J. Indian Chem. Soc.,
Vol. 93, July 2016, pp. 775-780

# Computational deign and QSPR study on carbonic anhydrase mitochondrial isozymes VA inhibitors : As an anti obesity agent†

**Shalini Singh**

QSAR & Cheminformatics Laboratory, Department of Chemistry, Bareilly College, Bareilly-243 001, Uttar Pradesh, India

*E-mail* : shalinisingh_15@yahoo.com *Fax* : 91-581-2567808

**Abstract : This paper presents result of quantitative structure-activity relationships (QSAR) study realized with the PRECLAV, omega, brood and MOPAC software. The dependent property is the inhibitory activity against human carbonic anhydrase mitochondrial isoforms VA. The calibration set includes 12 2-substituted-1,3,4-thiadiazole-5-sulfamides heterocyclic with two clinically used CA inhibitors namely AZA, and ZNS molecules. The prediction set contains nine others not yet synthesized substituted heterocyclic sulphonamides having unknown observed values of activity. In the presence of prediction set, the predictive quality of QSAR of hCA VA ($r^2$ = 0.9528, $F$ = 60.5698, $r^2_{CV}$ = 0.9052) is large. The obtained models suggest a slightly different inhibition mechanism for the isoforms. Large percentage, in weight, of $C_2HN_3$ molecular fragments seems to be not favorable to inhibitory activity of VA.**

**Keywords : QSAR, omega, brood, PRECLAV, carbonic anhydrase VA.**

## Introduction

There are 16 α-carbonic anhydrase (CA, EC 4.2.1.1) isoforms present in mammals, CA I-CA XV, A two CA isozymes, VA and VB, are present in mitochondria[1–5]. These two isozymes are involved in several biosynthetic processes, such as ureagenesis, gluconeogenesis and lipogenesis[1,2,6–8]. As hCA VA/VB are involved in several biosynthetic processes catalyzed by pyruvate carboxylase, acetyl carboxylase and carbamoyl phosphate synthetases I and II, providing the bicarbonate substrate to these carboxylating enzymes involved in fatty acid biosynthesis, their inhibition may lead to the development of anti-obesity agents possessing a new mechanism of action[8]. Inhibition data for classical sulphonamide CA inhibitors (CAIs) used clinically, such as AZA. Supuran *et al.* reported the synthesis of 1,3,4-thidiazole sulfamides possessing various 2-substitutents[9,10]. This scaffold has been chosen as it is present in one of the most investigated and powerful CAI, acetazolamide AZA, used clinically since 1956 and also because its binding to the enzyme is effective, sulphonamides possessing varied structures, incorporating act as potent inhibitors of human mitochondrial isoforms VA. Such compounds may be useful for the development of novel anti-obesity therapies[9]. The prediction set includes molecules having unknown observed values of dependent property. The quantitative structure-activity relationships (QSAR) studies can be made in absence or in presence of certain prediction set. In the absence of the prediction set, the purpose of QSAR studies is the identification of the molecular features with the highest impact (favourable or unfavourable) on the biochemical activity. In the presence of the prediction set, the purpose is to identify the prediction set molecules having the largest computed activity. The search for new human mitochondrial isoforms VA inhibitors is important for medicinal chemistry. Therefore, the structures of the prediction set molecules were selected mainly by their possibility to be synthesized in laboratory conditions and taking into account the commercial availability of the raw materials.

