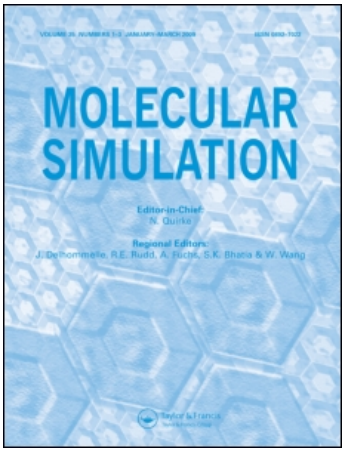


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

MIA-QSAR modelling of activities of a series of AZT analogues: bi- and multilinear PLS regression

Mohammad Goodarzi[a]; Matheus Puggina de Freitas[b]
[a] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), La Plata, Argentina [b] Departamento de Química, Universidade Federal de Lavras, Lavras, MG, Brazil

First published on: 23 September 2009

To cite this Article Goodarzi, Mohammad and de Freitas, Matheus Puggina(2010) 'MIA-QSAR modelling of activities of a series of AZT analogues: bi- and multilinear PLS regression', Molecular Simulation, 36: 4, 267 — 272, First published on: 23 September 2009 (iFirst)
To link to this Article: DOI: 10.1080/08927020903278001
URL: http://dx.doi.org/10.1080/08927020903278001

Molecular Simulation
Vol. 36, No. 4, April 2010, 267–272

Taylor & Francis
Taylor & Francis Group

MIA-QSAR modelling of activities of a series of AZT analogues: bi- and multilinear PLS regression

Mohammad Goodarzi[a] and Matheus Puggina de Freitas[b]*

[a]*Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), UNLP, CCT La Plata-CONICET, Diag. 113 y 64, C.C. 16, Suc. 4, 1900, La Plata, Argentina;* [b]*Departamento de Química, Universidade Federal de Lavras, UFLA, Caixa Postal 3037, Lavras, MG 37200-000, Brazil*

(*Received 23 June 2009; final version received 20 August 2009*)

The activities of a series of azidothymidine derivatives, compounds with anti-HIV potency, were computationally modelled using multivariate image analysis applied to quantitative structure–activity relationships (MIA-QSAR). Two regression methods were tested in order to find the best correlation between actual and predicted activities: bilinear (traditional) partial least squares (PLS), applied to the unfolded dataset, and multilinear PLS (N-PLS), applied to the three-way array. The predictive abilities of the PLS- and N-PLS-based models were found to be nearly equivalent, and both the methods derived QSAR models that are statistically superior to conventional QSAR, in which physicochemical descriptors and multiple linear regression were applied.

Keywords: MIA-QSAR; AZT analogues; HIV; PLS; N-PLS

1. Introduction

Azidothymidine (AZT; Figure 1) was the first antiretroviral drug approved for the treatment of AIDS. The development of novel anti-HIV compounds has been progressively focused in order to improve the potency and selectivity of ligands towards specific binding sites, and also to find target compounds against resistant viruses. Therefore, modification of the AZT structure, e.g. the replacement of 3′-hydroxyl of thymidine by an azido group, results in an alternate substrate for the inhibition of HIV reverse transcription [1]. A series of 2-amino-9-(azido-2,3-dideoxy-β-D-*erythro*-pentofuranosyl)-6-substituted-9*H*-purines (Table 1) [2], analogues of AZT, have shown anti-human immunodeficiency virus activity (anti-HIV-1_{IIIB} activity), and are representative systems to derive useful QSAR models.

The most recently applied structure-based QSAR analyses are based on three- and multi-dimensional approaches [3–8], which can give important insights about which interaction of a given structure moiety provides negative or positive influences on the activity. However, such methods require conformational screening and 3D alignment of ligands, and are supposed to be not superior to classical 2D methods, at least in many practical cases [9]. Furthermore, multivariate image analysis (MIA) descriptors have a proportioned satisfactory correlation between chemical structures and biological activities in a variety of studies [10–16]. The so-called multivariate image analysis applied to quantitative structure–activity relationships (MIA-QSAR) method is an image-based methodology in which descriptors are pixels (binaries) of 2D chemical structures, and has demonstrated some operational advantages over commonly used 3D methods and requires just a modest investment.

