bond donor at the C-4 position on the isoxazole is vital for retaining the activity. Davies et al. [36] observed that S-acetamide derivatives of compounds have better bioactivity profile than the S-alkylamines. The importance of **da_amdN_6B** was further confirmed by comparing following pair of the molecules: 856 ($pIC_{50} = 6.848$, **da_amdN_6B** = 0) with 861 ($pIC_{50} = 7.114$, **da_amdN_6B** = 1). The earlier work identified the role of amide group, and in the present work, we successfully identified that a combination of amide group with H-bond donor/acceptor within six bonds is a better strategy to have better Hsp90 inhibitory activity. Therefore, such a combination of the amide nitrogen atom and H-bond donor/acceptor should be retained in future optimizations.

In short, three molecular descriptors emphasize the importance of ring carbon atoms, especially aromatic carbon atoms. This could be attributed to the lipophilic character of the active site of Hsp90. Likewise, four molecular descriptors underline the significance of different types of nitrogen atoms, which are responsible for the establishment of the polar or H-bond interactions with polar residues and water molecules present inside the active site of Hsp90. Hence, the present work is successful in identifying reported as well as novel pharmacophoric features of Hsp90 inhibitors.

4. Materials and Methods

The OECD (Organization for Economic Cooperation and Development) guidelines and a standard protocol recommended by different researchers [11–13,16,18,25,26,29,30,37] involve the sequential execution of (1) data collection and its curation, (2) structure generation and calculation of molecular descriptors, (3) objective feature selection (OFS), (4) splitting the dataset into training and external validation sets, (5) subjective feature selection involving building a regression model and validation of the developed model, which have all been followed to build a widely applicable QSAR model for Hsp-90 inhibitory activity. This also ensures thorough validation and successful application of the model.

4.1. Data Collection and Its Curation

The dataset of Hsp-90 inhibitory activity used for building, training, and validating the QSAR model in the present work was downloaded from BindingDB (<https://www.bindingdb.org/bind/index.jsp>, accessed on 24 December 2021), which is a free and publicly accessible database. Initially, the dataset comprised 1839 molecules. Then, as a part of data curation, entries with ambiguous IC_{50} values, duplicates, salts, metal-based inhibitors, etc.

were omitted [11–13,16,18,25,26,29,30,37]. The final dataset comprises 1141 structurally diverse molecules with remarkable variation in structural scaffolds, which were tested experimentally for potency in terms of IC_{50} (nM) (see the MS Excel file ‘SupplementaryMaterial-Final’ in the Supplementary Materials). The dataset includes N-terminal inhibitors of Hsp90. The experimental IC_{50} values have a sufficient variation ranging from 5 to 350,000 nM. After that, IC_{50} values were converted to their negative logarithmic value ($pIC_{50} = -\log_{10}IC_{50}$) so that a comparison of their values became easier. In Table 1 and Figure 6, some of the most and least active molecules are included as examples only.

Table 1. SMILES notation, IC_{50} (nM) and pIC_{50} (M) of the five most and least active molecules of the selected dataset.

S.N.	Ligand SMILES	IC_{50} (nM)	pIC_{50} (M)
308	<chem>COC1cccc(n1)-c1cc(F)ccc1[C@H]1Cc2nc(N)nc(C)c2C(NOC2C[C@H](O)[C@H](O)C2)=N1</chem>	5	8.301
908	<chem>CCNC(=O)c1noc(c1NC(=O)[C@H]1CC[C@H](CNS(=O)(=O)c2cc(F)cc2)CC1)-c1cc(C(C)C)c(O)cc1O</chem>	5.4	8.268
770	<chem>CCNC(=O)c1nnn(c1-c1ccc(CNC2CCCC2)cc1)-c1cc(C(C)C)c(O)cc1O</chem>	6.8	8.167
767	<chem>CCNC(=O)c1nnn(c1-c1ccc(CN2CCCC2CCO)cc1)-c1cc(C(C)C)c(O)cc1O</chem>	10	8
749	<chem>CCNC(=O)c1nnn(c1-c1ccc(CNCCCN(CC)CC)cc1)-c1cc(C(C)C)c(O)cc1O</chem>	12	7.921
775	<chem>Oc1cc(O)c2C[C@@H](OC(=O)[C@H]3CC[C@H](F)CC3)[C@H](Oc2c1)c1ccc(O)c(O)c1</chem>	69,000	4.161
1073	<chem>COC(COCCOc1ccc(Br)cc1)CN1CCN(CC1)c1cccc1C(C)(C)C</chem>	70,430	4.152
1141	<chem>CO[C@H]1C[C@H](C)C2c(OC)c(O)cc3NC(=O)\C(C)=C\C[C@H](O)C[C@H](OC)[C@H](OC(N=O)\C(C)=C\C[C@H](C)[C@H]1O)c23</chem>	96,000	4.018
778	<chem>Oc1cc(O)c2C[C@H](OC(=O)c3cccc3)[C@H](Oc2c1)c1cccc1</chem>	120,000	3.921
207	<chem>CSc1nc(C)nc(N)n1</chem>	350,000	3.456

