Applied Medical Informatics
Original Research Vol. 25, No. 3-4/2009, pp: 65-74

Molecular Descriptors Family on Vertex Cutting: Relationships between Acelazolamide Structures and their Inhibitory Activity

Sorana D. BOLBOACĂ[1,*], Monica M. MARTA[1], Carmen E. STOENOIU[2], Lorentz JÄNTSCHI[2]

[1] "Iuliu Haţuieganu" University of Medicine and Pharmacy, 13 Emil Isac, 400023 Cluj-Napoca, Romania.
[2] Technical University of Cluj-Napoca, 15 C-tin Daicoviciu, 400020 Cluj-Napoca, Romania.
E-mail(s): sbolboaca@umfcluj.ro; mmarta@umfcluj.ro; carmen@j.academicdirect.org; lori@academicdirect.org

* Author to whom correspondence should be addressed; Tel.: +4-0264431697; Fax: +4-0264-593847.

Abstract: *Aim*: To investigate the relationship between the structural information of acetazolamides and their inhibitory activity on carbonic anhydrase II. *Material and Method*: A sample of previously reported acetazolamides was studied. A pool of descriptors was calculated based on matrix representation and vertex cut in order to be included in the multiple linear regression analysis. The best performing model in terms of goodness-of-fit was analysed in order to assess its validity and reliability. The model was compared with previously reported models using a series of information and prediction criteria besides the Steiger's Z test. *Results*: A model with a 99.77% determination coefficient proved to be the best performing model. The obtained model proved to have a less than 5% average of the absolute difference between the observed and the estimated inhibitory activity. The information and prediction criteria showed that the obtained model was better than the previously reported models. This conclusion is also sustained by the results of Steiger's Z test (7.78; $p = 3.66 \cdot 10^{-15}$). *Conclusion*: The inhibitory activity on carbonic anhydrase II of acetazolamides proved to be of geometric and topologic nature and depended on the compounds' melting point, relative atomic mass and atomic electronegativity.

Keywords: Sulfonamides; Carbonic Anhydrase II Inhibitors; quantitative Structure-Activity Relationships (qSAR); Molecular Descriptors Family on Vertex (MDFV).

Introduction

Acetazolamides (sulfonamide derivatives) are used to treat glaucoma [1,2], epilepsy [3,4], benign intracranial hypertension [5,6], vertigo and dizziness [7,8], augmented breaths during exposure to hypoxia [9], degenerative ataxias [10], cryptococcosis [11], and post meningitis subdural effusion [12]. They are also used to evaluate chronic cerebral ischemia [13], high altitude sickness (such as acute mountain sickness, high-altitude cerebral oedema, and high-altitude pulmonary oedema) [14,15], cystinuria [16], and dural ectasia [17]. Moreover, acetazolamides are also used as diuretics [18] in various clinical situations such as hypertension, heart failure, renal failure, nephritic syndrome and cirrhosis.

Virtual approaches are used in drug design for the development of new active drugs with fewer side effects compared with the existent ones [19]. The modification of the structure of a known drug using for example the tail approach is one way of developing new compounds [20]. Quantitative structure-activity relationships (qSAR) approaches are tools used to estimate and

predict the activity of biological active molecules [21]. The numerical representation of the molecular structure is used in these approaches [22].

Eroğlu et al. studied a sample of 18 acetazolamides (5 acetazolamide derivatives, eight sulfonamide derivatives, acetazolamide, methazolamide, dichloriphenamide, ethozolamide and dorzolamide) [23]. The best performing model when the whole sample of compounds was used is presented in Eq(1).

$$\text{LogKI} = 5.869 + 0.0017T_e - 0.225\mu + 0.0091S + 0.403\chi \quad \text{Eq(1)}$$

$n = 18$; $R^2 = 0.857$; $F = 19.4$; $s^2 = 0.137$; $R^2_{CV} = 0.789$

where T_e = total energy at 0 K (a.u.); μ = dipole moment (debye); S = entropy at 298 K (cal/M-K); χ = electronegativity (eV); n = sample size; R^2 = determination coefficient; F = Fisher-value; s^2 = variance; R^2_{CV} = determination coefficient in cross-validation (leave-one-out).

The best performing model, after removal of two outliers, and its characteristics is presented in Eq(2).

$$\text{LogK}_I = 3.071 - 0.0020T_e - 0.244\mu + 0.0019S + 0.253\chi - 0.202\varepsilon_H \quad \text{Eq(2)}$$

$n = 16$ (C3 and C17 outliers); $R^2 = 0.943$; $F = 33.2$; $s^2 = 0.067$; $R^2_{CV} = 0.855$

where ε_H = energy of HOMO (eV).

Note that Eq(2) did not respect the relation $n = v \cdot 5$ [24] where n = sample size and v = number of variables in the model (in case of Eq(2) the sample size needed is of $25 = 5 \cdot 5$ compounds).