## The calibration set and the prediction set

Supuran reported for the first inhibition study against mitochondrial isoform hCA VA with 2-substituted-1,3,4-thiadiazole-5-sulfamides with clinically use CA inhibitors

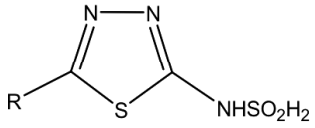
†**In honour of Professor M. C. Chattopadhyaya on the occasion of his 70th birth anniversary.**

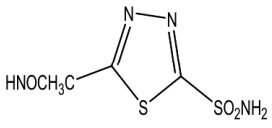
such as AZA and ZNS (Table 1) were included in the calibration set. The inhibitory activity (as KI values, in the micromolar range for isozymes) was expressed by means of the equation $A = \log (c/K)$, where $c$ was taken as 42000 in order to obtain large values of '$A$'. The inhibition constant value '$A$' of the molecules under the study spanned a range from 1 to 4 is more suggestive. In the above mentioned study[9] it has been also observed that sulphonamides of type **1-12** (Table 1) act as hCA VA inhibitors, with variable efficacies, depending on their chemical structure. The prediction set contains 8 other not yet synthesized substituted heterocyclic sulphonamides generated by Brood10 software (version 2.0.0, open eye science software, Santa Fe, USA), having unknown observed values of activity (Table 1) and structure presented in Fig. 2. Brood[11] uses the shape and attachment geometry of the query fragment to identify a family of similar fragments. The discovery of novel bioactive molecule is the primary goal of computational drug discovery.

**Descriptor calculation and quality of the model**

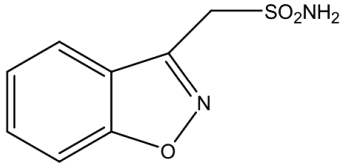
The minimum energy geometry for each of the molecule in the calibration and prediction set was obtained by the conformational search ability of the Omega v.2.4. 3[12–14] program. The isomeric SMILES notation was used as program input in order to avoid any influences on conformational model generation by presenting 3D seed structures. The omega employs a rule-based algorithm in combination with variants of the MMMF 94. The force field used was the 94s variant of the MMMF_NoEstat[12–14] includes all MMFF terms except coulomb interactions. The conformations of minimum energy obtained by molecular mechanics calculations were further optimized by quantum chemical calculations. The semi empirical PM6 method[15] included in the MOPAC2009 software[16] optimized the geometry more rigorously. The criterion for the energy cutoff during PM6 minimization is completed with the following sequence of keyword. Based on the output files created by MOPAC, the PRECLAV software calculated, for each molecule, more than 1000 *whole molecule* descriptors, specific to this program[17,18]. The PRECLAV software was used for identification of the 'significant' molecular fragments and for statistical computations. PRECLAV divides the analyzed molecules into

**Table 1.** Value of the predictors, observed inhibition constant (in μM) and their corresponding $A$ value where $A = \log (42000/KA)$, estimated inhibition constant ($A$), hat diagonal, standardized residual, $|RStudent|$ of the calibration set molecules **1-12** of hCA VA and predicted value and hat diagonal of **13–20** bot yet synthesized compounds




AZA acetazolamide




ZNS zonisamide

| Compd. no. | R | Obs. | Pred. | Res. | Residual | *RStudent* | Diagonal | Compd. no. | Predicted. | Hat Diagonal |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | H | 3.172 | 3.253 | –0.082 | –1.1553 | –1.1839 | 0.3066 | **13** | 3.884 | 0.2728 |
| **2*** | **Et** | **3.351** | 3.357 | –0.006 | –0.0766 | –0.0717 | 0.1499 | **14** | 4.126 | 0.5496 |
| **3** | t-Bu | 3.606 | 3.55 | 0.056 | 0.6974 | 0.6732 | 0.0985 | **15** | 4.016 | 0.3989 |
| **4** | $CF_3$ | 3.76 | 3.725 | 0.035 | 0.4744 | 0.4501 | 0.2576 | **16** | 3.67 | 0.1467 |
| **5*** | **MeS** | **3.118** | 3.166 | –0.048 | –0.861 | –0.8455 | 0.566 | **17** | 4.126 | 0.5503 |
| **6** | EtS | 3.655 | 3.503 | 0.152 | 1.9275 | 2.4637 | 0.1411 | **18** | 3.666 | 0.1457 |
| **7*** | **Ph** | **3.66** | **3.67** | –0.011 | –0.1362 | –0.1276 | 0.1467 | **19** | 3.977 | 0.5075 |
| **8** | 4-$MeOC_6H_4$ | 3.72 | 3.726 | –0.005 | –0.0775 | –0.0725 | 0.3242 | **20** | 3.577 | 1.0959 |
| **9** | 4-Br-$C_6H_4$ | 4 | 4.095 | –0.095 | –1.7001 | –1.9898 | 0.5639 | | | |
| **10** | $MeSO_2$ | 3.684 | 3.607 | 0.077 | 1.0259 | 1.0297 | 0.2265 | | | |
| **11** | AZA | 2.824 | 2.818 | 0.006 | 1.3248 | 1.4026 | 0.9968 | | | |
| **12** | ZNS | 3.322 | 3.401 | –0.078 | –1.0462 | –1.0533 | 0.2223 | | | |

