

# 3D-QSAR Modeling on 2-Pyrimidine Carbohydrazides as Utrophin Modulators for the Treatment of Duchenne Muscular Dystrophy by Combining CoMFA, CoMSIA, and Molecular Docking Studies

Reza Mahmoudzadeh Laki and Eslam Pourbasheer\*



Cite This: *ACS Omega* 2024, 9, 24707–24720



Read Online

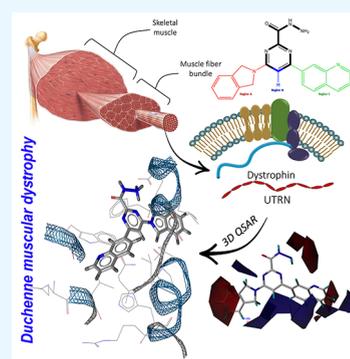
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** The 3D-QSAR models were developed using CoMFA and CoMSIA techniques to investigate essential molecular fields, optimization strategies, and structure–activity relationships for utrophin-modulating compounds. The data set (71 molecules) was divided into two training and test sets using the hierarchical clustering approach. The training set was aligned based on the most active compound. The built and optimized models based on the PLS approach provided acceptable results. The results were  $q^2 = 0.528$  and  $r^2 = 0.776$  for CoMFA and  $q^2 = 0.600$  and  $r^2 = 0.811$  for CoMSIA models. According to the statistical results, it was found that both the CoMFA models with and without regional focusing and also the CoMSIA model have good estimation ability. Molecular docking was also performed with high-activity compounds (as ligands) and target receptors (protein), and its results, together with the results of 3D-QSAR, give new insights for the design of compounds with higher biological activity. Finally, based on the overall results, the design of new compounds with higher utrophin modulation activity was carried out.



## 1. INTRODUCTION

Duchenne muscular dystrophy (DMD) is a debilitating condition associated with a progressive genetic disorder that leads to severe muscle weakness and wasting over time.<sup>1</sup> According to studies, the main cause of this disease is the mutation in the dystrophin gene, which disrupts the functional dystrophin protein. Since the dystrophin gene is known as a very large gene, it therefore exhibits an extraordinarily high rate of new mutations.<sup>2</sup> Dystrophin is the protein responsible for the aforementioned condition that stabilizes muscle fibers during movement and acts as a critical structural link between the actin cytoskeleton and the extracellular matrix.<sup>3</sup> Although significant advancements are being made in the development of promising approaches such as the removal of specific exons and their adjacent introns from mRNA structures before translation, stop-codon read-through, and gene modification, there is presently no universally effective treatment available for the entire target community. In addition, to solve the lack of functional dystrophin in these patients, oral small molecules have been made using utrophin.<sup>4</sup> The mentioned therapeutic approach has the potential to be beneficial for all sufferers, regardless of specific dystrophin mutations. Despite the progress made in targeting specific patient subsets with the aforementioned therapies, their limited efficacy and potential toxicity pose significant challenges.<sup>5</sup> Therefore, it is crucial to explore alternative therapeutic techniques that offer greater efficacy and are not limited to specific mutations. Among the studies that have been reported about solving these challenges, we can mention the study of Hadwen and his colleagues<sup>6</sup> who,

by identifying the factors that cause the expression of utrophin, conducted an *in silico* study about reporting a method for potential treatment of DMD. Using systematic databases of human gene expression changes, they performed an Affymetrix array-based analysis on different cell lines using 1600 compounds and identified a list of small compounds that increase utrophin mRNA. Also, anisomycin, a p38-activating antibiotic, was one of the most promising utrophin mRNA regulators that they reported. Another study by Echigoya et al.<sup>7</sup> reported a potential therapeutic approach for DMD using antisense “modifying” oligonucleotides to induce exon skipping. They proposed an *in silico* prescreening approach based on predictive statistical modeling. Although these studies show that mRNA control and removal of specific exons or skipping exons can be an effective way to treat DMD, as said, universal effective treatment is not available for the entire target population, so studying a new method such as modulating the expression of utrophin can be promising.

Utrophin (UTRN) upregulation represents a mechanism-based approach that holds promise for treating all patients with DMD, regardless of their specific mutations. As mentioned, utrophin is an autosomal paralog of dystrophin protein and has

**Received:** February 7, 2024

**Revised:** April 29, 2024

**Accepted:** May 17, 2024

**Published:** May 31, 2024



Table 1. Statistical Results Obtained from CoMFA and CoMSIA Models

parameter	CoMFA-1	CoMFA-2 (focusing)	CoMSIA
PLS statistics			
LOO cross $q^2$ /SEP	0.528/0.371	0.553/0.361	0.600/0.341
group cross $q^2$ /SEP	0.555/0.357	0.522/0.373	0.586/0.347
nonvalidated $r^2$ /SEE	0.776/0.255	0.750/0.270	0.811/0.235
$F$	61.324	53.126	75.797
$r_{\text{bootstrap}}^2$	0.858 $\pm$ 0.019	0.826 $\pm$ 0.048	0.857 $\pm$ 0.049
$S_{\text{bootstrap}}$	0.209 $\pm$ 0.085	0.217 $\pm$ 0.091	0.190 $\pm$ 0.095
optimal compounds	3	3	3
$r_{\text{test}}^2$	0.674	0.695	0.529
CCC <sub>training</sub>	0.8741	0.8574	0.8956
CCC <sub>test</sub>	0.7567	0.7800	0.6589
field distribution			
steric	0.525	0.593	0.134
electrostatic	0.475	0.407	0.482
hydrophobic			0.269
H-bond acceptor			0.115

been identified as a potential operational substitute. Studies conducted on DMD animal models have demonstrated that utrophin can compensate for the absence of dystrophin. As a result, there is considerable interest within the scientific community to explore approaches that can modulate utrophin levels in dystrophic tissues. Such an approach has the potential to provide a comprehensive treatment option for individuals with DMD, but there are still obstacles that need to be overcome. A modulator must interact with the correct part of the target molecule to be effective, but designing a drug that can only interact with the desired region (without altering other important functions) seems difficult. This can be considered an important challenge for a complex molecule like utrophin, which is composed of different parts. On the other hand, since the modulation of utrophin expression is done in parts of the body where direct drug delivery is difficult (embryonic tissue, heart fibers, etc.), computer modeling studies can be of great help regarding these challenges. Examining drugs designed as modulators of utrophin expression to reduce unwanted side effects due to cellular processes with similar proteins seems to be another challenge that can be overcome based on studies of structure–activity relationships and the design of new compounds.<sup>4,8</sup> Recently, Vuorinen and co-workers have introduced a new series of utrophin modulators in a novel cellular screening course.<sup>9</sup> In their report, they investigated how utrophin is transcribed in firefly genes. According to the Vuorinen group study, Chatzopoulou and colleagues reported further optimization and structure–activity relationships (SAR) for their main compound (OX01914).<sup>10</sup> This approach could enhance the intrinsic potency of the compound and improve its overall physicochemical properties.

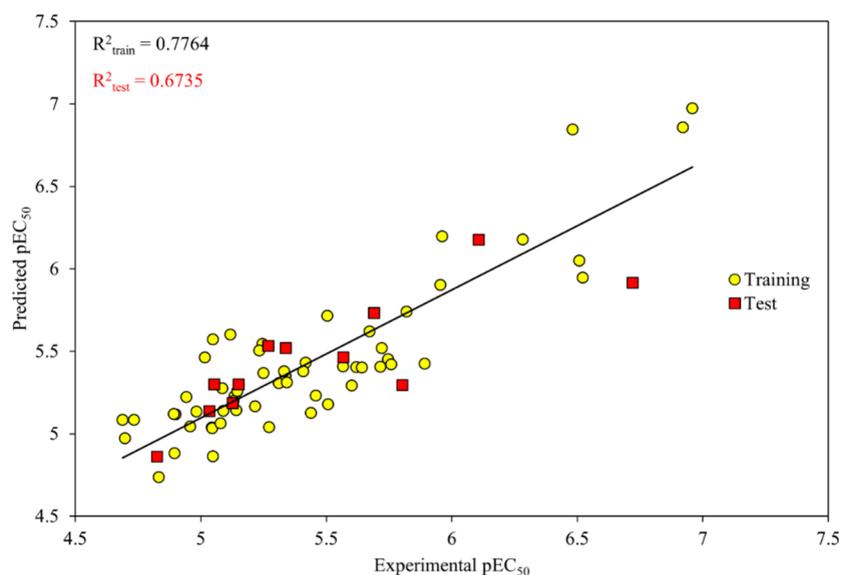
Quantitative structure–activity relationship (QSAR) modeling has gained considerable popularity as a suitable alternative to practical studies in predicting changes in the activity of molecules by changes in their structure.<sup>11–14</sup> The 3D-QSAR methods have emerged as a more sophisticated approach for modeling structure–activity relationships in drug discovery.<sup>15–17</sup> This method considers the complete 3D representation of molecules, taking into account their steric characteristics and the associated electrostatic effects that influence their biological activity.<sup>18–20</sup> Investigating changes in the biological activity of molecules by changing their structure, based on

quantitative relationships using computational algorithms and statistical techniques, is the goal of 3D-QSAR studies.<sup>21,22</sup> Comparative molecular field analysis (CoMFA)<sup>23</sup> and comparative molecular similarity indices analysis (CoMSIA)<sup>24</sup> are two prominent analysis techniques among the various methods used in 3D-QSAR. In the CoMFA technique, by preparing contour maps and examining the steric and electrostatic fields around a compound, it is possible to obtain the highest biological activity available for the molecule in a virtual Cartesian grid.<sup>23</sup> The number of interactions of these fields around the molecules can be seen in these contour maps. Contrarily, CoMSIA broadens the scope of the study by integrating new parameters such as hydrophobicity, hydrogen-bond donor, and hydrogen-bond acceptor characteristics.<sup>24</sup> By combining 3D-QSAR and molecular docking studies, a pervasive approach can be presented to detect the effects between different functional groups in the molecular structure and its interaction with the receptor in the presence of fields created around the molecule.<sup>25</sup>

