



Dihydrofolate reductase inhibitors: a quantitative structure–activity relationship study using 2D-QSAR and 3D-QSAR methods

Juan C. Garro Martinez^{1,2} · Matias F. Andrada¹ · Esteban G. Vega-Hissi^{1,2} · Francisco M. Garibotto^{1,2} · Manuel Noguerras³ · Ricaurte Rodríguez^{3,4} · Justo Cobo³ · Ricardo D. Enriz^{1,2} · Mario R. Estrada¹

Received: 14 June 2016 / Accepted: 3 November 2016
© Springer Science+Business Media New York 2016

Abstract In this work, we study the structure–activity relationship of a series of Dihydrofolate reductase inhibitors by two-dimensional quantitative activity–structure relationship and three-dimensional quantitative activity–structure relationship techniques. The two-dimensional quantitative activity–structure relationship models were developed by using two different types of topological molecular descriptors, PaDEL and Dragon descriptors. The models showed an excellent predictive power, $R^2_{\text{train}} = 0.916$ and $R^2_{\text{val}} = 0.806$ for the PaDEL, and $R^2_{\text{train}} = 0.952$ and $R^2_{\text{val}} = 0.963$ for those obtained with Dragon descriptors. Simple molecular descriptors as maxHCsats, IC3, SPI, SIC2, and GATS5p were adequate to obtain predictive models. The three-dimensional quantitative activity–structure relationship was performed through three variable selected approaches, Partial Linear Square (PLS), Fractional Factorial Design (FFD) and Uninformative Variable Elimination-Partial Linear Square (UVE-PLS) using the Open3DQSAR software.

All the 2D and 3D models were validated using two compounds (number **24** and **25**), which were synthesized and presented here for the first time. Their biological activities were correctly predicted by all the quantitative activity–structure relationship models. Finally, we proposed three compounds (**26**, **27**, and **28**), which showed a high predicted Dihydrofolate reductase inhibitory activity. Molecular docking study suggested that compounds bind to receptor similarly to the most active inhibitors.

Keywords Dihydrofolate reductase · QSAR · 3D-QSAR · Topological descriptors · DHFR inhibitors synthesis

Introduction

The enzyme Dihydrofolate reductase (DHFR) plays a key role in the folate pathway, responsible for the biosynthesis of deoxythymidine monophosphate, as well as purine nucleotides and the amino acids histidine and methionine (Beierlein et al. 2008; Li et al. 2011; Nammalwar et al. 2012). The DHFR catalyzes the transformation of dihydrofolate to tetrahydrofolate (Beierlein et al. 2009). The literature shows that DHFR is important in the studies of several human diseases such as protozoal, bacterial and fungal infections, psoriasis, autoimmune disease, and neoplastic diseases (Li et al. 2011; Sharma and Chauhan 2012).

The availability of high resolution crystal structures of DHFR of *P. carinii* (pcDHFR) and human DHFR (hDHFR), as well as a plethora of information in regard to the active and inactive compounds, has provided a solid platform for ligand-based drug design of potent and specific DHFR inhibitors (Champness et al. 1994; Klon et al. 2002). Our research group has previous experience in the synthesis of pyrimidine derivatives (Quiroga et al. 2009; De la Torre

Electronic supplementary material The online version of this article (doi:10.1007/s00044-016-1742-4) contains supplementary material, which is available to authorized users.

✉ Juan C. Garro Martinez
jcgarro@unsl.edu.ar

- ¹ Departamento de Química, Facultad de Química, Bioquímica y Farmacia, Universidad Nacional de San Luis, Chacabuco 915, San Luis 5700, Argentina
- ² IMIBIO-CONICET, UNSL, Chacabuco 915, San Luis 5700, Argentina
- ³ Departamento de Química Inorgánica y Orgánica, Universidad de Jaén, 23071 Jaén, Spain
- ⁴ Departamento de Química, Universidad Nacional de Colombia, Ciudad Universitaria, Carrera 30, No. 45-03 Bogotá, Colombia

et al. 2014; Olivella et al. 2012) and we have reported a new series of DHFR inhibitor, which possesses a moderate inhibitory effect against hDHFR (Tosso et al. 2013). More recently (Tosso et al. 2014) we have synthesized and tested seven novel compounds structurally related to those previously reported in Tosso et al. (2013).

The quantitative activity–structure relationship (QSAR) is one of the most utilized tools in rational drug design aided by computers. This approach is a mathematical hypothesis substantiated on the supposition that molecular structure (properties or descriptors of the molecules) is responsible for the biological activity of a compound. Thus, entities with analogous structure would present a similar biological activity. Various QSAR approaches [as three-dimensional QSAR (3D-QSAR)] have been developed gradually over a time span from the called “classic QSAR” [two-dimensional QSAR (2D-QSAR)], and served as a valuable predictive tool, particularly in the design of pharmaceuticals. The main difference between them resides in the manner as the structural properties are represented and the quantitative relationships with the biological activities are extracted (Verma et al. 2010).

In the past years, several QSAR studies specifically targeted to DHFR inhibitors have been performed. These investigations span from methods which use simple 2D descriptors to more complex and advanced receptor based 3D-QSAR methods. One of the most relevant papers is published by Agrawal et al. (2002), who reported a QSAR model using a series of 19 2,6-substituted 2,4-diaminopyrido [3,2-d]pyrimidine derivatives against pcDHFR using topological indexes. In another paper, (Mattioni and Jurs 2003) the author used a data set of 345 diverse DHFR inhibitors to build QSAR models using artificial neural networks. Deb-nath et al. (2003) reported a study of nonparabolic Hansch QSAR models with the aim to obtain physicochemical and structural features of pyrimidine derivatives. Manchester and Czerminski (2008) used a data set of DHFR inhibitors from the work of Sutherland and Weaver (2004) to compare simple atom mapping following alignment (SAMFA), a newly proposed 3D-QSAR method, with CoMFA. More recently, Gangjee et al. (2010) have described a CoMFA analysis of tgDHFR and rIDHFR based on 80 antifolates. This brief discussion shows the current interest in to elucidate the key structural characteristics for the inhibition through developing new QSAR models specific for DHFR.

In the present work, we performed a QSAR analysis on a series of compounds (Imino and Guanine derivatives) that possess DHFR inhibitory activity (Tosso et al. 2013, 2014). In contrast with other papers, we presented a combined 2D-QSAR and 3D-QSAR analysis in which two types of topological molecular descriptors (PaDEL and Dragon) and different variable selection methods are utilized. We think that information provided in this work would help in

understanding of molecular structural requirements necessary for developing new potential DHFR inhibitors.

Material and methods

Data set

The data set contains 23 DHFR inhibitors including Imino and Guanine derivatives and classical inhibitors as Methotrexate (MTX). These compounds were synthesized and evaluated as DHFR inhibitors in previous works (Tosso et al. 2013, 2014). We separated the training and validation set compounds applying a procedure that is based on the combination of *k*-means clustering and Kennard–Stone algorithms (MacQueen 1967) implemented in Matlab. Thus, we performed a rational sets partition avoiding the random selection which is sometimes inappropriate or unsuitable (Andrada et al. 2015). To achieve comparable results all developed 2D-QSAR and 3D-QSAR models were obtained using the same training and validation sets.

The IC₅₀ values expressed in micromole (concentration of a compound required to inhibit 50 % of the DHFR activity) were converted to log(1/IC₅₀) and used as dependent variable. The compounds of the training and validation sets and their biological activity values are presented in Table 1.

2D-QSAR analysis

The structures of the 23 DHFR inhibitors were obtained from Molecular Dynamics calculations using the human DHFR as the starting structure (PDB entry code 2W3M) and subsequent optimization into a reduced model of the binding comprising 23 amino acids (Tosso et al. 2013). The optimizations were performed at the PM6 level of theory using the MOPAC2009 program (Stewart 2008).

The optimized structures were used as input in the PaDEL and Dragon software to calculate the topological and 2D-molecular descriptors (Yap 2011; Tetko et al. 2005). A total of 1444 and 762 molecular descriptors were obtained with PaDEL and Dragon software, respectively. To avoid the redundant information the descriptors with a correlation higher than 0.9 were removed from total set.

