Applied Medical Informatics
Original Research **Vol. 17, No. 3, 4/2005, pp: 12 - 21**

Complex Structural Information Integration: Inhibitor Activity on Carbonic Anhydrase II of Substituted Disulfonamides

Lorentz Jäntschi [1], Mihaela Ligia Ungureşan [1], Sorana-Daniela Bolboacă [2]
[1]Technical University of Cluj-Napoca, România
[2]Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, România

Abstract **-** *The inhibition activity on carbonic anhydrase II of forty substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides has been studies using a molecular descriptor family on structure-activity relationship approach. The approach use the structure information of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides for generating and calculating the molecular descriptors family. Computed molecular descriptors entered into a multiple linear regression analysis and the models with grater performance in modeling of inhibition activity on carbonic anhydrase II are analyzed. The analysis of the models was performed through correlation coefficients, squared correlation coefficients, and regressions parameters. The predictivity of the models was evaluated by cross-validation leave-one-out and training versus test (starting with 50% and up to 25% of compounds in training sets) analysis. The comparison between bi- and four-varied models was performed through a correlated correlation analysis and the comparison between the best performing model and previous reported models by Fisher Z test. The four-varied model allows acquiring some knowledge about the relationships between complex structural information of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides and theirs inhibition activity on carbonic anhydrase II.*

Keywords **-** *Molecular Descriptors Family (MDF), Quantitative Structure-Activity Relationship (QSAR), carbonic anhydrase II (CA II), substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides*

Introduction

Quantitative structure-activity relationships (QSAR), a mathematical approaches of linking chemical structure and biological activity in a quantitative manner [1], is widely used in modeling of biological activities of compounds and in drug design [2-4].

The inhibition activity of different chemicals on CA II, an enzyme with primary localization on cytoplasm related with CA II deficiency syndrome [5], osteopetrosis [6], renal acidosis and cerebral calcification [7], mental retardation [8], nephrocalcinosis, urolithiasis, and hypercalciuria [9], has been studied by the use of quantitative structure-activity relationships since 1980's [10-13].

Forty substituted disulfonamides, twenty 1,3,4-thiadiazole- and twenty 1,3,4-thiadiazoline-disulfonamides, with inhibition properties on CA II were previous studied by the use of quantum chemical quantitative structure-activity relationships [14]. The published models and the related statistical parameters are in table 1.

Table 1. QSAR models and related statistical parameters of CA II inhibitions

Model no.	Expression / Statistical parameters
1	$\log IC_{50} = 8.92 \cdot 10^{-3} \cdot \Pi_{xx} - 6.68 \cdot Q_{Cr1} + 18.97 \cdot Q_{S1} - 0.736 \cdot E_H + 0.0667 \cdot \mu_x - 0.0417 \cdot \mu_z + 0.0275 \cdot \Delta H_S - 64.15$ $n = 40, R^2 = 0.719, Q^2 = 0.475, s = 0.304, F = 11.70$
2	$\log IC_{50} = -7.05 \cdot Q_{Cr1} + 13.19 \cdot Q_{S1} - 0.677 \cdot E_H + 0.126 \cdot \mu_x - 0.0421 \cdot \mu + 0.298 \cdot LogP + 0.0302 \cdot \Delta H_S - 46.52$ $n = 36, R^2 = 0.876, Q^2 = 0.777, s = 0.152, F = 28.66$

Πzz = the polarizability tensor, QCr1 = the changes of the atoms of the attached ring carbon, QS1 = the changes of the atoms of the primary sulfonamide group, μx = the dipole moment, QNr1 = the charges on one azot atom and ΔHS = the salvation energy for the sulfonamide group
n = the studied sample size, *R2* = the square of the multiple correlation coefficients, *Q2* = the same quantity base on the predicted errors (the leave-one-out techniques), *s* = the standard errors of estimate of the equation, *F* = the Fisher variance ratio

In accordance with the estimation and prediction abilities of activities obtained by applying the molecular descriptors family on structure-activity relationships (MDF-SAR) method [15,16], the approach was applied on a sample of substituted 1,3,4-thiadiazole- and thiadiazoline-disulfonamides with inhibition activity on CA II and the abilities in estimation and prediction of the obtained models were investigated.

Materials and Methods

Substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides

Twenty 1,3,4-thiadiazole- and twenty 1,3,4-thiadiazoline-disulfonamides, with inhibition properties on CA II was included into the study. The measured inhibitory activity of compounds, express as concentration of the compound that is required for fifty percent inhibition in vitro, transformed in logarithmical scale (log IC_{50}), was taking from a previous study [14].

The generic structure of compounds, abbreviation, substituents, and measured inhibition activity are in table 2.

Modeling method

The inhibition activity of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides was modeled through integration of complex structure information. The modeling process uses an original molecular descriptors family on structure-activity relationship approach in order to estimate and predict the inhibition activity of studied compounds. The steps applied [17] were:

- *Step 1 (3D representation of compounds)* The three-dimensional representations of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides were building up with ***HyperChem*** software [18].
- *Step 2 (creation of measured properties file)* The measured inhibition concentration 50% of each compounds was transformed in logarithm scale and was stored in a *.txt file.
- *Step 3 (generation of molecular descriptors family)* All forty compounds were used in construction and generation of molecular descriptors family. The algorithm of generation the list of molecular descriptors family for substituted 1,3,4-thiadiazole- and thiadiazoline-disulfonamides is strictly based on compounds structure. A bias algorithm with a significance level equal with 10^{-9} was applied after the generation of molecular descriptors family in order to discard redundant information.

