Journal of Molecular Graphics and Modelling 31 (2011) 10–19

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases

Javier García[a], Pablo R. Duchowicz[a,*], María F. Rozas[a], José A. Caram[a,b], María V. Mirífico[a,c], Francisco M. Fernández[a], Eduardo A. Castro[a]

[a] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA, CCT La Plata-CONICET), Casilla de Correo 16, Sucursal 4, 1900 La Plata, Argentina
[b] Facultad de Ciencias Exactas, Departamento de Química, Universidad Nacional de La Plata, Calle 1 y 47, 1900 La Plata, Argentina
[c] Facultad de Ingeniería, Área Departamental Ingeniería Química, Universidad Nacional de La Plata, Calle 1 y 47, 1900 La Plata, Argentina

ARTICLE INFO

Article history:
Received 7 June 2011
Received in revised form 28 July 2011
Accepted 31 July 2011
Available online 19 August 2011

Keywords:
QSAR theory
1,2,5-Thiadiazolidin-3-one 1,1-dioxide
Serine proteases
Molecular Dynamics
Flexible descriptors

ABSTRACT

Selective inhibitors of target serine proteinases have a potential therapeutic role for the treatment of various inflammatory and related diseases. We develop a comparative quantitative structure–activity relationships based analysis on compounds embodying the 1,2,5-thiadiazolidin-3-one 1,1-dioxide scaffold. By means of classical Molecular Dynamics we obtain the conformation of each lowest-energy molecular structure from which we derive more than a thousand of structural descriptors necessary for building predictive QSAR models. We resort to two different modeling approaches with the purpose of testing the consistency of our results: (a) multivariable linear regressions based on the replacement method and forward stepwise regression, and (b) the calculation of flexible descriptors with the CORAL program. All the models are properly validated by means of standard procedures. The resulting QSAR models are supposed to be of great utility for the rational search and design (including synthesis and/or *in vitro* biochemical studies) of new effective non-peptidyl inhibitors of serine proteinases.

1. Introduction

Various proteolytic enzymes, including the human leukocytes elastase (HLE), cathepsin G (Cat G), and proteinase 3 (PR 3) [1] are (chymo)trypsin-like proteases with implications in the etiology and/or pathophysiology of a range of inflammatory diseases, including pulmonary emphysema [2], chronic bronchitis [3], adult respiratory distress syndrome [4], etc. The existence of a protease–antiprotease imbalance is generally associated to depressed levels of physiological protein inhibitors. It is for this reason that there has been so much interest in developing highly selective and potent irreversible inhibitors of serine proteases [1].

The 1,2,5-thiadiazolidin-3-one 1,1-dioxide structural scaffold (Fig. 1) has been recognized as a key structural constituent due to its high versatility for appending peptidyl or non-peptidyl recognition elements, which in turn favors the optimization of multiple binding interactions to several enzyme subsites that allow suppressing their activities. Other examples of inactivators of this kind include haloenol and ynenol lactones [5,6], substituted isocoumarins [7], 3-alkyl-*N*-hydroxysuccinimide derivatives [8–10], substituted dihydrouracils [11], β-lactams [12], and saccharin derivatives [13–15].

Among the main drawbacks of resorting to orally administered peptide and protein drugs appears the underlying compromise between efficiency and poor absorption, low metabolic stability and rapid excretion. Furthermore, it has been observed that HLE and Cat G resist the inhibition by proteins, although they are inhibited by low molecular weight compounds [16]. The design of effective non-peptidyl inhibitors of proteases has been commonly achieved by searching a molecule that mimics the backbone conformation of a protein inhibitor. It should also be capable of orienting recognition elements appended to it in the same vector relationship as the amino acid side chains of the protein inhibitor, thus making it possible the exploitation of favorable substrate–enzyme binding interactions.

The well known theory of quantitative structure–activity relationships (QSAR) [17–19] is based on the hypothesis that the biological activity of a chemical compound is mainly determined by its molecular structure [17]. It does not offer specific details on the usually complex mechanism/path of the process. However, it is possible to get some insight into the underlying mechanism by means of the QSAR-based predicted activities. As far as we are aware none of the previous *in vitro* biochemical studies was complemented by the application of QSAR Theory to model the structure–activity relationships (SAR) exhibited by 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as inhibitors of serine proteinases. Quite on the contrary, different molecular modeling docking studies have analyzed the energy-minimized enzyme-inhibitor complexes by means of

* Corresponding author. Tel.: +54 221 425 7430/7291; fax: +54 221 425 4642.
E-mail address: pabloducho@gmail.com (P.R. Duchowicz).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.07.007

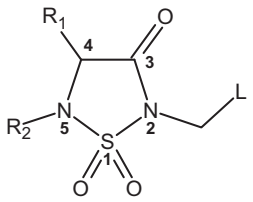

Fig. 1. The 1,2,5-thiadiazolidin-3-one 1,1-dioxide structural scaffold.

the Tripos Force Field of SYBYL software (Tripos Associates, St. Louis, MO) and the available experimental information on the crystal structures of the enzymes bound to the inhibitors [1,20–23].

In this work we carry out a QSAR analysis on several compounds with the 1,2,5-thiadiazolidin-3-one 1,1-dioxide scaffold that could serve as a rational guide for the design of potent and selective therapeutic agents. It is our purpose to develop useful QSAR models for predicting active and inactive molecular structures, which allows us to describe the biochemical properties of thiadiazolidin-3-one 1,1-dioxides.

2. Materials and methods

2.1. Experimental data set

The *in vitro* activities of 1,2,5-thiadiazolidin-3-one 1,1 dioxide compounds against the panel of serine proteases HLE, Cat G and PR 3 are extracted from available biochemical studies [1,16,20–25]; they are displayed in supplementary Table 1S. Those inhibitory potencies, which are expressed in terms of the apparent second-order inactivation rate constant $k^*_{inact}[M^{-1}\,s^{-1}]$ measured with the progress curve method [26], are then converted into logarithm form $\log_{10} k^*_{inact}$ for modeling purposes. All the heterocyclic compounds under analysis exhibit the particularity of being readily synthesized using aminoacid precursors [1,16,20–25].

The molecular set includes the 1,2,5-thiadiazolidin-3-one 1,1-dioxide scaffold with different substituents, such as sulfones, sulfides, sulfonamides, phosphates, carboxylates, etc., on 2 and 5N-atoms, and 4C-atom on the heterocycle (refer to Fig. 1).

2.2. Geometry optimization and calculation of molecular descriptors

We keep the S-configuration for the sp^3 carbon atom in the 1,2,5-thiadiazolidin-3-one 1,1-dioxide for all the molecular structures (including racemic mixtures), except when the chirality of the sp^3 C-atom of the heterocycle changes in Table 1S, in which case we choose the R-configuration. The initial conformations of the compounds are drawn with the aid of the "Model Build" modulus of the HyperChem 6.03 program for Windows [27].

The conformational space of the molecules is scanned by means of the Molecular Dynamics module of the HyperChem. The MM^+ Molecular Mechanics Force Field available in that package is used for the simulations. The starting geometries are heated from 0 to 900 K in 0.1 ps. After that, the temperature is kept constant by coupling the system to a simulated bath with a relaxation time of 0.5 ps. After an equilibration period of about 5 ps, a 500 ps simulation is carried out saving the coordinates every 10 ps. The simulation time step is 1 fs. The saved geometries are then minimized to an energy gradient smaller than 0.01 kcal mol^{-1} $Å^{-1}$, using the Semiempirical Method PM3 from the Molecular Orbitals Theory with the Polak–Ribiere algorithm. The lowest-energy conformers found in such simulations are employed as models for the 3D-structure.

We then compute 1497 molecular descriptors using the Dragon program [28], including descriptors of all types such as Constitutional, Topological, Geometrical, Charge, GETAWAY (Geometry, Topology and Atoms-Weighted AssemblY), WHIM (Weighted Holistic Invariant Molecular descriptors), 3D-MoRSE (3D-Molecular Representation of Structure based on Electron diffraction), Molecular Walk Counts, BCUT descriptors, 2D-Autocorrelations, Aromaticity Indices, Randic Molecular Profiles, Radial Distribution Functions, Functional Groups, Atom-Centred Fragments, Empirical and Properties [29].

We also calculate atomic charge density-based descriptors by means of the Recon 5.5 software [30], encoding electronic and structural information relevant to the chemistry of intermolecular interactions. This sort of computed descriptors is not provided by Dragon software, and the robustness of Recon has been demonstrated elsewhere [31,32]. Recon is an algorithm for the reconstruction of molecular charge densities, and charge density-based electronic properties of molecules, using atomic charge density fragments precomputed from *ab initio* wavefunctions. The method is based on the quantum theory of atoms in molecules [33]. A library of atomic charge density fragments has been built in a form that allows for the rapid retrieval of the fragments and molecular assembly. In the present case, the smiles chemical notation is employed as input for the generation of 248 transferable atom equivalent (TAE) descriptors, developed by Breneman and Weber [34]. In this way, the total number of calculated structural descriptors amounts to 1745 variables.

