Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 5165–5169
Bioorganic & Medicinal Chemistry Letters

Quantitative structure–activity relationship studies of vitamin D receptor affinity for analogues of 1α,25-dihydroxyvitamin D_3. 1: WHIM descriptors

Maykel Pérez González,[a,b,c,*] Pedro Lois Suárez,[a] Yagamare Fall[a] and Generosa Gómez[a]

[a]*Department of Organic Chemistry, Vigo University, C.P. 36200, Vigo, Spain*
[b]*Chemical Bioactive Center, Central University of Las Villas, Santa Clara, Villa Clara, C.P. 54830, Cuba*
[c]*Service Unit, Experimental Sugar Cane Station 'Villa Clara-Cienfuegos', Ranchuelo, Villa Clara, C.P. 53100, Cuba*

Received 14 July 2005; revised 10 August 2005; accepted 22 August 2005
Available online 3 October 2005

Abstract—The weighted holistic invariant molecular (WHIM) approach has been applied to the study of the VDR affinity of 86 vitamin D analogues. A model able to describe more than 71% of the variance in the experimental activity was developed with the use of the mentioned approach. In contrast, none of three different approaches, including the use of BCUT, Galvez topological charge indices, and 2D autocorrelations descriptors, was able to explain more than 38% of the variance in the mentioned property, even with more variables in the equation.

1α,25-Dihydroxyvitamin D_3 (calcitriol) is the hormonally active form of vitamin D_3. Besides regulating calcium homeostasis, this multifunctional hormone is also involved in other cellular processes, including cell differentiation, immune system regulation, and gene transcription.[1,2] These biological activities are mediated by the vitamin D receptor (VDR),[3] which is a member of the nuclear receptor superfamily.[4] However, the clinical utility of calcitriol for treatment of cancers and skin disorders is limited by its hypercalcemic effects. There is accordingly much interest in the design and synthesis of vitamin D analogues with more selective biological effects. To date, thousands of vitamin D analogues have been synthesized and their biological activities evaluated.[5] Many of them have already been developed, and some are under development as clinical agents for treating metabolic bone diseases, skin diseases such as psoriasis, immune disorders, or malignant tumors.[6]

Most of the analogues of calcitriol synthesized so far show modifications at the side chain and conformational analysis has been carried out in order to predict a potentially active side-chain structure.[7,8]

However, a limited number of studies have been devoted to the search for new potent vitamin D analogues using quantitative structure activity relationship (QSAR).

Our research group has demonstrated that QSAR studies are a powerful method for the design of bioactive compounds and the prediction of activity according to physical and chemical properties.[9–20] In this paper, we describe for the first time the QSAR between the vitamin D receptor and some analogues of 1α,25-dihydroxyvitamin D_3 using the weighted holistic invariant molecular (WHIM) descriptors. This molecular descriptor compares favorably with the other three kinds of molecular descriptors and proved to be superior in modeling the biological activity under study.

The WHIM descriptors are built in such a way as to capture the relevant molecular 3-D information regarding the molecular size, shape, symmetry, and atom distribution with respect to some invariant reference frame. The algorithm consists of performing a principal component analysis (PCA) on the centered molecular coordinates by using four different weighting schemes. The four weighting schemes are: (1) the unweighted case *U*

Keywords: QSAR; VDR; WHIM descriptors; Vitamin D; Dragon software.
* Corresponding author. E-mail: mpgonzalez76@yahoo.es

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.08.065

($w_i = 1$, $i = l,n$, where n is the number of atoms for each compound), (2) the atomic masses M ($w_i = m_i$), (3) the van der Waals volume V ($w_i = vdw_i$), and (4) the Sanderson atomic electronegativity E ($w_i = e\ln_i$). For each weighting scheme, a set of statistical indices is calculated on the atoms projected onto each principal component t_m ($m = 1,2,3$).[21]

In this study, we have selected a data set of 86 analogues of 1α,25-dihydroxyvitamin D_3 whose affinity for VDR was reported.[22–35] 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 bovine thymus 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.

