European Journal of Medicinal Chemistry 44 (2009) 3439–3444

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

A QSAR study on relationship between structure of sulfonamides and their carbonic anhydrase inhibitory activity using the eigenvalue (EVA) method

Oral Oltulu*, Mehmet M. Yaşar, Erol Eroğlu

Harran University, Physics Department, Osmanbey Kampus, 6300 Sanliurfa, Turkey

ARTICLE INFO

Article history:
Received 12 November 2008
Received in revised form
23 January 2009
Accepted 7 February 2009
Available online 21 February 2009

Keywords:
EVA
Sulfonamides
DFT
Carbonic anhydrase inhibition

ABSTRACT

In this study, we present an application of EVA descriptors for a QSAR model of inhibition of carbonic anhydrase isozyme CA II by an heterogeneous set of 66 sulfonamide compounds. For each of the compounds, geometry optimization and frequency calculations have been performed using the DFT/B3LYP level of the theory in conjugated with the 6-31G* basis set. Different numbers of EVA descriptors for each structure were produced by applying various values of Gaussian kernel of a fixed standard deviation, σ (cm^{-1}) and sampled at fixed increments of L (cm^{-1}) during the evaluation of the descriptors based on their vibrational frequencies. The set of compounds was divided into two subsets. The first subset contained the 22 compounds that were used as the test compounds. The remaining 44 compounds were used as the training set. Several QSAR models have been developed using these calculated EVA descriptors and the carbonic anhydrase isozyme CA II inhibitory data (K_i) of the compounds. Among the QSAR models evaluated, the one that produced the best statistical results had the parameters σ and L both equal to 5 cm^{-1}. This model produced correlation values (R^2) of 0.777 and 0.616 for the training and test sets, respectively. The results of this study showed that EVA descriptors perform well as explanatory and predictive tools for modeling the inhibition activity of carbonic anhydrase by a set of sulfonamide compounds.

1. Introduction

The quantitative structure–activity relationship (QSAR) process is used to quantitatively correlate chemical structure with a well-defined process, such as biological activity, chemical reactivity, physico-chemical properties, or environmental behavior. Molecular descriptors are mathematical values that describe the structures or shapes of the molecules, and they are used to predict the activity of molecules as well as their physical and chemical properties. These descriptors may be empirical or theoretical values. A QSAR model can be constructed by using many types of descriptors, such as topological, quantum chemical, geometrical, and spectral descriptors. In the present study, we used spectral-type eigenvalue (EVA) descriptors to model the extent to which sulfonamide compounds inhibit carbonic anhydrase isozyme CA II. The EVA concept was originally developed by scientists at Shell Research, Limited [1,2]. The main motivation for developing the EVA concept was to overcome the alignment-related limitations of the CoMFA (Comparative Molecular Field Analysis) method [3]. EVAs are built from normal coordinate analysis, namely vibrational eigenvalues, and derived from either a quantum theoretical or molecular mechanical treatment of molecular structures. Early applications of EVAs by Turner *et al.* [4–7] showed that EVAs can be successfully used to construct QSAR models based on a variety of biological endpoints. It was concluded that, in most cases, the EVA models have been found to yield statistically robust QSAR models that are comparable, in statistical terms, to those obtained using CoMFA but without the difficulties associated with aligning the structures concerned. Furthermore, comparisons of 3-D QSAR methods (EVA, CoMFA, and CoMSIA (Comparative Molecular Similarity Indices Analysis)) have been studied for different biological endpoints by different groups [8–11] These groups have also confirmed that the statistical quality of the EVA models is similar to that obtained from CoMFA or CoMSIA approaches. Moreover, the EVA method was successfully applied to the QSAR modeling of HIV-1 integrase inhibitors [12], calcium channel agonists [13], structure–anti-oxidant activity [14,15], peroxisome proliferator-activated receptor gamma agonists [16], and structure–rate-constant relationships of (4 + 2) cycloadditions [17].

The carbonic anhydrases (CAs, EC 4.2.1.1) are biologically ubiquitous, zinc-containing enzymes (present in prokaryotes and eukaryotes) and are encoded by four distinct, evolutionarily

* Corresponding author. Tel.: +90 4143440050; fax: +90 4143440051.
E-mail address: oltulu@harran.edu.tr (O. Oltulu).