2.3. Model development

In order to verify the consistency of our results we compare the QSAR models obtained by means of two different approaches: (a) the search for the best molecular descriptors via multivariable linear regressions based on the replacement method (RM) and on forward stepwise regression (FSR); (b) the calculation of flexible descriptors with the CORAL (CORrelation And Logic) program. All the routines necessary for present calculations were written in the language of technical computing Matlab 7.0 [35]. In every QSAR model displayed in this paper N denotes the number of training set molecules, *range* is the experimental range of activities covered by the model, d is the number of descriptors of the model, R^2 is the squared correlation coefficient, S is the standard deviation of the model when applied on the training set, F is the Fisher parameter, res is the residual for a given molecule (difference between the experimental and predicted activity), outliers > $x.S$ indicates the number of molecules with a predicted *res* greater than x times S, $Corr^{max}$ represents the maximum squared correlation coefficient between two given descriptors of the model, VIF is the variance inflation factor, loo subscript belongs to the leave-one-out cross validation result, and Rand superscript stands for Y-Randomization.

2.3.1. Linear descriptors search

2.3.1.1. Replacement method. In recent years theoretical and experimental researchers have focused an increasing attention on finding the most efficient tools for selecting molecular descriptors in QSAR studies. Therefore, there are many methods for the selection of the best structural descriptors from a large pool of them. One of such approaches is the replacement method (RM) [36,37] that has already proved successful in earlier studies [38–42]. In brief, the RM is an efficient optimization tool that generates multi-parametric linear regression QSAR models on a training (calibration) molecular set by searching the set **D** of D descriptors for an optimal subset **d** of $d \ll D$ ones with minimum model's standard deviation (S). The quality of the RM results is satisfactorily close to the one obtained from an exact (combinatorial) full search of molecular descriptors, although with a much smaller CPU time. Our RM results take into account the variance inflation factor (VIF_{ij}), a method for detecting the severity of multicollinearity or high degree of correlation

Table 1
Best QSAR found with the replacement method on HLE dataset.

d	S	R^2	S_{loo}	R^2_{loo}	S_{test}	R^2_{test}	$Corr^{max}$	Molecular descriptors
1	1.20	0.460	1.23	0.430	0.95	0.686	–	*C*-003
2	1.06	0.587	1.10	0.555	1.01	0.643	0.097	*BIC*3, *C*-003
3	0.87	0.721	0.92	0.689	0.98	0.669	0.166	*IC*3, *Mor*15*p*, *C*-003
4	0.80	0.766	0.87	0.728	0.95	0.697	0.166	*BIC*3, *Mor*15*p*, *C*-003, *MLOGP*
5	0.74	0.801	0.81	0.764	0.87	0.757	0.166	*BIC*3, *Mor*15*p*, *HATS*8*e*, *C*-003, *MLOGP*
6	0.71	0.821	0.79	0.781	0.83	0.790	0.526	*IC*2, *piPC*05, *BEHp*8, *RDF*055*e*, *C*-003, *MLOGP*
7	0.70	0.830	0.79	0.783	0.83	0.797	0.178	*SIC*3, *PCD*, *BEHp*1, *Mor*13*u*, $R5^{+}_{v}$, *C*-003, *MLOGP*

(linear dependency) among two or more supposedly independent variables [43,44]. VIF_{ij} for a given descriptor i can be easily computed if we know the squared correlation coefficient R^2_{ij} between this descriptor and the remaining j ones of the model:

$$VIF_{ij} = \frac{1}{1 - R^2_{ij}} \quad (1)$$

In practice, a value $VIF_{ij} > 10$ indicates that there may be significant multicollinearity among the chosen subset of descriptors.

2.3.1.2. Forward stepwise regression. The forward stepwise regression (FSR) [45] consists of a step-by-step addition of the best descriptors to the linear model so that they lead to the smallest S in the training set, until there is no-other variable outside the equation that satisfies the selection criterion. The FSR requires fewer linear regressions than RM.

2.3.2. The CORAL method

CORAL version 1.4 [46] is a freeware for Windows. Each molecular structure must be represented by SMILES (Simplified Molecular Input Line Entry System) notation, calculated with ACD/ChemSketch software [47]. CORAL is based on the presence of certain SMILES attributes occurring in the molecule which can be associated to its activity [48–51]. Symbols representing chemical elements, cycles, branching of molecular skeleton, charges, etc., are used as SMILES attributes. The CORAL modeling process not only considers the presence of individual elements SMILES attributes (s_k), but also clusters of two (ss_k) and three (sss_k) of them. For example, SMILES = Clc1ccccc1 then s_k = (Cl, c, 1, c, c, c, c, c, 1); ss_k = (Clc, c1, cc, cc, cc, cc, cc, c1); sss_k = (Clc1, c1c, ccc, ccc, ccc, ccc, cc1).

The model is a one-variable correlation between the activity values and the flexible descriptor (DCW) defined as:

$$DCW(\text{threshold}) = \alpha \sum_k CW(s_k) + \beta \sum_k CW(ss_k) + \gamma \sum_k CW(sss_k) \quad (2)$$

where α, β, γ are 1 or 0, and CW is the correlation weight for the element/s of the SMILES. The threshold is the parameter used to define rare (noise) SMILES attributes. The rare SMILES attributes may lead to overtraining: excellent correlation for the training set accompanied by poor correlation for the validation set. The threshold is an integer j with the meaning that all SMILES attributes that take place in less than j SMILES notations of the training set are classified as rare. In present study, numerical data for CW are calculated by Monte Carlo simulation maximizing the correlation coefficient between the activity values and the DCW descriptor defined in Eq. (2) for the training set. The quality of the predictions depends on the selected options/parameters in the algorithm, such as the number of epochs used during the Monte Carlo optimization, D_{start}, $d_{precision}$, dR_{weight}, dC_{weight}, threshold range and others, which should be correctly specified in order to calculate the DCW values. More specific details of the CORAL algorithm are available in the recent literature [48–51].

2.3.3. Analysis of the happenstance of the model

Another simple way of proving that the structure–activity relationships derived in this study do not result from happenstance comes from checking their robustness by means of the so-called Y-randomization [52]. This technique consists of scrambling the experimental values of the property in such a way that they do not longer correspond to the respective compounds. The smallest standard deviation S^{Rand} obtained after analyzing 1000 cases of Y-randomization for each developed QSAR turned out to be poorer (greater) than the one found in the true calibration (S). This result supports the assumption that the correlations derived here are not fortuitous but the result of actual structure–activity relationships.

2.3.4. Model validation

In addition to provide a satisfactory correlation for the training set, each QSAR should be properly validated in order to test its predictive performance. For example, we can carry out the test known as Leave-One-Out Cross Validation (loo) [53]. Statistical parameters R^2_{loo} and S_{loo} measure the stability of the developed QSAR upon inclusion/exclusion of compounds selected randomly and, according to the specialized literature, R^2_{loo} should be greater than 0.7 for obtaining a validated model [54].

We also apply a more realistic validation that consists of omitting from the complete molecular set (Table 1S) some compounds which constitute the 'test set' (denoted as 'test'). By performing such a splitting one estimates whether the QSAR found have any capability to estimate the activities of the compounds in the "fresh" test set that have never been used in the construction of the model. We randomly choose the molecules for the training and test sets before starting the search for the optimal model.

2.3.5. Degree of contribution of selected descriptors

In order to determine the relative importance of each descriptor in the linear regression model, we calculate standardized regression coefficients:

$$b^s_j = \frac{s_j\, b_j}{s_Y} \quad j = 1, \ldots, d \quad (3)$$

where b_j is the regression coefficient for the descriptor j, and s_j and s_Y are the standard deviations for that descriptor and for the experimental activity, respectively. The larger the value of b^s_j the greater the importance of the descriptor [45].

3. Results and discussion

We apply FSR and RM to the three datasets (HLE, PR 3 and Cat G) and obtain the best linear regressions with 1–7 variables extracted from the set of D = 1745 descriptors. As discussed in earlier papers [36,37,55] the RM provides various final solutions with minimum S for the training set, from which one has to select the model with the best predictive value.