Before calculating the molecular descriptors, we carried out geometry optimization calculations for each compound of this study using the quantum chemical semiempirical method AM1[36] included in Mopac 6.0 computer software.[37] Dragon[38] computer software was employed to calculate the molecular descriptors. Four kinds of molecular descriptors including different groups: BCUT, Galvez topological charge indices, 2D autocorrelations, and WHIM descriptors[21] were calculated for this set of compounds. Descriptors with constant or near constant values inside each group of descriptors were discarded. The statistical processing to obtain the QSAR models was carried out using regression multivariate technique employing the forward stepwise method for the selection of variables.[39] Analysis of residuals and deleted residuals from the regression equations was used to identify outliers. The statistical significance of the models was determined examining the regression coefficient, the standard deviation, the number of variables, and the proportion between the cases and variables in the equation. All statistical analyses and data exploration were carried out using the STATISTIC 6.0 software.[40] The models obtained were validated by calculating q^2 values. The q^2 values are calculated from 'leave-one-out' (LOO) and 'leave-group-out' (LGO) test known as cross-validation; in this last case, the 25% of the data was selected for the analysis. The q^2 values can be considered a measure of the predictive power of a regression equation, whereas R^2 can always be increased artificially by adding more parameters (descriptors); q^2 decreases if a model is overparameterized and is therefore a more meaningful summary statistic for QSAR models.

In this sense, the best QSAR model obtained with the WHIM descriptors is given below together with the statistical parameters of the regression:

$$\begin{aligned} -\log(\mathrm{VDR}) = {} & -3.648 - 7.853 \cdot L2m + 36.997 \\ & \cdot P1m - 8.645 \cdot L1v + 25.155 \\ & \cdot E1v - 24.251 \cdot E1e + 3.245 \\ & \cdot P2s + 7.603 \cdot Tu + 70.642 \cdot Dm \end{aligned} \quad (1)$$

$$\begin{aligned} & N = 86 \quad R = 0.857 \quad S = 0.734 \quad F_{\mathrm{exp}} = 26.72 \\ & p < 10^{-5} \quad q^2_{\mathrm{LOO}} = 0.675 \quad S_{\mathrm{LOO}} = 0.813 \\ & q^2_{\mathrm{LGO}} = 0.653 \quad S_{\mathrm{LGO}} = 0.849, \end{aligned}$$

where N is the number of compounds used, R is the correlation coefficient, S is the standard deviation of the regression, F_{exp} is the Fisher ratio at the 95% confidence level, p is the significance of the variables in the model, q^2_{LOO}, S_{LOO} and q^2_{LGO}, S_{LGO} are the square of the correlation coefficient and the standard deviation of the leave-one-out and leave-group-out cross-validation, respectively. However, after an analysis of the correlation between the variables in the equation, one can conclude that there is collinearity among the descriptors, as can be observed in the following Table 1.

For that reason, an orthogonalization process is necessary to eliminate this undesirable effect. Here, we used the Randić orthogonalization method to correct this problem.[41–43] The main philosophy of this approach is to avoid the exclusion of descriptors on the basis of its collinearity with other variables previously included in the model. The acceptable level of collinearity to avoid is a more subjective issue. In the view of the authors, 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.

The QSAR model obtained with the WHIM descriptors after the orthogonalization process is given below, together with the statistical parameters:

Table 1. Correlation coefficients among the eight most significant variables

	L2m	*P1m*	*L1v*	*E1v*	*E1e*	*P2s*	*Tu*	*Dm*
L2m	1.00	−0.89	0.04	−0.31	−0.38	0.94	0.44	0.23
P1m		1.00	0.38	0.19	0.33	−0.97	−0.03	−0.23
L1v			1.00	−0.19	−0.10	−0.26	0.91	−0.19
E1v				1.00	0.90	−0.19	−0.33	0.52
E1e					1.00	−0.30	−0.27	0.58
P2s						1.00	0.15	0.28
Tu							1.00	−0.12
Dm								1.00

Marked correlations are significant at $p < .05000$.

$$-\log(\mathrm{VDR}) = -1.032 - 0.358 \cdot {}^{1}\Omega L2m - 0.714 \cdot {}^{2}\Omega E1e - 0.273 \cdot {}^{3}\Omega L1v + 0.631 \cdot {}^{4}\Omega E1v + 0.459 \cdot {}^{5}\Omega Dm \quad (2)$$

$$N = 86 \quad R = 0.847 \quad S = 0.744 \quad F_{\mathrm{exp}} = 40.537$$
$$p < 10^{-5} \quad q^2 = 0.679 \quad S_{\mathrm{cv}} = 0.793$$
$$q^2_{\mathrm{LGO}} = 0.652 \quad S_{\mathrm{LGO}} = 0.841$$

As a result of the orthogonalization process of variables of model 1, three variables resulted in being statistically not significant, and these were excluded in the final model (see Eqs. 1 and 2). These variables are *P*1*m*, *P*2*s*, and *Tu*. The *P*1*m* is strongly correlated with *L*2*m* containing 89.12% of duplication. Also *P*2*s* is collinear with *L*2*m* having a coefficient of correlation of 94.34. The *L*2*m* was the variable taken as the first orthogonal descriptor, and the rest of the descriptors were orthogonalized with respect to it; therefore, those variables that have elevated correlation with it will have a major probability of being excluded from the model.

