

4D-QSAR Study of 17 β -Hydroxysteroid Dehydrogenase Type 3 InhibitorsJoão Paulo A. Martins^{✉*,a} and Eduardo B. de Melo^b^aDepartamento de Química, Universidade Federal de Minas Gerais,
31270-901 Belo Horizonte-MG, Brazil^bLaboratório de Química Medicinal e Ambiental Teórica,
Universidade Estadual do Oeste do Paraná, 85819-110 Cascavel-PR, Brazil

Prostate cancer is one of the most common form of cancer among men, and its incidence has been increasing progressively. Since the prostate depends on androgenic hormones to regulate its growth and development, interference with androgenic biosynthesis are important to control the disease. Thus, in this work, a 4D-quantitative structure-activity relationship (QSAR) receptor independent study was carried out using the Laboratório de Quimiometria Teórica e Aplicada (LQTA)-QSAR approach and the new tool Web-4D-QSAR with a set of benzylidene oxazolidinedione and thiazolidinedione inhibitor of the 17 β -hydroxysteroid dehydrogenase type 3. The obtained model is robust, free from chance correlation, and with good predictability in all tests performed. In addition to this result, a good mecanistic interpretation related to the binding with the biological target could be traced, which strengthens its potential application in virtual screening and design of new inhibitors of the studied enzyme with the possibility of use in the pharmacological treatment of prostate cancer.

Keywords: prostate cancer, 17 β -HSD3, computer-aided drug design, Web-4D-QSAR

Introduction

Prostate cancer is the most common form of cancer among men above 50 years of age and the second most common in Western men, and its incidence has been increasing progressively in the last few decades.^{1,2} Surgical removal of the prostate, or radiotherapy, is indicated for localized tumors. In the cases involving metastasis, the treatment is based on androgenic suppression with hormone therapy, using analogs from luteinizing hormone (LHRH).³⁻⁵ However, all approaches pose considerable risks and side effects and negatively influence the patient's quality of life.⁴

Since the prostate depends on androgenic hormones to regulate its growth and development, regulation and inhibition of the androgenic receptor action and/or inhibition of the androgenic biosynthesis are important mechanisms to control the disease.⁶ Drugs such as flutamide, bicalutamide, enzalutamide, and apalutamide (ARN-509) function via the first mechanism of action, while abiraterone acts through the second mechanism (Figure 1).⁷ In this context, enzyme 17 β -hydroxysteroid dehydrogenase type 3 (17 β -HSD3),

which converts androstenedione to active androgen testosterone, that is over-expressed in hormone-dependent prostate cancer, is considered a potential target to the chemotherapy of prostate cancer.^{2,6,8,9}

Computer-aided drug design (CADD) methods include a series of techniques to discover, design, and improve chemicals *in silico* with the purpose of being used as new drugs.¹⁰⁻¹² The correct use of these techniques could mitigate general costs, time required to obtain positive results, use of animals in the laboratory, and chemical and biological residues.^{13,14} Among CADD approaches, quantitative structure-activity relationships (QSAR) look for obtaining prediction models that correlate the chemical structure, through molecular descriptors, with a determined biological activity.^{15,16}

Considering the increasing interest in new chemotherapeutic agents for the prostate cancer treatment, this study aims to obtain a 4D-QSAR receptor independent (RI) model with a set of 49 derivatives of benzylidene oxazolidinedione and thiazolidinedione available in the literature,⁶ which show the ability to inhibit the 17 β -HSD3 enzyme.

*e-mail: joaopauloam@ufmg.br

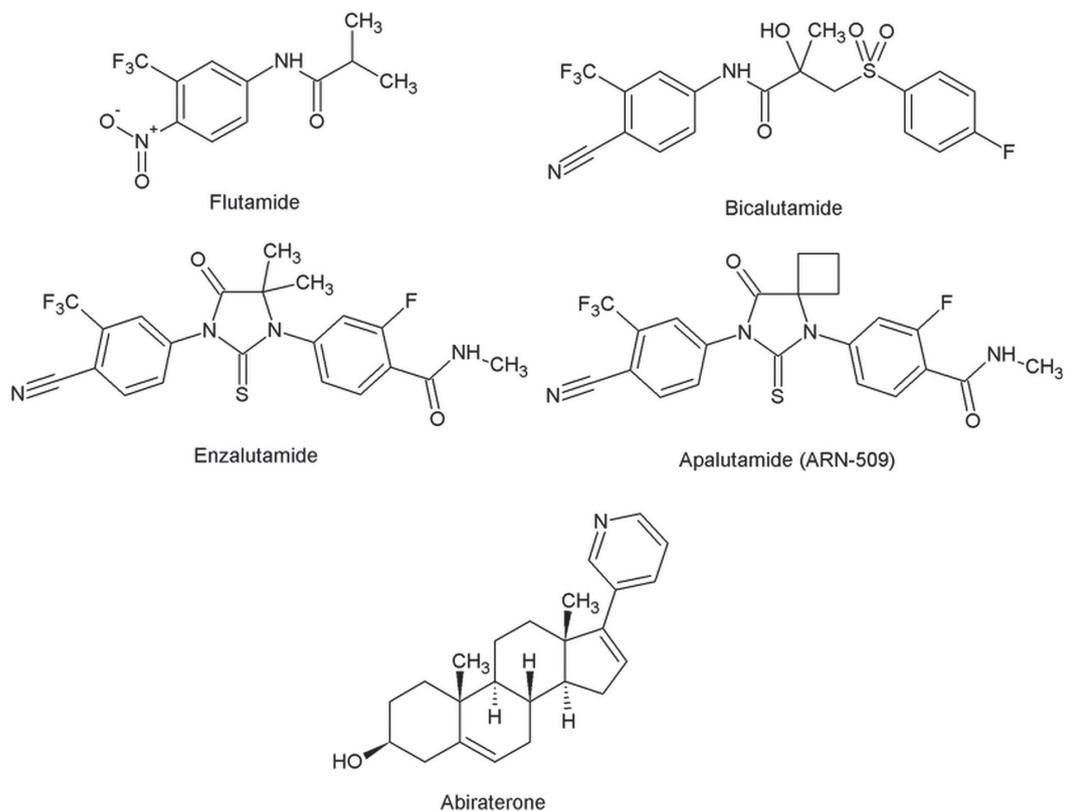


Figure 1. Structure of approved and under study anti-androgens.

