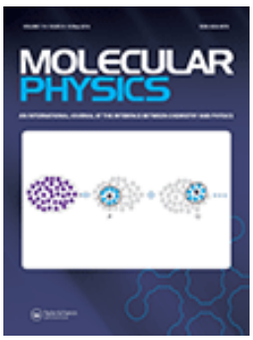


Taylor & Francis
Taylor & Francis Group

Molecular Physics

An International Journal at the Interface Between Chemistry and Physics

ISSN: 0026-8976 (Print) 1362-3028 (Online) Journal homepage: http://www.tandfonline.com/loi/tmph20

Towards molecular design using 2D-molecular contour maps obtained from PLS regression coefficients

Cleber N. Borges, Stephen J. Barigye & Matheus P. Freitas

To cite this article: Cleber N. Borges, Stephen J. Barigye & Matheus P. Freitas (2017): Towards molecular design using 2D-molecular contour maps obtained from PLS regression coefficients, Molecular Physics, DOI: 10.1080/00268976.2017.1347294

To link to this article: http://dx.doi.org/10.1080/00268976.2017.1347294

MOLECULAR PHYSICS, 2017
https://doi.org/10.1080/00268976.2017.1347294

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

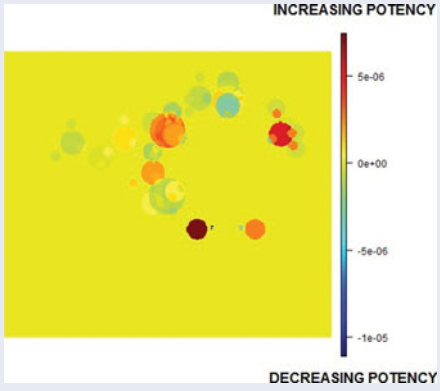
Towards molecular design using 2D-molecular contour maps obtained from PLS regression coefficients

Cleber N. Borges[a], Stephen J. Barigye[a,b] and Matheus P. Freitas[a]

[a]Department of Chemistry, Federal University of Lavras, Lavras, Brazil; [b]Facultad de Medicina, Universidad de Las Américas, Quito, Ecuador

ABSTRACT

The multivariate image analysis descriptors used in quantitative structure-activity relationships are direct representations of chemical structures as they are simply numerical decodifications of pixels forming the 2D chemical images. These MDs have found great utility in the modeling of diverse properties of organic molecules. Given the multicollinearity and high dimensionality of the data matrices generated with the MIA-QSAR approach, modeling techniques that involve the projection of the data space onto orthogonal components e.g. Partial Least Squares (PLS) have been generally used. However, the chemical interpretation of the PLS-based MIA-QSAR models, in terms of the structural moieties affecting the modeled bioactivity has not been straightforward. This work describes the 2D-contour maps based on the PLS regression coefficients, as a means of assessing the relevance of single MIA predictors to the response variable, and thus allowing for the structural, electronic and physicochemical interpretation of the MIA-QSAR models. A sample study to demonstrate the utility of the 2D-contour maps to design novel drug-like molecules is performed using a dataset of some anti-HIV-1 2-amino-6-arylsulfonylbenzonitriles and derivatives, and the inferences obtained are consistent with other reports in the literature. In addition, the different schemes for encoding atomic properties in molecules are discussed and evaluated.



ARTICLE HISTORY
Received 20 March 2017
Accepted 11 June 2017

KEYWORDS
QSAR; multivariate image analysis; partial least squares; drug design

1. Introduction

Quantitative structure-activity relationships (QSARs) have long been used in the rational design of chemical compounds of therapeutic interest by correlating bioactivities and chemical structural features, through the so-called molecular descriptors (MDs). This approach has in the recent times gained renewed interest, thanks to the advancements in the QSAR methods for codifying chemical information as well as the regularisation of the guidelines (or principles) for good practices in molecular modelling [1]. However, chemical interpretation of the QSAR models (or MDs) in terms of structural characteristics that influence the studied biological activities continues to be a challenge, and at most indirect relationships are afforded. For instance, the well-known descriptor log*P* describes the overall hydrophobicity of a molecule, which indeed correlates with several bioactivities [2], but the contribution of the different atom (or group) types in a molecule, as well as their corresponding

CONTACT Matheus P. Freitas matheus@dqi.ufla.br

substituent positions to the modelled property is rather obliterated, due to the global nature of this MD.