Our research reports the results of the inhibition constant and molecular descriptors calculated based on matrix representation and vertex cut for a sample of 18 acetazolamides. The best performing model was compared with previously reported models in order to identify the method with the highest performances.

Material and Method

Compounds Set

The sample previously studied by Eroğlu et al. [23] was investigated. The chemical name of the compounds, abbreviation (round brackets) and 2D structure are presented in Figure 1. The structures of the compounds were drawn using Symyx Draw software (version 3.2.0.352)[1]. The inhibition constant on carbonic anhydrase II, expressed in logarithmic scale, is presented in Table 3.

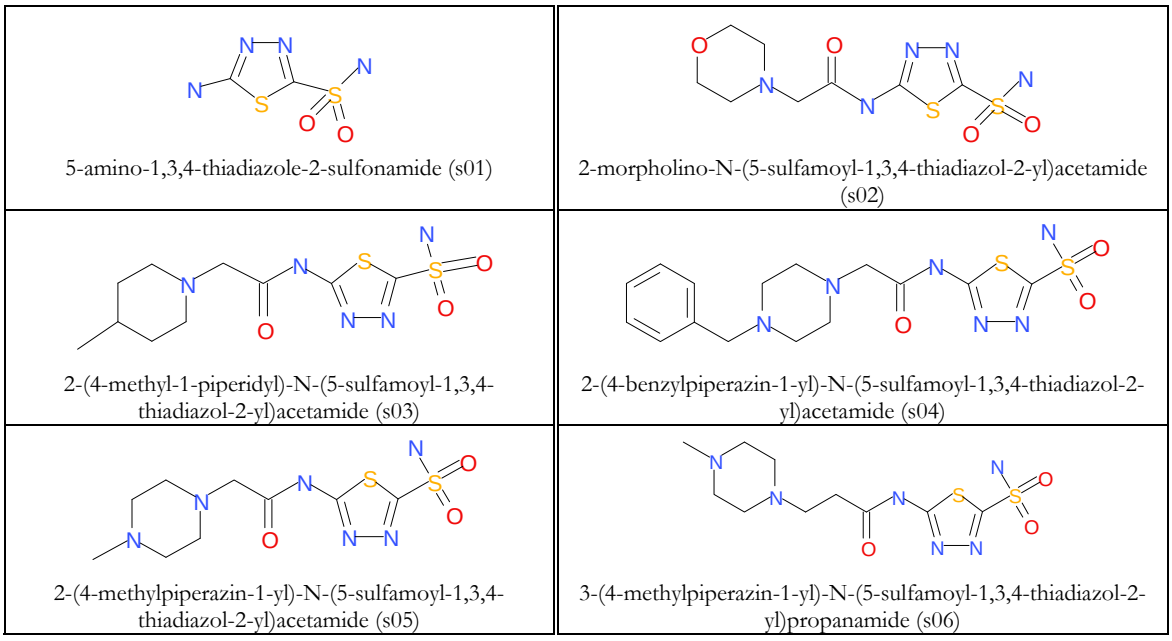

Figure 1. General structure, chemical name and abbreviation of acetazolamide

[1] http://www.symyx.com/micro/getdraw/

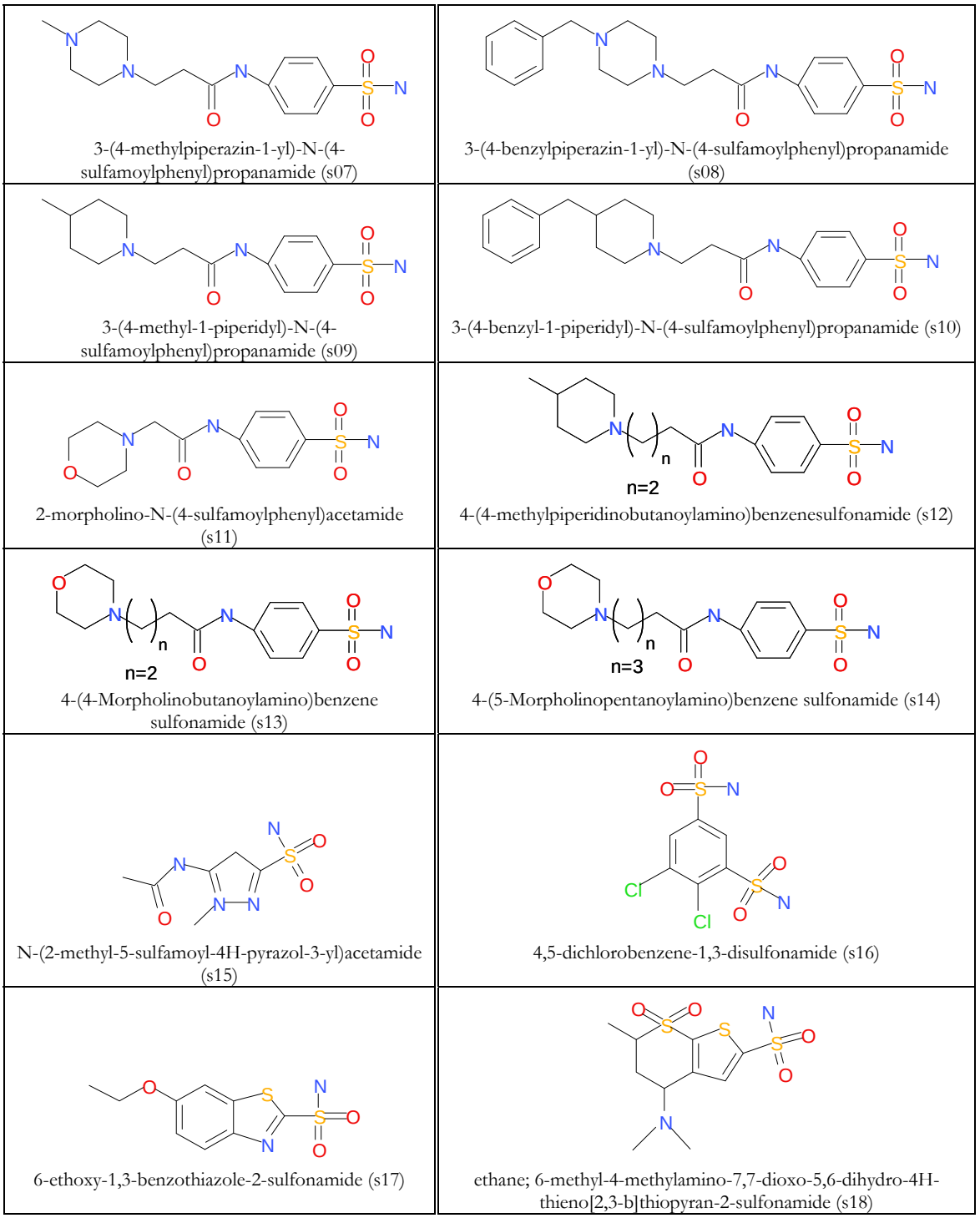

Figure 1. (continuation) General structure, chemical name and abbreviation of acetazolamide

Molecular Modelling: Molecular Descriptors Family on Vertex Cutting

The steps used in the modelling process were as follows:

- *Step 1: Normal Distribution.* Test the normal distribution of experimental data using EasyFit software[2]. Three tests were applied: Kolmogorow-Smirnov [25], Anderson Darling [26] and Chi Squared [27]. The data were considered normally distributed if an agreement between all tests at a 5% significance level was obtained (see Table 1).

[2] http://www.mathwave.com/

Table 1. Tests of Normality

	Kolmogorov-Smirnov	Anderson-Darling	Chi-Squared
Sample size	18	18	18 (df=2)
Statistics	0.1744	0.6114	3.9098
p-value	0.5844	n.a.	0.1416
Critical value (5%)	0.3094	2.5018	5.9915
Reject H_0?	No	No	No

df = degree of freedom; n.a. = not available
H_0 = normally distributed data.