Methodology

Data set

The 49 benzylidene oxazolidinedione and thiazolidinedione (Figure 2 and Table 1) used in this work were synthesized and published by Harada *et al.*⁶ All compounds were assayed for their *in vitro* inhibition of the enzyme 17 β -HSD3. The data reported as half-maximal inhibitory concentration (IC_{50} , in nM) were converted to $-\log IC_{50}$ (pIC_{50}) to obtain symmetrically distributed data. The data set was divided into training and test sets of 39 and 10 molecules, respectively.

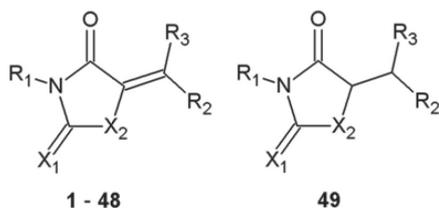


Figure 2. Basic structure of the data set.

Construction of the ligands' three-dimensional molecular models

The 49 derivatives were built with HyperChem 7

software,¹⁷ from the geometry of the file 01059739 obtained in ZINC database.¹⁸ Using the same program, geometry optimizations were performed using molecular mechanics (MM+) and, in sequence, using the semi empirical level Austin Model 1 (AM1). Further, using the Gaussian 09 program,¹⁹ new calculations were carried out at the Hartree-Fock (HF)/6-31G(d) level and then using density functional theory (DFT) (B3LYP/6-311G(d,p)).

Conformational ensemble profile (CEP) of each ligand, CEP alignment and calculation of the descriptors

Molecular descriptors were obtained from the CEP for the RI-4D-QSAR study using the Laboratório de Quimiometria Teórica e Aplicada (LQTA)-QSAR methodology.²⁰ The choice of an RI approach is justified by the fact that crystallographic structures of 17 β -HSD3 are not yet available, although the use of protein structures is not essential for a study involving molecular interaction fields (MIFs). According to LQTA-QSAR methodology, the molecular set was submitted to a molecular dynamics (MD) simulation using the GROMACS molecular package²¹ according to the scheme described by Martins *et al.*²² The trajectory file was recorded every 1000 simulation steps. The CEPs of all ligands were assembled in the same file considering the ligand conformations recorded from 50 to

Table 1. Dataset studied and the respective pIC₅₀ values^a

Compound	R ₁	R ₂	R ₃	X ₁	X ₂	pIC ₅₀
1	4-OCH ₃ -Ph-	3-Br-4-OH-Ph-	H	S	S	7.85
2	CH ₃	3-Br-4-OH-Ph-	H	S	S	6.05
3	allyl	3-Br-4-OH-Ph-	H	S	S	6.92
4	propargyl	3-Br-4-OH-Ph-	H	S	S	6.52
5	CH ₃ OCH ₂ CH ₂	3-Br-4-OH-Ph-	H	S	S	6.72
6	C ₆ H ₁₁	3-Br-4-OH-Ph-	H	S	S	8.52
7	tetrahydropyran-4-yl	3-Br-4-OH-Ph-	H	S	S	7.12
8	4-OCH ₃ -Ph-	3-Br-4-OH-Ph-	H	S	S	7.05
9	4-OCH ₃ -Ph-	3-Br-4-OH-Ph-	H	S	O	7.89
10	Ph-	3-Br-4-OH-Ph-	H	S	O	8.22
11	3-OCH ₃ -Ph-	3-Br-4-OH-Ph-	H	S	O	8.15
12	2-OCH ₃ -Ph-	3-Br-4-OH-Ph-	H	S	O	7.22
13	3,4-OCH ₂ O-Ph-	3-Br-4-OH-Ph-	H	S	O	8.10
14	4-F-Ph-	3-Br-4-OH-Ph-	H	S	O	8.70
15	4-Cl-Ph-	3-Br-4-OH-Ph-	H	S	O	8.22
16	4-CH ₃ -Ph-	3-Br-4-OH-Ph-	H	S	O	8.70
17	4-CF ₃ -Ph-	3-Br-4-OH-Ph-	H	S	O	8.52
18	4-CN-Ph-	3-Br-4-OH-Ph-	H	S	O	7.72
19	4-OCH ₃ -Ph-	3-Br-4-OH-Ph-	H	O	O	7.64
20	4-COOCH ₃ -Ph-	3-Br-4-OH-Ph-	H	O	O	7.62
21	4-C(O)NHCH ₃ -Ph-	3-Br-4-OH-Ph-	H	O	O	7.25
22	4-C(O)N(CH ₃) ₂ -Ph-	3-Br-4-OH-Ph-	H	O	O	7.60
23	4-C(O)-piperazine-Ph-	3-Br-4-OH-Ph-	H	O	O	7.38
24	4-C(O)-morpholine-Ph-	3-Br-4-OH-Ph-	H	O	O	6.52
25	4-N(CH ₃) ₂ -Ph-	3-Br-4-OH-Ph-	H	O	O	7.60
26	4-NHAc-Ph-	3-Br-4-OH-Ph-	H	O	O	7.24
27	4-NHCO-(c-hex)	3-Br-4-OH-Ph-	H	O	O	7.68
28	4-SCH ₃ -Ph-	3-Br-4-OH-Ph-	H	O	O	7.05
29	4-SO ₂ CH ₃ -Ph-	3-Br-4-OH-Ph-	H	O	O	6.05
30	pyridin-2-yl	3-Br-4-OH-Ph-	H	O	O	7.13
31	pyridin-3-yl	3-Br-4-OH-Ph-	H	O	O	7.67
32	4-OCH ₃ -pyridin-3-yl	3-Br-4-OH-Ph-	H	O	O	8.40
33	4-CH ₃ -pyridin-3-yl	3-Br-4-OH-Ph-	H	O	O	7.17
34	4-F-pyridin-3-yl	3-Br-4-OH-Ph-	H	O	O	7.68
35	4-Cl-pyridin-3-yl	3-Br-4-OH-Ph-	H	O	O	7.70
36	4-N(CH ₃) ₂ -pyridin-3-yl	3-Br-4-OH-Ph-	H	O	O	8.40
37	4-OCH ₃ -Ph-	4-OH-Ph-	H	S	S	6.91
38	4-OCH ₃ -Ph-	3-CH ₃ -4-OH-Ph-	H	S	S	7.70
39	4-OCH ₃ -Ph-	3-OCH ₃ -4-OH-Ph-	H	S	S	6.60
40	4-OCH ₃ -Ph-	3-F-4-OH-Ph-	H	S	S	8.70
41	4-OCH ₃ -Ph-	3-Cl-4-OH-Ph-	H	S	S	8.40
42	4-OCH ₃ -Ph-	2-Cl-4-OH-Ph-	H	S	S	8.30
43	4-OCH ₃ -Ph-	3,5-Cl-4-OH-Ph-	H	S	S	7.92
44	4-OCH ₃ -Ph-	3,5-F-4-OH-Ph-	H	S	S	8.30
45	4-OCH ₃ -Ph-	3-Cl-4-OH-5-F-Ph-	H	S	S	9.00
46	4-OCH ₃ -Ph-	4-OH-pyridin-2-yl	H	S	S	7.05
47	4-OCH ₃ -Ph-	4-OH-3-Br-pyridin-2-yl	H	S	S	7.40
48	4-OCH ₃ -Ph-	3-Br-4-OH-Ph-	CH ₃	S	S	7.05
49	4-OCH ₃ -Ph-	3-Br-4-OH-Ph-	H	S	S	7.05