While efforts have been made to deal with the challenge of interpretability, particularly through more local definitions of structural features, for example, using the comparative molecular field analysis (CoMFA) [3] which computes steric and electrostatic descriptors by scanning a grid cell containing the three-dimensional (3D) molecule, these have only been partial solutions, particularly due to the very conceptual nature of these alternatives. First, the optimised geometries used to build 3D-QSAR models may not correspond to the actual bioactive conformation and, therefore, chemical interpretation based on a 3D space may not make sense [4]. Additionally, these methods are time consuming and computationally costly due to the conformational screening and 3D alignment procedures involved. On the other hand, these 3D MDs have in fact not been found to be superior to 2D ones, at least in many practical cases [5]. In this sense, a topo-chemical method proposed about a decade ago by Freitas *et al.* [6], and denominated as MIA-QSAR (acronym for multivariate image analysis applied to QSAR), has found utility as an important tool for modelling the chemical and biological properties of molecules. This method is based on 2D projections of chemical structures (images) in the sense that the numerically decoded pixels forming the images are considered as the descriptors. Therefore, the pixel coordinates in the 2D space give insight on the chemical groups and their respective positions in a molecule framework.

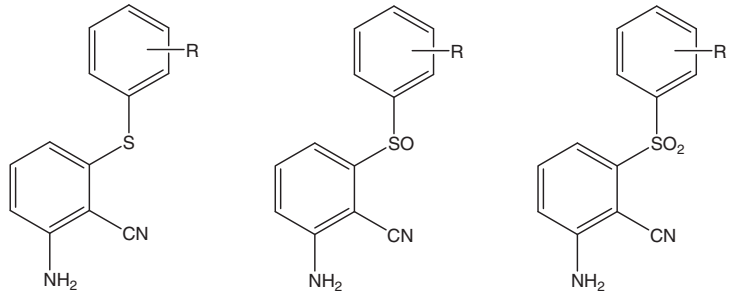
The pioneer MIA-QSAR descriptors were derived from molecules represented as black and white (binary) wireframes, and these will hereafter be denominated as the traditional MIA-QSAR descriptors. Later, generalisations based on colour schemes to improve the modelling power and the interpretability of the obtained descriptors were subsequently introduced, and named aug-MIA-QSAR [7,8]. According to the colour-based approach, atom types in a molecule are distinguished using dissimilar colours, which are numerically encoded using the RGB (red-green-blue) additive colour model. As anticipated, the use of pixels as descriptors comes along with the curse of dimensionality, characterised by many observations (pixels) for a much less number of sample points (molecules). Classically, dimensionality reduction methods based on the mapping of the features into orthogonal projections have been employed in model building. Nonetheless, despite the satisfactory statistical performance obtained with the MIA-QSAR models, chemical interpretation of these models in terms of the influence of the different functional groups/atom types to the modelled bioactivity has been elusive particularly because the identity of the original features (pixel coordinates) are concealed in the orthogonal projections (latent variables (LVs)). As a consequence, the practical utility of the MIA-QSAR models in spearheading the rational *in silico* design of new chemical entities of therapeutic interest has been diminutive.

In previous reports [7,8], efforts were made to offer some degree of interpretability to the MIA-QSAR models using the original features. To achieve this objective, feature selection tools e.g. Shannon's entropy rank-based filters were used to tremendously reduce the dimensionality of the MIA-QSAR data matrices to more manageable proportions, and regression methods such as multiple linear regression which use the original variables in model building were applied. However, an important bottleneck arose from this approach: only a very limited number of variables (pixel coordinates) from the 10,000s of coordinates contained in a multivariate image (MVI) could be used in model building consistent with the Topliss and Costello rule [9], and this often resulted in gaps in the chemical interpretation workflow. It was thus concluded that just as in microarray data, where a coherent description of the studied genes or their functions requires numerous observations, a wholesome understanding of the contribution of the different atom types and/or groups to the studied properties required that the entire **X** data matrix obtained from the MVI be employed. In this sense, the partial least squares (PLS) statistical technique has been revisited as it is unaffected by the high dimensionality of data matrices and thus all data points retrieved from an MVI may be used. Therefore, the present report introduces an innovative initiative to allow for the extraction of relevant chemical information from the aug-MIA-QSAR models based on contour maps. These maps offer graphical evidence on the structural moieties responsible for enhanced or attenuated biological activities, and thus establishing a platform for rational drug design. An example study is provided to illustrate the utility of the contour maps in the chemical interpretation of the aug-MIA-QSAR models, using the 2-amino-6-arylsulfonylbenzonitriles and derivatives, which are known to possess anti-HIV-1 activity. In addition, an evaluation of the effect of integrating steric and electrostatic aspects in the aug-MIA-QSAR colour scheme is performed.