Different regression methods may be utilised in the MIA-QSAR modelling. The superposition of K images (2D chemical structures) results in a three-way array, which can be unfolded to a two-way array (an $\mathbf{X}$-matrix); both bilinear (traditional) [17] and multilinear [18] partial least-squares (PLS) regression methods may then be applied to correlate descriptors with the corresponding dependent variables. Multilinear PLS (N-PLS) has been supposed to be superior to the unfolding PLS due to its simplicity (the number of variables can be effectively reduced) and predictive ability [19,20], but one still needs to know how it performs and compares with bilinear PLS when coupled to MIA descriptors.

Thus, the goal of this work was to develop PLS- and N-PLS-based MIA-QSAR models to an important class of anti-HIV compounds, as well as to compare their predictive ability with a conventional approach, in which representative physicochemical descriptors, together with multiple linear regression (MLR), were used to correlate compounds with the corresponding bioactivities.

*Corresponding author. Email: matheus@ufla.br

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927020903278001
http://www.informaworld.com

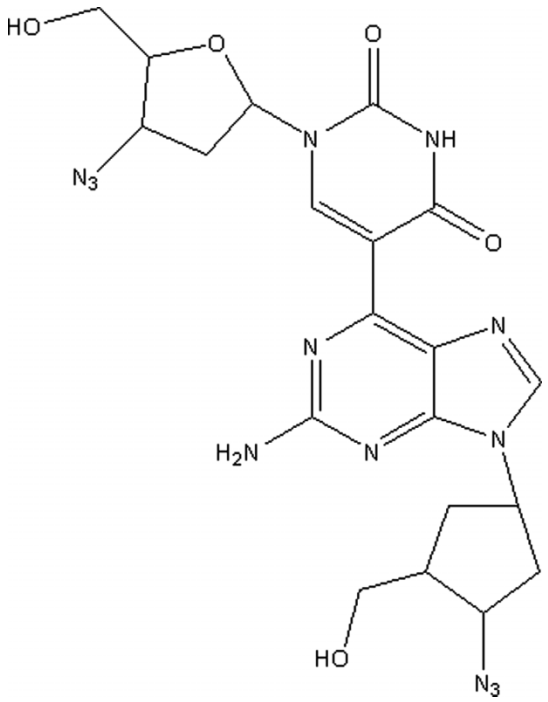


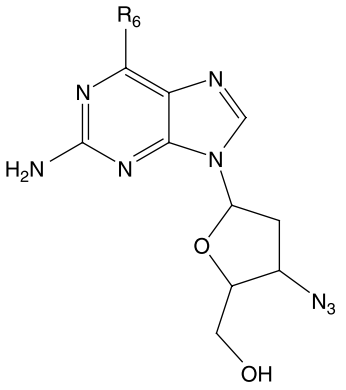
Figure 1. Structure of 3′-azido-3′-deoxythymidine (AZT).

2. Computational methods

2.1 MIA-QSAR

MIA descriptors are binaries obtained from the pixels of 2D chemical structures, which must be drawn by using any appropriate drawing program. In this work, 25 anti-HIV compounds (Table 1) obtained from the literature [2] were drawn through the ChemDraw program [21]. Each 2D chemical structure was systematically built, i.e. the scaffold (the fused rings) was kept unaltered and the substituents of Table 1 were added accordingly. The chemical structures were then transformed into bitmaps and saved in a 600×650 pixels workspace. Since the dataset used is a congeneric series, chemical structures possess a common substructure, which was superimposed for the entire series. This was achieved by taking a pixel in common among the whole series of compounds, as shown in Figure 2, and fitting it in a given coordinate of the defined workspace. This 2D alignment is rapid and requires a simple manual precision. The 25 2D images were read and converted into double arrays by using Matlab [22]. These 25 samples were then grouped to give a $25 \times 600 \times 650$ three-way array, which was calibrated with the corresponding dependent variables (anti-HIV-1_{IIIB} activity – pIC_{50}, IC_{50} in μM) using the N-PLS regression method [17]. The three-way array was unfolded to a two-way array (**X**-matrix) and another calibration was carried out through PLS. The advantage of this latter procedure is that the **X**-matrix can be reduced in dimension by removing columns with zero variance (common pixels and blank workspace are deleted). However, this makes back-folding impossible. Validation of both models was achieved through leave-one-out (LOO) and leave-20%-out (to simulate test sets) cross-validations (CVs), and the predictive ability was statistically evaluated through the root mean square errors of calibration (RMSEC) and validation (RMSECV), as well as by the squared correlation coefficients of the regression line of experimental vs. fitted and predicted activity values (r^2 and q^2, respectively).

Table 1. Experimental and calibrated activities (pIC_{50}, IC_{50} in $mol\,l^{-1}$) for a series of 2-amino-9-(azido-2,3-dideoxy-β-D-*erythro*-pentofuranosyl)-6-substituted-9*H*-purines used in the MIA-QSAR modelling.



