Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 1360–1365
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

Radial Distribution Function descriptors for predicting affinity for vitamin D receptor

Maykel Pérez González [a,*], Zoila Gándara [b], Yagamare Fall [b], Generosa Gómez [b]

[a] *Molecular Simulation and Drug Design Group, Chemical Bioactive Center, Central University of Las Villas, Autopista Nacional Km 246, Santa Clara, Villa Clara C.P. 54830, Cuba*
[b] *Department of Organic Chemistry, Vigo University, C.P. 36200 Vigo, Spain*

Received 14 August 2007; received in revised form 12 October 2007; accepted 15 October 2007
Available online 23 October 2007

Abstract

The QSAR is an alternative method for the research of new and better Vitamin D analogues with affinity for the VDR receptor. This paper describes the results of applying the Radial Distribution Function (RDF descriptors) approach for predicting the VDR affinity of 38 vitamin D analogues. The model described 80% of the experimental variance, with a standard deviation of 0.35. Leave-one-out, bootstrapping and external set validation were carried out with the aim of evaluating the predictive power of the model. The values of their respective squared correlations coefficients were 0.72, 0.70 and 0.79. The RDF approach was compared with four other predictive models, but none of these could explain more than 71.0% of the variance with six variables in their respective models.

Keywords: VDR affinity; RDF descriptors; 3D Descriptors; Vitamin D; QSAR

1. Introduction

The active vitamin D_3, 1α,25-dihydroxyvitamin D_3, is a multi-functional hormone. In addition to its central action in regulating calcium and phosphorus homeostasis, this hormone plays a role in controlling growth and cell differentiation and in immunomodulation [1,2]. 1α,25-Dihydroxyvitamin D_3 has been widely used as a potential therapeutic agent in the treatment of bone metabolic diseases and the skin disorder psoriasis [3,4]. However, the use of this vitamin is limited because a high dose causes critical hypercalcemia. Vitamin D analogues which have a selective activity profile, such as a high cell differentiating potential with little calcemic activity, have been developed by structural modification of the parent 1α,25-dihydroxyvitamin D_3. However, a limited number of studies have been devoted to the search for new potent vitamin D analogues using Quantitative Structure Activity Relationship (QSAR) [5,6].

QSARs have been broadly used for some years in medical research [7–9]. This methodology makes use of the molecular descriptors offering valuable and simple information about the structure of the molecules which is used later in the elaboration of the predictive models. The use of this methodology allows cost savings by reducing the laboratory resources needed, and the time required to create and investigate new drugs with certain desired biological activity [10].

For these reasons, our aim is to continue developing useful QSAR models for predicting the VDR affinity for different Vitamin D analogues in order to design new compounds with better pharmacological and biological profiles.

2. Materials and methods

2.1. Data set

In this study, we have selected a data set of 38 analogues of 1α,25-dihydroxyvitamin D_3 whose affinity for VDR was reported [11–14]. One of the most important steps in computer-aided drug design is to have a representative and

* Corresponding author.
E-mail address: mpgonzalez76@yahoo.es (M.P. González).

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2007.10.020

randomized training and predicting series. With this aim, we selected training (31 compounds) and test set (7 compounds) using K-means cluster analysis [15–18]. The application of this statistical technique guarantee that the test set be representative of the training set. In other words all representative compound of the test set in the multidimensional descriptor space must be close to those of the training set.

The relative potency of the analogues was calculated from their concentration needed to displace 50% of [^{3}H]-1α,25-dihydroxyvitamin D_3 from its receptor VDR in porcine intestinal compared with the activity of 1α,25-dihydroxyvitamin D_3 (assigned a 100% value). The experimental values of this biological activity and the structures of these analogues are provided as supporting information.

2.2. Radial Distribution Function descriptors

The 3D coordinates of the atoms of molecules can be transformed into a structure code that has a fixed number of descriptors irrespective of the size of a molecule. This task is performed by a structure coding technique referred to as Radial Distribution Function code (RDF code) [19,20]. In general, there are some prerequisites for a structure code:

- independent from the number of atoms, i.e., the size of a molecule,
- unambiguity regarding the three-dimensional arrangement of the atoms, and
- invariance against translation and rotation of the entire molecule.