Table 2. Structure of compounds and measured log IC_{50}

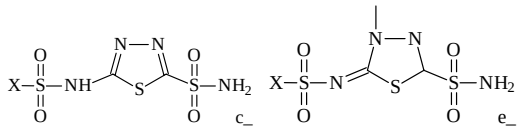

Abb.	X	log IC_{50} (nM)
c_01	Me	0.7782
c_02	$PhCH_2$	0.6990
c_03	4-Me-C_6H_4	0.6021
c_04	4-F-C_6H_4	0.6021
c_05	4-Cl-C_6H_4	0.4771
c_06	4-Br-C_6H_4	0.3010
c_07	4-MeO-C_6H_4	0.4771
c_08	4-AcNH-C_6H_4	0.4771
c_09	4-H_2N-C_6H_4	0.3010
c_10	3-H_2N-C_6H_4	0.0000
c_11	4-O_2N-C_6H_4	0.0000
c_12	3-O_2N-C_6H_4	-0.0458
c_13	2-O_2N-C_6H_4	0.4771
c_14	Me_2N	0.9031
c_15	2-$HO_2CC_6H_4$	-0.3010
c_16	4-(2,4,6-Me_3Py^+)C_6H_4	0.6021
c_17	4-(2,4,6-Ph_3Py^+)C_6H_4	2.0414
c_18	2,4-$(O_2N)_2C_6H_3$	0.6990
c_19	4-Cl-3-O_2N-C_6H_3	0.4771
c_20	2,4,6-$Me_3C_6H_4$	0.9542

e_01	Me	0.6021
e_02	$PhCH_2$	0.9031
e_03	4-Me-C_6H_4	0.4771
e_04	4-F-C_6H_4	0.6021
e_05	4-Cl-C_6H_4	0.4771
e_06	4-Br-C_6H_4	0.3010
e_07	4-Me0-C_6H_4	0.4771
e_08	4-AcNH-C_6H_4	-0.1549
e_09	4-H_2N-C_6H	-0.2218
e_10	3-H_2N-C_6H_4	-0.3010
e_11	4-O_2N-C_6H_4	0.6021
e_12	3-O_2N-C_6H_4	0.3010
e_13	2-O_2N-C_6H_4	0.0000
e_14	Me_2N	0.6990
e_15	2-$HO_2CC_6H_4$	-0.6990
e_16	4-(2,4,6-Me_3Py^+)C_6H_4	0.6021
e_17	4-(2,4,6-Ph_3Py^+)C_6H_4	2.0414
e_18	2,4-$(0_2N)_2C_6H_3$	0.6021
e_19	4-Cl-3-O_2N-C_6H_3	0.3010
e_20	2,4,6-$Me_3C_6H_4$	0.8451

Me = methyl; Ph = phenyl; Ac = acetyl; Py+ = pyridine

- *Step 3 (continuation)* Each calculated descriptor has an individual name of seven letters, which express the modality of construction. Thus, the 7th letter describes the compound characteristic relative to its geometry (*g*) or topology (*t*); the 6th letter denotes the atomic property (the cardinality - *C,* the number of directly bonded hydrogen's - *H*, the atomic relative mass - *M*, the atomic electronegativity - *E*, the group electronegativity - *G*, and the partial charge - *Q*); the 5th letter denotes the atomic interaction descriptor; the 4th letter denotes the overlapping interaction model; the 3rd letter denotes the fragmentation criterion (the minimal fragments - *m*, the maximal fragments - *M*, the Szeged fragments criterion – *D*, and the Cluj fragments criterion – *P* [19,20]); the 2nd letter denotes the cumulative method of fragmentation properties (one out of nine selectors [17]), and the 1st letter denotes the linearization procedure applied in global molecular descriptor generation (identity - *I*, inverse - *i*, absolute - *A*, a inverse of absolute - *a*, natural logarithm of absolute value - *L*, simple natural logarithm - *l*).
- *Step 4 (search and identification of SAR models)* The highest values for correlation and squared correlation coefficients were the criterions imposed in searching and identifying the MDF-SAR models.
- *Step 5 (models validation)* Analysis of predictive abilities of the choused MDF-SAR models was performed through model validation analysis by computing: the cross-validation leave-one-out (loo) score ($r^2_{cv\text{-}loo}$), Fisher parameter (F_{pred}), probability of wrong model for loo analysis (p_{pred}), and standard error (s_{loo}). In leave-one-out analysis the property of each compound was predicted by the regression equation calculated based on the all other compounds. The analysis of predictive abilities of MDF-SAR models were performed with ***Leave-one-out Analysis*** application [21].
- *Step 6 (SAR models analysis)* The choused MDF-SAR models were analyzed through computing and interpreting the following parameters: the correlation coefficient (r), the squared correlation coefficient (r^2), the adjusted squared correlation coefficient (r^2_{adj}), the standard error of estimation (s_{est}), the Fisher parameter (F_{est}) and its significance (p_{est}), the significance ($p_{descriptor}$) and 95% confidence intervals ($95\%CI_{descriptor}$) associated with Student parameter ($t_{descriptor}$) of regression coefficients, the co-linearity parameters (analysis of the squared correlation coefficients between descriptors, r^2(*descriptor*, *descriptor*), and between one descriptor and measured IC_{50} - r^2(*log* IC_{50}, *descriptor*)), and the model stability (defines as the difference between squared correlation coefficient and leave-one-out correlation coefficient score, the model is consider stable if the difference has lower value, r^2 - $r^2_{cv\text{-}loo}$). The comparison between bi- and four-varied models was performed through a correlated correlation analysis by applying the Steiger test [22].

The estimation abilities of the model with the highest squared correlation coefficient was

analyses in training and test sets with ***Training vs. Test*** application [23]. There were analyzed eleven situations, starting with sample sizes in training sets from twenty to thirty and corresponding sample sizes in test sets from twenty to ten.

