Full Paper

www.molinf.com

molecular informatics

DOI: 10.1002/minf.201700080

# Multi-Objective Optimization of Benzamide Derivatives as Rho Kinase Inhibitors

Giovanna Cardoso Gajo,[a] Daniela Rodrigues Silva,[a] Stephen J. Barigye,[a] and Elaine Fontes Ferreira da Cunha*[a]

**Abstract:** Despite recent advances in Computer Aided Drug Discovery and High Throughput Screening, the attrition rates of drug candidates continue to be high, underscoring the inherent complexity of the drug discovery paradigm. Indeed, a compromise between several objectives is often required to obtain successful clinical drugs. The present manuscript details a multi-objective workflow that integrates the 4D-QSAR and molecular docking methods in the simultaneous modeling of the Rho Kinase inhibitory activity and acute toxicity of Benzamide derivatives. To this end, the $pIC_{50}/pLD_{50}$ ratio is considered as the response variable, permitting the concurrent modeling of both properties and representing a shift from classical step-by-step evaluations. The 4D-QSAR strategy is used to generate the Grid Cell Occupancy Descriptors (GCODs), and Stochastic Gradient Boosting (SGB) and Partial Least Squares (PLS) methods as the model fitting techniques. While the statistical parameters for the PLS model do not meet established criteria for acceptability, the SGB model yields satisfactory performance, with correlation coefficients $r^2 = 0.95$ and $r^2pred = 0.65$ for the training and test set, respectively. Posteriorly, the structural interpretation of the most relevant GCODs according to the SGB model is performed, allowing for the proposal of 139 novel benzamide derivatives, which are then screened using the same model. Of these 9 compounds were predicted to possess $pIC_{50}/pLD_{50}$ ratio values higher than those for the employed dataset. Finally, in order to corroborate the results obtained with the SGB model, a docking simulation was formed to evaluate the binding affinity of the proposed molecules to the ROCK2 active site and 3 chemical structures (i.e. p6, p14 and p131) showed higher binding affinity than the most active compound in the training set, while the rest generally demonstrated comparable behavior. It may therefore be concluded that the consensus models that intertwine the 4D-QSAR and molecular docking methods contribute to more reliable virtual screening and compound optimization experiments. Additionally, the use of multi-objective modeling schemes permits the simultaneous evaluation of different chemical and biological profiles, which should contribute to the control *a priori* of causative factors for the high attrition rates in later drug discovery phases.

**Keywords:** 4D-QSAR · Benzamide · Boosting Trees · ROCK2 · Virtual Screening · Protein-ligand Docking

## 1 Introduction

Rho kinase (EC: 2.7.11.1) is a serine / threonine protein kinase implicated in the regulation of the vascular tone, cell proliferation, inflammation and oxidative stress. It contains an N-terminal catalytic domain and can be expressed by two isoforms, i.e. ROCK1 and ROCK2, respectively. These isoforms exhibit 64% sequence identity and 79% overall similarity, while their kinase domains exhibit 92% identity and 97% similarity, respectively.[1] These isoforms are distributed in different tissues; ROCK1 is expressed in the lungs, liver, stomach, spleen, kidney and testis; while ROCK2 is highly expressed in the brain tissue, heart and muscle.[2] Experimental evidence suggests ROCK's involvement in various cellular functions such as cell shape, motility, secretion, proliferation and gene expression.[3] On the other hand, the RhoA/ROCK pathway has been reported to be involved in angiogenesis, cerebral ischemia, erectile dysfunction, glomerulosclerosis, hypertension, myocardial hypertrophy, myocardial ischemia-reperfusion injury, neointima formation, pulmonary hypertension and vascular remodeling.[4]

The implication of ROCK in mediating various cellular functions and its involvement in a lot of severe pathological processes makes the development of ROCK inhibitors attractive as it would certainly be beneficial in the treatment of multiple diseases.[5] Particularly, ROCK inhibitors have been suggested to be beneficial in the treatment of Alzheimer's disease, bronchial asthma, cancers, demyelinating diseases, glaucoma, and osteoporosis.[4] However, regardless of efforts in the search of novel ROCK inhibitors, Fasudil is the only approved ROCK inhibitor for clinical use in Japan for the treatment of subarachnoid hemorrhage.[6] This fact epitomizes the inherent challenges encountered in the drug discovery and development (DDD) process.

Indeed, despite the recent advances in Computer Aided Drug Discovery and High Throughput Screening, the high attrition rates in the DDD process continue to be a key

[a] *G. Cardoso Gajo, D. Rodrigues Silva, S. J. Barigye, E. F. F. da Cunha*
*Department of Chemistry (DQI), Federal University of Lavras (UFLA), P.O. Box 3037, CEP 37200–000, Lavras, Minas Gerais, Brazil*
*E-mail: elaine_cunha@ufla.br*
*elaineffdacunha@gmail.com*

challenge[7] The failure of a drug candidate is usually attributed to multiple factors, such as low bioactivity, poor pharmacokinetic (PK) and pharmacodynamics (PD) profiles, as well as high toxicity. In this sense, developing a successful clinical drug requires the best compromise between several objectives, which often compete with each other. A recent study demonstrated that 66% of the Clinical Phase III failures were due to poor efficacy, with the underlying reason being inadequate PK & PD profiles, while the rest of the failures were attributed to toxicological issues.[8] The ADME/Tox evaluations involve sequential and iterative assessments of the efficacy, PK, PD, metabolic and toxicological properties of drug candidates. This step-by-step approach may prove to be frustrating and time consuming, since initially promissory candidates may be abandoned at later stages. It would be therefore of interest to incorporate a multi-objective framework, where different properties are simultaneously modeled in the very preliminary stages of the DDD process to provide hints on favorable lead compound optimization modes.

Given the enormous costs and prolonged time frames experienced during *in vitro* and *in vivo* ADME/Tox evaluations, in addition to the adoption of stringent policies aimed at regulating the use of laboratory animals for experimentation, there has been a fundamental shift to *in silico* methods, as a quick and cost-effective alternative in the screening of New Molecular Entities (NMEs) and lead compound optimization.[9] In this sense, the QSAR (Quantitative Structure-Activity Relationship) methodology has become an essential tool for *in silico* ADME/Tox evaluations, and has been adopted by pharmaceutical companies, academia, well as many regulatory agencies (for predictive toxicology), contributing to the reduction in time, cost and animal testing during the DDD process.[10]

The QSAR modeling paradigm is comprised of several approaches ranging from simple evaluations of molecular similarity to more complex methods such as 3D/4D QSAR methods, which incorporate information on the conformational behavior of ligands as well as their target-ligand interaction modes through molecular dynamics simulations.[11] The utility of these techniques in the predicting of biological properties of molecules has been extensively reported in the literature.[12–15] The 4D-QSAR method provides information on the 3D molecular conformations and interaction modes, thus constituting an important advantage over other 0D-3D QSAR methods.

However, with the arsenal of currently available HTS protocols coupled with the often frustrating performance of classical QSAR methods in general, the 4D-QSAR methodology has become rather deprecated in DDD. As a strategy to revitalize the utility and reliability of this method in virtual screening experiments, we propose a consensus workflow which intertwines the 4D-QSAR and molecular docking methods. In principle, coupling the two conceptually different approaches should help to offset the weakness inherent to each of these methods and thus enhancing their collective performance. For instance, compounds with structural moieties deemed as favorable for a particular bioactivity profile, may not possess the adequate binding affinity due to factors such as steric or entropic penalties. On the other hand, compounds with good binding affinities may be agonists instead of antagonists and vice-versa. In this sense, coupling the 4D-QSAR and molecular docking methods helps to evaluate both the bioactivity and binding affinity profiles of screened compounds, thus enhancing the reliability of the obtained results.