No.	R_6	Exp.	Fitted PLS	Fitted N-PLS
10	$N(CH_2)_4$	5.10	5.10	5.14
11	OCH_3	5.04	4.89	4.88
12	OCH_2CH_3	4.64	4.80	4.74
13	$OCH_2CH_2CH_3$	5.04	4.95	4.88
14	$OCH(CH_3)_2$	4.66	4.74	4.72
15	$O(CH_2)_3CH_3$	4.72	4.77	4.91
16	O-cyclobutyl	4.82	4.92	4.93
17	OC_6H_5	4.37	4.29	4.28
18	$OCH_2C_6H_5$	5.10	5.09	5.05
20	NH_2	5.70	5.48	5.61
21	$NHCH_3$	4.28	4.58	4.61
22	$NHCH_2CH_3$	4.77	4.57	4.48
23	$NH(CH_2)_2CH_3$	4.64	4.65	4.62
24	NH-cyclopropyl	4.74	4.92	4.73
25	$NH(CH_2)_3CH_3$	4.54	4.47	4.64
26	NH-cyclobutyl	4.85	4.61	4.65
27	NHC_6H_5	3.95	4.10	4.05
29	$NH(CH_2)_2C_6H_5$	4.68	4.60	4.58
31	OH	5.30	5.42	5.14
32	$N(CH_3)_2$	4.92	4.88	5.10
33	$CH_3NCH_2CH_3$	4.96	4.85	4.96
34	$CH_3N(CH_2)_2CH_3$	4.92	4.95	5.10
35	NCH_3-cyclopropyl	5.40	5.22	5.22
37	$N(CH_2)_3$	5.04	5.19	5.17
38	Cl	5.22	5.38	5.21

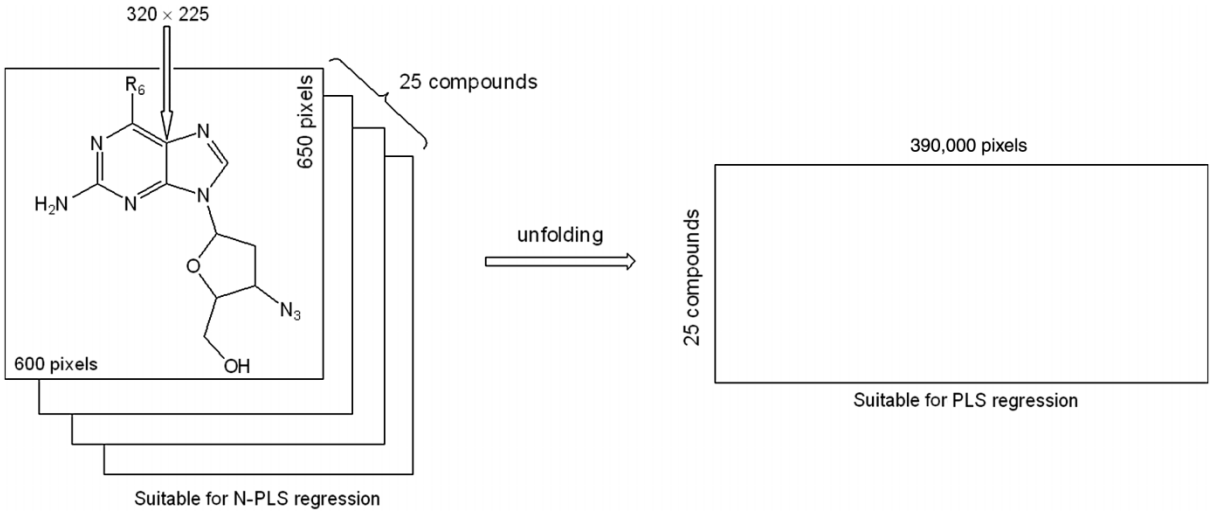


Figure 2. Generic molecular structure of 2-amino-9-(azido-2,3-dideoxy-β-D-*erythro*-pentofuranosyl)-6-substituted-9*H*-purines, anti-HIV compounds, drawn in 600 × 650 pixels windows. The arrow in the structure indicates the point fixed at the 320 × 225 coordinate. A three-way array was obtained by grouping the 25 2D images (treated with N-PLS), and then unfolded to a two-way array (treated with PLS).

