International Journal of Quantitative Structure-Property Relationships
Volume 1 • Issue 1 • January-June 2016

Ten Years of the MIA-QSAR Strategy: Historical Development and Applications

Stephen Jones Barigye, Federal University of Lavras, Lavras, Brazil
Matheus Puggina de Freitas, Federal University of Lavras, Lavras, Brazil

ABSTRACT

It has been a decade since the introduction of the MIA-QSAR (acronym of Multivariate Image Analysis applied to Quantitative Structure Activity Relationship) method. While many successes have been achieved over the years, the path for the progressive development of this method has been far from straight; several hurdles have been encountered and corresponding solutions devised, some immediately and others gradually. This report offers a comprehensive treatise of the most relevant stages in the historical development of the MIA-QSAR strategy. New challenges and future prospects are also discussed.

KEYWORDS

2D-Discrete Fourier Transform, Ball & Stick, Chemoface, MIA-QSAR, Molecular Modeling, Multivariate Image, Van der Waals radii, Wireframes

INTRODUCTION

One of the fundamental questions that has lived on throughout the history of chemistry has been how to best extract relevant information from structural representations of chemicals (Boeyens & Ogilvie, 2008), as a means of understanding the factors/processes that determine observed phenomena and in order to predict future tendencies or to alter known compounds to yield desired properties. This constitutes an important task for theoretical chemists, requiring a token of ingenuity, deductive reasoning, analysis, obstinacy and, perhaps, coincidence.

It has been ten years since the introduction of the MIA-QSAR (acronym of Multivariate Image Analysis applied to Quantitative Structure Activity Relationship) method. Looking back revokes the memories of the pioneering efforts to define a molecular modeling approach that would serve as an alternative to the then costly commercial programs at the peak of their popularity, such as CoMFA, CoMSIA and SOMFA (Hwan Kim, Greco, & Novellino, 1998; Klebe, Abraham, & Mietzner, 1994; Robinson, Winn, Lyne, & Richards, 1999). So, the idea was to come up with a simple and computationally cheap approach that would encode structural information of chemicals. While there existed popularized chemical structural representations such as SMILES, SMARTS, CML, MDL MOL/SDF, etc., chemical images had not yet been used in molecular modeling. The initial attempt to use infrared spectral images proved futile, as these yielded very poor correlations with molecular properties and this effort was thus abandoned. So going back to the drawing table, an interesting interrogative arose: what would be a better source of chemical information than the molecular structure itself? This interrogative is in fact the one that inaugurated this interesting journey in the use of chemical structure images in the modeling of physical, chemical, physicochemical and biological properties of chemical compounds.

DOI: 10.4018/IJQSPR.2016010103

The MIA-QSAR strategy was motivated from the reasoning that for a given series of congruent chemical structures, the variation in the observed properties is function of the non-congruent substructures/groups. Over the years, the MIA-QSAR approach has progressed, overcoming the hurdles particular to the method and incorporating strategies to codify more chemically meaningful information. This review offers a historical treatise of the different phases that this method has gone through, enlightening the challenges encountered, efforts devised and the innovations that have been incorporated to improve its usability.