Formally, the Radial Distribution Function of an ensemble of N atoms can be interpreted as the probability distribution to find an atom in a spherical volume of radius r [21]. The equation represents the Radial Distribution Function code as it is used in this investigation:

$$g(r) = f \times \sum_{i}^{N-1} \sum_{j}^{N} A_i A_j e^{-B\left(r-r_{ij}\right)^2}$$

where f is a scaling factor and N is the number of atoms. By including characteristic atomic properties A of the atoms i and j, the RDF codes can be used in different tasks to fit the requirements of the information to be represented. The exponential term contains the distance r_{ij} between the atoms i and j and the smoothing parameter B, which defines the probability distribution of the individual distances. $g(r)$ was calculated at a number of discrete points with defined intervals.

The atomic properties A_i and A_j used in this equation (see supporting information) enable the discrimination of the atoms of a molecule for almost any property that can be attributed to an atom. Such distribution function provides, besides information about interatomic distances in the whole molecule, the opportunity to gain access to other valuable information, e.g., bond distance, ring types, planar and non-planar systems and atom types. This fact is a most valuable consideration for a computer-assisted code elucidation. The Radial Distribution Function in this form meets the entire requirement mentioned above, especially invariance against linear translations.

2.3. Computational strategies

The structures of the compounds are firstly pre-optimized with the Molecular Mechanics Force Field (MM+) procedure and after that the geometries are further refined by means of the semi-empirical method AM1 [22] included in MOPAC 6.0 [23] using the algorithm Polak–Ribiere and a gradient norm limit of 0.01 kcal/Å.

The DRAGON [24] computer software was employed to calculate the Constitutional [25], Geometrical [25], Weighted Holistic Invariant Molecular Descriptors (WHIM) [26,27], GEometry, Topology, and Atom-Weights AssemblY (GETAWAY) [28,29] and RDF [20,21] descriptors used in this study. Descriptors with constant or near constant values inside each group of descriptors were discarded.

The mathematical models were obtained by means of the Multiple linear Regression Analysis (MRA) implemented in the STATISTICA software version 6.0. The Genetic Algorithm (GA) was used as the variable selection strategy. The GA simulation conditions were 1000 generations, number of crossovers = 5000, smoothness factor = 1, mutation probability for adding new term = 50% and 300 models populations. Analysis of residuals and deleted residuals from the regression equations was used to identify outliers. The statistical significance of the models was determined by examining the squared regression coefficient (R^2), the standard deviation (S), the number of variables, the Fisher ratio (F), the Akaike information criterion (AIC) [30,31] and Kubinyi function (FIT) [32,33]. Also we developed a Randic orthogonalization process of molecular descriptors in all models to avoid the collinearity among variables [34–38]. The main philosophy of this approach is to avoid the exclusion of descriptors on the basis of their collinearity with other variables previously included in the model. In our view, the collinearity of the variables should be as low as possible because the interrelatedness among the different descriptors can result in a highly unstable regression coefficient, which makes it impossible to know the relative importance of an index and underestimates the utility of the regression coefficient model [5,18,39–41].

2.4. Validation of the models

The robustness of the models and their predictivity were evaluated by both q^2 Leave-One-Out" (LOO) cross-validation, bootstrap and a test set that was selected using K-means cluster analysis [15–18]. In addition the proposed models were also checked for reliability by permutation testing: new models are recalculated for randomly reordered response (Y scrambling). Evidence that the proposed model is well founded, and not just the result of chance correlation, is provided by obtaining new models on the set with randomized responses that have significantly lower R^2 and q^2 than those of the original model.

Finally, the RDF descriptors model for the prediction of the affinity for the VDR, explained in the previous section, was

compared with the rest of the methodologies. The development of the other models involved the use of the same data set that we reported in the previous section. The comparison was based on the quality of the statistical parameters of the regression as well as the predictive capability of the models generated.

