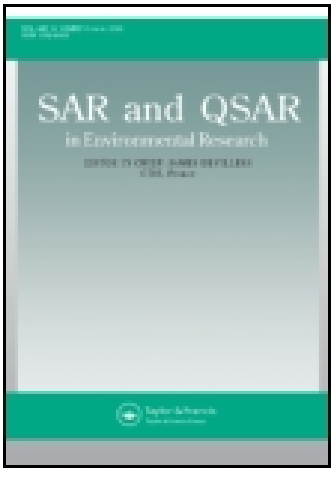


SAR and QSAR in Environmental Research

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gsar20

Computational modelling of the antischistosomal activity for neolignan derivatives based on the MIA-SAR approach

M.H. Duarte[a], S.J. Barigye[a], E.G. da Mota[a] & M.P. Freitas[a]

[a] Department of Chemistry, Federal University of Lavras, Lavras, Minas Gerais, Brazil

Published online: 16 Mar 2015.

SAR and QSAR in Environmental Research, 2015
Vol. 26, No. 3, 205–216, http://dx.doi.org/10.1080/1062936X.2015.1018942

Computational modelling of the antischistosomal activity for neolignan derivatives based on the MIA-SAR approach

M.H. Duarte, S.J. Barigye, E.G. da Mota and M.P. Freitas*

Department of Chemistry, Federal University of Lavras, Lavras, Minas Gerais, Brazil

(*Received 27 November 2014; in final form 21 January 2015*)

Theoretical models for exploring the antischistosomal activity of a dataset of 18 synthetic neolignans are built using the multivariate image analysis applied to structure–activity relationships (MIA-SAR) approach. The obtained models were validated using the accuracy (Acc) in leave-one-out cross-validation, external validation and Y-randomization procedures, yielding correct classification superior to 80%, 70% and 60%, respectively. Additionally, a comparison was made of the models obtained from binary (black and white) and coloured images; the colours (pixel values) were selected to correspond to chemical properties. It was observed that the models obtained from coloured images with pixel values corresponding to electronegativity (known as the aug-MIA-SAR$_{colour}$ approach) generally yielded superior statistical parameters compared with those obtained from binary images (MIA-SAR) and randomly coloured images (atoms are coloured according to their type) with atomic sizes corresponding to Van der Waals radius (aug-MIA-SAR), respectively. Mechanistic interpretation of the influence of different substituents on the antischistosomal activity revealed that methoxy substituents in the R1 (or R2) and R5 positions of the neolignan scaffold are indispensable for the antischistosomal activity. The obtained results provide knowledge of the possible structural modifications to yield novel neolignan compounds with antischistosomal activity.

Keywords: *Schistosoma mansoni*; PLS-DA; PCA; HCA; MIA-SAR

1. Introduction

Schistosomiasis is a neglected disease caused by helminths of the genus *Schistosoma*. Endemic in 78 countries worldwide, the number of infected people in 2012 was 249 million, of which 45% were children of aged 5–14 [1]. Schistosomiasis has been mainly controlled using Praziquantel, a drug known to be active against all species of *Schistosoma* [2]. However, in view of the persistent and prolonged use of this drug (for over 40 years) in the control and prevention of *Schistosoma mansoni*, the risk of parasite resistance is eminent [3]. Additionally, Praziquantel does not prevent the recurrence of the disease [4], it is inactive against juvenile schistosomes and has limited effect in patients with advanced lesions in the liver and spleen. It is thus important to seek for more efficient drugs active against all stages in schistosomiasis therapy [5,6].

In a previous study on phytochemicals of therapeutic interest, it was shown that neolignans extracted from leaves of the plant *Virola surinamensis* exhibit biological activity against several microorganisms, such as bacteria [7], fungi [8] and other parasites, such as

*Corresponding author. Email: matheus@dqi.ufla.br

Leishmania spp. [9,10] *Trypanosoma cruzi* [11] and *S. mansoni* [12,13]. *In vivo* studies using mice as model organisms revealed that neolignans isolated from the leaves of *V. surinamensis* possess high activity against the penetration of cercariae of *S. mansoni*. Motivated by this finding, studies on several neolignan derivatives have been performed with the aim of gaining greater understanding of their activity mechanisms and structure–activity relationships (SARs) in order to suggest strategies for the structure optimization to yield more effective and broad spectrum drugs.

