A DFT-based QSAR Study on Inhibition of Human Dihydrofolate Reductase

Sedat Karabulut[a], Natalia Sizochenko[b], Adnan Orhan[c], Jerzy Leszczynski[b]

[a] *Department of Chemistry, Faculty of Science and Literature, Balikesir University, Balikesir, 10145, Turkey*

[b] *Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, Jackson, 39217-0510 Mississippi, USA*

[c] *School of Medicine, Department of Obstetrics and Gynaecology, Uludag University, Gorukle Kampusu, Bursa, 16120, Turkey*

corresponding author: Sedat Karabulut, email: sedatk@balikesir.edu.tr, address: Faculty of Science and Literature, Department of Chemistry, Balikesir University, Balikesir, 10145, Turkey, phone: +905301180284

Abstract

Diaminopyrimidine derivatives are frequently used as inhibitors of human dihydrofolate reductase, for example in treatment of patients whose immune system are affected by human immunodeficiency virus. Forty-seven dicyclic and tricyclic potential inhibitors of human dihydrofolate reductase were analyzed using the quantitative structure-activity analysis supported by DFT-based and DRAGON-based descriptors. The developed model yielded an RMSE deviation of 1.1 a correlation coefficient of 0.81. The prediction set was characterized by $R^2 = 0.60$ and RMSE = 3.59. Factors responsible for inhibition process were identified and discussed. The resulting model was validated via cross validation and Y-scrambling procedure.

From the best model, we found several mass-related descriptors and Sanderson electronegativity-related descriptors that have the best correlations with the investigated inhibitory concentration. These descriptors reflect results from QSAR studies based on characteristics of human dihydrofolate reductase inhibitors.

Keywords: Dihydrofolate reductase, diaminopyrimidine, DFT, descriptors, QSAR, QSARins.

1. Introduction

Among various chemical compounds responsible for biochemical processes folates are different by their unique metabolism. They have been recognized as an attractive and effective chemotherapeutic target. Folate metabolism plays an essential role in nucleic acid synthesis, methionine regeneration, shuttling and redox reactions of one carbon units required for normal metabolism, regulation and organization [2]. Several antifolates have been developed to attack the key enzymes in the folate cycle. The folates, which are included into the group of B vitamins family are composed of an aromatic pteridine ring attached through a methylene group to p-aminobenzoate and a glutamate residue [1]. In this important group of enzymes, the enzyme dihydrofolate reductase was identified in 1958 as the target for methotrexate and also trimethoprim (TMP), which is a potent synthetic antibacterial agent [3]. Its essential function is to induce dihydrofolate to tetrahydrofolate within the thymidylate synthesis cycle. Dihydrofolate reductase (DHFR) along with thymidylate synthase (TS) is included in the part of the cycle responsible for the synthesis of thymidylate monophosphate (dTMP), which is involved in DNA biosynthesis and cell replication [4]. The first DHFR inhibitors were methotrexate, trimethoprim and pyrimethamine [5].

Inhibitory effects have been at the forefront of many quantitative structure-activity relationship (QSAR) and pharmacophore studies [6-14]. Back to early 80's, well-known *Corwin Hansch* analyzed the structure-activity relation of triazines inhibiting DHFR [6]. In 20 years *King et al.* published article about trimethoprim analogues binding to DHFR [7]. *Andrea* and *Kalayeh* used neural networks to investigate DHFR inhibition by 256 5-phenyl-3,4-diamino-6,6-dimethyldihydrotriazines [8]. *So* and *Richards* also applied the neural networks to find structure-activity relationships between structural features of the 2,4-diamino-5-(substituted-benzyl)pyrimidines with inhibitory activity of towards DHFR [9]. Later, pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors was aimed to create suitable pharmacophore hypothesis [10]. Hist published two papers in this area: he investigated the inhibition of dihyd DHFR by pyrimidines [11] and by triazines [12]. Scientists still have a keen interest to computational analysis of DHFR inhibitors. Recently (in 2016), *Singla et al.* applied QSAR approach to study the DHFR inhibition by triazine–benzimidazoles with 4-fluoroaniline substitution [13].

Diaminopyrimidine inhibitors of DHFR such as trimethoprim, pyrimethamine, trimetrexate and piritrexim (**Figure 1**) represent frequently used drugs in treatment of patients whose immune system are affected by human immunodeficiency virus (HIV) [15,16].