3. Results and discussion

The preliminary models that we found taking into account the descriptors derived by the RDF approach were the following:

$$-\log(\mathrm{VDR}) = 0.2666 \times \mathrm{RDF140m} - 0.236 \times \mathrm{RDF055m} + 0.119 \times \mathrm{RDF095e} + 0.095 \times \mathrm{RDF035m} - 0.268 \times \mathrm{RDF095m} - 1.547 \quad (1)$$

$N = 31$, $R^2 = 0.70$, $S = 0.49$, $F(5, 25) = 11.54$, $p < 10^{-5}$, AIC = 0.36, FIT = 1.03, $Q^2_{\mathrm{LOO}} = 0.56$, $S_{\mathrm{LOO}} = 0.54$, $q^2_{\mathrm{BOOT}} = 0.50$, $a(R^2) = 0.12$, $a(Q^2) = -0.37$, and $R^2_{\mathrm{ExT}} = 0.33$.

$$-\log(\mathrm{VDR}) = 0.337 \times \mathrm{RDF140m} - 0.205 \times \mathrm{RDF055m} + 0.160 \times \mathrm{RDF095e} + 0.427 \times \mathrm{RDF035m} - 0.241 \times \mathrm{RDF035v} - 0.311 \times \mathrm{RDF095m} - 1.648 \quad (2)$$

$N = 31$, $R^2 = 0.78$, $S = 0.42$, $F(6, 24) = 14.42$, $p < 10^{-5}$, AIC = 0.29, FIT = 1.29, $Q^2_{\mathrm{LOO}} = 0.67$, $S_{\mathrm{LOO}} = 0.57$, $q^2_{\mathrm{BOOT}} = 0.65$, $a(R^2) = 0.14$, $a(Q^2) = -0.45$, and $R^2_{\mathrm{ExT}} = 0.66$.

We selected the model subjected to the principle of parsimony [42]. Then, we chose a function with higher statistical signification but having as few parameters as possible. For that reason, although several models were developed for the RDF, changing the number of variables in every step of the analysis the best model obtained was the model with six variables.

This affirmation is supported by the statistical and validation parameters of the models obtained. In this connection, another criterion that we used is the ratio between the number of cases and variables included in the model. It has been reported that this relation is appropriate when there are five cases per variable (5:1) has a minimum value [43]. As our training set consisted of 31 compounds, the maximum number of variables is six due to the seven variables in the model could be considered as overfit. We also applied the AIC and FIT to determine if a variable should be included in the model. That is to say, if the Akaike's information criterion decreases in value when adding an additional variable and the Kubinyi function increases in value, then, the introduction of this new variable is justified.

The quality of the statistical parameters for the model with six variables was adequate. The introduction of new variable improved the AIC (decrement from 0.36 to 0.29) and the FIT (increased from 1.03 to 1.29). The statistical parameters in this model improved with increases in the value of the R^2, and F, and a decrease in the value of S. Finally, the validation parameters of model (2) are superior to model (1).

In spite of achieving adequate values of statistical parameters, we thought that it was not enough to say that our model was appropriate. Therefore, we carried out a comparison with other methodologies to demonstrate the superiority of our model. The results obtained from this comparison are given in Tables 1 and 2. The meaning of the descriptors used in these models is provided as supporting information.

As we can see in a general way, the model proposed using the RDF descriptors shows a better statistical significance. That is to say, the RDF ones explain more of the variance of the data than the other ones using the same number of descriptors in the equation (six variables). The value of the squared regression coefficient for the RDF approach is 0.78 while the other methodologies have values never higher than 0.71.

The RDF also presents the best validation parameters, regarding the better predictive power. The parameter q^2, equivalent statistically to R^2, shows its greater value in this methodology for the two types of cross-validation LOO and Bootstrapping (0.67 and 0.65, respectively). The rest of the methodologies have lower values of q^2 than the RDF for both cross-validation techniques. Finally, Eq. (2) has a better prediction of the test set with a $R^2_{\mathrm{ExT}} = 0.66$ than the others methodologies. For that reasons, we demonstrate once again the superiority of the RDF methodology.

Once the superiority of our model was demonstrated for predicting this biological activity, we will analyze some characteristics of this one with the aim of improving its statistical parameters.

The collinearity among the variables is an undesirable effect because it can result in a highly unstable regression coefficient of these ones, which makes it impossible to know the relative importance of an index and underestimates the utility of the regression coefficient model [5,18,39–41].