In recent times, computational methods have emerged as useful alternatives in the drug discovery process, with their key advantages being the reduced time and cost incurred in characterizing chemical structures and/or their interaction with other ligands/target sites. Discussions on the different approaches are beyond the scope of the present report (for details see ref. [14]). Nonetheless, it is important to highlight that models derived from topological or topochemical (two-dimensional) configurations of chemical structures have surprisingly yielded good results in correlation studies comparable and sometimes superior to more complex methods, such as molecular dynamics, quantum computing and molecular field analyses of chemical structures. In this sense, our research group has successfully used the multivariate image analysis applied to quantitative structure–activity relationships (MIA-QSAR) methodology for a decade in correlation studies with results comparable with more sophisticated QSAR techniques, such as comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) [15,16]. In this method, chemical structures are represented as multivariate images, aligned in reference to the basic scaffold in the series of congeneric compounds and the pixels in each image are used as chemical descriptors. The reasoning behind this approach is that the progressive variation in the biological activity of a series of compounds is a function of the substituents bonded to the main scaffold.

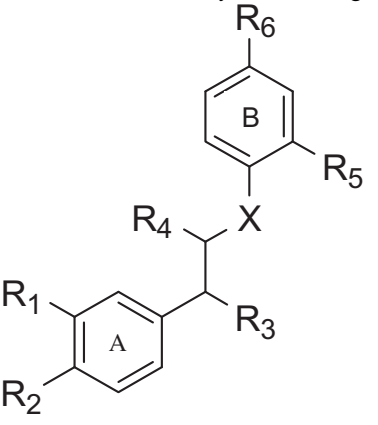
Previous applications of the MIA-QSAR method were based on binary grey scale images (black and white) [14,17]. However, with the aim of incorporating greater chemical information to the molecular structural images (e.g. steric and electronic properties), extensions of the traditional MIA-QSAR method have been performed. In this sense, two colour schemes denominated as 'aug-MIA-QSAR' and 'aug-MIA-QSAR$_{colour}$' have been defined to codify useful chemical information. In the former, atom types are randomly assigned dissimilar colours and their sizes designed to be proportional to the Van der Waals radius, while in the latter, the colours are carefully selected so that their pixel values (according to the RGB system) possess perfect correlation with Pauling's electronegativity for atoms. Note that solid colours are employed in both schemes to guarantee proper correspondence with the selected atomic property [17,18]. It is anticipated that the proposed extensions aid in yielding MIA-QSAR models with greater interpretability and predictive power.

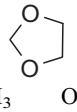
The objective of the present report is to build classification models for a series of neolignans based on the extended MIA-SAR methodology and subsequently employ these models in the study of the intrinsic structural characteristics critical in their antischistosomal activity.

2. Material and methods

A dataset consisting of 18 neolignans with reported activity against *S. mansoni* cercaria in rats was employed (see Table 1) [19]. The neolignans are dimers formed from oxidative coupling of allyl and propenil phenols that occur in the Myristicaceae family of plants. The neolignans employed in this paper are those of the type 8.O.4′, with an O bridge between carbons 8 and 4 [20], where five have been reported to be active (5, 6, 8, 9, 17) and 13 are reported to be inactive (1, 2, 3, 4, 7,10, 11,12, 13, 14, 15, 16, 18) to *S. mansoni*.

Table 1. Chemical structure and antischistosomal activity of the neolignan compounds.



Compound	*Activity*	*R1*	*R2*	*R3*	*R4*	*R5*	*R6*	*X*
1	Inactive	H	H	=O	H	H	H	O
2*	Inactive	H	H	=O	CH_3	H	H	O
3	Inactive	H	H	=O	CH_3	OCH_3	H	O
4	Inactive	H	H	=O	CH_3	OCH_3	$CH_2CH{=}CH_2$	O
5*	Active	OCH_3	OCH_3	=O	CH_3	OCH_3	$CH{=}CHCH_3$	O
6	Active	OCH_3	OCH_3	=O	CH_3	OCH_3	$CH_2CH{=}CH_2$	O
7	Inactive	H	H	OH	CH_3	OCH_3	H	O
8	Active	OCH_3	OCH_3	OH	CH_3	OCH_3	$CH_2CH{=}CH_2$	O
9	Active	OCH_3	OCH_3	OH	CH_3	OCH_3	$CH{=}CHCH_3$	O
10*	Inactive	H	H	=O	CH_3	OCH_3	H	S
11	Inactive			=O	CH_3	OCH_3	H	S
12	Inactive	OCH_3	OCH_3	=O	CH_3	H	H	S
13	Inactive	OCH_3	OCH_3	=O	CH_3	H	CH_3	S
14	Inactive	H	H	=O	H	H	H	NH
15	Inactive	OCH_3	OCH_3	=O	CH_3	H	H	NH
16*	Inactive	OCH_3	OCH_3	=O	CH_3	CH_3	H	NH
17	Active	OCH_3	OCH_3	=O	CH_3	OCH_3	H	NH
18	Inactive	OCH_3	OCH_3	=O	CH_3	H	H	NCH_3

*Compounds selected for test set.