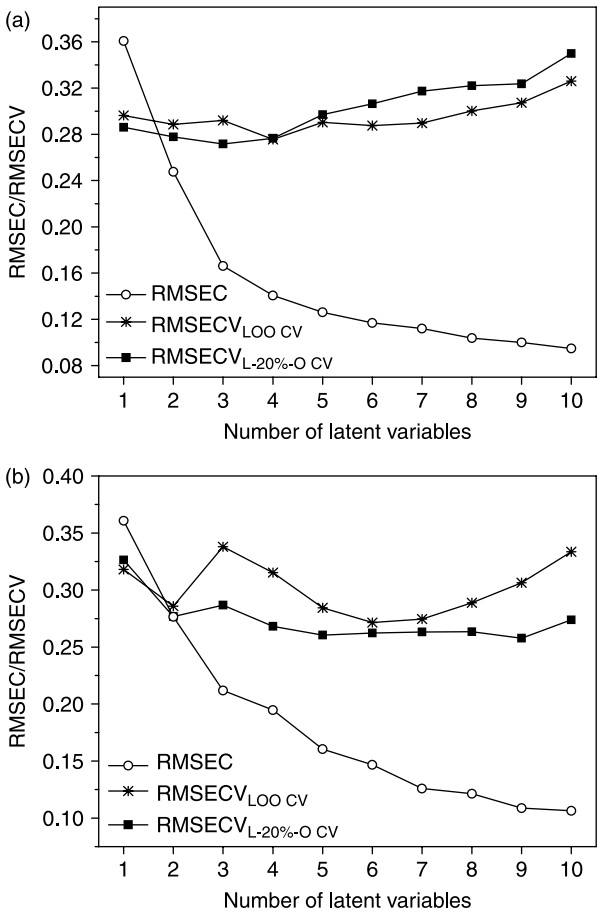


Figure 3. Plots of number of LVs vs. RMSEC/RMSECV for (a) MIA-QSAR/PLS model and (b) MIA-QSAR/N-PLS model.

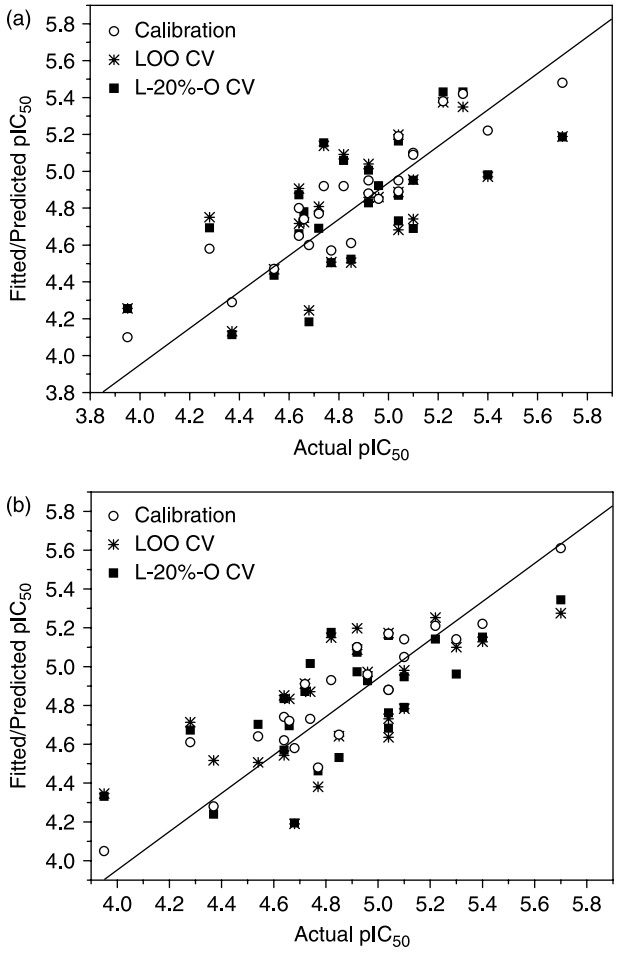


Figure 4. Plots of experimental vs. fitted (calibrated) and predicted pIC_{50} for (a) MIA-QSAR/PLS model and (b) MIA-QSAR/N-PLS model.

Table 2. Statistical parameters of the PLS- and N-PLS-based MIA-QSAR models.