0223-5234/$ – see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2009.02.016

unrelated gene families. Sixteen different α-CA isozymes or CA-related proteins (CARP), with variability in their subcellular localization and differences in tissue distribution, have been isolated and described in mammals [18–22]. These enzymes efficiently catalyze a very simple physiological reaction, i.e., the chemical interconversion of carbon dioxide to bicarbonate ions, and, thus, are involved in critical physiological processes. These crucial processes are involved in respiration and the transport of CO_2/bicarbonate between metabolizing tissues and the lungs, as well as in pH changes and the regulation of CO_2 fixation. They are also involved in electrolyte secretion in a variety of tissues and in many biosynthetic reactions, such as gluconeogenesis, lipogenesis, ureagenesis, bone resorption, calcification, tumorigenicity, and many other physiological or pathological processes [22]. CAs are important targets for the design of novel pharmacological agents useful in the treatment or prevention of a variety of disorders, such as glaucoma, acid–base disequilibria, epilepsy, other neuromuscular diseases, as well as altitude sickness and obesity [22,23]. Selectivity of enzyme inhibitors for closely related enzyme isoforms is particularly important for the design of potential drug candidates. For example, all the drugs used for the treatment of glaucoma have some systemic side effects [24]. To reduce the side effects associated with these drugs, it is of interest to develop new CA II inhibitor agents that can be used topically for the long-term management of glaucoma. The prospect of overcoming the systemic side effects of a drug and achieving the desired effect at lower dosages is very attractive to researchers. QSAR studies are tools for predicting the endpoints of interest when organic molecules are used as drugs [25]. Many physiological activities of molecules can be related to their compositions and structures. Since the first discovery of the QSAR by Hansch in early sixties, many types of molecular descriptors have been developed and used to perform QSAR analysis [26]. In the literature, there have been a number of QSAR studies of CA inhibition by sulfonamides using different types of descriptors namely quantum chemical [27] and topological [28] descriptors and the 3-D approach of CoMFA and CoMSIA [29]. In the present study, we present an application of EVA descriptors for the QSAR modeling of the inhibition of carbonic anhydrase isozyme CA II by a heterogeneous set of 66 sulfonamide compounds.

2. Computational details

2.1. CA II inhibition data

Experimental inhibition data (inhibition constant, K_i) of the 66 sulfonamides, which were obtained by the CO_2 hydration reaction catalyzed by CA II isozyme, are shown in Fig. 1. These data were collected from five different literature sources [30–34] and were converted into values on a logarithmic scale for convenience.

2.2. Computational details of descriptor and statistical analysis

For all the molecules studied, 3-D modeling and calculations were performed using the Gaussian 03 quantum chemistry package [35]. To save computational time, initial geometry optimizations were conducted with the molecular mechanics (MM) method, using the MM+ force fields. The lowest energy confirmations of the molecules obtained by the MM method were further optimized by the DFT [36] method by employing Becke's three-parameter hybrid functional (B3LYP) [37] and the 6-31G* basis set. Their fundamental vibrations were also calculated using the same method to check if there were true minima. All the computations were conducted for the ground states of these molecules as single states.

2.3. Calculation of the EVA descriptors

The EVA descriptors were derived from calculated vibrational frequencies of the compounds obtained through the application of a classical normal coordinate analysis (NCA) to an appropriately energy minimized structure calculated using the DFT method B3LYP functional in conjugated with the 6-31G* basis set.

The number of vibrational normal modes varies with the number of atoms N in a molecule. A molecule without axial symmetry has $3N-6$ normal modes of vibration, and the eigenvalues from the NCA correspond to the vibrational wavenumbers (vwns). Once determined, from whatever source, the set of vwns for a given structure is projected onto a linear-bounded frequency scale (BFS) that typically ranges from 1 to 4000 cm^{-1}. The use of this range means that all fundamental vibrational normal modes of the molecules are included in the analysis. If a vwn exceeds 4000 cm^{-1}, then, either the BFS can be extended, or all vwns from all molecules can be scaled according to scale factors. Next, a Gaussian kernel of fixed standard deviation (σ) is placed over each frequency value. The BFS is then sampled at fixed increments of L (cm^{-1}), and the value of the resulting EVA descriptor at each sample point x, EVA(x), is the sum of the amplitudes of the overlaid kernels at that point.

$$\mathrm{EVA}(x) = \sum_{i=1}^{3N-6} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(x-f_i)^2/2\sigma^2}$$

where f_i is the ith frequency of the structure.