2.1 *Preparation of chemical images and data matrix*

For the traditional MIA-SAR based approach, the chemical structures were drawn using the ChemSketch program and each image was aligned in a space with predefined dimensions (300 × 300 pixels) in the Microsoft Windows Paint program and saved as bitmaps (bmp) files (see the supplementary material which is available via the multimedia link on the online article webpage). In the aug-MIA-SAR model, the chemical structures were designed in GaussView 5.0 [21], using spheres with radii proportional to the Van der Waals radii of atoms and random solid colours to represent different atoms. The images were aligned and saved following the same procedure as in the traditional MIA-SAR model. The alignment of the two-dimensional (2D) molecular images in the MIA-SAR approach is a simple manual procedure, where a common pixel for the compound series is fitted at the same coordinate in the

workspace and the structures are consequently aligned with respect to this coordinate. This procedure guarantees the superposition of the basic structural scaffold. Finally, for the aug-MIA-SAR$_{colour}$ model, similar steps were followed as in the aug-MIA-SAR approach, with a particularity that the solid colours were selected to be proportional to the Pauling's electronegativity scale (see the supplementary material which is available via the multimedia link on the online article webpage).

After aligning the images for the 18 chemical structures, the ensuing multivariate image structure was unfolded to yield a bi-dimensional data matrix. Subsequently, the zero variance columns (variables) were deleted to yield matrices with dimensions: 18 × 1149, 18 × 5242 and 18 × 5349 for the MIA-SAR, aug-MIA-SAR and aug-MIA-SAR$_{colour}$ approaches, respectively.

The chemical image and data matrix processing were performed using an in-house program denominated Chemoface [22]. This is a free user-friendly program for chemical image processing and chemoinformatic analysis downloadable from http://ufla.br/chemoface/.

2.2 *Principal component analysis and hierarchical cluster analysis*

Principal component analysis (PCA) is an unsupervised data analysis method aimed at stratifying information contained in a data matrix into latent variables (known as principal components), which are orthogonal in nature. The first factor explains the highest dataset variance and each successive component seeks to maximize the variance not explained by the preceding factor and thus obtain a lower dimensional data structure. It follows that variables or instances strongly loaded in the same factor are highly correlated, while those loaded in different factors are uncorrelated. In this report, PCA was performed to analyse the intrinsic differences/similarities of the chemical structures, using the instance score plots. The eigenvalue criterion ($\lambda > 1.00$) was used to determine the number of important principal components.

As a complementary study, hierarchical cluster analysis (HCA) was performed to acquire greater understanding of the underlying structure of the dataset matrix. Note that although HCA also belongs to the unsupervised data analysis methods, it possesses two major differences from PCA. First, HCA employs distance metrics (e.g. Euclidean distance) as the similarity/dissimilarity measures and thus does not entail an *a priori* assumption of a linear relationship between the variables/instances, while the principal components are derived from the correlation/covariance matrix, which implicitly suggests a linear relationship. Second, because the amalgamation algorithms do not seek to maximize the dataset variance in a determined cluster, but rather to maximize the inter-cluster distance, HCA yields a more spread structure, which enables the detection of the natural tendency of the data structure, as well as the detection on anomalous behaviour.

In this study, complete linkage and the Euclidean distance were employed as the union criteria and similarity function, respectively. The distance value corresponding to steepest ascent in the amalgamation schedule was considered as cut-off to determine the natural number of agglomerations.

2.3 *Partial least squares-discriminant analysis (PLS-DA)*

PLS-DA is a useful tool in classification tasks when dealing with high dimensionality data matrices comprised of extremely correlated variables [23,24]. This method is based on the transformation of the original variables in latent variables, and these are linearly independent. These latent variables are linear combination of original variables. However, a previous study

by Topliss and Edwards [25] demonstrated that, when there are few original variables in the high dimensional matrix, which are in fact relevant for the modelled property, PLS is more prone to Type I errors (omitting structural features important for the Y-block). Therefore, bearing in mind that the MIA-SAR method is centred on the synthesis that the classification of compounds as active/inactive is a function of chemical moieties in particular positions of the basic scaffold in congeneric series of compounds, the risk of Type I errors is largely possible. In this sense, dimension reduction procedures were carried out on the original data matrix, using the information theoretic parameter, information gain (IG). Details of this method are reported elsewhere [26]. Afterwards, PLS-DA was applied to the reduced data matrix for the traditional MIA-SAR, aug-MIA-SAR and aug-MIA-SAR$_{colour}$ approaches, respectively, using 14 (77.8%) compounds in training series and 4 (22.2%) in the test series, randomly chosen.