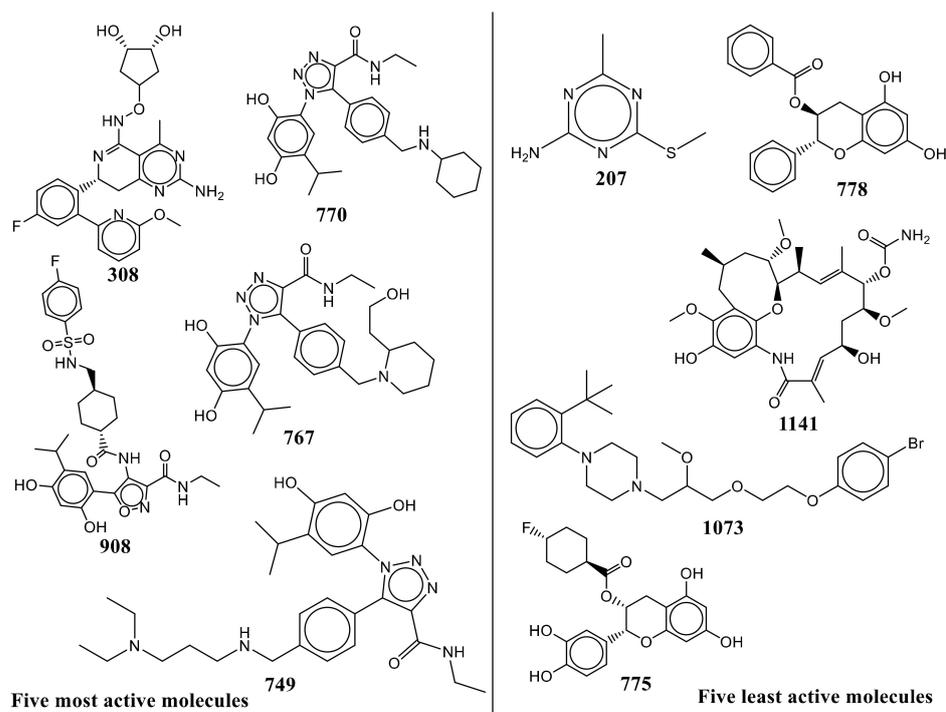


Figure 6. Representative examples from the selected dataset (the five most active and five least active molecules).

4.2. Calculation of Molecular Descriptors and Objective Feature Selection (OFS)

A crucial step before the calculation of molecular descriptors is to convert the SMILES notations to 3D-optimized structures and partial charge assignment, which was accomplished using OpenBabel 3.1 [38] using MMFF94 force field. In the present work, the X-ray-resolved structure of molecule **1007** (pdb 6YE8) was used to identify the parameter tuning in OpenBabel, required to get a better optimized structure, until there was a high similarity between the MMFF94-optimized structure and X-ray-resolved structure. This enhances the chances of getting a bioactive conformer, which in turn is highly beneficial for further optimization of Hsp90 inhibitors in the drug discovery pipeline. A comparison of the X-ray-resolved structures of molecules **1007** and **33** (pdb 2VCJ) and their respective MMFF94-optimized structures are represented in Figure 7.