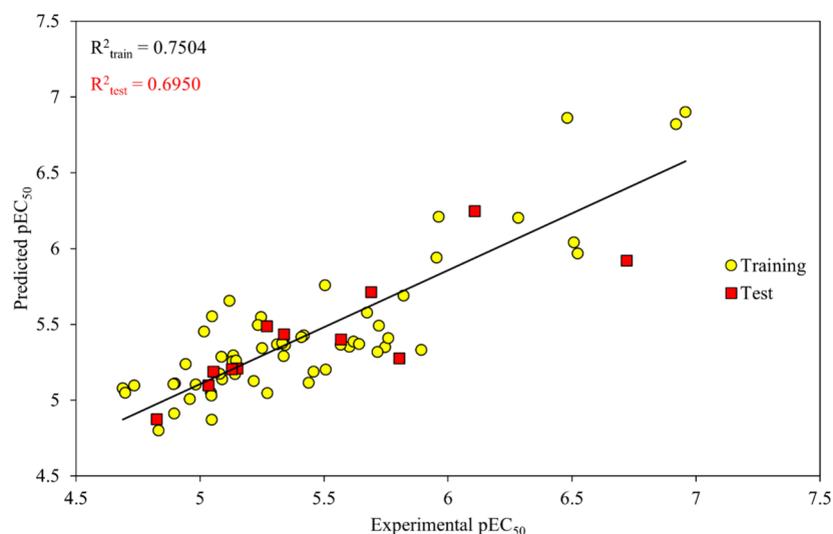
The present work aims to create a 3D-QSAR model utilizing CoMFA and CoMSIA techniques to investigate the essential molecular fields, optimization strategies, and design of new compounds with higher utrophin modulation activity. The molecular docking technique was also used to study the effective interaction between the modulator and its associated receptor.

## 2. RESULTS AND DISCUSSION

**2.1. CoMFA Analysis.** To study the relationship between chemical structures and the UTRN modulation activity of compounds, the CoMFA model named CoMFA-1 consisting of a training set with 57 compounds was used. Also, to establish a relationship between the structural features of the compounds and their UTRN modulation activity, the PLS approach was used to build the model. Many statistical parameters were selected and used to evaluate the estimation power of the model and its performance in accurately predicting modulation activity based on chemical structure descriptors. The results obtained from the CoMFA analysis are listed in Table 1. The most favorable result is when the number of components is three, according to which the highest value of  $q^2$  was obtained. Based on the results of the PLS technique in nonvalidation mode, the  $r^2$  is 0.776, the standard



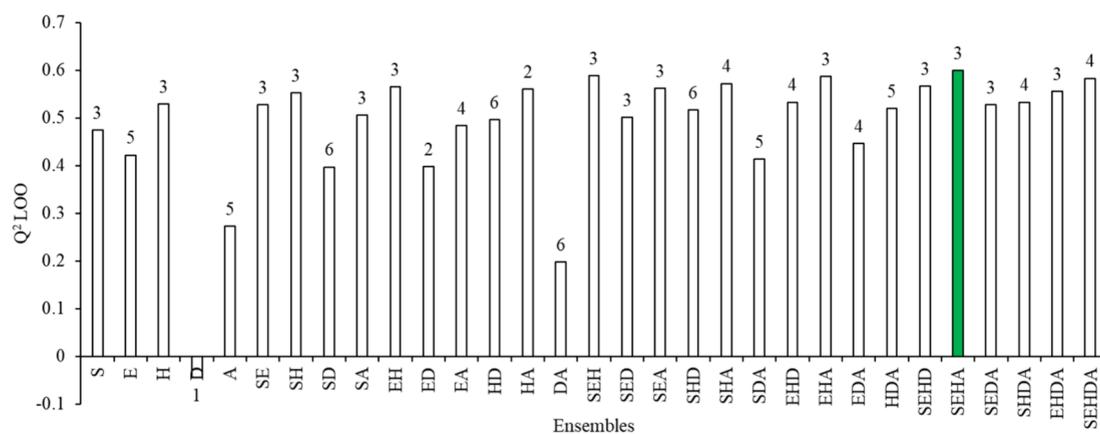
**Figure 1.** Graph obtained from the correlation of experimental values against predicted values of pEC<sub>50</sub> by the CoMFA model.



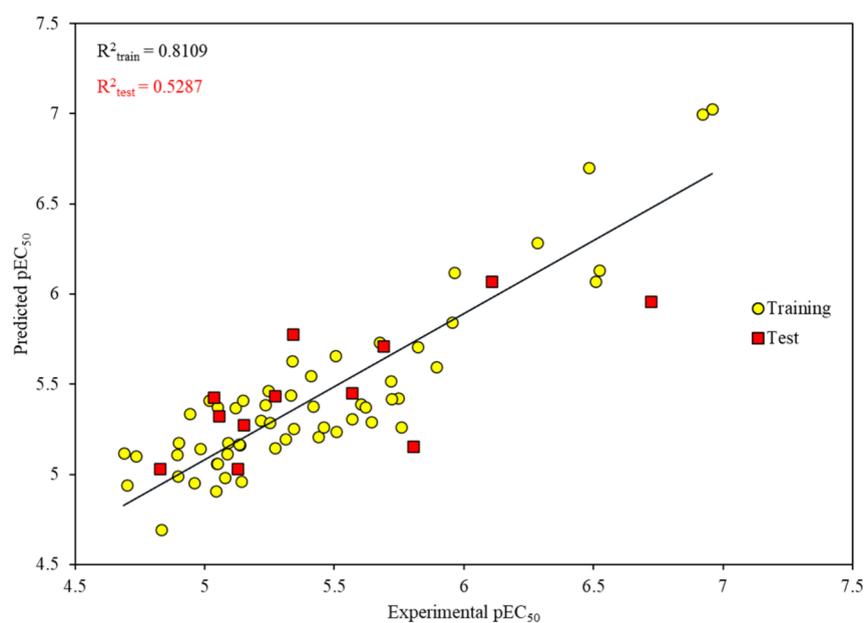
**Figure 2.** Graph obtained from the correlation of experimental values against predicted values of pEC<sub>50</sub> by the CoMFA focusing model.

error of estimate (SEE) is 0.255, and the  $F$  value is equal to 61.324. Also, the bootstrap  $r^2$  value was 0.848, which shows the good performance and reliability of the model. CoMFA-1 demonstrated that 52.5% of the variance is attributed to the steric field, while the remaining 47.5% of the variance is related to the electrostatic field. Finally, the obtained results, including high  $F$  value, high  $r_{\text{bootstrap}}^2$  value, and low SEE, confirm the reliability and accuracy of the model. There is a statistical relationship between the experimental pEC<sub>50</sub> values and the values predicted by the CoMFA-1 model as shown in Figure 1. This plot serves as a visual representation of how well the model's predictions align with the actual experimental data. In some cases, applying focusing to the CoMFA model can improve the results of the model; for this reason, the new CoMFA-2 model was built by applying focusing. This approach resulted in enhancements to several statistical parameters (such as the  $q_{\text{LOO}}^2$ ), but overall, it provided similar results to the initial model. The results obtained from this model are also shown in Table 1. Additionally, the models underwent extensive validation protocols, as listed in the

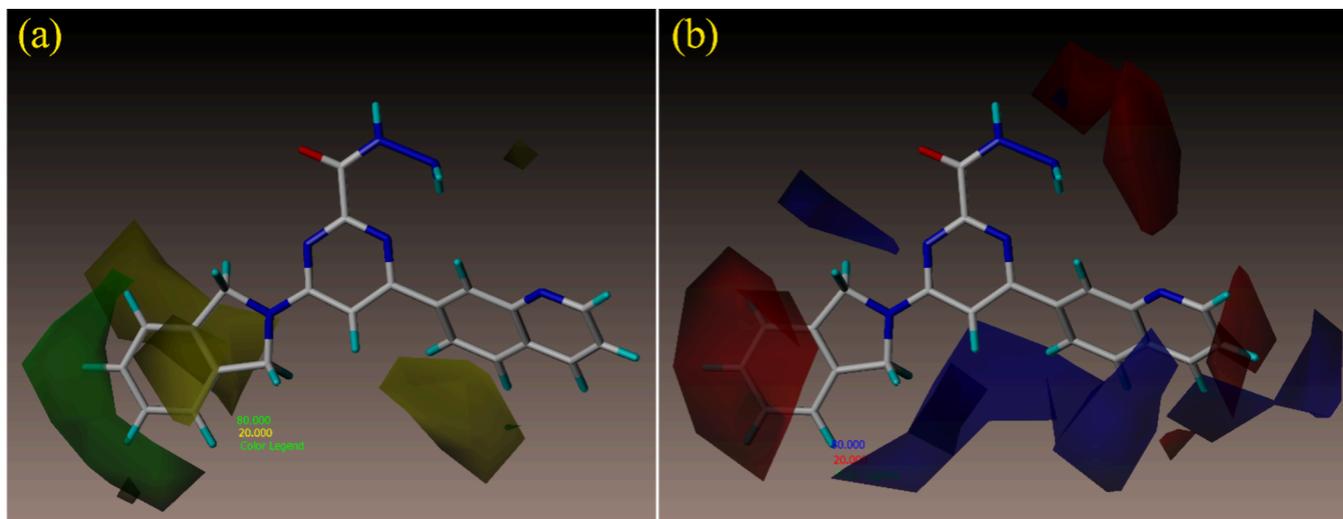
relevant table, which ultimately resulted in the acceptance and validation of the proposed models. Also, there is a statistical relationship between the experimental pEC<sub>50</sub> values and the values predicted by the CoMFA-2 model as shown in Figure 2. According to the figure, the closeness of the data points to the diagonal line reflects the model's ability to make accurate predictions. Also, for the CoMFA-2 model, the variance ratio of the fields was checked, which was 40.7 for electrostatic fields and 59.3 for steric fields, which shows that steric fields have a greater effect on changes in the activity of structures. In Table 5, the estimated pEC<sub>50</sub> values obtained through the CoMFA-1 and CoMFA-2 models are reported. Also, Table S2 lists the calculated residual values for the reported models. Data for which the calculated residual value is greater than three times the standard deviation of the residuals are considered outliers.<sup>26</sup> Therefore, two molecules 48 and 52 based on the residual values and the graphs obtained from the correlation of the predicted pEC<sub>50</sub> values against the calculated residual values (Figures S1–S3) were removed from the set due to the distance from the line (as outlier data). Also, these compounds