The most relevant descriptors were chosen through multiple linear regression (MLR) using the enhanced replacement method (ERM) as the molecular descriptor selection approach (Mercader et al. 2011). The ERM is an optimization tool that generates QSAR models by searching an optimal subset of *d* descriptors from a set of *D* descriptors ($d \ll D$) with minimum standard deviation (*S*) of the model (Mercader et al. 2010). This method produces multivariable linear QSAR models close to the full search methods with lower computational cost (Paz et al. 2015;

Table 1 Structure and experimental biological activity of the data set

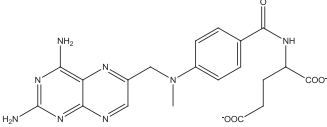
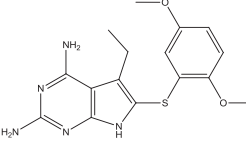
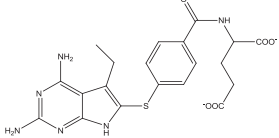
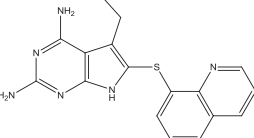
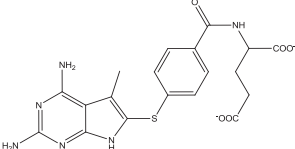
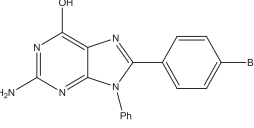
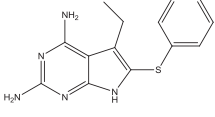
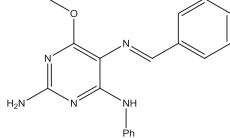
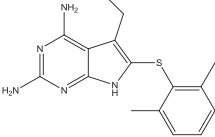
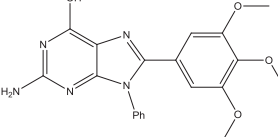
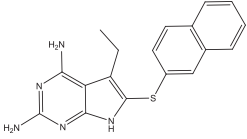
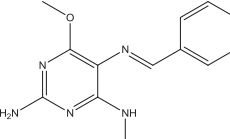
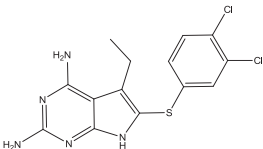
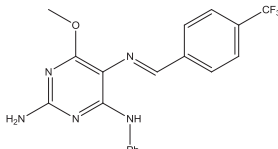
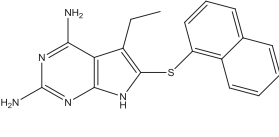
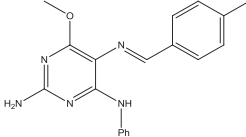
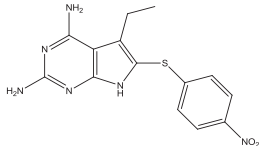
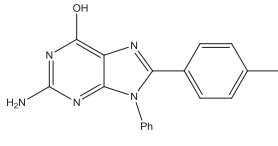
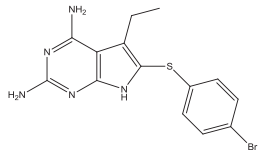
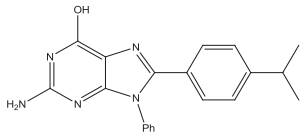
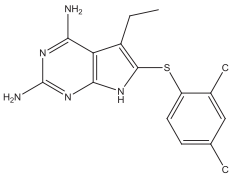
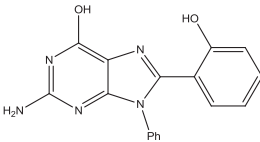
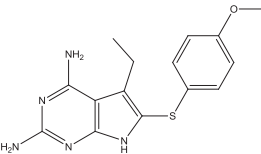
ID	structure	IC50 μ M	$\log(1/IC_{50})$	ID	structure	IC50 μ M	$\log(1/IC_{50})$
1		0.022	1.658	13		5.8	-0.763
2		0.066	1.180	14 ^a		1.4	-0.146
3 ^a		0.21	0.678	15		54.45	-1.736
4		18	-1.255	16		68.01	-1.833
5		32	-1.505	17		49.39	-1.694
6		6.0	-0.778	18		63.05	-1.800
7		2.8	-0.447	19		64.54	-1.810
8		2.9	-0.462	20 ^a		65.98	-1.819
9		5.8	-0.763	21		27.87	-1.445

Table 1 continued

ID	structure	IC50 μM	$\log(1/\text{IC}_{50})$	ID	structure	IC50 μM	$\log(1/\text{IC}_{50})$
10 ^a		2.7	-0.431	22		43.84	-1.642
11		1.4	-0.146	23		77.09	-1.887
12		2.9	-0.462				

^a Compounds of validation set

Garro Martinez et al. 2014). The quality of the models was quantified with the traditional statistic parameters; regression coefficient (R), determination coefficient (R^2), the standard deviation (S), and root-means square error (RMS).

The validation was performed using a set of compounds (validation set), which was selected from the data set through k -means clustering (using $k = 2$ and 3) and Kennard–Stone algorithms. In addition, the optimal QSARs were also validated using the leave-one-out (loo) and the leave-more-out (lmo) cross-validation procedures, generating 100,000 cases of random data removal for lmo, (m is $\approx 20\%$) and y -randomization. The y -randomization involves the interchange of the dependent variable (biological activity) between the compounds. To establish a meaningful validation, we performed 100,000 cases of y -randomization. All the calculations were computed using the algorithms included in Matlab software (Matlab, The MathWorks Inc. 2008).

3D-QSAR method

3D-QSAR study was performed with Open3DQSAR packages developed by Tosco and Balle (2011). Open3DQSAR is free available software for statistical manipulation of the molecular interaction fields (MIF). Recent investigations indicate that results obtained using this method is comparable to CoMFA/CoMSIA (Ghasemi and Shiri 2012).

The 3D structures of the compounds were aligned with Open3DALIGN (Tosco et al. 2011). The alignment was

carried out using all available molecules (N) as possible templates. Therefore, N alignments were produced, each obtained by superimposition of all molecules on the corresponding template compound of each alignment. A score value was computed for each alignment which indicates the quality of the molecular overlap.

The next step to generate a 3D-QSAR model was gathering MIF information for all analyzed compounds. We computed a steric and an electrostatic field. The steric field was based on AMBERFF99 Van der Waals parameters (Wang et al. 2000) and was calculated using a Lennard–Jones' 6–12 potential between the n atoms of the molecule and a sp^3 carbon. Equation 1, while the electrostatic field was based on a point charge model and was obtained through Coulomb interactions between a positively charged probe and the n atoms of the molecules (Eq. 2).

$$E_{\text{VDW}} = \sum_{i=1}^n \left[\frac{A_i}{r_i^{12}} - \frac{B_i}{r_i^6} \right] \quad (1)$$

$$E_{\text{cle}} = k \sum_{i=1}^n \left[\frac{q_i}{r_i^m} \right] \quad (2)$$

In order to carry out the MIF calculation, we generated a grid box with a step size of 1.0 and a 5.0 Å outgap around of the compounds. Then, we performed pre-processing of data to exclude uninformative variables, which only add noise or even can negatively condition the model. The following steps were performed (Tosco et al. 2011):

- Zeroing [sets to zero grid values that are close to zero (value ≤ 0.05)];
- Max/Min cut-off [sets to user-defined maximum/minimum threshold values the grid points lying above or below these boundaries, respectively (Max = 30, Min = -30)];
- Standard deviation cut-off [removes variables having a standard deviation among different objects lower than a user-defined threshold, in order to improve the signal to-noise ratio (value = 0.1)];
- N*-level variable elimination (removes variables assuming only a few different values across the different objects to prevent them from biasing the model).

Once pre-treatment and scaling operations were completed, Partial Linear Square (PLS) models were built examining from 1 to 5 principal components (PC). As well as in 2D-QSAR model, predicted power of the obtained 3D-QSAR models was challenged by an external validation set and cross-validation (loo and lmo).

The predictivity of a 3D-QSAR model can be significantly improved by appropriate variable clustering and selection methods or procedures (Baroni et al. 1993). Herein, we implemented the Fractional Factorial Design (FFD) and UVE-PLS (uninformative variable elimination-partial linear square) variables selection methods. The FFD selection method, described by Baroni et al. (1992), aims at selecting the variables which have the largest effect on predictivity and can operate on both, single variables or groups of variables identified by a previous smart region definition (SRD) run. The UVE-PLS variable selection removes the least informative variables, i.e., those characterized by small PLS pseudo-coefficients (Centner et al. 1996).

We selected the determination coefficient (R^2) and the standard deviation (*S*) as parameters, which suggest the quality of training and validation.

Experimental section

Synthesis of the new DHFR inhibitors

In this paper, two new compounds were synthesized. The synthetic route of the Imino and Guanine derivative compounds is shown in the Fig. 1. The experimental data are available in supplementary material.

General Procedure for the synthesis of imine derivatives

To a solution of triaminopyrimidine, 1 (1 mmol) in 15 mL of methanol, 1 mmol of appropriate aryl aldehyde was added, and this solution was stirred overnight at room temperature. The solid formed was filtered off, washed with

fresh methanol and dried at 50 °C, if necessary it can be recrystallized from MeOH.