The purpose of the above EVA smoothing procedure is not an attempt to simulate the infrared spectrum of the molecule of interest, since the transition dipole data are ignored, but rather to provide a basis upon which vibrations occurring at slightly different frequencies may be compared to one another. The Gaussian function applied to define peak shapes adds a probabilistic element, in that the peak maxima are centered at each of the calculated frequency values (f_i), and, thus, these points are taken to be the most probable values of the respective frequencies. An EVA descriptor sampled at a point $x \neq f_i$ is thus considered to be a less probable value of the ith frequency, and the corresponding contribution of f_i to the final value of EVA(x) will be less than the maximum possible contribution.

2.4. Statistical analysis

CODESSA PRO [38] (**Co**mprehensive **De**scriptors for **S**tructural and **S**tatistical **A**nalysis), Version 2.7.2 was used for statistical analysis. This code uses diverse statistical structure property/activity correlation techniques for the analysis of experimental data in combination with the calculated molecular descriptors. The heuristic method (HM) [39] implemented in CODESSA PRO was employed for selecting the 'best' regression model.

In this study, statistical fitness of the derived equations is given with the following parameters; the squared correlation coefficient (goodness of fit), R^2, the standard deviation of the regression, s^2, the F-value for the regression, F. Predictive power of the derived equations is given with the cross-validated correlation coefficient, R^2_{CV}, for internal validation and the predicted square of the correlation coefficient, R^2_{pred}, which is calculated from the test set by applying the equation developed on the training set. Reliability of the derived equations is given by the Y-randomization test.