2. Results and discussion

Four aug-MIA-QSAR models based on the defined colour schemes for atomic properties were built and then compared with the traditional MIA-QSAR model reported earlier, in which chemical structures were drawn as black and white wireframes [10]. Since pixel colours of atoms in MIA-QSAR are obtained considering the contribution of

Table 1. Data-set of 64 2-amino-6-arylsulfonylbenzonitriles and thio and sulfinyl derivatives used in the MIA-QSAR modelling, with the respective anti-HIV-1 activities.



Cpd	R	pIC_{50}	Cpd	R	pIC_{50}
1	H	1.836	**33**	3-Cl, 5-Me	3.495
2	2-OMe	2.367	**34**	3-OMe, 5-CF_3	2.684
3	3-OMe	2.222	**35**	H	2.699
4	2-Me	1.796	**36**	2-OMe	3.222
5	3-Me	2.215	**37***	3-OMe	3.046
6	4-Me	0.939	**38**	4-OMe	1.602
7	2-Cl	2.387	**39**	2-Me	2.638
8*	3-Cl	2.131	**40***	3-Me	3.398
9	2-Br	1.523	**41**	4-Me	2.022
10*	3-Br	2.292	**42**	2-Cl	2.387
11	3-F	2.009	**43**	3-Cl	3.229
12	3-CN	2.762	**44**	4-Cl	2.523
13*	4-CN	1.359	**45**	2-Br	2.301
14*	3-CF_3	1.893	**46**	3-Br	3.268
15	3-NH_2	1.502	**47**	4-Br	1.699
16	3,5-Me_2	3.367	**48**	2-F	2.523
17*	3-Cl, 5-Me	2.754	**49***	3-F	2.523
18	3-OMe, 5-Me	2.699	**50***	2-CN	2.268
19	3-OMe, 5-CF_3	2.292	**51**	3-CN	2.620
20*	2-OMe	2.319	**52**	4-CN	1.097
21	3-OMe	1.796	**53**	3-CF_3	2.456
22	2-Me	1.032	**54**	2,5-Cl_2	3.523
23	3-Me	1.534	**55***	3,5-Cl_2	4.155
24*	4-Me	1.310	**56**	3,5-Me_2	5.000
25	2-Br	1.407	**57***	3-Br, 5-Me	4.699
26	3-Br	4.097	**58**	3-Cl, 5-Me	4.523
27*	4-Br	1.694	**59**	3-OMe, 5-Me	4.301
28	2-CN	2.409	**60**	3-OMe, 5-CF_3	4.046
29	3-CN	1.848	**61**	3-OH, 5-Me	3.367
30	3-CF_3	1.398	**62**	3-OCH_2CH_3, 5-Me	4.222
31*	3,5-Me_2	3.469	**63***	3-$O(CH_2)_2CH_3$, 5-Me	4.222
32	2,5-Cl_2	2.007	**64**	3-$O(CH_2)_3CH_3$, 5-Me	3.222

*Test set compounds.

the RGB channels, they may be numerically manipulated to achieve proportionality to known atomic properties, particularly those related with the factors most relevant to the ligand–enzyme interactions and, therefore, biological activity, i.e. steric and electrostatic interactions. While the former are dependent on atomic size, the latter are due to polar interactions, which originate from the bonding of atoms with different electronegativities. Thus, the chemical structures were drawn with atoms coloured according to the van der Waals radii (r_{vdW}), Pauling's electronegativity (ε) and the relationships $r_{vdW} \times \varepsilon$ and r_{vdW}/ε to account for both effects. The images corresponding to the 64 compounds of Table 1 were aligned (Figure 1) and the descriptors matrix was subsequently regressed against the pIC_{50} values using PLS regression.