Similar to the previous case, the *Tu* descriptor is highly correlated with *L*1*v* (correlation coefficients of 91.61). The *Tu* is a total size unweighted index; much related to the index *L*1*v* that is the first component size directional WHIM index weighted by atomic van der Waals volumes. This is probably the main cause for the lack of *Tu* in the description of the VDR affinity.

The variables in the model (Eq. 2) encoded specific structure information such as atomic distribution (*E*1*v*, *E*1*e*, and *Dm*) and size (*L*1*v* and *L*2*m*) of the whole molecular structure of the vitamin D analogues present in this study.

This structure information could be very important in the design of new ligands and had been proved previously in the X-ray crystal structure of VDR complex with the vitamin D indicated, in addition to the cavity in which the flexible C17 side chain moiety of this compound is accommodated, the presence of an extra space in the vicinity of the A-ring. It was suggested that substituents of synthetic A-ring analogues could occupy this additional space. In this connection, another example of the importance of the size and the atomic distribution in this biological property was demonstrated by Fujishima et al.[22] These authors found that the simple introduction of the second methyl group in a specific position can cause approximately 100-fold reduction of the affinity to VDR.

On the other hand, the variables in this model are related to the electronegativity and the atomic mass. In a general way, the electronegativity is the power of an atom in a molecule to attract electrons itself. These electronegativity values are useful in determining the bond polarity for the molecule. The bond polarity is a vector, pointing from the atom with less electronegativity to the atom with larger values. If the vector sum of bond polarities of a molecule is not zero, the molecule is said to

Table 2. The statistical parameters of the lineal regressions models obtained for the four kinds of descriptors

Descriptors	Variables	R^2	S	F	q^2_{LOO}	S_{LOO}	q^2_{LGO}	S_{LGO}
BCUT	BEHm2; BEHm8; BEHv1; BELv8; BELe1; BELe2; BELe7; BEHp4	0.347	1.061	5.195	0.238	1.143	0.134	1.276
Galvez topological charge indices	GGI1; GGI2; GGI4; GGI8; GGI9; GGI10; JGI4; JGI8	0.291	1.107	4.132	0.153	1.321	0.117	1.298
2D autocorrelations	ATS4v; ATS3p; ATS5p; MATS4m; MATS1v; MATS3v; GATS2e; GATS4e	0.376	1.035	5.912	0.258	1.136	0.212	1.251
WHIM	*L2m*; *L1v*; *E1v*; *E1e*; *Dm*	0.717	0.774	40.537	0.679	0.793	0.652	0.841

have dipole moment. The affinity of the VDR for these compounds depends on the bond dipole moment. This result is explained by the fact that when the total polarity of molecules is increased, the hydrophobicity is lower, affecting the capacity of the molecule to permeate across the biological membranes to reach intracellular targets like VDR in different tissues and organs in the body.

As we previously pointed out, one of the objectives of the current work was to compare several kinds of molecular descriptors for describing the property under study. Consequently, we have developed three other models using the same data set that was included in our QSAR study with the WHIM descriptors. For this aim, the BCUT, Galvez topological charge indexes, and 2D autocorrelation descriptors were included. The results of these models are shown in Table 2.

The statistical information for the best regressions of affinity for VDR with these molecular descriptors shows that the WHIM descriptors explain the experimental variance of the data better than the other approximations. This behavior should be due to the fact that the 3D descriptors are better for modeling this biological property than the 2D descriptors because it is well known that the spatial configuration and the stereochemistry of the ligands are determinant factors for the affinity at the VDR.

On the other hand, the model presented by the 2D kinds of descriptors included eight variables in their equations, in comparison with only five in our model. From a mechanistic point of view, inclusion of a variable into a QSAR model suggests that the variable is related directly as the chemistry of a molecule influences its biological activity. Whilst the affinity for VDR is certainly a multivariate process, there are not an infinite number of controlling factors, and QSAR would be expected to have no more than a few factors controlling this biological activity.

Finally, cross-validation LOO and LGO types were developed, and it was possible to confirm that the model obtained using the WHIM descriptors had a greater coefficient of correlation (q^2_{LOO}, q^2_{LGO}) and showed a minor standard deviation (S_{LOO}, S_{LGO}) for these tests (Table 2). For those reasons, we considered that WHIM methodology can be a very useful tool for the prediction of affinity for VDR of different vitamin D analogues.