^aAccording to Harada *et al.*,⁶ all tests were carried out using human cell-based assay, with HeLa cells transiently transfected with 17 β -HSD3, and the deviations were within $\pm 5\%$. pIC₅₀: $-\log IC_{50}$ (IC₅₀: half-maximal inhibitory concentration).

500 ps, and these data were used for building the QSAR models. Atoms used for the alignment of the different conformers are presented in Figure 3 and the CEPs for the most and least active compounds from the molecular set are presented in Figure 4.

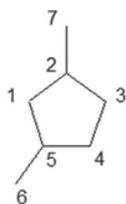


Figure 3. Representation of the atoms used for the alignment of the CEP.

The CEPs, resulting from GROMACS MD simulations, were used to generate the energy descriptors of intermolecular interactions with LQTAgrid program.²⁰ In this approach a probe selected by the user runs over a grid that contains the CEP, and the electrostatic and steric 3D properties are computed for each individual grid point, based on the Coulombic (equation 1) and Lennard-Jones potential functions (equation 2), respectively.

$$E_{\text{ele}} = \frac{1}{n} \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}} \quad (1)$$

$$E_{\text{LJ}} = \frac{1}{n} \left(\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6} \right) \quad (2)$$

with

$$C_{ij}^{(12)} = \sqrt{C_{ii}^{(12)} C_{jj}^{(12)}} \quad (3)$$

$$C_{ij}^{(6)} = \sqrt{C_{ii}^{(6)} C_{jj}^{(6)}} \quad (4)$$

where q_i is the charge of the i th probe; q_j is the charge of the j th atom from CEP; ϵ_0 is the vacuum permittivity; $C_{ii}^{(12)}$, $C_{ii}^{(6)}$, $C_{jj}^{(12)}$ and $C_{jj}^{(6)}$ are parameters for the probes and the atoms in the CEP, respectively; n indicates the number of frames aligned in the CEP; and r_{ij} represents the distances between the i th probe and the j th atom of CEP.

The probe used for this study, based on the ff43a1 force field parameterization to simulate atoms or molecular fragments, was NH_3^+ , which corresponds to the amino-terminal portion of peptides. The probe explored every point of a 1 Å resolution grid, and the grid size was $30 \times 34 \times 36 \text{ \AA}^3$. This study was accomplished with a new web based software running LQTA-QSAR methodology called Web-4D-QSAR.²² The generated 3D-energy interaction descriptors were exploited further in the variable selection procedure.

Variable reduction and selection

As a step of variable reduction, the absolute values of the correlation coefficients between each descriptor and the biological activity ($|r|$) were calculated, and those with coefficients lower than 0.3 were eliminated from the analysis. The variable selection was carried out using the ordered predictor selection (OPS) algorithm.^{23,24} This algorithm attributes an importance to each descriptor based on an informative vector. The columns of the matrix are rearranged in such a way that the most important descriptors are presented in the first columns. Then, successive partial least squares (PLS) regressions are performed with an increasing number of descriptors in order to find the best PLS model. In this analysis, the product vector was used as the informative vector and the correlation coefficient of cross-validation (Q^2_{LOO}) as the criterion to select the best models.

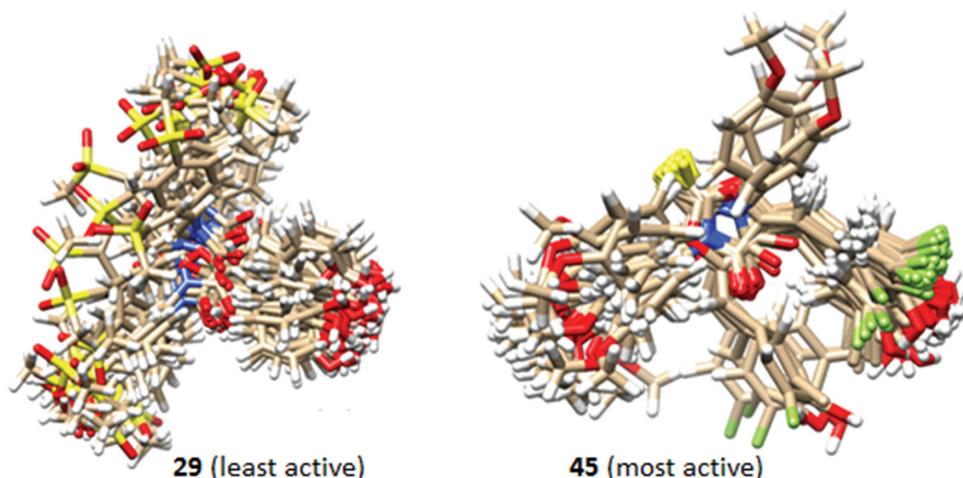


Figure 4. Comparison of the CEPs resulting from MD simulations of the least active (left) and most active (right) compounds.