Table 2 shows the statistical parameters for the built MIA-QSAR models. As can be observed, all four models were predictive, robust and not prone to chance correlation [11,12]. Figure 2 shows the scatter plots of the experimental and calculated values for the four aug-MIA-QSAR models. In a comparison of the four models, it is observed that although they generally demonstrate comparable behaviour for most of the statistical parameters, the aug-MIA-QSAR model based on the r_{vdW}/ε relationship is least prone to fortuitous correlation [$r^2_{y\text{-rand}} = 0.238(r_{vdW})$, $0.281(\varepsilon)$, $0.257(r_{vdW} \times \varepsilon)$, $0.169(r_{vdW}/\varepsilon)$], justified by its lower degree of freedom (three LVs) relative to the rest of the models (four LVs). Therefore, according to the parsimonious principle, the r_{vdW}/ε-based aug-MIA-QSAR model is chosen for subsequent

Table 2. Statistical data for the four MIA-QSAR models.

Parameter	Model r_{vdW}	Model ε	Model $r_{vdW} \times \varepsilon$	Model r_{vdW}/ε	Literature [10]
LV´s	4	4	4	3	3
RMSEC	0.46	0.45	0.51	0.46	0.43–0.48
r^2	0.77	0.78	0.74	0.77	0.80–0.81
$RMSE_{y\text{-rand}}$	0.85	0.81	0.84	0.90	
$r^2_{y\text{-rand}}$ [a]	0.24	0.28	0.26	0.17	
[c] r^2p (y-rand)	0.56	0.55	0.51	0.59	
RMSECV	0.66	0.62	0.70	0.66	0.52–0.60
q^2	0.57	0.60	0.56	0.55	0.62–0.71
RMSEP	0.54	0.55	0.63	0.53	0.48–0.55
r^2_{test}	0.78	0.80	0.75	0.79	0.75–0.82
r^2m (test)	0.77	0.75	0.72	0.77	

[a] Mean of 10 repetitions.

Figure 1. (Colour online) Superposed chemical images used in the MIA-QSAR modelling. Congruent and varying moieties can be observed (the varying moieties explain the variance in the bioactivities block **Y**). In this example, the colours assigned to each atom type were numerically described by the RGB scale and are proportional to r_{vdW}/ε (see Section 3 for details).

analysis. Bearing in mind that the colour scheme r_{vdW}/ε incorporates information on both the atomic size and electronegativity, it may be inferred that there exists an influence of the two factors on the activity mechanism of the 2-amino-6-arylsulfonylbenzonitriles and derivatives.

The statistical performance of the four aug-MIA-QSAR models is compared with that of the previously reported traditional MIA-QSAR model [10] and similar behaviour is observed (see Table 2). Nonetheless, it should be noted that aug-MIA-QSAR models offer an important advantage as they may be more interpretable in chemical/structural terms, given the direct relationship between the considered atomic properties and the defined colour schemes. Using the procedure described in the sub-section **3.2**, a row vector **b** of PLS regression coefficients with dimension 1 × 102,600 was obtained, where the value of each vector component indicates the importance of the corresponding predictor (pixel coordinate) in increasing or attenuating the modelled bioactivity. These PLS regression coefficients have been previously used in omics-type data modelling to determine the relevance of single variables with respect to individual responses, as a strategy for variable selection [13,14]. The vector **b** is independent of algorithmic details such as the procedure followed in the normalisation of the weights, scores and loadings, and is thus directly comparable to the regression coefficients obtained in other methods [15]. In order to relate the importance of the pixel coordinates with the functional groups and/or atom types they represent, **b** was reshaped to a 342 × 300 array (the initial size of each chemical image in pixels) yielding the 2D-contour map in Figure 3.

As can be observed in Figure 3, variables corresponding to substituent positions 3 and 5 possess strong positive correlation with the anti-HIV 1 activity. This implies that high values for r_{vdW}/ε (corresponding large van der Waals radius/size and/or low electronegativity) are favourable for these positions. Analysing Table 3, it is inferred that bromine, chlorine and methyl (carbon) groups favour an increase in bioactivity. Steric effects have been cited in the literature to explain the high activity observed with bromine and chlorine (i.e. bulky groups in this position favour activity), while the high activity of the methyl group was attributed to hydrophobic interactions, observed with a COMSIA-based model (see ref. [16]). Note that while the data-set employed in the present study does not contain sulphur groups (e.g. –SH, $–SCH_3$, etc.) in the substituent positions 3 and 5, the high r_{vdW}/ε value for sulphur suggests these to be an interesting moiety to evaluate for anti-HIV 1 activity. Additionally, from Figure 3 it is observed that bonding oxygen to the sulphur atom constituting the sulphide bridge (orange subspace A), typical of sulfinyl and sulfonyl compounds, is related with an increase in bioactivity. Even more interesting is the observation that the aggregation of a second oxygen atom (red–brown subspace B), as in sulfonyl moieties, is related to a much higher influence in enhancing the anti-HIV 1 activity. Therefore, the influence of the sulphur functional groups in enhancing the modelled bioactivity is sorted in the following order: $–SO_2– > –SO– > –S–$. This result is consistent with previous theoretical (molecular dynamics) and experimental (*in vitro*) studies