2-[(E)-{[2-amino-4-methoxy-6-(phenylamino)pyrimidin-5-yl]imino}methyl]phenol (**24**) 78 % yield; yellow solid; mp 178–180 °C; R_f 0.29 (CHCl_3); IR (ν cm^{-1}): 3462, 3345, 3229, 1632, 1601, 1555, 1120, 1082; ^1H NMR (DMSO- d_6): δ 3.89 (s, 3H), 6.50 (s, 2H), 6.90–6.98 (m, 3H), 7.25–7.29 (m, 3H), 7.70–7.74 (m, 3H), 8.43 (s, 1H), 9.08 (s, 1H), 11.83 (s, 1H). ^{13}C NMR (DMSO- d_6): δ 53.2, 103.5, 116.2, 119.0, 120.2, 121.7, 121.8, 128.3, 129.8, 131.5, 140.3, 157.07, 158.0, 158.6, 159.5, 160.9. MS (*m/z*, %) (assignment, abundance %): 335 (M^+ , 54.8), 334 (M-1, 16.2), 242 (M-93, 100), 215 (M-120, 8.8), 77 (C_6H_5^+ , 11.9). Calculated HRMS for $\text{C}_{18}\text{H}_{17}\text{N}_5\text{O}_2$: 335.1382; found: 335.1380.

Procedure for the synthesis of guanine derivative

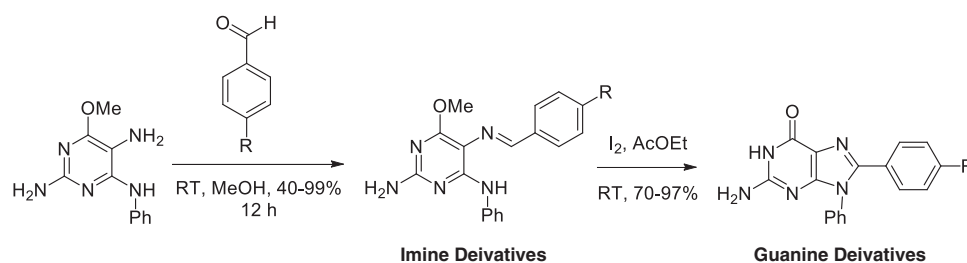
To a solution of triaminopyrimidine, 1 (1 mmol) in 15 mL of methanol, 1 mmol of appropriate aryl aldehyde was added, and this solution was stirred overnight at room temperature. The solid formed was filtered off, washed with fresh methanol and used directly without further purification with 1 mmol of iodine in 20 mL of AcOEt; the reaction mixture was stirred for 24 h at room temperature. The solid formed is filtered off and washed with a solution of NaHCO_3 , then with a solution of sodium thiosulfate and, dried in an oven at 100 °C (see Fig. 1).

2-amino-8-(1,3-benzodioxol-5-yl)-9-phenyl-1,9-dihydro-6H-purin-6-one (**25**) 52 % global yield; beige solid; mp > 300 °C; R_f 0.82 ($\text{CHCl}_3/\text{CH}_3\text{OH}$, 9:1); IR (ν cm^{-1}): 3422, 3186, 1694, 1645, 1232, 1032; ^1H NMR (DMSO- d_6): δ 6.00 (s, 2H), 6.49 (s, 2H), 6.80–6.84 (m, 3H, broad band), 7.33 (s, 1H, broad band), 7.48 (s, 4H, broad band), 10.63 (s, 1H). ^{13}C NMR (DMSO- d_6): δ **101.3, 108.2, 108.3, 122.4, 123.8, 123.9, 128.2, 128.7, 129.4, 135.6, 144.9, 147.0, 147.6, 153.6, 153.7, 156.7**. MS (*m/z*, %) (assignment, abundance %): 347 (M^+ , 100), 346 (M-1, 45), 304 (M-43, 4), 224 (M-123, 13), 77 (C_6H_5^+ , 5). Calculated HRMS for $\text{C}_{18}\text{H}_{13}\text{N}_5\text{O}_3$: 347.1018; found: 347.1015.

Bioassays of the new DHFR inhibitors

The bioassay is based on the capacity of DHFR to catalyze the NADPH-dependent reduction of dihydrofolic acid to tetrahydrofolic acid, such as it is described by Tosso et al. (2013, 2014). The rate of NADPH consumption in the presence of the compound is monitored by the decrease in absorbance at 340 nm (Greedy 1980; Blakley 1995; Costi and Ferrari 2001; Schweitzer et al. 1990; Mathews et al. 1963; Hillcoat et al. 1967). The reactions were carried out in a solution with saturating concentrations of the cofactor (80 μM NADPH) and the substrate (50 μM dihydrofolate),

Fig. 1 The common synthetic route of Imine and Guanine derivatives



50 mM Tris-HCl, 0.001 M 2-mercaptoethanol, and 0.001 M EDTA at pH 7.4 and 30 °C. The enzyme was purchased from Sigma-Aldrich Co. (St. Louis, MO).

Results and discussion

Clustering and sets selection

The training and validation sets were designated through *k*-means clustering and Kennard–Stone algorithms. Dendrogram and silhouette plots were performed using the PaDEL and Dragon molecular descriptors information, Fig. 2. The dendrograms indicate the data set could be clustered according to the type of descriptors, two clusters for PaDEL and three clusters for Dragon. In support of these results, we carried out silhouette plots with *k* = 2 and 3 (number of clusters). The silhouette values near to +1 are points that are very distant from neighboring clusters and values close to -1 indicate points that are probably assigned to the wrong cluster. A value equal to 0 (zero) are up to points that cannot be assigned to any cluster.

The negative values in PaDEL silhouette plots indicate that there exists a compound which cannot be correctly assigned to a specific cluster. The other compounds could be possibly grouped in two cluster or less likely in three clusters.

In contrast, the Dragon silhouette plots confirm the three clusters found in the respective dendrogram only with an unassigned compound.

Then, we used the Kennard–Stone algorithm to find the most representative compounds for each cluster to build a validation set that depicted the whole data set. The selected compounds are: **3**, **10**, **14**, **20** (see Table 1). Thus, we shape a training set containing 19 compounds ($\approx 82\%$) and a validation set with four compounds ($\approx 18\%$).

Development of 2D-QSAR models

The search of predictive QSAR models was carried out using two different families of topological and 2D molecular descriptors: PaDEL and Dragon (Yap 2011; Tetko 2005). In each case, 2D-QSAR regression containing from 1 to 3 of the most representative descriptors were constructed.

All the constructed models comply with the classic semi-empirical “rule of thumb”, which indicates that at least six or seven data points (i.e., compounds) should be present by descriptor (Hansch 1990). These descriptors were selected from the total set using MLR approach. The statistic parameters of the models with lower standard deviation are shown in Table 2.

Both type of descriptors (PaDEL and Dragon) showed an excellent correlation with the biological activity. A smaller number of PaDEL descriptors (two descriptors) were required to achieve a predictive model. However, a significant difference of predictability, which allows us to rule out any of models, is not appreciated. Thus, we selected the PaDEL model with two descriptors and the Dragon model with three descriptors as the most predictive 2D-QSAR models. The mathematical equations are:

PaDEL model

$$\log(1/IC_{50}) = -11.504 + 1.847(\text{maxHCsats}) + 2.042(IC3) \quad (3)$$

$$N = 19, R^2_{\text{train}} = 0.916, S_{\text{train}} = 0.319, R_{\text{loo}} = 0.939, S_{\text{loo}} = 0.379, R_{\text{lmo}} = 0.879, S_{\text{lmo}} = 0.528, S_{\text{rand}} = 0.474$$

Dragon model

$$\log(1/IC_{50}) = -8.885 + 0.0001(\text{SPI}) + 8.490(\text{SIC2}) + 0.961(\text{GATS5p}) \quad (4)$$

$$N = 19, R^2_{\text{train}} = 0.952, S_{\text{train}} = 0.248, R_{\text{loo}} = 0.963, S_{\text{loo}} = 0.306, R_{\text{lmo}} = 0.938, S_{\text{lmo}} = 0.395, S_{\text{rand}} = 0.396$$

The selected models were validated using loo, lmo, external set and *y*-randomization. All the validation regression coefficients (R_{loo} , R_{lmo} , and R_{val}) exceed amply the accepted value of 0.50. Also, we performed 100,000 cases of *y*-randomization obtaining standard deviation (*S*) values greater (PaDEL; 0.474 and Dragon; 0.396) than those found, when true calibration was considered (PaDEL;

