17th European Symposium on Computer Aided Process Engineering – ESCAPE17
V. Plesu and P.S. Agachi (Editors)

Modelling the Inhibition Activity on Carbonic Anhydrase I of Some Substituted Thiadiazole- and Thiadiazoline- Disulfonamides: Integration of Structure Information

Sorana-Daniela BOLBOACĂ[a], Lorentz JÄNTSCHI[b]

[a]*„Iuliu Hatieganu" University of Medicine and Pharmacy, 13 Emil Isac, 400023 Cluj-Napoca, Romania, sbolboaca@umfcluj.ro*
[b]*Technical University of Cluj-Napoca, 15 Constantin Daicoviciu, 400020 Cluj-Napoca, Romania, lori@academicdirect.org*

Abstract

A structure-activity relationships based on an original molecular descriptors family method has been developed and applied on a sample of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides. Forty compounds were studied for their inhibition activity on carbonic anhydrase I. The molecular descriptors family was generated based on complex information obtained from compounds structure. The structure-activity relationships models were built using the generated descriptors. Significant models with best performances in estimation were identified. The prediction abilities of two multivariate models were analyzed, and the correlation coefficients were compared with the correlation coefficients obtained by previously reported models. The results revealed that the molecular descriptors family on structure-activity relationships is a useful approach in characterization of inhibition activity on carbonic anhydrase I of studied substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides.

Keywords: Molecular Descriptors Family on Structure-Activity Relationships (MDF-SAR), Substituted 1,3,4-Thiadiazole- and 1,3,4-Thiadiazoline-Disulfonamides, Carbonic Anhydrase I (CA I), Inhibition Activity

1. Introduction

Carbonic anhydrases are ubiquitous metallo-enzymes that catalyze the hydration of carbon dioxide and the dehydration of bicarbonate. This reaction is ubiquitous in nature, involving the interchange of gaseous and ionic species crucial to a wide range of physiological and biochemical processes, being fundamental for example in respiration, renal tubular acidification and bone resorption [1].
There are known in human eleven active CA isozymes [2], some of which act in cytosol (I, II and III), others being membrane-bound isozymes (IV, VII, IX, XII and XIV), mitochondrial isozyme (V), and one secreted salivary isozyme (VI). The carbonic anhydrase I it is known to has low catalytic activity comparing with carbonic anhydrase II [3] and medium affinity for sulfonamides [4].
1,3,4-Thiadiazole- and 1,3,4-Thiadiazoline-Disulfonamides played an important role in development of classes of pharmacological agents based on their biological action of inhibition of the carbonic anhydrases enzymes [5].

2. Background, Problem Statement

A number of forty substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides were previously studied as inhibitors on carbonic anhydrase I [6]. The equations of two previously reported QSAR models (Eq.(1) and Eq.(2)) and their statistical characteristics are as follow:

$$\log IC_{50} = 9.29 \cdot 10 - 3 \cdot \Pi_{xx} - 5.72 \cdot 10\text{-}3 \cdot \Pi_{zz} - 13.04 \cdot Q_{Nr2} + 17.07 \cdot Q_{S1} + 1.560 \cdot Q_{S2} + 6.90 \cdot 10^{2} \cdot \mu_x - 50.83$$

$$R^2 = 0.753,\ Q^2 = 0.628,\ s = 0.289,\ F = 16.78,\ n = 40 \quad (1)$$

where Π_{xx}, and Π_{zz} = polarizability tensor, Q_{S1} and Q_{S2} = the changes of the atoms of the primary, respectively secondary sulfonamide group, μ_x = the dipole moment, and Q_{Nr2} = the charges on second N atoms; R^2 = the square of the multiple correlation coefficients, Q^2 = the leave-one-out score, s is the standard errors of estimate, F =the the Fisher variance ratio, and n is the sample size.