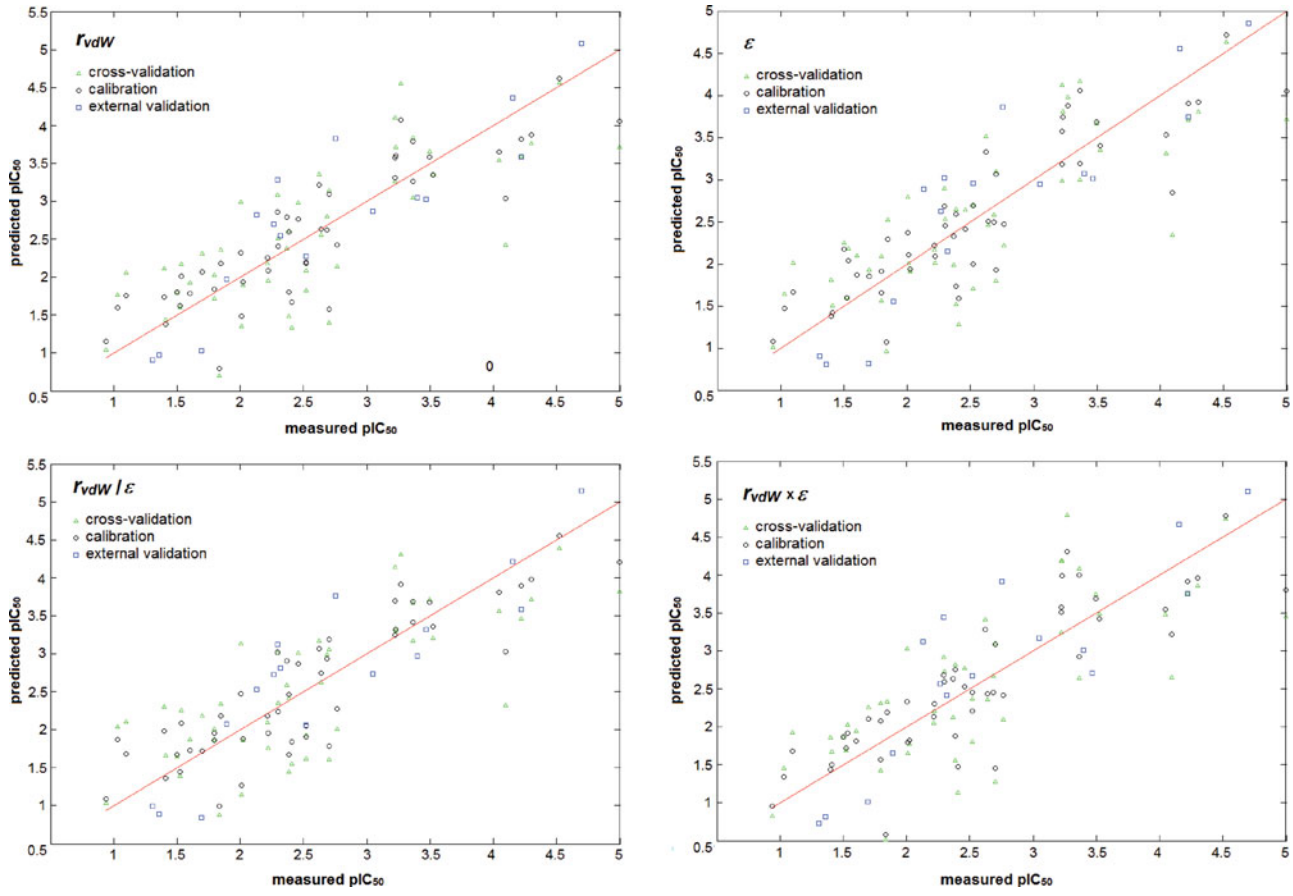


Figure 2. Plots of experimental × predicted pIC_{50} values using the MIA-QSAR models obtained from atomic colours proportional to r_{vdW}, ε, $r_{vdW} \times \varepsilon$ and r_{vdW}/ε.

reported in the literature [16,17]. X-ray crystallography revealed that the second oxygen in the sulfonyl moiety helps to maintain the side chain of the binding site Tyr181 of the HIV-1 reverse transcriptase in the correct position for tight binding and thus enhancing the inhibitory activity [17].