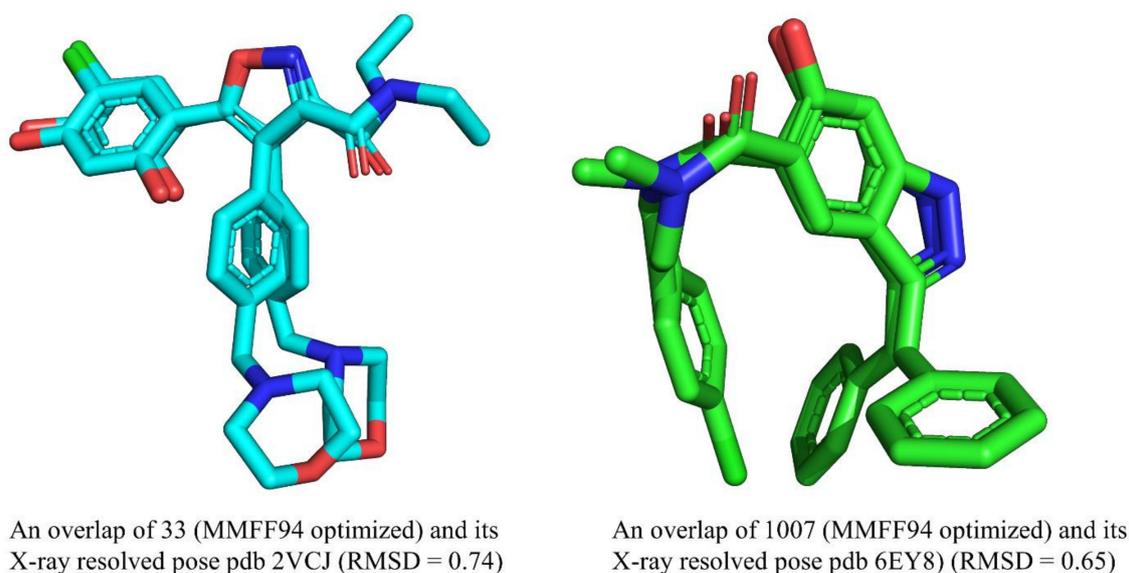


Figure 7. A comparison of X-ray-resolved and MMFF94-optimized structures of molecules 1007 and 33.

From Figure 7, it is clear that there is a high similarity between the X-ray-resolved and MMFF94-optimized structure of molecules **1007** and **33**, which indicates that appropriate parameter tuning was achieved to optimize the rest of the molecules. That is, the same parameter tuning in OpenBabel was used to optimize the other molecules of the selected dataset. The parameters are as follows: geometry optimization, steepest descent, number of steps: 1500; cut off: 0.01.

In the next step, the 3D-optimized structures of all molecules in the dataset were used to calculate a good number of molecular descriptors. It is important to note that calculation of diverse molecular descriptors enhances the chances of a successful QSAR analysis and significantly helps in mechanistic interpretation. However, descriptor pruning is very useful as it further strengthens the diminished risk of overfitting from noisy redundant descriptors. To fulfil these objectives, more than 40,000 molecular descriptors were generated using *PyDescriptor* [39]. After that, OFS involved elimination of the near constant (90% molecules) and highly intercorrelated ($|R| > 0.90$) molecular descriptors. For this, QSARINS-2.2.4 was used. The final set of molecular descriptors comprises 1228 molecular descriptors, which still comprise manifold descriptors (1D- to 3D-), leading to coverage of a broad descriptor space.

4.3. Splitting the Dataset into Training and External Sets and SFS (Subjective Feature Selection)

Subjective feature selection involves selection of appropriate number and set of molecular descriptors to build a model using suitable algorithm. Prior to SFS, it is essential to divide the dataset into training and test (also known as external or prediction set) sets with a proper composition and proportions to circumvent information leakage and to verify the

predictive ability of a model [11–13,16,18,25,26,29,30,37]. Hence, the dataset was randomly split into training (80% = 915 molecules) and prediction or external (20% = 226 molecules) sets. It is to be noted that the training set was used for the selection of optimum number of molecular descriptors, and the sole purpose of prediction/external set was to validate the external predictive ability of the model (Predictive QSAR). A GA-MLR-based QSAR model is free from over-fitting if it comprises an optimum number of molecular descriptors. Therefore, in the present work, a simple yet effective method of identifying the breaking point was used. Generally, the continuous inclusion of molecular descriptors in the GA-MLR model significantly increases the value of Q^2_{LOO} , but after the breaking point, the value of Q^2_{LOO} does not increase significantly [24]. The number of molecular descriptors corresponding to the breaking point was considered optimum for model building. A graph (see Figure 8) was plotted between the number of molecular descriptors involved in the model and Q^2_{LOO} values, which indicated that the breaking point agreed with the six molecular descriptors. Consequently, QSAR models comprising more than six descriptors were not considered. For SFS, the set of molecular descriptors was selected using the genetic algorithm integrated with multilinear regression (GA-MLR) method available in QSARINS-2.2.4 (generations per size: 10,000; population size: 50; mutation rate: 60; significance level: 0.05; fitness parameter: Q^2_{LOO}).