Results

Two MDF-SAR models, one bi- and one four-varied, proved to have abilities in estimation and prediction of 1,3,4-thiadiazole- and 1,3,4-thiadiazoline disulfonamides inhibition activity on CA II. The MDF-SAR models are in table 3.

The molecular descriptors used by the models, theirs calculated values and the estimated of the inhibition concentration 50% by each models ($\hat{Y}_{2\text{-v}}$, $\hat{Y}_{4\text{-v}}$) are in table 4.

Table 3. MDF-SAR models

Model no.	Equation
1	$\hat{Y}_{2\text{-v}} = -4.4479 + 2.4352 \cdot imDdSCg + 9.4635 \cdot 10^{-2} \cdot iiMrqQg$
2	$\hat{Y}_{4\text{-v}} = -9.9859 + 4.5643 \cdot imDdSCg + 2.945 \cdot 10^{-3} \cdot isDrqQg + 5.2036 \cdot IIMDQQg + 1.4832 \cdot lmMrsGg$

The statistical parameters associated with the analysis of the bi- and four-varied MDF SAR models are in tables 5 and 6.

The distributions of measured (log IC_{50}) and estimated by bi- ($\hat{Y}_{2\text{-v}}$) and four-varied ($\hat{Y}_{4\text{-v}}$) MDF-SAR models are in figure 1 and the distributions of residuals in figure 2.

The results of correlated correlation analysis are in table 7.

Table 4. Descriptors used in MDF-SAR models, theirs values and estimated inhibition activity

		Bi-varied model			Tetra-varied model			
No.	Abb.	$\hat{Y}_{2\text{-v}}$	iiMrqQg	imDdSCg	isDrqQg	IIMDQQg	lmMrsGg	$\hat{Y}_{4\text{-v}}$
1	c_01	0.7022	8.0614	1.8016	105.65	0.1203	1.0679	0.7582
2	c_02	0.6276	8.6874	1.7466	162.91	0.0405	1.3894	0.7374
3	c_03	0.5761	9.1547	1.7073	172.34	0.0951	1.0866	0.4209
4	c_04	0.2722	9.3193	1.5761	177.76	0.1023	1.5409	0.5489
5	c_05	0.5774	9.1583	1.7077	170.76	0.1152	1.0090	0.4074
6	c_06	0.5427	8.7839	1.7080	161.88	0.0973	1.0086	0.2890
7	c_07	0.5824	10.395	1.6617	200.39	0.0914	1.1896	0.4290
8	c_08	0.1158	14.978	1.2920	355.18	0.1477	1.7094	0.2610
9	c_09	0.3207	10.553	1.5481	219.09	0.1204	1.1992	0.1305
10	c_10	0.3314	10.676	1.5477	220.87	0.1374	1.2246	0.2599
11	c_11	0.3700	11.074	1.5481	242.14	0.1060	1.2809	0.2446
12	c_12	0.3575	10.952	1.5477	237.39	0.1252	1.2809	0.3284
13	c_13	0.2797	11.136	1.5086	244.70	0.1483	1.4664	0.5670
14	c_14	0.6869	9.2454	1.7493	132.88	0.1450	1.0871	0.7565
15	c_15	-0.4185	10.835	1.2336	261.86	0.1013	1.7094	-0.5220
16	c_16	0.4546	11.921	1.5499	373.43	0.0746	1.4068	0.6628
17	c_17	1.9327	28.797	1.5011	961.96	0.0452	1.3894	1.9945
18	c_18	0.3124	12.385	1.4735	337.21	0.2088	1.2809	0.7190
19	c_19	0.2432	10.529	1.5172	252.22	0.1182	1.2809	0.1965
20	c_20	0.7904	13.082	1.6427	262.52	0.1051	1.3656	0.8571
21	e_01	0.5344	6.7974	1.7818	92.229	0.0676	1.2671	0.6493
22	e_02	0.5952	7.4401	1.7818	149.79	0.0402	1.3894	0.8579
23	e_03	0.6359	7.8693	1.7818	142.78	0.0780	1.0866	0.5845
24	e_04	0.1903	8.4772	1.5752	167.93	0.0932	1.5409	0.4687
25	e_05	0.6545	8.0666	1.7818	155.24	0.0878	1.0085	0.5568
26	e_06	0.6370	7.8816	1.7818	151.04	0.0931	1.0093	0.5732
27	e_07	0.4772	9.3113	1.6606	182.88	0.0951	1.1896	0.3912
28	e_08	0.3150	17.054	1.2931	430.95	0.1147	1.7094	0.3176
29	e_09	0.1396	8.6623	1.5472	190.86	0.0675	1.1992	-0.2318
30	e_10	0.0954	8.4144	1.5387	189.87	0.0766	1.2246	-0.1887
31	e_11	0.4220	11.646	1.5472	335.20	0.1100	1.2809	0.5353
32	e_12	0.2555	10.106	1.5387	231.36	0.1101	1.2809	0.1914
33	e_13	0.0889	9.1145	1.5088	216.24	0.0941	1.4664	0.2020

34	e_14	0.6320	7.8282	1.7818	113.66	0.0735	1.1504	0.5702
35	e_15	-0.6248	8.8112	1.2275	238.79	0.0819	1.7094	-0.7184
36	e_16	0.8854	16.757	1.5389	392.97	0.0651	1.4068	0.6209
37	e_17	2.0926	30.486	1.5011	949.42	0.0543	1.3894	2.0050
38	e_18	0.4339	13.653	1.4741	442.43	0.1121	1.2809	0.5283
39	e_19	0.2318	10.676	1.5068	258.51	0.1175	1.2809	0.1644
40	e_20	0.6189	9.0003	1.7309	199.79	0.0597	1.3656	0.8391