On the other hand, an inverse relationship is observed for substituents in position 4 in the sense that high r_{vdW}/ε values in substituent position 4 are related to a decrease in biological activity (see light blue subspace in position 4 of Figure 3). For instance, in the sulfonyl compounds with substituents in position 4, the bioactivity decreases in this order (see Table 1): –Cl (compound **44**) > –Me (compound **41**) > –Br (compound **47**) > –OMe (compound **38**) > –CN (compound **52**), while according to Bond's method for the computation of the van der Waals volume (cm^3/mol) a reverse sequence is observed for these substituent groups [i.e. –Cl (11.62) < –Me (13.67) < –Br (14.40) < –CN (14.70)] [18]. This result is consistent with a previous CoMFA-based study where it was

Table 3. Atomic properties considered to colour the atoms in the MIA-QSAR models. An example of RGB composition is given to show how atoms were coloured according to the respective r_{vdW}/ε values.[a]

Atom	r_{vdW}	ε	$r_{vdW} \times \varepsilon$	r_{vdW}/ε	R	G	B	Total $\propto r_{vdW}/\varepsilon$
H	0.370	2.100	0.777	0.176	58	58	60	176 (dark grey)
C	0.910	2.500	2.275	0.364	121	121	122	364 (grey)
N	0.740	3.000	2.220	0.247	0	0	247	247 (blue)
O	0.920	3.500	3.220	0.263	255	8	0	263 (red)
F	0.710	4.000	2.840	0.178	18	80	80	178 (olive)
S	0.990	2.500	2.475	0.396	220	176	0	396 (yellow)
Cl	1.140	3.000	3.420	0.380	75	255	50	380 (green)
Br	1.270	2.800	3.556	0.454	254	0	200	454 (pink)

[a] The total pixel values for the chemical bonds and blank spaces were set to zero. The RGB composition is an example to assign an specific number to each atom.

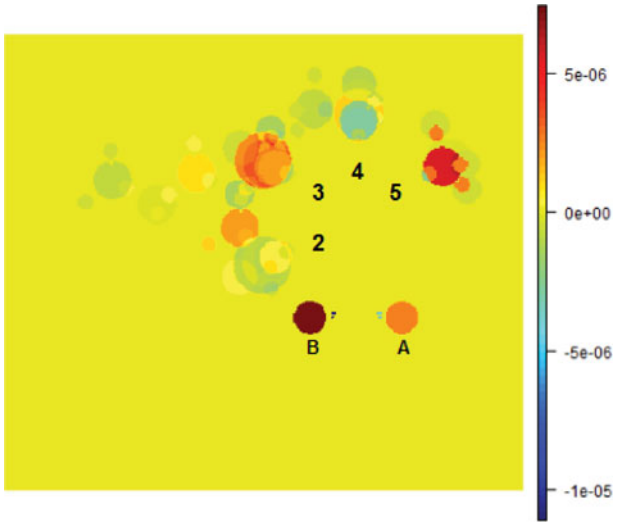


Figure 3. (Colour online) 2D-contour map obtained from regression coefficient analysis of the MIA-QSAR model based on atomic colours proportional to r_{vdW}/ε. The colour scale indicates the contribution (either negative or positive) of each descriptor for the bioactivity of 2-amino-6-arylsulfonylbenzonitriles and derivatives as anti-HIV-1 compounds.

demonstrated that the afore-mentioned bulky groups in the substituent position 4 did not favour the anti-HIV 1 activity, which validates the 2D-MIA-QSAR contour maps [16].