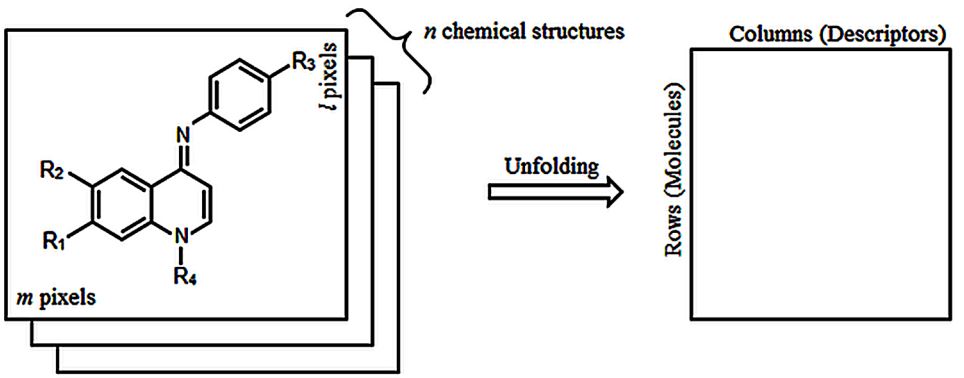
THEORETICAL STRUCTURE OF THE MIA-QSAR DESCRIPTORS

The MIA-QSAR descriptors are in fact pixels of chemical structural images (*i.e.* the image pixels are considered as the descriptors) and are obtained as follows. For a given set of n chemical structures, with a common substructure or moiety (*i.e.* basic molecular scaffold), these are drawn on an $m \times l$ (same size) canvas and saved as images, separately. Subsequently, the chemical images are aligned with respect to a common coordinate forming a 3-way $n \times m \times l$ Multi Variate Image (MVI). Later, the MVI is unfolded to form a 2-way $n \times (m \times l)$ data matrix; with n rows and $m \times l$ columns (see Figure 1 for illustration). Since pixels can be described numerically, their coordinates in an image give rise to chemical drawings (chemical structures with different substituents in varying positions) and, therefore, distinct data matrices for each compound are obtained, which explain the variance in the properties block (dependent variables).

Traditional MIA-QSAR Descriptors

The very first attempt to use images of chemical structures as a source of molecular descriptors involved a binary pixel scale, in that black and white images were considered, and the chemical structures were drawn as simple wire-frames. In this sense, the ensuing variables comprised of the values 0 (black) and 765 (white), exclusively, where 765 represented the blank spaces in the canvas, while 0 indicated the wire-frames that comprised the chemical structures. From here on, these will be dominated as the traditional MIA-QSAR descriptors. These molecular descriptors (MDs) were successfully used in the modeling of numerous bioactivities such as: affinity to the dopamine D_2 receptor subtype (Freitas, Brown, & Martins, 2005), glycogen synthase kinase 3 (GSK-3) inhibitors (Goodarzi, Freitas, & Jensen, 2009), antimalarials (Cormanich, Freitas, & Rittner, 2011; Goodarzi & Freitas, 2011), anxiolytic agents [5-HT_{2C} receptor antagonists] (Bitencourt & Freitas, 2008), HIV reverse transcriptase inhibitors (Freitas, 2006; Goodarzi & Freitas, 2008; Goodarzi & Freitas, 2010a), phosphodiesterase type 5 (PDE-5) inhibitors (Antunes, Freitas, & Rittner, 2008), antifungals

Figure 1. Workflow followed in the generation of the MIA-QSAR descriptors



(Bitencourt & Freitas, 2009), peptides for treatment of dengue (Silla et al., 2011), anti-inflammatory agents (Lloret et al., 2009) and acetylcholinesterase inhibitors (Bitencourt, Freitas, & Rittner, 2012), among others. The traditional MIA-QSAR descriptors did not only find utility in the modeling of bioactivity endpoints; applications in spectroscopy (in the prediction of ^{13}C chemical shifts) (Goodarzi, Freitas, & Ramalho, 2009) and agrochemistry (in the modeling of the phytotoxicity and soil sorption profiles of herbicides) may be found in the literature (Bitencourt & Freitas, 2008; Freitas, Matias, Macedo, Freitas, & Venturin, 2013; Goodarzi & Freitas, 2010b). These results demonstrate the utility of these MDs in codifying relevant chemical structure information. However, despite the successful results obtained with the traditional MIA-QSAR approach, a couple of challenges were encountered: 1) heteroatoms could not be discriminated and thus their symbols were used, which added a token of subjectivity to the method (Cormanich, Nunes, & Freitas, 2012); 2) as anticipated, data matrices from black and white images yield data matrices with highly correlated variables, and this meant that only modeling methods that entail the transformation of the dataset into orthogonal projections, e.g. Partial Least Squares (PLS) and Principal Component Regression (PCR), Least Squares-Support Vector Machine (LS-SVM), etc. could be used. While these methods have shown to yield good correlations, no claim of generality could be made on their performance in the modeling of all chemical structure endpoints, consistent with the no free lunch theorem (Wolpert & Macready, 1997); 3) the traditional MIA-QSAR method only codified topological information and there was no possibility of incorporating weighting schemes for the atoms and/or substructures in the black and white images, using known atomic properties such as electronegativity, polarizability, atomic hydrophobicity (Ghose & Crippen, 1987) and electron affinity, or size features, such as the atomic Van der Waals radius, as a means of integrating greater chemical information to the MIA-QSAR method; 4) mechanistic interpretation of the models built with the MIA-QSAR descriptors was deemed intractable. With the aim of providing solutions to these challenges, an extension of the MIA-QSAR method based on color schemes was introduced and denominated as aug-MIA-QSAR (acronym for augmented Multivariate Image Analysis applied to QSAR). In the next subsection, the particularities of this extension are discussed.