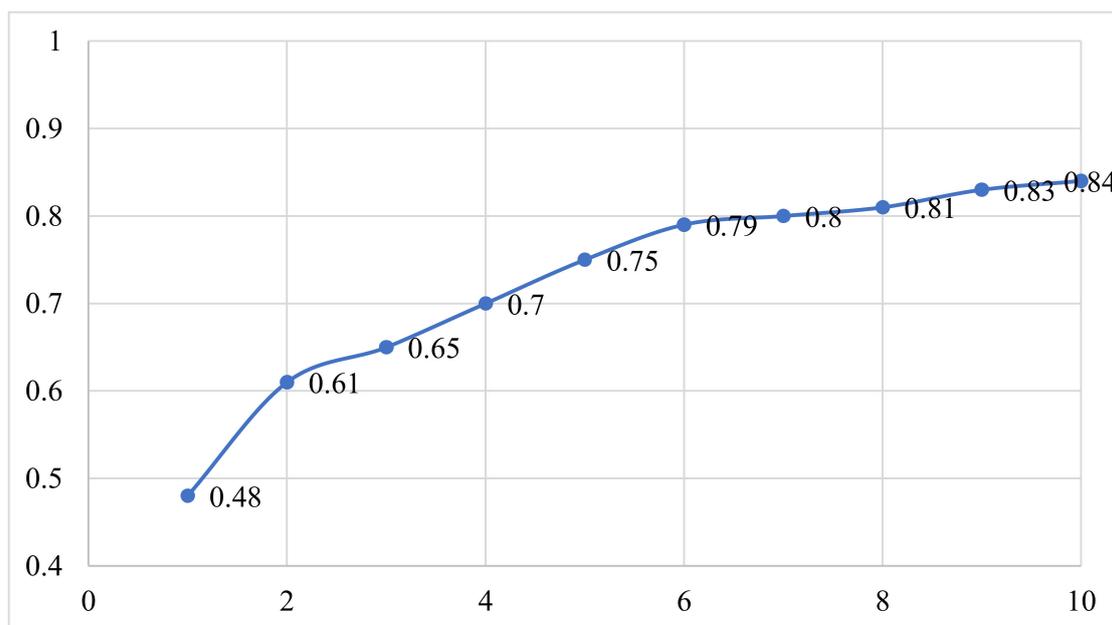


Figure 8. Plot of number of descriptors against leave-one-out coefficient of determination Q^2_{LOO} to identify the optimum number of descriptors.

4.4. Building Regression Model and Its Validation

The GA-MLR approach resulted in the generation of a good number of models having good to excellent statistical performance. Therefore, the following stringent parameters and criteria suggested by different researchers were used to select the best model [11–13,16,18,25,26,29,30,37,40]: $R^2_{tr} \geq 0.6$, $Q^2_{loo} \geq 0.5$, $Q^2_{LMO} \geq 0.6$, $R^2 > Q^2$, $R^2_{ex} \geq 0.6$, $RMSE_{tr} < RMSE_{cv}$, $\Delta K \geq 0.05$, $CCC \geq 0.80$, $Q^2-F^n \geq 0.60$, $r^2_m \geq 0.5$, $(1-r^2/r_o^2) < 0.1$, $0.9 \leq k \leq 1.1$ or $(1-r^2/r_o^2) < 0.1$, $0.9 \leq k' \leq 1.1$, $|r_o^2 - r_o'^2| < 0.3$, $RMSE_{ex}$, MAE_{ex} , R^2_{ex} , Q^2_{F1} , Q^2_{F2} , and Q^2_{F3} , and low R^2_{Yscr} , $RMSE$, and MAE . The details of these statistical parameters are available in the Supplementary Materials. An important aspect of validation of a QSAR model is to identify the applicability domain. In the present work, the William's plot was plotted to assess the applicability domain of the QSAR model [11–13,16,18,25,26,29,30,37,41,42].