3. Computational methods

3.1. Data-set construction and validation procedures

A data-set of 64 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl derivatives with the corresponding anti-HIV-1 IC_{50} values, was obtained from the literature [17]. The chemical structures for these compounds were drawn using the GaussView program [19] and saved separately as bitmap files of 342 × 300 pixels size, keeping the common substructure aligned. This procedure yielded an MVI of dimension 64 × 342 × 300. The chemical structures were drawn by considering atoms as spheres with sizes proportional to the corresponding van der Waals radii. In addition, the atom types were assigned dissimilar colours to distinguish them, since different numbers are allocated to each colour pixel, in agreement with the RGB system of colours. According to the RGB model, the entire colour spectrum is obtained from the contribution of red (255), green (255) and blue (255) components, thus varying from 0 (black, absence of colour) to 765 (white, the sum of all three original components). The overall colour values applied to each atom were made proportional to the corresponding van der Waals radii (r_{vdW}), electronegativity (ε), $r_{vdW} \times \varepsilon$ and r_{vdW}/ε values (see Table 3), respectively. It is worth mentioning that atomic colors*per se* do not matter, but rather the numerical data specified for each atom. Each image was converted to numbers and unfolded to form a row vector, yielding, therefore, a 64 row data matrix. Given that every pixel coordinate corresponds to a variable, the MIA-QSAR approach yields a high dimensionality matrix ($p >> n$) and, therefore, PLS was used as the statistical technique to build regression models for the four colour schemes defined. The chemical data-set was split into training (48 compounds) and test sets (16 compounds) with the observations in each set maintained identical to a previous study [10]. The MIA-QSAR models were validated using the leave-one-out cross-validation and external validation procedures. Other measures considered in the assessment of the quality of the built models include: the determination coefficient between actual and predicted pIC_{50} values (q^2 and r^2_{test}), root mean square error of prediction (RMSECV and RMSEP) and the modified r^2_{test} ($r_m{}^2$) parameter, according to the criteria established in the literature [20]. In addition, the reliability of the models was attested using the y-randomisation test [analysed in terms of the corrected penalised r^2 ($^cr_p{}^2$)] [21], in which the y-block is shuffled and regression performed to verify the inexistence of chance correlation. The image treatment and statistical analysis were performed using the Chemoface program [22].

3.2. Partial least squares for regression

The PLSR aims at constructing LVs that maximise the correlation (or covariance) between the matrix **X** and the vector **y**, respectively, where **y** is an $n \times 1$ response vector. Given the cross product vector $\mathbf{X}^T\mathbf{y}$, the LVs are extracted to correspond to the direction of most variation (characterised by vector **w**). The projections of **X** on **w** and denominated as **X**-scores are designated by **t**. As for the **X**- and **y**-loadings, these are computed by regressing the scores matrix **T** against **X** and **y**, respectively. However, the vectors **w** for successive LVs are not directly comparable as they are obtained from sequentially deflated **X** and **y**. To achieve comparability, the weights **w** are related to **X**, using the expression $\mathbf{R} = \mathbf{W}\,(\mathbf{P}^T\mathbf{W})^{-1}$, where **P** is the loadings matrix for **X**. It may also be verified that the scores matrix $\mathbf{T} = \mathbf{XR}$. Finally, regressing **y** against **T** yields the regression coefficients vector **c**, of dimension $1 \times l$ vector (where l is the number of LVs), and the relevance of each of the original variables to **y** is given by the regression coefficients $\mathbf{b} = \mathbf{cR}$, where **b** is a $1 \times p$ vector, **c** is a $1 \times l$ vector and **R** is an $l \times p$ matrix, respectively. For details see refs. [13–15]

The 2D-contour maps were obtained according to the following procedure: (1) given the three LV aug-MIA-QSAR model selected for the interpretation experiment, the corresponding regression coefficients row vector **b**, with dimension 1 × 102,600, was determined using the SIMPLS algorithm [23]; (2) the vector **b** was refolded to a 342 × 300 array (the dimension of each bitmap image corresponding to the molecules);(3) a colour scale for the **b** coefficients was defined to allow for a visual graphical analysis of the substituents/atom types related with an increase or decrease in the anti-HIV1 activity and the magnitude of this influence (in terms of the absolute value of the coefficients).