Therefore, the objective of the present manuscript is to evaluate the utility of a consensus workflow integrating the 4D-QSAR and molecular docking methods in the modeling of novel benzamide derivatives with potential ROCK2 inhibitory activity and a low acute toxicity profile. Several reports in the literature have highlighted the risk of off-target toxicity by Kinase inhibitors in general.[16] Indeed, compound Y27637 which is a widely used standard in the ROCK inhibitory tests has not yet been approved for clinical use up to date, probably due to its high acute toxicity (Category 4).[17] Additionally, this compound has been shown to mimic the hydrophobic statin-induced apoptosis, and thus causing myotoxicity in fibroblasts and rat models.[18] Taking into account these precedents, it would therefore be of interest to steer the optimization process towards enhanced ROCK2 inhibitory activity and lower toxicity, simultaneously. To this end, the $pIC_{50}/pLD_{50}$ ratio is considered as the response variable and the 4D-QSAR strategy is employed to generate the Grid Cell Occupancy Descriptors (GCODs), posteriorly used in the model building based on the Stochastic Gradient Boosting (SGB) and Partial Least Squares (PLS) model fitting techniques,[19,20] respectively. Lastly, a protein-ligand docking simulation is performed to verify the binding modes of the proposed derivatives.

## 2 Materials and Methods

### 2.1 Biological Data

For this study, a series of 81 benzamide derivatives were employed. The ROCK2 inhibitory activity of these compounds has been experimentally evaluated and reported in the literature.[21] (The chemical structures of the benzamide derivatives employed in the present study are provided as Supplementary Information, Table S1). The 3D structures of the benzamide derivatives were optimized using the CHARMM force field[22] available in the *Discovery Studio software*.[23] On the other hand, due to the unfortunate inexistence of experimentally determined acute toxicity data for the benzamide derivatives employed in the present study, the $LD_{50}$ values for these compounds were estimated using the rat model based admetSAR predictor, freely available online at http://lmmd.ecust.edu.cn:8000/.

The $pLD_{50}$ was calculated as suggested by Europe Union, 2010[24] and the ratio of $pIC_{50}/pLD_{50}$ determined. The overall data set was divided into a training set (70% of the overall set) and an external evaluation set (30% of the overall set) using a random selection procedure. External validation has been strongly emphasized in the literature as the only standard practice for assessing the earnest predictive ability of QSAR models. In fact, it has been recommended that test set selection and external validation should be carried out a number of times in order to identify the ranges of external predictivity, instead of conditioning the model to a single external set.[25,26]

### 2.2 Molecular Dynamics Simulation

The molecular dynamics simulation (MDS) process is aimed at generating a conformational ensemble profile (CEP) for each compound and thus investigate ligand conformational flexibility . This method calculates the system's behavior based on Newton's equations of motion for a discrete time interval. The set of simulation settings yields a trajectory of system atoms providing thermodynamically accessible conformations for the compounds under analysis.[27]

The MDS was carried out using the MOLSIM program with an extended MM2 force field[28] available in 4D-QSAR software[29]. The temperature for dynamics was adjusted to 300 K in order to guarantee proximity to the temperature used in biological assays. The simulation sampling time was 100 ps with 0.001 ps intervals. A distance dependent dielectric function was also applied in order to model the explicit solvent effect. In addition, common atoms (C1, C4, C7, C10, C12) were fixed to prevent a large conformational variation of the ligand, which could impair alignment analysis (Figure 1).

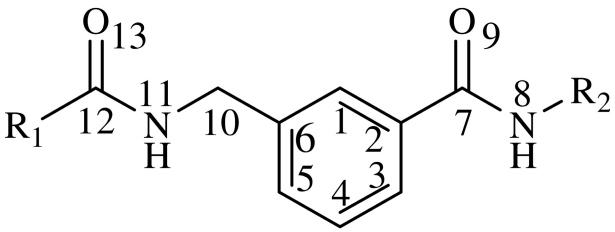



**Figure 1.** Representation of atoms used in different alignments.

### 2.3 Alignment and Interaction Pharmacophore Elements Definition

The alignments are chosen to span the common framework of the molecules. To guarantee a full analysis, the atoms to the right, left and middle of the common part of the compounds should be used in alignment.[30] Six alignments were performed using the atoms C1, N8 and C12 in different orders (ali1: 12, 1, 8; ali2: 1, 8, 12; ali3: 8, 12, 1; ali4: 12, 8, 1; ali5: 1, 12, 8; ali6: 8, 1, 12) (Figure 1). Then, each conformation from the CEP was placed in a reference cubic cell space according to the alignment under consideration. The cubic lattice serves to record the spatial occupancy distribution of each atom of the ligands. The grid cell size was defined to be 1 Å.

The atoms of each compound were classified into seven types of interaction pharmacophore elements (IPE) corresponding to the possible types of interactions that occur in the cell box in accordance with the 4D-QSAR methodology.[11] The IPEs are classified into: any type (any); nonpolar (np); positively charged polar (p+); negatively charged polar (p-); hydrogen bond acceptors (ha); hydrogen bond donors (hd) and aromatic systems (ar). The grid cell occupancy profiles for each of the chosen IPEs were computed and used as the basis set for the trial cell occupation descriptors (GCOD, Grid Cell Occupancy Descriptor) that will be used in the construction of QSAR models.

### 2.4 Obtaining the 4D-QSAR Regression Models

The matrix of GCODs was filtered to exclude variables with zero variance as these do not provide any relevant information. Additionally, unsupervised and supervised feature selection procedures were performed over the reduced GCOD matrix using Shannon's Entropy (SE) and Differential Shannon's Entropy (DSE) parameters, respectively, and 400 variables were finally retained. For this dimensionality reduction procedure, the IMMAN software was used.[31]

Posteriorly, ten 4D-QSAR models for each alignment were built using the Stochastic Gradient Boosting (SGB) method, based on regression trees (RT) base learners. The RTs are obtained by growing the largest trees via greedy recursive partitioning and then pruning the grown tree to yield the final appropriate fit.[32] The RTs possess many attractive features, including simplicity, interpretability, capacity in handling large data sets and modeling nonlinearities, as well as immunity to outliers. Moreover, the RTs do not entail any assumption on the data distribution, collinearity and heteroscedasticity.[33] However, it is known that RTs are highly instable, that is, a small variation in training data leads to a huge change in the computed results.[34] In order to correct this failure, SGB is applied. In this approach, a fraction $\eta$ of the training data is sampled at random (without replacement) from the full training dataset at each iteration. These randomly selected sub samples are used in place of the full samples to fit the base learner and compute the model update for current iteration.[35] In this sense, a sequence of very simple RTs is obtained, where each successive tree is built to predict the residuals of the preceding tree. Then the average prediction over many trees is computed.[36] This implementation not only reduces

the computing time, but also actually yields more accurate models.[35]

The SGB based 4D-QSAR models were built according to the configurations: the number of initial additive terms was set at 400, the learning rate at 0.15, the subsample proportion at 0.50 and the random test data proportion at 0.30. In addition, an early stopping criterion was adopted based on the following parameters: the minimum number of cases was set at 5, the maximum number of levels at 10, the minimum number in child node at 1 and the maximum number of nodes at 3. The Mean Absolute Error for the training and test sets was used to assess the predictive ability of the models. The SGB based regression model was developed using the Statistic Data Mining software, version 8.0.[37]