The quality of the obtained classification models was evaluated considering the parameters: correct classification (accuracy) in prediction [Acc(p)]; leave-one-out cross validation [Acc(CVloo)]; y-randomization (mean of 10 repetitions); and external validation [Acc(ext)]. In the CV_{loo} procedure, a sample is taken at random from the dataset and a model is developed with the $(n - 1)$ remaining samples. Then, a prediction of the excluded sample is performed. This process is continued until n validated models are obtained. The accuracy the classification model is then determined as the average performance for the n validated models. A limitation of this cross-validation procedure is that it alone does not provide criteria on the earnest predictive power of the estimated model and, therefore, validation on a set of compounds not employed in the model building (i.e. external validation) is necessary. Bearing in mind that few data points are employed in the present study, a fivefold random split of the dataset into training and test sets was performed and the average accuracy reported.

One of the key limitations of PLS-DA as a statistical classification method is interpretability, given that the original variables are transformed into latent variables to maximize the communality of the independent and dependent variable blocks, thus 'masking' the original information contained in the variables. However, it is known that PLS-DA and linear discriminant analysis (LDA) yield the same result when the number of latent variables considered in PLS-DA is equal to the number of original variables. Therefore, in order to yield more interpretable models, small subsets of orthogonal variables were extracted from the reduced data matrices following a stepwise variable selection algorithm in which correlation with the classification variable is maximized, while multicollinearity among independent variables is minimized. It is anticipated that models built on this subset of orthogonal variables should provide greater insight on the influence of different substituents on the antischistosomal activity.

3. Results and discussion

3.1 *Principal component analysis*

Figure 1 shows the score plots for the first two principal components (PC1 and PC2) for MIA-SAR, aug-MIA-SAR and aug-MIA-SAR$_{colour}$ strategies, explaining 63.14%, 72.52% and 66.29% of the data variance, respectively. As can be observed, with the exception of compound 17, the three score plots achieve data class delineation of active neolignans from the inactives (i.e. the active compounds are viewed to be clustered together).

The key difference between molecule 17 and the cluster of active compounds (C1) lies in substituent positions X (O for NH) and R6 (propenyl group for H), while the inactive compounds in cluster C2 (12, 13, 15, 16, 18) differ in their totality from molecule 17 only in

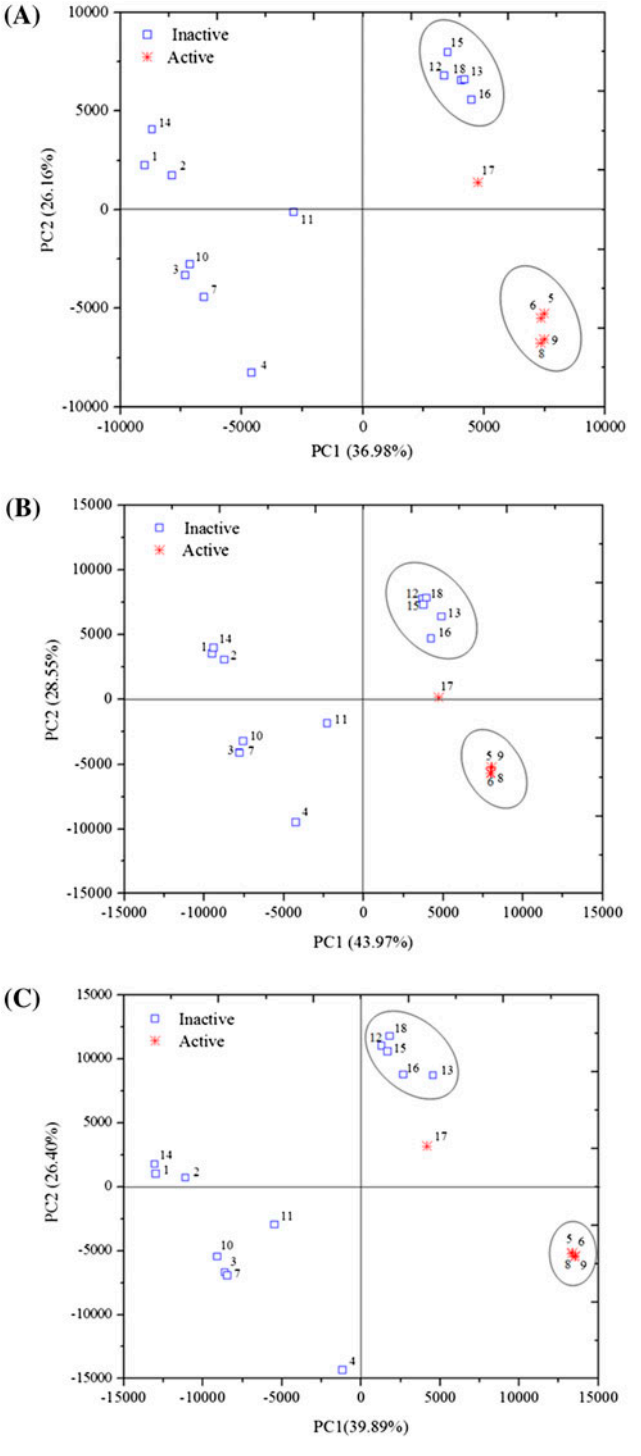


Figure 1. Graph of PCA series of neolignans using (A) traditional MIA-SAR, (B) aug-MIA-SAR and (C) aug-MIA-SAR$_{colours}$.