5. Conclusions

In the present work, a relatively large and structurally diverse dataset of 1141 Hsp90 inhibitors was used for developing a six-descriptor-based and extensively validated GA–MLR QSAR model with $R^2_{tr} = 0.78$, $Q^2_{LMO} = 0.77$, $R^2_{ex} = 0.78$, and $CCC_{ex} = 0.88$. The inclusion of easily understandable descriptors resulted in identification of important pharmacophoric features that are correlated with Hsp90 inhibitory activity. The present QSAR analysis effectively captured a mixture of reported as well as novel significant structural features. The analysis vindicates that ring and aromatic carbons are important in deciding the activity. In addition, different types of nitrogen atoms in correlation with different types of carbon atoms influence the Hsp90 inhibitory activity. A good balance of external predictive ability and mechanistic interpretations, which are further supported by the reported crystal structures of Hsp90 inhibitors, make the QSAR model useful for the future optimization of molecules in the pipeline as a better Hsp90 inhibitor.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ph15030303/s1>.

Author Contributions: Conceptualization, V.H.M., M.E.A.Z. and S.A.A.-H.; formal analysis and data curation, V.H.M. and V.M.P.; writing, M.E.A.Z., V.H.M., M.M.R., S.N.A.B. and S.D.T.; Revisions, M.E.A.Z., V.H.M., S.D.T., M.M.R. and S.N.A.B.; editing and proofreading, V.H.M., M.E.A.Z. and S.N.A.B. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia, for its support of this research through research group number RG-21-09-76.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article and Supplementary Materials.

Acknowledgments: The authors acknowledge the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia, for its support of this research through research group number RG-21-09-76. V. H. Masand is thankful to Paola Gramatica (Italy) and her team for providing the free copy of QSARINS 2.2.4.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

SMILES	Simplified molecular-input line-entry system
GA	Genetic algorithm
MLR	Multiple linear regression
QSAR	Quantitative structure–activity relationship
WHO	World Health Organization
ADMET	Absorption, distribution, metabolism, excretion, and toxicity
OLS	Ordinary least square
QSARINS	QSAR Insubria
OECD	Organization for Economic Cooperation and Development

References

1. Ho, N.; Li, A.; Li, S.; Zhang, H. Heat Shock Protein 90 and Role of Its Chemical Inhibitors in Treatment of Hematologic Malignancies. *Pharmaceuticals* **2012**, *5*, 779–801. [[CrossRef](#)]
2. Li, L.; Wang, L.; You, Q.-D.; Xu, X.-L. Heat Shock Protein 90 Inhibitors: An Update on Achievements, Challenges, and Future Directions. *J. Med. Chem.* **2019**, *63*, 1798–1822. [[CrossRef](#)]
3. Bhat, R.; Tummalapalli, S.R.; Rotella, D.P. Progress in the Discovery and Development of Heat Shock Protein 90 (Hsp90) Inhibitors. *J. Med. Chem.* **2014**, *57*, 8718–8728. [[CrossRef](#)] [[PubMed](#)]