Construction of QSAR models

4D-QSAR models were built using PLS regression²⁵ using QSAR modeling program.²³ Since in QSAR the variables normally have different numerical ranges, it was necessary to perform a preprocessing of the data using the auto-scaling approach: the variables were mean-centered and scaled by variance.²⁶ The quality of samples was evaluated through the difference between the predicted activity and the observed activity in the data set. Considering the degree of this difference, the sample could be considered an outlier.²⁷

Statistical quality of the QSAR models

In this study, statistical parameters recommended for the validation of QSAR models^{23,26,28-30} were used to ensure the internal and external prediction quality of the obtained models. The explained variances of the models were evaluated using the coefficient of determination (R^2) and the root mean square error of calibration (RMSEC). For adjusted QSAR models, we adopted the most recommended limit for R^2 (> 0.6), while the values of RMSEC should be as close as possible to 0.^{28,29} The significance of the models was evaluated using the F-test at a 5% level.²⁶

Leave-one-out (LOO) cross-validation was performed by calculating the Q^2_{LOO} and the root mean square error of cross-validation (RMSECV). For a model to be approved, the RMSECV should be close to 0, and $Q^2_{LOO} > 0.5$.^{28,29} The RmSquare metrics³¹ for internal validation, where the average $r_m^2(LOO)$ -scaled > 0.5 , and $\Delta r_m^2(LOO)$ -scaled < 0.2 were also carried out. As a robustness test of internal quality, we carried out the leave-N-out (LNO) cross-validation and achieved a systematic removal of up to N elements of the training set. In this study, N = 10 (6 replicates for each N) for all models. For a robust model, the average value of Q^2_{LNO} should be as close as possible to the value of Q^2_{LOO} .²⁶ The chance correlation was also verified by the y-randomization test, another internal quality test. The models were recalculated after randomization of the vector y , and a significant worsening in these new regressions was expected. This procedure was repeated 50 times. The results were evaluated according to the method proposed by Eriksson *et al.*¹³

The external validation of the obtained models was performed by predicting the pIC_{50} values of the test set. The quality was assessed using the coefficient of determination of external validation (R^2_{pred}), $r_m^2(pred)$ -scaled metrics, and Golbraikh-Tropsha metrics (k , k' and $|R^2_0 - R'^2_0|$). It is recommended that $R^2_{pred} > 0.5$, the root mean square error of prediction (RMSEP) be close to 0, average

$r_m^2(pred)$ -scaled > 0.5 , and $\Delta r_m^2(pred)$ -scaled < 0.2 . The k and k' metrics should be in the range 0.85-1.15, and $|R^2_0 - R'^2_0| < 0.3$.³¹⁻³⁴

The second test involved the Prediction Reliable Indicator, a tool recently presented by Roy *et al.*³⁵ In this tool, predictive ability of the model is evaluated regarding new compounds that were not part of the original data set. The predictions for this set (named real external set) are ranked into good, moderate, and poor/unreliable. To make this test feasible, a 2D similarity search (DICE approach of 60% based on the basic structure of compounds **1** to **49**) was carried out in the ZINC15 database.³⁶ Excluding duplicated results, eight compounds (**50** to **57**, Figure 5) were submitted to the test. As it was done for the original data set, molecular descriptors were calculated in LQTagrid program.²⁰ In order to increase the reliability of these predictions, applicability domain was also calculated using the Euclidean Applicability Domain 1.0 tool.³⁷ This approach is based on the similarity between the descriptors used in the model and in real external set, and a compound is considered within the domain if the normalized average distances of each compound in the original set to it fall between 0-1.³⁸

Results and Discussion

The obtained model (equation 5) corresponds to the results of the 4D-QSAR study obtained using the OPS-PLS approach, and it was built based on eight field descriptors (three Coulomb descriptors and five Lennard-Jones descriptors, where the numbers indicate the coordinates in the Cartesian space of each descriptor) that are projected to two latent variables (LV) accounting for 32.937% (LV1: 18.889%; LV2: 14.048%) of the original information. This model was obtained with 39 compounds (no outliers were identified) and ten compounds (**6**, **13**, **19**, **23**, **25**, **30**, **39**, **43**, **47** and **48**) were selected to form the external test set. This selection was carried out in such a way that all the range of the biological activity of the test set could be covered. The statistical results of the obtained model are presented in Table 2. The values of the descriptors selected in the model as well as the predicted values of the pIC_{50} in the cross validation for the training set and in the external validation for the test set are available in the Supplementary Information section (Table S1).

$$pIC_{50} = 8.76 - 0.012 \times (20_17_17_LJ) - 0.056 \times (11_10_12_C) - 0.19 \times (18_24_22_LJ) + 0.013 \times (17_15_15_LJ) + 0.0020 \times (19_10_19_LJ) + 0.0025 \times (10_18_23_C) - 0.010 \times (20_17_16_LJ) + 0.020 \times (12_21_17_C) \quad (5)$$