substituent position R5 (none possesses a methoxy group). Note the greater proximity of molecule 17 to C2 rather than C1, which suggests greater structural resemblance to the former. This trend yields three important inferences: (1) isosteric replacement in substituent position X seems to have no effect on antischistosomal activity, i.e. it shifts molecule 17 away from C1 but activity is maintained; (2) the methoxy group in substituent position R5 is important for antischistosomal activity, i.e. while all inactive compounds in cluster C2 have methoxy groups in R1 and R2, their only global dissimilarity with molecule 17 lies with an additional methoxy group in R5 for the latter, while all members of C1 possess the R5 methoxy substituent; and (3) replacement of the propenyl group in R6 with H has no deleterious effect on the antischistosomal activity. It is important to note that given that PC1 and PC2 account for more than 63% of the variance in all cases, the original variables (>1000) appear to be highly correlated. This is in fact a common characteristic of data matrices obtained with the MIA-SAR method. As a consequence, PCA results should be analysed with precaution.

3.2 *Hierarchical cluster analysis*

Figure 2 shows the dendograms obtained with PCA according to the MIA-SAR, aug-MIA-SAR and aug-MIA-SAR$_{colour}$ strategies, respectively.

Considering cut-off distances delineated by the vertical lines in the dendograms, three clusters are obtained for each of the MIA-SAR strategies. An analysis of these clusters reveals that with the exception of compound 17, the active compounds are clustered together in the dendograms obtained according to the MIA-SAR, aug-MIA-SAR and aug-MIA-SAR$_{colour}$ strategies, respectively, and thus corroborating the result obtained with PCA. Additionally, the dendograms for MIA-SAR (Figure 2A) and aug-MIA-SAR (Figure 2B) suggest atypical structural characteristics for compound 11, which is a logical result given that it possesses a unique substituent (1,3-dioxolanyl group) in substituent positions R1 and R2. However, the dendogram for aug-MIA-SAR$_{colour}$ approach indicates atypical behaviour for compound 4, which may be attributed to the combined presence of the propenyl group in R6 and H-atoms in R1 and R2, since compounds with H substituents in R1, R2 and R6 do not exhibit atypical behaviour, neither do those with a propenyl group in R6 and methoxy groups in R1 and R2. The results obtained with PCA generally corroborate the inferences obtained with PCA on the structural characteristics necessary for the antischistosomal activity for the neolignans.

3.3 *Partial least squares-discriminant analysis*

Table 2 shows the statistical parameters for the MIA-SAR, aug-MIA-SAR and aug-MIA-SAR$_{colour}$ models built using descriptor subsets obtained with the IG filter and the stepwise variable selection procedures, respectively.

As can be observed, the obtained models generally possess good statistical behaviour evidenced by the high percentages for correct classification (accuracy) in prediction [Acc(p)], internal [Acc(CV$_{loo}$)] and external cross-validation [Acc(ext)], respectively. Additionally, the low Y-randomization values suggest that the obtained models are not prone to the chance correlation phenomenon. In regard to the models obtained using the IG filtered subset and the stepwise variable selection procedure, these are observed to be generally comparable, although an improvement in accuracy for the external validation set is observed with the aug-MIA-SAR$_{colour}$ strategy (i.e. from 75% to 100%). Note that the key advantage of using a handful of variables is that the mechanistic interpretation of the built models is straightforward.