With the goal of comparing the results obtained using SGB with other statistical model building methods, ten Partial Least Squares (PLS) models were built for each alignment using the filtered GCOD matrix. The PLS is a statistical modeling technique based on the transformation of the original dataset matrix into orthogonal projections, also known as latent variables (LVs). The LVs are linear combinations of the original variables and are aimed at maximizing the covariance between the **X** matrix and the response variable.[38] The PLS-based model was validated using the leave-one-out cross ($LOO_{CV}$) validation and external validation ($r^2ext$) procedures, respectively. Additionally, the model was assessed for possible fortuitous correlation using the Y-randomization procedure, where the following parameters were considered: the determination coefficient for y-randomization, penalized determination coefficient and root-mean-square-error of randomization, denoted by $r^2$y-rand, $R^2$p(y-rand) and RMSEy-rand, respectively. Other parameters that were considered to assess the quality of the PLS-based model included the root-mean-square-error of cross-validation ($RMSE_{CV}$) and prediction ($RMSE_P$), respectively. The PLS modeling was performed using the Chemoface program.[39]

One of the key advantages offered by the 4D-QSAR strategy is the interpretability of the GCOD variables, in terms of the chemical environment and its relationship to the modeled property. In sense, in order to interpret the built QSAR model, a subset of the most relevant descriptors was selected and analyzed. Each descriptor represents a particular IPE and therefore from an examination of the chemical information encoded in the most relevant GCODs, it is possible to identify key elements essential for the biological response of the analyzed compounds and, eventually, to propose novel derivatives with an improved biological profile. Consequently, novel benzamide derivatives were proposed following classical medicinal chemistry principles such as bioisosterism and molecular hybridization.

### 2.5 Molecular Docking

The molecular docking methodology aims to predict the ligand binding modes for biomolecular target sites of therapeutic interest based on the hypothesis that a lower energy score represents a better protein-ligand binding pose, thereby helping to elucidate the main elements that determine the protein-ligand interactions.[40] Ideally, molecules predicted to have high activity should demonstrate favorable protein-ligand interaction profiles. Therefore, in order to evaluate the binding affinity of proposed molecules relative to those experimentally determined to possess high inhibitory activity, as a means of validating the results obtained with the 4D-QSAR regression model, a docking experiment was performed using the ROCK2 target protein.

The molecular docking was performed using the Molegro Virtual Docker (MVD) software.[41] The X-ray crystallographic structure of the ROCK2 enzyme with resolution of 2.93 Å, in complex with methyl 3-[({2′-(aminomethyl)-5′-[(3-fluoropyridin- 4-yl)carbamoyl] biphenyl-3-yl}carbonyl) amino]- 4-fluorobenzoate was obtained from the *Protein Data Bank* (PDB ID code: 4WOT).[42] Prior to docking, geometry optimization and partial charge distribution calculations of the compounds were carried out using semi-empirical AM1 method. Only the ligands were considered flexible during the docking simulations. For each ligand, 100 independent docking runs were performed using a guided differential evolution algorithm, wherein each run returns to one solution (pose). Default parameters were used, except for the population size which was increased from 50 to 100. The MolDock score was employed as the scoring function with a grid resolution of 0.30 Å. MolDock is an empirical score function derived from the piecewise linear potential (PLP) scoring function proposed by Gehlhaar et al.[43,44] and afterwards extended by Yang et al.[45] MolDock further improves these functions by considering a new hydrogen bonding term and charge schemes. The $E_{PLP}$ function takes into account steric interactions and hydrogen bonds. In order to increase docking accuracy, the results were re-ranked using a more complex score function, which considers additional terms for $sp^2$-$sp^2$ torsion and the Lennard-Jones 12–6 potential.[41] The best poses were selected from the lowest MolDock re-ranking score.

## 3 Results and Discussions

### 3.1 Model Analysis

#### *3.1.1 Boosting Tree Model*

The model was built using the ratio $pIC_{50}/pLD_{50}$ as response variable and the 4D-QSAR descriptors as the independent variables. Based on the reduction in average squared error over the test set, the optimum number of trees for the

ensemble was determined to be 76 for an average squared error (ASE) values of 0.229 and 1.667 for the training and test sets, respectively (see Figure 2). The selected descriptors along with their importance are given in Table S2 (Supplementary Information).

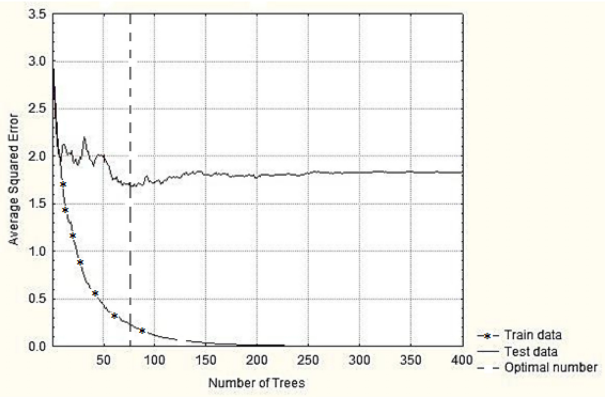



**Figure 2.** The optimal number of trees.

Table 1 shows the statistical parameters of the SGB model. As can be observed, this model yields satisfactory performance with correlation coefficients of 0.956 and 0.685 for the training and test sets, respectively. Additionally, the SGB model yielded RMSE values for calibration and external validation of 0.4784 and 0.7662, respectively. These residual values may be considered as very low. Figure 3 shows the correlation between experimental and predicted values for the training and test set for the boosted tree model. As can be observed there exists a good correspondence between the experimental and predicted $pIC_{50}/pLD_{50}$ values.

### *3.1.2 PLS*

The 81 compounds of Table S1 (chemical structures provided as Supplementary Information) were grouped in a training set consisting of 57 compounds and a prediction set comprised of the remaining 24 compounds. The leave-one-out cross validation procedure was applied to the training set, in order to determine the optimal number of latent variables to be used in the PLS regression. Consequently, the number of optimal latent variables was determined to be 4 as it corresponded to a minimum in the RMSEcv plot (see supporting information Figure S1). Table 1 shows the statistical parameters of the PLS-based model. As

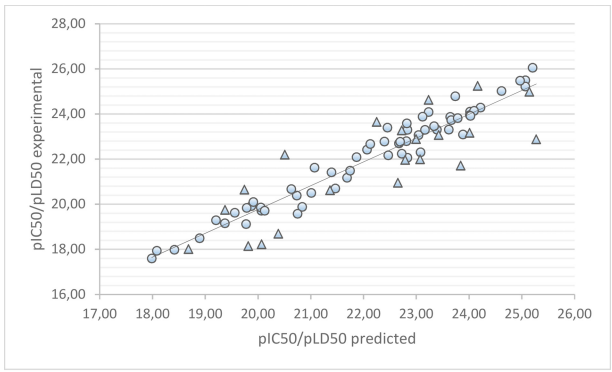



**Figure 3.** Correlation of the experimental vs predicted $pIC_{50}/pLD_{50}$ values based on the Boosted tree model [spheres – training set, triangles – test set].

can be observed, the obtained statistical parameters are generally unsatisfactory according to established criteria for evaluating the predictive power and robustness of QSAR models, i.e. the correlation coefficients for calibration ($r^2$) and $LOO_{CV}$ were 0.557 (recommended $r^2 \geq 0.6$[46]) and 0.306 (recommended $q^2 \geq 0.5$), respectively; external validation parameter $r^2test = 0.311$ (recommended $r^2test \geq 0.5$) and $R^2p(y\text{-rand}) = 0.266$ [recommended $R^2p(y\text{-rand}) \geq 0.5$ . It may therefore be concluded that PLS, as the classical statistical technique for modeling high dimensionality data is not suitable for the present study and the modeled response variable does not lend itself to a linear relationship with the GCOD descriptors.