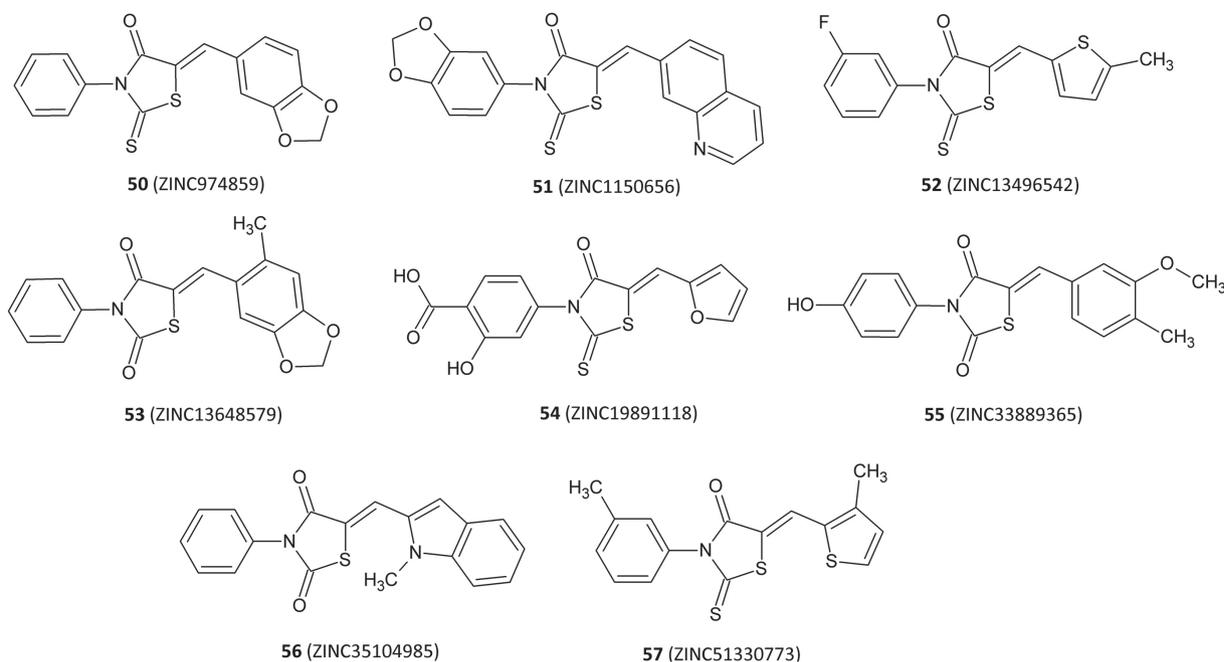


Figure 5. Compounds selected in the virtual screening step using 2D similarity.

Table 2. Results obtained in internal and external validations

Parameter	Value
N	39
R ²	0.77
RMSEC	0.35
F (cF)	60.98
Q ² _{LOO}	0.61
RMSECV	0.44
Average r _m ² (LOO)-scaled	0.50
Δr _m ² (LOO)-scaled	0.14
R ² -Q ² _{LOO}	0.16
R ² _{pred}	0.74
RMSEP	0.35
Average r _m ² (pred)-scaled	0.61
Δr _m ² (pred)-scaled	0.20
k	1.00
k'	0.10
R ² ₀ - R ² ₀ '	0.13

N: number of elements; R²: coefficient of determination; RMSEC: root mean square error of calibration; F: F-ratio test; Q²_{LOO}: square correlation coefficient for leave-one-out cross-validation; RMSECV: root mean square error of cross-validation; r_m²(LOO)-scaled and Δr_m²(LOO)-scaled: average r_m²(LOO)-scaled > 0.5, and Δr_m²(LOO)-scaled < 0.2: average and difference between r_m²(LOO)-scaled and r'_m²(LOO)-scaled values, where r_m² is obtained from predicted values in LOO validation plotted in **x** axis while real values are plotted in **y** axis and r'_m² is similar but the predicted and real values are plotted in opposite axis; R²_{pred}: coefficient of determination of external validation; RMSEP: root mean square error of prediction; r_m²(pred)-scaled and Δr_m²(pred)-scaled: average r_m²(pred)-scaled > 0.5, and Δr_m²(pred)-scaled < 0.2: same meaning as for LOO metrics, but applied to the test set of the external validation; k, k' and |R²₀ - R²₀'|: Golbraikh-Tropsha metrics.

The results of internal validation statistics show that the model is able to explain and predict enough information. The result of the F-test is higher than the corresponding tabulated value, thus the obtained regression can be considered significant. The average result of LNO (Q²_{LNO} = 0.59) is close to Q²_{LOO}, with the largest deviation observed for Q²_{LSO} (0.57 ± 0.055), showing that the model is robust (Figure 6a). The intercepts of the models obtained through the regressions built in the **y**-randomization test (Figure 6b) also show that the model is robust and free from chance correlation.

One of the advantages of a 4D model is the possibility of better interpretation of the model. These type of models are able to encode information regarding how the spatial features arising from conformational flexibility may be important for a specific biological activity.^{16,39-42} Previous studies using LQTA-QSAR approach^{16,43} showed that it is possible to obtain models such that even if it is an RI model, the selected descriptors may be related to amino acid residues that form binding sites of biological targets, and this may be useful in the interpretation of results. However, unlike the cited studies, the crystallographic structure of 17β-HSD3 is not available. According to Lukacik *et al.*,⁴⁴ its hydrophobic nature has thus far prevented its structure determination. Nevertheless, other proteins from the same family are already available and may be used as a tool to aid interpretation due to high homology between them. The human 17β-HSD1 presents 74% of homology with type 3,⁸ and thus was used to help in the interpretation process.

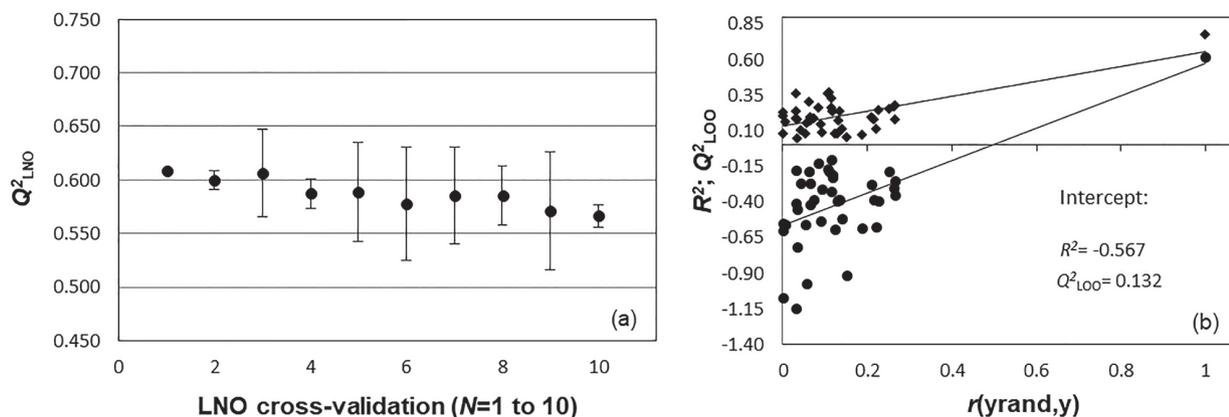


Figure 6. Plots of LNO cross-validation (a) and y-randomization tests (b) of model 1.