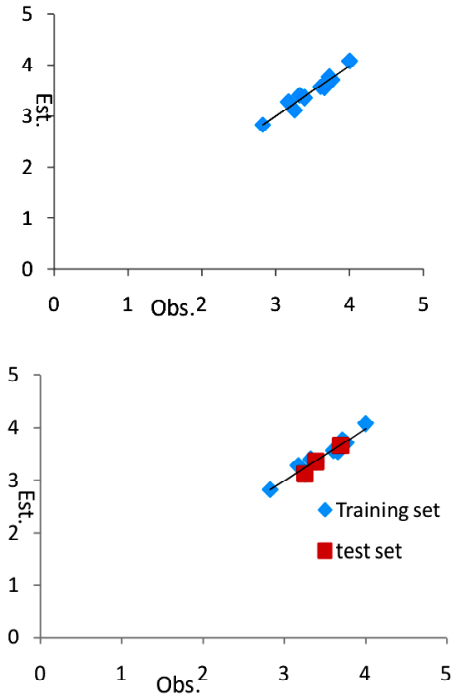



**Fig. 1.** Correlation of observed vs estimated KI in the calibration set and validation set.

virtual fragments using an algorithm reported earlier[19,20]. The virtual fragments identified by PRECLAV do not always coincide with the classical functional groups. The presence of a significant fragment in the molecule greatly influences (in a positive or negative way) the inhibitory activity of the molecule.

The program PRECLAV computes type (2) multilinear QSARs.

$$A = C_0 + \sum_{i=1}^{k} C_i \bullet D_i \quad (1)$$

where $A$ is (the value of) activity; $C_0$ is the free term (intercept); $C_i$ are coefficients (weighting factors); $D_i$ are (the value of) significant descriptors; $k$ is the number of descriptors.

The square of Pearson linear correlation $r^2$ of observed/computed values, the Fisher function $F$, the standard error of estimation *SEE*, and the quality function $Q$[18] are criteria for the quality of prediction for the molecules in calibration set.