In light of the greater robustness and predictive power obtained with the SGB model relative to the PLS model, the former may thus be considered as more suitable for virtual screening of novel benzamide derivatives.

### 3.2 Structural Interpretation of the GCOD Model

The aim of the present subsection is to interpret the variables contained in the SGB model to gain greater understanding on the structural characteristics that favor an increase in potency of the ROCK2 inhibitors and at the same time lowering the toxicity. However, the matrix of the 400 GCOD descriptors employed to build the boosted tree model was quite large for a tractable interpretation. Therefore, this subset of molecular descriptors was ordered according to the importance scores, ranging from 0.3 to 1.0, and was further refined to filter out descriptors that codified

**Table 1.** Summary of parameters obtained for PLS and boosted tree models.

| Model | Size | $r^2$ | $q^2loo$ | RMSEc | $r^2pred$ | $R^2p(y\text{-rand})$ | RMSEp |
|---|---|---|---|---|---|---|---|
| PLS | 4 LV | 0.557 | 0.306 | 1.383 | 0.311 | 0.266 | 1.322 |
| SBG | 76 Trees | 0.956 | – | 0.478 | 0.654 | – | 0.766 |

similar information (i.e. highly correlated and with a similar importance score). For the importance scores of the 400 GCOD descriptors, see Table S2 in Supplementary Information. Consequently, 9 variables considered as the most representative ofthe SGB model were retained and employed for the structural interpretation. Figure 4 shows the positions of the retained descriptors using the compound 70, which has the highest biological response in the training set. In 4D-QSAR, descriptors are represented by Cartesian coordinates and interaction pharmacophore elements, "(x, y, z, IPE)", which corresponds to the interactions that may occur between ligand-protein.

The GCODs 1, 2, 10, 3, 25 and 26 are located around the R1 substituent, where 1, 2 and 10 are coordinates over the phenyl group and 3, 25 and 26 over the imidazole moiety. On the other hand, GCODs 8, 21 and 22 are located around the R2 substituent. The descriptor **1** (1,–2,–3,np) represents a non-polar interaction and possesses the highest importance score (1.000, Figure 4) in the data set. This GCOD has greater occupation frequency for molecule 24 (Figure 5) and is located adjacent to the indole carbon, indicating that non-polar interactions in this portion are relevant to the activity of these compounds. In fact, if a comparison is performed for the molecules 22 and 24 which only differ in this coordinate, in that 22 has a nitrogen atom in place of a carbon, the biological response for 24 is observed to be greater than that of 22 by approximately two logarithm units.

The descriptor **2** (−1,−3,0,any) represents any type of interaction and has greater occupation frequency for molecules 1 and 3 (Figure 5). This GCOD is located on a phenyl hydrogen atom, suggesting that non-polar groups in this region may increase the potency of these compounds. The GCOD **10** (−1,0,−3,any) also represents a non-specific interactions. This descriptor has greater occupation frequency in compound 58 (Figure 5) and is located adjacent to the methyl linked to triazole group thus, indicating that non-polar substituents in this position are beneficial as well.

The GCOD **25** (−1,−2,−6,any) has greater occupation for molecule 19 (Figure 5) and represents any type of interaction. This descriptor is located on the pyridine C–O bond, underscoring the importance of different interactions in this region. The descriptor **26** (2,−3,−5,any) also

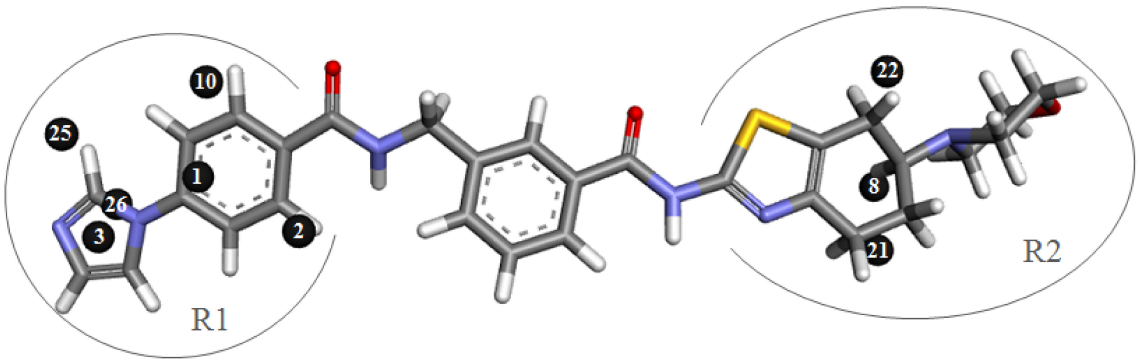



**Figure 4.** The most important descriptors represented in the compound 70. 1: (1,−2,−3,np)[1.000]; 2: (−1,−3,0,any)[0.963]; 3: (1,−4,−5,any)[0.932]; 8: (3,2,14,any)[0.811]; 10: (−1,0,−3,any)[0.801]; 21: (1,0,14,np)[0.740]; 22: (−1,4,13,np)[0.737]; 25: (−1,−2,−6,any)[0.717].

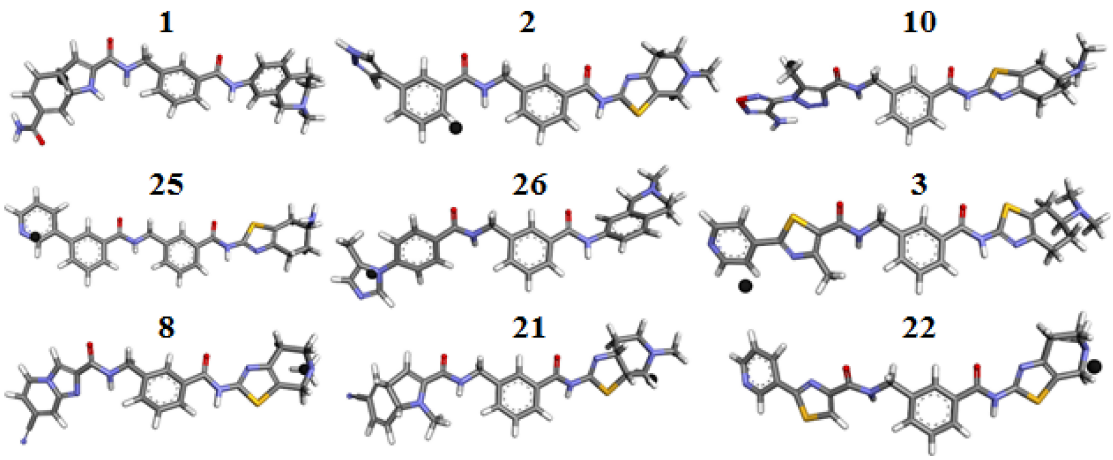



**Figure 5.** The nine most important descriptors are represented in their respective molecules with the highest occupation frequency.

represents non-specific interaction types, its occupation frequency values are low with the largest being for the compound 76 (Figure 5) where it is positioned on the imidazole nitrogen atom. The descriptor **3** (1,−4,−5,any) represents interactions of any type and the greatest occupation frequency value is for compound 79 (Figure 5). However, because of the system motion during MDS, the position of this descriptor varied widely, not allowing for a clear interpretation of the information codified.