Parameter	MIA-QSAR/PLS	MIA-QSAR/N-PLS	2D QSAR/MLR
LVs	4	6	–
r^2	0.849	0.835	0.512
RMSEC	0.14	0.15	0.25
$q^2_{\text{LOO CV}}$	0.475	0.463	–
RMSECV	0.28	0.27	1.28
$q^2_{\text{L-20\%-O CV}}$	0.493	0.504	–
RMSECV	0.27	0.27	–

2.2 Conventional QSAR

The structures of the compounds were pre-optimised with the molecular mechanics force field (MM+) procedure included in HyperChem version 7.0 [23], and the resulting geometries were further refined by means of the semi-empirical method parametric method-3 (PM3) using the Polak–Ribiere algorithm and a gradient norm limit of 0.01 kcal Å^{-1}. Representative descriptors, namely $\log P$, surface area, molecular volume, molar refractivity, polarisability, hydration energy and molecular weight, were then calculated through this software, and regressed against the experimental bioactivities using MLR.

3. Results and discussion

The chemical structures of the 25 anti-HIV-1_{IIIB} compounds were grouped in such a way that the coincident substructures (the basic scaffold) were all congruent, i.e. the substituents accounted for the explained variance of data – the basis of a structure-based QSAR analysis. The subsequent analyses differed in how the generated descriptors (pixels) were treated: in the first approach, the three-way array originated from the superimposition of K 2D images (chemical structures with $I \times J$ pixels dimension each), giving a $K \times I \times J$ three-way array ($25 \times 600 \times 650$), was regressed against the bioactivities through N-PLS. The three-way array was then unfolded to a $K \times (I \times J)$ two-way array ($25 \times 390{,}000$), in which traditional (bilinear) PLS was the regression method used.

The optimum number of latent variables (LVs) in each model was achieved by jointly analysing RMSEC and RMSECV (Figure 3). It was found that both models are stable (q^2 does not vary, independent of how many groups were left out each time), and that RMSEC and RMSECV minimised, or at least did not decrease significantly, after four PLS and six N-PLS components; thus, these were the LVs used in calibration and validation. Calibration with both PLS and N-PLS gave reasonable correlation ($r^2 > 0.8$), with low residuals between experimental and fitted pIC_{50}, as depicted in Tables 1 and 2. Also, both models were validated through two different procedures: LOO CV, where 25 models were developed with one different prediction sample at a time, and leave-20%-out CV, where random test sets, composed of five prediction samples each, were utilised. These procedures gave q^2

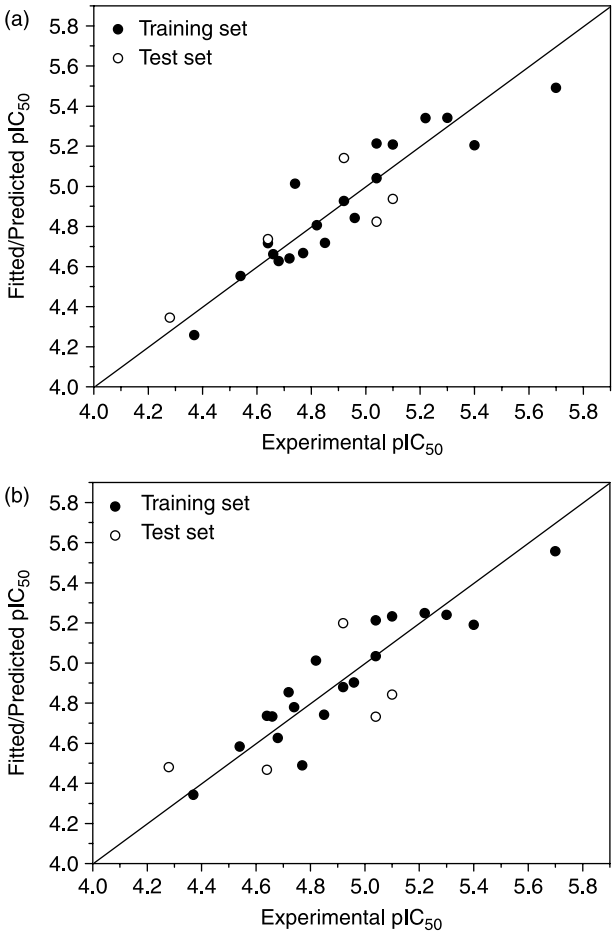


Figure 5. Plots of experimental vs. fitted (training set) and predicted (test set) pIC_{50} for (a) MIA-QSAR/PLS model and (b) MIA-QSAR/N-PLS model.

Table 3. Statistical results for the dataset split into training and test sets.