4. Zhao, H.; Moroni, E.; Colombo, G.; Blagg, B.S.J. Identification of a New Scaffold for Hsp90 C-Terminal Inhibition. *ACS Med. Chem. Lett.* **2013**, *5*, 84–88. [[CrossRef](#)]
5. Li, Y.; Zhang, T.; Schwartz, S.J.; Sun, D. New developments in Hsp90 inhibitors as anti-cancer therapeutics: Mechanisms, clinical perspective and more potential. *Drug Resist. Updates* **2009**, *12*, 17–27. [[CrossRef](#)]
6. Hoter, A.; El-Sabban, M.; Naim, H. The HSP90 Family: Structure, Regulation, Function, and Implications in Health and Disease. *Int. J. Mol. Sci.* **2018**, *19*, 2560. [[CrossRef](#)]
7. Zuehlke, A.D.; Moses, M.A.; Neckers, L. Heat shock protein 90: Its inhibition and function. *Philos. Trans. R. Soc. B Biol. Sci.* **2017**, *373*, 20160527. [[CrossRef](#)]
8. Biamonte, M.A.; Van de Water, R.; Arndt, J.W.; Scannevin, R.H.; Perret, D.; Lee, W.-C. Heat Shock Protein 90: Inhibitors in Clinical Trials. *J. Med. Chem.* **2009**, *53*, 3–17. [[CrossRef](#)] [[PubMed](#)]
9. Patil, V.M.; Masand, N.; Gupta, S.P.; Blagg, B.S.J. QSAR Studies to Predict Activity of HSP90 Inhibitors. *Curr. Top. Med. Chem.* **2021**, *21*, 2272–2291. [[CrossRef](#)] [[PubMed](#)]
10. Jhaveri, K.; Taldone, T.; Modi, S.; Chiosis, G. Advances in the clinical development of heat shock protein 90 (Hsp90) inhibitors in cancers. *Biochim. Biophys. Acta BBA Mol. Cell Res.* **2012**, *1823*, 742–755. [[CrossRef](#)] [[PubMed](#)]
11. Gramatica, P. Principles of QSAR Modeling. *Int. J. Quant. Struct.-Prop. Relatsh.* **2020**, *5*, 61–97. [[CrossRef](#)]
12. Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R.; et al. QSAR modeling: Where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57*, 4977–5010. [[CrossRef](#)]
13. Gramatica, P. On the development and validation of QSAR models. *Methods Mol. Biol.* **2013**, *930*, 499–526. [[CrossRef](#)]
14. Gramatica, P.; Cassani, S.; Roy, P.P.; Kovarich, S.; Yap, C.W.; Papa, E. QSAR Modeling is not Push a Button and Find a Correlation: A Case Study of Toxicity of (Benzo-)triazoles on Algae. *Mol. Inform.* **2012**, *31*, 817–835. [[CrossRef](#)]
15. Li, J.; Gramatica, P. The importance of molecular structures, endpoints' values, and predictivity parameters in QSAR research: QSAR analysis of a series of estrogen receptor binders. *Mol. Divers.* **2010**, *14*, 687–696. [[CrossRef](#)]
16. Muratov, E.N.; Bajorath, J.; Sheridan, R.P.; Tetko, I.V.; Filimonov, D.; Poroikov, V.; Oprea, T.I.; Baskin, I.I.; Varnek, A.; Roitberg, A.; et al. QSAR without borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564. [[CrossRef](#)]
17. Fujita, T.; Winkler, D.A. Understanding the Roles of the “Two QSARs”. *J. Chem. Inf. Model.* **2016**, *56*, 269–274. [[CrossRef](#)]
18. Zaki, M.E.A.; Al-Hussain, S.A.; Masand, V.H.; Sabnani, M.K.; Samad, A. Mechanistic and Predictive QSAR Analysis of Diverse Molecules to Capture Salient and Hidden Pharmacophores for Anti-Thrombotic Activity. *Int. J. Mol. Sci.* **2021**, *22*, 8352. [[CrossRef](#)]
19. Zhao, H.; Moroni, E.; Yan, B.; Colombo, G.; Blagg, B.S.J. 3D-QSAR-Assisted Design, Synthesis, and Evaluation of Novobiocin Analogues. *ACS Med. Chem. Lett.* **2012**, *4*, 57–62. [[CrossRef](#)]
20. Barta, T.E.; Veal, J.M.; Rice, J.W.; Partridge, J.M.; Fadden, R.P.; Ma, W.; Jenks, M.; Geng, L.; Hanson, G.J.; Huang, K.H.; et al. Discovery of benzamide tetrahydro-4H-carbazol-4-ones as novel small molecule inhibitors of Hsp90. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3517–3521. [[CrossRef](#)]
21. Bussenius, J.; Blazey, C.M.; Aay, N.; Anand, N.K.; Arcalás, A.; Baik, T.; Bowles, O.J.; Buhr, C.A.; Costanzo, S.; Curtis, J.K.; et al. Discovery of XL888: A novel tropane-derived small molecule inhibitor of HSP90. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 5396–5404. [[CrossRef](#)]
22. Abbasi, M.; Sadeghi-Aliabadi, H.; Amanlou, M. Prediction of new Hsp90 inhibitors based on 3,4-isoxazolidinamide scaffold using QSAR study, molecular docking and molecular dynamic simulation. *DARU J. Pharm. Sci.* **2017**, *25*, 17. [[CrossRef](#)]
23. Gramatica, P. External Evaluation of QSAR Models, in Addition to Cross-Validation Verification of Predictive Capability on Totally New Chemicals. *Mol. Inform.* **2014**, *33*, 311–314. [[CrossRef](#)]
24. Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem.* **2013**, *34*, 2121–2132. [[CrossRef](#)]
25. Chirico, N.; Gramatica, P. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *J. Chem. Inf. Model.* **2012**, *52*, 2044–2058. [[CrossRef](#)]
26. Chirico, N.; Gramatica, P. Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J. Chem. Inf. Model.* **2011**, *51*, 2320–2335. [[CrossRef](#)]
27. Gramatica, P.; Pilutti, P.; Papa, E. Approaches for externally validated QSAR modelling of Nitrated Polycyclic Aromatic Hydrocarbon mutagenicity. *SAR QSAR Environ. Res.* **2007**, *18*, 169–178. [[CrossRef](#)]
28. Gramatica, P. Principles of QSAR models validation internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694–701. [[CrossRef](#)]
29. Martin, T.M.; Harten, P.; Young, D.M.; Muratov, E.N.; Golbraikh, A.; Zhu, H.; Tropsha, A. Does rational selection of training and test sets improve the outcome of QSAR modeling? *J. Chem. Inf. Model.* **2012**, *52*, 2570–2578. [[CrossRef](#)]
30. Tropsha, A.; Gramatica, P.; Gombar, V.K. The Importance of Being Earnest Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003**, *22*, 69–77. [[CrossRef](#)]
31. Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.D.; Lee, K.H.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 241–253. [[CrossRef](#)]
32. Polishchuk, P. Interpretation of Quantitative Structure–Activity Relationship Models: Past, Present, and Future. *J. Chem. Inf. Model.* **2017**, *57*, 2618–2639. [[CrossRef](#)]
33. Jackson, S.E. Hsp90: Structure and Function. In *Molecular Chaperones*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 155–240.