Table 5. Quality of MAD SAR models

	StdError	r^2(Y, *desc*)	t	95%CI_{lower}	95%CI_{upper}	p (%)
Bi-varied model						
Intercept	0.5000		-8.8959	-5.4610	-3.4348	$1.01 \cdot 10^{-8}$
imDdSCg	0.2836	0.1265	8.5882	1.8607	3.0099	$2.45 \cdot 10^{-8}$
iiMrqQg	0.0089	0.3573	10.654	0.0766	0.1126	$7.92 \cdot 10^{-11}$
Four-varied model						
Intercept	0.9280		-10.757	-11.869	-8.1013	$1.21 \cdot 10^{-10}$
imDdSCg	0.3613	0.1265	12.634	3.8308	5.2977	$1.34 \cdot 10^{-12}$
isDrqQg	0.0002	0.3231	16.314	0.0026	0.0033	$6.23 \cdot 10^{-16}$
IIMDQQg	1.0205	0.0568	5.0990	3.1319	7.2754	$1.19 \cdot 10^{-3}$
lmMrsGg	0.2430	0.0210	6.1028	0.9897	1.9763	$5.65 \cdot 10^{-5}$

StdError = standard error, Y = log IC^{50}, desc = molecular descriptor, t = parameter of the Student test

Table 6. The statistical characteristics of MDF-SAR models

Parameter	Value	
n	40	40
v	2	4
r	0.8862	0.9506
95%CI_{r_lower}	0.7937	0.9079
95%CI_{r_upper}	0.9385	0.9737
r^2	0.7853	0.9037
r^2_{adj}	0.7737	0.8927
s_{est}	0.2477	0.1706
F_{est}	68	82
p_{est} (%)	$4.4 \cdot 10^{-11}$	$2.7 \cdot 10^{-15}$
$r^2_{cv\text{-}loo}$	0.7564	0.8804
s_{loo}	0.2640	0.1902
F_{pred}	57	64
p_{pred} (%)	$4.6 \cdot 10^{-10}$	$1.2 \cdot 10^{-13}$
r^2 - $r^2_{cv\text{-}loo}$	0.0289	0.0234
r^2(*imDdSCg*, *iiMrqQg*)	0.1643	n.a.
r^2(*imDdSCg*, *isDrqQg*)	n.a.	0.1960
r^2(*imDdSCg*, *IIMDQQg*)	n.a.	0.0836
r^2(*imDdSCg*, *lmMrsGg*)	n.a.	0.5933
r^2(*isDrqQg*, *IIMDQQg*)	n.a.	0.0259
r^2(*isDrqQg*, *lmMrsGg*)	n.a.	0.1062
r^2(*IIMDQQg*, *lmMrsGg*)	n.a.	0.1062

n = number of compounds, v = number of descriptor, n.a. = not applicable

Table 7. Results of correlated correlation analysis

Parameter	Value
$r(\log IC_{50} - \hat{Y}_{4\text{-}v})$	0.95064
$r(\log IC_{50} - \hat{Y}_{2\text{-}v})$	0.88617
$r(\hat{Y}_{four\text{-}v} - \hat{Y}_{2\text{-}v})$	0.93439
Steiger's Z	3.17474
p (%)	0.075

Comparing the performances of bi- (model 1, table 3) and four-varied (model 2, table 3) MDF-SAR models with previous reported models (model 1 and 2, table 1) was observed that the four-varied model obtained statistically

significant grater correlation coefficients (model 2, table 3 – model 1, table 1, p-value = 0.0056).

The results obtained in training versus test analysis are in table 8, and the plot of squared correlation coefficient in figure 4.

The graphical representation of the performances of four-varied model is in figure 3.

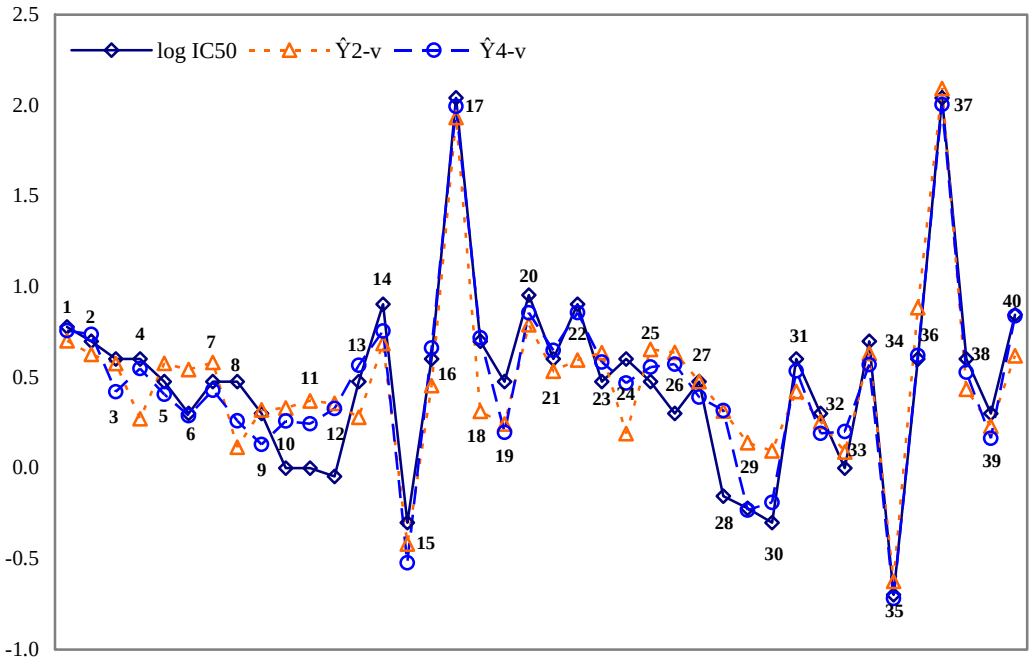

Figure 1. The distributions of measured and estimated by MDF-SAR models inhibition activity