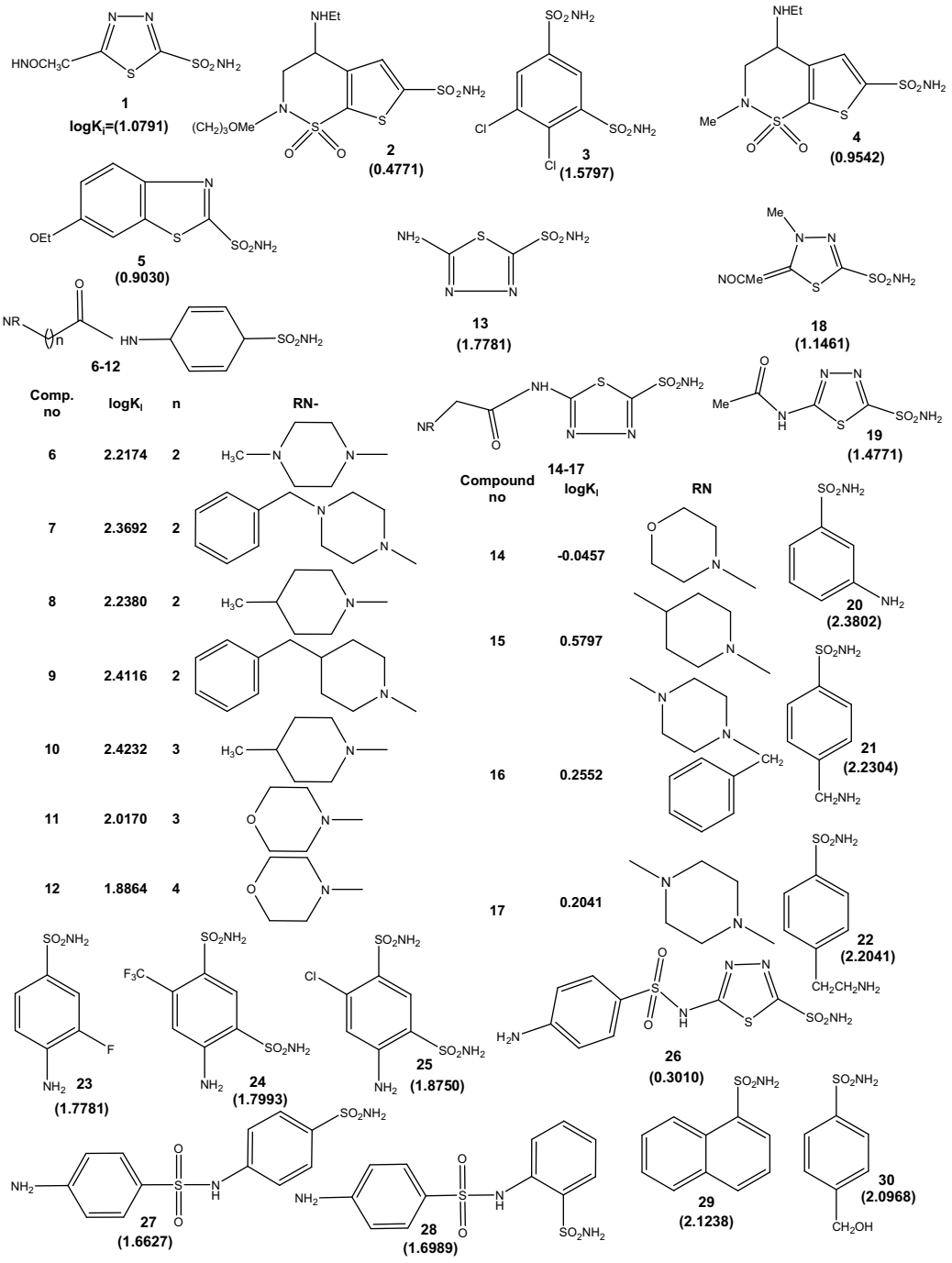

Fig. 1. The structures and biological activity values (K_i) of 66 aromatic and heterocyclic sulfonamide compounds.

3. Results and discussion

The structures and biological activity values (K_i) of 66 aromatic and heterocyclic sulfonamides are given in Fig. 1. The compounds were divided into two subsets (a training set of $N = 44$ compounds and a test set of $N = 22$ compounds). The second **2**, third **3**, fifth **5**, sixth **6**, etc., data points went into the training set, and the first **1**, fourth **4**, seventh **7**, etc., went into the test set. Using the heuristic method, several regression equations were obtained for various σ and L values in this study. Among the regression results, six equations were selected as models, and they are presented in Table 1. The plot of observed versus calculated K_i for CA II using the best statistical model ($\sigma = 5\ \mathrm{cm}^{-1}$ and $L = 5\ \mathrm{cm}^{-1}$) is shown in Fig. 2.

Table 1 shows that the model with $\sigma = 5\ \mathrm{cm}^{-1}$ and $L = 5\ \mathrm{cm}^{-1}$ gives the best statistical parameters. This QSAR model which is given in bold type in Table 1 has a squared correlation coefficient, R^2, of approximately 0.777 and a cross-validated correlation coefficient, R^2_{CV}, of approximately 0.693. Generally, when R^2_{CV} values are larger than 0.5, the models are considered to have sound predictive power [40]. The accuracy of cross-validation results is extensively accepted in the literature based on the R^2_{CV} value. In this sense, a high value of the statistical characteristic ($R^2_{CV} > 0.5$) is considered to be indicative of the highly accurate predictive ability of the