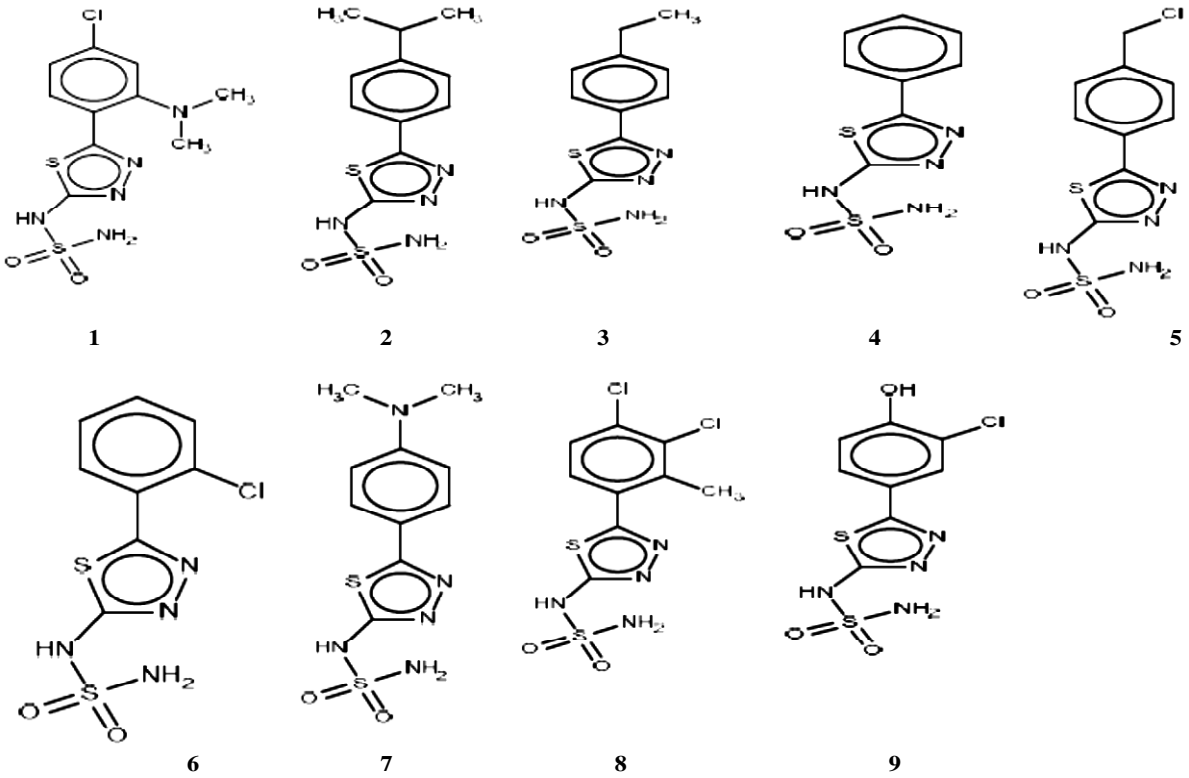



**Fig. 2.** Predicted compound.

$$F = r^2/(1 - r^2) \bullet (N - p)/p \quad (2)$$

$$SEE = [(\Sigma\Delta^2)/(N - 1)]^{1/2} \quad (3)$$

$$Q = r^2 \bullet (1 - p/N) \quad (4)$$

where $p$ is number of descriptors; $N$ is number of molecules in the calibration set; $\Delta$ is difference $A_{obs} - A_{calc}$.

The descriptors included in the best (by $Q$ function) QSAR are named 'predictors'.

The relative utility of predictors is computed by the formula (6).

$$U = (R^2 - r^2)/(1 - r^2) \quad (5)$$

where $R^2$ is the square of Pearson correlation between the observed values and the computed values (using $p$ predictors), $r^2$ is the square of Pearson correlation between the observed values and the calculated values (using the $p$-1 predictors, i.e. the QSAR equation without the analyzed predictor).

After computation of $U$[18] for each predictor, the values of $U$ are normalized by the highest of them (the highest value for $U$ becomes 1000). The predictors with high enough value of $U$ ($U > 500$) can be considered 'with high relative utility'. These predictors are useful because they correlate well with $A_{obs}$ values and present low correlation with other descriptors.

PRECLAV[18] calculates square of cross-validated correlation $r^2_{CV}$ using LHO (Leave Half Out) method. However, this usual method is applied after ordering of molecules in calibration set according to the observed values of activity. Therefore, the cross-validated function $r^2_{CV}$ is a measure of homogeneity of calibration set from the point of view of predictors' set, i.e. from the point of view of structure-activity relationship. A low value ($< 0.4$) of $r^2_{CV}$ means 'the QSAR for molecules having high values of activity and the QSAR for molecules having low values of activity include the same descriptors, but very different weighting factors'. Actually, the computation of $r^2_{CV}$ is a very drastic 'internal validation test'.