34. Vallée, F.; Carrez, C.; Pilorge, F.; Dupuy, A.; Parent, A.; Bertin, L.; Thompson, F.; Ferrari, P.; Fassy, F.; Lambertson, A.; et al. Tricyclic Series of Heat Shock Protein 90 (Hsp90) Inhibitors Part I: Discovery of Tricyclic Imidazo[4,5-c]pyridines as Potent Inhibitors of the Hsp90 Molecular Chaperone. *J. Med. Chem.* **2011**, *54*, 7206–7219. [[CrossRef](#)]
35. Baruchello, R.; Simoni, D.; Grisolia, G.; Barbato, G.; Marchetti, P.; Rondanin, R.; Mangiola, S.; Giannini, G.; Brunetti, T.; Alloatti, D.; et al. Novel 3,4-Isoxazolidiamides as Potent Inhibitors of Chaperone Heat Shock Protein 90. *J. Med. Chem.* **2011**, *54*, 8592–8604. [[CrossRef](#)]
36. Davies, N.G.M.; Browne, H.; Davis, B.; Drysdale, M.J.; Foloppe, N.; Geoffrey, S.; Gibbons, B.; Hart, T.; Hubbard, R.; Jensen, M.R.; et al. Targeting conserved water molecules: Design of 4-aryl-5-cyanopyrrolo[2,3-d]pyrimidine Hsp90 inhibitors using fragment-based screening and structure-based optimization. *Bioorg. Med. Chem.* **2012**, *20*, 6770–6789. [[CrossRef](#)]
37. Fourches, D.; Muratov, E.; Tropsha, A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J. Chem. Inf. Model.* **2010**, *50*, 1189–1204. [[CrossRef](#)]
38. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [[CrossRef](#)]
39. Masand, V.H.; Rastija, V. PyDescriptor: A new PyMOL plugin for calculating thousands of easily understandable molecular descriptors. *Chemom. Intell. Lab. Syst.* **2017**, *169*, 12–18. [[CrossRef](#)]
40. Zaki, M.E.A.; Al-Hussain, S.A.; Masand, V.H.; Akasapu, S.; Bajaj, S.O.; El-Sayed, N.N.E.; Ghosh, A.; Lewaa, I. Identification of Anti-SARS-CoV-2 Compounds from Food Using QSAR-Based Virtual Screening, Molecular Docking, and Molecular Dynamics Simulation Analysis. *Pharmaceuticals* **2021**, *14*, 357. [[CrossRef](#)]
41. Kar, S.; Roy, K.; Leszczynski, J. Applicability Domain: A Step Toward Confident Predictions and Decidability for QSAR Modeling. In *Computational Toxicology*; Humana Press: New York, NY, USA, 2018; pp. 141–169.
42. Gramatica, P.; Kovarich, S.; Roy, P.P. Reply to the comment of S. Rayne on “QSAR model reproducibility and applicability: A case study of rate constants of hydroxyl radical reaction models applied to polybrominated diphenyl ethers and (benzo-)triazoles”. *J. Comput. Chem.* **2013**, *34*, 1796. [[CrossRef](#)]