4. Conclusions

The fundamental goal of introducing colour schemes in the MIA-QSAR framework according to pre-defined atomic properties (or relations) was to enhance the utility of this method in the sense that more chemically meaningful models would be obtained. However, the use of dimensionality reduction techniques like PLS in the modelling of MIA-QSAR data matrices presented another challenge as the identity of the original features is masked in the orthogonal projections. The PLS regression coefficients enable the assessment of the relevance of single predictors to the modelled responses in high dimensionality data matrices. Since each variable in the MIA-QSAR method corresponds to a particular pixel coordinate, the transformation of the regression coefficients into 2D-contour maps allows for the direct analysis of the structural (functional groups/atom types), electronic and physicochemical properties affecting the modelled bioactivity. Consequently, direct interpretation of the built MIA-QSAR models is achieved and thus greatly augmenting the practical utility of this method in the screening and design of new molecules of interest in drug therapy, material sciences and agrochemistry. The MIA-QSAR method is much simpler to operate than nD-QSAR methods ($n \geq 3$) and involves a low computational cost. Further tasks include implementation of the 2D-contour maps functionality as a routine script in the Chemoface software [22].

Acknowledgments

The authors are thankful to FAPEMIG and CNPq for the financial support and fellowships (to S. J. Barigye and M. P. Freitas).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Fundação de Amparo à Pesquisa do Estado de Minas Gerais, Conselho Nacional de Desenvolvimento Científico e Tecnológico

References

[1] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, and A. Tropsha, J. Med. Chem. **57**, 4977 (2014).
[2] C. Hansch and T. Fujita, ρ-σ-π Analysis, J. Am. Chem. Soc. **86**, 1616 (1964).
[3] R.D. Cramer, D.E. Patterson, and J.D. Bunce, J. Am. Chem. Soc. **110**, 5959 (1988).
[4] M.P. Freitas and T.C. Ramalho, Cienc. Agrotecnol. **37**, 485 (2013).
[5] E. Estrada, E. Molina, and I. Perdomo-López, J. Chem. Inf. Comp. Sci. **41**, 1015 (2001).
[6] M.P. Freitas, S.D. Brown, and J.A. Martins, J. Mol. Struct. **738**, 149 (2005).
[7] C.A. Nunes and M.P. Freitas, Eur. J. Med. Chem. **62**, 297 (2013).
[8] M.R. Freitas, S.J. Barigye, and M.P. Freitas, RSC Adv. **5**, 7547 (2015).
[9] J.G. Topliss and J.D. Costello, J. Med. Chem. **15**, 1066 (1972).
[10] M.P. Freitas, Org. Biomol. Chem. **4**, 1154 (2006).
[11] A. Golbraikh and A. Tropsha, J. Mol. Graph. Model. **20**, 269 (2002).
[12] A. Tropsha, Mol. Inf. **29**, 476 (2010).
[13] G. Palermo, P. Piraino, and H.D. Zucht, Adv. Appl. Bioinform. Chem. **2**, 57 (2009).
[14] I.G. Chong and C.H. Jun, Chemometr. Intell. Lab. Syst. **78**, 103 (2005).
[15] R. Wehrens, *Chemometrics with R. Multivariate Data Analysis in the Natural Sciences and Life Sciences* (Springer, Heidelberg, 2011).
[16] J.H. Chan, J.S. Hong, R.N. Hunter III, G.F. Orr, J.L. Cowan, D.L. Sherman, S.M. Sparks, B.E. Reitter, C.W. Andrews III, R.J. Hazen, M. St. Clair, L.R. Boone, R.G. Ferris, K.L. Chech, G.B. Roberts, S.A. Short, K. Weaver, R.J. Ott, J. Ren, A. Hopkins, D.I. Stuart, and D.K. Stammers, J. Med. Chem. **44**, 1866 (2001).
[17] R. Hu, F. Barbault, F. Maurel, M. Delamar, and R. Zhang, Chem. Biol. Drug Des. **76**, 518 (2010).
[18] Y.H. Zhao, H.A. Michael, and M.Z. Andreas, J. Org. Chem. **68**, 7368 (2003).
[19] R.D. Dennington, T.A. Keith, and J.M. Millam, *GaussView 5.0* (Gaussian, Inc., Wallingford, CA, 2008).
[20] K. Roy, P. Chakraborty, I. Mitra, P.K. Ojha, S. Kar, and R.N. Das, J. Comp. Chem. **34**, 1071 (2013).
[21] I. Mitra, A. Saha, and K. Roy, Mol. Simul. **36**, 1067 (2010).
[22] C.A. Nunes, M.P. Freitas, A.C.M. Pinheiro, and S.C. Bastos, J. Braz. Chem. Soc. **23**, 2003 (2012).
[23] S. De Jong, Chemometr. Intell. Lab. Syst. **18**, 251 (1993).