After computing the $A_{calc}$ values of the activity for the prediction set molecules, the program computes the average value $A_{calc}{}^m$ and the standard deviation $\sigma$ of the estimated values. The program considers 'high values' the values fulfilling the criterion (7) and 'low values' the values fulfilling the criterion (8).

$$A_{calc} > A_{calc}{}^m + 0.5 \bullet \sigma \quad (6)$$

$$A_{calc} < A_{calc}{}^m - 0.5 \bullet \sigma \quad (7)$$

Therefore, each estimated value $A_{calc}$ is 'high' or 'low' not in absolute manner, but relating to the other estimated values[18].

**Applicability of domain and detection of outliers**

A QSAR model can be used for showing new compounds if its domain of application is defined[19,20]. The need to characterize the model applicability domain is also reflected in the OECD guiding principle for QSAR model validation[21,22]. QSAR model should only be used for making predictions of compounds fall within the specified domain may be considered reliable. Extent of extrapolation[23,24] is one simple approach to define the applicability of the domain. It is based on the calculation of the hat diagonal (leverage) $h_i$ for each chemical, where the QSAR model is used to predict its activity :

$$h_i = ¼ x_i{}^T (X^T X)^{-1} x_i \quad (8)$$

In eq. (8), $x_i$ is the descriptor-row vector of the query molecule and $X$ is the $k$ x $n$ matrix containing the $k$ descriptor values for each one of the $n$ training molecules. A hat diagonal (leverage) value $> 3(k + 1)/n$ leverage warning limit[22] is considered large.

Outliers are compounds that are poorly fit by the regression model. Outlying compounds should not be removed unless a good reason for their removal can be given. The variance of the observed residuals is not constant. This makes comparisons among the residuals difficult. One solution is to standardize the residuals[25,26] by dividing by their standard deviations. This gives a set of standardized residuals. The cross-validated LOO standardized residuals is a $|RStudent|$ that has the impact of a single observation removed from the mean square error. A molecule is defined as an outlier in which $|RStudent| > 2$[26].

To visualize the applicability of domain of a developed QSAR model, William plot was used. In the William plot, $|RStudent|$ versus leverage values ($h_i$) are plotted. This plot could be used for an immediate and simple graphical detection of both the response outliers and structurally influential compounds in a model. It must be noted that compounds with high value of leverage and good fitting in the developed model can stabilize the model. On the other hand, compounds with bad fitting in the

developed model may be outliers. Thus, combination of leverage and the |*RStudent*| could be used for assigning the applicability of domain.

## Results and discussion

The statistical computations were conducted using the specific formulas and procedures of PRECLAV program algorithm. Using only the "significant" descriptors. PRECLAV computes ten thousand QSAR type (3) multilinear equations. The quality of the obtained equation is reflected by the value of the *Q* function and also by values of some usual statistical functions. During the PRECLAV MLR analysis, I observed that the 3-parametric model has the highest value of the *Q* function for hCA VA inhibitors and also has the highest predictive power as follows :

Dependent property : Inhibition constant (*A*) for hCAVA

Molecules number in calibration set : 12

Number of "significant" descriptors in presence of set : 277

$A = -0.4755 - 2.3604(acc) + 0.6371(mot) - 0.168\ (scj)$

acc = Minimum net charge of heteroatoms ($U = 1000$),

mot = Molecular mass weighted moment of inertia *C* ($U = 993$),

scj = Number of single conjugated bonds ($U = 859$);

*SEE* = 0.0725 $r^2$ = 0.9528 $F$ = 60.5698, $r^2_{CV}$ = 0.9052, $rpred^2$ : 0.89914

*SEE* = Standard error of estimation, $r^2$ = Pearson square correlation, $F$ = Fisher function, $r^2_{CV}$ = Pearson cross validated square correlation (Leave one out method), $rpred^2$ = predictive $r^2$.