The prediction of affinity for VDR of vitamin D analogues has been a goal of pharmaceutical companies due to the importance of this biological property for many pharmacological targets. For this reason, this in silico method has been applied in order to predict this property in the early stage of drug development and some of them have become important tools in the selection of new drug candidates. In this study, WHIM descriptor was used to predict this activity for a series of compounds. The procedure has shown that a good regression model can be obtained using the weighted holistic invariant molecular descriptor weight by the atomic Sanderson electronegativity, van der Waals volume, and atomic masses. The linear model developed in the current work is easily calculated and suitable for the rapid prediction of VDR affinity, and the cross-validation of the final model supports this claim. This suggests that the present method should be regarded as one of the choices for lead optimization programs in the drug discovery process.

Acknowledgments

Maykel Pérez González thanks the University of Vigo for financial support. This work was supported by a grant from the Xunta de Galicia (PGIDT04 BTF301031PR).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2005.08.065.

References and notes

1. Masuno, H.; Yamamoto, K.; Wang, X.; Choi, M.; Ooizumi, H.; Shinki, T.; Yamada, S. *J. Med. Chem.* **2002**, *45*, 1825.
2. *Vitamin D*; Feldman, D., Glorieux, F. H., Pike, J. W., Eds.; Academic Press: San Diego, CA, 1997; For a general review of vitamin D chemistry and biology, see (a) *Vitamin D: Chemistry, Biology and Clinical Applications of the Steroid Hormone*; Norman, A. W., Bouillon, R., Thomasset, M., Eds.; Vitamin D Workshop: Riverside, CA, 1997; (b) Feldman, D.; Glorieux, F. H.; Pike, J. W. *Vitamin D*; Academic Press: San Diego, 1997; (c) Dai, H.; Posner, G. H. *Synthesis* **1994**, *1383*; (d) Zhu, G.-D.; Okamura, W. H. *Chem. Rev.* **1995**, *95*, 1877; (e) Posner, G. H.; Kahraman, M. *Eur. J. Org. Chem.* **2003**, *3889*.
3. Pike, J. W. *Annu. Rev. Nutr.* **1991**, *11*, 189.
4. Mangelsdorf, D. J.; Thummel, C.; Beato, M.; Herrlich, P.; Schutz, G.; Umesono, K.; Blumberg, B.; Kastner, P.; Mark, M.; Chambon, P.; Evans, R. M. *Cell* **1995**, *83*, 835.
5. Bouillon, R.; Okamura, W. H.; Norman, A. W. *Endocrine Rev.* **1995**, *16*, 200.
6. Verstuyf, A.; Segaert, S.; Verlinden, L.; Bouillon, R.; Mathieu, C. *Exp. Opin. Invest. Drugs* **2000**, *9*, 443.
7. Yamamoto, K.; Takahashi, J.; Hamano, K.; Yamada, S. *J. Org. Chem.* **1993**, *58*, 2530.
8. Yamada, S.; Yamamoto, K.; Masuno, H.; Ohta, M. *J. Med. Chem.* **1998**, *41*, 1467.
9. Ríos-Santamaría, I.; García-Doménech, R.; Gálvez, J.; Cortijo, J.; Santamaría, P.; Morcillo, E. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 477.
10. González, M. P.; Terán, C.; Fall, Y.; Teijeira, M.; Besada, P. *Bioorg. Med. Chem.* **2005**, *13*, 601.
11. González, M. P.; Terán, C. *Bioorg. Med. Chem.* **2004**, *12*, 2985.
12. González, M. P.; Terán, C. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3077.
13. González, M. P.; Terán, C. *Bull. Math. Biol.* **2004**, *66*, 907.
14. González, M. P.; González, H. D.; Cabrera, M. A.; Molina, R. R. *Bioorg. Med. Chem.* **2004**, *12*, 735.