Table 1
The statistical parameters of the linear regression models obtained for the five kinds of descriptors involved in the comparison

Descriptors	Variables	R^2	S	F	P	AIC	FIT
RDF	RDF035m, RDF055m, RDF095m, RDF140m, RDF035v, RDF095e	0.78	0.42	14.42	$<10^{-5}$	0.29	1.29
Constitutional	Ss, Ms, nBM, nCIR, RBF, nH	0.34	0.74	2.03	$<10^{-3}$	0.87	0.18
Geometrical	H3D, SPAN, SPAM, ASP, L/Bw, SEig	0.39	0.71	2.57	$<10^{-4}$	0.80	0.23
WHIM	G2u, E3e, P1p, G1p, Kp, Ds	0.71	0.49	9.56	$<10^{-5}$	0.39	0.85
GETAWAY	HGM, H8v, H8p, RARS, R4m$^+$, R6e	0.53	0.62	4.59	$<10^{-5}$	0.61	0.41

Table 2
The validation parameters of the linear regression models obtained for the five kinds of descriptors involved in the comparison

Descriptors	Variables	$q^2_{\text{CV-LOO}}$	$S_{\text{CV-LOO}}$	q^2_{BOOT}	R^2_{EXT}
RDF	RDF035m, RDF055m, RDF095m, RDF140m, RDF035v, RDF095e	0.67	0.57	0.65	0.66
Constitutional	Ss, Ms, nBM, nCIR, RBF, nH	0.06	1.65	0.01	0.01
Geometrical	H3D, SPAN, SPAM, ASP, L/Bw, SEig	−0.33	1.86	0.01	0.05
WHIM	G2u, E3e, P1p, G1p, Kp, Ds	0.53	0.92	0.45	0.01
GETAWAY	HGM, H8v, H8p, RARS, R4m$^+$, R6e	0.23	1.17	0.13	0.01

The QSAR optimal model obtained with the RDF descriptors (Eq. (2)) after orthogonalization and standardization is given below, together with the statistical parameters of regression analysis.

$$-\log(\text{VDR}) = 1.119 - 0.181 \times {}^{1}\Omega\text{RDF055m} + 0.128 \times {}^{2}\Omega\text{RDF140m} - 0.171 \times {}^{5}\Omega\text{RDF035v} - 0.315 \times {}^{6}\Omega\text{RDF095m} \quad (3)$$

$N = 31$, $R^2 = 0.75$, $S = 0.44$, $F(4, 26) = 19.73$, $p < 10^{-5}$, AIC = 0.26, FIT = 1.67, $Q^2_{\text{LOO}} = 0.66$, $S_{\text{LOO}} = 0.59$, $q^2_{\text{BOOT}} = 0.64$, $a(R^2) = 0.08$, $a(Q^2) = -0.31$, and $R^2_{\text{EXT}} = 0.73$.

As can be seen, variables ${}^{3}\Omega$RDF095e and ${}^{4}\Omega$RDF035m were found to be not statistically significant: they are highly correlated with the other variables and, at the same time, they have the lowest contribution to the regression coefficients in the third equation. Further, the significance of adding ${}^{3}\Omega$RDF095e and ${}^{4}\Omega$RDF035m into the model remains unclear as seen from the modest improvements in R^2.

On the other hand, the presence of outliers in QSAR models can become a serious problem due to the fact that the model is unable to predict its "real" biological activity. In this connection, we looked for the presence of outliers in Eq. (3), where one outlier, compound **16**, was found because this compound included in the training set had a standard residual value higher than 2δ, where δ is equivalent to the standard deviation and a relative high deleted residual, as can be seen in the supporting information.