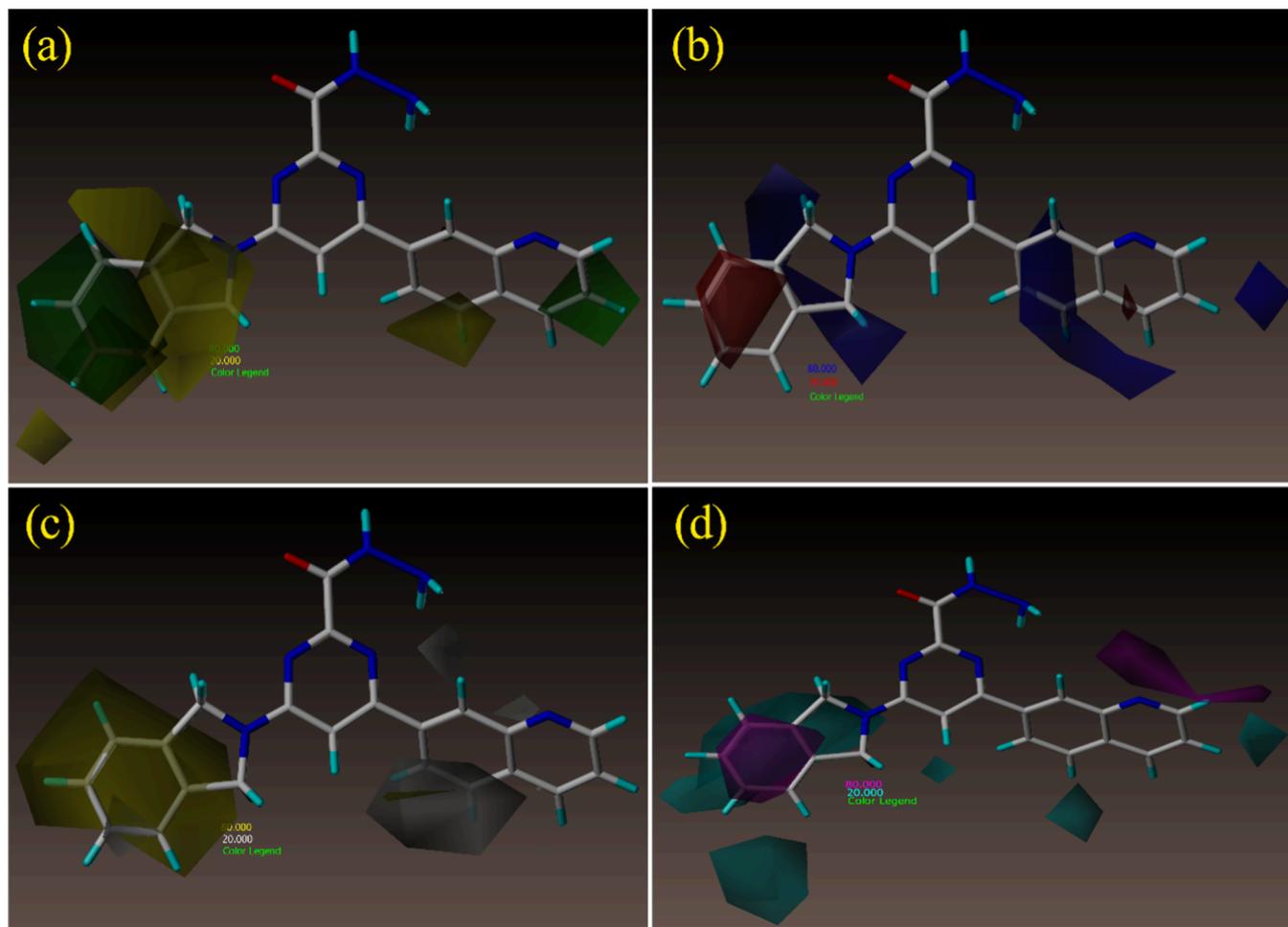
**Figure 3.** Diagram of 31 different states of CoMSIA descriptors (S, steric; E, electrostatic; H, hydrophobic; D/A, H bond donor/acceptor) with their corresponding  $q^2$  values and the number of principal components used in the model.



**Figure 4.** Graph obtained from the correlation of experimental values against predicted values of  $pEC_{50}$  by the CoMSIA model.



**Figure 5.** Contour maps obtained from the CoMFA model: (a) steric and (b) electrostatic based on molecule 50.



**Figure 6.** Contour maps obtained from the CoMSIA model: (a) steric, (b) electrostatic, (c) hydrophobic, and (d) HBA based on molecule 50.

are in the experimental set and are not effective in predicting the activity by the models. This inconsistency could be due to experimental error in measuring the  $EC_{50}$  value for the target structures.

**2.2. CoMSIA Analysis.** To explore the impact of each CoMSIA field on prediction accuracy, a comprehensive analysis of all possible combinations of CoMSIA descriptors was conducted. The training set with 57 compounds was used to optimize the number of components by  $q_{LOO}^2$  values, as depicted in Figure 3. This process aimed to identify the most influential CoMSIA fields and their respective contributions toward improving the accuracy of predictions. Among the different combinations of descriptors, the optimal statistical results were achieved when there was a combination of four descriptors: steric, electrostatic, hydrophobic, and hydrogen bond acceptor. As per the data presented in Table 1, the built model showed good statistical results for  $q_{LOO}^2$  (0.600) and  $r^2$  (0.811) when utilizing three principal components. The results of this validation process indicated both CoMSIA and CoMFA are better models for investigating structure–activity relationships for the modulation of UTRN. Also, there is a statistical relationship between the experimental  $pEC_{50}$  values and the values predicted by the CoMSIA model as shown in Figure 4. The CoMFA-1 and CoMSIA models, accompanied by their explained contour maps, were employed to justify the modulation of UTRN.

**2.3. Steric Contour Map.** The contour maps around molecule 50 (the most active compound) for steric and electrostatic fields based on the CoMFA-1 model can be seen in Figure 5a,b, respectively. To better describe the contours of different fields (to replace suitable functional groups), molecule 50 was divided into three distinct regions A, B, and C according to Figure 1. As shown in Figure 5a, the steric interactions have green and yellow default color contours. Areas where bulky groups have a favorable effect on compound modulator activity and help increase activity are marked with green contours. These green regions account for 80% of the steric field's contribution, signifying the significance of these areas in promoting favorable interactions. Conversely, the yellow contours indicate regions where bulky substitutes are unfavorable within the template molecule. These yellow regions constitute 20% of the steric field's contribution, suggesting that substituents in these areas may lead to reduced activity or unfavorable interactions with the target. Also, according to Figure 5b, electrostatic interactions have blue and red default color contours. In the areas where the blue contours are spread, by replacing the positive groups, the modulatory activity of the compound increases. These areas, contributing 80% to the electrostatic field, suggest that incorporating positively charged moieties at these positions can enhance the interaction with the target. However, using the negatively charged groups in the areas marked by red contours can improve the modulatory activity of the

compounds. These red areas account for 20% of the electrostatic field's contribution, suggesting that incorporating negatively charged substituents in these regions can lead to improved activity by facilitating specific electrostatic interactions. As can be seen, from Figures 5a and 10, around region A and related to  $R_1$  and  $R_2$  positions in the CoMFA-1 steric contour map, a large green contour is marked. This contour is accompanied by smaller yellow contours in region C near  $R_3$ , indicating that the use of bulky groups in these areas decreases the molecules' modulation activity. Based on Table 5, the use of bulky groups in region A, specifically as  $R_1$  substitutions for molecules 8, 10, and 11, leads to increased modulation activity ( $pEC_{50}$ ) values ( $5.57 < 5.61 < 5.75$ ). A similar trend is observed for molecules 31 and 32, where utilizing the bulky group ( $OCH_3 < Ph$ ) in region A and near  $R_1$  and  $R_2$  results in better modulation activity ( $4.83 < 5.08$ ). These observations demonstrate that incorporating specific bulky groups in region A and near  $R_1$  and  $R_2$  can positively influence the molecules' modulation activity, as indicated by higher  $pEC_{50}$  values. Also, by comparing molecules 33 and 34, it is clear that the biological activity increases as the molecule becomes larger in the  $R_1$  and  $R_2$  positions of region A ( $5.05 < 5.96$ ).