Several scientific groups have been performed synthetic studies to develop new inhibitors of *P. carinii* and *T. gondii* DHFR inhibitors, which could work without additional sulfa- drugs or leucovorin (as it works usually) [17]. Hundreds of diaminopyrimidine antifolates were tested as inhibitors of DHFR [18, 19]. In 2001 R. G. Nelson et al. published one of the largest databases on lipophilic polycyclic diaminopyrimidines as inhibitors of *C. parvum* and human DHFR enzymes [17].
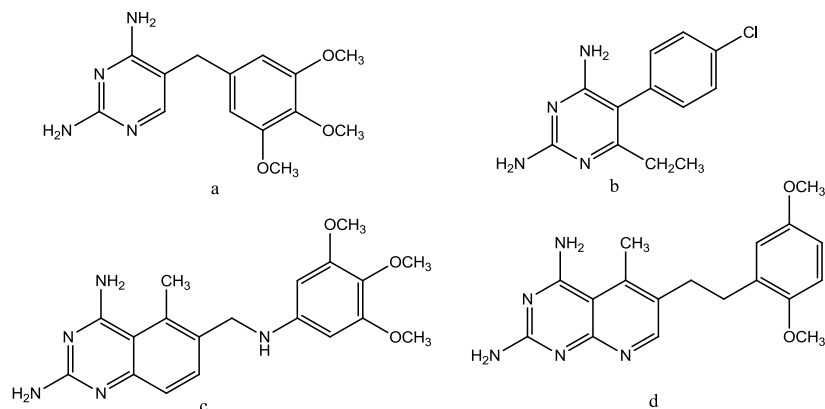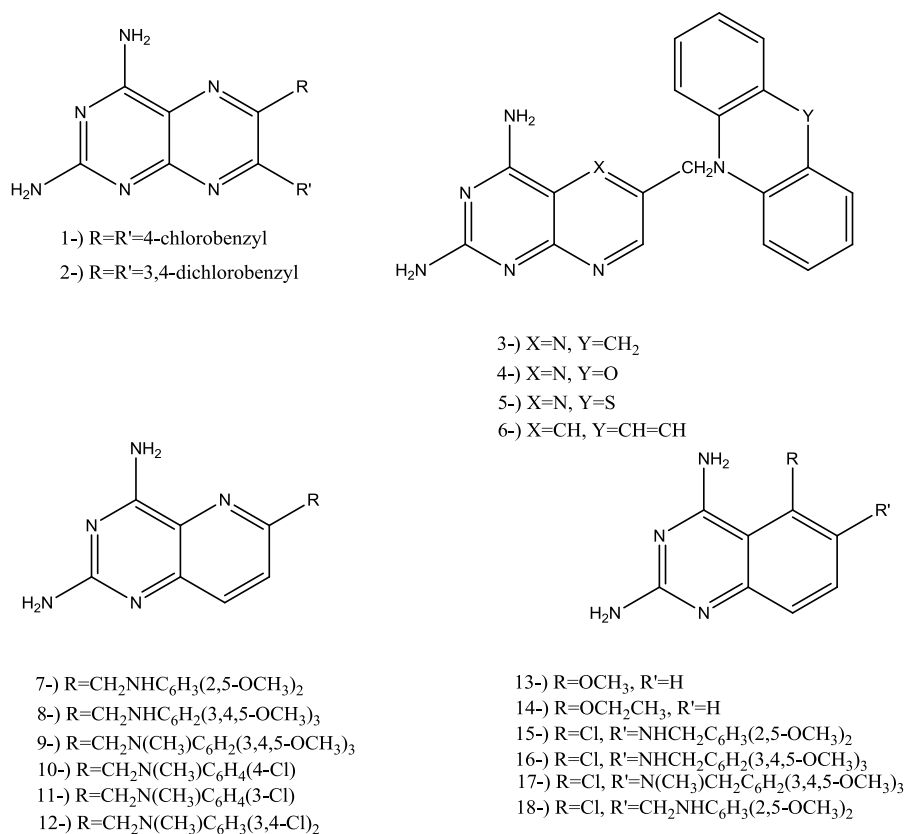
Figure 1. Molecular structures of trimethoprim (a), pyrimethamine, trimetrexate and pritrexim (d)

Despite the fact that the inhibition activity of diaminopyrimidine derivatives has been widely studied, there are no QSAR studies presented in literature. To fill this gap in the present study a QSAR model for 47 dicyclic and tricyclic diaminopyrimidine derivatives – human DHFR inhibitors has been developed and used to describe investigated phenomena.

2. Materials and methods

2.1 Dataset

A dataset of 47 potential inhibitors of human DHFR was gathered from literature [17]. In **Table 1** and **Figure 2** chemical structures and experimental data is presented. The inhibitory concentration values were expressed as micromolar (μM) units (**Table 1**). Each titration was performed twice, and the mean DHFR inhibitory activity was plotted against the inhibitor concentration.

1-) R=R'=4-chlorobenzyl
2-) R=R'=3,4-dichlorobenzyl

3-) X=N, Y=CH$_2$
4-) X=N, Y=O
5-) X=N, Y=S
6-) X=CH, Y=CH=CH

7-) R=CH$_2$NHC$_6$H$_3$(2,5-OCH$_3$)$_2$
8-) R=CH$_2$NHC$_6$H$_2$(3,4,5-OCH$_3$)$_3$
9-) R=CH$_2$N(CH$_3$)C$_6$H$_2$(3,4,5-OCH$_3$)$_3$
10-) R=CH$_2$N(CH$_3$)C$_6$H$_4$(4-Cl)
11-) R=CH$_2$N(CH$_3$)C$_6$H$_4$(3-Cl)
12-) R=CH$_2$N(CH$_3$)C$_6$H$_3$(3,4-Cl)$_2$

13-) R=OCH$_3$, R'=H
14-) R=OCH$_2$CH$_3$, R'=H
15-) R=Cl, R'=NHCH$_2$C$_6$H$_3$(2,5-OCH$_3$)$_2$
16-) R=Cl, R'=NHCH$_2$C$_6$H$_2$(3,4,5-OCH$_3$)$_3$
17-) R=Cl, R'=N(CH$_3$)CH$_2$C$_6$H$_2$(3,4,5-OCH$_3$)$_3$
18-) R=Cl, R'=CH$_2$NHC$_6$H$_3$(2,5-OCH$_3$)$_2$