The new equation obtained with their statistical parameters appears below:

$$-\log(\text{VDR}) = 0.864 - 0.160 \times {}^{1}\Omega\text{RDF055m} + 0.096 \times {}^{2}\Omega\text{RDF140m} - 0.170 \times {}^{5}\Omega\text{RDF035v} - 0.311 \times {}^{6}\Omega\text{RDF095m} \quad (4)$$

$N = 30$, $R^2 = 0.80$, $S = 0.35$, $F(4, 25) = 24.91$, $p < 10^{-5}$, AIC = 0.17, FIT = 2.16, $Q^2_{\text{LOO}} = 0.72$, $S_{\text{LOO}} = 0.54$, $q^2_{\text{BOOT}} = 0.70$, $a(R^2) = 0.07$, $a(Q^2) = -0.38$, and $R^2_{\text{EXT}} = 0.79$, where ${}^{1}\Omega$RDF055m, ${}^{6}\Omega$RDF095m, and ${}^{2}\Omega$RDF140m are the Radial Distribution Function weighted by atomic masses at 5.5, 9.5 and 14.0 Å, respectively. In addition ${}^{5}\Omega$RDF035v is the Radial Distribution Function weighted by atomic van der Waals volumes at 3.5 Å.

As can be seen, the statistical parameters of the equation improved significantly from a statistical point of view, in comparison with the reported for Eq. (3). In addition, the results of the LOO and bootstrapping validations demonstrated that the results of the model were greatly improved. An additional and conclusive probe that the new model is better from a statistical point of view is the improvement of the AIC = 0.17 and the FIT = 2.16 values. Therefore, this is new evidence that compound **16** should be considered as potential outlier.

Although the biological phenomena are complex by nature, in this work the VDR affinity of a set of 38 vitamin D analogues was successfully modeled through multilinear regression analysis using RDF descriptors and Genetic algorithm as method of variables selection. However, none of the other four kinds of descriptors demonstrated potentialities for modeling this biological property from a statistical point of view. The results obtained here strongly suggest that the main features controlling the VDR affinity are the molecular size and shape of the Vitamin D analogues.

Interpreting a QSAR model in terms of the specific contribution of substituents and other molecular features to the modeled activity is always a difficult work. In this paper, RDF descriptors in model (4) suggest the occurrence of some linear dependence between the affinity for VDR, the Vitamin D analogues and the 3D molecular distribution of mass and van der

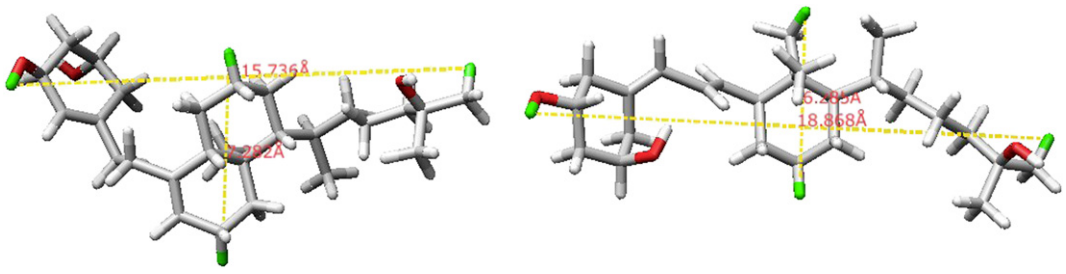

Fig. 1. Distances between different critical atoms that represent the relative values of mass and shape distributions at 3.5, 5.5, 9.5 and 14.0 Å according to the calculated RDF descriptors included in model (4).

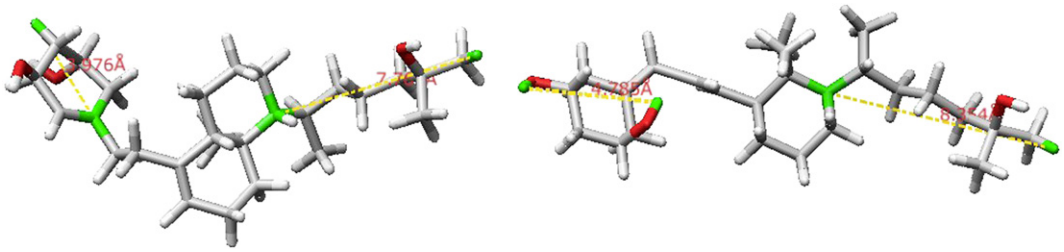

Fig. 2. Distances between different critical atoms that represent the relative values of mass and shape distributions at 3.5, 5.5, 9.5 and 14.0 Å according to the calculated RDF descriptors included in model (4).

Waals volumes calculated at radius ranging from 3.5 to 14.0 Å from the specific geometrical centers of each molecule.