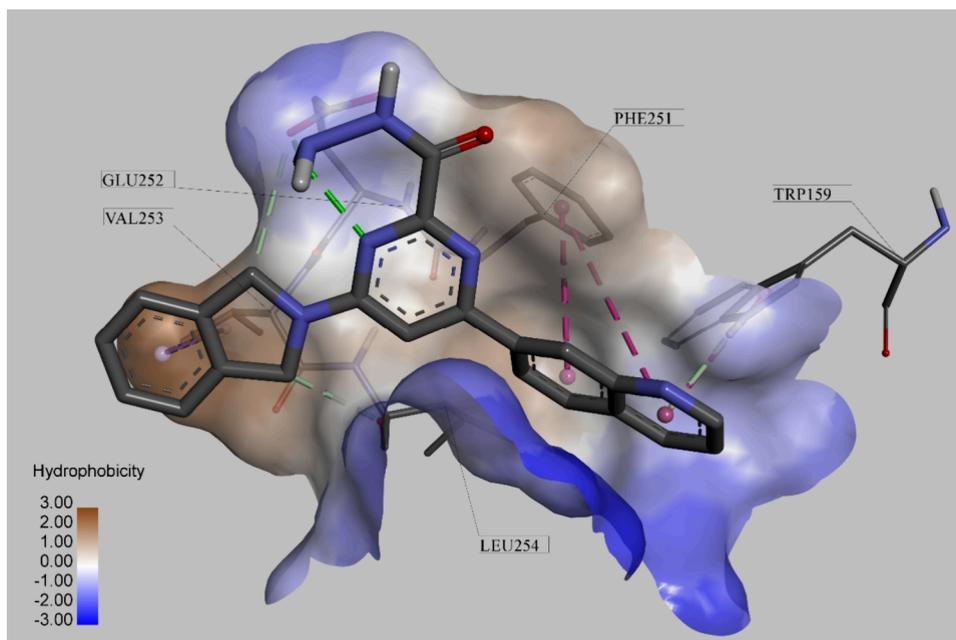
Replacing a phenol group with naphthalene in molecules 47 and 48 shows that the use of a bulky group in  $R_1$  of region C reduces the modulation activity of the compound. This is due to the presence of a yellow contour around the  $R_1$  position of region C. The presence of a relatively large yellow contour spanning between regions A and B leads to a reduction in  $pEC_{50}$  when utilizing larger groups in these regions. Analyzing compounds 20 and 23 reveals that the substitution of a phenyl group (considered a bulky group) results in a decrease in modulation activity from 5.62 to 5.51. Similarly, the comparison of compounds 19 and 22 shows that the substitution of a bulky group (phenyl) in these regions decreased the biological activity of the compounds. Also, the contour maps for spatial, electrostatic, hydrophobic, and hydrogen bond acceptor fields based on the CoMSIA model can be seen in Figure 6a–d. As shown in Figure 6a for the steric fields around molecule 50, green contours (favorability of bulky groups) and yellow contours (favorability of small groups) are similar to the distribution of contours in the CoMFA-1 model. With the comparisons made, it was concluded that changes in region A can significantly increase the modulator activity of the compounds.

**2.4. Electrostatic Contour Map.** According to Figure 5b, the contour maps for the electrostatic field around molecule 50 based on the CoMFA-1 model are known. In this figure, two red and blue contours are shown. The red contour indicates the positions where the modulation activity of compounds increases by replacing an electron-withdrawing group in these areas. But the concept of the blue contour is that by replacing an electron-donating group in these areas, the modulatory activity of the compounds increases. Looking at Figure 5b, there is a relatively large red contour in region A, in which, based on the comparison of the structures in Table 5 with their corresponding  $pEC_{50}$ , replacing the electron-withdrawing groups can increase the  $pEC_{50}$  of the compounds. According to the above points, a clear illustration of the impact of electron-withdrawing groups in enhancing the modulation activity can be observed by comparing molecules 9 ( $R_1 = 3-Ome$ ) and 11 ( $R_1 = 3-OCHF_2$ ). In molecule 11, where the electron-withdrawing group ( $3-OCHF_2$ ) is substituted in  $R_1$ , the modulation activity significantly increases to a value of

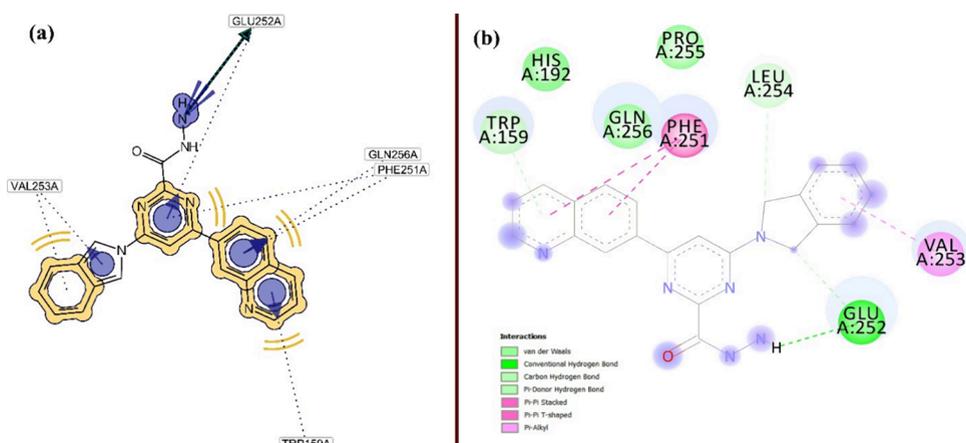
5.75. However, in molecule 9, where a 3-Ome group (with less electron-withdrawing properties than  $3-OCHF_2$ ) is substituted at the same position, the activity value is 5.15. Also, the comparison of compounds 14 and 15 shows that increasing the electron-withdrawing property in the  $R_1$  position of region A increases the modulation activity of UTRN. By adding an  $-OH$  group to the para position in compound 15,  $pEC_{50}$  increases from 4.90 (in compound 14) to 5.44. Another comparison can be made between compounds 16 and 18, where the F group was substituted by the  $OCHF_2$  group, and the modulation activity increased from 5.27 to 5.51, indicating that the modulation activity increased by replacing the electron-withdrawing groups into these regions.

If we give attention to Figure 5b again, a small red contour in region C is evident in which the substitution of the electron-withdrawing group in these regions may increase the activity of the molecules. Comparison of compounds 48 and 49 shows such an effect in which the biological activity increases from 5.45 to 6.92 by substituting an isoquinoline group instead of a naphthalene group in the  $R_1$  position of region C. Also, by adding an electron-withdrawing group instead of H in compounds 1 and 70, the modulation activity increased from 4.67 to 5.14, respectively. The mentioned comparisons confirm that the substitution of electron-withdrawing groups in the  $R_1$  position of region C increases the modulation activity of the compound. Again, according to Figure 5b, there is a wide blue contour that extends from region A to region C and also includes region B. By examining two molecules 37 and 52, it can be concluded that by adding negatively charged groups ( $-N-$  and  $CF_3$ ) to  $R_1$  of region C, the activity of the molecule decreases from 5.69 (in molecule 37) to 4.92 (in molecule 52). It is clear that the addition of negatively charged groups in this region is undesirable and reduces the modulation activity of the compound. The contour map of the electrostatic field for the CoMSIA model is shown in red and blue colors in Figure 6b. According to the figure, it can be understood that the presence of a red contour in region A, like the CoMFA-1 model, increases the modulation activity of the compound when using electron-withdrawing groups.

**2.5. Hydrophobic and HBA Contour Maps.** Hydrophobic contour maps generated using CoMSIA are shown in Figure 6c. Based on this figure, the areas where the hydrophobic properties have a favorable effect on the biological activity of the compound are marked with gray color, and the areas where the modulation activity is reduced by replacing the hydrophobic groups are marked with a yellow contour. In conclusion, regions A and C could be considered as the main substructures to increase the modulation of UTRN by manipulating the hydrophobicity of the functional groups used. Comparing compounds 4 and 2, where the addition of a  $CF_3$  group instead of CN in the  $R_1$  position of the C region, indicates that the substitution of a hydrophobic group increases the modulation activity from 5.09 to 5.25. Figure 6d shows the map of hydrogen bond acceptor contours with purple and cyan colors. Purple contours are favorable positions for hydrogen bond acceptor groups, and cyan contours are unfavorable positions for these groups. According to the data in Table 1, the field distribution percentage for different descriptors of CoMSIA has been determined, which is 13.4 for steric fields, 48.2 for electrostatic fields, 26.9 for hydrophobic fields, and 11.5 for hydrogen bond acceptor fields. These results show that the effect of electrostatic and hydrophobic



**Figure 7.** Representation of H-bonding and hydrophobic interactions with residues in the docking of molecule 50 to the receptor along with a graphical representation of the hydrophobic surface.



**Figure 8.** Interactions between molecule 50 and the corresponding receptor and their types as determined by (a) LigandScout and (b) Discovery Studio.

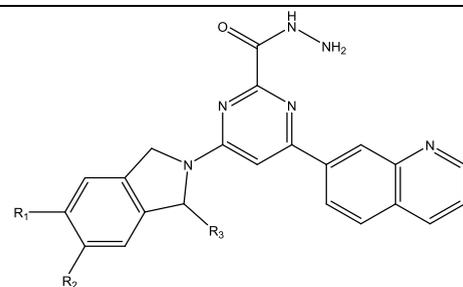
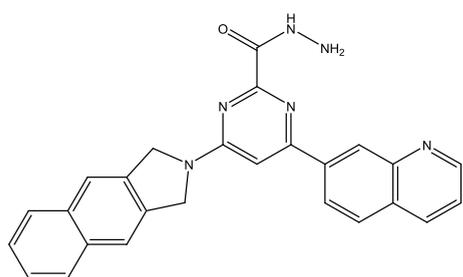
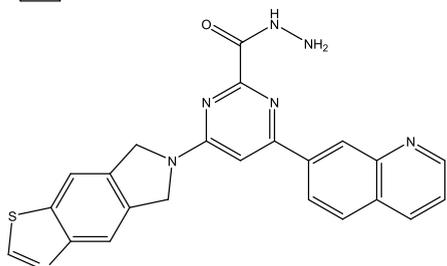
fields is more than the other descriptors of the presented model in increasing the modulation activity of compounds.

**2.6. Molecular Docking Analysis.** Compounds with  $pEC_{50}$  higher than 6 (molecules 53, 54, 51, 45, 47, 49, and 50) along with some compounds with the lowest modulating activity (molecules 1, 71, 42, 69, and 31) were selected and prepared for binding to the receptor binding site. The docking results of molecule 50 (the most potent modulator of UTRN) showed H-bond interaction with Glu252A residue. In addition, residues Leu254A, Phe251A, Val253A, and Trp159A were involved in hydrophobic interactions, which are shown in Figure 7 along with the surface display of hydrophobicity. To simplify and further identify interactions as well as to compare compounds with high activity and compounds with low activity, a pharmacophore was created using LigandScout 2.03 for the presented compounds.