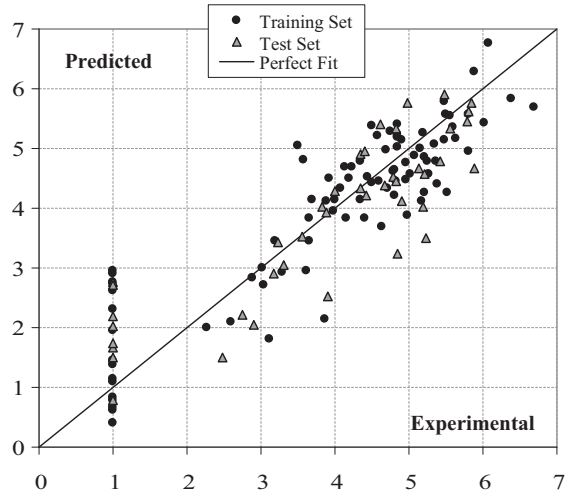

Fig. 2. Predicted $\log_{10}\ k^*_{\text{inact}}$ for HLE according to Eq. (4) as function of experimental values.

3.1. QSAR on HLE data

The best RM models for HLE appear highlighted in supplementary Table 2S and are copied into Table 1. We appreciate that in this case, as well as for the other datasets, the various RM models with different dimensions exhibit roughly the same value of S. From them we select the following five-descriptors model:

$$\log_{10}\ k^*_{\text{inact}} = -11.496(\pm 3) + 20.301(\pm 3)\cdot BIC3 - 2.280(\pm 0.3)\cdot Mor15p - 4.611(\pm 1)\cdot HATS8e + 1.274(\pm 0.2)\cdot C\text{-}003 + 0.501(\pm 0.09)\cdot MLOGP \quad (4)$$

$N = 90$, range = 1.000–6.693, $d = 5$, $N/d = 18$, $R^2 = 0.801$, $S = 0.74$, $F = 67.9$, outliers > 3.S = 0, $\text{Corr}^{\text{max}} = 0.167$, $R^2_{\text{loo}} = 0.764$, $S_{\text{loo}} = 0.81$, $R^2_{\text{l-20\%-o}} = 0.682$, $S_{\text{l-20\%-o}} = 0.95$, $S^{\text{Rand}} = 140$, $N_{\text{test}} = 42$, $R^2_{\text{test}} = 0.757$, $S_{\text{test}} = 0.87$.

It represents a compromise between the statistical performance achieved on both the training and test sets, given by the values of S and S_{test}, respectively. It is worth noting that the S_{test} is the value of the standard deviation resulting from the application of the chosen model to the test set with compounds that were not used in the construction of the model. We follow the common practice of keeping a relatively small number of descriptors in the model in order to avoid any possible fortuitous correlation. At the same time we want the model with the best predictive value which is expected to be the one with the smallest value S_{test}. The application of all these considerations to Table 1 leads to the $d = 5$ model shown in Eq. (4). Further proof of the predictive value of the chosen model is provided by contrasting the predicted and experimental values of the activities as shown in Fig. 2, the dispersion plot of residuals (i.e. residuals as function of predicted activities) shown in Fig. 1S, and the absence of interrelationships between the descriptors of Eq. (4) (see the correlation matrix, Table 3S).

Table 2 shows the activities predicted by each QSAR as well as the experimental values for comparison. We appreciate that Eq. (4) predicts the experimental activities of the compounds in the training and test sets reasonably well (test data marked with $\hat{}$). It is worth mentioning that the model predicts low activities for the inactive compounds in both sets. Compounds for which activities have not yet been measured do not exhibit favorable predicted inhibitory potencies, which is in line with previous SAR observations [1,16,20–25]. All these facts strongly suggest that the models derived in this work may be useful, predictive and properly validated.

The descriptors in Eq. (4) embody multidimensional features of the molecular structure, where the 3D conformation-dependent descriptors were obtained by Molecular Dynamics. The parameters can be classified as follows:

- Topologicals (2D): *BIC*3, bond information content (neighborhood symmetry of 3-order), obtained from elements of Graph Theory.
- 3D-MoRSE (3D): *Mor*15*p*, signal 15/weighted by atomic polarizabilities, obtained from the 3D-molecule representation of structure based on electron diffraction.
- GETAWAY (3D): *HATS*8*e*, leverage-weighted autocorrelation of lag 8/weighted by atomic Sanderson electronegativities, derived from the molecular influence matrix.
- Atom-Centred Fragments (1D): *C*-003, number of CHR_3 groups, where R represents any group linked through carbon.
- Property (1D): *MLOGP*, Moriguchi octanol–water partition coefficient, which is a measure of the lipophilic character of the compounds.

Specific details of such theory-based Dragon descriptors are well-known in the literature [29]. Application of Eq. (3) leads to the following order of contributions to the inhibition of HLE:

$$Mor15p\,(0.45) > BIC3\,(0.39) > C\text{-}003\,(0.35) > MLOGP\,(0.32) > HATS8e\,(0.22) \quad (5)$$

The relative magnitudes of the coefficients b^s_j (shown between parentheses) suggest that the numerical variables complement each other and that the selected structural attributes are similarly relevant for predicting the biological activity. The application of FSR leads to the models in Table 4S. It is clear that none of them yields a better result than the one provided by RM for a given d on the test set.

In what follows we discuss an optimal flexible descriptor calculated with the CORAL program. Upon inserting into Eq. (2) the correlation weights produced by a Monte Carlo simulation, we obtain the following QSAR:

$$\log_{10}\ k^*_{\text{inact}} = -4.643(\pm 0.3) + 0.218(\pm 0.007)\cdot \text{DCW}_1(0) \quad (6)$$

$N = 90$, range = 1.000–6.693, $d = 5$, $N/d = 18$, $R^2 = 0.801$, $S = 0.74$, $F = 67.9$, outliers > 3.S = 0, $\text{Corr}^{\text{max}} = 0.167$, $R^2_{\text{loo}} = 0.764$, $S_{\text{loo}} = 0.81$, $R^2_{\text{l-20\%-o}} = 0.682$, $S_{\text{l-20\%-o}} = 0.95$, $S^{\text{Rand}} = 140$, $N_{\text{test}} = 42$, $R^2_{\text{test}} = 0.757$, and $S_{\text{test}} = 0.87$; $N = 90$, range = 1.000–6.693, $d = 1$, $N/d = 90$, $R^2 = 0.918$, $S = 0.47$, $F = 979.9$, outliers > 3.S = 1, $R^2_{\text{loo}} = 0.914$, $S_{\text{loo}} = 0.48$, $R^2_{\text{l-20\%-o}} = 0.884$, $S_{\text{l-20\%-o}} = 0.51$, $S^{\text{Rand}} = 148$, $N_{\text{test}} = 42$, $R^2_{\text{test}} = 0.790$, and $S_{\text{test}} = 0.78$.

The numerical parameters used in the CORAL calculation of $\text{DCW}_1(0)$ are: number of epochs: 8, number of probes: 3, threshold value adopted for the training and test set: 0, $D_{\text{start}} = 0.5$, $d_{\text{precision}} = 0.01$, $dR_{\text{weight}} = 0$, $dC_{\text{weight}} = 0$, and $\alpha = \beta = 0$ (refer to Eq. (2)). The QSAR given by Eq. (6) exhibits a slightly better prediction value than Eq. (4). Although it is based on a quite different modeling methodology, it also predicts low activities for those compounds with yet unknown experimental values. The predictions of Eq. (6) are plotted in Fig. 3.

3.2. QSAR on PR 3 data

From the RM results in Table 5S we derive the best models with 1–7 descriptors shown in Table 3 and then the following optimal QSAR:

$$\log_{10}\ k^*_{\text{inact}} = 1.638(\pm 0.9) + 0.0448(\pm 0.01)\cdot G(N..Cl) + 0.0357(\pm 0.009)\cdot Mor02u - 15.841(\pm 3)\cdot HATS8p + 92.968(\pm 28)\cdot R5^+_v + 1.242(\pm 0.1)\cdot C\text{-}003a \quad (7)$$

Table 2
The best models obtained with the RM technique on HLE, PR 3 and Cat G datasets.