Figures 7a and 7b present the eight descriptors selected in the space around the CEPs of the least active molecule (29) and the most active (45), respectively. It can be observed that the concentration of Coulomb descriptors at the most flexible end corresponds to position R₁, while the Lennard-Jones descriptors are concentrated at the center of the molecule (with the exception of 18_24_22_LJ).

This distribution reminds the organization of the cavity of 17 β -HSD, which is characterized by a central hydrophobic region and ends formed by groups capable of forming hydrogen bonds (Figure 7c).

It is also interesting to observe that most of the descriptors in the space presents patterns very similar to the positioning of the amino acid residues that form the cavity

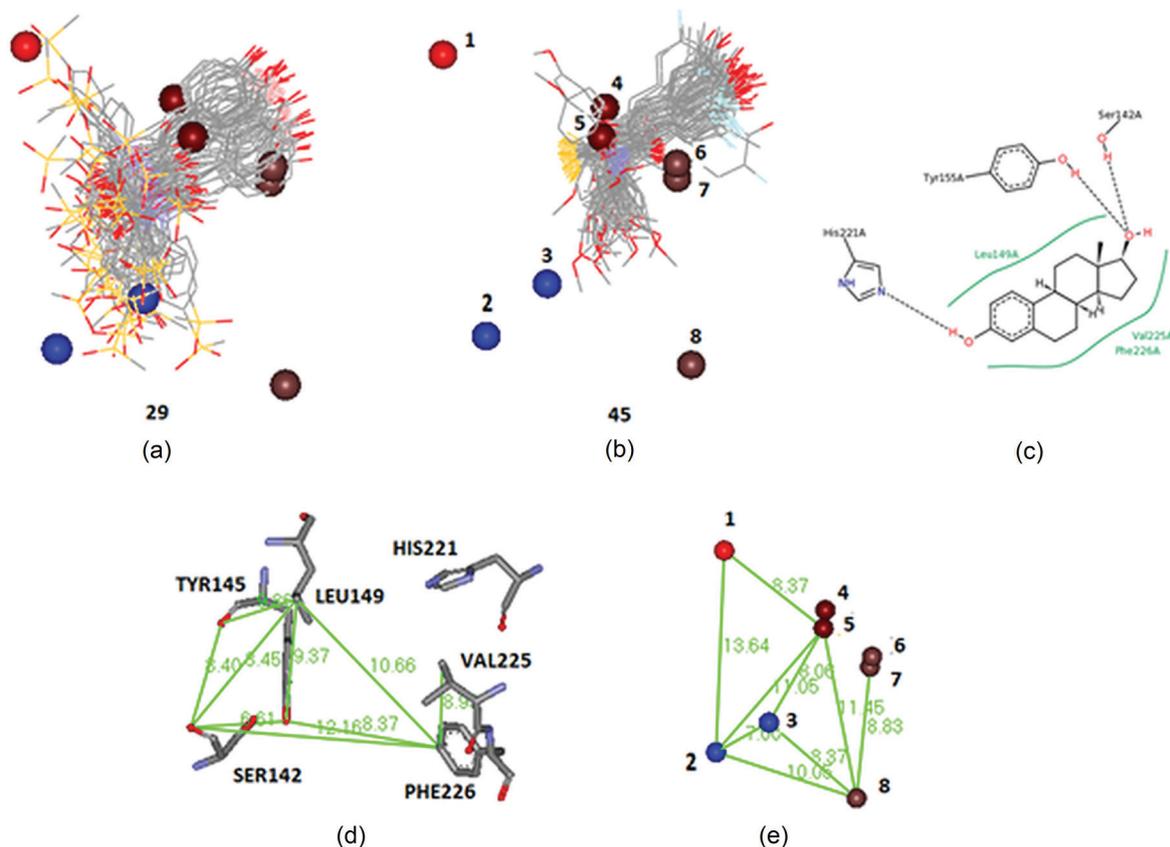


Figure 7. (a) and (b) Full model in three-dimensional space surrounding the CEP of the least and most active derivatives. Descriptors are 1 (11_10_12_C), 2 (10_18_22_C), 3 (12_21_17_C), 4 (19_10_19_LJ), 5 (17_15_15_LJ), 6 (20_17_16_LJ), 7 (20_17_17_LJ), and 8 (18_24_22_LJ); red: negative Coulomb descriptors; blue: positive Coulomb descriptors; brown: positive Lennard-Jones descriptors; light brown: negative Lennard-Jones descriptors; (c) PoseView diagram of cavity of Protein Data Bank (PDB) human 17 β -HSD1 (1FDT) complexed with estradiol; (d) distances between the principal amino acids of active site of PDB 1FDT; (e) distances between the descriptors of the model for comparison with (d). Except for (c) (1FDT 2D diagram and interactions obtained from PDB), all items were built using ViewerLite 4.2.⁴⁵

Table 3. Data and results for the real external test set

Compound	Molecular descriptor								Pred. pIC ₅₀	Prediction quality
	20_17_17_LJ	11_10_12_C	18_24_22_LJ	17_15_15_LJ	19_10_19_LJ	10_18_23_C	20_17_16_LJ	12_21_17_C		
50	142.87	0.66	-0.15	141.03	27.92	-4.24	144.57	-71.64	5.86	moderate
51	135.55	-5.27	-0.09	184.04	-1.02	0.32	127.22	-9.59	8.31	moderate
52	132.28	0.49	-0.84	155.76	32.42	-5.65	136.92	-46.99	6.87	moderate
53	151.23	8.40	-1.10	157.02	-0.57	-26.02	149.28	-18.65	6.13	moderate
54	147.13	10.64	-0.61	166.03	133.41	-2.93	147.17	-15.18	6.97	moderate
55	152.72	-1.86	-0.16	173.34	135.89	-2.16	155.34	-7.86	7.69	moderate
56	132.55	-2.84	-0.14	173.34	131.15	-14.66	135.05	-79.27	6.46	moderate
57	18.06	-0.73	3.95	156.91	42.30	0.44	133.42	-35.47	7.87	moderate

Pred. pIC₅₀: predicted -log IC₅₀ (IC₅₀: half-maximal inhibitory concentration).