Figure 2. Molecular structures of investigated inhibitors

Table 1: IUPAC names, modeling status and IC_{50} values of 47 molecules

#	Name	Status	IC_{50} (µM)
1	6,7-bis(4-chlorobenzyl)pteridine-2,4-diamine	Prediction	5.7
2	6,7-bis(3,4-dichlorobenzyl)pteridine-2,4-diamine	Training	3.5
3	6-(acridin-10(9H)-ylmethyl)pteridine-2,4-diamine	Training	0.56
4	6-(10H-phenoxazin-10-ylmethyl)pteridine-2,4-diamine	Training	0.23
5	6-(10H-phenothiazin-10-ylmethyl)pteridine-2,4-diamine	Training	0.81
6	6-(5H-dibenzo[b,f]azepin-5-ylmethyl)pyrido[2,3-d]pyrimidine-2,4-diamine	Training	1.4
7	6-{[(2,5-dimethoxyphenyl)amino]methyl}pyrido[3,2-d]pyrimidine-2,4-diamine	Prediction	0.83
8	6-{[(3,4,5-trimethoxyphenyl)amino]methyl}pyrido[3,2-d]pyrimidine-2,4-diamine	Prediction	0.49
9	6-{[methyl(3,4,5-trimethoxyphenyl)amino]methyl}pyrido[3,2-d]pyrimidine-2,4-diamine	Training	0.0089
10	6-{[(4-chlorophenyl)(methyl)amino]methyl}pyrido[3,2-d]pyrimidine-2,4-diamine	Training	0.31
11	6-{[(3-chlorophenyl)(methyl)amino]methyl}pyrido[3,2-d]pyrimidine-2,4-diamine	Training	0.027
12	6-{[(3,4-dichlorophenyl)(methyl)amino]methyl}pyrido[3,2-d]pyrimidine-2,4-diamine	Training	0.0004
13	5-methoxyquinazoline-2,4-diamine	Training	2.8
14	5-ethoxyquinazoline-2,4-diamine	Training	0.75
15	5-chloro-N^6-(2,5-dimethoxybenzyl)quinazoline-2,4,6-triamine	Training	0.0039
16	5-chloro-N^6-(3,4,5-trimethoxybenzyl)quinazoline-2,4,6-triamine	Training	0.0013
17	5-chloro-N^6-methyl-N^6-(3,4,5-trimethoxybenzyl)quinazoline-2,4,6-triamine	Prediction	0.01
18	5-chloro-6-{[(2,5-dimethoxyphenyl)amino]methyl}quinazoline-2,4-diamine	Training	0.01
19	6-ethyl-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	9.4
20	6-tert-butyl-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.022
21	6-(3-thienylmethyl)-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.59
22	6-(2-methoxybenzyl)-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.094
23	6-(3-methoxybenzyl)-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Prediction	0.23
24	6-(4-methoxybenzyl)-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.29
25	6-(2,5-dimethoxybenzyl)-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Prediction	0.074
26	6-(3,4-dimethoxybenzyl)-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.19
27	6-(3,4,5-trimethoxybenzyl)-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Prediction	0.16
28	6-(2-methylbenzyl)-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.094
29	6-(3-methylbenzyl)-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.38
30	6-[3-(trifluoromethyl)benzyl]-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.19
31	6-[3-(trifluoromethoxy)benzyl]-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.15
32	6-[4-(trifluoromethoxy)benzyl]-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.31
33	6-(3,4-dichlorobenzyl)-5,6,7,8-tetrahydroquinazoline-2,4-diamine	Training	0.094
34	6-(2,5-dimethoxyphenyl)-5-methylthieno[2,3-d]pyrimidine-2,4-diamine	Training	0.98
35	6-(2,5-dimethoxybenzyl)-5-methylthieno[2,3-d]pyrimidine-2,4-diamine	Prediction	0.64
36	5-methyl-6-(3,4,5-trimethoxybenzyl)thieno[2,3-d]pyrimidine-2,4-diamine	Training	3
37	6-(2-bromo-3,4,5-trimethoxybenzyl)-5-methylthieno[2,3-d]pyrimidine-2,4-diamine	Training	1.6
38	6-[2-(2-bromo-3,4,5-trimethoxyphenyl)ethyl]thieno[2,3-d]pyrimidine-2,4-diamine	Training	7.3
39	6-(2-bromo-3,4,5-trimethoxybenzyl)-5,6,7,8-tetrahydropyrido[4,3-d]pyrimidine-2,4-diamine	Training	2.8
40	9-chlorobenzo[f]quinazoline-1,3-diamine	Training	0.012
41	9-methoxybenzo[f]quinazoline-1,3-diamine	Training	0.17
42	4-[[(2,4-diaminopteridin-6-yl)methyl](methyl)amino]-N-(-1-methylbutyl)benzamide	Training	0.6
43	4-[[(2,4-diaminopteridin-6-yl)methyl](methyl)amino]-N-[3-hydroxy-1-(hydroxymethyl)propyl]benzamide	Training	1.9
44	N-1-adamantyl-4-[[(2,4-diaminopteridin-6-yl)methyl](methyl)amino]benzamide	Training	0.77
45	6-({methyl[4-(morpholin-4-ylcarbonyl)phenyl]amino}methyl)pteridine-2,4-diamine	Prediction	1.9
46	N-cyclohexyl-4-[[(2,4-diaminopteridin-6-yl)methyl](methyl)amino]benzamide	Training	0.58
47	4-{4-[[(2,4-diaminopteridin-6-yl)methyl](methyl)amino]benzoyl}piperazine-1-carboxylate	Training	0.81

2.2 DFT modeling

Quantum chemical techniques are widely used in combination with QSAR modeling [20]. Density functional theory (DFT)-related techniques have been advocated to be quite useful for such purposes. The quantum chemical calculations were performed using the molecular modeling package GAUSSIAN 09 [21]. The 3D structures of all molecules were constructed using the GaussView 5.0. Structural energy minimizations were performed at b3lyp/6-31+g(d,p) level. An energy scan for all considered compounds has been performed by rotation of the selected single bonds. Step sizes were selected as 60 degrees and energy calculation has been performed for five steps at AM1 level for all molecular structures to gain the best input geometry before geometry optimization using DFT approach. No imaginary frequencies were observed for optimized structures of all 47 molecules.