No.	HLE			PR 3			Cat G		
	Exp.	Eq. (4)	Eq. (6)	Exp.	Eq. (7)	Eq. (9)	Exp.	Eq. (10)	Eq. (12)
1	3.915	4.497	3.795	2.633	2.938	3.015	Inactive[a]	0.730	1.659
2	4.831	4.443	4.115	3.107	3.783	3.123	2.898	2.173	3.171
3	3.892	4.118	3.791	2.924	3.357	2.687	1.845	2.197	2.344
4	4.803	4.225	5.006	3.850	3.582	3.960	1.903^[b]	1.617	2.882
5	4.588	4.441	4.683	4.013^	3.320	3.524	2.544	2.688	2.056
6	4.412	3.831	4.445	3.387	3.096	3.450	1.778	1.590	1.743
7	3.886^	3.924	3.847	3.021	2.170	2.991	2.176^	1.747	1.703
8	4.427^	4.216	4.254	3.248	3.665	3.826	2.462	2.460	2.185
9	3.305^	3.058	3.555	3.262^	2.951	3.568	1.477	1.149	1.025
10	3.233^	3.435	4.597	–[c]	3.189	3.898	Inactive^	1.791	0.962
11	3.021	3.008	2.575	2.491	2.682	2.538	Inactive	1.087	1.719
12	3.662	3.444	2.832	2.568^	3.691	3.058	Inactive	1.240	1.156
13	4.508	4.438	4.737	3.798	4.212	4.187	Inactive	1.376	1.228
14	5.340	5.074	4.914	4.210	3.816	3.857	2.255	2.185	1.722
15	3.977	3.954	3.795	3.352	3.764	3.398	Inactive^	1.736	0.861
16	4.979	3.885	3.972	3.719	2.905	3.067	2.041	1.985	1.355
17	3.819^	4.034	3.692	–	2.876	2.927	Inactive	2.072	1.112
18	4.348	4.150	4.045	3.981	3.842	3.500	2.000	1.743	1.777
19	4.677^	4.373	4.987	4.228	4.313	4.290	2.204^	0.975	2.143
20	5.217	4.257	5.156	4.307	3.852	4.251	1.845^	2.211	2.976
21	4.188	4.509	4.715	3.904^	3.886	4.071	1.301	1.125	1.219
22	4.000^	4.283	4.883	3.512^	3.807	4.032	1.778^	2.374	2.052
23	4.582	5.207	4.437	3.679	3.865	3.688	Inactive^	0.582	1.254
24	5.380	4.396	5.540	3.352^	2.861	3.864	Inactive	0.539	2.183
25	Inactive	0.661	0.684	–	1.788	0.599	1.477	2.514	2.187
26	Inactive	1.454	1.212	–	1.433	0.699	2.079	2.191	2.080
27	Inactive	0.400	0.828	Inactive	1.388	0.953	4.049	3.633	3.132
28	Inactive^	0.795	0.828	–	1.423	0.953	3.199	3.292	3.132
29	Inactive	2.302	1.212	–	1.249	0.316	2.505^	2.433	2.877
30	Inactive^	1.660	1.233	–	2.018	0.669	2.881^	3.453	2.904
31	Inactive	1.381	0.973	–	2.089	0.869	3.053	2.653	2.243
32	Inactive^	1.511	0.973	–	1.150	0.869	2.447^	2.361	2.243
33	Inactive	1.139	1.109	Inactive^	1.421	1.711	3.575	4.028	3.963
34	2.892	2.834	2.955	3.262	2.257	2.263	Inactive	1.081	1.248
35	3.861	2.137	3.133	3.695^	1.481	1.932	2.114	1.704	1.741
36	Inactive^	2.202	0.970	Inactive	1.851	1.127	Inactive	1.203	1.289
37	Inactive	2.762	1.148	Inactive	1.857	0.797	Inactive^	2.090	1.782
38	Inactive^	2.028	1.968	Inactive	1.415	1.174	Inactive	1.985	1.979
39	2.279	2.002	2.901	2.301^	1.801	2.545	Inactive	0.662	1.070
40	2.908^	2.045	3.078	1.903	1.183	2.214	Inactive	1.397	1.564
41	Inactive	2.956	1.828	Inactive^	2.954	1.645	Inactive^	0.472	0.580
42	Inactive	2.731	2.005	Inactive	2.000	1.314	Inactive	1.332	1.074
43	3.033	2.706	3.010	Inactive^	2.178	1.981	Inactive	1.461	1.425
44	3.906^	2.514	3.187	Inactive	1.476	1.651	2.785	1.804	1.919
45	–	2.625	2.280	–	1.424	1.052	4.025	4.012	3.896
46	–	2.675	3.117	–	1.501	2.379	4.356^	3.973	4.068
47	–	1.175	2.624	–	1.508	1.113	4.107^	3.522	4.444
48	–	2.573	3.117	–	1.188	0.994	4.103	3.241	4.340
49	–	1.594	2.534	–	1.548	1.283	4.091^	3.754	4.104
50	–	0.559	1.410	–	0.501	0.663	3.929^	3.973	3.368
51	–	0.757	2.247	–	0.724	1.991	4.402^	3.212	3.541
52	–	2.362	2.302	–	1.525	1.709	4.240	4.206	3.718
53	–	0.438	−0.329	–	1.275	0.573	2.544	3.330	3.160
54	–	1.653	0.646	–	1.635	0.573	4.316	3.452	4.122
55	–	0.528	0.646	–	1.394	0.573	4.824	3.769	4.122
56	–	3.011	2.481	–	1.759	1.305	2.633^	2.818	2.552
57	–	1.654	0.713	–	1.301	1.032	4.234	3.909	4.041
58	–	2.707	−0.506	–	1.313	0.314	4.197	3.787	4.194
59	–	0.458	1.080	–	1.222	0.988	2.690	3.300	2.868
60	5.186	5.260	5.424	4.214^	3.977	3.739	–	1.388	0.645
61	4.831^	4.446	4.023	3.179	3.148	3.196	–	2.175	1.778
62	4.828^	5.337	2.976	3.777	3.490	3.771	–	2.170	0.532
63	4.349^	4.896	3.592	3.029^	2.496	2.178	–	2.594	2.618
64	4.906	5.134	4.889	4.015	3.851	3.999	1.477^	1.650	0.476
65	5.242	4.791	5.241	3.960	2.968	3.673	1.778	2.170	1.827
66	3.188	3.454	3.176	3.193	3.175	3.206	–	2.468	1.587
67	2.748^	2.214	2.425	2.531	2.698	2.478	–	3.304	1.850
68	4.403^	4.954	4.405	4.101^	3.640	4.094	–	1.545	0.488
69	5.226^	3.497	3.120	3.555^	3.463	3.376	–	1.858	1.749
70	4.355	4.786	4.192	4.045^	3.869	3.949	–	2.387	2.130
71	Inactive	2.900	1.397	3.037	2.708	3.008	1.699^	2.996	2.811
72	Inactive^	1.750	0.360	Inactive	1.440	1.008	2.690	2.414	2.474
73	Inactive	0.611	1.080	Inactive	1.456	0.988	4.242^	3.694	2.868
74	Inactive^	2.759	2.481	Inactive	1.744	1.305	2.633	2.630	2.552

Table 2 (*Continued*)