$$\log IC_{50} = - 3.68 \cdot 10^{-3} \cdot \Pi_{zz} + 3.152 \cdot Q_{Cr2} + 0.157 \cdot \mu_x + 0.400 \cdot LogP - 24.62 \cdot Q_{O1} - 44.1$$

$$R^2 = 0.700,\ Q^2 = 0.570,\ s = 0.201,\ F = 13.98,\ n = 36 \quad (2)$$

where Q_{Cr2} = the charges on second C atom, $LogP$ = the partition coefficient, and Q_{O1} = the changes of the atoms of the primary sulfonamide group.
Starting from the hypothesis that there is a relationship between the structure of biological active compounds and their structure, an original method called molecular descriptors family on structure-activity relationships (MDF-SAR) was developed. The MDF-SAR method proved its usefulness in estimation and prediction of inhibition activity on CA IV [7] and CA II [8]. The aim of the research was to study the estimation and prediction abilities of the MDF-SAR methodology in modelling of the inhibition activity on carbonic anhydrase I of a sample of forty substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides.

3. Paper approach

3.1. Substituted 1,3,4-Thiadiazole- and 1,3,4-Thiadiazoline-Disulfonamides

A sample of twenty 1,3,4-thiadiazole disulfonamides and twenty 1,3,4-thiadiazoline disulfonamides, with inhibition activity on carbonic anhydrase I was included into the study. The measured inhibition activity on CA I, expressed as logarithm of concentration of the 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides that is required for fifty percent inhibition in vitro (log IC_{50}), was taking from a previously reported study [6].

3.2. MDF-SAR Methodology

The MDF-SAR method integrate the complex information obtained from the structure of the substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides into models in order to explain the inhibition activity of these compounds on carbonic anhydrase I (CA I). A number of six steps were used into modelling process [9].

The compounds preparation for the modelling process was done in the first step. In this step, the three-dimensional structure of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides were built up by using HyperChem software and the file with measured inhibition on CA I was created.

In the second step**,** the Molecular Descriptors Family (MDF) was generated and the values for each descriptor were calculated for the studied compounds. The resulted descriptors have a name of seven letters that explained the modality of descriptor construction: the compound characteristic relative to its geometry (*g*) or topology (*t*) - the 7th letter; the atomic property - the 6th letter [9]; the atomic interaction descriptor - the 5th letter [9]; the overlapping interaction model - the 4th letter [9]; the fragmentation criterion used in calculations - the 3rd letter [10,11]; the cumulative method of fragmentation properties - the 2nd letter [9], and the linearization procedure applied in generation of molecular descriptors - the 1st letter.

The best performing MDF-SAR models were selected in the third step. Three criterion were used in this process: (1) the goodness-to-fit of the model (the correlation coefficient and the squared correlation coefficient; the values closest to ±1 indicated a good model); (2) the co-linearity between pairs of descriptors (a value less tha 0.5 indicated the absence of co-linearity between descriptors); and (3) the significance of the regression model (a p-value less than 0.05 indicating a significant model). Internal validation of the MDF-SAR models was analyzed in the fourth step by using the Leave-one-out Analysis application[1].

[1] http://vl.academicdirect.org/molecular_topology/mdf_findings/loo/

The comparison between the MDF-SAR model and previously reported models was done in the fifth step by using the Steiger's Z test at a significance level of 5% [12].
The prediction ability of the best performing MDF-SAR model was analyzed in the sixth step by using the Training vs. Test application[2]. There were analyzed twelve situations, starting with sample sizes in training set from twenty and increasing with one until thirty-one and corresponding sample sizes in test sets from twenty to nine.

3.3. Results & Discussions

One MDF-SAR model with four descriptors proved to be able to estimate and predict the inhibition activity on CA I of studied substituted 1,3,4-thiadiazole disulfonamides and 1,3,4-thiadiazoline disulfonamides. The MDF-SAR model has the following equation:

$$\hat{Y}_{4d} = 1.14 + 8.79 \cdot 10^{-2} \cdot inPRlQg + 3.52 \cdot 10^{-3} \cdot lPDMoMg + 2.43 \cdot iAMRqQg + 1.04 \cdot inMRkQt \quad (3)$$

where $\hat{Y}_{4d}$ = the estimated inhibition activity on CA I by using the MDF-SAR model with four descriptors, and *inPRlQg*, *lPDMoMg*, *iAMRqQg*, *inMRkQt* are molecular descriptors.
The statistical characteristics of the MDF-SAR model with four molecular descriptors are presented in table 1.