The extracted DFT descriptors are as follows: free Gibbs energy (E(RB3LYP), dipole moment (Dip), sum of electronic and zero-point energies (ZPE), sum of electronic and thermal Energies (TE), sum of electronic and thermal enthalpies (TEnt), sum of electronic and thermal free energies (TFE), HOMO energy (HOMO), LUMO energy (LUMO), hardness, softness, electronegativity and electrophilicity. Hardness, softness, electronegativity and electrophilicity have been calculated by known calculation schema. All energy values were defined in atomic unit, without any conversion.

2.3. DRAGON descriptors

E-DRAGON package was applied to calculate more than 1600 descriptors [22]. Highly correlated descriptors have been eliminated and more than 600 remaining descriptors were combined with DFT descriptors. DRAGON software has calculated a variety of molecular descriptors derived from different types molecular representations (from 0D to 3D) [23, 24]. It allows choosing the molecular descriptors which are more suitable for each specific study. Descriptors, selected to generate QSAR model will be described more specifically.

2.3 QSAR model development and validation

There are several steps in QSAR analysis [25]. The datasets used in such process is a combination of characteristics (so-called descriptors) that should be correlated with the experimental activities. Aforementioned descriptors are calculated using different quantum chemical, mathematical, or physical method. At the next step, a big pool of descriptors obtained during the first step is reduced using specific procedures of descriptor selection. Then, initial pool of target species is split between training and prediction sets to ensure the quality of developed model. Using statistical techniques and measures, the QSAR model is built, validated, and interpreted.

The initial dataset was divided into training and test sets [26]. These sets were selected manually (every 5th selected as test molecule while ascending activity), and structurally diverse molecules covering a wide range of activities were included in both sets [27]. Therefore, we had 38 compounds in training set

and 9 compounds in prediction set (**Table 1**). We utilized the training set to generate QSAR model, and then validated our results using the test set.

QSAR models were calculated in QSARINS by means of the multiple linear regression (MLR) method [28]. MLR is a multivariate technique that generates a multiple linear combination of descriptors [29]. The model fitting was evaluated using the coefficient of determination R^2, and a modified form R^2_{adj}, root-mean square error RMSE, and Q^2_{loo} – validation coefficient of leave-one out for both training and prediction sets [30]. Y-randomization (R^2_y) procedure was performed for training set [31]. All these criteria describe how well the model reproduces the data used.

Y-scrambling procedure was performed to check the descriptors used in the model [31]. To perform Y-scrambling, the bioactivities are randomized and the new model is created. In this way one could test the validity of developed QSAR model and ensure that the selected descriptors are not random. Y-scrambling models are built using the same descriptors as the original model. Statistical parameters – R^2 and Q^2 values – are also calculated for scrambled models. This process is performed from 5 to 500 times to ensure that random model is truly random. The statistical parameters for such type of model should be as low as possible.

PCA modeling is the widely used multivariate exploratory techniques [32]. Each Principal Components (PCs) detects the internal relations inside the set of objects (chemical structures), decreasing the original dimensionality of the data. This reduction transformed the original matrix to a new one, composed by PCs orthogonal to each other.

The last parameter evaluated was the applicability domain [33]. Activity of the entire universe of chemicals cannot be predicted even by a robust and validated QSAR model. Therefore, each compound should be checked using space of accessible descriptors. The prediction is valid only if the compound is within the applicability domain of the model. To visualize the applicability domain of a QSAR model, the Williams plot—the plot of standardized cross-validated residuals versus leverage values was used [33].

3. Results and discussion

The best model performance for training set was: $R^2 = 0.81$, $R^2_{adj} = 0.78$, RMSE = 0.83, $R^2_y = 0.18$, while results for validation were: $Q^2_{loo} = 0.68$ RMSE = 1.1 $Q^2_y = 0.35$. The prediction set was characterized by following statistics: $R^2 = 0.60$, RMSE = 3.59. Plot of experimentally determined (observed) versus predicted values of developed model is presented in **Figure 3.** Black line represents perfect agreement between observed and predicted values. Selected descriptors are presented in **Table 2**.

The plot of the two PC scores (**Supplementary Information, plot S1**) showed that the data are well clustered. The two principal components (PC1 and PC2) explained 55.81% (25.29% + 30.52%) of the total variance in the data. The score plot of PC1 versus PC2 (**SI, plot S2**) is expected to provide a reasonably accurate representation of the whole space defined by the descriptors.