No.	HLE			PR 3			Cat G		
	Exp.	Eq. (4)	Eq. (6)	Exp.	Eq. (7)	Eq. (9)	Exp.	Eq. (10)	Eq. (12)
75	Inactive	1.105	0.713	Inactive	1.074	1.032	4.234^	3.634	4.041
76	Inactive^	2.713	0.018	Inactive^	1.475	0.715	4.197	4.419	4.473
77	4.797	4.608	4.588	3.230^	4.455	3.772	2.176	1.755	2.004
78	3.699	4.143	3.407	–	4.210	3.233	1.845	1.576	1.743
79	5.068	4.870	5.116	4.447^	5.043	3.872	1.778	1.385	1.896
80	4.228	4.693	3.935	–	4.595	3.333	1.301^	1.085	1.635
81	4.846	5.180	4.913	3.255^	4.034	4.254	Inactive	1.355	1.497
82	4.695	4.977	4.701	–	3.378	4.633	2.000^	2.071	1.889
83	5.554	5.554	5.442	4.810^	4.810	4.354	Inactive^	1.566	1.389
84	5.788^	5.454	5.442	–	4.741	4.354	1.602	1.875	1.389
85	5.815	4.947	5.610	4.076	4.340	4.315	2.602	2.432	2.222
86	5.877	6.277	5.610	–	4.110	4.315	1.778	1.987	2.222
87	5.805^	5.618	5.394	5.101	4.540	4.635	2.146	1.955	2.132
88	6.024	5.423	5.394	5.197^	4.641	4.635	1.699	1.823	2.132
89	4.978^	5.769	5.229	–	4.279	4.733	1.699	1.985	1.781
90	5.273	4.575	5.229	4.130	4.401	4.733	1.602	1.496	1.781
91	5.483^	5.911	5.280	–	4.127	4.765	Inactive	1.161	0.712
92	4.844	5.402	5.067	–	4.107	5.144	Inactive	1.354	1.104
93	5.197^	4.031	4.643	–	4.487	3.764	1.778^	1.693	1.510
94	Inactive	0.828	1.216	2.519^	2.511	2.658	–	4.455	3.640
95	Inactive	0.788	1.004	3.365	1.944	3.037	–	3.382	4.033
96	Inactive	1.960	1.216	2.380	2.356	2.658	–	4.383	3.640
97	4.076	4.335	4.962	3.079^	4.852	4.795	–	0.294	0.144
98	4.607^	5.412	4.177	3.531	3.414	3.378	–	0.831	0.708
99	4.342^	4.340	3.785	3.477^	3.156	3.919	–	1.611	0.741
100	4.708	4.327	4.324	3.785^	3.671	3.262	–	0.306	0.082
101	4.848	5.250	5.139	3.806	4.209	3.717	–	0.613	0.413
102	4.851^	3.237	4.598	4.204^	2.677	3.889	–	2.330	1.024
103	4.452	4.527	3.937	4.004	3.261	3.815	2.301	1.930	2.124
104	3.996	4.137	3.847	3.505	3.495	3.493	–	1.152	0.277
105	5.148	4.998	4.993	4.439	4.775	3.897	–	0.747	0.351
106	5.173	4.122	5.237	3.380	3.360	3.403	–	0.941	0.569
107	5.361	4.797	5.034	4.438	4.023	4.324	1.778	1.215	1.449
108	5.127^	4.671	5.936	4.479	3.974	4.791	1.845	2.431	1.604
109	4.967	4.769	5.930	4.465	3.899	4.826	Inactive	0.919	1.429
110	4.149	3.841	4.015	3.716^	3.556	4.747	1.845	1.789	1.578
111	–	1.810	1.357	Inactive	1.379	0.817	1.954	2.139	2.353
112	–	3.091	2.173	Inactive	1.800	1.273	1.903	3.089	2.684
113	3.294	2.939	2.755	Inactive	1.442	0.983	3.732	3.203	2.920
114	2.477^	1.496	2.662	2.954	2.077	3.063	Inactive^	1.555	1.267
115	2.602	2.093	3.191	2.477	2.162	3.163	Inactive	1.074	1.159
116	3.176^	2.902	3.434	2.602	2.169	2.670	Inactive	1.142	1.378
117	3.568	4.806	4.452	3.531	3.012	3.773	Inactive	0.675	1.703
118	3.114	1.818	2.717	3.708	2.843	2.781	Inactive^	1.752	1.445
119	3.653	3.831	3.770	3.301	3.176	3.146	2.301	2.333	2.385
120	Inactive	2.610	1.590	Inactive^	3.152	2.163	Inactive	0.659	0.777
121	3.613	2.962	3.556	3.380	4.100	3.916	1.845^	2.047	1.058
122	4.630	3.698	4.085	3.771	4.590	4.016	Inactive	1.666	0.950
123	4.782^	4.512	4.328	3.255	3.473	3.522	2.301^	2.324	1.169
124	4.137	4.693	4.471	–	4.419	3.578	2.477	2.659	2.985
125	5.213^	4.570	3.946	–	2.822	2.953	–	2.587	1.902
126	4.962	4.477	4.137	3.415	3.729	3.686	3.041	2.753	1.497
127	5.427^	4.788	4.665	4.107	4.306	3.786	2.477	2.345	1.389
128	5.526	4.258	4.842	4.415	3.743	3.456	1.954	2.291	1.882
129	5.852^	5.759	5.346	4.945	4.680	4.626	2.699^	1.535	1.494
130	6.693	5.698	5.874	4.524	4.917	4.726	1.778	1.396	1.386
131	6.377	5.835	6.124	4.158	4.786	4.293	1.477	1.684	2.083
132	6.086	6.773	6.406	5.293	5.371	5.051	3.362	2.668	2.914
133	4.496	5.378	4.805	3.978	3.965	3.980	3.079^	2.445	1.573
134	5.503	5.563	5.333	3.987^	4.512	4.080	2.000	2.196	1.465
135	4.751	5.282	5.035	–	4.045	3.740	–	2.575	1.598
136	5.798	5.573	5.563	–	4.470	3.840	2.000^	1.959	1.490
137	5.472	5.779	5.531	–	4.529	4.742	–	1.838	1.569
138	3.505	5.041	5.699	–	3.160	4.703	–	2.269	2.402
139	5.553^	5.330	4.915	–	3.835	3.888	2.301	2.426	2.304
140	4.805	4.639	5.084	–	3.298	3.850	–	2.282	3.137
141	5.022	4.571	5.023	–	4.116	3.686	1.954^	2.286	1.327
142	5.597	5.359	5.533	–	4.066	3.613	1.845	1.658	1.973
143	5.877^	4.676	4.667	–	4.023	3.950	–	0.671	2.537
144	5.208	4.864	5.308	–	3.978	3.519	–	1.796	1.797
145	5.644	5.160	5.501	–	4.219	3.728	1.301	1.872	1.833
146	5.489	5.131	5.501	–	4.238	3.728	2.000	1.712	1.833
147	4.852	5.034	5.164	–	4.752	4.793	–	0.964	3.588
148	4.901^	4.112	5.776	–	5.129	5.663	–	0.756	4.718
149	3.556^	3.530	3.941	3.544	3.526	3.525	3.342^	3.792	3.944
150	–	2.114	2.280	–	0.931	1.052	4.025	4.054	3.896

[a] Modeled with the value 1.000.
[b] Denotes test set compound.
[c] Not determined.

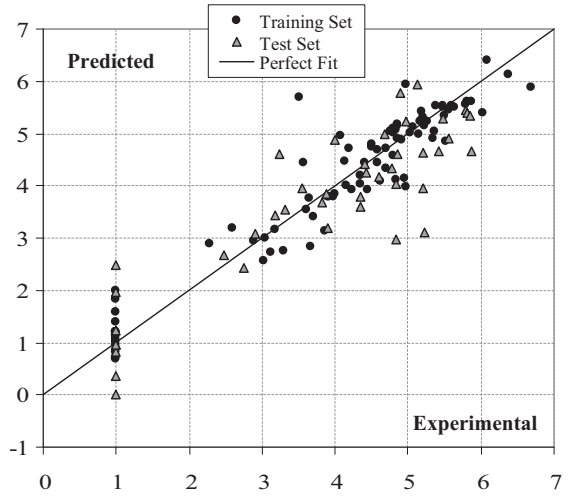

Fig. 3. Predicted $\log_{10}\ k^{*}_{\text{inact}}$ for HLE according to Eq. (6) as function of experimental values.

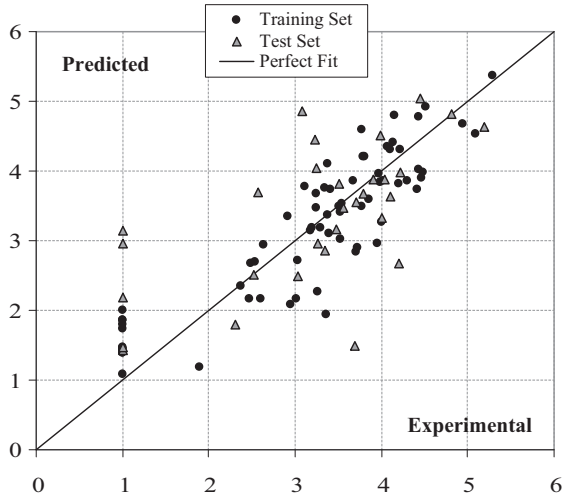

Fig. 4. Predicted $\log_{10}\ k^{*}_{\text{inact}}$ for PR 3 according to Eq. (7) as function of experimental values.

$N = 70$, range = 1.000–5.293, $d = 5$, $N/d = 14$, $R^2 = 0.806$, $S = 0.55$, $F = 53.2$, outliers > 3.$S = 0$, $\text{Corr}^{\max} = 0.236$, $R^2_{\text{loo}} = 0.770$, $S_{\text{loo}} = 0.60$, $R^2_{\text{l–20\%–o}} = 0.656$, $S_{\text{l–20\%–o}} = 0.75$, $S^{\text{Rand}} = 0.99$, $N_{\text{test}} = 30$, $R^2_{\text{test}} = 0.406$, and $S_{\text{test}} = 1.06$.