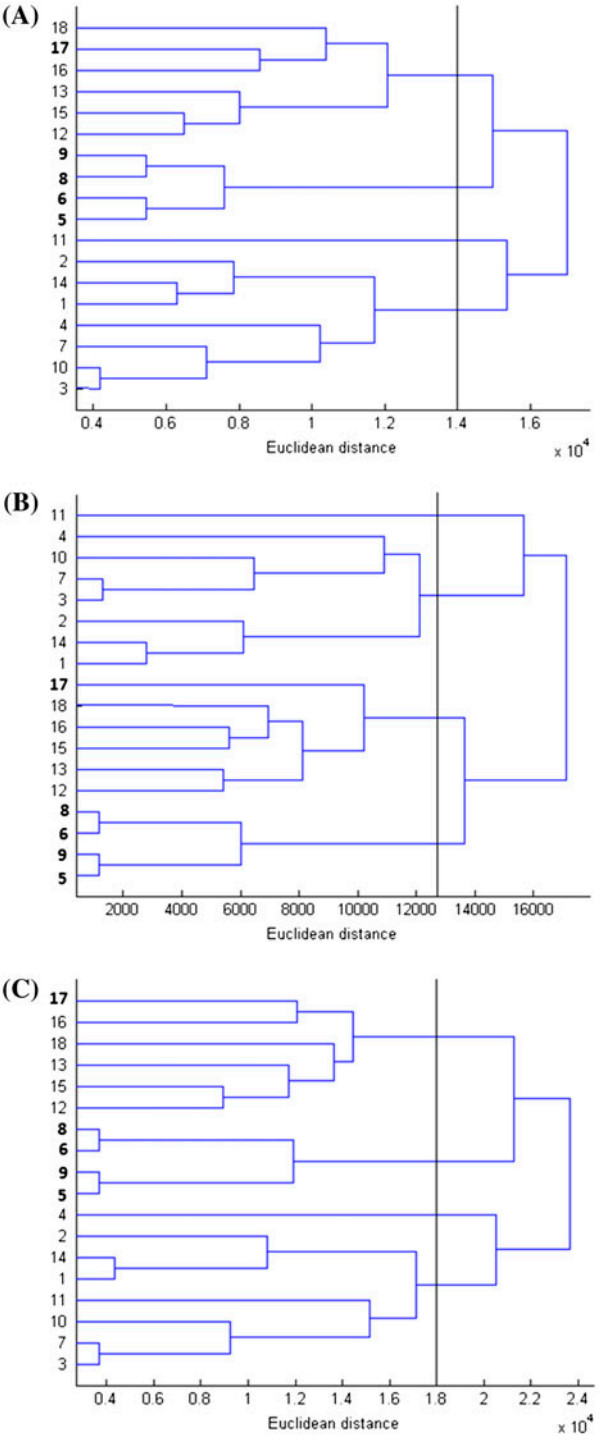


Figure 2. Dendrograms of (A) MIA-SAR, (B) aug-MIA-SAR(B) and (C) aug-MIA-SAR$_{colour}$ model for 18 neolignans.

Table 2. Statistical parameters for the MIA-SAR, aug-MIA-SAR and aug-MIA-SAR$_{colour}$ models.

IG filtered subset			
Parameter (%)[†]	MIA-SAR	aug-MIA-SAR	aug-MIA-SAR$_{colour}$
Acc(p)	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
Acc(Y-rand)	69.57 ± 8.25	70.54 ± 8.83	69.43 ± 8.16
Acc(CV$_{loo}$)	94.29 ± 5.35	96.43 ± 3.57	95.71 ± 3.50
Acc(ext)	90.00 ± 12.25	87.50 ± 12.50	90.00 ± 12.50
Stepwise variable selection			
Acc(p)	94.64 ± 3.09	89.29 ± 3.57	100.00 ± 0.00
Acc(Y-rand)	71.43 ± 0.00	68.04 ± 2.11	70.00 ± 0.71
Acc(CV$_{loo}$)	94.64 ± 3.09	82.14 ± 6.19	100.00 ± 0.00
Acc(ext)	93.75 ± 12.25	75.00 ± 0.00	100.00 ± 0.00

[†]Average values are reported.

However, if the three MIA-SAR approaches are compared, it is observed that the best results are obtained with the aug-MIA-SAR$_{colour}$ strategy, while comparable behaviour is observed between the MIA-SAR and aug-MIA-SAR approaches. The superior performance of the aug-MIA-SAR$_{colour}$ approach suggests greater relevance to the electronegativity (chemical property considered in determining the colours for the atoms) in determining the antischistosomal activity. In fact, this is in agreement with previous experimental reports that suggested the importance of electronic properties in the antischistosomal activity of neolignans [19].

In light of the superior performance obtained with the aug-MIA-SAR$_{colour}$ based model, an effort is made to explore the information codified by the constituent variables in terms of the influence of different atoms/functional groups on the modelled biological activity.

3.4 *Interpretation of the aug-MIA-SAR$_{colour}$ model*

The incorporation of colour schemes to the MIA-SAR approach allows for the straightforward identification of atoms represented in the variables and greater understanding of their influence on the modelled bioactivity. Table 3 shows the variables contained in the final models obtained according the MIA-SAR, aug-MIA-SAR and aug-MIA-SAR$_{colour}$ approaches, respectively. Here, mechanistic interpretation of the aug-MIA-SAR$_{colour}$ model is performed. This model is built from two orthogonal variables, obtained by stepwise variable selection, with each variable representing a particular pixel position in the multivariate image collage.