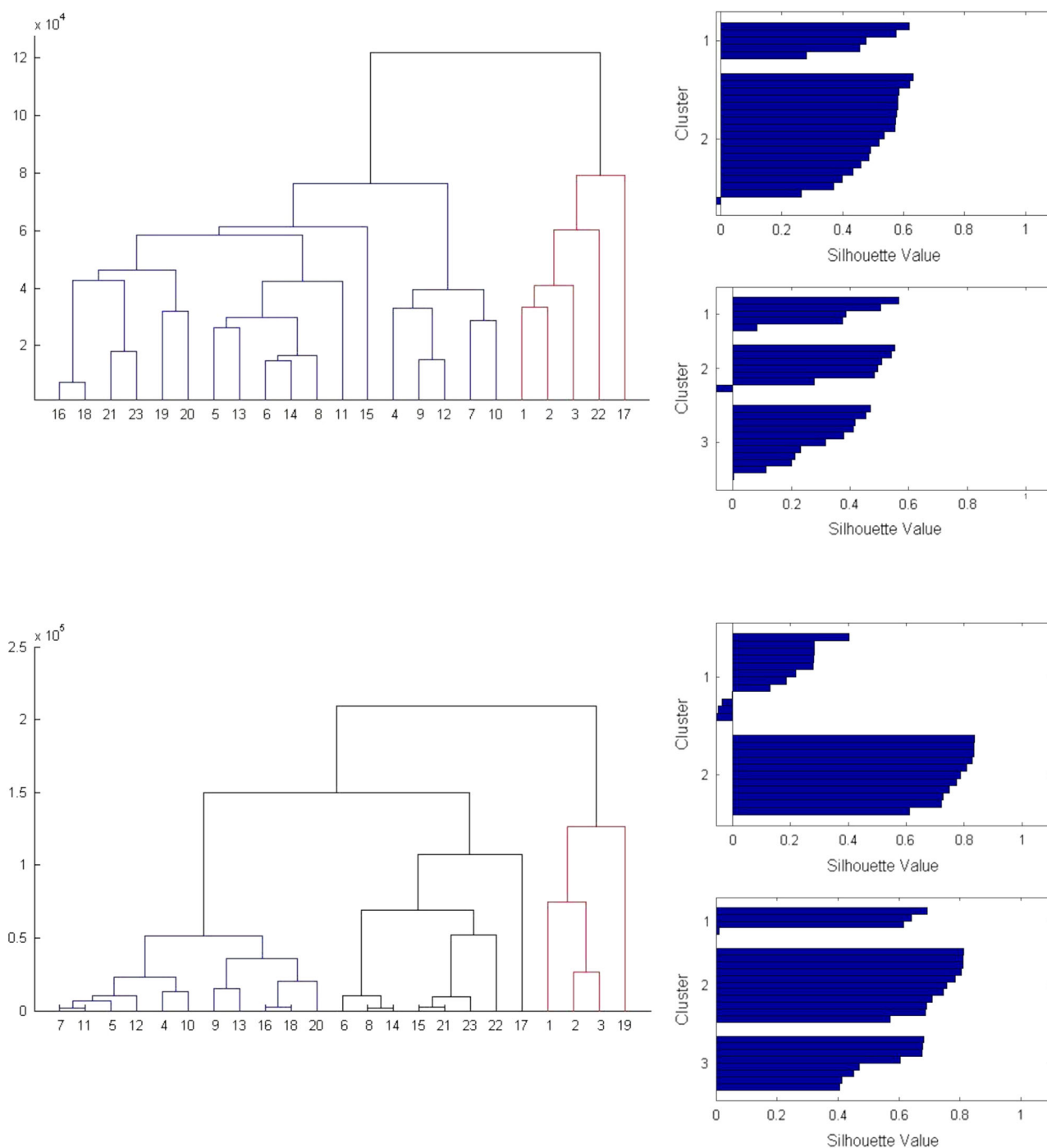


Fig. 2 Dendrogram and silhouette plots. (*Top*): Using PaDEL descriptors. (*Bottom*): Using Dragon descriptors

0.319 and Dragon; 0.248). The y -randomization results prove that the best models were not found by chance.

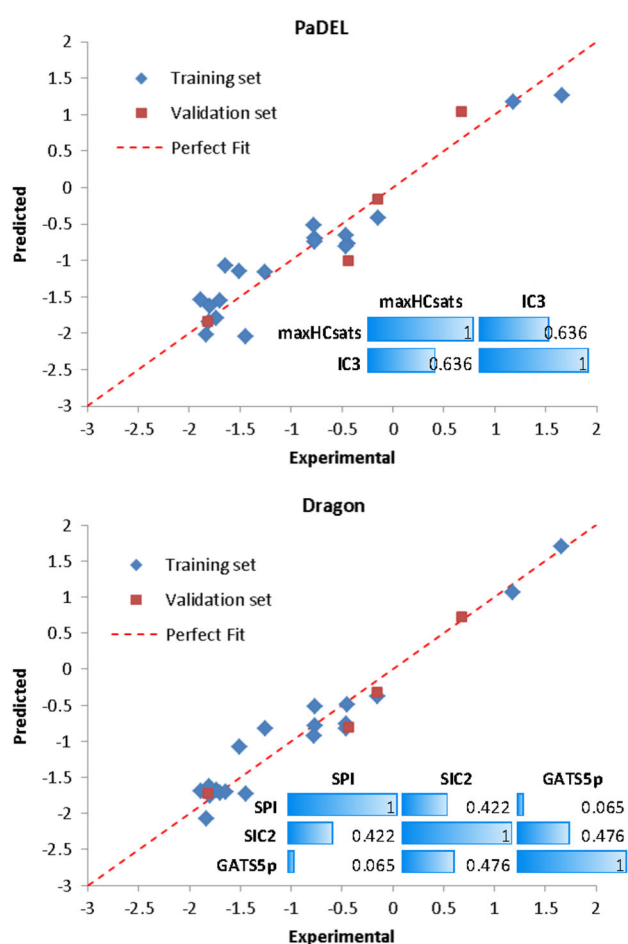
The plots shown in Fig. 3 illustrate the correlation between the predicted and experimental $\log(1/IC_{50})$. As can be appreciated, the training and validation data set values are close to the perfect fit, showing the high predictive power of the models. The predicted $\log(1/IC_{50})$ values are shown in Table 3.

Molecular descriptors analysis

The Table 4 lists a brief description of selected MDs. In the PaDEL model, the most relevant molecular descriptors related to the DHFR inhibitory activity were maxHCsats and IC3 (Hall and Kier 1995; Todeschini and Consonni 2009), with correlation of 63 % (0.636), Fig. 3 (top). For Dragon model, the most significant molecular

Table 2 Statistic parameters of the 2D-QSAR models

N ^o	Descriptors	R ² _{train}	S _{train}	RMS	R ² _{val}	S _{val}	RMS
PaDEL models							
1	nHBDOn_Lipinski	0.804	0.476	0.452	0.803	0.428	0.331
2 ^a	maxHCsats, IC3	0.916	0.319	0.294	0.806	0.475	0.300
3	ATSC4p, MATS4c, SsNH2	0.984	0.140	0.126	0.645	1.236	0.553
Dragon models							
1	IC2	0.892	0.352	0.334	0.927	0.378	0.293
2	SPI, IC2	0.926	0.300	0.276	0.930	0.518	0.328
3 ^a	SPI, SIC2, GATS5p	0.952	0.248	0.222	0.963	0.447	0.200

^a Selected models**Fig. 3** Predicted vs. experimental log(1/IC₅₀) plots and descriptors correlation matrix. (Top): PaDEL model, Eq. (1). (Bottom): Dragon model, Eq. (2)

descriptors were SPI, SIC2, and GATS5p (Todeschini and Consonni 2009). The maximum correlation (47%) was found between SIC2 and GATS5p (Fig. 3 (bottom)). The numerical values of the molecular descriptors are listed in Table 1S of supplementary material.