The pharmacophore analysis of compounds with high modulatory activity highlighted a hydrogen bond donor interaction with Glu252A, and some number of aromatic

ring interactions were found involving Trp159A, Phe251A, Gln256A, Val253A, Leu254A, and Glu252A. Based on the results, it was found that the pharmacophore created on molecule 50 (the highest activity) covers all the features identified from the pharmacophore analysis on compounds with high activity and provides more useful information than other compounds. The results of compounds with high  $pEC_{50}$  are shown in Figure S4, and the results of molecule 50 are shown in Figure 8a. Also, comparing the pharmacophore analysis of compounds with high activity and compounds with low activity showed that the information obtained from the pharmacophore results of compounds with low activity is weak and not reliable (it does not confirm the information obtained from other molecules with low activity). On the other hand, the results of the analysis of pharmacophore compounds with high activity confirm the identified characteristics and the information obtained from other molecules with high activity. The results of compounds with low  $pEC_{50}$  are given in Figure S5. The combined docking and pharmacophore analysis

Table 2. Newly Designed Utrphin Modulators, with Their Corresponding Estimated Activity (pEC<sub>50</sub>) by the CoMFA Model

No.	Substituents			Predicted pEC <sub>50</sub> CoMFA-1
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	
				
D1	CF <sub>3</sub>	Br	H	7.14
D2	CF <sub>3</sub>	NH <sub>2</sub>	CH <sub>3</sub>	7.11
D3	Tert-Butyl	H	H	7.08
D4	Br	CF <sub>3</sub>	H	7.05
D5				6.99
D6				6.97

provided valuable insights into the interactions between high activity molecules and the receptor binding site, guiding the systematic design of potential new modulators. Based on Figure 7 and Figure S4, focusing on the distribution of hydrophobic properties around high activity molecules during docking to the receptor can serve as a valuable guide for replacing appropriate functional groups to enhance compounds' biological activity. However, comprehending how hydrophobicity may cause modulation activity in other regions throughout the data set seems challenging. This problem may be due to the lack of sufficient information in the receptor to understand the structural features of the compounds and their relationship with the changes in the activity of the molecules. The interactions identified in Figures 7 and 8a,b and Figure S4 as well as the hydrophobic contour maps of CoMSIA, which are shown in Figure 6c, match and can be useful in identifying regions involved in hydrophobic interactions.

**2.7. Design of New UTRN Modulators.** To design new structures with high modulation power, studying contour maps can be of great help. Figures 5a,b and 6a–d obtained from the CoMFA-1 and CoMSIA models provided valuable insights into the effects of different substituents in regions A, B, and C

of the reference molecule. According to the figures, by adding bulky or electron-withdrawing groups such as *tert*-butyl, phenyl, –COH, –NH<sub>2</sub>, –CF<sub>3</sub>, and Br to the R<sub>1</sub> and R<sub>2</sub> positions in region A, the activity of the compound increases. As mentioned, another solution to increase the modulation activity of compounds is to avoid the use of undesirable groups in different regions. For instance, using the positive groups, especially at positions R<sub>1</sub> and R<sub>2</sub>, or incorporating negative groups at position R<sub>3</sub> should be avoided. Based on the comprehensive analysis, the six new compounds with estimated pEC<sub>50</sub> values, higher than the reference molecule (molecule number 50), have been designed and are listed in Table 2. As mentioned, molecular docking together with 3D-QSAR results provided new insights to design compounds with higher biological activity. Docking and pharmacophore analysis were also performed on the newly designed compounds to prove their pharmacophoric features and the descriptors appearing in the QSAR model. Based on the results of this analysis, the relationships identified between the ligand and the receptor in the newly designed compounds were also confirmed. The results of pharmacophore analysis on new compounds are shown in Figure S6.

**2.8. Pseudo-Drugs and ADME Assessment.** A general method for assessing drug-likeness or figuring out whether a chemical compound with a particular pharmacological or biological activity has chemical and physical characteristics that would likely make it an orally active drug in humans is Lipinski's rule of five, also referred to as Pfizer's rule of five (RO5). According to this guideline, an orally active drug should not exceed one violation of the following criteria: (a) a maximum of five H-bond donors (N–H and O–H bonds), (b) a maximum of 10 H-bond acceptors (N or O atoms), (c) a molecular weight of no more than 500 Da, (d) the log *P* no more than five for the compounds, and (e) not more than 120 cm<sup>3</sup>/mol molar refractivity for the compound. A chemical is considered high risk in terms of oral bioavailability if it breaks more than one of these guidelines. The guidelines outline molecular characteristics that are critical to a drug's pharmacokinetics, that is, how it is absorbed, distributed, metabolized, and excreted in the human body in the "ADME" section. The rule, however, is unable to determine if a substance is pharmacologically active.<sup>27,28</sup> Using the Swiss ADME web server (<http://www.swissadme.ch/>), the physical properties and ADME parameters of the designed compounds were calculated.<sup>29,30</sup> Based on these findings summarized in Table 3, except for two compounds (D1 and D4) that only violated the molecular weight guideline, the rest of the designed compounds obeyed Lipinski's rule of five. As can be seen in Table 3, all the compounds designed in this work included the properties of a drug and showed very good synthetic accessibility. Also, Veber's rule calls into further question a 500 molecular weight cutoff. The polar surface area and the number of rotatable bonds have been found to better discriminate between orally active compounds and those that are not for a large data set of compounds. According to this rule, compounds that meet only the two criteria of 10 or fewer rotatable bonds and polar surface area no greater than 140 Å<sup>2</sup> are predicted to have good oral bioavailability.<sup>31</sup> The summary of obtained results is indicated in Table 4.

### 3. PRINCIPLES OF THE METHOD

**3.1. Data Set.** The different structures of the main compound, along with the utrophin modulatory activity (EC<sub>50</sub>), were collected from the literature.<sup>9,10</sup> The utrophin modulatory activity of all compounds, which was reported as EC<sub>50</sub> (μM), was converted to the corresponding logarithmic value [pEC<sub>50</sub> (M)]. The data set was divided into two training and test sets using the hierarchical clustering approach. All compounds used in this study with their corresponding biological activity are listed in Table 5. Also, to easily reproduce the models by researchers, the SMILES code of drawn molecules is given in Table S1.

**3.2. Hierarchical Clustering Approach.** One of the important steps in creating a QSAR model is how to divide the data set. This divide is essential to increase the test set's variety and assess the model's predictive power when extrapolating outside of the training set.<sup>32</sup> For this purpose, a hierarchical clustering approach was employed to partition the data set into the test and training sets. The CoMFA descriptors and biological activity have been used as the data set. In this approach, each molecule is categorized based on its similarities with other molecules. Eventually, this process generates clusters that group molecules with similar characteristics together.<sup>33</sup> The dendrogram resulting from hierarchical clustering is shown in Figure 9. Based on the distribution of

**Table 3. Drug-likeness Properties of the Newly Designed Utrophin Modulators Predicted by Swiss ADME**

molecule	molecular weight	rotatable bonds	H-bond acceptors	H-bond donors	heavy atoms	TPSA <sup>a</sup>	consensus log <i>P</i>	ESOL class <sup>b</sup>	GI absorption	BBB permeant	Lipinski violations	bioavailability score	PAINS <sup>c</sup>	synthetic accessibility
D1	529.31	5	8	2	34	97.03	3.86	moderately soluble	high	no	1	0.55	0	3.26
D2	479.46	5	8	3	35	123.05	3.05	moderately Soluble	high	no	0	0.55	0	3.86
D3	438.52	5	5	2	33	97.03	3.36	moderately soluble	high	no	0	0.55	0	3.42
D4	529.31	5	8	2	34	97.03	4.00	moderately soluble	high	no	1	0.55	0	3.26
D5	432.48	4	5	2	33	97.03	3.28	moderately soluble	high	no	0	0.55	0	3.19
D6	439.49	4	6	2	32	138.16	2.69	moderately soluble	low	no	0	0.55	0	3.26

<sup>a</sup>Topological polar surface area. <sup>b</sup>Estimated solubility. <sup>c</sup>Pan-assay interference structure.

**Table 4. Summary of the Results Obtained from ADME Predictions**

ADME parameter	observed range	standard Range
physicochemical property		
molecular weight	432–529	below 500
H-bond donors	two to three	not more than 5
H-bond acceptors	five to eight	not more than 10
rotatable bonds	four to five	not more than 10
heavy atoms	32 to 35	20 to 70
log <i>P</i>	2.69–4.00	less than 5
water solubility	moderately soluble	
polar surface area	97–138	no greater than 140 Å <sup>2</sup>
pharmacokinetic property		
GI absorption	low to high	
BBB permeant	no	
drug-likeness		
PAINS	0	
synthetic accessibility	3.2 to 3.9	1 (very easy) to 10 (very difficult)
Lipinski	no violations	
bioavailability score	0.55	0.55 (sufficiently absorbable oral route)

biological activity data over the full data set and the variety of chemical structures included within it, the test set compounds were chosen. Using the information obtained from hierarchical clustering, out of a total of 71 structures, 57 compounds (80%) were selected as the training set, and 14 compounds (20%) were used as the test set.