- *Step 2: Test for Outliers*. The Grubbs test [28] was applied in order to identify the presence of outliers. Since data proved to have no outliers (G_{max} = 1.58 ($G_{5\%}$ = 2.78); G_{min} = 0.25 ($G_{5\%}$ = 2.78)) all compounds were introduced in the modelling approach.
- *Step 3: Structure representation*. Draw the 2D structures of acetazolamides using HyperChem[3]; the hydrogen atoms were added and the molecular geometry was built. The molecule structures were saved as *.mol files.
- *Step 4: Structure optimization*. The structures of compounds were optimize as described bellow:
  - Conformational analysis using Molecular Modeling Pro Plus[4].
  - Application of moderate changes using Moly Minimizer (Molecular Modeling Pro Plus).
  - Optimization of compounds geometry using Hyper Chem (PM3 semi-empirical method [29]).
- *Step 5: Compounds validation*. Validate the compounds and compute the partial charges (where necessary) using HyperChem software.
- *Step 6: Calculate MDFV descriptors*. Compute the Molecular Descriptors Family on Vertex Cutting (MDFV) based on the molecular graphs. A detailed description of the MDFV approach is presented in [30].
- *Step 7: Identification of MLR*. Identify the best MDFV multi-linear regression models.
- *Step 8: Models validation*. Validate the identified MDFV models. The criteria used for validation were [31]:▪ agreement of all correlation coefficients between the observed and the estimated activity; ▪ highest explanation of the observed variances; ▪ lowest standard error of the estimated; ▪ highest Fisher value (lowest p-value); ▪ internal validation (leave-one-out); ▪ absence of collinearity between pairs of descriptors.
- *Step 9: Models comparison*. Compare the best performing MDFV model with the previously reported model using: ▪ Akaike information criteria [30];▪ Kubinyi function [30];▪ Steiger's Z test [32].