Model	Sets	r^2	r^2_m	r^2_0	r'^2_0	k	k'	$(r^2 - r^2_0)/r^2$	$(r^2 - r'^2_0)/r^2$	$\lvert r^2_0 - r'^2_0 \rvert$
PLS	Training set	0.884	0.775	0.868	0.884	0.999	1.000	0.017	0.000	0.015
	Test set	0.704	0.485	0.607	0.703	0.999	1.000	0.137	0.001	0.096
N-PLS	Training set	0.895	0.795	0.882	0.895	0.999	1.000	0.014	0.000	0.012
	Test set	0.414	0.220	0.195	0.349	0.988	1.010	0.528	0.157	0.154

Table 4. Physicochemical descriptors[a] calculated through HyperChem for the 25 compounds of dataset, and fitted activities obtained through MLR regression[b].

No.[c]	Log *P*	SA (approx.)	SA (graid)	MV	MR	α	HE	MW	Fitted pIC_{50}
10	−0.65	554.84	773.9	1320.03	140.67	49.69	−35.90	525.48	5.07
11	4.57	197.90	281.94	460.68	75.08	28.64	−20.20	345.36	4.93
12	0.93	372.98	517.35	833.21	79.96	28.95	−23.56	306.28	4.99
13	1.27	413.47	553.87	890.83	84.71	30.78	−22.53	320.31	4.76
14	1.74	452.02	581.9	944.48	89.24	32.62	−21.91	334.33	4.80
15	1.68	432.05	567.8	927.73	89.13	32.62	−21.23	334.33	4.62
16	2.13	490.92	613.7	997.7	93.84	34.46	−21.43	348.36	5.14
17	1.72	408.04	584.61	957.56	91.78	33.68	−21.38	346.34	4.19
18	1.42	412.80	592.16	977.68	104.55	36.78	−23.65	368.35	5.14
20	1.94	411.34	612.29	1024.92	108.72	38.61	−22.13	382.38	5.17
21	0.20	327.64	495.23	790.63	77.14	27.83	−27.17	291.27	4.82
22	0.61	374.05	518.32	836.54	81.91	29.66	−24.19	305.30	4.80
23	0.95	407.55	549.43	891.26	86.66	31.50	−23.13	319.32	4.83
24	1.42	442.64	579.08	944.75	91.19	33.33	−22.47	333.35	4.87
25	1.00	386.65	563.53	921.43	89.12	32.56	−22.74	331.33	4.60
26	1.82	482.28	609.67	998.18	95.79	35.17	−21.95	347.38	5.03
27	1.40	399.65	579.12	954.42	93.73	34.40	−22.11	345.36	4.34
29	1.24	404.87	582.53	968.89	106.80	37.49	−24.24	367.37	4.48
31	1.87	447.24	642.07	1076.55	115.42	41.16	−23.29	395.42	5.20
32	0.89	335.28	487.96	779.32	75.20	27.11	−28.79	292.25	4.94
33	0.97	413.14	544.18	885.76	87.21	31.50	−20.74	319.32	4.94
34	1.31	441.15	570.66	930.82	91.96	33.33	−20.17	333.35	4.72
35	1.78	477.94	598.15	984.07	96.48	35.17	−19.68	347.38	4.86
37	1.37	422.10	584.4	962.86	94.42	34.40	−20.00	345.36	4.87
38	0.84	389.47	560.79	917.79	90.11	32.56	−20.73	331.33	5.27

[a] SA, surface area; MV, molar volume (cm^{-3}); MR, molar refractivity; α, polarisability; HE, hydration energy ($kcal\,mol^{-1}$); MW, molecular weight.
[b] MLR equation: $pIC_{50} = 26.67 + 2.3482 \log P - 0.03539\,SA\,(approx.) + 0.07613\,SA\,(graid) - 0.02769\,MV + 0.24347\,MR + 0.23857\alpha + 0.16973\,HE - 0.16132\,MW$.
[c] Compounds numbered according to the literature [2].

marginally inferior or superior to 0.5 (Table 2), which is in the range of acceptable values. Since the range in the activity values was not large, one would expect correlation not too high, but the low residuals illustrated in Figure 4, together with the small RMSECV found (0.27–0.28), confirm the satisfactory predictive ability of both models.

Despite the similar prediction power achieved in this work using PLS and N-PLS regression methods, the MIA-QSAR/PLS model was more parsimonious (four PLS LVs against six N-PLS LVs), while N-PLS has been found to be more advantageous when compared with PLS by several factors, such as improved interpretability, increased predictive ability and it is more stable [19]. However, the great operational advantage of using PLS instead of N-PLS is that the $25 \times 390{,}000$ **X**-matrix was reduced to a 25×3386 dimension after deleting columns with zero variance, i.e. those columns corresponding to common pixels (scaffold and blank workspace) among the whole series of compounds (2D images). This step minimises largely the memory used and reduces the computational costs; the drawback is that this procedure makes back-folding impossible, which makes interpretation of results more difficult.