**3.3. Computational Analysis and Structural Alignment of Molecules.** The 3D chemical structures of 71 compounds were constructed and analyzed using molecular modeling methods with the aid of SYBYL-X.<sup>34</sup> The Gasteiger–Hückel technique was used to determine the atomic charges.<sup>35</sup> Subsequently, all compounds underwent optimization, and default force field parameters were used for energy minimization. The Powell-CG algorithm was utilized with 0.01 kcal/mol Å convergence criteria.<sup>36</sup> In 3D-QSAR analysis, choosing a template molecule is a critical step as it greatly influences the alignment and subsequent correct prediction of the model, and it helps in establishing a reliable and accurate model. Typically, the compound chosen as the reference molecule for alignment purposes is the most active one. Considering that in this research, compound 50 had the highest biological activity, this compound was chosen as the templet molecule for alignment purposes. Then, based on the common part of the compounds, all of them were aligned on compound 50. Figure 10 displays the reference molecule's chemical structure together with its common substructure, while Figure 11 showcases the aligned compounds.

**3.4. CoMFA/CoMSIA Procedures.** Using a probe atom in SYBYL-X, the steric and electrostatic fields were calculated for each compound everywhere in the lattice. The probe atom used in this network is an sp<sup>3</sup> carbon atom with the default values of the software (van der Waals radius of 1.52 Å and charge of +1.0). A threshold of 30 kcal/mol was set for both steric and electrostatic fields. If the value of either field at any point exceeds this threshold, the threshold replaces the field

value. Region focusing is a repeatable technique that enhances a model by assigning weights to lattice points. CoMFA focusing enhances the model compared to the initial CoMFA by assigning greater weight to specific grid points. As a result, these points contribute more significantly to improving the model's resolution and predictive power. Through careful selection of the focusing value, lattice resolution can be considerably increased compared to approaches that do not employ focusing.<sup>23</sup> In addition to grid spacing, standard deviation coefficients were utilized as distinct weighting factors to achieve a more suitable model.<sup>37</sup> In addition to CoMFA, CoMSIA was employed to generate supplementary fields, offering additional information such as hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), and hydrophobic descriptors. These extra descriptors contribute to a more comprehensive understanding of the molecular properties and interactions being analyzed. The identical lattice box that was utilized for constructing the CoMFA model was used to calculate the CoMSIA model. This ensures consistency and comparability between the two methods by maintaining the same spatial parameters and grid dimensions. As a probe atom, a carbon atom that has undergone sp<sup>3</sup> hybridization was used to calculate the five CoMSIA descriptors.<sup>24</sup>

The steric descriptors for the CoMSIA model are related to the third power of the atomic radius. The atomic partial charges serve as the basis for the electrostatic descriptions. Atomic parameters provided by Viswanadhan et al. are used to determine hydrophobic interactions.<sup>38</sup> According to Folkers et al.,<sup>39</sup> the HBD and HBA indices are calculated from experimental findings. These various descriptor types capture different aspects of molecular interactions and properties, providing a comprehensive representation for CoMSIA analysis.

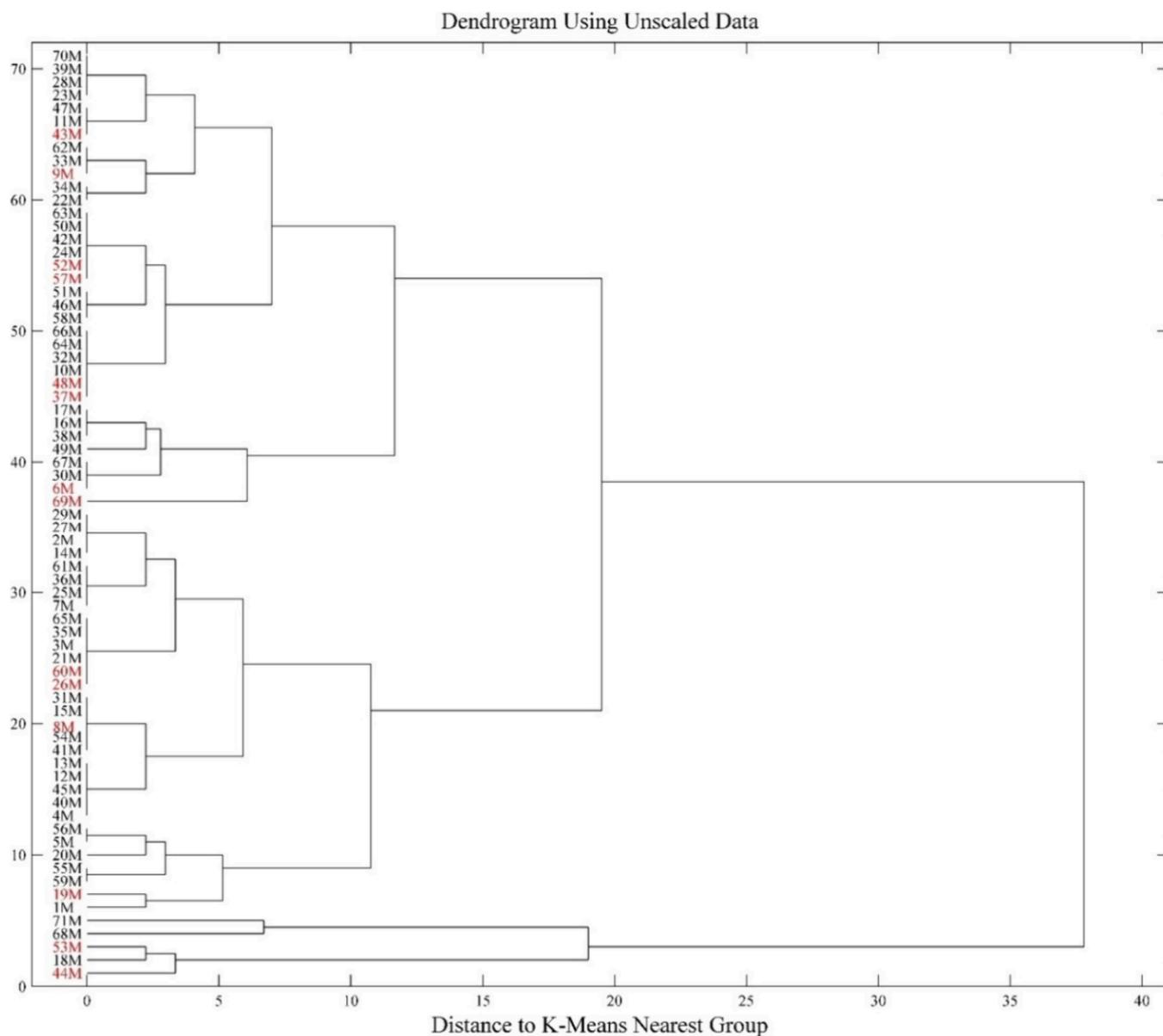
**3.5. Partial Least-Squares Calculations.** Partial least-squares (PLS) analysis is a modeling method that can be used to analyze the relationship between two sets of variables.<sup>40–42</sup> This technique is an advanced form of MLR analysis that provides the identification of the smallest set of points in the lattice by establishing the relationship between CoMFA descriptors as independent variables and the biological activity of compounds (pEC<sub>50</sub> in this case) as dependent variables. In this work, the PLS method was employed to construct various 3D-QSAR models, and the leave-one-out (LOO) cross-validation method was utilized to assess the predictive capability of these models.<sup>43</sup> This assessment involved obtaining the *q*<sup>2</sup> value as a measure of the model's estimation accuracy. Additionally, to determine the optimal number of components and minimize the standard error of predictions (SEP), the PLS approach was examined using cross-validation. Then, PLS was performed with the nonvalidation method with the optimal number of components and appropriate column filtering, and *r*<sup>2</sup>, SE, and *F* values were obtained. To confirm the validity of the created model, *q*<sup>2</sup> and *r*<sup>2</sup> coefficients are good criteria and, if their results are in the range of 0.5 and 1.0, it is considered that the model is satisfactory.<sup>44</sup> After building the final PLS model for the training set, *r*<sub>tes</sub><sup>2</sup> of the test set was also calculated.

**3.6. Molecular Docking.** In this study along 3D-QSAR research, molecular docking analysis was also performed to investigate the behavior between the modulator molecules of UTRN and its corresponding receptor. The AutoDock 4.2 software was employed for this purpose.<sup>45</sup> The receptor utilized in the docking analysis was obtained from RCSB PDB

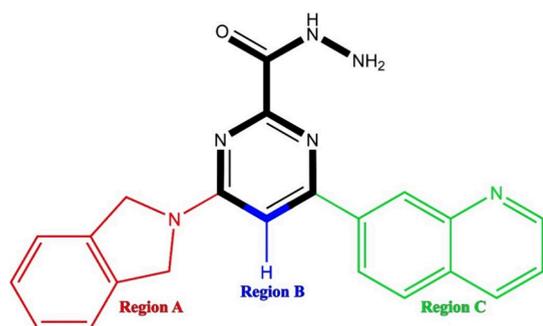
Table 5. Chemical Structure of Utrophin Modulators and Experimental and Predicted pEC<sub>50</sub> by CoMFA and CoMSIA Models