As can be observed, the variable X78 possesses pixel values corresponding to blank spaces (750) for the molecules (1–4, 7, 10–11, 14) and pixel values for bonds (612) for the rest of the molecules (i.e. 5–6, 8–9, 12–13, 15–18), which suggests that the blank spaces correspond to a pixel in the vicinity of small size substituents superposed with a pixel for bond linking the basic scaffold to bigger size substituents. A painstaking analysis of this pattern reveals that pixel X78 corresponds to the substituent R1/R2 (note that R1 and R2 are identical), where the smaller and bigger size substituents are hydrogen and the methoxy groups, respectively (see also Table 1). Interestingly, all the compounds without methoxy groups for this variable are all inactive. Note that the reverse is not all true, i.e. not all compounds with methoxy groups in this position are active. To obtain complementary information in this line,

Table 3. Selected variables for MIA-SAR, aug-MIA-SAR and aug-MIA-SAR$_{colour}$ models.

	MIA-SAR			*aug-MIA-SAR*			*aug-MIA-SAR$_{colour}$*	
Compound	*X1*	*X422*	*X596*	*X3520*	*X4996*	*X5237*	*X78*	*X4392*
1	765	0	765	750	750	750	750	210
2	765	0	765	750	750	750	750	210
3	765	0	765	750	426	612	750	350
4	765	0	0	426	426	612	750	350
5	0	765	0	426	426	612	612	350
6	0	765	0	426	426	750	612	350
7	765	765	765	750	426	612	750	350
8	0	765	0	426	426	612	612	350
9	0	765	0	426	426	612	612	350
10	765	0	765	750	426	612	750	350
11	765	0	765	750	426	750	750	350
12	0	0	765	750	750	750	612	210
13	0	0	765	750	750	750	612	210
14	765	0	765	750	750	750	750	210
15	0	0	765	750	750	750	612	210
16	0	765	765	750	612	750	612	250
17	0	765	765	750	426	750	612	350
18	0	765	765	750	750	750	612	210

the second variable, i.e. X4392, is analysed. This variable contains pixel values for hydrogen (210) for the following molecules (1–2, 12–15, 18), pixel values for oxygen (350) for the molecules (3–11, 17) and a pixel value for carbon (250) for molecule 16. A scrutiny of this pattern indicates that the pixel X4392 corresponds to the R5 substituent position.

This result suggests that antischistosomal activity is primarily determined by substituents in R1 (or R2) and R5 positions, where methoxy groups in R1 (or R2) and R5 yield active compounds, while replacement of any of the two eliminates the antischistosomal activity. In other words, it is not imperative for neolignans to possess methoxy groups in both the R1 and R2 positions to guarantee the desired bioactivity. While previous experimental studies have emphasized the importance of methoxy groups in the ring B of neolignans to the biological activity of neolignans [13], the expendability of one of the methoxy groups represents a novelty of the present study.

4. Conclusions

This work presents unsupervised (PCA and HCA) and supervised (PLS-DA) models for evaluating the antischistosomal activity of neolignans based on the MIA-SAR approach. The study reveals that the incorporation of colour schemes in the molecular structure images yields superior performance than when binary (black and white) images are used. Additionally, interpretation of the influence of different substituents on the antischistosomal activity has been performed, highlighting the importance of methoxy substituents in the R1 (or R2) and R5 positions of the neolignan scaffold in determining their bioactivity. The obtained results provide critical knowledge on the possible structural modifications that may be performed to yield novel neolignan compounds with antischistosomal activity.

Acknowledgements

Authors are thankful to FAPEMIG for the financial support of this research and for the studentship (M.H.D.), to CAPES for studentship (E.G.M.) and to CNPq for fellowships (to S.J.B. and M.P.F.).