Table 3 Experimental and predicted values of biological activity for the data set

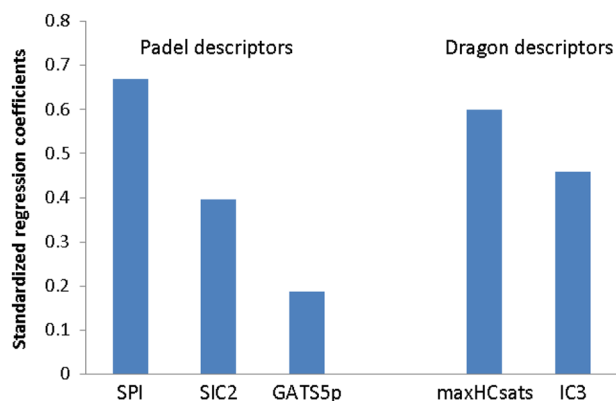
ID	Exp. log (1/IC ₅₀)	2D-QSAR		3D-QSAR		
		PaDEL	Dragon	PLS	FDD	UVE-PLS
1	1.658	1.265	1.706	1.452	1.319	1.590
2	1.180	1.174	1.059	1.029	1.231	1.199
3 ^a	0.678	1.038	0.724	0.854	0.803	0.706
4	-1.255	-1.163	-0.819	-0.734	-0.619	-1.143
5	-1.505	-1.145	-1.079	-1.022	-1.099	-1.401
6	-0.778	-0.515	-0.919	-0.709	-0.842	-0.558
7	-0.447	-0.776	-0.486	-0.585	-0.945	-0.447
8	-0.462	-0.656	-0.826	-0.629	-0.604	-0.626
9	-0.763	-0.745	-0.518	-0.457	-0.438	-0.731
10 ^a	-0.431	-1.012	-0.812	-0.564	-0.519	-0.740
11	-0.146	-0.414	-0.376	-0.488	-0.391	-0.191
12	-0.462	-0.806	-0.765	-0.671	-0.944	-0.486
13	-0.763	-0.690	-0.788	-0.891	-1.035	-0.733
14 ^a	-0.146	-0.158	-0.328	-0.974	-0.933	-0.986
15	-1.736	-1.796	-1.683	-1.360	-1.163	-1.732
16	-1.833	-2.024	-2.074	-1.952	-1.967	-1.890
17	-1.694	-1.551	-1.719	-1.648	-1.542	-1.834
18	-1.800	-1.631	-1.749	-1.810	-1.736	-1.807
19	-1.810	-1.843	-1.631	-1.696	-1.558	-1.846
20 ^a	-1.819	-1.843	-1.728	-1.847	-1.804	-1.545
21	-1.445	-2.049	-1.733	-1.971	-1.884	-1.510
22	-1.642	-1.075	-1.705	-1.541	-1.559	-1.585
23	-1.887	-1.539	-1.685	-1.939	-1.824	-1.775

^a Compounds of validation set

The standardization of the regression coefficients allowed assigning a greater importance to the molecular descriptors (Draper and Smith 1981). The coefficients values are presented in the Fig. 4. For the descriptors in the PaDEL model, the absolute standardized regression coefficients were: maxHCsats = 0.598 and IC3 = 0.457. Thus, the maxHCsats [an electrotopological state atom-type descriptor (Hall and Kier 1995)] was found as the most important PaDEL descriptor. These topological descriptors arise from the electronic environment of each atom due to its intrinsic electronic properties and the influence of other atoms in the molecule, and they are computed to parameterize key properties such as hydrogen bonds, molecular polarity, etc. In addition, atom-type and the group-type electrotopological state sums are computed for a number of atoms and functional groups (Hall and Kier 1995). Specifically, maxHCsats is associated to hydrogen atoms on carbons bonded to saturated carbon atoms, such as the ones present in compounds 1, 2, and 3 structures, Table 1. The positive sign of this descriptor in the Eq. (3) indicates that presence

Table 4 Brief descriptors of MDs of PaDEL and Dragon models

Descriptors	Type	Description
PaDEL models		
nHBDon_Lipinski	PaDEL HBond donor count	Number of hydrogen bond donors (using Lipinski's definition: any OH or NH. Each available hydrogen atom is counted as one hydrogen bond donor)
maxHCsats	Electrotopological state atom type	Maximum atom-type H E-State
SsNH2	Electrotopological state atom type	Count of atom-type H E-State: -NH2+
IC3	Information content	Information Content index (neighborhood symmetry of 3-order)
ATSC4p	Autocorrelation	Centered Broto-Moreau autocorrelation—lag 4/weighted by polarizabilities
MATS4c	Autocorrelation	Moran autocorrelation—lag 4/weighted by charges
Dragon models		
IC2	Information content	Information Content index (neighborhood symmetry of 2-order)
SIC2	Information content	Structural Information Content index (neighborhood symmetry of 2-order)
SPI	Topological indices	Superpendentic index
GATS5p	2D autocorrelations	Geary autocorrelation of lag 5 weighted by polarizability

**Fig. 4** Standardized regression coefficient for PaDEL and Dragon descriptors

of this class of hydrogen atoms in the molecule improve the biological activity.

IC3 is an Information Content index (neighborhood symmetry of 3-order) descriptor. It is a topological descriptor which indicates the complexity of graph based on the 3rd order neighborhood of vertices. It describes the connectivity and branching in a molecule and can be related to molecular shape and symmetry. The relative number of rings in the fragments can also be related to molecular shape. A higher positive value of the information content would be beneficiary to the activity (Todeschini and Consonni 2009). Similarly, the compounds **1**, **2**, and **3** present the highest values of this descriptor, see Table S1.

In the Dragon model, the absolute standardized regression coefficients for each descriptor were: SPI = 0.669, SIC2 = 0.396, GATS5p = 0.187. In this case, the main descriptor was SPI [superpendentic index, topological descriptors (Todeschini and Consonni 2009)]. The value of this descriptor changes significantly with a small change in the branching of a molecule. The high sensitivity towards branching, exceptionally high discriminating power coupled with extremely low degeneracy, suggest that superpendentic indexes are the promising tools for isomer discrimination (Gupta and Singh 1999). The second descriptors SIC2 is the Structural Information Content index (neighborhood symmetry of 2-order). These types of descriptors are modifications of information content index (IC_n) and are also related to the complexity of the molecular graph. In agreement with the PaDEL model, compounds **1**, **2**, and **3**, present high values SPI and SIC2 due to branching of their molecular structures. GATS5p (Geary autocorrelation of lag 5 weighted by atomic polarizabilities, autocorrelation descriptors) is associated to the presence of polarizable atoms at a topological distance of 5. This descriptor has a little influence on the activity. The positive sign of the three Dragon descriptors in the Eq. 4 indicates that $\log(1/IC_{50})$ values are directly related to these descriptors.

Development of 3D-QSAR model

The 3D-QSAR models were built based on the same training and validation sets utilized in the 2D-QSAR

analysis. The PLS, FFD, and UVE-PLS variable selection methods were used to correlate chemical structures to DHFR inhibitory activity (Baroni et al. 1993, 1992; Centner et al. 1996).

The molecular alignment is the key procedure to develop 3D-QSAR models (Li et al. 2013). The ligand-based alignment was carried out using the 19 compounds of training set as possible templates. According to the score values, the best overlap is obtained when compound number 6 is used as template, the Fig. 5.

We constructed five 3D-QSAR models (from 1 to 5 principal component) for each variable selection method (PLS, FFD, and UVE-PLS). The statistical results are summarized in the Table 5. The PLS exhibited satisfactory results when three principal components (PC=3) were considered ($R^2_{\text{train}} > 0.9$ and $R^2_{\text{val}} > 0.6$). The calibration and external validation statistic parameters in FFD and UVE-PLS methods were similar to PLS. However, we found an improvement in the internal validation quality (loo and lmo) when these two approaches were applied.

The model with PC=3 was the most predictive PLS model with values of $R^2_{\text{loo}} > 0.5$ for internal validation (loo and lmo) and $R^2_{\text{val}} = 0.694$ for the external validation. In FFD, two principal components (PC=2) were adequate to build a predictive model. This presented acceptable validation parameter statistic: $R^2_{(\text{loo}, \text{lmo})} > 0.7$ for the loo and lmo and R^2_{val} similar to PLS (0.693). When we used UVE-PLS, the best model was found with PC=5. This variable selection method presents validation parameters ($R^2_{\text{loo}} = 0.820$, $R^2_{\text{lmo}} = 0.801$, and $R^2_{\text{val}} = 0.626$) that demonstrate the robustness and the excellent predictive power of the derived model.

The predicted $\log(1/IC_{50})$ values are listed in the Table 3, while Fig. 6 shows the correlation between the predicted and experimental values of $\log(1/IC_{50})$ obtained for the most predictive models.

The Fig. 7 shows the steric and electrostatic contour maps of the aligned compounds (Fig. 7a, c) and the same type of map obtained for the most active compound (Fig. 7b, d). In the steric contour map, the interaction regions are represented by green and yellow contours, Fig. 7a, b. The bulky groups, near the green regions, increase the activity but negatively affect when they are located near the yellow regions. The electrostatic maps represent the regions where an increase in the positive charge (blue regions) and/or negative charge (red regions) is favorable to the biological activity (Fig. 7c, d).