Results

The best performing MDFV model in terms of goodness-of-fit is presented in Eq(3):

$$\hat{Y}_{MDFV} = 1.01(\pm 0.22) + TLvFFAdR*(2.01 \cdot 10^{-5})(\pm 1.17 \cdot 10^{-6}) + GMpFFIdI*(204.77)(\pm 17.96) + TEmFIIDI*(-0.90)(\pm 0.10) \quad \text{Eq(3)}$$

where the numbers in round brackets represent the parameter needed to compute the 95% confidence intervals for the slope parameters.

The statistical characteristics of the model from Eq(3) are presented in Eq(4).

[3] http://www.hyper.com/
[4] http://www.chemsw.com/

$r^2 = 0.9977$; $r^2_{adj} = 0.9972$; $s_{est} = 0.07$; F (p) = 764 ($9.99 \cdot 10^{-16}$); Eq(4)
$r^2_{cv\text{-}loo} = 0.9898$; $s_{cv\text{-}loo} = 0.09$; $F_{cv\text{-}loo}$ (p) = 453 ($3.56 \cdot 10^{-14}$); E = 0.022
TLvFFAdR: tolerance = 0.905; VIF = 1.105;
GMpFFIdI: tolerance = 0.112; VIF = 8.966;
TEmFIIDI: tolerance = 0.109; VIF = 9.142;
where r^2 = determination coefficient; r^2_{adj} = adjusted determination coefficient; s_{est} = standard error of the estimated; F = F-value; p = p-value; cv-loo = leave-one-out cross-validation; E = error; VIF = variance inflation factor.

The analysis of agreement among the correlation coefficients is presented in Table 2. . The graphical representations of the residuals and of the observed vs estimated activity of acetazolamides are presented in Figures 2 and 3.

Table 2. Correlation analysis between pairs of descriptors and between descriptors and the observed activity

	logKI - ElogKI	logKI - TLvFFAdR	logKI - GMpFFIdI	logKI - TEmFIIDI	TLvFFAdR - GMpFFIdI	TLvFFAdR - TEmFIIDI	GMpFFIdI - TEmFIIDI
r	0.99 ($1.8 \cdot 10^{-22}$)	0.84 ($1.3 \cdot 10^{-5}$)	0.56 (0.02)	0.42 (0.08)	0.27 (0.27)	0.31 (0.22)	0.94 ($5.0 \cdot 10^{-9}$)
ϱ	0.99 ($1.1 \cdot 10^{-16}$)	0.84 ($1.1 \cdot 10^{-5}$)	0.61 (0.01)	0.51 (0.03)	0.30 (0.23)	0.31 (0.21)	0.96 ($5.2 \cdot 10^{-10}$)
semiQ	0.99 ($1.7 \cdot 10^{-18}$)	0.84 ($1.1 \cdot 10^{-5}$)	0.58 (0.01)	0.47 (0.05)	0.29 (0.25)	0.31 (0.21)	0.95 ($5.0 \cdot 10^{-9}$)
τ_a	0.96 ($2.6 \cdot 10^{-8}$)	0.66 ($1.3 \cdot 10^{-4}$)	0.44 (0.01)	0.37 (0.03)	0.20 (0.24)	0.18 (0.31)	0.82 ($1.8 \cdot 10^{-6}$)
τ_b	0.96 ($2.6 \cdot 10^{-8}$)	0.66 ($1.3 \cdot 10^{-4}$)	0.44 (0.01)	0.37 (0.03)	0.20 (0.24)	0.18 (0.31)	0.82 ($1.8 \cdot 10^{-6}$)
τ_c	0.91 ($1.5 \cdot 10^{-7}$)	0.62 ($3.0 \cdot 10^{-4}$)	0.41 (0.02)	0.35 (0.02)	0.19 (0.27)	0.17 (0.33)	0.78 ($6.6 \cdot 10^{-6}$)
γ	0.96 ($8.8 \cdot 10^{-8}$)	0.68 (0.01)	0.45 (0.25)	0.37 (0.42)	0.21 (0.79)	0.18 (0.85)	0.86 ($1.6 \cdot 10^{-5}$)

R = Pearson's correlation coefficient; ϱ = Spearman's rank correlation coefficient;
semiQ = semi-quantitative correlation coefficient;
τ_a, τ_b, τ_c = Kendall's tau a, b, and c correlation coefficient; γ = Gamma correlation coefficient;
Blue = Hypothesis of linear dependence can be accepted at a significance level of 1%.
Red = Hypothesis of linear dependence is rejected at a significance level of 10%.