Golbraikh and Tropsha [24] have suggested real test compounds for external validation to achieve a reliable QSAR model; thus, the whole dataset was split into training and test sets, composed of 20 (80%) and 5 (20%, compounds **13**, **18**, **21**, **23** and **32**) compounds, respectively. The best PLS and N-PLS calibration models were found using four LVs, whose results are depicted in Table 3 and Figure 5. According to Golbraikh and Tropsha [24], models are considered acceptable, if they satisfy all of the following conditions: (i) $q^2 > 0.5$, (ii) $r^2 > 0.6$, (iii) r_0^2 or r'^2_0 is close to r^2, such that $[(r^2 - r_0^2)/r^2]$ or $[(r^2 - r'^2_0)/r^2] < 0.1$, and $0.85 \le k \le 1.15$ or $0.85 \le k' \le 1.15$, where r^2 and r_0^2 are squared correlation coefficient values between observed and predicted values of the test set compounds with and without an intercept, respectively, and k and k' are the slopes of regression lines through the origin. Adequacy was found for both the PLS and N-PLS models based on these criteria. Additional statistic has been proposed by Roy and Roy [25] in order to test the external predictability, namely r_m^2, which is defined as $r^2[1 - (r^2 - r_0^2)^{1/2}]$. For a model with good external predictability, the r_m^2 value should be greater than 0.5. The corresponding value for the PLS model was found to be acceptable according to this criterion, while the N-PLS-based model gave a r_m^2 smaller than 0.5; however, it should be borne in mind the diminutive amount of compounds used in the external set.

Overall, the PLS-based model was found to be marginally superior to the N-PLS one, but although the simplicity of building MIA-QSAR models when compared to the most used multidimensional approaches, they are not easier to manipulate than conventional QSARs. Thus, it must be proved that MIA-QSAR is a suitable alternative and can be useful when physicochemical descriptors fail in predicting the bioactivities. Parameters related to lipophilicity/hydrophilicity and steric volume, such as $\log P$ and molar refractivity, have been successfully applied to derive useful 2D QSAR models [26,27]. Therefore, a series of representative physicochemical parameters were calculated, and the MLR using the data of Table 4 resulted in a poor correlation ($r^2 = 0.512$) and high residuals (see Table 2 for errors), which is insufficient to be used in the prediction of new drug candidates.

4. Conclusions

The above results show that both MIA-QSAR models built for the important class of AZT analogues are statistically significant and suggest that this image-based method may be reliably applied to predict the activities of proposed congeners. An important finding is that the incipiently explored N-PLS regression provided results comparable to the well-known bilinear PLS, with the advantage of being applied directly to the 3D array, avoiding the unfolding step and increasing the potential for interpretation, since descriptors (pixels giving rise to chemical structures) are not disconnected to form a two-way array. Finally, the results were superior to the ones found through conventional QSAR, demonstrating that the presented approach may serve at least as a complementary tool when existing methods do not work well.

Acknowledgements

CNPq is gratefully acknowledged for the fellowship (to M.P.F.), and FAPEMIG of Brazil and the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) of Argentina for the financial support.

References

[1] P.A. Furman, J.A. Fyfe, M.H. St. Clair, K. Weinhold, J.L. Ridoout, G.A. Freeman, S.N. Lehrman, D.P. Bolognesi, S. Broder, H. Mitsuya, and D.W. Barry, *Phosphorylation of 3′-azido-3′-deoxythymidine and selective interaction of the 5′-triphosphate with human-immunodeficiency-virus reverse-transcriptase*, Proc. Natl Acad. Sci. USA 83 (1986), pp. 8333–8337.

[2] G.A. Freeman, S.R. Shaver, J.L. Rideout, and S.A. Short, *2-Amino-9-(3-azido-2,3-dideoxy-beta-D-erythro-pentofuranosyl)-6-substituted-9H-purines – synthesis and anti-HIV activity*, Bioorg. Med. Chem. 3 (1995), pp. 447–458.

[3] R.D. Cramer III, D.E. Patterson, and J.D. Bunce, *Comparative molecular-field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins*, J. Am. Chem. Soc. 110 (1988), pp. 5959–5967.