Table 2. Descriptors from developed QSAR model

Descriptor	Type	Value
Intercept	-	26.1768
ARR	aromatic ratio	-11.7825
GATS5m	Geary autocorrelation of lag 5 weighted by mass	2.0467
G3m	3rd component symmetry directional WHIM index / weighted by mass	-50.4002
P2e	2nd component shape directional WHIM index / weighted by electronegativity	10.2443
E2e	2nd component accessibility directional WHIM index / weighted by electronegativity	4.0286
HATS7u	leverage-weighted autocorrelation of lag 7 / unweighted	17.1069
R6e	R autocorrelation of lag 6 / weighted by electronegativity	-19.8151
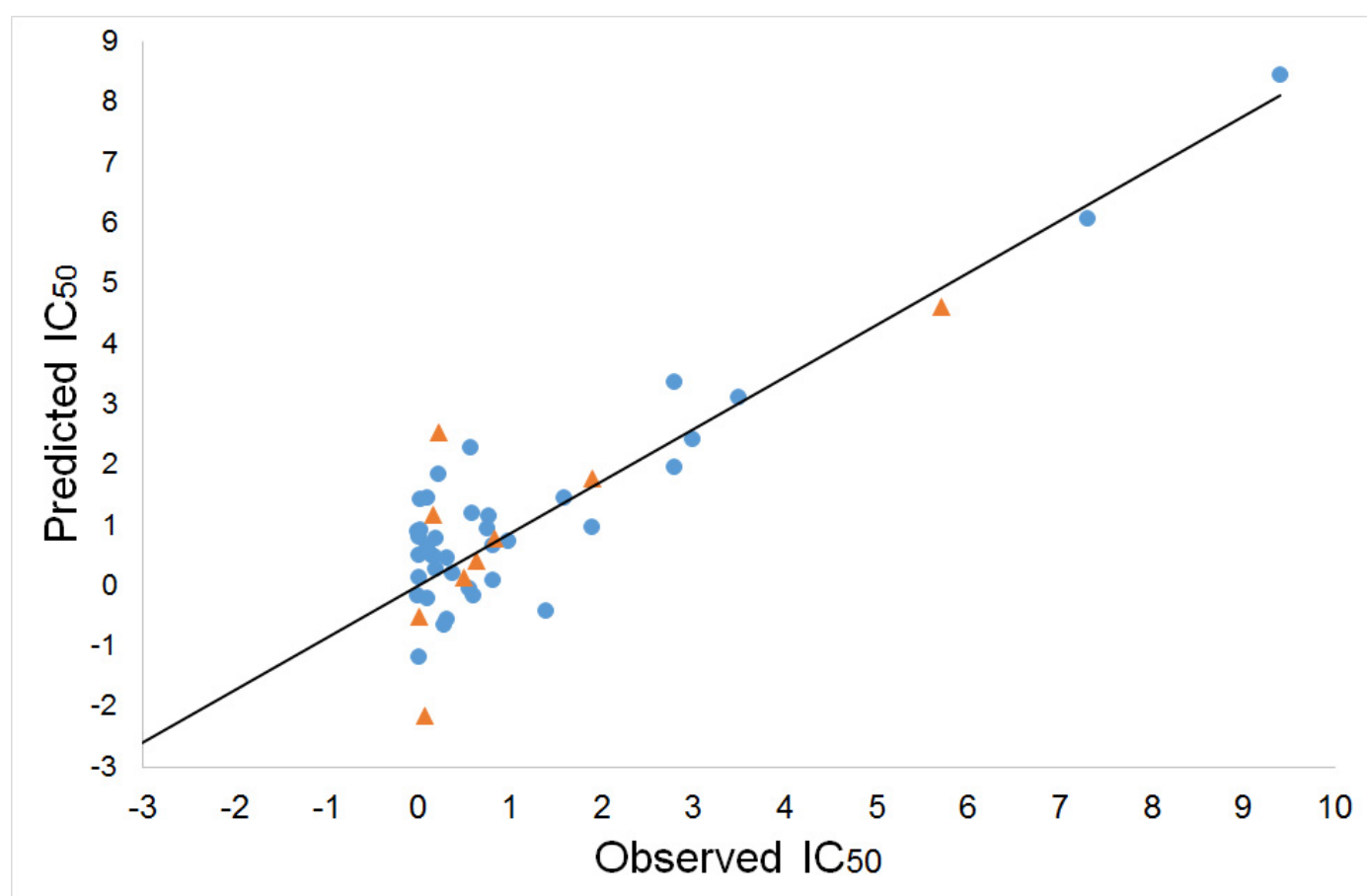
Figure 3. Experimental versus predicted data from MLR modeling

The results of Y-scrambling procedure shown that all scrambled models had statistically less significant values that the developed model (**SI, plot S3**). Plot of domain applicability demonstrates the relationship between the leverages and standardized residuals (**Figure 4**). As one can see, all studied compounds lie within the allowed boundary.
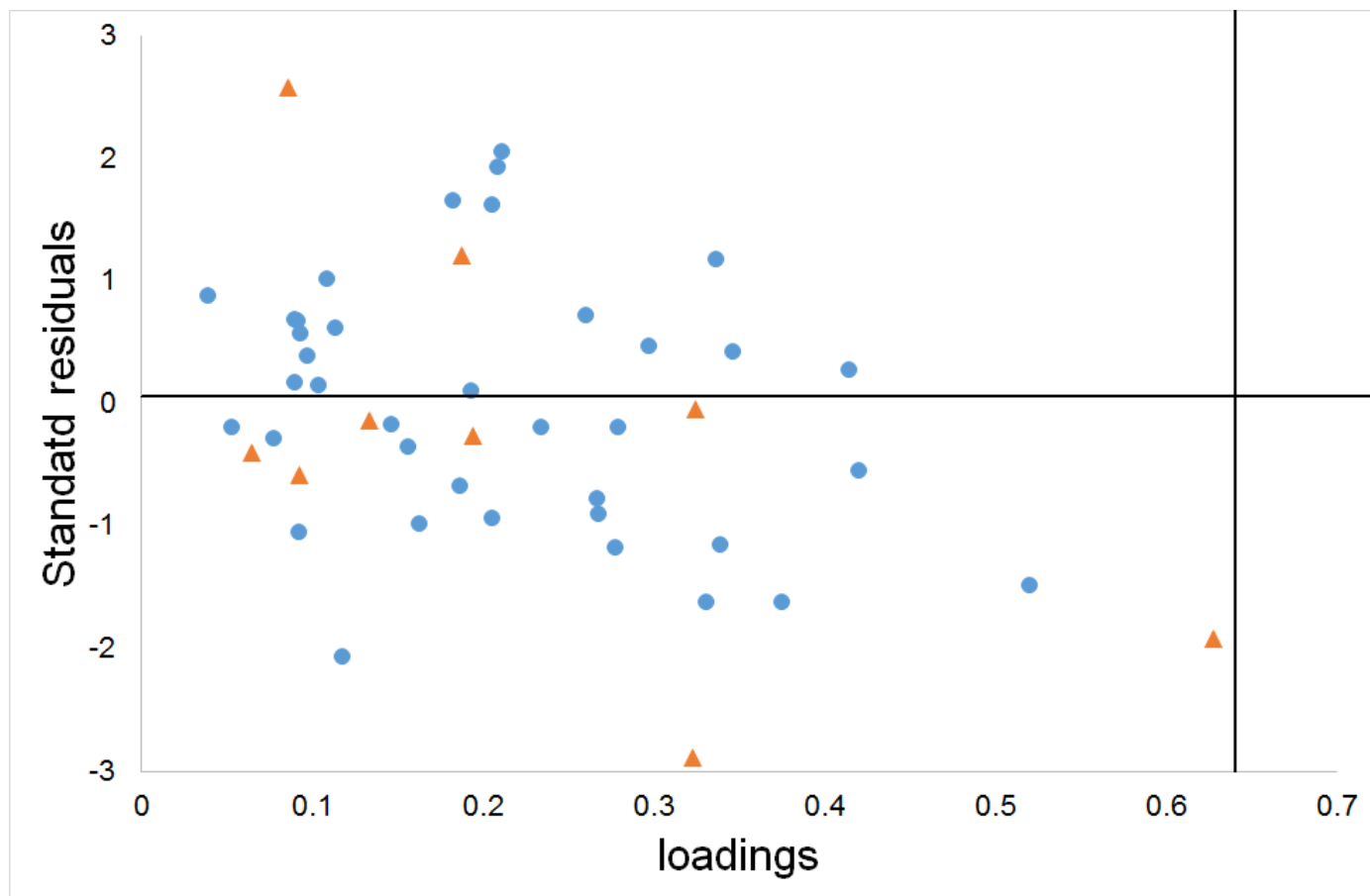
**Figure 4.** Williams plots of QSAR model