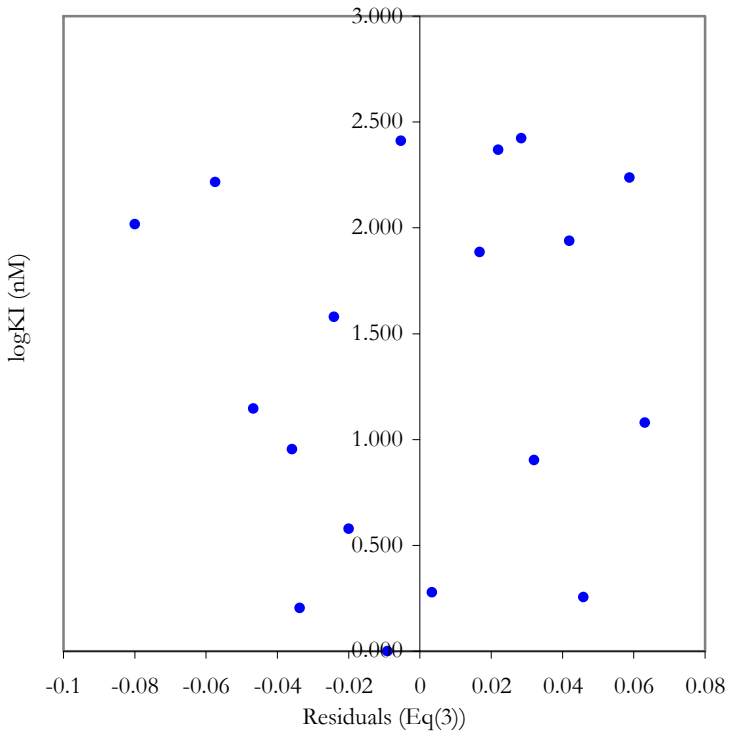

Figure 2. Observed inhibitory activity vs residuals

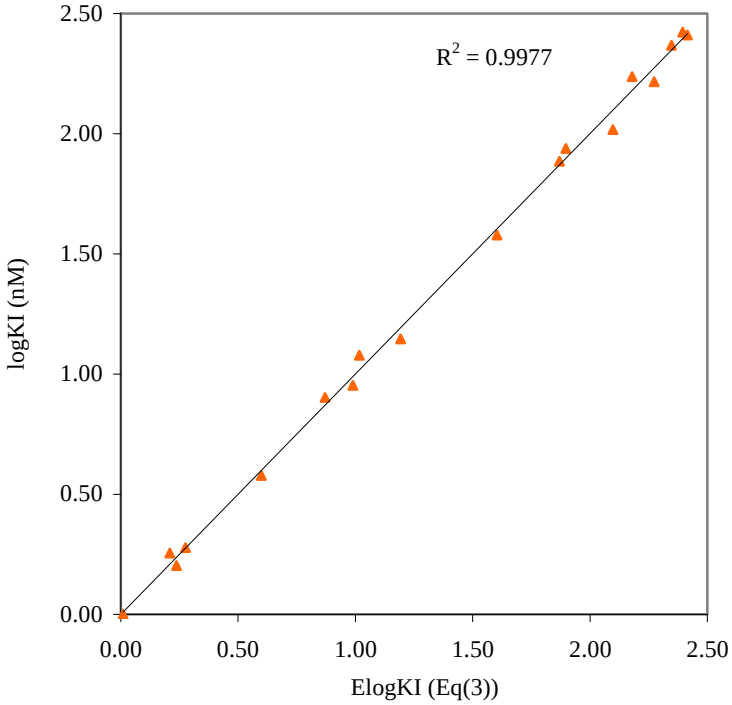

Figure 3. Goodness-of-fit of MDFV model (Eq(3))

Table 3 presents the values of the MDFV descriptors used by the model presented in Eq(3) as well as the observed and estimated activity of acetazolamides.

A collinearity analysis was performed and the results, expressed as correlation coefficients and associated probabilities, are presented in Table 2, while Table 4 includes the parameters of collinearity diagnosis.

The obtained MDFV model (Eq(3)) was compared with the previously reported models (Eq(1), Eq(2)) in terms of information criteria. The results are presented in Table 5. The goodness-of-fit of the MDFV model (Eq(3)) was compared with the goodness-of-fit of the previously reported model (Eq(4)) using the Steiger's Z test and a value of 7.78 (p –value = $3.66 \cdot 10^{-15}$) was obtained.