According to algebraic sign of coefficients in QSAR formula and the value of utility *U* the main factor in influence on activity value is the minimum net charge of heteroatom's and molecular mass weighted moment of inertia *C* and number of single conjugated bonds play dominant role to activity. Molecular fragments $C_2HN_3$ not favorable for the activity and this fragment cotain AZA so this compound has low activity .

## Validation of the computation procedure

For the validation of the method, I have proceeded to a QSAR study with a validation set and reduced calibration set. The validation set was extracted from the homogenized calibration set. In bold letter with star containing compound in Table 1 are selected for validation set (Compd. no **2, 5** and **7**). The selection of the validation set should be such that it captures all the features and characteristics of the whole set of molecules. From the point of view of them considerations discussed here, the most significant is the correlation between the calculated and the experimental values of hCA VA inhibitory activity for the molecules in the validation set. In Table 2 and Fig. 1 there are listed the calculated values and the experimental values of the hCA VA inhibitory power for the molecules in the validation set. I have found good predictive $r^2$.

*Applicability domain* :

I used |*RStudent*| of observed inhibitory activity calculated by the obtained models and hat diagonal (leverage) for assigning applicability of domain (AD). Values for leverage have been calculated for both calibration set and prediction set compounds showing in Table 1. Applicability of domain for the developed model of calibration set is shown in William plot Fig. 3. Influential compounds are points with leverage value higher than the warning

**Table 2.** Observed/estimated values of hCA V inhibitory activity (A) for the molecules in the training and test set

| Training set | | | Test set | | |
|---|---|---|---|---|---|
| Compd. no. | Obs. | Est. | Compd. no. | Obs. | Est. |
| **1** | 3.172 | 3.274 | **2** | 3.39 | 3.351 |
| **3** | 3.606 | 3.566 | **5** | 3.256 | 3.118 |
| **4** | 3.76 | 3.705 | **7** | 3.696 | **3.66** |
| **6** | 3.655 | 3.543 | | | |
| **8** | 3.72 | 3.763 | | | |
| **9** | 4 | 4.075 | | | |
| **10** | 3.684 | 3.592 | | | |
| **11** | 2.824 | 2.819 | | | |
| **12** | 3.322 | 3.405 | | | |

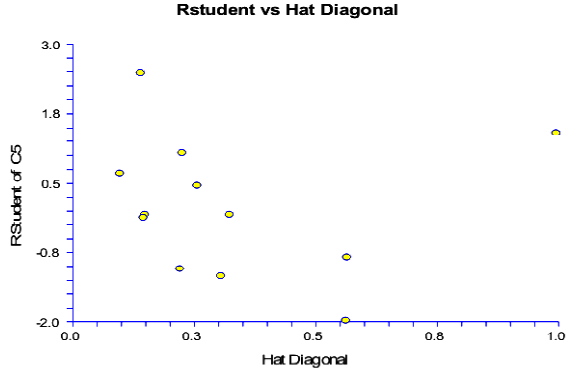



**Fig. 3.** |*RStudent*| of observed vs hat diagonal applicability domain.

leverage limit 1, so outliers are not present. It can be seen in the William plot; all molecules in calibration set lie in the application domain of the developed model. The computed activity of the prediction set compounds is also within limit of hat diagonal.

## Conclusions

In calibration set minimum net charge of heteroatom's and molecular mass weighted moment of inertia *C* and number of single conjugated bonds play dominant role to activity. Molecular fragments $C_2HN_3$ not favorable for the activity. Many molecules in proposed prediction set have much higher computed activity than observed value and hat diagonal limit.