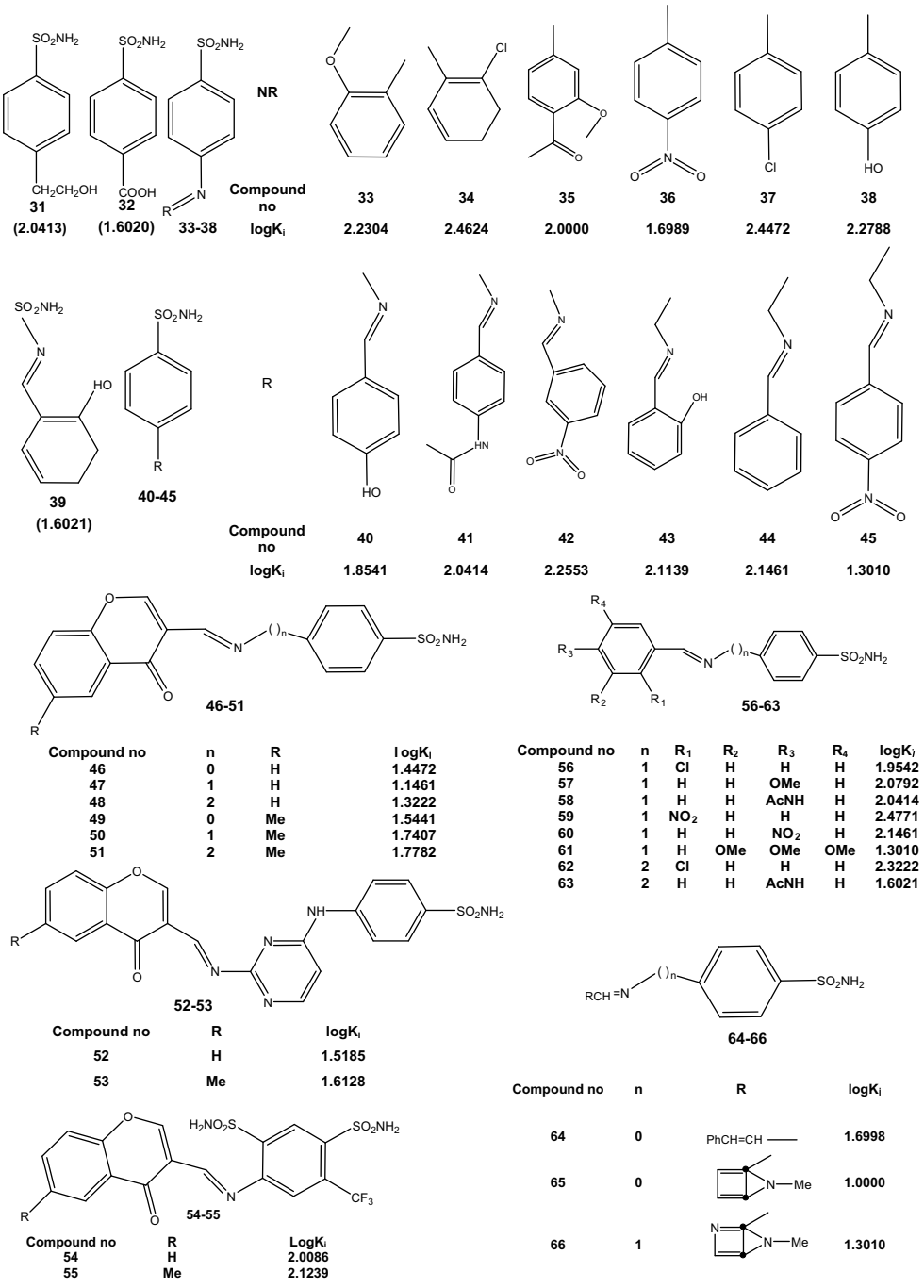

Fig. 1. (*continued*).

model. However, several authors suggest that a high R^2_{CV} value appears to be a necessary, but not sufficient, condition for a model to have a highly accurate predictive power [41]. The predictive ability of a QSAR model can be estimated only by using a sufficiently large collection of compounds that was not used for building the model. Therefore, we have tested various models against our external test set. When the predictive power of the best model was tested against the external (test) set, it gave the best statistical parameters, predicting a correlation coefficient value, R^2_{pred}, of approximately 0.616. Studies of the use of EVA for QSAR have shown that the use of various σ and L values in the generation of the descriptor can, in some instances, have a significant effect on the statistical quality of the derived QSAR models [4–7]. In our case, $\sigma = 5\ cm^{-1}$ and $L = 5\ cm^{-1}$ gave the best statistical parameters. In Table 1, two other models with ($\sigma = 2.5\ cm^{-1}$ and $L = 5\ cm^{-1}$) and ($\sigma = 10\ cm^{-1}$ and $L = 5\ cm^{-1}$) also gave statistically satisfactory results as evidenced by their R^2 values of approximately 0.736 and 0.735, R^2_{pred} values of approximately 0.571 and 0.561, and R^2_{CV} values of approximately 0.642 and 0.660, respectively.

To check the reliability of the proposed model, we used the *Y*-randomization test [42]. In any predictive model, excellent adjustments and even satisfactory predictions can be achieved without the existence of a real relationship between the molecular structure description and the property of the studied set [43].

Table 1
Obtained QSAR model with statistical parameters for various σ and L values.