No.	Substituents			Exp. (pEC <sub>50</sub> )	Predicted activity		
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>		CoMFA-1	CoMFA-2	CoMSIA
1	Ph	H	Ph	4.67	5.08	5.08	5.12
2	Ph	H	Ph	5.25	5.55	5.55	5.47
3	Ph	H	Ph	5.09	5.14	5.14	5.18
4	Ph	H	Ph	5.09	5.28	5.29	5.11
5	Ph	H	Ph	5.46	5.23	5.19	5.26
6 <sup>a</sup>	Ph	H	Ph	5.34	5.52	5.43	5.78
7	3-OH	H	CF <sub>3</sub>	5.67	5.62	5.58	5.73
8 <sup>a</sup>	4-OH	H	CF <sub>3</sub>	5.57	5.46	5.40	5.45
9 <sup>a</sup>	3-OMe	H	CF <sub>3</sub>	5.15	5.30	5.21	5.27
10	4-OMe	H	CF <sub>3</sub>	5.61	5.29	5.35	5.39
11	3-OCHF <sub>2</sub>	H	CF <sub>3</sub>	5.75	5.45	5.35	5.42
12	3-Cl	H	CF <sub>3</sub>	5.22	5.17	5.13	5.30
13	4-Cl	H	CF <sub>3</sub>	5.42	5.43	5.43	5.38
14	3-F	H	CF <sub>3</sub>	4.90	5.12	5.11	5.18
15	3-F,4-OH	H	CF <sub>3</sub>	5.44	5.13	5.11	5.21
16	3F	H	CF <sub>3</sub>	5.27	5.04	5.05	5.15
17	2,5-F	H	CF <sub>3</sub>	5.04	5.04	5.05	4.91
18	3-OCHF <sub>2</sub>	H	CF <sub>3</sub>	5.51	5.18	5.20	5.24
19 <sup>a</sup>	H	H	-	5.13	5.18	5.20	5.03
20	H	H	-	5.62	5.40	5.39	5.37
21	H	H	-	5.76	5.42	5.41	5.26
22	H	H	-	5.12	5.60	5.66	5.37
23	H	H	-	5.51	5.72	5.76	5.66
24	H	H	-	5.31	5.31	5.37	5.20
25	H	H	-	5.02	5.46	5.45	5.41
26 <sup>a</sup>	H	H	-	5.06	5.30	5.19	5.33
27	H	H	-	5.34	5.35	5.29	5.63
28	H	H	-	5.41	5.38	5.41	5.55
29	H	H	-	4.98	5.13	5.10	5.14
30	H	H	H	4.94	4.94	5.22	5.24
31	H	H	H	4.83	4.83	4.74	4.80
32	H	H	H	5.08	5.08	5.06	5.17
33	H	H	H	5.05	5.05	5.57	5.55
34	H	H	H	5.96	5.96	5.90	5.94
35	H	H	H	5.14	5.14	5.23	5.29
36	H	H	H	5.82	5.82	5.74	5.69
37 <sup>a</sup>	H	H	H	5.69	5.69	5.73	5.71
38	H	H	H	5.13	5.13	5.20	5.26
39	H	H	H	5.89	5.42	5.33	5.60
40	H	H	H	5.72	5.41	5.32	5.52
41	H	H	H	5.23	5.50	5.50	5.38
42	H	H	H	4.74	5.08	5.10	5.10
43 <sup>a</sup>	H	H	-	5.27	5.53	5.49	5.44
44 <sup>a</sup>	H	H	-	6.72	5.92	5.92	5.96
45	H	H	-	6.51	6.05	6.04	6.07
46	H	H	-	5.96	6.20	6.21	6.12
47	H	H	-	6.52	5.95	5.97	6.13
48 <sup>a</sup>	H	H	-	5.45	6.58	6.64	6.57
49	H	H	-	6.92	6.86	6.82	7.00
50	H	H	-	6.96	6.97	6.90	7.03
51	H	H	-	6.48	6.85	6.86	6.70
52 <sup>a</sup>	H	H	-	4.92	6.10	6.15	5.98
53 <sup>a</sup>	H	H	-	6.11	6.18	6.25	6.07
54	H	H	-	6.28	6.18	6.20	6.28
55	2-F	H	2-F	4.96	5.043	5.007	4.954
56	4-F	H	4-F	4.89	5.118	5.106	5.11
57 <sup>a</sup>	3-OMe	H	3-OMe	5.04	5.136	5.095	5.426
58	4-OMe	H	4-OMe	5.15	5.257	5.261	5.41
59	N	N	N	5.05	5.03	5.03	5.06
60 <sup>a</sup>	OH	H	H	5.81	5.29	5.27	5.16
61	OMe	H	H	5.64	5.40	5.37	5.29
62	OEt	H	H	5.57	5.41	5.37	5.31
63	OPr	H	H	5.25	5.37	5.34	5.29
64	*Pr	H	H	5.34	5.31	5.36	5.25
65	CH <sub>2</sub> OH	H	H	5.72	5.52	5.49	5.42
66	CH <sub>2</sub> OMe	H	H	5.33	5.38	5.38	5.44
67	OCH <sub>2</sub> CH <sub>2</sub> OMe	H	H	5.05	4.86	4.87	5.06
68	OCH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> NHAc	H	H	4.90	4.88	4.91	4.99
69 <sup>a</sup>	OCH <sub>2</sub> CH <sub>2</sub> OCH <sub>2</sub> CH=CH	H	H	4.83	4.86	4.87	5.03
70	H	H	H	5.14	5.14	5.17	4.86
71	OCH <sub>2</sub> CH <sub>2</sub> OCH <sub>2</sub> CH=CH	H	H	4.70	4.97	5.05	4.94

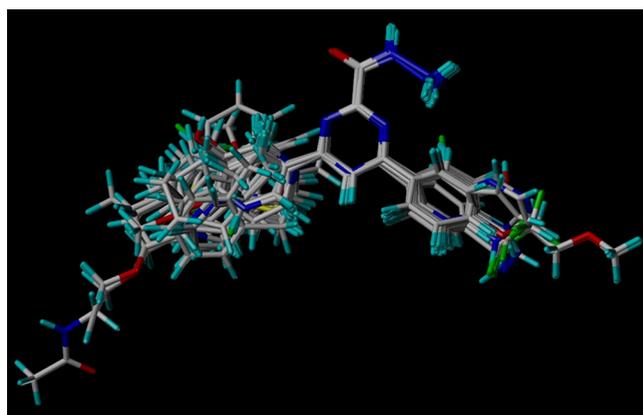
<sup>a</sup>Test set molecules. The asterisk symbol (\*) indicates outliers.



**Figure 9.** Hierarchical clustering dendrogram for two training and testing sets (testing set molecules are marked in red).



**Figure 10.** Chemical structure of the target molecule (molecule 50). Three regions A, B, and C are marked with different colors, and the common area is bold.



**Figure 11.** Molecule 50 on which all the structures are aligned.

(<https://www.rcsb.org/>) with PDB ID 1QAG, representing the actin-binding region of the dystrophin homologue utrophin at a resolution of 3.0 Å. Before the docking procedures, appropriate preparations were made for the protein structure, involving the removal of ions and water molecules. Then, to prepare the protein, Kollman charges and polar-H were added to it.<sup>46</sup> For correct docking, a grid box with dimensions of 65 × 65 × 65 Å, a grid spacing of 1 Å, and Cartesian [*x*, *y*, *z*] coordinates of [29.9885, 32.2400, 71.9803] was defined to effectively cover the docking site. A number of compounds with high pEC<sub>50</sub> along with a few compounds with low UTRN modulatory activity were selected as ligands during docking analysis, and the number of 100 runs based on the genetic algorithm was considered to select a conformation with suitable binding. These conformations were then evaluated to identify orientations with lower binding energy levels and examine conformational clusters with more members, and the best conformation was selected for further investigation by LigandScout 2.03.<sup>47</sup>

#### 4. CONCLUSIONS

In this work, for the development of UTRN modulators, the 3D-QSAR studies based on CoMFA and CoMSIA models were used. By using the PLS approach, the relationship between the biological activity of molecules and CoMFA and CoMSIA descriptors was investigated and the estimation power and accuracy of the created models were confirmed. To achieve better results, CoMFA focusing was used, and the obtained results were similar to the CoMFA model without focusing. Analyzing the built models as well as examining the contour maps of CoMFA and CoMSIA descriptors and their effects on the modulation activity of molecules could provide a new path for the preparation of UTRN modulators with higher potency. In addition, molecular docking results identified a number of hydrophobic interactions associated with the corresponding receptor. By examining the results, it was understood that the positions of R<sub>1</sub> and R<sub>2</sub> in region A have a good potential to increase the biological activity of the molecules by replacing the bulky and electron-withdrawing groups. This conclusion is confirmed by the combination of molecular docking and 3D-QSAR analysis and results in the design of six new compounds with higher modulation activity.

#### ■ ASSOCIATED CONTENT

##### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.4c01225>.

SMILES code table for all compounds and residual values and graph obtained from the correlation of predicted values of pEC<sub>50</sub> against calculated residual values for all models (PDF)

#### ■ AUTHOR INFORMATION

##### Corresponding Author

Eslem Pourbasheer – Department of Chemistry, Faculty of Science, University of Mohaghegh Ardabili, Ardabil 56199-11367, Iran; [orcid.org/0000-0001-8969-4341](https://orcid.org/0000-0001-8969-4341);  
Email: [e.pourbasheer@uma.ac.ir](mailto:e.pourbasheer@uma.ac.ir)

##### Author

Reza Mahmoudzadeh Laki – Department of Chemistry, Faculty of Science, University of Mohaghegh Ardabili,

Ardabil 56199-11367, Iran; [orcid.org/0000-0001-7061-1542](https://orcid.org/0000-0001-7061-1542)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acsomega.4c01225>

#### Author Contributions

The manuscript was written through the contributions of all the authors. R.M.L. did all the drawing of molecules and calculations with software. E.P. acted as the corresponding author, supervising the overall research and the manuscript preparation. All authors reviewed the manuscript.

#### Notes

The authors declare no competing financial interest.

#### ■ ACKNOWLEDGMENTS

Respectfully, we have undertaken this project without any financial support and we were able to complete it without the need for external financial assistance. We sincerely appreciate all individuals who contributed in any way to this project and aided us along the way.