The contours for the steric field showed an important green region, which involves the phenyl group and its substituents. The presence of bulky groups in this region could improve the activities of the compounds; the most active compounds (1, 2, 3, 8, 11, and 14) have bulky substituents on this region. In contrast, the compounds with low

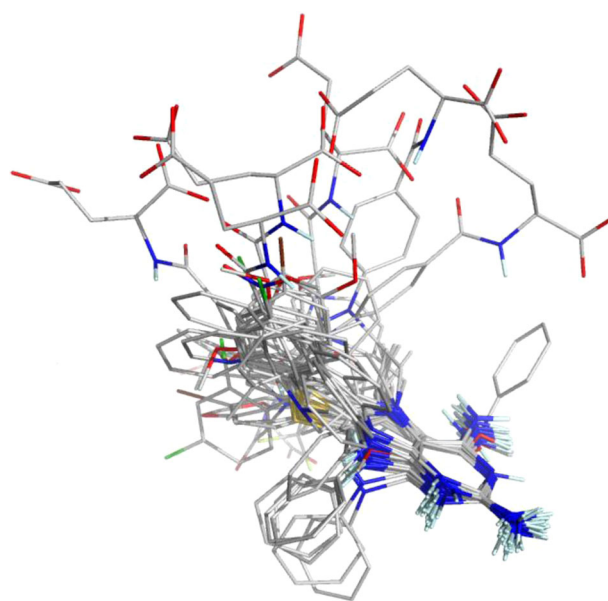


Fig. 5 The molecular alignment of the 23 compounds

Table 5 Statistic results of 3D-QSAR models

PC	S_{train}	R^2_{train}	S_{loo}	R^2_{loo}	S_{lmo}	R^2_{lmo}	S_{val}	R^2_{val}
PLS								
1	0.846	0.314	0.983	0.074	0.973	0.091	0.793	0.336
2	0.428	0.824	0.644	0.602	0.655	0.588	0.609	0.212
3	0.268	0.931	0.689	0.545	0.694	0.535	0.379	0.694
4	0.177	0.970	0.731	0.488	0.728	0.489	0.446	0.577
5	0.120	0.986	0.753	0.457	0.748	0.459	0.409	0.645
FFD								
1	0.819	0.356	0.947	0.140	0.930	0.156	0.803	0.369
2	0.329	0.896	0.528	0.733	0.539	0.720	0.381	0.693
3	0.228	0.950	0.596	0.659	0.603	0.650	0.381	0.692
4	0.160	0.976	0.661	0.581	0.655	0.588	0.340	0.754
5	0.117	0.987	0.680	0.557	0.670	0.568	0.413	0.637
UVE-PLS								
1	0.824	0.349	0.952	0.132	0.945	0.141	0.793	0.336
2	0.308	0.909	0.497	0.764	0.504	0.756	0.541	0.379
3	0.180	0.969	0.471	0.787	0.474	0.784	0.434	0.600
4	0.120	0.986	0.439	0.815	0.457	0.798	0.455	0.561
5	0.089	0.992	0.433	0.820	0.454	0.801	0.420	0.626

activity (compounds 16, 18, 20, and 23) have bulky groups in the yellow regions.

According to the electrostatic contour maps of Fig. 7c, d, there is a noticeable influence of the electropositive groups on the activity. The presence of electropositive substituents in the blue region could increase the biological activity. This is evidenced by comparing the activity of the

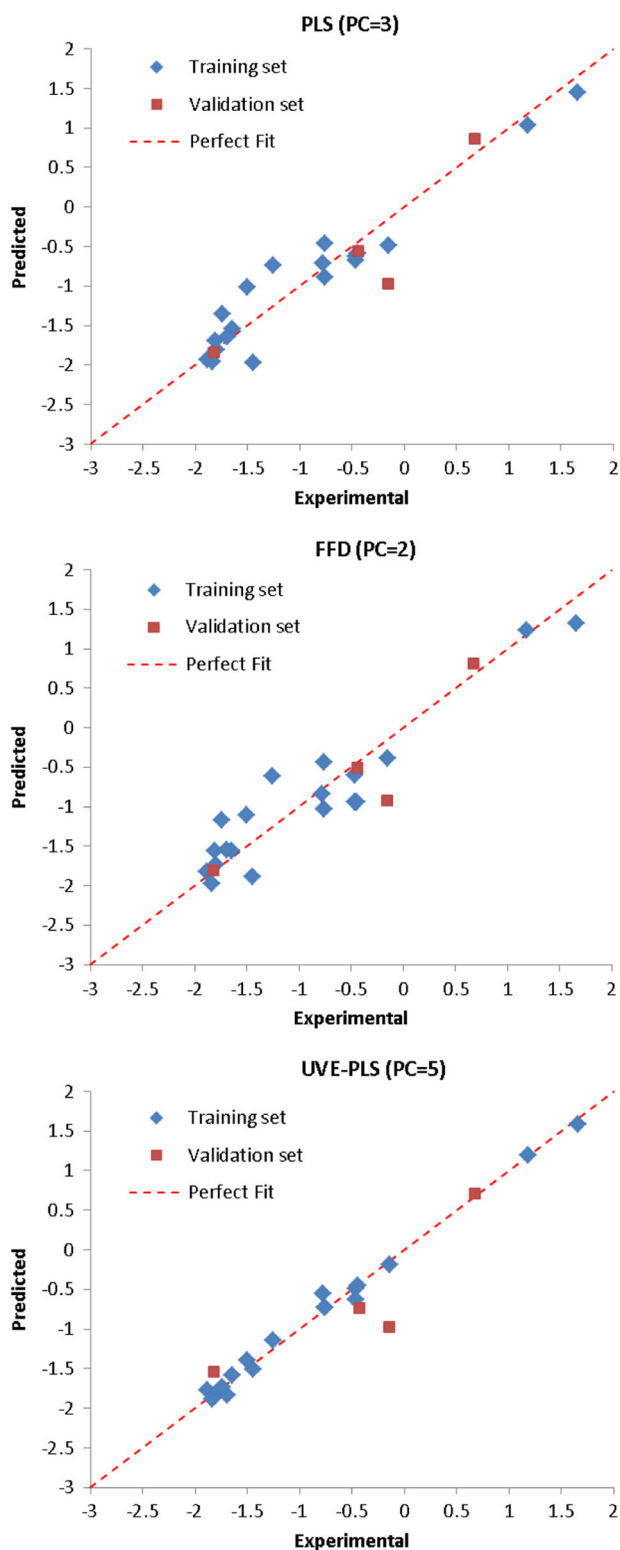


Fig. 6 Predicted vs. experimental $\log(1/IC_{50})$ plots. (Top): PLS model using PC = 3. (Middle): FFD model using PC = 2. (Bottom): UVE-PLS model using PC = 5

compounds with $-NH_2$ in the guanine group with respect to the compounds with $-OH$ or $-OCH_3$ as substituents.

Proposal, bioassay and molecular docking of new DHFR inhibitors

The information provided from the structure-activity relationship studies (2D-QSAR and 3D-QSAR) was very helpful to design and propose new compounds as DHFR inhibitor. Two new compounds (**24** and **25**) were synthesized and their activities were experimentally evaluated (52.89 and 20.81 μM , respectively). The general procedure for the synthesis and bioassay was showed in the experimental section. The biological activities of the compounds **24** and **25** are in a complete agreement with the predicted values by the QSAR models, providing an additional support and an experimental corroboration for such studies. The Table 6 presents the structure, experimental and predicted activities values.

In addition, we propose three new structures (**26**, **27**, and **29**) as possible potent DHFR inhibitors. These were designed with a molecular structure similar to the most active compounds maximizing the maxHCsats, IC3 and SPI descriptors values and considering the effect of size and electronegativity of the substituents. All the QSAR models predicted the $\log(1/IC_{50})$ values that indicate these compounds could have an important biological activity. The IC_{50} mean values (including all QSAR models) of compounds **26**, **27**, and **28** are 0.13, 0.16, and 0.06 μM , respectively, which are similar to the most active compound of the data set. The structure and the predicted activity are show in the Table 7.

The inhibitor-enzyme interaction was compared to MTX by means of molecular docking study. The structure of human DHFR (PDB entry code 2W3M) and Autodock Vina (Trott and Olson 2008) software were used for docking analysis. The results indicate that the three new structures bind to the active site of DHFR similarly to MTX (score: 9.9 kcal/mol). The Fig. 8 illustrates the molecular docking of the three compounds and MTX in the active site of DHFR. The 3D-QSAR models predicted that electropositive substituents located in the blue region (Fig. 7) would increase inhibitor activity. This is in agreement with the docking complexes where the orientation of ligands point electropositive groups towards electrotronegative residues of the active site (see GLU30 in Fig. 8). Moreover, electronegative substituents, located in red region of Fig. 7, establish favorable interactions with ARG70 (electropositive residue). The implementation of QSAR models and molecular docking about the three compounds suggests that they could be used as lead compounds.