As in the precedent case, this model exhibits the best balance between the statistical parameters obtained on the training and the test sets. Additional proof on the validity of Eq. (7) is provided by the statistical results in Fig. 4 and Fig. 3S, as well as through the correlation matrix in Table 3S. The FSR does not improve the results produced by Eq. (7) as shown in Table 6S. Present QSAR correctly predicts active and inactive thiadiazolidine derivatives. Those compounds with yet unmeasured experimental activities do not display

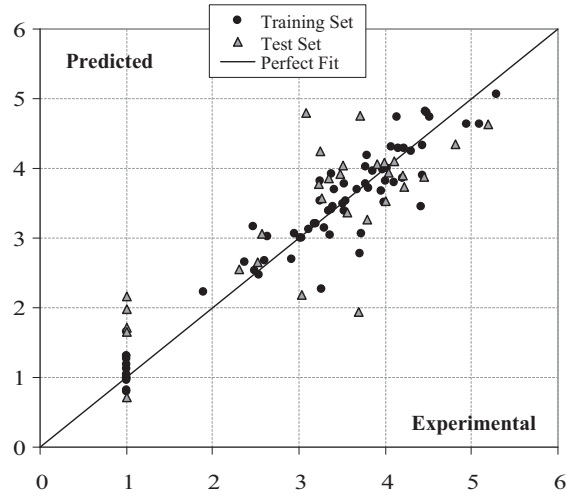

Fig. 5. Predicted $\log_{10}\ k^{*}_{\text{inact}}$ for PR 3 according to Eq. (9) as function of experimental values.

attractive predicted inhibitory potencies with the exception of **84** (4.741), **147** (4.752), and **148** (5.129).

All descriptors in Eq. (7) are 3D, with the exception of the atom-centred fragment *C*-003. The *G*(*N.Cl*) stands for the sum of geometrical distances between nitrogen and chlorine atoms, while *Mor*02*u* is the 3D-MoRSE signal 02/unweighted. There are two GETAWAY: *HATS*8*p*, the leverage-weighted autocorrelation of lag 8/weighted by atomic polarizabilities, and $R5^{+}_{v}$, the *R* maximal autocorrelation of lag 5/weighted by atomic van der Waals volumes. The ranking of contributions of these descriptors reveals that the number of CHR_3 groups contributes most to the predicted PR 3 activities:

$$C\text{-}003\,(0.49) > HATS8p\,(0.32) > G(N..Cl)\,(0.28) > Mor02u\,(0.27) > R5^{+}_{v}\,(0.25) \quad (8)$$

Fig. 5 shows the predictions of the optimized single-descriptor model

$$\log_{10}\ k^{*}_{\text{inact}} = -3.120(\pm 0.2) + 0.145(\pm 0.005) \cdot \text{DCW}_2(0) \quad (9)$$

$N = 70$, range = 1.000–5.293, $d = 1$, $N/d = 70$, $R^2 = 0.920$, $S = 0.34$, $F = 781.6$, outliers > 3.$S = 0$, $R^2_{\text{loo}} = 0.916$, $S_{\text{loo}} = 0.35$, $R^2_{\text{l–20\%–o}} = 0.884$, $S_{\text{l–20\%–o}} = 0.38$, $S^{\text{Rand}} = 0.90$, $N_{\text{test}} = 30$, $R^2_{\text{test}} = 0.643$, and $S_{\text{test}} = 0.74$.

The numerical parameters used in the CORAL calculation of $\text{DCW}_2(0)$ are identical to those mentioned for obtaining Eq. (6). This QSAR leads to a somewhat better prediction of the PR 3 inhibitory activities in both the training and test sets and corroborates that some of the unmeasured compounds may have considerable activities, like **84** (4.354), **147** (4.793) and **148** (5.663).

Table 3
Best QSAR found with the replacement method on PR 3 dataset.

d	*S*	R^2	S_{loo}	R^2_{loo}	S_{test}	R^2_{test}	$\text{Corr}^{\max}$	Molecular descriptors
1	0.88	0.480	0.91	0.445	1.07	0.306	–	*nCt*
2	0.70	0.671	0.74	0.642	1.13	0.270	0.001	*MATS*2*m*, *nCt*
3	0.62	0.748	0.66	0.712	1.17	0.245	0.086	*nCl*, *HATS*8*p*, *C*-003
4	0.60	0.771	0.65	0.729	1.17	0.334	0.203	*Mor*15*m*, *Mor*16*m*, *HATS*8*u*, *nCt*
5	0.55	0.806	0.60	0.769	1.06	0.406	0.236	*G*(*N.Cl*), *Mor*02*u*, *HATS*8*p*, $R5^{+}_{v}$, *C*-003
6	0.54	0.817	0.61	0.773	1.15	0.387	0.542	*VRA*1, *MATS*2*m*, *Mor*16*m*, *R*3*m*, $R5^{+}_{v}$, *nCt*
7	0.54	0.824	0.61	0.771	1.12	0.428	0.362	*MATS*2*m*, *Mor*02*u*, *Mor*16*v*, *ISH*, $R5^{+}_{v}$, *nCt*, *C*-002

Table 4
Best QSAR found with the replacement method on Cat G dataset.

d	S	R^2	S_{loo}	R^2_{loo}	S_{test}	R^2_{test}	$Corr^{max}$	Molecular descriptors
1	0.76	0.462	0.78	0.437	1.27	0.120	–	*piPC*10
2	0.66	0.604	0.69	0.570	0.77	0.563	0.832	*MPC*09, *SRW*10
3	0.61	0.664	0.65	0.624	0.68	0.672	0.819	*X*3*sol*, *MPC*10, *BELv*3
4	0.55	0.729	0.59	0.696	0.55	0.803	0.857	*X*3*sol*, *MPC*07, *BELv*3, *MATS*5*m*
5	0.50	0.778	0.55	0.740	0.82	0.551	0.767	*MPC*08, *BELv*3, *ATS*3*m*, *Mor*02*m*, *Mor*09*m*
6	0.47	0.812	0.50	0.781	0.67	0.704	0.767	*MPC*08, *BELv*3, *ATS*3*m*, *GATS*5*m*, *RDF*085*v*, *Mor*09*m*
7	0.45	0.826	0.50	0.789	0.74	0.658	0.803	*SEige*, *MPC*07, *BELv*3, *ATS*3*m*, *Mor*02*m*, *Mor*09*m*, *R*4*u*

3.3. QSAR on cat G data

The analysis of the various models in Table 4 suggests that we may choose the following QSAR equation:

$$\log_{10}\ k^*_{inact} = -6.896(\pm 3) + 0.0285(\pm 0.003)\cdot MPC08 + 8.0830(\pm 2)\cdot BELv3 - 0.0964(\pm 0.009)\cdot ATS3m - 223.483(\pm 48)\cdot GATS5m + 0.0857(\pm 0.02)\cdot RDF085v + 0.730(\pm 0.1)\cdot Mor09m \quad (10)$$

$N = 84$, range = 1.000–4.824, $d = 6$, $N/d = 14$, $R^2 = 0.812$, $S = 0.47$, $F = 55.7$, outliers > 3.S = 0, $Corr^{max} = 0.767$, $R^2_{loo} = 0.782$, $S_{loo} = 0.50$, $R^2_{l-20\%-o} = 0.706$, $S_{l-20\%-o} = 0.58$, $S^{Rand} = 0.80$, $N_{test} = 36$, $R^2_{test} = 0.704$, and $S_{test} = 0.67$.

In this case, the order of the contributions to the Cat G activities:

$$MPC08\,(1.28) > ATS3m\,(1.25) > Mor09m\,(0.41) > BELv3\,(0.40) > RDF085\,v\,(0.37) > GATS5m\,(0.28) \quad (11)$$

where the two most relevant molecular descriptors have a topological origin: *MPC*08, the molecular path count of order 8, and *ATS*3*m*, the Broto-Moreau 2D-Autocorrelation of a topological structure-lag 3/weighted by atomic masses. Other contributing descriptors are: *Mor*09*m*, the 3D-MoRSE signal 09/weighted by atomic masses; a BCUT (2D): *BELv*3, the lowest eigenvalue no. 3 of Burden matrix, weighted by atomic van der Waals volumes; a 3D radial distribution function (RDF): *RDF*085*v*, RDF 8.5/weighted by atomic van der Waals volumes, and the 2D-Autocorrelation *GATS*5*m*, Geary autocorrelation-lag 5/weighted by atomic masses. All these quantities are well-defined in the literature.

In this case CORAL yields the following equation:

$$\log_{10}\ k^*_{inact} = -1.695(\pm 0.2) + 0.129(\pm 0.007)\cdot DCW_3(1) \quad (12)$$

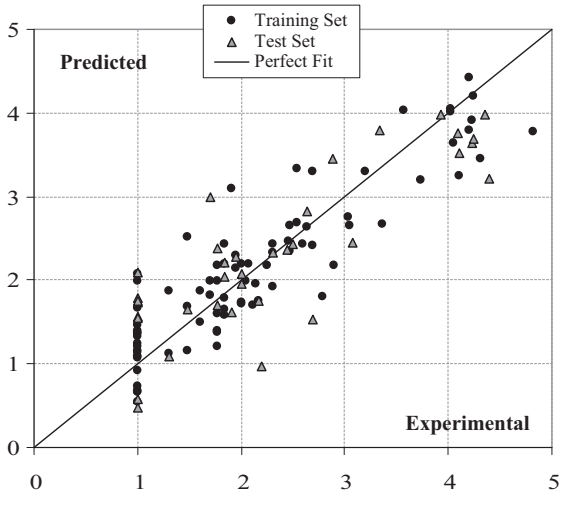

Fig. 6. Predicted $\log_{10}\ k^*_{inact}$ for Cat G according to Eq. (10) as function of experimental values.