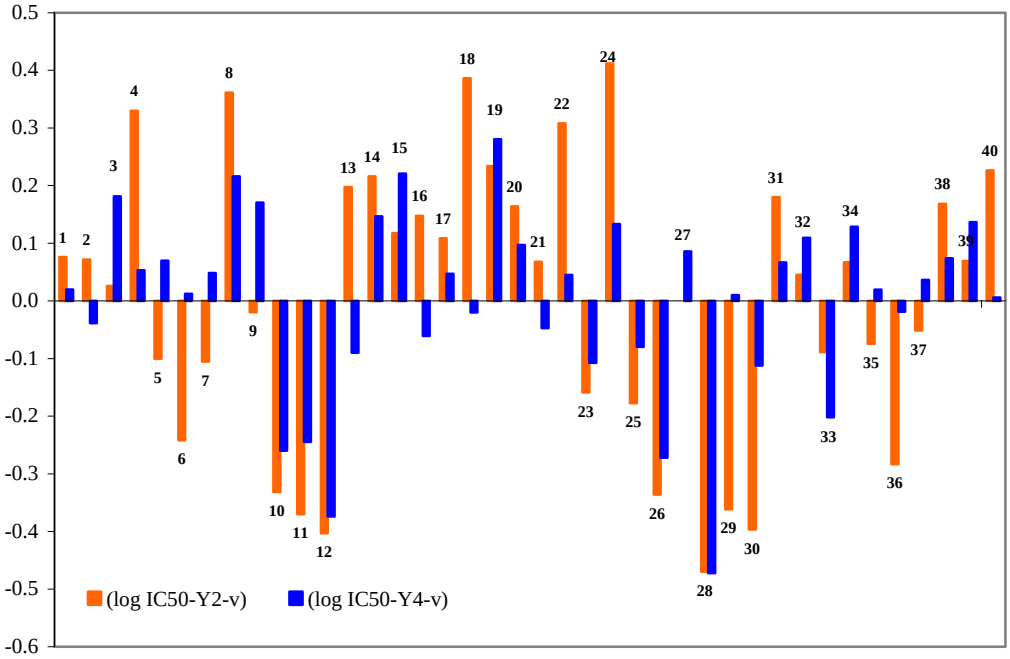

Figure 2. The distribution of the residuals for bi- and four-varied MDF-SAR models

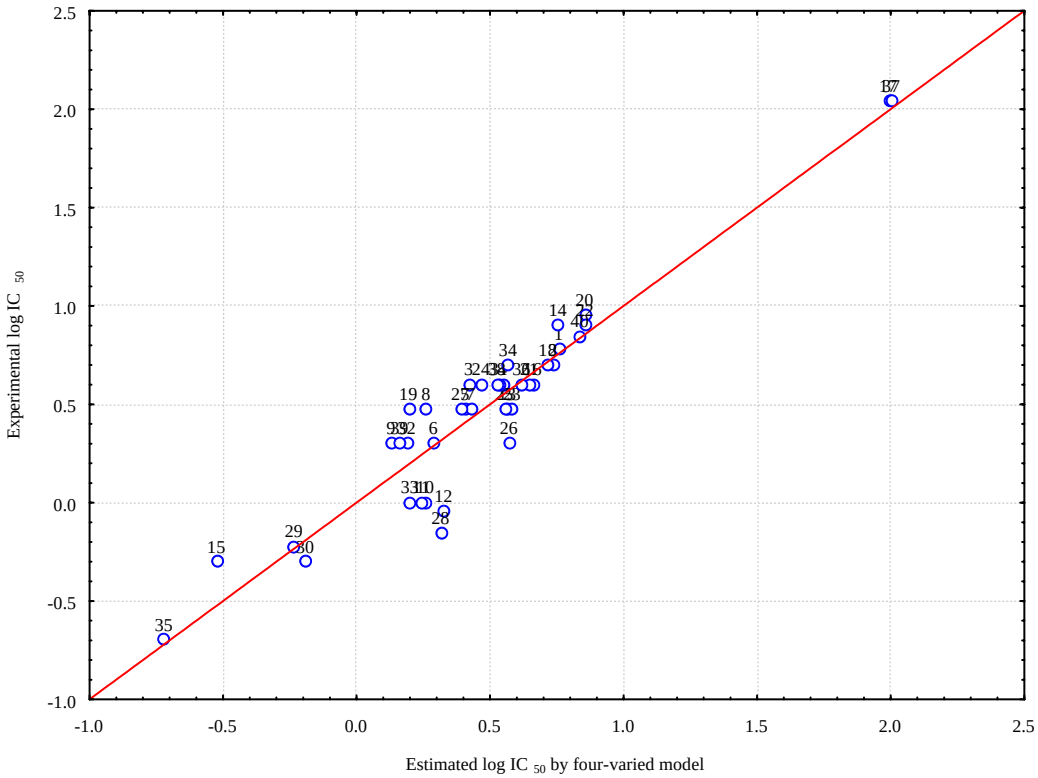

Figure 3. Estimated with the four-varied model versus measured log IC_{50} of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides

Table 8. Quality of models in training versus test analysis

	MDF-SAR equation					Training				Test				r_{tr}vs r_{ts}
No.	Intercept	imDdSCg	isDrqQg	IIMDQQg	lmMrsGg	No_{tr}	r_{tr}	95%CI_{rtr}	F_{tr}	No_{ts}	r_{ts}	95% CI_{rts}	F_{ts}	$F_{Z\text{-test}}$
1	-10.808	4.8727	$3.03 \cdot 10^{-3}$	5.6355	1.7079	20	0.9179	[0.801, 0.967]	20**	20	0.9604	[0.901, 0.984]	40**	1.10†
2	-9.7091	4.4868	$2.55 \cdot 10^{-3}$	5.5492	1.3963	21	0.9137	[0.796, 0.965]	20**	19	0.9559	[0.887, 0.983]	28**	1.01†
3	-8.5266	4.1108	$1.87 \cdot 10^{-3}$	3.6226	1.2308	22	0.8950	[0.760, 0.956]	17**	18	0.9076	[0.765, 0.965]	11**	0.20†
4	-10.714	4.9179	$3.03 \cdot 10^{-3}$	5.3688	1.5626	23	0.9498	[0.883, 0.979]	41**	17	0.9547	[0.876, 0.984]	24**	0.15†
5	-10.659	4.6905	$2.97 \cdot 10^{-3}$	5.6210	1.8122	24	0.9399	[0.864, 0.974]	36**	16	0.9497	[0.858, 0.983]	25**	0.26†
6	-8.9966	4.3022	$2.89 \cdot 10^{-3}$	5.0608	1.0878	25	0.9624	[0.915, 0.984]	63**	15	0.9011	[0.722, 0.967]	9*	1.39†
7	-9.2808	4.3025	$2.86 \cdot 10^{-3}$	4.1786	1.3324	26	0.9416	[0.872, 0.974]	41**	14	0.9589	[0.872, 0.987]	17**	0.49†
8	-9.2919	4.2780	$2.89 \cdot 10^{-3}$	5.3242	1.3077	27	0.9406	[0.872, 0.973]	42**	13	0.9649	[0.884, 0.990]	24**	0.72†
9	-9.8900	4.4827	$2.42 \cdot 10^{-3}$	5.5150	1.5653	28	0.9090	[0.811, 0.957]	27**	12	0.9710	[0.897, 0.992]	12*	1.51†
10	-9.4715	4.3030	$2.35 \cdot 10^{-3}$	5.8316	1.4684	29	0.9142	[0.824, 0.959]	31**	11	0.9820	[0.929, 0.995]	10*	1.97†
11	-10.194	4.6598	$2.97 \cdot 10^{-3}$	5.1893	1.5175	30	0.9483	[0.893, 0.975]	56**	10	0.9662	[0.859, 0.992]	15*	0.51†

* $0.001 < p < 0.01$, ** $p < 0.001$, † $p > 0.05$

95% CI = 95% confidence intervals; r_{tr} = correlation coefficient – training set; r_{ts} = correlation coefficient – test set; FZ-test = Fisher's Z test; No_{tr} = number of compounds in training sets; No_{ts} = number of compounds in test sets

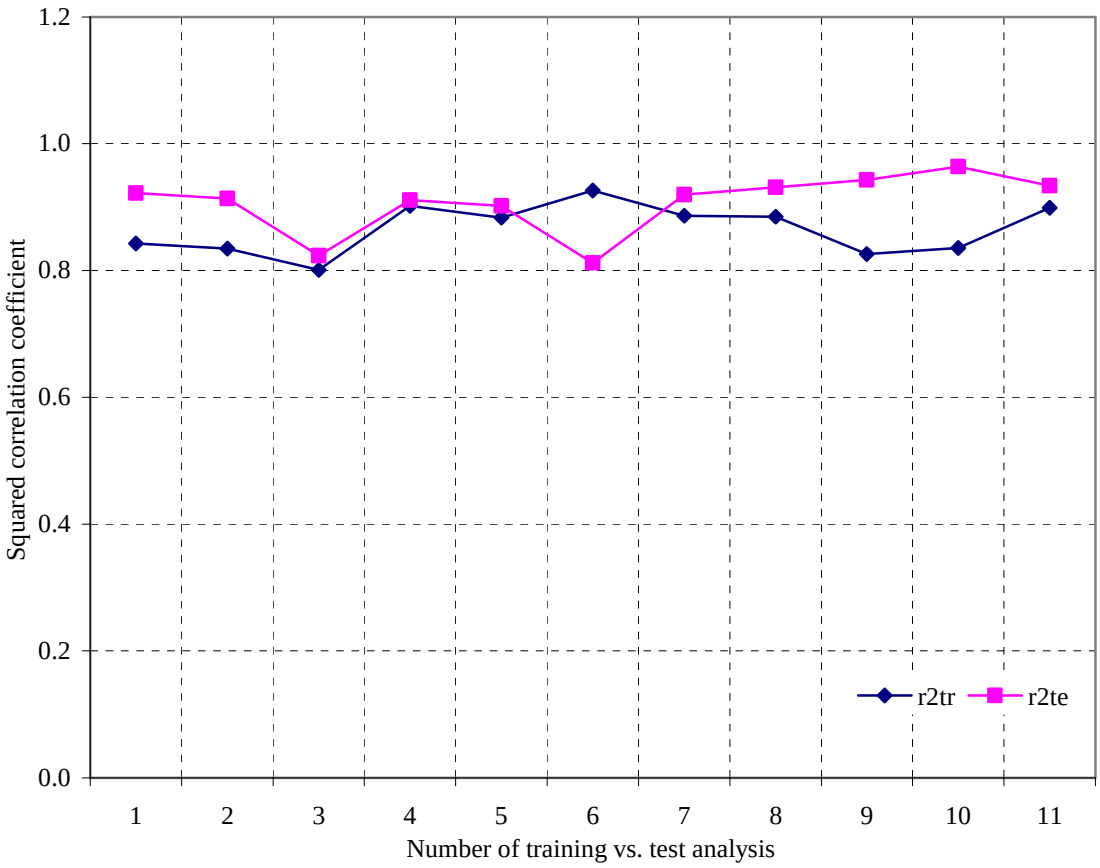

Figure 4. Training and test squared correlation coefficients with the four-varied model