References

[1] World Health Organization, *Schistosomiasis: Number of people receiving preventive chemotherapy in 2012*, Weekly Epidemiol, Rec. 89 (2014), pp. 21–28.
[2] D. Cioli and L. Pica-Mattoccia, *Praziquantel*, Parasitol. Res. 90 (2003), pp. S3–S9.
[3] W. Wang, L. Wang, and Y.-S. Liang, *Susceptibility or resistance of praziquantel in human schistosomiasis: A review*, Parasitol. Res. 111 (2012), pp. 1871–1877.
[4] D.P. McManus and A. Loukas, *Current status of vaccines for schistosomiasis*, Clin. Microbiol. Rev. 21 (2008), pp. 225–242.
[5] L.M. Silva and R. Menezes, S.A.d. Oliveira, and Z.A. Andrade, *Chemotherapeutic effects on larval stages of* Schistosoma mansoni *during infection and re-infection of mice*, Rev. Soc. Bras. Med. Trop. 36 (2003), pp. 335–341.
[6] S. Chatterjee, A. Mbaye, J.G. De Man, and E.A. Van Marck, *Does the neuropeptide somatostatin have therapeutic potential against schistosomiasis?* Trends Parasitol. 18 (2002), pp. 295–298.
[7] A.O. de Souza, J.B. Alderete, P.R.R. Minarini, and P. da Silva Melo, I. Ferreira, L.E.S. Barata, and C.L. Silva, *Structure activity relationship, acute toxicity and cytotoxicity of antimycobacterial neolignan analogues*, J. Pharm. Pharmacol. 63 (2011), pp. 936–942.
[8] S.A. Zacchino, S.N. López, G.D. Pezzenati, R.L. Furlán, C.B. Santecchia, L. Muñoz, F.A. Giannini, A.M. Rodríguez, and R.D. Enriz, In vitro *evaluation of antifungal properties of phenylpropanoids and related compounds acting against dermatophytes*, J. Nat. Prod. 62 (1999), pp. 1353–1357.
[9] L.E. Barata, L.S. Santos, P.H. Ferri, J.D. Phillipson, A. Paine, and S.L. Croft, *Anti-leishmanial activity of neolignans from* Virola *species and synthetic analogues*, Phytochemistry 55 (2000), pp. 589–595.
[10] M. Aveniente, E.F. Pinto, L.S. Santos, B. Rossi-Bergmann, and L.E. Barata, *Structure–activity relationship of antileishmanials neolignan analogues*, Bioorg. Med. Chem. 15 (2007), pp. 7337–7343.
[11] I. Nocito, M.V. Castelli, S.A. Zacchino, and E. Serra, *Activity of 8.O.4'-neolignans against* Trypanosoma cruzi, Parasitol. Res., 101 (2007), pp. 1453–1457.
[12] L.E.S. Barata, *Isolamento e sintese de neolignanas de* Virola Surinamensis *(Rol.) Warb.*, Instituto de Química, Universidade Estadual de Campinas, Brazil, 1976.
[13] L.D.S. Santos, *Síntese e atividade biologica de neolignanas 8.O.4' derivados e compostos correlatos*, Instituto de Química, Universidade Estadual de Campinas, Brazil, 1991.
[14] R. Todeschini and V. Consonni, *Molecular Descriptors for Chemoinformatics*, John Wiley & Sons, Weinheim, 2009.
[15] M.P. Freitas, *Multivariate image analysis applied to QSAR: Evaluation to a series of potential anxiolytic agents*, Chemom. Intell. Lab. 91 (2008), pp. 173–176.
[16] R.A. Cormanich, M. Goodarzi, and M.P. Freitas, *Improvement of multivariate image analysis applied to quantitative structure–activity relationship (QSAR) analysis by using wavelet–principal component analysis ranking variable selection and least-squares support vector machine regression: QSAR study of checkpoint kinase WEE1 inhibitors*, Chem. Biol. Drug Des. 73 (2009), pp. 244–252.
[17] C.A. Nunes and M.P. Freitas, *Introducing new dimensions in MIA-QSAR: A case for chemokine receptor inhibitors*, Eur. J. Med. Chem. 62 (2013), pp. 297–300.
[18] M.C. Guimarães, E.G. da Mota, D.G. Silva, and M.P. Freitas, *aug-MIA-QSPR modelling of the toxicities of anilines and phenols to* Vibrio fischeri *and* Pseudokirchneriella subcapitata, Chemom. Intell. Lab. Sys. 134 (2014), pp. 53–57.

[19] C.N. Alves, L.G. de Macedo, K.M. Honório, A.J. Camargo, L.S. Santos, I.N. Jardim, L.E. Barata, and A.B. da Silva, *A structure-activity relationship (SAR) study of neolignan compounds with anti-schistosomiasis activity*, J. Braz. Chem. Soc. 13 (2002), pp. 300–307.
[20] S. Apers, A. Vlietinck, and L. Pieters, *Lignans and neolignans as lead compounds*, Phytochem. Rev. 2 (2003), pp. 201–217.
[21] R. Dennington, T. Keith, and J. Millam, GaussView, version 5, Gaussian Inc, Wallingford, 2009.
[22] C.A. Nunes, M.P. Freitas, A.C.M. Pinheiro, and S.C. Bastos, *Chemoface: A novel free user-friendly interface for chemometrics*, J. Braz. Chem. Soc. 23 (2012), pp. 2003–2010.
[23] S. Chevallier, D. Bertrand, A. Kohler, and P. Courcoux, *Application of PLS-DA in multivariate image analysis*, J. Chemom. 20 (2006), pp. 221–229.
[24] J.M.M. Neto and G.C. Moita, *Uma introdução à análise exploratória de dados multivariados*, Quím. Nova 21 (1998), pp. 467–469.
[25] J.G. Topliss and R.P. Edwards, *Chance factors in studies of quantitative structure-activity relationships*, J. Med. Chem. 22 (1979), pp. 1238–1244.
[26] L. Yu and H. Liu, *Efficient feature selection via analysis of relevance and redundancy*, J. Mach. Learn. Res. 5 (2004), pp. 1205–1224.