Let us discuss the mechanistic interpretation of the developed QSAR model. The selected descriptors allow providing elucidation of the studied phenomena. For instance, aromatic ratio (ARR) is related to π-π stacking during interactions between diaminopyrimidine derivatives with target protein. GATS5m descriptor represents the Geary 2D-autocorrelation descriptor, which describe the topology of the peptide in association with atomic masses. G3m is the component symmetry directional weighted holistic invariant molecular descriptor weighted by mass. P2e encodes information on atomic symmetry weighted by electronegativity; whereas E2e encodes information on atomic shape weighted by Sanderson electronegativity. HATS7u and R6e are GETAWAY descriptors, which are calculated from the leverage matrix obtained by the centered atomic coordinates (molecular influence matrix). HATS7u represents the leverage-weighted autocorrelation GETAWAY descriptor. R6e is the autocorrelation descriptors weighted by electronegativity.

Overall, interestingly enough, the generated model employs most of descriptors (**Table 2**) that are related to mass or electronegativity. Mass-related descriptors seem to be important because of similarity of chemical compounds within the initial dataset. Electronegativity-related descriptors are also related to intermolecular interactions between diaminopyrimidine derivatives with DHFR. Summarizing our results one can conclude that DRAGON descriptors are suitable for reliable modeling of human DHFR.

The molecular structures of the best (12, 16, 15 and 9) and worst (19, 38, 1 and 2) human DHFR inhibitors were summarized respectively in **Figure 5**. One can conclude that the best inhibitors possess

aromatic rings which have polar substituents (chlorine or methoxy) and are connected to the diaminopyrimidine ring with a –CH$_2$-NH- or –NCH$_3$-CH$_2$- linkage. The lack of these structural properties on the worst inhibitors gives clues that the linkage type of two ring systems (substituted benzene and diaminopyrimidine) may have an important role on inhibition activity of mentioned molecules.
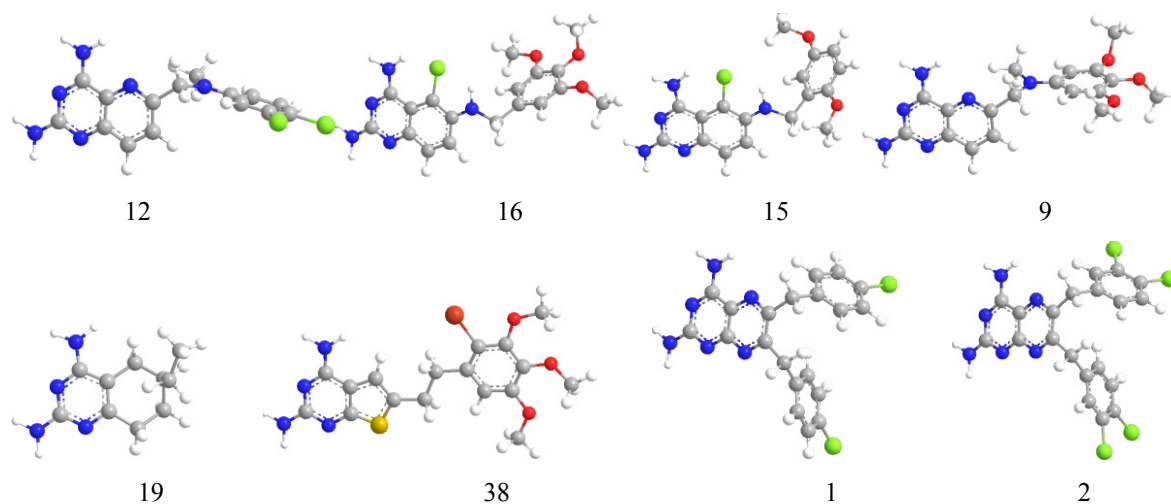

Figure 5. Optimized molecular structures of best (first line) and worst (second line) human DHFR inhibitors.