Table 1. MDF-SAR model with four descriptors: statistical characteristics

Parameter (abbreviation)	Value Model with four descriptors (n = 40, v = 4)
Correlation coefficient (r)	0.9579
95% CI for correlation coefficient (95%CI_r)	[0.9212-0.9776]
Squared correlation coefficient (r^2)	0.9175
Adjusted squared correlation coefficient (r^2_{adj})	0.9081
Standard error of estimation (s_{est})	0.1624
Fisher parameter (F_{est})	97†
Cross-validation leave-one-out score ($r^2_{cv\text{-}loo}$)	0.8911
Standard error of leave-one-out analysis (s_{loo})	0.1869
Fisher parameter of loo analysis (F_{pred})	71†
$r^2 - r^2_{cv\text{-}loo}$	0.0264

n = the number of compound;
v = the number of descriptors used by the model; † $p < 0.001$

Analyzing the name of descriptors used by MDF-SAR model it can be say that the inhibition activity on CA I is likely to be of geometry (*inPRlQ**g***, *lPDMoM**g***, *iAMRqQ**g***) as well as topology (*inMRkQ**t***) nature, being depend by atomic relative mass (*lPDMoM**g***) and strongly dependent by the partial charge of the compounds (*inPRl**Q**g*, *iAMRq**Q**g*, *inMRk**Q**t*). Regarding the co-linearity between two descriptors, all squared correlation coefficients had values less than or equal with 0.49. The goodness-of-fit of MDF-SAR model with four descriptors is sustained by the correlation coefficient which is equal with 0.9579 and its squared value (r2 = 0.9175). Almost ninety-two percent from the variation of inhibition

[2] http://vl.academicdirect.org/molecular_topology/qsar_qspr_s/

activity on CA I of studied compounds can be explained by its linear relationship with the variation of the four molecular descriptors used by the model (Eq.(3)).

The value of the cross-validation leave-one-out score sustains the predictive ability of the MDF-SAR model with four descriptors while the value of difference between the cross-validation leave-one-out score and the squared correlation coefficient sustained the stability of the model.

The power of the MDF-SAR model with four descriptors in prediction of inhibition activity on CA I of studied compounds is sustained by the absence of multi co-linearity between descriptors used by the model. The internal validation of the MDF-SAR model with four descriptors was analyzed by splitting the whole set of compounds into training and test, and the results are presented in table 2.

Table 2. Training versus test analysis: results

n_{tr}	r_{tr}	95% CI_{rtr}	F_{tr}	n_{ts}	r_{ts}	95% CI_{rts}	F_{ts}	Zr_{tr}-r_{ts}
20	0.936	[0.842-0.975]	27†	20	0.972	[0.929-0.989]	60†	1.23
21	0.961	[0.905-0.984]	49†	19	0.954	[0.881-0.982]	34†	0.27
22	0.984	[0.961-0.993]	129†	18	0.901	[0.750-0.963]	14†	2.69‡
23	0.945	[0.873-0.977]	38†	17	0.965	[0.902-0.987]	38†	0.65
24	0.934	[0.851-0.971]	33†	16	0.942	[0.837-0.980]	18†	0.18
25	0.892	[0.766-0.951]	19†	15	0.962	[0.881-0.988]	6‡	1.53
26	0.945	[0.880-0.975]	44†	14	0.951	[0.848-0.985]	18†	0.15
27	0.945	[0.882-0.975]	46†	13	0.972	[0.905-0.992]	32†	0.88
28	0.934	[0.860-0.969]	39†	12	0.988	[0.957-0.997]	70†	2.24‡
29	0.958	[0.911-0.982]	67†	11	0.966	[0.872-0.991]	13‡	0.28
30	0.916	[0.830-0.960]	33†	10	0.976	[0.897-0.994]	15‡	1.49
31	0.945	[0.887-0.973]	55†	9	0.981	[0.908-0.996]	21‡	1.18

number of compounds in training (n_{tr}) and test (n_{ts}) sets; correlation coefficient obtained in training (r_{tr}) and test (r_{ts}) sets with associated 95% confidence intervals (95%CI_{rtr} respectively 95%CI_{rts}); Fisher parameter associated with training (F_{tr}) and test (F_{ts}) models; Fisher's Z parameter of correlation coefficients comparison ($Z_{rtr\text{-}rts}$) ; † $p \leq 0.001$; ‡ $0.001 < p < 0.05$