#### ■ REFERENCES

- (1) Mercuri, E.; Bönnemann, C. G.; Muntoni, F. Muscular dystrophies. *Lancet* **2019**, *394* (10213), 2025–2038.
- (2) Santos, R.; Goncalves, A.; Oliveira, J.; Vieira, E.; Vieira, J. P.; Evangelista, T.; Moreno, T.; Santos, M.; Fineza, I.; Bronze-da-Rocha, E. New variants, challenges and pitfalls in DMD genotyping: implications in diagnosis, prognosis and therapy. *Journal of human genetics* **2014**, *59* (8), 454–464.
- (3) Guiraud, S.; Davies, K. E. Pharmacological advances for treatment in Duchenne muscular dystrophy. *Current opinion in pharmacology* **2017**, *34*, 36–48.
- (4) Guiraud, S.; Roblin, D.; Kay, D. E. The potential of utrophin modulators for the treatment of Duchenne muscular dystrophy. *Expert Opinion on Orphan Drugs* **2018**, *6* (3), 179–192.
- (5) Iftikhar, M.; Frey, J.; Shohan, M. J.; Malek, S.; Mousa, S. A. Current and emerging therapies for Duchenne muscular dystrophy and spinal muscular atrophy. *Pharmacology & Therapeutics* **2021**, *220*, No. 107719.
- (6) Hadwen, J.; Farooq, F.; Witherspoon, L.; Schock, S.; Mongeon, K.; MacKenzie, A. Anisomycin activates utrophin upregulation through a p38 signaling pathway. *Clinical and Translational Science* **2018**, *11* (5), 506–512.
- (7) Echigoya, Y.; Mouly, V.; Garcia, L.; Yokota, T.; Duddy, W. In silico screening based on predictive algorithms as a design tool for exon skipping oligonucleotides in Duchenne muscular dystrophy. *PLoS One* **2015**, *10* (3), No. e0120058.
- (8) Soblecheró-Martín, P.; López-Martínez, A.; de la Puente-Ovejero, L.; Vallejo-Illarramendi, A.; Arechavala-Gomez, V. Utrophin modulator drugs as potential therapies for Duchenne and Becker muscular dystrophies. *Neuropathology and Applied Neurobiology* **2021**, *47* (6), 711–723.
- (9) Vuorinen, A.; Wilkinson, I. V. L.; Chatzopoulou, M.; Edwards, B.; Squire, S. E.; Fairclough, R. J.; Bazan, N. A.; Milner, J. A.; Conole, D.; Donald, J. R.; Shah, N.; Willis, N. J.; Martínez, R. F.; Wilson, F. X.; Wynne, G. M.; Davies, S. G.; Davies, K. E.; Russell, A. J. Discovery and mechanism of action studies of 4, 6-diphenylpyrimidine-2-carbohydrazides as utrophin modulators for the treatment of Duchenne muscular dystrophy. *Eur. J. Med. Chem.* **2021**, *220*, No. 113431.
- (10) Chatzopoulou, M.; Conole, D.; Emer, E.; Rowley, J. A.; Willis, N. J.; Squire, S. E.; Gill, B.; Brough, S.; Wilson, F. X.; Wynne, G. M.; Davies, S. G.; Davies, K. E.; Russell, A. J. Structure-activity relationships of 2-pyrimidinecarbohydrazides as utrophin modulators for the potential treatment of Duchenne muscular dystrophy. *Bioorg. Med. Chem.* **2022**, *69*, No. 116812.

- (11) Nekoei, M.; Mohammadhosseini, M.; Pourbasheer, E. A quantitative structure-activity relationship study on CXL017 derivatives as effective drugs for cancer treatment. *Journal of the Chinese Chemical Society* **2021**, *68* (10), 1972–1986.
- (12) Pourbasheer, E.; Aalizadeh, R.; Ganjali, M. R. QSAR study of CK2 inhibitors by GA-MLR and GA-SVM methods. *Arabian Journal of Chemistry* **2019**, *12* (8), 2141–2149.
- (13) Rafiei, H.; Khanzadeh, M.; Mozaffari, S.; Bostanifar, M. H.; Avval, Z. M.; Aalizadeh, R.; Pourbasheer, E. QSAR study of HCV NSSB polymerase inhibitors using the genetic algorithm-multiple linear regression (GA-MLR). *Excli Journal* **2016**, *15*, 38–53.
- (14) Pourbasheer, E.; Vahdani, S.; Aalizadeh, R.; Banaei, A.; Ganjali, M. R. QSAR study of prolylcarboxypeptidase inhibitors by genetic algorithm: Multiple linear regressions. *Journal of Chemical Sciences* **2015**, *127* (7), 1243–1251.
- (15) Pourbasheer, E.; Shokouhi Tabar, S.; Masand, V. H.; Aalizadeh, R.; Ganjali, M. R. 3D-QSAR and docking studies on adenosine A(2A) receptor antagonists by the CoMFA method. *SAR QSAR Environ. Res.* **2015**, *26* (6), 461–477.
- (16) Pourbasheer, E.; Aalizadeh, R.; Ebadi, A.; Ganjali, M. R. 3D-QSAR Analysis of MCD Inhibitors by CoMFA and CoMSIA. *Combinatorial Chemistry & High Throughput Screening* **2015**, *18* (8), 751–766.
- (17) Pourbasheer, E.; Aalizadeh, R.; Shokouhi Tabar, S.; Ganjali, M. R.; Norouzi, P.; Shadmanesh, J. 2D and 3D Quantitative Structure–Activity Relationship Study of Hepatitis C Virus NSSB Polymerase Inhibitors by Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis Methods. *J. Chem. Inf. Model.* **2014**, *54* (10), 2902–2914.
- (18) Pourbasheer, E.; Aalizadeh, R. 3D-QSAR and molecular docking study of LRRK2 kinase inhibitors by CoMFA and CoMSIA methods. *SAR and QSAR in Environmental Research* **2016**, *27* (5), 385–407.
- (19) Pourbasheer, E.; Ahmadpour, S.; Zare-Dorabei, R.; Nekoei, M. Quantitative structure activity relationship study of p38 $\alpha$  MAP kinase inhibitors. *Arabian Journal of Chemistry* **2017**, *10* (1), 33–40.
- (20) Pourbasheer, E.; Aalizadeh, R.; Ganjali, M. R.; Norouzi, P. QSAR study of IKK $\beta$  inhibitors by the genetic algorithm: multiple linear regressions. *Medicinal Chemistry Research* **2014**, *23*, 57–66.
- (21) Verma, J.; Khedkar, V. M.; Coutinho, E. C. 3D-QSAR in drug design—a review. *Current topics in medicinal chemistry* **2010**, *10* (1), 95–115.
- (22) Kubinyi, H.; Folkers, G.; Martin, Y. C. *3D QSAR in drug design: recent advances*. 2006.
- (23) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110* (18), 5959–5967.
- (24) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *Journal of medicinal chemistry* **1994**, *37* (24), 4130–4146.
- (25) Pourbasheer, E.; Amanlou, M. 3D-QSAR analysis of anti-cancer agents by CoMFA and CoMSIA. *Medicinal Chemistry Research* **2014**, *23*, 800–809.
- (26) Lehmann, R. 3  $\sigma$ -rule for outlier detection from the viewpoint of geodetic adjustment. *Journal of Surveying Engineering* **2013**, *139* (4), 157–165.
- (27) Shetty, P. R.; Shivaraja, G.; Krishnaswamy, G.; Pruthviraj, K.; Mohan, V. C.; Sreenivasa, S. Pyrazole schiff bases: synthesis, characterization, biological screening, in silico ADME and molecular docking studies. *J. Heterocycl Chem.* **2020**, *30*, 123–130.
- (28) Lipinski, C. A. Lead-and drug-like compounds: the rule-of-five revolution. *Drug discovery today: Technologies* **2004**, *1* (4), 337–341.
- (29) Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep* **2017**, *7*, 42717.
- (30) Daina, A.; Zoete, V. A boiled-egg to predict gastrointestinal absorption and brain penetration of small molecules. *ChemMedChem* **2016**, *11* (11), 1117–1121.
- (31) 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. *Journal of medicinal chemistry* **2002**, *45* (12), 2615–2623.
- (32) Pirhadi, S.; Ghasemi, J. B. 3D-QSAR analysis of human immunodeficiency virus entry-1 inhibitors by CoMFA and CoMSIA. *European journal of medicinal chemistry* **2010**, *45* (11), 4897–4903.
- (33) Hartigan, J. A. *Clustering algorithms*; John Wiley & Sons, Inc.: New York, NY, 1975.
- (34) SYBYL-X, T. 1.2; Tripos International, St.: Louis, USA, 2010.
- (35) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **1980**, *36* (22), 3219–3228.
- (36) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. Validation of the general purpose tripos 5.2 force field. *Journal of computational chemistry* **1989**, *10* (8), 982–1012.
- (37) Aalizadeh, R.; Pourbasheer, E.; Ganjali, M. R. Analysis of B-Raf<sup>V600E</sup> V 600 E inhibitors using 2D and 3D-QSAR, molecular docking and pharmacophore studies. *Molecular diversity* **2015**, *19*, 915–930.
- (38) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *Journal of chemical information and computer sciences* **1989**, *29* (3), 163–172.
- (39) Folkers, G.; Merz, A.; Rognan, D. In *3D QSAR in Drug Design: Theory. Methods and Applications*, 583.
- (40) Geladi, P. Notes on the history and nature of partial least squares (PLS) modelling. *Journal of Chemometrics* **1988**, *2* (4), 231–246.
- (41) Khajehsharifi, H.; Pourbasheer, E.; Tavallali, H.; Sarvi, S.; Sadeghi, M. The comparison of partial least squares and principal component regression in simultaneous spectrophotometric determination of ascorbic acid, dopamine and uric acid in real samples. *Arabian Journal of Chemistry* **2017**, *10*, S3451–S3458.
- (42) Mahramyari, S.; Pourbasheer, E.; Banaei, A.; Ganjali, M. R.; Norouzi, P. Simultaneous spectrophotometric determination of ceftazidime and sulbactam using multivariate calibration methods. *Rsc Advances* **2014**, *4* (77), 41039–41044.
- (43) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5* (3), 735–743.
- (44) Wehrens, R.; Putter, H.; Buydens, L. M. The bootstrap: a tutorial. *Chemometrics and intelligent laboratory systems* **2000**, *54* (1), 35–52.
- (45) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791. From NLM
- (46) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *Journal of computational chemistry* **1984**, *5* (2), 129–145.
- (47) Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* **2005**, *45* (1), 160–169.