Table 3. Acetazolamides: MDFV Descriptors from Eq(3), Observed and Estimated Activity

Mol	logKI (nM)	TLvFFAdR	GMpFFIdI	TEmFIIDI	ElogKI ($\hat{Y}_{MDFV}$)(nM)	Diff%
s01	1.079	$5.82 \cdot 10^{4}$	0.004	2.180	1.016	6
s02	0.000	$3.16 \cdot 10^{4}$	0.010	4.093	0.009	0
s03	0.579	$3.55 \cdot 10^{4}$	0.015	4.608	0.599	3
s04	0.255	$3.09 \cdot 10^{4}$	0.020	6.086	0.209	18
s05	0.204	$3.11 \cdot 10^{4}$	0.013	4.423	0.238	17
s06	0.278	$3.21 \cdot 10^{4}$	0.014	4.700	0.275	1
s07	2.217	$9.01 \cdot 10^{4}$	0.020	5.193	2.274	3
s08	2.369	$8.94 \cdot 10^{4}$	0.028	6.856	2.347	1
s09	2.238	$1.12 \cdot 10^{5}$	0.017	5.100	2.179	3
s10	2.411	$1.12 \cdot 10^{5}$	0.027	7.040	2.416	0
s11	1.939	$8.78 \cdot 10^{4}$	0.016	4.586	1.897	2
s12	2.423	$9.58 \cdot 10^{4}$	0.020	5.140	2.395	1
s13	2.017	$9.58 \cdot 10^{4}$	0.019	5.140	2.097	4
s14	1.886	$9.58 \cdot 10^{4}$	0.018	5.140	1.869	1
s15	1.146	$3.61 \cdot 10^{4}$	0.011	3.084	1.193	4
s16	0.903	$5.44 \cdot 10^{4}$	0.010	3.682	0.871	4
s17	1.579	$1.40 \cdot 10^{5}$	0.006	3.774	1.603	2
s18	0.954	$8.24 \cdot 10^{4}$	0.012	4.606	0.990	4

Table 4. Collinearity Diagnostic

Dimension	Eigenvalue	Condition Index	Variance Proportions			
			(Constant)	TLvFFAdR	GMpFFIdI	TEmFIIDI
1	3.797	1.000	0.00	0.01	0.00	0.00
2	0.131	5.384	0.00	0.85	0.02	0.00
3	0.068	7.491	0.27	0.13	0.07	0.00
4	0.004	30.494	0.72	0.01	0.91	1.00

Table 5. Validation and comparison of the models

Parameter	**Model**		
	MDFV-Eq(3)	**PREV[23]-Eq(1)**	**PREV[23]-Eq(2)**
AIC_c (corrected Akaike information criterion)	-104.49	-26.52	-39.50
$w_i(AIC_c)$	1.00	$1.17 \cdot 10^{-17}$	$7.71 \cdot 10^{-15}$
AIC_{R2} (AIC based on determination coefficient)	-0.92	5.17	6.24
$w_i(AIC_{R2})$	0.93	0.04	0.03
AIC_u (McQuarrie and Tsai corrected AIC)	-4.34	0.11	-0.48
$w_i(AIC_u)$	0.80	0.09	0.12
BIC (Schwarz (or Bayesian) Information Criterion)	-99.48	-21.21	-34.50
APC (Amemiya prediction criterion)	0.00	0.18	0.07
HQC (Hannan-Quinn Criterion)	-107.08	-30.91	-46.40
FIT (Kubinyi function)	159.03	1.67	3.38

w_i = Akaike weights for model *i*.
Parameters: The smallest is the best except for FIT and w_i (where the largest the best);
PREV[23] = previously reported regression models

Discussion

The inhibitory activity on carbonic anhydrase II of a sample of 18 acetazolamides was successfully modelled using the molecular descriptors family on the vertex cutting approach.

A model using three molecular descriptors with good estimation and prediction abilities was obtained (Eq(3)) based on cutting the vertices of acetazolamide graphs. The MDFV model was selected from a pool of models by applying Hawkins's criteria [24]: highest correlation coefficient, highest Fisher parameter, lowest standard error of the estimated, and the smallest possible number of significant parameters ($n = 5 \cdot v$, where n = sample size and v = number of variables in the model) (see Eq(4)). The MDFV model used three descriptors with the following contributions [30] to the inhibitory activity of acetazolamides:

qSAR model	Eq(3)
Interaction Via	Bonds (topology - *T*) & Space (geometry - *G*)
Dominant Atomic Property	Melting point (*L*) & Relative atomic mass (*M*) & Atomic electronegativity (E)
Structure on Activity Scale	Reciprocal (*R*) & Identity (*I*) & Identity (*I*)

The MDFV model proved capable of estimating the inhibitory activity of acetazolamides. The combination of the descriptors used proved able to estimate 99.77% of the inhibitory activity. The prediction ability of the model could be assessed through the determination coefficient obtained in leave-one-out cross-validation analysis. Thus, the dataset of 18 acetazolamides was subject to 18 experiments; for each experiment 17 compounds were used in training sets and 1 compound in the test set (the true error is estimated as the average error rate on test examples $E = (1/18) \cdot \sum_{i=1}^{18} E_i$, where E = the true error, E_i = the error in experiment i_{th}, $1 \le i \le 17$).