(Figures 7d and 7e), including the distances among them, with the largest difference corresponding to the distance between descriptors 1 and 2 (13.64 Å), and the distance between the residues SER142 and the acid group of the backbone of TYR145 (8.40 Å). In the other cases, the maximum difference observed is only 2.6 Å. It is important to observe that the enzyme used in the comparison is not 17β-HSD3, but a homolog, which can explain the observed differences. Thus, it may be proposed that the 4D-QSAR model presents the interpretation related to the blocking mechanism of 17β-HSD3 by the studied benzylidene oxazolidinedione and thiazolidinediones.

Considering the autoscaled values of the selected descriptors (11_10_12_C (-0.48), 12_21_17_C (0.28), 10_18_23_C (0.27), 18_24_22_LJ, (-0.26), 20_17_17_LJ (-0.21), 17_15_15_LJ (0.19), 19_10_19_LJ (0.16) and 20_17_16_LJ (-0.13)), it can be observed that the most important (11_10_12_C) is occupied by the R₁ side chain of the most active compound. Since this descriptor has a negative coefficient in the model, this fact indicates that the occupation of this region of the target tends to make the compounds be less active. A similar interpretation could be given for descriptors 10_18_23_C and 12_21_17_C. Despite the positive values that the coefficients of these descriptors present, the values of the descriptors are predominantly negative, especially for 10_18_23_C, which is the furthest from the most active compound **45**.

A similar analysis could be done for descriptor 18_24_22_LJ, the most important Lennard-Jones descriptor and that also presents negative coefficient. High values for this descriptor are favored by bigger R₁ substituents, which in turn decrease the biological activity. The result of this interpretation, in addition to good global statistics, strengthens the hypothesis that the model adequately describes the inhibition of this enzyme.

Referring to the test with the Prediction Reliable

Indicator tool, all compounds had their predictions classified as moderate (Table 3) using the standard weighing scheme 0.5-0-0.5.³⁵ But the Euclidean applicability domain (Figure 8) shows that one compound (**57**) has a normalized value greater than 1. Thus, the prediction obtained for this compound is not reliable, while the prediction applied to the other compounds, which corresponds to 87.5% of the real external validation set, is reliable. These results, combined with the results of internal and external validations, and mechanistic interpretation, show that the model has a real potential to predict with confidence the potential for inhibition of the enzyme 17β-HSD3 for new hits that are identified, or even synthesized, and which are related to the original structure.

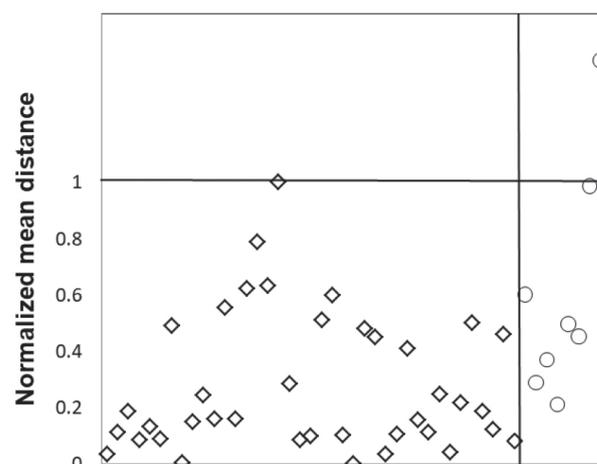


Figure 8. Plot of Euclidean applicability domain. Diamonds: training set; circles: real external set.

Conclusions

In this study, we have developed a 4D-QSAR model using a set of 49 benzylidene oxazolidinedione and thiazolidinedione derivatives with inhibitory activity

against 17 β -HSD3. The model was validated both internally and externally, indicating that it is significant, robust, had no chance correlation, and presented good predictive ability. The test carried out with the Prediction Reliability Indicator tool indicated that the model presents sufficient reliability to predict the inhibition of new potential inhibitors of the enzyme of interest that are structurally related to the data set. Besides, the interpretation could be directly related to the interaction with the binding site of 17 β -HSD3. Thus, the obtained 4D model may be a useful tool for aiding virtual screening and design studies of new inhibitors of 17 β -HSD3, so they can be used as lead compounds or new drugs for potential application in prostate cancer treatment.

Supplementary Information

Supplementary data are available free of charge at <http://jbcs.sbq.org.br> as PDF file.

Acknowledgments

The authors wish to thank Fundação Araucária (grant 2010/7354), PROAP/CAPES, and Pró-Reitoria de Pesquisa (PRPq) of Universidade Federal de Minas Gerais (UFMG) for the support received.