### References


1. C. T. Supuran, *Nat. Rev. Drug Discov.*, 2008, **7**, 168.
2. C. T. Supuran and A. Scozzafava, *Bioorg. Med. Chem.*, 2007, **15**, 4336.
3. C. T. Supuran, A. Scozzafava and J. Conway (eds.), "Carbonic anhydrase – its inhibitors and activators", CRC Press, Boca Raton (FL), USA, 2004, pp. 1-376.
4. C. T. Supuran, A. Scozzafava and A. Casini, *Med. Res. Rev.*, 2003, **23**, 146.
5. J. Y. Winum, M. Rami, A. Scozzafava, J. L. Montero and C. T. Supuran, *Med. Res. Rev.*, 2008, **28**, 445.
6. C. J. Lynch, H. Fox, S. A. Hazen, B. A. Stanley, S. Dodgson and K. F. Lanoue, *Biochem. J.,* 1995, **310**, 197.
7. S. A. Hazen, A. Waheed, W. S. Sly, K. F. LaNoue and C. J. Lynch, *Faseb. J.*, 1996, **10**, 481.
8. O. Güzel, A. Innocenti, A. Scozzafava, A. Salman and C. T. Supuran, *Bioorg. Med. Chem.,* 2009, **17**, 4894.
9. F. Z. Smaine, F. Pacchiano, M. Rami, V. Barraga, D. Vullo, A. Scozzafava, J. Y. Winum and C. T. Supuran, *Bioorg. Med. Chem. Lett.,* 2008, **18**, 6332.
10. J. Y. Winum, L. Toupet, V. Barragan, G. Dewynter and J. L. Montero, *Org. Lett.*, 2001, **3**, 2241.
10. BROOD (version 2.0.0), OpenEye Science Software, 3600 Cerrillos Road, Suite 1107, Santa Fe, USA, 2010.
11. J. J. P. Stewart, MOPAC2012, Stewart Computational Chemistry, Colorado Springs, Co, USA, http:// OpenMOPAC.net 2012.
12. OMEGA (version 2.4.3), OpenEye Science Software, 3600 Cerrillos Road, Suite 1107, Santa Fe, USA, 2010.
13. G. Tresadern, D. Bemporad and T. A. Howe, *J. Mol. Graph Model*, 2009, **27**, 860.
14. T. A. Halgre, *J. Comput. Chem.*, 1999, **20**, 720.
15. J. J. P. Stewart, *J. Mol. Model*, 2007, **113**, 1173.
16. J. J. P. Stewart (2008), MOPAC2009, Stewart Computational Chemistry, Colorado Springs, Co, USA, http:/ /OpenMOPAC.net.
17. L. Tarko and C. T. Supuran, *Bioorg. Med. Chem.*, 2013, **21**, 1404.
18. PRECLAV and DESCRIPT programes are available from Center of Organic Chemistry, Bucharest, Romanian Academy, ltarko@cco.ro; tarko_laszlo@yahoo.com.
19. L. Tarko, *Rev. Chim. (Bucuresti)*, 2004, **55**, 539.
20. L. Tarko, *Arkivoc*, 2004, **xiv**, 74.
19. A. Golbraikh and A. Tropsha, *J. Mol. Graph Model*, 2002, **20**, 269.
20. D. W. Osten, *J. Chemom.*, 1998, **2**, 39.
21. Organization for Economic Co-operation and Development Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship QSAR Models, 2007, OECD Document ENV/JM/MONO (2007) 2.
22. A. P. Worth, T. Aldenberg, I. Benjamin and M. T. D. Cronin, *ATLA*, 2005, **33**, 155.
23. A. Tropsha, P. Gramatica and V. K. Gombar, *Comb. Sci.*, 2003, **22**, 69.
24. S. Weaver and M. P. Gleeson, *J. Mol. Graph Model*, 2008, **26**, 1315.
25. NCSS (Statistical Software Delux package), 329 North 1000 East, Kaysville, UT, USA, 2004.
26. Dennis R. Cook, "Residuals and Influence in Regression", Chapman and Hall, New York, 1982.