[4] G. Klebe, U. Abraham, and T. Mietzner, *Molecular similarity indexes in a comparative-analysis (CoMSIA) of drug molecules to correlate and predict their biological-activity*, J. Med. Chem. 37 (1994), pp. 4130–4146.

[5] P.J. Goodford, *GRID*, University of Oxford, Oxford, 1995.

[6] A.J. Hopfinger, S. Wang, J.S. Tokarski, B. Jin, M. Albuquerque, P.J. Madhav, and C. Duraiswami, *Construction of 3D-QSAR models using the 4D-QSAR analysis formalism*, J. Am. Chem. Soc. 119 (1997), pp. 10509–10524.

[7] A. Vedani and M. Dobler, *5D-QSAR: The key for simulating induced fit?* J. Med. Chem. 45 (2002), pp. 2139–2149.

[8] A. Vedani, M. Dobler, and M.A. Lill, *Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor*, J. Med. Chem. 48 (2005), pp. 3700–3703.

[9] F. Tian, P. Zhou, and Z. Li, *A novel atom-pair hologram (APH) and its application in peptide QSARs*, J. Mol. Struct. 871 (2007), pp. 140–148.

[10] M.P. Freitas, S.D. Brown, and J.A. Martins, *MIA-QSAR: A simple 2D image-based approach for quantitative structure–activity relationship analysis*, J. Mol. Struct. 738 (2005), pp. 149–154.

[11] M.P. Freitas, *MIA-QSAR modelling of anti-HIV-1 activities of some 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners*, Org. Biomol. Chem. 4 (2006), pp. 1154–1159.

[12] M.P. Freitas, *Multivariate QSAR: From classical descriptors to new perspectives*, Curr. Comput.-Aid. Drug Des. 3 (2007), pp. 235–239.

[13] J.R. Pinheiro, M. Bitencourt, E.F.F. da Cunha, T.C. Ramalho, and M.P. Freitas, *Novel anti-HIV cyclotriazadisulfonamide derivatives as modeled by ligand- and receptor-based approaches*, Bioorg. Med. Chem. 16 (2008), pp. 1683–1690.

[14] M.P. Freitas, *Multivariate image analysis applied to QSAR: Evaluation to a series of potential anxiolytic agents*, Chemom. Intell. Lab. Sys. 91 (2008), pp. 173–175.

[15] M.P. Freitas and R. Rittner, *MIA-QSAR as an alternative approach for modeling some antifungals*, QSAR Comb. Sci. 27 (2008), pp. 582–585.

[16] M.P. Freitas, *A 2D image-based approach for modelling some glycogen synthase kinase 3 inhibitors*, Med. Chem. Res. 16 (2008), pp. 461–467.

[17] R. Bro, *Multiway calibration. Multilinear PLS*, J. Chemom. 10 (1996), pp. 47–61.

[18] H. Wold, *Estimation of principal components and related models by iterative least squares*, in *Multivariate Analysis*, K.R. Krishnaiah, ed., Academic Press, New York, NY, 1966, pp. 391–420.

[19] J. Nilsson, E.J. Homan, A.K. Smilde, G.J. Grol, and H. Wikström, *A multiway 3D QSAR analysis of a series of (S)-N-[(1-ethyl-2-pyrrolidinyl)methyl]-6-methoxybenzamides*, J. Comput.-Aid. Mol. Des. 12 (1998), pp. 81–93.

[20] M.M.C. Ferreira, *Multivariate QSAR*, J. Braz. Chem. Soc. 13 (2002), pp. 742–753.

[21] *ChemDraw Ultra version 7.0*, CambridgeSoft, Cambridge, MA (2001)

[22] *Matlab version 7.5*, MathWorks, Inc., Natick, MA, 2007.

[23] *HyperChem version 7.0*, Hypercube, Inc., Gainesville, FL, 2007.

[24] A. Golbraikh and A. Tropsha, *Beware of* q^{2}*!*, J. Mol. Graph. Modell. 20 (2002), pp. 269–276.

[25] P.P. Roy and K. Roy, *On some aspects of variable selection for partial least squares regression models*, QSAR Comb. Sci. 27 (2008), pp. 302–313.

[26] C. Hansch and T. Fujita, *ρ-σ-π Analysis. Method for correlation of biological activity – chemical structure*, J. Am. Chem. Soc. 86 (1964), pp. 1616–1626.

[27] J.T. Leonard and K. Roy, *Classical QSAR modeling of HIV-1 reverse transcriptase inhibitor 2-amino-6-arylsulfonylbenzonitriles and congeners*, QSAR Comb. Sci. 23 (2004), pp. 23–35.