Training set ($N = 44$)				Gaussian kernel and sampling parameters		Test set ($N = 22$)	
R^2	R^2_{CV}	F	s^2	σ (cm^{-1})	L (cm^{-1})	R^2 (pred)	RMSE
0.736	0.642	21.25	0.120	2.5	5	0.571	0.467
0.777	**0.693**	**26.48**	**0.101**	**5**	**5**	**0.616**	**0.453**
0.735	0.660	21.09	0.120	10	5	0.561	0.400
0.678	0.572	16.07	0.146	20	10	0.302	0.498
0.626	0.502	12.76	0.170	40	20	0.250	0.679
0.400	0.249	6.52	0.266	80	40	0.278	0.493

R^2; the square of the regression correlation coefficient. R^2_{CV}; the cross-validated square of the regression correlation coefficient. F; the F-value for the regression. s^2; the standard deviation of the regression. σ; the fixed Gaussian kernel. L; the sampling increment. R^2 (pred.); the predicted square of the regression correlation coefficient. RMSE; the root mean square error.
Training set: **2**, **3**, **5**, **6**, **8**, **9**, **11**, **12**, … **65**, **66**; total 44 compounds. Test set: **1**, **4**, **7**, **10**, **11**, …, **64**; total 22 compounds.

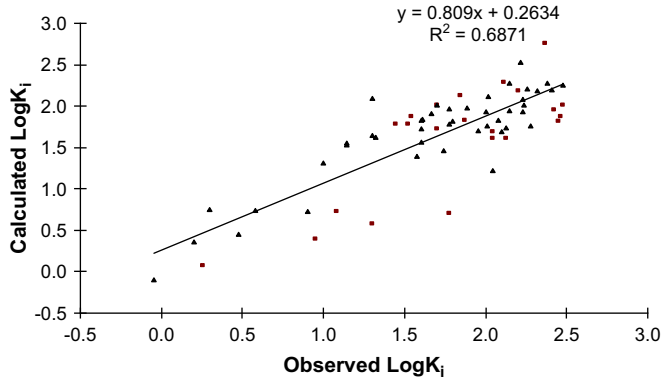

Fig. 2. The plot of observed versus calculated K_i for CA II using statistically the best QSAR model ($\sigma = 5\ cm^{-1}$ and $L = 5\ cm^{-1}$). Triangles represent the training compounds, and the squares represent the test set compounds.

A method capable of distinguishing between a real structure–activity correlation and a chance description is then needed in every case [40]. Several random shuffles of the $Y - (\log K_i)$ were chosen, and the modeling process was performed for all cases. The results are shown in Fig. 3. The lower values of R^2 and R^2_{CV} in comparison with the real model's results support the hypothesis that the good statistical results obtained by the QSAR model are not due to a chance correlation or structural dependency of the training set.

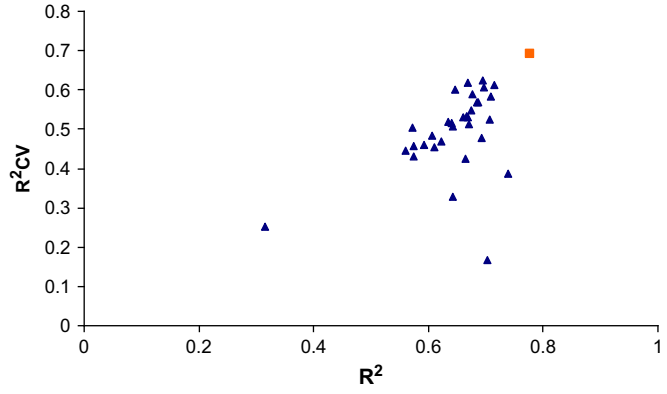

Fig. 3. Y-randomization test associated to the QSAR model ($\sigma = 5\ cm^{-1}$ and $L = 5\ cm^{-1}$). Triangles represent the randomly ordered activities, and the square corresponds to the real activities.

4. Conclusions

The results of this study show that EVA descriptors perform well as explanatory and predictive tools for modeling the inhibition activity of carbonic anhydrase by a set of sulfonamide compounds. The reliability of the proposed QSAR model has been checked by the Y-randomization test, and the results of the test support the hypothesis that the good statistical parameters obtained by the QSAR model are not due to a chance correlation or structural dependency of the training set. The results of this study show that optimization of the EVA descriptors by adjusting σ and L can influence the predictive power of the QSAR model.

When a well established QSAR model is obtained, it can be used to screen a virtual library. One can easily create a virtual library of sulfonamides containing thousands of new compounds by changing the functional groups of compounds shown in Fig. 1. It should be noted that, in principle, a QSAR model is only considered as a reliable screening tool within its domain of application, not in the entire chemical space. Therefore, its domain of application should be defined with a great caution.