As it can be observed (see table 2) that for all sample sizes of training and test sets the regression models were statistical significant. In 95% of the cases, the correlation coefficients obtained in training and test sets did not exceed the 95%CI of the correlation coefficient of the MDF-SAR model with four variables, this observation sustaining the stability of the model. Just in two cases (for sample sizes in training set equal with 22, respectively 28) there were observed statistical significant differences between correlation coefficients obtained in training and respectively in test sets , but always the values were greater than 0.900 (see table 2).The correlation coefficient obtained by MDF-SAR model proved to be statistical significant greater comparing with the correlation coefficients obtained by previously reported models (Steiger's $Z_{Eq.(1)\ -\ Eq.(3)} = 2.563$, $p = 0.0052$; Steiger's $Z_{Eq.(2)\ -\ Eq.(3)} = 2.965$, $p = 0.0015$). Thus, the MDF-SAR model is able to obtained better results comparing with previously reported models by using a less number of variables.

Comparing the MDF-SAR model (Eq.(3)) with previously reported models (Eq.(1) and Eq.(2)) some remarks can be made: (1) the number of variable is less (four comparing with five - Eq.(2), resperctively six - Eq.(1)); (2) the squared correlation coefficient and the cross-validation leave-one-out score is greater; (3)

in construction of the model are used all compounds (the Eq.(2) used a sample size of 36 compounds).

4. Conclusions and future work

The inhibition activity on CA I of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides can be characterized starting from the complex information obtained from the compounds structures by using the MDF-SAR method. The MDF-SAR model reveal that the inhibition activity on CA I of substituted 1,3,4-thiadiazole- and 1,3,4-thiadiazoline-disulfonamides is likely to be of geometry and topology nature, being related with two atomic properties, the partial charge and relative atomic mass.
The MDF-SPR model can be used in order to predict the inhibition on CA I of new compounds from the same class without any experiments and measurements, by using the MDF SAR Predictor application[3].
Even if the MDF-SAR models are stable and valid, future studies on new external compounds from same class are necessary in order to assess the robustness and predictivity of the MDF-SAR models.

Acknowledgements

The research was supported in part by UEFISCSU Romania through projects ET36/2005 & ET108/ 2006.

References

1. H.F. Deutsch, Int. J. Biochem., 19 (1987) 101.
2. D. Hewett-Emmett and R.E. Tashian, Mol. Phylogen. Evol., 5 (1996) 50.
3. B.C. Tripp, K. Smith and J.G. Ferry, J. Biol. Chem., 276 (2001) 48615.
4. X. de Leval, M. Ilies, A. Casini, J.M. Dogne, A. Scozzafava, E. Masini, F. Mincione, M. Starnotti and C.T. Supuran, J. Med. Chem., 47 (2004) 2796.
5. C.T. Supuran, A. Scozzafava and J. Conway (eds),. CRC - Taylor & Francis, 2004.
6. C.T. Supuran and W.B. Clare, Eur. J. Med. Chem., 34 (1999) 41.
7. L. Jäntschi and S. Bolboacă, Electronic Journal of Biomedicine, 2 (2006) 22.
8. L. Jäntschi, M.L. Ungureşan and S.D. Bolboacă, Applied Medical Informatics, 17 (2005) 12.
9. L. Jäntschi, Leonardo Electronic Journal of Practices and Technologies, 6 (2005) 76.
10. P.V. Khadikar, N.V. Deshpande, P.P. Kale, A. Dobrynin, I. Gutman and G. Domotor, J. Chem. Inf. Comput. Sci., 35 (1995) 547.
11. L. Jäntschi, G. Katona and V.M. Diudea, Commun. Math. Comput. Chem., 41 (2000) 151.
12. J.H. Steiger, Psychol. Bull., 87 (1980) 245.

[3] http://vl.academicdirect.org/molecular_topology/mdf_findings/sar