Around the R2 substituent are the descriptors 8, 21 and 22 (see Figure 4). The descriptor **8** (3,2,14, any) has a non-specific IPE and higher occupation frequency for the molecules 28, 30 and 34 (Figure 5). These molecules have the same R2 substituent and the descriptor is located over the methyl group attached to the piperidine. The GCOD **21** (1,0,14, np) codifies information on non-polar interactions and is located near the hydrogen atom of the piperidine group of the compound 48 (Fig. 5), which has the highest frequency of occupation for this descriptor. Similar to GCOD 21, the descriptor **22** (-1,4,13, np) is related with non-polar IPE and is located near the hydrogen piperidine group. Its highest occupation frequency value is for compound 6 (Fig. 5), suggesting that hydrophobic interactions in this region favor the bioactivity of these compounds. Figure 5 shows all descriptors represented in their respective molecules with the highest occupation frequency.

### 3.3 Proposal of Novel Benzamide Derivatives and Drug-likeness Filtering

Analyzing the results obtained from the 4D-QSAR study, 139 new structures were proposed with the aim of yielding novel benzamide derivatives with a greater $pIC_{50}/pLD_{50}$ ratio. The structures were constructed using molecules **70** and **79** as a base, which are the most active of the series and with the highest predicted value of ratio $pIC_{50}/pLD_{50}$, i.e. 26.058 and 25.273, respectively. These proposed chemical structures were screened using the boosting model and their $pIC_{50}/pLD_{50}$ values predicted. From these proposals, the molecular structures p1, p3, p6, p90, p130, p131, p133, p134, and p135 were predicted to possess $pIC_{50}/pLD_{50}$ ratios greater than that predicted for the compound **79**, while the structures p14 and p96 showed a predicted ratio $pIC_{50}/pLD_{50}$ similar to the value predicted of compound **79** (see Table 2). For the design of the p1 molecule the compound **70** was used. As can be observed in Table 2, the addition of the methyl group at GCOD2 position resulted in an increase of the $pIC_{50}/pLD_{50}$ ratio for the compound from 26.058 to 26.541. For the proposed structure p133, naphthalene was used in R1 position and the R2 substituent from the compound 79, see Table 2. Furthermore, the predictions for p1 and p133 are superior to the experimental value of the compound **70**, which is the most active compound of the studied series.

Posteriorly, the ten proposals predicted to possess the best $pIC_{50}/pLD_{50}$ ratios were evaluated using the "Lipinski rule of five"[48] for drug likeness using the Molinspiration Online Property Calculation Software Toolkit (http://www.molinspiration.com/). This rule helps to define the limits on physicochemical and structural properties of chemicals to screen for molecules with satisfactory oral absorption. These boundaries are defined to be: milog$P \leq 5$, molecular weight (MW) $\leq 500$ u, number of hydrogen bond acceptors ($n_{ON}$) $\leq 10$, and number of hydrogen bond donors ($n_{OHNH}$) $\leq 5$ and polar surface area (TPSA) no greater than 140 Å$^2$. The "Lipinski rule of five" considers that poor absorption is more likely when two or more violations are observed.[49] The results for the proposed compounds are encouraging because almost all proposed molecules satisfy the conditions stipulated by Lipinski (see Table 3), suggesting no absorption problems. Only the p3 compound commits two violations, for this reason it is not considered in posterior evaluations.

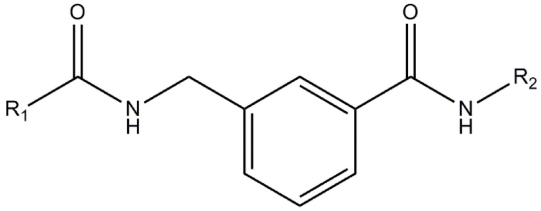
### 3.4 Molecular Docking

Molecular docking can be helpful in DDD by evaluating the binding affinity of the most promising hits. The training set compounds 79 and 64, which have the highest and lowest biological response, respectively, along with the new proposed compounds were docked inside de active site of ROCK2 (PDB ID: 4WOT). These training set compounds were selected for the purpose of comparison. Prior to performing the actual docking simulation, the ligand binding process was validated using the "redocking" procedure of the co-crystallized ligand in the ROCK2. Figure 6 illustrates the superposition of the crystal and docked structures, and it can be observed that the main interactions are preserved. The Root Mean Square Deviation (RMSD) between the crystal and docked structures was 1.06 Å, reinforcing the validity of the docking procedure. The key amino acid residues in the active site identified by MVD simulation were Met172, Phe103, Ala102, Lys121, Asp232 and Thr235.

Table 4 shows the energy values obtained from the docking simulation. As can be observed, the interaction energy values ($E_{INTER}$, kcal mol$^{-1}$) decreased in the following order: p1, p90, p14, 79, p133, p134, p131, p6, p130, p135 and 64. In this regard, the proposed molecules p1, p90 and p14 exhibit more energetically favorable interactions (more negative values) than the most active compound of the training set (79), but similar to the co-crystallized ligand (−222.856 kJ mol$^{-1}$). Moreover, the calculated binding affinity values for p6, p14 and p131 were also slightly more energetically favorable than compound 79, which in turn is similar to p1. In addition, all proposed compounds possess better interaction energies than the least active training set compound (64). These outcomes are in agreement with those obtained by the SGB based model, since these three compounds (p1, p90 and p14) were predicted to have the

**Table 2.** Structures of the best proposed molecules along with compounds 70 and 79.



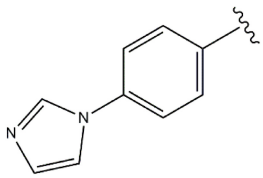
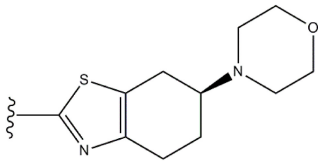
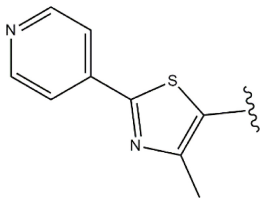
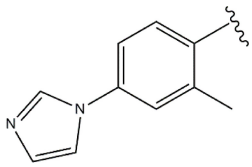
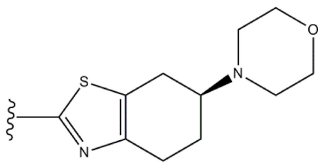
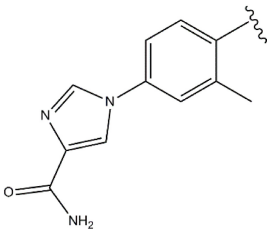
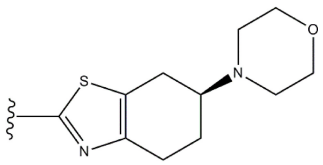
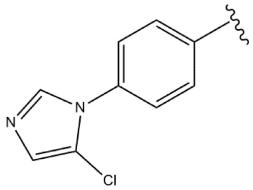
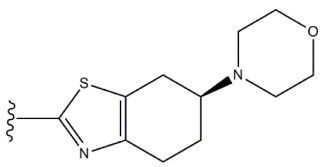
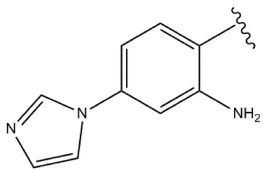
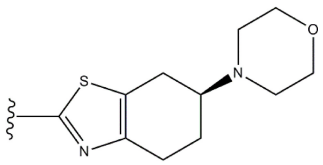
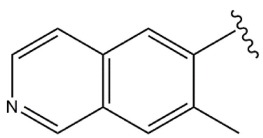
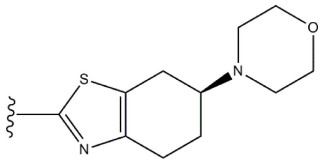
| Molecule | R1 | R2 | $pIC_{50}/pLD_{50}$ predict |
|---|---|---|---|
| 70 |  |  | 26.058 |
| 79 |  |  | 25.273 |
| p1 |  |  | 26.541 |
| p3 |  |  | 25.300 |
| p6 |  |  | 25.316 |
| p14 |  |  | 25.260 |
| p90 |  |  | 25.570 |