By considering the warnings above and the results achieved in this study, it is the authors' opinion that the models developed in this study can be used to calculate the EVA descriptors and estimate the inhibition activity of novel virtual sulfonamide compounds prior to their synthesis.

References

[1] P. Jonathan, W.V. McCarthy, A.M.I. Roberts, J. Chemom. 10 (1996) 189–213.
[2] A.M. Ferguson, T. Heritage, P. Jonathon, S.E. Pack, L. Phillips, J. Rogan, P.J. Snaith, J. Comput. Aided Mol. Des. 11 (1997) 143–152.
[3] R.D. Cramer, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959.
[4] D.B. Turner, P. Willett, A.M. Ferguson, T. Heritage, J. Comput. Aided Mol. Des. 13 (1999) 271–296.
[5] D.B. Turner, P. Willett, A.M. Ferguson, T.W. Heritage, J. Comput. Aided Mol. Des. 11 (1997) 409–422.
[6] D.B. Turner, P. Willett, J. Comput. Aided Mol. Des. 14 (2000) 1–21.
[7] D.B. Turner, P. Willett, Eur. J. Med. Chem. 35 (2000) 367–375.
[8] X.Q. Xiong, D.M. Zhao, P.F. Bu, Y. Liu, J.H. Ren, J. Wang, M.S. Cheng, Molecules 13 (2008) 1822–1839.
[9] W.M.F. Fabian, W. Stampfer, M. Mazur, G. Uray, Chirality 15 (2003) 271–275.
[10] A.H. Asikainen, J. Ruuskanen, K.A. Tuppurainen, SAR QSAR Environ. Res. 16 (2005) 555–565.
[11] J.J. Sutherland, J. Med. Chem. 47 (2004) 5541–5554.
[12] M.T. Makhija, J. Chem. Inf. Comput. Sci. 41 (2001) 1569–1577.
[13] M. Fort, Org. Biomol. Chem. 2 (2004) 3301–3311.
[14] D.X. Kong, T. Jiang, H.S. Guan, Chin. J. Chem. 23 (2005) 816–822.
[15] D.X. Kong, T. Jiang, H.S. Guan, Chem. J. Chin. Univ. 25 (2004) 713–716.
[16] C.X. Liao, J. Chem, Inf. Comput. Sci. 44 (2004) 230–238.
[17] A.P. Borosy, J. Mol. Struct. 729 (2005) 169–176.
[18] C.T. Supuran, A. Scozzafava, J. Conway (Eds.), Carbonic Anhydrase—Its Inhibitors and Activators, CRC Press, Boca Raton (FL), USA, 2004, pp. 1–363, and references cited therein.
[19] (a) C.T. Supuran, A. Scozzafava, Expert Opin. Ther. Pat. 10 (2000) 575;
(b) C.T. Supuran, Nat. Rev. Drug Discov. 7 (2008) 168–181.
[20] C.T. Supuran, A. Scozzafava, A. Casini, Med. Res. Rev. 23 (2003) 146.
[21] K.S. Smith, J.G. Ferry, FEMS Microbiol. Rev. 24 (2000) 335.
[22] C.T. Supuran, A. Scozzafava, Bioorg. Med. Chem. 15 (2007) 4336–4350.
[23] S. Pastorekova, S. Parkkila, J. Pastorek, C.T. Supuran, J. Enzyme Inhib. Med. Chem. 19 (2004) 199–229.
[24] N. Pfeiffer, Surv. Ophthalmol. 42 (1999) 137.
[25] M. Karelson, Molecular Descriptors in QSAR/QSPR, J. Wiley and Sons, New York, 2000.
[26] R. Tadeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.
[27] (a) B.W. Clare, C.T. Supuran, Eur. J. Med. Chem. 35 (2000) 859–865;
(b) B.W. Clare, C.T. Supuran, Eur. J. Med. Chem. 34 (1999) 463–474;
(c) C.T. Supuran, B.W. Clare, Eur. J. Med. Chem. 34 (1999) 41–50;
(d) C.T. Supuran, B.W. Clare, Eur. J. Med. Chem. 33 (1998) 489–500;
(e) B.W. Clare, C.T. Supuran, Eur. J. Med. Chem. 32 (1997) 311–319;
(f) C.T. Supuran, B.W. Clare, Eur. J. Med. Chem. 30 (1995) 687–696;
(g) E. Eroglu, H. Türkmen, J. Mol. Graph. Model. 26 (2007) 701–708;
(h) E. Eroglu, H. Türkmen, S. Güler, S. Palaz, O. Oltulu, Int. J. Mol. Sci. 8 (2007) 145–155.