The distance between the determination coefficient of the MDFV model and the determination coefficient obtained in leave-one-out cross-validation analysis was able to predict the ability of each model. In our case, a 1% difference between these two determination coefficients supported the prediction ability of the MDFV model (Eq(3)). Moreover, the analysis of the absolute difference

between experimental and estimated logKI sustains the validity of the model, the difference being of 4.11%.

A collinearity analysis was conducted in order to validate the model. Several diagnosis methods were used: bivariate correlation matrix (significant correlation coefficients; see Table 2), tolerance ($\leq$ 0.01 indicate multicolinearity) and variance inflation factor ($\geq$ 10 indicates the presence of collinearity [33,34]) (see Eq(4)), condition index and variance proportions. A conditional index (a measure of how "dependent" one independent variable is on the others) $\geq$ 30 and the presence of at least two variance proportions for a particular independent variable > 50 indicate multicollinearity [35,36]. The analysis of all these parameters obtained for our regression model (Eq(4), Table 2 & 4) indicates the absence of multicollinearity. The highest value of one condition index was slightly higher than 30 (obtained at the 4th dimension with a significant correlation between the GMpFFIdI and the TEmFIIDI descriptors). Thus, since all the above-described criteria were not met, we concluded that there was no collinearity in the model from Eq(3).

A valid and reliable regression model able to explain the relationship between the structure of acetazolamides and their inhibitory activity on carbonic anhydrase II was obtained. Since this sample of compounds was previously studied, the main question was: Is the MDFV model significantly better than the previously reported models? A comparison analysis between the MDFV model (Eq(3)) and the previously reported models [23] (Eq(1) and Eq(2)) was conducted using a series of information and prediction criteria. The analysis of the results obtained showed that the MDFV model performed better in terms of goodness-of-fit compared with the previously reported and investigated models (Eq(1)& Eq(2)) (see Table 5). This conclusion is also supported by the result of the Steiger's Z test when the MDFV model was compared with the model from Eq(1). The model from Eq(2) was not subject to Steiger's Z test since it failed to meet Hawkins's criterion referring to sample size (the sample size should not be lower than five fold the number of parameters in the equation).

Our study aimed to model the inhibitory activity of a sample of acetazolamides by using as much information as possible from the structure of the compounds. A valid and reliable model with three descriptors was obtained. Useful information related to the structural nature of the inhibitory activity on carbonic anhydrase II of acetazlamides was obtained. Since the analysis was conducted on such a small sample size of acetazolamides, the obtained results should be assessed on a larger sample in order to allow generalization. The analyzed sample size and the absence of the external validation set are the main limitations of the present study. Further investigations needed to validate the approach are currently being carried out.

Conclusions

The MDFV approach provided a reliable and valid model able to explain the relationship between the structure of studied acetazolamides and their inhibitory activity on carbonic anhydrase II. The MDFV model proved to be the best model for the studied acetazolamides compared to previously reported models in terms of information and prediction criteria, Kubinyi function, Akaike's weights, and Steiger's Z test.

Acknowledgements

The research was supported by the Executive Unit for Financing Higher Education and Scientific Research University Project no: ID0458 (206/01.10.2007).

References

1. Funk J. Glaucoma. Therapeutische Umschau 2009;66(3):173-181.
2. Law S.K. Switching within glaucoma medication class. Current Opinion in Ophthalmology 2009;20(2):110-115.