References

- Marta, M. G.; Hanna, A. S.; Silva, J. L. F.; Carvalho, H. A.; *Diagn. Trat.* **2012**, *17*, 90.
- Vicker, N.; Sharland, C. M.; Heaton, W. B.; Gonzalez, A. M.; Bailey, H. V.; Smith, A.; Springall, J. S.; Day, J. M.; Tutill, H. J.; Reed, M. J.; Purohit, A.; Potter, B. V.; *Mol. Cell. Endocrinol.* **2009**, *301*, 259.
- Damião, R.; Sarkis, A. S.; Jacobino, A. O.; Carrerette, F. B.; Rocha, G.; *Câncer de Próstata Metastático: Tratamento e Complicações*; Sociedade Brasileira de Urologia: São Paulo, 2006.
- Clark, J. A.; Inui, T. S.; Silliman, R. A.; Bokhour, B. G.; Krasnow, S. H.; Robinson, R. A.; Spaulding, M.; Talcott, J. A.; *J. Clin. Oncol.* **2003**, *21*, 3777.
- Berger, A. P.; Spranger, R.; Kofler, K.; Steiner, H.; Bartsch, G.; Horninger, W.; *Prostate* **2003**, *57*, 93.
- Harada, K.; Kubo, H.; Abe, J.; Haneta, M.; Conception, A.; Inoue, S.; Okada, S.; Nishioka, K.; *Bioorg. Med. Chem.* **2012**, *20*, 3242.
- Trewartha, D.; Carter, K.; *Nat. Rev. Drug Discovery* **2013**, *12*, 823.
- Marchais-Oberwinkler, S.; Henn, C.; Möller, G.; Klein, T.; Negri, M.; Oster, A.; Spadaro, A.; Werth, R.; Wetzel, M.; Xu, K.; Frotscher, M.; Hartmann, R. W.; Adamski, J.; *J. Steroid Biochem. Mol. Biol.* **2011**, *125*, 66.
- Ning, X.; Yang, Y.; Deng, H.; Zhang, Q.; Huang, Y.; Su, Z.; Fu, Y.; Xiang, Q.; Zhang, S.; *Steroids* **2017**, *121*, 10.
- Sant'Anna, C. M. R.; *Quim. Nova* **2002**, *25*, 505.
- Santos Filho, A. O.; Alencastro, R. B.; *Quim. Nova* **2003**, *26*, 253.
- Silva, V. B.; Silva, C. H. T. P.; *Rev. Eletrônica Farm.* **2007**, *4*, 15.
- Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P.; *Environ. Health Perspect.* **2003**, *111*, 1361.
- Organization for Economic Co-Operation and Development (OECD); *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*; OECD: Paris, 2006.
- Almeida, V. L.; Lopes, J. C. D.; Oliveira, S. R.; Donnici, C. L.; *Quim. Nova* **2010**, *7*, 1482.
- Andrade, C. H.; Pasqualoto, K. F. M.; Ferreira, E. I.; Hopfinger, A. J.; *Molecules* **2010**, *15*, 3281.
- HyperChem(TM) Professional 7.51*; Hypercube, Inc., Gainesville, Florida, USA.
- Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G.; *J. Chem. Inf. Model.* **2012**, *52*, 1757; <http://zinc.docking.org>, accessed in March 2019.
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; *Gaussian 09, Revision A.02*; Gaussian, Inc., Wallingford CT, 2009.
- Martins, J. P. A.; Barbosa, E. G.; Pasqualoto, K. F. M.; Ferreira, M. M. C.; *J. Chem. Inf. Model.* **2009**, *49*, 1428.
- Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E.; *SoftwareX* **2015**, *1-2*, 19.
- Martins, J. P. A.; Oliveira, M. A. R.; Queiroz, M. S. O.; *J. Comput. Chem.* **2018**, *39*, 917; <https://github.com/rougeth/Web-4D-QSAR>, accessed in March 2019.

23. Ferreira, M. M. C.; Martins, J. P. A.; *Quim. Nova* **2013**, *36*, 554.
24. Teófilo, R. F.; Martins, J. P.; Ferreira, M. M. C.; *J. Chemom.* **2009**, *23*, 32.
25. Wold, S.; Sjöström, M.; Eriksson, L.; *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109.
26. Kiralj, R.; Ferreira, M. M. C.; *J. Braz. Chem. Soc.* **2009**, *20*, 770.
27. Gramatica, P.; *QSAR Comb. Sci.* **2007**, *26*, 694.
28. Chirico, N.; Gramatica, P.; *J. Chem. Inf. Model.* **2011**, *51*, 2320.
29. Chirico, N.; Gramatica, P.; *J. Chem. Inf. Model.* **2012**, *52*, 2044.
30. Roy, K.; Mitra, I.; Kar, S.; Ojha, P. K.; Das, R. N.; Kabir, H.; *J. Chem. Inf. Model.* **2012**, *52*, 396.
31. Ojha, P. K.; Mitra, I.; Das, R. N.; Roy, K.; *Chemom. Intell. Lab. Syst.* **2011**, *107*, 194.
32. Golbraikh, A.; Tropsha, A.; *J. Mol. Graphics Modell.* **2002**, *20*, 269.
33. Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.; Lee, K.; Tropsha, A.; *J. Comput.-Aided Mol. Des.* **2003**, *17*, 241.
34. Tropsha, A.; Gramatica, P.; Gombar, V. K.; *QSAR Comb. Sci.* **2003**, *22*, 69.
35. Roy, K.; Ambure, P.; Kar, S.; *ACS Omega* **2018**, *3*, 11392; http://teqip.jdvu.ac.in/QSAR_Tools, accessed in March 2019.
36. Sterling, T.; Irwin, J. J.; *J. Chem. Inf. Model.* **2015**, *55*, 2324; <http://zinc15.docking.org>, accessed in March 2019.
37. http://teqip.jdvu.ac.in/QSAR_Tools, accessed in March 2019.
38. Nandy, A.; Kar, S.; Roy, K.; *Mol. Simul.* **2014**, *40*, 261.
39. Damale, M. G.; Harke, S. N.; Khan, F. A. K.; Shinde, D. B.; Sangshetti, J. N.; *Mini-Rev. Med. Chem.* **2014**, *14*, 35.
40. Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; *J. Am. Chem. Soc.* **1997**, *119*, 10509.
41. Vedani, A.; Dobler, M.; *Quant. Struct.-Act. Relat.* **2002**, *21*, 382.
42. Ghasemi, J. B.; Safavi-Sohi, R.; Barbosa, E. G.; *Mol. Diversity* **2012**, *16*, 203.
43. de Melo, E. B.; Ferreira, M. M. C.; *J. Chem. Inf. Model.* **2012**, *52*, 1722.
44. Lukacik, P.; Kavanagh, K. L.; Oppermann, U.; *Mol. Cell. Endocrinol.* **2006**, *248*, 61.
45. *ViewerLite 4.2*, Accelrys Inc., San Diego, 2001.

Submitted: October 24, 2018

Published online: April 1, 2019