The two sp^3 hybridized bridge atoms (carbon and nitrogen) may facilitate rotation around three single bonds which allows reach to the best geometry to drive through the active site of the enzyme. The compound 38 also has two atoms between two ring systems but there is a thienyl ring and this may be one of the factors which reduce inhibition capacity. Aforementioned findings are in agreement with developed QSAR model.

As we can see from **Table 1**, compounds with lipophilic substituents demonstrate higher inhibitoty potential than compounds with polar substituents. In seems that the alkyl chain of the substituent interacts with the enzyme active site. Our resolts demonstrated that this interaction appears for a chain length of 4– 6 carbon atoms. This conclusion raised from descriptors related to topology of the peptide in association with atomic masses.

The electrostatic potential energy maps of two best (12 and 16) and two worst (19 and 38) compounds have been calculated (**Figure 6**). While the three dimensional structure is an important driving factor for the docking of the substrates to the active site of the enzymes, distribution of electrons affects the reactivity. A strongly localized negative charge can be seen on the α nitrogen of diaminopyrimidine ring of compound 12 and 16. This localized negative charge may allow the molecule to attack an electrophilic center from α nitrogen of diaminopyrimidine ring.

The negative charge has been delocalized on diaminopyrimidine ring and whole molecule in compound 19 and 38 respectively (**Figure 6**). This delocalization may reduce chance of nucleophilic attack from the α nitrogen of diaminopyrimidine ring and this phenomenon can be another reason for inhibition capacity of related molecules.
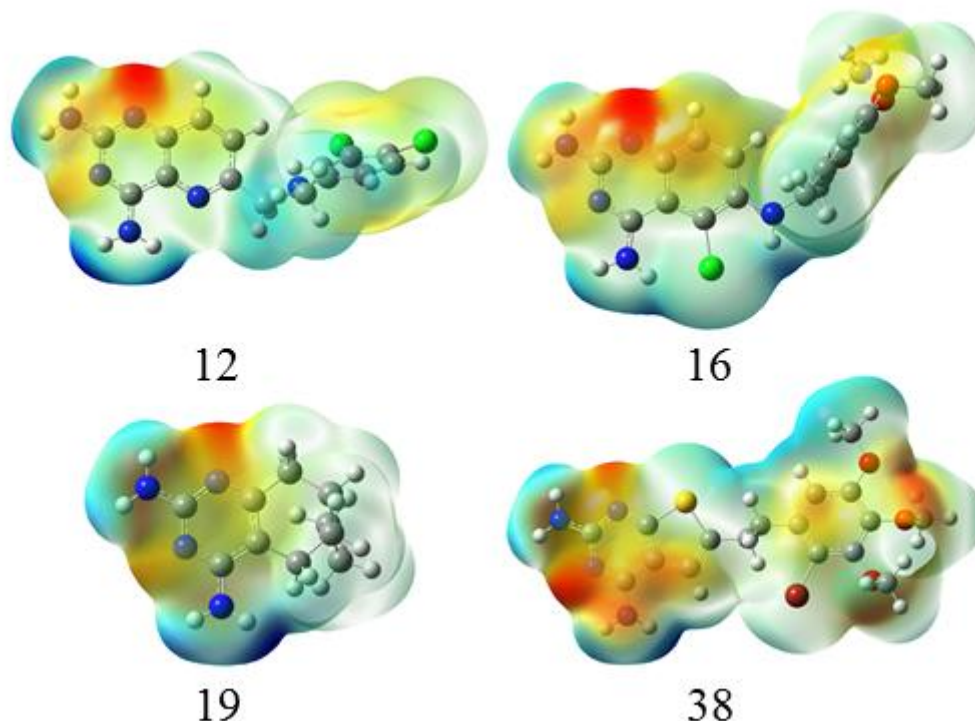
Figure 6. Calculated electrostatic potential energy maps of best (12 and 16) and worst (19 and 38) compounds (isovalue=0.02).

4. Conclusion

We have employed a QSAR approach to generate model for 47 chemically diverse dicyclic and tricyclic diaminopyrimidine derivatives tested for their inhibitory activity against human dihydrofolate reductase. Multiple linear QSAR model was used in combination with DFT modeling and DRAGON-based descriptors. We have found that majority of important descriptors are related to mass or to electronegativity of investigated compounds. Mass-related descriptors could reflect the similarity of chemical species within the initial dataset. Electronegativity-related descriptors seem to be are related to intermolecular interactions. Summarizing our results one can conclude that DRAGON descriptors, encompassed into a suitable QSAR model are suitable for reliable modeling of human dihydrofolate reductase.

We found that the best inhibitors possess aromatic rings which have polar substituents (chlorine or methoxy) and are connected to the diaminopyrimidine ring with a –CH2-NH- or –NCH3-CH2- linkage. We suggested that linkage type of two ring systems (substituted benzene and diaminopyrimidine) may have an important role on inhibition activity. Our results are in agreement with literature sources. These encouraging results could be used for comprehensive search of other potential dicyclic and tricyclic inhibitors of human dihydrofolate reductase.

References

1. B. Shane, Folate chemistry and metabolism. In: Folate in health and disease, L.B. Bailey, editor. New York, NY: Marcel Dekker, 1995.

2. L.B. Bailey, J.F. Gregory III, Folate metabolism and requirements. J Nutr. 129 (1999) 779-82.

3. R.L. Then, History and future of antimicrobial diaminopyrimidines. J Chemother. 5 (1993) 361-368.

4. E.M. Berman, L.M. Werbel, The renewed potential for folate antagonists in contemporary cancer chemotherapy. J. Med. Chem. 34 (1991) 479-85.