**Table 2.** continued

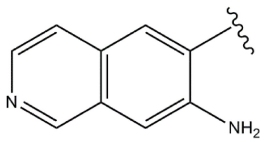
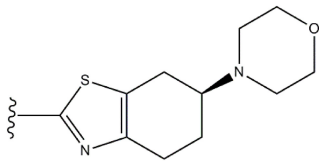
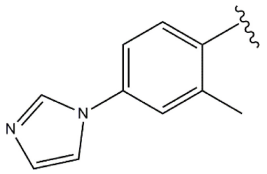
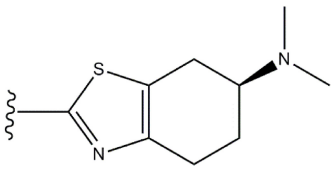
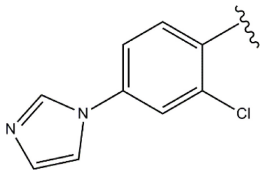
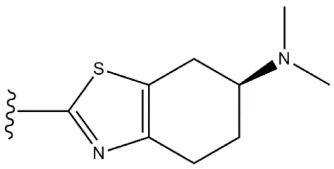
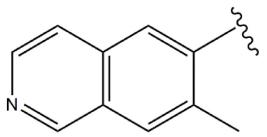
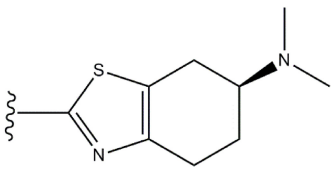
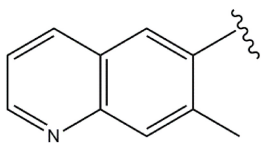
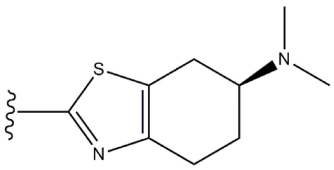
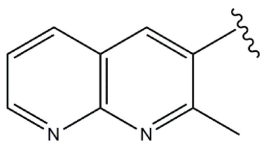
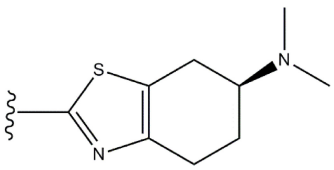
| Molecule | R1 | R2 | $pIC_{50}/pLD_{50}$ predict |
|---|---|---|---|
| p96 |  |  | 25.215 |
| p130 |  |  | 25.883 |
| p131 |  |  | 25.414 |
| p133 |  |  | 26.109 |
| p134 |  |  | 25.285 |
| p135 |  |  | 25.597 |

**Table 3.** Calculated parameters of the "Lipinski rule of five" for the proposed and 70 and 79 compounds.

| Molecule | miLogP | TPSA | MW | nON | nOHNH | nviolations |
|---|---|---|---|---|---|---|
| 70 | 2.61 | 101.39 | 542.66 | 9 | 2 | 1 |
| 79 | 5.14 | 100.11 | 608.79 | 8 | 2 | 2 |
| p1 | 3.20 | 101.39 | 556.69 | 9 | 2 | 1 |
| p3 | 2.26 | 144.48 | 599.72 | 11 | 4 | 2 |
| p6 | 3.43 | 101.39 | 577.11 | 9 | 2 | 1 |
| p14 | 2.24 | 127.41 | 557.68 | 10 | 4 | 1 |
| p90 | 3.64 | 96.45 | 541.68 | 8 | 2 | 1 |
| p130 | 3.35 | 92.15 | 514.65 | 8 | 2 | 1 |
| p131 | 3.58 | 92.15 | 535.07 | 8 | 2 | 1 |
| p133 | 3.90 | 87.22 | 499.64 | 7 | 2 | 0 |
| p134 | 3.79 | 87.22 | 499.64 | 7 | 2 | 0 |
| p135 | 3.21 | 100.11 | 500.63 | 8 | 2 | 1 |

highest $pIC_{50}/pLD_{50}$ ratio values. Conversely, the molecules p6, p130 and p135 have much lower interaction energy, although they were predicted to have high $pIC_{50}/pLD_{50}$ values, and thus suggesting that they may probably have

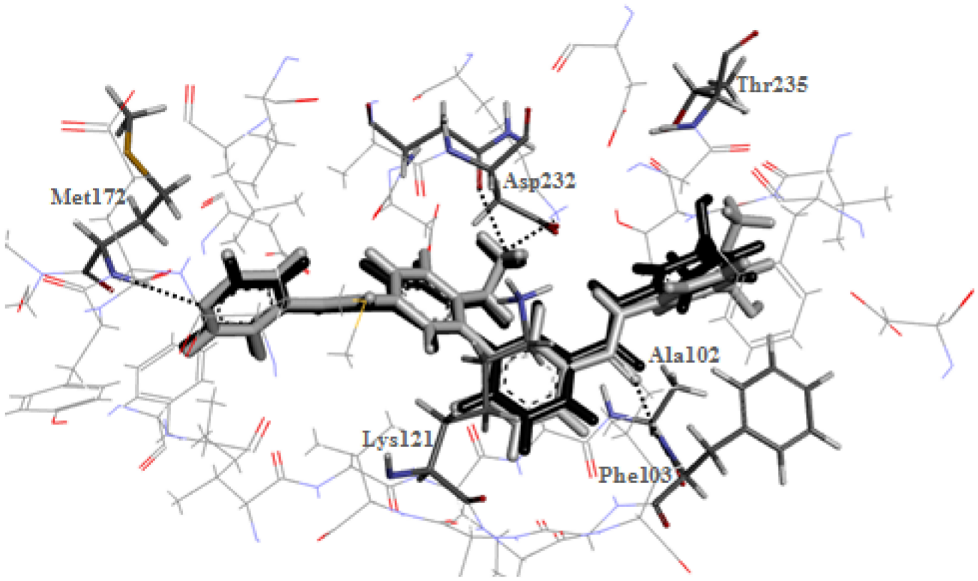



**Figure 6.** Redocking analysis, superposition of crystallized (black) and docked (gray) structures.

**Table 4.** Energy values from molecular docking for the analyzed compounds. MolDock and Rerank Score (a.u), interaction energy ($E_{INTER}$, kcal $mol^{-1}$), hydrogen bonding (HBond, kcal $mol^{-1}$) and binding affinity (kJ $mol^{-1}$).

| Compounds | $pIC_{50}/pLD_{50}$ | MolDock Score | Rerank Score | $E_{INTER}$ | HBond | Binding Affinity* |
|---|---|---|---|---|---|---|
| ligand | – | −197.178 | −169.098 | −222.856 | −8.699 | −35.586 |
| p1 | 26.541 | −209.775 | −163.016 | −221.824 | −3.563 | −29.002 |
| p133 | 26.109 | −188.346 | −164.917 | −209.076 | −5.298 | −29.009 |
| p130 | 25.883 | −175.358 | −134.372 | −188.061 | −3.926 | −27.737 |
| p135 | 25.597 | −165.142 | −145.431 | −187.510 | −2.440 | −28.186 |
| p90 | 25.570 | −192.363 | −153.471 | −214.612 | −3.096 | −30.849 |
| p131 | 25.414 | −189.463 | −159.147 | −201.545 | −7.057 | −31.814 |
| p6 | 25.316 | −169.666 | −131.769 | −192.954 | −6.021 | −34.754 |
| p134 | 25.285 | −185.464 | −160.410 | −203.885 | −2.654 | −27.712 |
| p14 | 25.260 | −186.066 | −157.677 | −213.429 | −6.043 | −32.410 |
| 79 | 25.273 | −197.147 | −163.781 | −210.867 | −3.897 | −30.190 |
| 64 | 17.601 | −169.998 | −143.251 | −185.721 | −2.085 | −24.386 |

* In MVD, the binding affinities are estimated using a multiple linear regression model composed by energy terms and descriptors, and which was calibrated using diverse complexes with known binding affinities.

extremely low $pLD_{50}$ values and thus compensating for their lower activity. Figure 7 illustrates the interaction profile for compound p1 (most negative MolDock score) inside the ROCK2 active site along with the main amino acid residues that interact through hydrogen bond.