Fig. 7 The steric and electrostatic contour maps. *Green* and *yellow* represent sterically favored and disfavored regions, respectively. *Blue* and *red* represent electrically favored and disfavored regions. **a** and **c**: Contour maps based on the training set compounds. **b** and **d**: Contour maps based on the most active compound (color figure online)

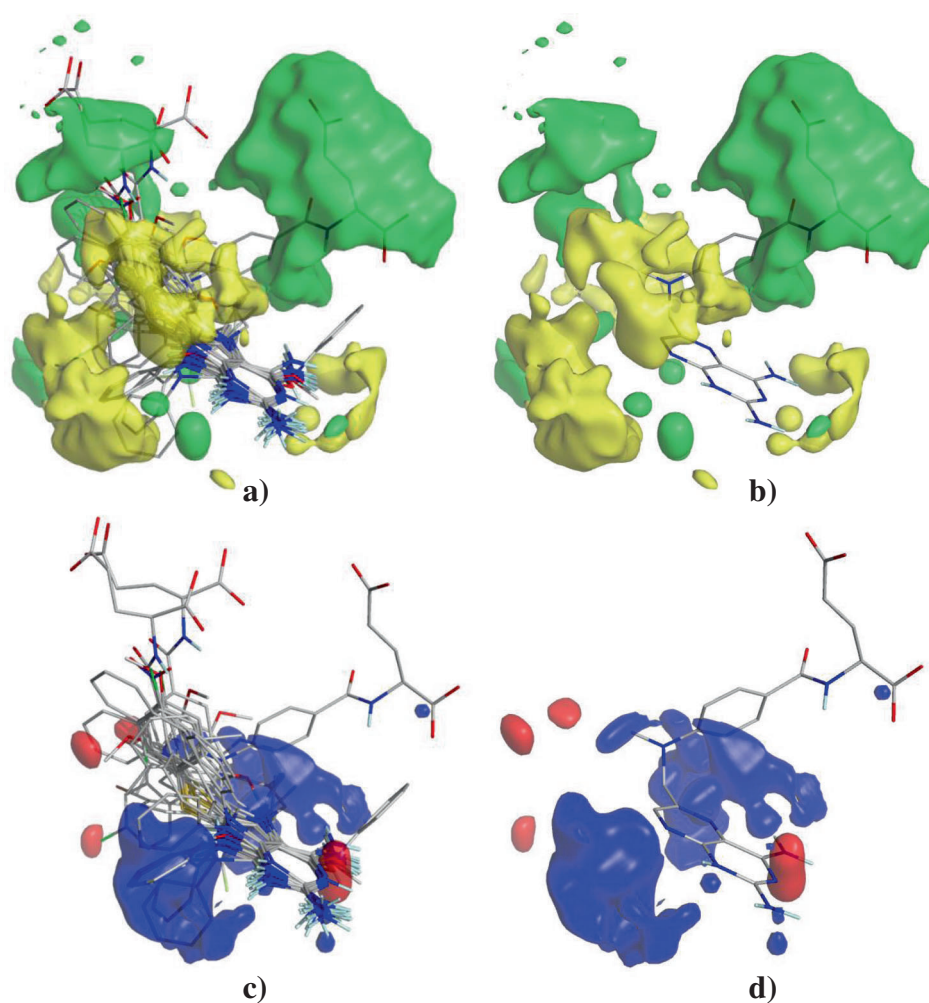


Table 6 Experimental and predicted values of biological activity for two new compounds

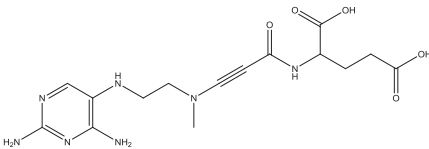
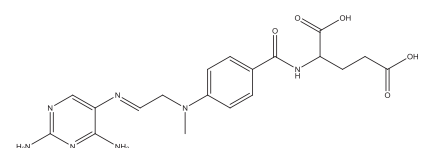
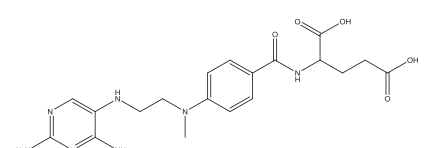
ID	Structure	Exp.log(1/IC ₅₀)	2D-QSAR		3D-QSAR		
			PaDEL	Dragon	PLS	FFD	UVE-PLS
24		-1.723	-1.393	-1.607	-1.732	-1.571	-1.592
25		-1.318	-1.217	-1.191	-1.582	-1.707	-1.401

Conclusion

In this manuscript, we performed 2D-QSAR and 3D-QSAR analysis applied to a series of DHFR inhibitors. In the 2D-

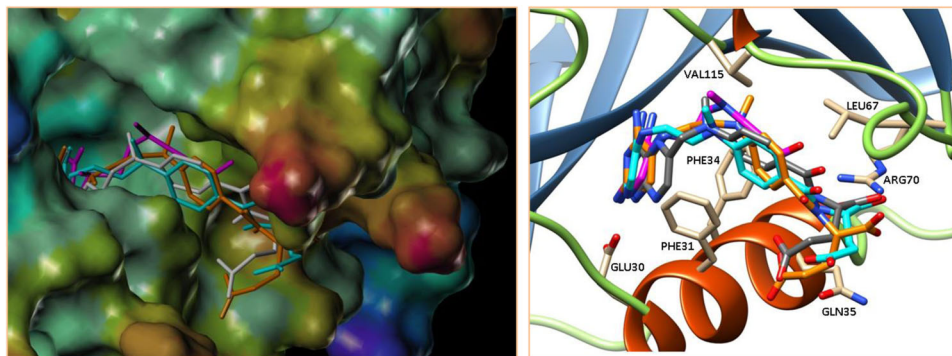
QSAR, we used two types of molecular descriptors that provided different information to the models: PaDEL and Dragon. Simple topological and 2D descriptors (maxHCsats, IC3, SPI, SIC2, and GATS5p) were adequate to obtain

Table 7 Predicted values of biological activity expressed as $\log(1/IC_{50})$ and IC_{50} (μM) between parentheses

ID	Structure	2D-QSAR		3D-QSAR			Score (kcal/mol)
		PaDEL	Dragon	PLS	FFD	UVE-PLS	
26		0.996 (0.101)	0.795 (0.160)	0.910 (0.123)	0.985 (0.104)	0.747 (0.179)	7.4
27		1.183 (0.066)	1.270 (0.054)	0.754 (0.176)	0.584 (0.261)	0.611 (0.245)	8.9
28		1.238 (0.058)	1.229 (0.059)	1.147 (0.071)	1.215 (0.061)	1.176 (0.067)	8.9

Docking score of the three proposed compounds

Fig. 8 Molecular docking of the compounds **26** (magenta), **27** (cyan), **28** (orange), and MTX (gray) (color figure online)



models with excellent predictive power. The 3D-QSAR analysis was developed using different selection approaches (PLS, FDD, and UVE-PLS). The results helped to explain that the biological activity is affected by the size and electronegativity of the substituents. The statistic parameters of the all developed models exceed with amplitude the accepted value of the calibration and validation steps.

The models were experimentally corroborated with two new compounds, **24** and **25**. These were synthesized and their experimental activities were correctly predicted by all QSAR models. In addition, three new compounds (**26**, **27**, and **28**) with possible high DHFR inhibitory activity are proposed in this investigation.

We think that the information provided in this report could be a useful tool for a better understanding of structural requirements of substituents to develop new potential DHFR inhibitors.

Acknowledgments This work was supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de San Luis (UNSL), Universidad de Jaén and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain). We thank "Centro de Instrumentación Científico-Técnica" of Universidad de Jaén and the staff for the data collection.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

- Agrawal VK, Sohgaora R, Khadikar PV (2002) QSAR studies on biological activity of piritrexim analogues against pc DHFR. *Bioorg Med Chem* 10:2919–2926
- Andrada MF, Vega-Hissi EG, Estrada MR, Garro Martinez JC (2015) Application of k-means clustering, linear discriminant analysis