5. J.R. Bertino, Karnofsky memorial lecture. Ode to methotrexate. J. Clin. Oncol. 11 (1993) 5-14.

6. C. Silipo, C. Hansch, Correlation analysis. Its application to the structure-activity relation of triazines inhibiting dihydrofolate reductase. J. Am. Chem. Soc. 97 (1975) 6849–6861.

7. R.D. King, S. Muggleton, R. A. Lewis, M. J. Sternberg, Drug design by machine learning: the use of inductive logic programming to model the structure-activity relationships of trimethoprim analogues binding to dihydrofolate reductase. PNAS 89 (1992), 11322–11326.

8. T.A. Andrea, H. Kalayeh, Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. J. Med. Chem. 34 (1991) 2824–2836.

9. S.S. So, W.G. Richards, Application of neural networks: quantitative structure-activity relationships of the derivatives of 2,4-diamino-5-(substituted-benzyl)pyrimidines as DHFR inhibitors. J. Med. Chem. 35 (1992) 3201–3207.

10. A.K. Debnath, L.F. Kimball, Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of mycobacterium avium complex dihydrofolate reductase. J. Med. Chem. 45 (2002) 41–53.

11. J.D. Hirst, R.D. King, M.J.E. Sternberg, Quantitative structure-activity relationships by neural networks and inductive logic programming. I. The inhibition of dihydrofolate reductase by pyrimidines. J. Comput.-Aided Mol. Des. 8 (1994) 405–420.

12. J.D. Hirst, R.D. King, M.J.E. Sternberg, Quantitative structure-activity relationships by neural networks and inductive logic programming. II. The inhibition of dihydrofolate reductase by triazines. J. Comput.-Aided Mol. Des. 8 (1994) 421–432.

13. P. Singla, V. Luxami, K. Paul, Synthesis, in vitro antitumor activity, dihydrofolate reductase inhibition, DNA intercalation and structure–activity relationship studies of 1,3,5-triazine analogues. Bioorg. Med. Chem. Lett. 26 (2016) 518–523.

14. H. Yilmaz, N. Sizochenko, B. Rasulev, A. Toropov, Y. Guzel, V. Kuz'min, D. Leszczynska, J. Leszczynski, Amino substituted nitrogen heterocycle ureas as kinase insert domain containing receptor (KDR) inhibitors: Performance of structure–activity relationship approaches. J. Food Drug Analysis. 23 (2015) 168–175.

15. M.C. Broughton, S.F. Queener, Pneumocystis carinii dihydrofolate reductase used to screen potential antipneumocystis drugs. Antimicrob. Agents Chemother. 35 (1991) 1348-55.

16. L.C. Chio, S.F. Queener, Identification of potent inhibitors of Toxoplasma gondii dihydrofolate reductase. Antimicrob. Agents Chemother. 37 (1993) 1914-23.

17. R.G Nelson, A. Rosowsky, Dicyclic and tricyclic diaminopyrimidine derivatives as potent inhibitors of cryptosporidium parvum dihydrofolate reductase: structure-activity and structure-selectivity correlations, Antimicrob. Agents Chemother. 45 (2001) 3293-303.

18. M.S. Bartlett, M. Shaw, P. Navaran, J.W. Smith, S.F. Queener, Evaluation of potent inhibitors of dihydrofolate reductase in a culture model for growth of Pneumocystis carinii. Antimicrob. Agents Chemother. 39 (1995) 2436-41.

19. R.L. Then, Antimicrobial dihydrofolate reductase inhibitors--achievements and future options: review. J Chemother. 16 (2004) 3-12.

20. N. Sizochenko, D. Majumdar, S. Roszak, J. Leszczynski. Application of quantum mechanics and molecular mechanics in chemoinformatics. In: Handbook of computational chemistry, J. Leszczynski, editor, Springer Netherlands, 2016, DOI: 10.1007/978-94-007-6169-8_52-1.

21. GAUSSIAN 09 code, http://www.gaussian.com/g_prod/g09.htm (accessed 01.05.2016)

22. e-DRAGON, Virtual Computational Chemistry Laboratory, http://www.vcclab.org/lab/edragon/ (accessed 01.05.2016).

23. A.K. Haghi, Methodologies and Applications for Chemoinformatics and Chemical Engineering, IGI Global, University of Guilan, Iran, 2013.

24. R. Todeschini, M. Lasagni, E. Marengo, New molecular descriptors for 2D and 3D structures, Theory. J. Chemom. 1994 8(4) 263–272.

25. OECD, Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR]. http://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm (accessed 01.05.2016)

26. T. Puzyn, J. Leszczynski, M.T.D. Cronin, Recent advances in QSAR studies: methods and applications. Springer Dordrecht Heidelberg London New York, 2010.

27. K. Roy, Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment. IGI Global; 2015.

28. P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS: A new software for the development, analysis, and validation of QSAR MLR models, J. Comput. Chem. 34 (2013) 2121–2132.

29. A.C. Rencher, W.F. Christensen, Multivariate regression, Section 10.1, Introduction, Methods of Multivariate Analysis, 3rd ed. Wiley Series in Probability and Statistics, John Wiley & Sons, 2012.

30. A. Tropsha, D.J. Abraham, Recent advances in development, validation, and exploitation of QSAR models, in: Burger's Medicinal Chemistry and Drug Discovery. John Wiley & Sons, Inc. 2003.

31. C. Rücker, G. Rücker, M. Meringer, y-Randomization and its variants in QSPR/QSAR. J. Chem. Inf. Model. 47 (2007) 234–557.

32. H. Hotelling, Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24 (1933) 417–441.

33. P. Gramatica, Principles of QSAR models validation: internal and external. QSAR & Comb Sci. 2007 26(5) 694-701.

Graphical Abstract