Discussions

The inhibition activity on CA II of substituted 1,3,4-thiadiazole and 1,3,4-thiadiazoline-disulfonamides proved to be in relationship with the compounds structural information. In both equations (model 1, 2, table 3) the inhibition activity on CA II is in relationship with the geometry of compounds (*imDdSC**g**, iiMrqQ**g**, isDrqQ**g**, IIMDQQ**g**, lmMrsG**g***) and is depend by the compounds partial charge (*iiMrq**Q**g, isDrq**Q**g, IIMDQ**Q**g*), the cardinality (*imDdS**C**g*) and the group electronegativity (*lmMrs**G**g*).

In both MDF-SAR models, all descriptors have positive regression coefficients and descriptor *imDdSCg* have an important contribution in modeling of inhibition activity on CA II, being present in both models.

Analyzing the performances of bi-varied model it can be observed that is statistically significant in estimation as well as in prediction (see the squared correlation coefficients and theirs adjusted values and leave-one-out score, table 5). Almost seventy-eight percent (seventy-seven if we looked at the adjusted value of squared correlation coefficient) of inhibition activity on CA II of studied substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides can be explained by its linear relationship with the variation of *imDdSCg* and *iiMrqQg* molecular descriptors (bi-varied model, table 5). The bi-varied model proved to be a valid and stable model (p_{pred} (%) = $4.4 \cdot 10^{-11}$; $r^2_{cv\text{-}loo}$ = 0.7564; r^2 - $r^2_{cv\text{-}loo}$ = 0.0289).

The estimation abilities of the four-varied model are sustained by the value of correlation coefficient (r^2 = 0.9037, table 5), confidence boundaries associated with the regression coefficients and probabilities associated with Student tests (for all coefficients less than 0.001 - see table 6). Almost ninety percent from variation of inhibition activity on CA II can be explained its linear relationship with the variation of the four molecular descriptors used in the model (model 2, table 3). The probability of wrong model for leave-one-out analysis (p_{pred}(%) = $2.7 \cdot 10^{-15}$) and its associated Fisher

parameter (F_{pred} = 82) sustain the estimation abilities of the model. The stability of the four-varied model is sustained by the values of difference between correlation coefficient and cross validation leave-one-out correlation score (r^2 - $r^2_{cv(loo)}$ = 0.0234), the value of cross validation score being very close to the value of adjusted squared correlation coefficient. The power of the four-varied model in prediction of inhibition on CA II of studied disulfonamides is sustained by the absence of co-linearity of descriptors (see the squared correlation coefficients between pairs of descriptors, which with an exception is less than 0.20 - table 5).

The residuals of the bi-varied model vary from -0.4699 to 0.4118 while the residual of the four-varied model vary from -0.4725 to 0.2806, in thirty out of forty cases the residuals of four-varied model having the smallest values comparing with the bi-varied model. The comparison between bi and four varied model proved that the four-varied one has a significantly greater correlation coefficient comparing with the bi-varied model (p(%) = 0.075, table 7) and statistically significant higher correlation coefficients comparing with previous reported model which took into consideration all forty compounds (p = 0.0056). Note that, the MDF-SAR model with the great squared correlation coefficient and cross validation leave-one-out score (four-varied model) is able to estimate the inhibition activity on CA II of studied compounds by the use of a half number of descriptors comparing with previous reported models, in condition in which the previous reported model used a greater number of descriptors that is acceptable [24].

The goodness-of-fit of the four-varied model was assessed in training versus test analysis. After we did study the robustness of model parameters by the used of cross validation leave-one-out analysis we decide to assess the internal predictivity in training and test sets. This task was performed by splitting randomly the compounds in training and test sets. Looking at the intercept and at the coefficients of descriptors in training sets it can be observed that all values did not exceed the 95% confidence intervals (see tables 8 and 6). The values of the correlation coefficients obtained in test sets are included into the 95% confidence intervals of the correlation coefficients obtained in training sets, and there were not identify statistical significances between the correlation coefficients on training and test sets (table 8, Fisher's Z parameter always greater than 0.05). More, all the correlation coefficients obtained in training and test sets are included into the 95% confidence intervals of four-varied model.

Looking at the graphical representation of the estimated log IC_{50} on CA II with four-varied model and measured log IC_{50} (figure 3) it can be observed that the model can be unstable, because the response of the compounds c_17 and e_17 are too isolated in the superior part of the regression line, meaning that the model can be 'driven' by these compounds.

Even if the internal validation results sustain the stability of the model, providing an approximation of the predictive ability of the four-varied MDF-SAR model, future studies are necessary in order to assess the influence of two compounds specified above in the four-varied model stability by the use of new experimentally tested 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides with inhibition properties on CA II. The external validation can be performed by the use of original software [25], which provide an environment able to compute in a short time, without any experiments the inhibition activity on CA II of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides.

Conclusions

The inhibition activity on CA II of studied substituted 1,3,4-thiadiazole and 1,3,4-thiadiazoline-disulfonamides reveal that the activity is related with compounds structure information, being in relationship with the geometry of compounds and depending by the partial charge, the cardinality and the group electronegativity.