[28] (a) V.K. Agrawal, J. Singh, P.V. Khadikar, C.T. Supuran, Bioorg. Med. Chem. Lett. 16 (2006) 2044–2051.(b) G. Melagraki, A. Afantitis, H. Sarimveis, O. Iglessi-Markopoulou, C.T. Supuran, Bioorg. Med. Chem. 14 (2006) 1108–1114;
(c) V.K. Agrawal, M. Banerji, M. Gupta, J. Singh, P.V. Khadikar, C.T. Supuran, Eur. J. Med. Chem. 40 (2005) 1002–1012;
(d) P.V. Khadikar, V. Sharma, S. Karmarkar, C.T. Supuran, Bioorg. Med. Chem. Lett. 15 (2005) 923–930;
(e) M. Jaiswal, P.V. Khadikar, C.T. Supuran, Bioorg. Med. Chem. Lett. 14 (2004) 5661–5666;
(f) V.K. Agrawal, S. Bano, C.T. Supuran, P.V. Khadikar, Eur. J. Med. Chem. 39 (2004) 593–600;
(g) M. Jaiswal, P.V. Khadikar, A. Scozzafava, C.T. Supuran, Bioorg. Med. Chem. Lett. 14 (2004) 3283–3290;
(h) A. Thakur, M. Thakur, P.V. Khadikar, C.T. Supuran, P. Sudele, Bioorg. Med. Chem. 12 (2004) 789–793;
(i) V.K. Agrawal, R. Sharma, P.V. Khadikar, Bioorg. Med. Chem. 10 (2002) 2993–2999;
(j) P.V. Khadikar, B.W. Clare, A.T. Balaban, C.T. Supuran, V.K. Agrawal, J. Singh, A.K. Joshi, M. Lakwani, Rev. Roum. Chim. 51 (2006) 703–717.
[29] H. Huang, X. Pan, N. Tan, N. Zeng, C. Ji, Eur. J. Med. Chem. 42 (2007) 365–372.
[30] L. Puccetti, G. Fasolis, D. Vullo, Z.H. Chohan, A. Scozzafava, C.T. Supuran, Bioorg. Med. Chem. Lett. 15 (2005) 3096–3101.
[31] J. Krungkrai, A. Scozzafava, R. Reungprapavut, S.R. Krungkrai, R. Rattanajak, S. Kamchonwongpaisand, C.T. Supuran, Bioorg. Med. Chem. 13 (2005) 483–489.
[32] C.T. Supuran, A. Scozzafava, A. Popescu, R. Bobes-Tureac, A. Banciu, G. Bobes-Tureac, M.D. Banciu, Eur. J. Med. Chem. 32 (1997) 445–452.
[33] D. Vullo, M. Franchi, E. Gallori, J. Pastorek, A. Scozzafava, S. Pastorekovac, C.T. Supurana, Bioorg. Med. Chem. Lett. 13 (2003) 1005.
[34] H. Turkmen, M. Durgun, S. Yilmaztekin, M. Emul, A. Innocenti, D. Vullo, A. Scozzafava, C.T. Supuran, Bioorg. Med. Chem. Lett. 15 (2005) 367.
[35] Gaussian 03, Revision B.05, Gaussian, Inc., Pittsburgh, PA, 2003.
[36] R.G. Parr, W. Yang, Density Functional Theory of Atoms and Molecules, Oxford University Press, New York, 1989.
[37] A.D. Becke, J. Chem. Phys. 98 (1993) 5648.
[38] CODESSA PRO software version 2.7.2, University of Florida (1994–95) and Semichem Inc. (1995–2004).
[39] CODESSA™, References Manual, V. 2.13 (PC) Semichem and the University of Florida (1995–1997)
[40] A. Tropsha, P. Gramatica, V.K. Gombar, QSAR Comb. Sci. 22 (2003) 69–77.
[41] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269–276.
[42] S. Wold, L. Eriksson, Statistical validation of QSAR results, in: H. van der Waterbeem (Ed.), Chemometrics Methods in Molecular Design (1995), pp. 309–318 VCH, Weinheim (Germany).
[43] J.G. Topliss, R.P. Edwards, J. Med. Chem. 22 (1979) 1238–1244.