The model was able to establish a reliable linear dependence between descriptors just encoding the size and shape of the studied molecules and their VDR affinity. This fact strongly suggests that the main features controlling the VDR affinity are the molecular size and shape of the Vitamin D analogues of the whole molecule or a specific substituent or the interactions among different sub-structure in the entire molecule. Clear examples of this are shown in Figs. 1 and 2.

As can be seen, in these figures our model explains three important sub-structures that are responsible for the biological activity of these molecules. These vitamin analogues are the best biological activities in the set used in this study and their sizes are about 15.73 and 18.87 Å, respectively. These distances are greater than the distance codified by ^{2}ΩRDF140m descriptor in the equation which means that all the sub-structures of the whole molecule are interrelated. For that reason we think that a little modification in any of these sub-structures may have large effects in the contribution to the biological property of the others ones. On the other hand, there exist three parts well defined in these molecules that play an important role in the explanation of their biological activities and that should be the starting point for potential studies, such as the A, B rings and the lateral chain of these analogues according to model (4) (^{5}ΩRDF035v; ^{1}ΩRDF055m and ^{6}ΩRDF095m). These descriptors codified these steric and shape factors at a radius more or less coincident with the length of the side chain, the ring A and B in Vitamin D analogues (see Figs. 1 and 2).

In the last 10 years, different protocols have been developed for the measurement of the affinity of Vitamin D analogues for the VDR. One of the principal tissues used in these protocols is porcine intestinal VDR. This means that the biological activities obtained with the use of different VDR tissues may not be optimal for QSAR models when included in the training set. Every QSAR investigation is based on an assumption of certain homogeneity. This implies similarity in the biological mode of action and measurement of biological activity of all the investigated compounds. The infringement of this rule could without doubt lead to bad interpretation of the models and the action mechanism of the compounds in an established target. For that reason, QSAR models should be realized for each VDR type in order to facilitate the research in this topic. They can without doubt be useful in the design of new Vitamin D analogues with better biological and pharmacokinetic profiles.

Acknowledgements

This work was supported by a grant from the Xunta de Galicia (PGIDIT04BTF301031PR). We also acknowledge the Universidad de Vigo and the Cuban Higher Education Ministry (R&D project number 6.181-2006).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejmech.2007.10.020.

References

[1] C.F. Garland, F.C. Garland, E.D. Gorham, M. Lipkin, H. Newmark, S.B. Mohr, M.F. Holick, Am. J. Public Health 96 (2006) 252–261.
[2] S. Vijayakumar, R.R. Mehta, P.S. Boerner, S. Packianathan, R.G. Mehta, Cancer J. 11 (2005) 362–373.
[3] K. Fogh, K. Kragballe, Curr. Drug Targets Inflamm. Allergy 3 (2004) 199–204.
[4] J. Koo, Cutis 70 (2002) 21–24.
[5] M.P. González, P.L. Suarez, Y. Fall, G. Gomez, Bioorg. Med. Chem. Lett. 15 (2005) 5165–5169.
[6] M.P. González, M. Puente, Y. Fall, G. Gomez, Steroids 71 (2006) 510–527.
[7] M.P. González, C. Teran, M. Teijeira, Bioorg. Med. Chem. Lett. 16 (2006) 1291–1296.
[8] M.P. González, C. Teran, M. Teijeira, A.H. Morales, Eur. J. Med. Chem. 41 (2006) 56–62.
[9] M.P. González, J. Caballero, A. Tundidor-Camba, A.H. Morales, M. Fernandez, Bioorg. Med. Chem. 14 (2006) 200–213.
[10] A.H. Morales, M.A. Cabrera, R.D. Combes, M.P. González, Curr. Comput-Aided Drug Des. 1 (2005) 237–255.
[11] R.R. Sicinski, H.F. DeLuca, Bioorg. Med. Chem. 7 (1999) 2877–2889.
[12] P.J. De Clercq, I. Murad, L.J. Gao, Y.J. Chen, D. Van Haver, M. Vandewalle, A. Verstuyf, L. Verlinden, C. Verboven, R. Bouillon, J. Steroid Biochem. Mol. Biol. 89–90 (2004) 61–66.
[13] S. Yamada, K. Yamamoto, H. Masuno, M. Choi, Steroids 66 (2001) 177–187.
[14] Y.J. Chen, L.J. Gao, I. Murad, A. Verstuyf, L. Verlinden, C. Verboven, R. Bouillon, D. Viterbo, M. Milanesio, D. Van Haver, et al., Org. Biomol. Chem. 1 (2003) 257–267.
[15] M.P. González, A.M. Helguera, M.A. Cabrera, Bioorg. Med. Chem. 13 (2005) 1775–1781.
[16] M.P. González, L.C. Dias, A.M. Helguera, Y.M. Rodriguez, L.G. de Oliveira, L.T. Gomez, H.G. Diaz, Bioorg. Med. Chem. 12 (2004) 4467–4475.
[17] E. Molina, H.G. Diaz, M.P. González, E. Rodriguez, E. Uriarte, J. Chem. Inf. Comput. Sci. 44 (2004) 515–521.
[18] M.P. González, H. Gonzalez Diaz, R. Molina Ruiz, M.A. Cabrera, R. Ramos de Armas, J. Chem. Inf. Comput. Sci. 43 (2003) 1192–1199.