- and multivariate linear regression for the development of a predictive QSAR model on 5-lipoxygenase inhibitors. *Chemometr Intel Lab* 143:122–129
- Baroni M, Clementi S, Cruciani G, Costantino G, Riganelli D, Oberrauch E (1992) Predictive ability of regression models. Part II: Selection of the best predictive PLS model. *J Chemom* 6:347–56
- Baroni M, Costantino G, Cruciani G, Riganelli D, Valigi R, Clementi S (1993) Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant Struct-Act Rel* 12:9–20
- Beierlein JM, Deshmukh L, Frey KM, Vinogradova O, Anderson AC (2009) The solution structure of *Bacillus anthracis* dihydrofolate reductase yields insight into the analysis of structure-activity relationships for novel inhibitors. *Biochemistry* 48:4100–4108
- Beierlein JM, Frey KM, Bolstad DB, Pelphrey PM, Joska TM, Smith AE, Priestley ND, Wright DL, Anderson AC (2008) Synthetic and crystallographic studies of a new inhibitor series targeting *Bacillus anthracis* dihydrofolate reductase. *J Med Chem* 51:7532–7540
- Blakley RL (1995) Eukaryotic dihydrofolate reductase. *Adv Enzymol Relat Areas Mol Biol* 70:23–102
- Centner V, Massart DL, de Noord OE, de Jong S, Vandeginste BM, Sterna C (1996) Elimination of uninformative variables for multivariate calibration. *Anal Chem* 68:3851–8
- Champness JN, Achari A, Ballantine SP, Bryant PK, Delves CJ, Stammers DK (1994) The structure of *Pneumocystis carinii* dihydrofolate reductase to 1.9 Å resolution. *Structure* 2:915–924
- Costi MP, Ferrari S (2001) Update on antifolate drugs targets. *Curr Drug Targets* 2:135–166
- De la Torre JM, Nogueras M, Borkowski EJ, Suvire FD, Enriz RD, Cobo J (2014) Easy synthesis of new series of pteridine analogs: di- and tetra- hydroxypyrimido[4,5-d]pyrimidines via 5-pyrimidinecarbaldehydes. *Arkivoc* 5:42–63
- Debnath B, Vishnoi SP, Sa B, Jha T (2003) QSAR Study on Some Dihydrofolate Reductase Inhibitors, *Internet Electron. J Mol Des* 2:128–136
- Draper NR, Smith H (1981) *Applied regression analysis*. 2nd Edition, John Wiley & Sons, New York
- Gangjee A, Lin X, Biondo LR, Queener SF (2010) CoMFA analysis of tgdHFR and rldHFR based on antifolates with 6-5 fused ring system using the all-orientation search (AOS) routine and a modified cross-validated r(2)-guided region selection (q(2)-GRS) routine and its initial application. *Bioorg Med Chem* 18:1684–1701
- Garro Martinez JC, Vega-Hissi EG, Andrada MF, Duchowicz PR, Torrens F, Estrada MR (2014) Lacosamide derivatives with anticonvulsant activity as carbonic anhydrase inhibitors. Molecular modeling, docking and QSAR analysis. *Curr Comput Aided Drug Des* 10:160–167
- Ghasemi J, Shiri F (2012) Molecular docking and 3D-QSAR studies of falcipain inhibitors using CoMFA, CoMSIA, and Open3DQSAR. *Med Chem Res* 21:2788–2806
- Gready JE (1980) Dihydrofolate reductase: binding of substrates and inhibitors and catalytic mechanism. *Adv Pharmacol Chemother* 17:37–102
- Gupta S, Singh M (1999) Superpendent Index: A Novel Topological Descriptor for Predicting Biological Activity. *J Chem Inf Comput Sci* 39:272–277
- Hall LH, Kier LB (1995) Electrotological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J Chem Inform Comput Sci* 35:1039–1045
- Hansch C (1990) *Comprehensive drug design*. Pergamon Press, New York, NY
- Hillcoat BL, Nixon PF, Blakley RL (1967) Effect of substrate decomposition on the spectrophotometric assay of dihydrofolate reductase. *Anal Biochem* 21:178–189
- Klon AE, Heroux A, Ross LJ, Pathak V, Johnson CA, Piper JR, Borhani DW (2002) Atomic structures of human dihydrofolate reductase complexed with NADPH and two lipophilic antifolates at 1.09 Å and 1.05 Å resolution. *J Mol Biol* 320:677–693
- Li X, Hilgers M, Cunningham M, Chen Z, Trzoss M, Zhang J, Kohnen L, Lam T, Creighton C, G C K, Nelson K, Kwan B, Stidham M, Brown-Driver V, Shaw KJ, Finn J (2011) Structure-based design of new DHFR-based antibacterial agents: 7-aryl-2,4-diaminoquinazolines. *Bioorg Med Chem Lett* 21:5171–5176
- Li X, Ye L, Shi W, Liu H, Liu C, Qian X, Zhu Y, Yu H (2013) In silico study on hydroxylated polychlorinated biphenyls as androgen receptor antagonists. *Ecotoxicol Environ Saf* 92:258–64
- MacQueen J (1967) *Some methods for classification and analysis of multivariate observations*. University of California Press, Berkeley, CA, p 281–297
- Manchester J, Czerminski R (2008) SAMFA: simplifying molecular description for 3D-QSAR. *J Chem Inf Model* 48:1167–1173
- Mathews CK, Scrimgeour KG, Huennekens FM (1963) Dihydrofolate reductase: *Methods enzymol* 6:364–368
- Mattioni BE, Jurs PC (2003) Prediction of dihydrofolate reductase inhibition and selectivity using computational neural networks and linear discriminant analysis. *J Mol Graph Model* 21:391–419
- Mercader AG, Duchowicz PR, Fernandez FM, Castro EA (2010) Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories. *J Chem Inf Model* 50:1542–1548
- Mercader AG, Duchowicz PR, Fernandez FM, Castro EA (2011) Advances in the replacement and enhanced replacement method in QSAR and QSPR theories. *J Chem Inf Model* 51:1575–1581
- Nammalwar B, Bunce RA, Berlin KD, Bourne CR, Bourne PC, Barrow EW, Barrow WW (2012) Synthesis and biological activity of substituted 2,4-diaminopyrimidines that inhibit *Bacillus anthracis*. *Eur J Med Chem* 54:387–396
- Olivella M, Marchal A, Nogueras M, Sanchez A, Melguizo M, Raimondi M, Zacchinoc S, Giannini F, Cobo J, Enriz RD (2012) Structure-activity relationship study of nitrosopyrimidines acting as antifungal agents. *Bioorg Med Chem* 20:6109–6122
- Paz P, Vega-Hissi E, Andrada M, Estrada M, Garro Martinez J (2015) Quantitative structure activity relationship and binding investigation of N-alkyl glycine amides as inhibitors of Leukotriene A4 hydrolase. *Med Chem Res* 24:496–504
- Quiroga J, Trilleras J, Abonía R, Insuasty B, Nogueras M, Cobo J, de la Torre JM (2009) 2,4-Aminopyrimidine-5-carbaldehydes as intermediates in a Friedländer type synthesis of 7-arylpyrido[2,3-d]pyrimidines. *Arkivoc* 2009:9–27
- Schweitzer BI, Dicker AP, Bertino JR (1990) Dihydrofolate reductase as a therapeutic target. *FASEB J* 4:2441–2452
- Sharma M, Chauhan PM (2012) Dihydrofolate reductase as a therapeutic target for infectious diseases: opportunities and challenges. *Future Med Chem* 4:1335–1365
- Stewart JJ (2008) *MOPAC2009*. S.C. Chemistry, Colorado Springs, CO
- Sutherland JJ, Weaver DF (2004) Three-dimensional quantitative structure-activity and structure-selectivity relationships of dihydrofolate reductase inhibitors. *J Comput Aided Mol Des* 18:309–31
- Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin VA, Radchenko EV, Zefirov NS, Makarenko AS, Tanchuk VY, Prokopenko VV (2005) Virtual computational chemistry laboratory-design and description. *J Comput Aid Mol Des* 19:453–63
- The MathWorks Inc. Matlab 7.0. Available online: <http://www.mathworks.com> (accessed on April 2016)

- Todeschini R, Consonni V (2009) *Molecular Descriptors for Chemoinformatics (Methods and Principles in Medicinal Chemistry)*; Wiley-VCH:Weinheim, Germany
- Tosco P, Balle T (2011) Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. *J Mol Model* 17:201–208
- Tosco P, Balle T, Shiri F (2011) Open3DALIGN: an open-source software aimed at unsupervised ligand alignment. *J Comput Aided Mol Des* 25:777–783
- Tosso RD, Andujar SA, Garro AD, Suvire FD, Garro Martinez JC, Enriz RD (2014) The electron density obtained from QTAIM analysis acting as a strong molecular descriptor. A molecular modeling study performed in DHFR inhibitors. *WATOC*, p 371. Chile
- Tosso RD, Andujar SA, Gutierrez L, Angelina E, Rodriguez R, Nogueras M, Baldoni H, Suvire FD, Cobo J, Enriz RD (2013) Molecular modeling study of dihydrofolate reductase inhibitors. Molecular dynamics simulations, quantum mechanical calculations, and experimental corroboration. *J Chem Inf Model* 53:2018–2032
- Trott O, Olson AL (2008) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput. Chem* 31:455–461
- Verma J, Khedkar VM, Coutinho EC (2010) 3D-QSAR in drug design--a review. *Curr Top Med Chem* 10:95–115
- Wang J, Cieplak P, Kollman PA (2000) How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J Comput Chem* 21:1049–1074
- Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 32:1466–1474