As described above, in this work the ligand poses were selected based on MolDock reranking score and similarity with the ligand in the active site. The compounds adopted a geometry configuration such that the R1 substituent is located at the bottom of the binding site and the R2 substituent in the solvent-exposed region. Just like the co-crystallized ligand, most of the docked compounds present a hydrogen bond interaction with the Met172 residue, which between through the cyclic nitrogen atom and the Met172 amino group. The p90, p134 and p135 compounds do not present this interaction because the positioning of the corresponding nitrogen atoms does not favor this interaction. Additionally, p14 adopts a conformation that does not allow this hydrogen bond interaction; instead it interacts with Asp385 and forms an additional hydrogen bond with Arg100 through amino group. Another important hydrogen bond interaction is between the Phe103 and the carbonyl group (near to R2), except for compound 79 that instead interacts with Lys121. Lys121 also forms hydrogen interaction with p130 and p131, but in this case with the carbonyl group near the R1 substituent. The compounds p6, p131, p134 and p135 form a weak hydrogen bond with Ala102. On the other hand, Asp232 interacts with the amino group of the compounds p130, p133 and p135. The R2 cyclic nitrogen for p1 and p6 hydrogen bonds with

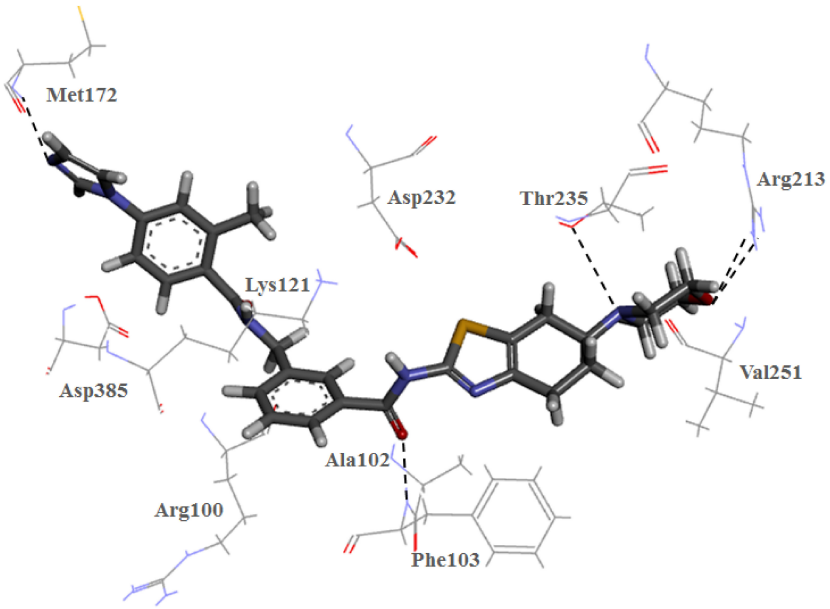



**Figure 7.** Docking structure of the p1 in the active site of ROCK2, the key amino acid residues and hydrogen bond interactions are depicted.

Thr235. Lastly, the morpholine group of p1, p6, p14 and p90 interacts with Arg213, and p6 forms an additional hydrogen bond with Val251.

Altogether, the docking simulation performed in the present section shows that the proposed molecules predicted by the SGB model to possess higher $pIC_{50}/pLD_{50}$ ratios generally exhibit the most favorable protein-ligand interaction profiles, evidenced by the lower interaction energy and binding affinity scores obtained. Although the modeled property is indeed a ratio where the $pIC_{50}$ is sought to be maximized while minimizing the $pLD_{50}$, the docking results generally corroborate the hypothesis that proposed molecules may in fact possess high inhibitory activity (i.e. high $pIC_{50}$ values) for which appropriate protein-ligand interactions are indispensable.

## 4 Conclusion

In the present study, we have demonstrated utility of a consensus molecular modeling scheme, coupling 4D-QSAR and molecular docking methods, in the multi-objective optimization of benzamide derivatives. To this end, a 4D-QSAR regression model was built using the SGB method, wherein the $pIC_{50}/pLD_{50}$ ratio was correlated to the interaction descriptors. An analysis of the most relevant variables in this model permitted the proposal of 139 candidates, 9 of which yielded better $pIC_{50}/pLD_{50}$ ratios when compared with the employed dataset. These compounds were evaluated for drug likeness using the Lipinski's Rule and were found to be within the established limits. Through the docking simulation, the binding affinity of proposed molecules was estimated and the chemical structures p6, p14 and p131 showed higher binding affinity for the ROCK2 active site than the most active in the training set, while the rest generally demonstrated comparable behavior. While the ultimate evidence of bioactivity and low toxicity requires *in vitro* and *in vivo* studies, the *in silico* ligand and structure-based approaches employed in the present study provide useful information to guide the design of novel compounds with possible better affinity to the ROCK2 target and low toxicity profiles.

## Conflict of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper .

## Acknowledgment

We thank the Brazilian agencies "Conselho Nacional de Desenvolvimento Científico e Tecnológico" (CNPq) and "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES) for their financial support.

## References

[1] K. Riento, A. J. Ridley, *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 446–456.
[2] M. Dong, B. P. Yan, J. K. Liao, Y.-Y. Lam, G. W. K. Yip, C.-M. Yu, *Drug Discov. Today* **2010**, *15*, 622–629.
[3] J. K. Liao, M. Seto, K. Noma, *J. Cardiovasc. Pharmacol.* **2009**, *50*, 17–24.