3. Stern J.M. Overview of evaluation and treatment guidelines for epilepsy. Curr Treat Options Neurol 2009;11(4):273-284.
4. Elger CE, Schmidt D. Modern management of epilepsy: A practical approach. Epilepsy Behav 2008;12(4):501-53.
5. Dhellemmes P, Defoort S, Vinchon M. Benign intracranial hypertension: The role of medical treatment. Neurochirurgie 2008;54(6):717-720.
6. Bershad EM, Humphreis III WE, Suarez JI. Intracranial hypertension. Semin Neurol 2008;28(5):690-702.
7. Strupp M, Brandt T. Diagnosis and therapy of dizziness and oculomotoric disorders. Nervenheilkunde 2009;28(7):456-462.
8. Fotuhi M, Glaun B, Quan SY, Sofare T. Vestibular migraine: A critical review of treatment trials. J Neurol 2009;256(5):711-716.
9. Bell HJ, Haouzi P. Acetazolamide suppresses the prevalence of augmented breaths during exposure to hypoxia. Am J Physiol Regul Integr Comp Physiol 2009;297(2):R370-R381.
10. Trujillo-Martín Ma del M, Serrano-Aguilar P, Monton-acAlvarez F, Carrillo-Fumero R. Effectiveness and safety of treatments for degenerative ataxias: A systematic review. Mov Disord 2009;24(8):1111-1124.
11. Huston SM, Mody CH. Cryptococcosis: An Emerging Respiratory Mycosis. Clin Chest Med 2009;30(2):253-264.
12. Bhat Y, Prakashini K, Sen S. Subdural effusion complicating neonatal meningitis: Successful treatment with acetazolamide. Indian J Pediatr 2009,76(1):103-105.
13. Vagal AS, Leach JL, Fernandez-Ulloa M, Zuccarello M. The acetazolamide challenge: Techniques and applications in the evaluation of chronic cerebral ischemia. Am J Neuroradiol 2009;30(5):876-884.
14. Peacock AJ. Medical problems of high altitude. J R Coll Physicians Edinb 2008;38(2):126-128.
15. Luks AM, Swenson ER. Medication and dosage considerations in the prophylaxis and treatment of high-altitude illness. Chest 2008;133(3):744-755.
16. Bandi G, Nakada SY, Penniston KL. Practical approach to metabolic evaluation and treatment of the recurrent stone patient. Wis Med J 2008;107(2):91-100.
17. Ahn NU, Ahn UM, Nallamshetty L, Springer BD, Buchowski JMS, Funches L, Garrett ES, Kostuik JP, Kebaish KM, Sponseller PD. Cauda equina syndrome in ankylosing spondylitis (the CES-AS syndrome): Meta-analysis of outcomes after medical and surgical treatments. J Spinal Disord 2001;14(5):427-433.
18. Reddy P, Mooradian AD. Diuretics: An update on the pharmacology and clinical uses. Am J Ther 2009;16(1):74-85.
19. Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. Brief Bioinform 2009;10(5): 579-591.
20. Turkmen H, Durgun M, Yilmaztekin S, Emul M, Innocenti A, Vullo A, Scozzafava A, Supuran CT. Carbonic anhydrase inhibitors. Novel sulfanilamide/acetazolamide derivatives obtained by the tail approach and their interaction with the cytosolic isozymes I and II, and the tumor-associated isozyme IX. Bioorg Med Chem Lett 2005;15(2):367-372.
21. Hammett LP. Some Relations between Reaction Rates and Equilibrium Constants. Chem Rev 1935;17:125-136.
22. Liu P, Long W. Current mathematical methods used in QSAR/QSPR studies. Int J Mol Sci 2009;10(5):1978-1998.
23. Eroğlu E, Türkmen H, Güler S, Palaz S, Oltulu O. A DFT-Based QSARs Study of Acetazolamide/Sulfanilamide Derivatives with Carbonic Anhydrase (CA-II) Isozyme Inhibitory Activity. Int J Mol Sci 2007;8(2):145-155.
24. Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci 2004;44:1-12.
25. Kolmogorov A. Confidence Limits for an Unknown Distribution Function. The Annals of Mathematical Statistics 1941;12(4):461-463.
26. Anderson TW, Darling DA. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Annals of Mathematical Statistics 1952;23(2):193-212.

27. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine 1900;50:157-175.
28. Grubbs F. Procedures for Detecting Outlying Observations in Samples. Technometrics 1969;11(1):1-21.
29. Clare BW, Supuran CT. Carbonic anhydrase inhibitors - Part 57: Quantum chemical QSAR of a group of 1,3,4-thiadiazole- and 1,3,4-thiadiazoline disulfonamides with carbonic anhydrase inhibitory properties. Eur J Med Chem 1999;34:41-50.
30. Bolboacă SD, Jäntschi L. Comparison of quantitative structure-activity relationship model performances on carboquinone derivatives. TheScientificWorldJOURNAL 2009;9:1148-1166.
31. Bolboacă S, Jäntschi L. Pearson Versus Spearman, Kendall's Tau Correlation Analysis on Structure-Activity Relationships of Biologic Active Compounds. Leonardo Journal of Sciences 2006;9:179-200.
32. Steiger JH. Tests for comparing elements of a correlation matrix. Psychol Bull 1980;87:245-251.
33. Norušis, MJ. The SPSS Guide to Data Analysis. for Release 4. Chicago: SPSS Inc., 1990
34. Stevens J. Applied Multivariate Statistics for the Social Sciences (4th ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002.
35. Belsley DA, Kuh E, Welsch RE. Regression Diagnostics: Identifying influential data and sources of collinearity. New York: John Wiley, 1980.
36. Tabachnick BG, Fidell LS. Using Multivariate Statistics, Fourth Edition. Needham Heights, MA: Allyn & Bacon, 2001.