The internal validation results sustain that the four-varied model is a stable and a valid one, but future studies on new external substituted 1,3,4-thiadiazole and 1,3,4-thiadiazoline-disulfonamides are necessary in order to assess the four-varied model robustness and predictivity.

Acknowledgement

Research was partialy supported through project ET36/2005 by UEFISCSU Romania.

References

[1] Hansch C. A Quantitative Approach to Biochemical Structure-Activity Relationships. Acct Chem Res 1969;2:232-239.

[2] Ng C, Xiao Y, Putnam W, Lum B, Tropsha A. Quantitative structure-pharmacokinetic parameters relationships (QSPKR) analysis of antimicrobial agents in humans using simulated annealing k-nearest-neighbor and partial least-square analysis methods. J Pharm Sci 2004;93:2535-44.

[3] Miri R, Javidnia K, Hemmateenejad B, Azarpira A, Amirghofran Z. Synthesis, cytotoxicity, QSAR, and intercalation study of new diindenopyridine derivatives. Bioorg Med Chem 2004;12:2529-36.

[4] Perkins R, Fang H, Tong W, Welsh WJ. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. Environ Toxicol Chem 2003;22:1666-79.

[5] Venta PJ, Welty RJ, Johnson TM, Sly WS, Tashian RE. Carbonic anhydrase II deficiency syndrome in a Belgian family is caused by a point mutation at an invariant histidine residue (107 His-Tyr): complete structure of the normal human CA II gene. Am J Hum Genet 1991;49:1082-90.

[6] Al Rajeh S, el Mouzan MI, Ahlberg A, Ozaksoy D. The syndrome of osteopetrosis, renal acidosis and cerebral calcification in two sisters. Neuropediatrics 1988;19:162-5.

[7] Al Rajeh S, el Mouzan MI, Ahlberg A, Ozaksoy D. The syndrome of osteopetrosis, renal acidosis and cerebral calcification in two sisters. Neuropediatrics 1988;19(3):162-5.

[8] Aramaki S, Yoshida I, Yoshino M, Kondo M, Sato Y, Noda K, et all. Carbonic anhydrase II deficiency in three unrelated Japanese patients. J Inherit Metab Dis 1993;16:982-90.

[9] Ismail EA, Abul Saad S, Sabry MA. Nephrocalcinosis and urolithiasis in carbonic anhydrase II deficiency syndrome. Eur J Pediatr 1997;156:957-62.

[10] Hansch C, McClarin J, Klein T, Langridge R. A Quantitative Structure-Activity Relationship and Molecular Graphics Study of Carbonic Anhydrase Inhibitors. Mol Pharmacol 1985;27:493-498.

[11] Gao H, Bajorath J. Comparison of binary and 2D QSAR analyses using inhibitors of human carbonic anhydrase II as a test case. Mol Divers 1998;4:115-30.

[12] Mattioni BE, Jurs PC. Development of quantitative structure-activity relationship and classification models for a set of carbonic anhydrase inhibitors. J Chem Inf Comput Sci 2002;42:94-102.

[13] Pastorekova S, Parkkila S, Pastorek J, Supuran CT. Carbonic anhydrases: current state of the art, therapeutic applications and future prospects. J Enzyme Inhib Med Chem 2004;19:199-229.

[14] Supuran TC, Clare WB. Carbonic anhydrase inhibitors - Part 57: Qunatum chemical QSAR of a group of 1,3,4-thiadiazole- and 1,3,4-thiadiazoline disulfonamide with carbonic anhydrase inhibitory properties. Eur J Med Chem 1999;34:41-50.

[15] Jäntschi L. Delphi Client - Server Implementation of Multiple Linear Regression Findings: a QSAR/QSPR Application. Applied Medical Informatics 2004;15:48-55.

[16] Bolboacă S, Jäntschi L. Molecular Descriptors Family on Structure Activity Relationships 3. Antituberculotic Activity of some Polyhydroxyxanthones, Leonardo Journal of Sciences 2005;7:58-64.

[17] Jäntschi L. Molecular Descriptors Family on Structure Activity Relationships 1. The review of Methodology. Leonardo Electronic Journal of Practices and Technologies 2005;6:76-98.

[18] ***, HyperChem, Molecular Modelling System [Internet page]; ©2003, Hypercube [about three screens]; [cited 2005 Sept]. Available from: URL: http://hyper.com/products/

[19] Jäntschi L, Katona G, Diudea VM. Modeling Molecular Properties by Cluj Indices. Commun Math Comput Chem (MATCH) 2000;41:151-188.

[20] Diudea M, Gutman I, Jäntschi L. Molecular Topology, 2nd Edition, Nova Science, Huntington, New York, 2002.

[21] ***, Leave-one-out Analysis. ©2005, Virtual Library of Free Software [cited 2006 March]. Available from: URL: http://vl.academicdirect.org/molecular_topology/mdf_findings/loo/

[22] Steiger JH. Tests for comparing elements of a correlation matrix. Psychol Bull 1980;87:245-51.

[23]***, Training vs. Test Experiment. ©2005, Virtual Library of Free Software [cited 2006 March]. Available from: URL: http://vl.academicdirect.org/molecular_topology/qsar_qspr_s/

[24] Hawkins DM. The Problem of Overfitting. J Chem Inf Comput Sci 2004;44:1-12. [Medline]

[25]***, MDF SAR Predictor, © 2005, Virtual Library of Free Software [cited 2006 March]. Available from: URL: http://vl.academicdirect.org/molecular_topology/mdf_findings/sar

APPLIED MEDICAL INFORMATICS Volume 17, No.3, 4 / 2005

CONTENTS