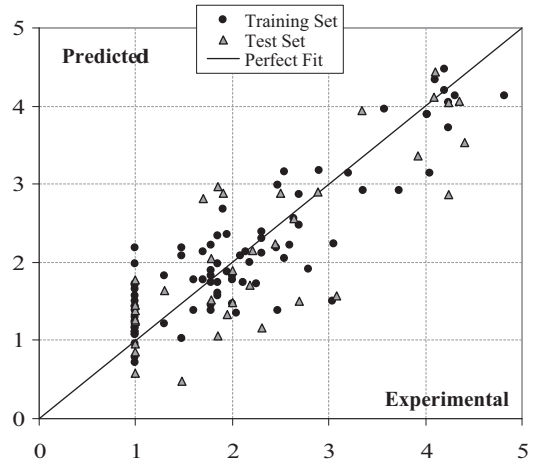

Fig. 7. Predicted $\log_{10}\ k^*_{inact}$ for Cat G according to Eq. (12) as function of experimental values.

$N = 84$, range = 1.000–4.824, $d = 1$, $N/d = 84$, $R^2 = 0.787$, $S = 0.48$, $F = 302.5$, outliers > 3.S = 1, $R^2_{loo} = 0.778$, $S_{loo} = 0.49$, $R^2_{l-20\%-o} = 0.741$, $S_{l-20\%-o} = 0.53$, $S^{Rand} = 0.91$, $N_{test} = 36$, $R^2_{test} = 0.672$, and $S_{test} = 0.69$.

The numerical parameters used in the CORAL calculation for $DCW_3(1)$ are the same as previously, with exception to the number of epochs that in this case is 6, while the threshold value adopted for the training and test set is 1. Table 2 shows that the predictions of Eqs. (10) and (12) are consistent, despite of coming from two utterly different modeling strategies. We plot the predictions of Eqs. (10) and (12) in Figs. 6 and 7, respectively, while the dispersion plot of the residuals is available in Supplementary Material section (Figs. 5S&6S). The numerical values for all the calculated descriptors appearing in Eqs. (4)–(12) are provided in Table 9S.

4. Conclusions

The 1,2,5-thiadiazolidin-3-one 1,1-dioxide based compounds have remarkable selectivity and are highly efficient inhibitors of the human serine proteinases HLE, Cat G and PR 3. We think that QSAR may be useful for a rational search of new heterocyclic inhibitors of this type with low molecular weights, making it possible to address the protease–antiprotease imbalance. In this work, we have developed predictive QSAR based on molecular descriptors calculated with Dragon, Recon and CORAL software by appropriately representing the chemical structures of the 1,2,5-thiadiazolidin-3-one 1,1-dioxides. Present results strongly suggest that such QSAR models are suitable for distinguishing between active or inactive structures beforehand, thus being powerful tools for the search of new compounds with satisfactory activity. One of the main contributions of this paper consists of proving that QSAR based straightforward multiparametric linear regression is as effective as the more complicated technique of molecular modeling docking studies.

Acknowledgments

We gratefully acknowledge the financial support by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), PIP11220100100151 project. We also thank the Universidad Nacional de La Plata.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.07.007.