15. González, M. P.; Días, L. C.; Morales, A. H.; Rodríguez, Y. M.; Gonzaga de Oliveira, L.; Gómez, L. T.; González, H. D. *Bioorg. Med. Chem.* **2004**, *12*, 4467.
16. Ramos de Armas, R.; González, H. D.; Molina, R. R.; González, M. P.; Uriarte, E. *Bioorg. Med. Chem.* **2004**, *12*, 4815.
17. González, M. P.; Morales, A. H.; Cabrera, M. A. *Bioorg. Med. Chem.* **2005**, *3*, 1775.
18. Morales, A. H.; Cabrera, M. A.; González, M. P.; Molina, R. R.; González, H. D. *Bioorg. Med. Chem.* **2005**, *13*, 2477.
19. Fernández, M.; Caballero, J.; Morales, A. H.; Castro, E.; González, M. P. *Bioorg. Med. Chem.* **2005**, *13*, 3269.
20. González, M. P.; Teran, C.; Teijeira, M.; Besada, P. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2641.
21. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
22. Fujishima, T.; Kittaka, A.; Yamaoka, K.; Takeyama, K.; Kato, S.; Takayama, H. *Org. Biomol. Chem.* **2003**, *1*, 1863.
23. Yamada, S.; Yamamoto, K.; Masuno, H.; Choi, M. *Steroids* **2001**, *66*, 177.
24. Saito, N.; Suhara, Y.; Kurihara, M.; Fujishima, T.; Honzawa, S.; Takayanagi, H.; Kozono, T.; Matsumoto, M.; Ohmori, M.; Miyata, N.; Takayama, H.; Kittaka, A. *J. Org. Chem.* **2004**, *69*, 7463.
25. Ono, K.; Yoshida, A.; Saito, N.; Fujishima, T.; Honzawa, S.; Suhara, Y.; Kishimoto, S.; Sugiura, T.; Waku, K.; Takayama, H.; Kittaka, A. *J. Org. Chem.* **2003**, *68*, 7407.
26. Konno, K.; Fujishima, T.; Maki, S.; Liu, Z.; Miura, D.; Chokki, M.; Ishizuka, S.; Yamaguchi, K.; Kan, Y.; Kurihara, M.; Miyata, N.; Smith, C.; DeLuca, H.; Takayama, H. *J. Med. Chem.* **2000**, *43*, 4247.
27. Fujishima, T.; Zhaopeng, L.; Konno, K.; Nakagawa, K.; Okano, T.; Yamaguchi, K.; Takayama, H. *Bioorg. Med. Chem.* **2001**, *9*, 525.
28. Fujishima, T.; Kojima, Y.; Azumaya, I.; Kittaka, A.; Takayama, H. *Bioorg. Med. Chem.* **2003**, *11*, 3621.
29. Fujishima, T.; Konno, K.; Nakagawa, K.; Kurobe, M.; Okano, T.; Takayama, H. *Bioorg. Med. Chem.* **2000**, *8*, 123.
30. Suhara, Y.; Kittaka, A.; Kishimoto, S.; Calverley, M.; Fujishima, T.; Saito, N.; Sugiura, T.; Waku, K.; Takayama, H. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3255.
31. Honzawa, S.; Suhara, Y.; Nihei, K.; Saito, N.; Kishimoto, S.; Fujishima, T.; Kurihara, M.; Sugiura, T.; Waku, K.; Takayama, H.; Kittaka, A. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3503.
32. Takayama, H.; Kittak, A.; Fujishima, T.; Suhara, Y.; Takayama, H. *Recent Results Cancer Res.* **2000**, *64*, 289.
33. Fujishima, T.; Kittaka, A.; Kurihara, M.; Saito, N.; Honzawa, S.; Kishimoto, S.; Sugiura, T.; Waku, K.; Takayama, H. *J. Steroid. Biochem. Mol. Biol.* **2004**, *89*(90), 89.
34. Odrzywolska, M.; Chodyski, M.; Halkes, S.; Van De Velde, J.; Fitak, H.; Kutner, A. *Chirality* **1999**, *11*, 249.
35. Fujishima, T.; Konno, K.; Nakagawa, K.; Tanaka, M.; Okano, T.; Kurihara, M.; Miyata, N.; Takayama, H. *Chem. Biol.* **2001**, *8*, 1011.
36. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *07*, 3902.
37. MOPAC version 6.0. Frank J. Seiler Research Laboratory, US Air Force Academy, Colorado Springs CO 1993.
38. Todeschini, R.; Consonni, V.; Pavan, M.; Dragon Software version 2.1, 2002.
39. Van der Waterbeemd, H.. In Manhnhold, R., Krogsgaard-Larsen, H., Van der Waterbeemd, H., Eds.; Discriminant analysis for activity prediction. Chemometric methods in molecular design. Method and Principles in Medicinal Chemistry; VCH: Weinhiem, 1995; Vol. 2, p 265.
40. STATISTICA version 6.0. Statsoft, Inc 2002.
41. Randić, M. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311.
42. Randić, M. *New J. Chem.* **1991**, *15*, 517.
43. Randić, M. *J. Mol. Struct. (Theochem)* **1991**, *233*, 45.