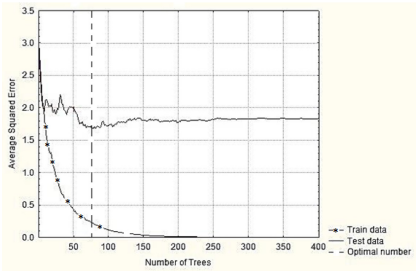
  

[4] P. Pan, M. Shen, H. Yu, Y. Li, D. Li, T. Hou, *Drug Discov. Today* **2013**, *18*, 1323–1333.
[5] P. Y. Mong, Q. Wang, *J. Immunol.* **2009**, *182*, 2385–94.
[6] M. J. Waring, J. Arrowsmith, A. R. Leach, P. D. Leeson, S. Mandrell, R. M. Owen, G. Pairaudeau, W. D. Pennie, S. D. Pickett, J. Wang, O. Wallace, A. Weir, *Nat. Rev. Drug Discov.* **2015**, *14*, 475–486.
[7] R. Guttendorf, *Bioanalysis* **2012**, *4*, 1395–1397.
[8] S. Adler, D. Basketter, S. Creton, O. Pelkonen, J. van Benthem, V. Zuang, K. E. Andersen, A. Angers-Loustau, A. Aptula, A. Bal-Price, E. Benfenati, U. Bernauer, J. Bessems, F. Y. Bois, A. Boobis, E. Brandon, S. Bremer, T. Broschard, S. Casati, S. Coecke, R. Corvi, M. Cronin, G. Daston, W. Dekant, S. Felter, E. Grignard, U. Gundert-Remy, T. Heinonen, I. Kimber, J. Kleinjans, H. Komulainen, R. Kreiling, J. Kreysa, S. B. Leite, G. Loizou, G. Maxwell, P. Mazzatorta, S. Munn, S. Pfuhler, P. Phrakonkham, A. Piersma, A. Poth, P. Prieto, G. Repetto, V. Rogiers, G. Schoeters, M. Schwarz, R. Serafimova, H. Tähti, E. Testai, J. van Delft, H. van Loveren, M. Vinken, A. Worth, J.-M. Zaldivar, *Arch. Toxicol.* **2011**, *85*, 367–485.
[9] OECD – Organization for Economic Co-operation and Development, Quantitative Structure-Activity Relationships Project [(Q) SARs], **2010**.
[10] J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav, C. Duraiswami, *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524.
[11] C. H. Andrade, K. F. M. Pasqualoto, E. I. Ferreira, A. J. Hopfinger, *4D-QSAR: Perspectives in Drug Design, Molecules.* **2010**, *15*, 3281–3294.
[12] T. M. de Assis, G. C. Gajo, L. C. de Assis, L. Santos-Garcia, D. R. Silva, T. C. Ramalho, E. F. F. da Cunha, *Chem. Biol. Drug Des.* **2015**, 1–12.
[13] L. Santos-Garcia, L. C. Assis, D. R. Silva, T. C. Ramalho, E. F. F. da Cunha, *J. Biomol. Struct. Dyn.* **2015**.
[14] D. R. Silva, T. C. Ramalho, E. F. F. da Cunha, *Lett. Drug Des. Discovery* **2014**, *11*, 649–664.
[15] S. Harmsen, R. J. Kok, *Curr. Pharm. Des.* **2012**, *18*, 2891–900.
[16] S.-A. B. Ltda, *Ficha de informações de segurança de produtos químicos*, Sigma-Y0503 Y-27632 Dicloridrato Monohidratado **2016**, pp. 1–7.
[17] M. Itagaki, A. Takaguri, S. Kano, S. Kaneta, K. Ichihara, K. Satoh, *J. Pharmacol. Sci.* **2009**, *109*, 94–101.
[18] J. H. Friedman, *Comput. Stat. Data Anal.* **2002**, *38*, 367–378.
[19] D. Rogers, A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **1994**, *34*,854–866.
[20] B. N. Cook, J. A. Kowalski, X. Li, D. R. Marshall, S. Shlyer, R. Sibley, L. L. Smith-Keenan, F. Soleymanzadeh, R. J. Sorcek, E. R. R. Young, Y. Zhang, WO_2012006203_A1.pdf, PCT/US2011/042508, **2012**.
[21] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187–217.
[22] A. S. INC, Discovery studio Modeling Environment, (**2007**).
[23] S. Lapenna, M. Fuart-gatnik, A. Worth, Review of QSAR Models and Software Tools for predicting Acute and Chronic Systemic Toxicity, **2010**, doi:10.2788/60766.
[24] A. Tropsha, A. Golbraikh, *Curr. Pharm. Des.* **2007**, *13*, 3494–3504.
[25] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'Min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, *J. Med. Chem.* **2014**, *57*, 4977–5010.
[26] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701–1718.
[27] N. L. Allinger, *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134.
[28] T. chem21 group INC, 4D-QSAR USER'S MANUAL, **1997**.
[29] G. B. Caldas, T. C. Ramalho, E. F. F. da Cunha, *J. Mol. Model.* **2014**, *20*.
[30] R. W. P. Urias, S. J. Barigye, Y. Marrero-Ponce, C. R. García-Jacas, J. R. Valdes-Martiní, F. Perez-Gimenez, *Mol. Diversity* **2015**, *19*, 305–319.
[31] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, *Classification And Regression Trees*, Wadsworth, **1984**.
[32] S. M. Tan, J. Jiao, X. L. Zhu, Y. P. Zhou, D. D. Song, H. Gong, R. Q. Yu, *Chemom. Intell. Lab. Syst.* **2010**, *103*,184–190.
[33] V. Svetnik, T. Wang, C. Tong, A. Liaw, R. P. Sheridan, Q. Song, *J. Chem. Inf. Model.* **2005**, *45*, 786–799.
[34] D. S. Cao, Q. S. Xu, Y. Z. Liang, L. X. Zhang, H. D. Li, *Chemom. Intell. Lab. Syst.* **2010**, *100*, 1–11.
[35] J. Jiao, S.-M. Tan, R.-M. Luo, Y.-P. Zhou, *J. Chem. Inf. Model.* **2011**, *51*, 816–28.
[36] I. StatSoft, STATISTICA (data analysis software system), **2007**. www.statsoft.com.
[37] E. Deconinck, M. H. Zhang, D. Coomans, Y. Vander Heyden, Evaluation of boosted regression trees (BRTs) and two-step BRT procedures to model and predict blood – brain barrier passage, **2007**, 280–291, doi:10.1002/cem.
[38] C. A. Nunes, M. P. Freitas, A. C. M. Pinheiro, S. C. Bastos, *J. Braz. Chem. Soc.* **2012**, *23*, 2003–2010.
[39] I. A. Guedes, C. S. de Magalhães, L. E. Dardenne, *Biophys. Rev.* **2014**, *6*, 75–87.
[40] R. Thomsen, M. H. Christensen, *J. Med. Chem.* **2006**, *49*, 3315–3321.
[41] S. Boland, A. Bourin, J. Alen, J. Geraets, P. Schroeders, K. Castermans, N. Kindt, N. Boumans, L. Panitti, S. Fransen, J. Vanormelingen, J. M. Stassen, D. Leysen, O. Defert, *J. Med. Chem.* **2015**, *58*, 4309–4324.
[42] D. K. Gehlhaar, G. Verkhivker, P. A. Rejto, D. B. Fogel, L. J. Fogel, S. T. Freer, Docking Conformationally Flexible Small Molecules into a Protein Binding Site through Evolutionary Programming, in *Proc. Fourth Int. Conf. Evol. Program*, **1995**, pp. 615–627.
[43] D. K. Gehlhaar, D. Bouzida, P. A. Rejto, Fully Automated and Rapid Flexible Docking of Inhibitors Covalently Bound to Serine Proteases, in *Proc. Seventh Int. Conf. Evol. Program*, **1998**, pp. 449–461.
[44] J.-M. Yang, C.-C. Chen, *Proteins Struct. Funct. Bioinf.* **2004**, *55*, 288–304.
[45] R. Kiralj, M. M. C. Ferreira, *J. Braz. Chem. Soc.* **2009**, *20*, 770–787.
[46] P. P. Roy, S. Paul, I. Mitra, K. Roy, Erratum: On two novel parameters for validation of predictive QSAR models (Molecules (2009) 14 (1660-1701)), Molecules. 15 (2010) 604–605, doi:10.3390/molecules15010604.
[47] C. a. Lipinski, *Drug Discov. Today Technol.* **2004**, *1*, 337–341.
[48] E. Deconinck, T. Hancock, D. Coomans, D. L. Massart, Y. Vander Heyden, *J. Pharm. Biomed. Anal.* **2005**, *39*, 91–103.

Received: June 8, 2017
Accepted: August 16, 2017
Published online on ■■ ▮▮, 0000

# FULL PAPER




*G. Cardoso Gajo, D. Rodrigues Silva, S. J. Barigye, E. F. F. da Cunha**