[19] J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, J. Chem. Inf. Comput. Sci. 36 (1996) 1030–1037.
[20] J. Gasteiger, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, Fresenius' J. Anal. Chem. 359 (1997) 50–55.
[21] M.C. Hemmer, V. Steinhauer, J. Gasteiger, Vib. Spectrosc. 19 (1999) 151–164.
[22] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902–3909.
[23] J. Frank, MOPAC (edn version 6.0), in: Seiler Research Laboratory (Ed.), US Air Force Academy, Colorado Springs CO, 1993.
[24] R. Todeschini, V. Consonni, M. Pavan, Dragon Software, ed. version 2.1, 2002.
[25] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors (vol. 1), first ed. Wiley–VCH, Mannheim, 2000.
[26] R. Todeschini, P. Gramatica, R. Provenzani, E. Marengo, Chemom. Intell. Lab. Syst. 27 (1995) 221–229.
[27] R. Todeschini, M. Lasagni, E. Marengo, J. Chemom. 8 (1994) 263–273.
[28] V. Consonni, R. Todeschini, M. Pavan, J. Chem. Inf. Comput. Sci. 42 (2002) 682–692.
[29] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, J. Chem. Inf. Comput. Sci. 42 (2002) 693–705.
[30] H. Akaike, IEEE Trans. Automat. Control AC-19 (1974) 716–723.
[31] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: B.N. Petrov, F. Csaki (Eds.), Second International Symposium on Information Theory, Akademiai Kiado, Budapest, 1973, pp. 267–281.
[32] H. Kubinyi, Quant. Struct. Act. Relat. 13 (1994) 393–401.
[33] H. Kubinyi, Quant. Struct. Act. Relat. 13 (1994) 285–294.
[34] D.J. Klein, M. Randić, D. Babić, B. Lučić, S. Nikolić, N. Trinajstić, Int. J. Quant. Chem. 63 (1991) 215–222.
[35] B. Lučić, S. Nikolić, N. Trinajstić, D. Jurić, J. Chem. Inf. Comput. Sci. 35 (1995) 532–538.
[36] M. Randić, J. Mol. Struct. (Teochem) 233 (1991) 45–59.
[37] M. Randić, New J. Chem. 15 (1991) 517–525.
[38] M. Randić, J. Chem. Inf. Comput. Sci. 31 (1991) 311–320.
[39] M.P. González, A.M. Helguera, R. Medina, R.M. Ruiz, Internet Electron. J. Mol. Des. 3 (2004) 200–208.
[40] M.P. González, C. Teran, Y. Fall, M. Teijeira, P. Besada, Bioorg. Med. Chem. 13 (2005) 601–608.
[41] M.P. González, C. Teran, M. Teijeira, M.J. Gonzalez-Moa, Eur. J. Med. Chem. 40 (2005) 1080–1086.
[42] D.M. Hawkins, J. Chem. Inf. Comput. Sci. 44 (2004) 1–12.
[43] J.G. Topliss, R.P. Edwards, J. Med. Chem. 22 (1979) 1238–1244.