References

[1] W.C. Groutas, R. Kuang, R. Venkataraman, J.B. Epp, S. Ruan, O. Prakash, Structure-based design of a general class of mechanism-based inhibitors of the serine proteinases employing a novel amino acid-derived heterocyclic scaffold, Biochemistry 36 (1997) 4739–4750.
[2] R. Dhami, B. Gilks, C. Xie, K. Zay, J.L. Wright, A. Churg, Acute cigarette smoke-induced connective tissue breakdown is mediated by neutrophils and prevented by 1-antitrypsin, Am. J. Respir. Cell Mol. Biol. 22 (2000) 244–252.
[3] R.A. Stockley, The role of proteinases in the pathogenesis of chronic bronchitis, Am. J. Respir. Crit. Care Med. 150 (1994) S109–S113.
[4] P.J. Barnes, Novel approaches and targets for treatment of chronic obstructive pulmonary disease, Respir. Crit. Care Med. 160 (1999) S72–S79.
[5] L.J. Copp, A. Krantz, R.W. Spencer, Kinetics and mechanism of human leukocyte elastase inactivation by ynenol lactones, Biochemistry 2 (1987) 169–178.
[6] J.A. Katzenellenbogen, R. Rai, W. Dai, Enol lactone derivatives as inhibitors of human neutrophil elastase and trypsin-like proteases, Bioorg. Med. Chem. Lett. 2 (1992) 1399–1404.
[7] M.A. Hernandez, J.C. Powers, J. Glinski, J. Oleksyszyn, J. Vijayalakshmi, E.F. Meyer, Effect of the 7-amino substituent on the inhibitory potency of mechanism-based isocoumarin inhibitors for porcine pancreatic and human neutrophil elastases: a 1.85-Å X-ray structure of the complex between porcine pancreatic elastase and 7-[(N-tosylphenylalanyl)amino]-4-chloro-3-methoxyisocoumarin, J. Med. Chem. 35 (1992) 1121–1129.
[8] W.C. Groutas, M.J. Brubaker, M.A. Stanga, J.C. Castrisos, J.P. Crowley, E.J. Shatz, Inhibition of human leukocyte elastase by derivatives of N-hydroxysuccinimide. A structure–activity-relationship study, J. Med. Chem. 32 (1989) 1607–1611.
[9] W.C. Groutas, R. Venkataraman, M.J. Brubaker, M.A. Stanga, Inhibition of human leukocyte elastase by phosphate esters of N-hydroxysuccinimide and its derivatives: direct observation of a phosphorylated enzyme by sup 31 P nuclear magnetic resonance spectroscopy, Biochemistry 30 (1991) 4132–4136.
[10] W.C. Groutas, R. Venkataraman, M.J. Brubaker, J.B. Epp, L.S. Chong, M.A. Stanga, J.J. McClenahan, F. Tagusagawa, 3-(Alkylthio)-N-hydroxysuccinimide derivatives: potent inhibitors of human leukocyte elastase, Biochim. Biophys. Acta 1164 (1993) 283–288.
[11] W.C. Groutas, H. Huang, J.B. Epp, R. Venkataraman, J.J. McClenahan, F. Tagusagawa, Mechanism-based inhibition of human leukocyte elastase and cathepsin G by susbtituted dihydrouracils, Biochim. Biophys. Acta 1227 (1994) 130–136.
[12] D.J. Underwood, B.G. Green, R. Chabin, S. Mills, J.B. Doherty, P.E. Finke, M. MacGoss, S.K. Shah, C.S. Burgey, T.A. Dickinson, P.R. Griffin, T.E. Lee, K.M. Swiderek, T. Covey, W.M. Westler, W.B. Knight, Mechanism of inhibition of human leukocyte elastase by β-lactams. 3. Use of electrospray ionization mass spectrometry and two-dimensional NMR techniques to identify β-lactam-derived E–I complexes, Biochemistry 34 (1995) 14344–14355.
[13] W.C. Groutas, N. Houser-Archield, L.S. Chong, R. Venkataraman, J.B. Epp, H. Huang, J.J. McClenahan, Efficient inhibition of human leukocyte elastase and cathepsin G by saccharin derivatives, J. Med. Chem. 36 (1993) 3178–3181.
[14] W.C. Groutas, L.S. Chong, R. Venkataraman, R. Kuang, J.B. Epp, N. Houser-Archield, H. Huang, J.R. Hoidal, Amino acid-derived phthalimide and saccharin derivatives as inhibitors of human leukocyte elastase, cathepsin G, and proteinase 3, Arch. Biochem. Biophys. 332 (1996) 335–340.
[15] D.J. Hlasta, C. Subramanyam, M.R. Bell, P.M. Carabateas, J.J. Court, R.C. Desai, M.L. Drozd, W.M. Eickhoff, E.W. Ferguson, R.J. Gordon, J.A. Johnson, V. Kumar, A.L. Maycock, K.R. Mueller, E.D. Pagani, D.T. Robinson, M.T. Saindane, P.J. Silver, S. Subramanian, A novel class of cyclic β-dicarbonyl leaving groups and their use in the design of benzisothiazolone human leukocyte elastase inhibitors, J. Med. Chem. 38 (1995) 739–744.
[16] W.C. Groutas, J.B. Epp, R. Kuang, S. Ruan, L.S. Chong, R. Venkataraman, J.Tu.S. He, H. Yu, Q. Fu, Y.H. Li, T.M. Truong, N.T. Vu, 1,2,5-Thiadiazolidin-3-one 1,1 dioxide: a powerful scaffold for probing the S9 subsites of (chymo)trypsin-like serine proteases, Arch. Biochem. Biophys. 385 (2001) 162–169.
[17] C. Hansch, A. Leo, Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, DC, 1995.
[18] H. Kubinyi, QSAR: Hansch Analysis and Related Approaches, Wiley-Interscience, New York, 2008.
[19] T. Puzyn, J. Leszczynski, M.T. Cronin, Recent Advances in QSAR Studies: Methods and Applications, 1st ed., Springer, New York, 2009.
[20] W.C. Groutas, R. Kuang, S. Ruan, J.B. Epp, R. Venkataraman, T.M. Truong, Potent and specific inhibition of human leukocyte elastase, cathepsin G and proteinase 3 by sulfone derivatives employing the 1,2,5-thiadiazolidin-3-one 1,1 dioxide scaffold, Bioorg. Med. Chem. 6 (1998) 661–671.
[21] R. Kuang, J.B. Epp, S. Ruan, L.S. Chong, R. Venkataraman, J. Tu, S. He, T.M. Truong, W.C. Groutas, Utilization of the 1,2,5-thiadiazolidin-3-one 1,1 dioxide scaffold in the design of potent inhibitors of serine proteases: SAR studies using carboxylates, Bioorg. Med. Chem. 8 (2000) 1005–1016.
[22] S. He, R. Kuang, R. Venkataraman, J. Tu, T.M. Truong, H.-K. Chan, W.C. Groutas, Potent inhibition of serine proteases by heterocyclic sulfide derivatives of 1,2,5-thiadiazolidin-3-one 1,1 dioxide, Bioorg. Med. Chem. 8 (2000) 1713–1717.
[23] W.C. Groutas, S. He, R. Kuang, S. Ruan, J. Tu, H.-K. Chan, Inhibition of serine proteases by functionalized sulfonamides coupled to the 1,2,5-thiadiazolidin-3-one 1,1 dioxide scaffold, Bioorg. Med. Chem. 9 (2001) 1543–1548.
[24] W.C. Groutas, N.M. Schechter, S. He, H. Yu, P. Huang, J. Tu, Human chymase inhibitors based on the 1,2,5-thiadiazolidin-3-one 1,1 dioxide scaffold, Bioorg. Med. Chem. Lett. 9 (1999) 2199–2204.
[25] R. Kuang, J.B. Epp, S. Ruan, H. Yu, P. Huang, S. He, J. Tu, N.M. Schechter, J. Turbov, C.J. Froelich, W.C. Groutas, A general inhibitor scaffold for serine proteases with a (Chymo) trypsin-like fold: solution-phase construction and evaluation of the first series of libraries of mechanism-based inhibitors, J. Am. Chem. Soc. 121 (1999) 8128–8129.
[26] J.F. Morrison, C.T. Walsh, The behavior and significance of slow-binding enzyme inhibitors, Adv. Enzymol. 61 (1988) 201–301.
[27] Hyperchem 7, Hypercube, Inc., Gainesville, 2007, http://www.hyper.com.
[28] Dragon Milano Chemometrics and QSAR Research Group, 2010, http://michem.disat.unimib.it/chm.
[29] R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics, Wiley–VCH, Weinheim, 2009.
[30] Recon Version 5.5, Rensselaer Polytechnic Institute, Troy, NY, USA, 2002, http://www.drugmining.com.
[31] B.K. Lavine, C.E. Davidson, C. Breneman, W.J. Katt, Electronic Van der Waals surface property descriptors and genetic algorithms for developing structure–activity correlations in olfactory databases, J. Chem. Inf. Comput. Sci. 43 (2003) 1890–1905.
[32] A. Worachartcheewan, C. Nantasenamat, T. Naenna, C. Isarankura-Na-Ayudhya, V. Prachayasittikul, Modeling the activity of furin inhibitors using artificial neural network, Eur. J. Med. Chem. 44 (2009) 1664–1673.
[33] R.F.W. Bader, Atoms in Molecules—A Quantum Theory, Clarendon Press, Oxford, 1990.
[34] C.M. Breneman, L.W. Weber, Transferable atom equivalents. Assembling accurate electrostatic potential fields for large molecules from ab initio and PROAIMS results on model systems, in: G.A. Jeffrey, J.F. Piniella (Eds.), The Application of Charge Density Research to Chemistry and Drug Design, Plenum, New York, 1991.
[35] Matlab 5.0, The MathWorks, Inc, Natick, MA, USA, 1996, http://www.mathworks.com.
[36] P.R. Duchowicz, E.A. Castro, F.M. Fernández, M.P. González, A new search algorithm of QSPR/QSAR theories: normal boiling points of some organic molecules, Chem. Phys. Lett. 412 (2005) 376–380.
[37] A.G. Mercader, P.R. Duchowicz, F.M. Fernández, E.A. Castro, Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories, Chemom. Intell. Lab. Syst. 92 (2008) 138–144.
[38] P.R. Duchowicz, M. Fernández, J. Caballero, E.A. Castro, F.M. Fernández, QSAR of non-nucleoside inhibitors of HIV-1 reverse transcriptase, Bioorg. Med. Chem. 14 (2006) 5876–5889.
[39] P.R. Duchowicz, M.G. Vitale, E.A. Castro, M. Fernandez, J. Caballero, QSAR analysis for heterocyclic antifungals, Bioorg. Med. Chem. 15 (2007) 2680–2689.
[40] P.R. Duchowicz, A. Talevi, L.E. Bruno-Blanch, E.A. Castro, New QSPR study for the prediction of aqueous solubility of drug-like compounds, Bioorg. Med. Chem. 16 (2008) 7944–7955.
[41] P.R. Duchowicz, M. Goodarzi, M.A. Ocsachoque, G.P. Romanelli, E.V. Ortiz, J.C. Autino, D.O. Bennardi, D. Ruiz, E.A. Castro, QSAR analysis on Spodoptera litura antifeedant activities for flavone derivatives, Sci. Total Environ. 408 (2009) 277–285.
[42] M. Goodarzi, P.R. Duchowicz, C.H. Wu, F.M. Fernández, E.A. Castro, New hybrid genetic based support vector regression as QSAR approach for analyzing flavonoids–GABA(A) complexes, J. Chem. Inf. Model. 49 (2009) 1475–1485.
[43] Multicollinearity.doc© 2002 Jeeshim and KUCC625 (2003-05-09).
[44] J.D. Curto, J.C. Pinto, New multicollinearity indicators in linear regression models, Int. Stat. Rev. 75 (2007) 114–121.
[45] N.R. Draper, H. Smith, Applied Regression Analysis, John Wiley & Sons, New York, 1981.
[46] Coral 1.5, 2010, http://www.insilico.eu/coral.
[47] ACD/ChemSketch Freeware, version 12.01, Advanced Chemistry Development, Inc, Toronto, ON, Canada, 2009, http://www.acdlabs.com.
[48] A.A. Toropov, E. Benfenati, SMILES in QSPR/QSAR modeling: results and perspectives, Curr. Drug Discov. Technol. 4 (2007) 77–116.
[49] A.A. Toropov, E. Benfenati, Additive SMILES-based optimal descriptors in QSAR modelling bee toxicity: using rare SMILES attributes to define the applicability domain, Bioorg. Med. Chem. 26 (2008) 4801–4809.
[50] A.A. Toropov, A.P. Toropova, E. Benfenati, Simplified molecular input line entry system-based optimal descriptors: quantitative structure–activity relationship modeling mutagenicity of nitrated polycyclic aromatic hydrocarbons, Chem. Biol. Drug Des. 73 (2009) 515–525.

[51] A.A. Toropov, A.P. Toropova, E. Benfenati, D. Leszczynska, J. Leszczynski, InChI-based optimal descriptors: QSAR analysis of fullerene [C60]-based HIV-1 PR inhibitors by correlation balance, Eur. J. Med. Chem. 45 (2010) 1387–1394.
[52] S. Wold, L. Eriksson, Statistical validation of QSAR results, in: H. van de Waterbeemd (Ed.), Chemometrics Methods in Molecular Design, VCH, Weinheim, 1995, pp. 309–318.
[53] D.M. Hawkins, S.C. Basak, D.J. Mills, Assessing model fit by cross validation, J. Chem. Inf. Model. 43 (2003) 579–586.
[54] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graphics Modell. 20 (2002) 269–276.
[55] P.R. Duchowicz, E.A. Castro, F.M. Fernández, Alternative algorithm for the search of an optimal set of descriptors in QSAR–QSPR studies, MATCH Commun. Math. Comput. Chem. 55 (2006) 179–192.