The aug-MIA-QSAR (SAR) Scheme

This approach, coined through the introduction of color schemes to the traditional MIA-QSAR method, marked an important step in the MVI-based modeling in the sense that it permitted, for the first time, the inclusion of weighting schemes to the chemical images using atomic properties. Consequently, this enabled the integration of vital chemical information and permitted the discrimination of heteroatoms in a more meaningful way, in addition to the greater modeling capacity ultimately obtained. Here the atoms are represented by solid color circles [the initial aug-MIA-QSAR set up used spotlighted circles to represent atoms in 3D dimension (Nunes & Freitas, 2013), which was subsequently abandoned because many pixel colors represented a single atom], contrary to the traditional MIA-QSAR approach, where wire-frames were used (see Figure 2 for comparison). The aug-MIA-QSAR scheme comprises of two approaches, i.e. aug-MIA-QSARvol and aug-MIA-QSARcolor, respectively.

aug-MIA-QSARvol

In this procedure, atoms constituting chemical structures are assigned default colors (as established by the chemical structure drawing program, e.g. GaussView) with the aim of solely discriminating the atom types, while the radii of the circles representing the atoms are defined to be proportional to the Van der Waals radius (Duarte, Barigye, da Mota, & Freitas, 2015; Duarte, Barigye, & Freitas, 2015; Freitas & Duarte, 2015; Freitas, Barigye, & Freitas, 2015; Guimarães, Mota, Silva, & Freitas, 2014). Table 1 shows the default colors and their corresponding pixel values as determined by the GaussView program for the most common atoms in organic compounds. The Van der Waals radius (and thus volume) is closely related to the size-dependent physicochemical properties, such as polarizability, diamagnetic susceptibility and chemical hardness (Freitas & Duarte, 2015), and thus

Figure 2. Aligned chemical images employed to generate the MIA-QSAR descriptors according to a) the traditional MIA-QSAR, b) aug-MIA-QSARvol (atoms spotlighted), c) aug-MIA-QSARvol (atoms with solid colors), and d) aug-MIA-QSARcolor approaches. The later three images were drawn using the GaussView program

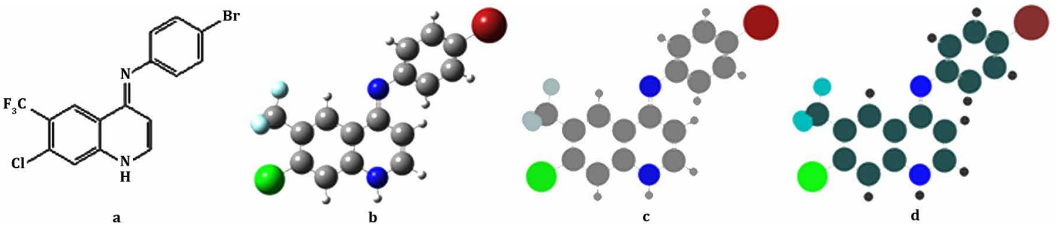


Table 1. Colors and respective pixels for atoms according to the aug-MIA-QSARvol and aug-MIA-QSAR$_{color}$ approaches, respectively

	aug-MIA-QSARvol		aug-MIA-QSAR$_{color}$		
Atom	**Pixel**	**Colors**	**E†**	**Pixel**	**Colors**
H	612	Silver	2.1	210	Charcoal
C	426	Gray	2.5	250	Teal
N	279	Blue	3.0	300	Persian blue
O	229	Red	3.5	350	Scarlet
F	688	Electric blue	4.0	400	Turquoise
S	493	Gold	2.5	250	Olive
Cl	289	Green	3.0	300	Green
Br	231	Carmine	2.8	280	Maroon
I	294	Byzantium	2.5	250	Purple

†E: Pauling's electronegativity

relevant chemical information is incorporated in the MIA-QSAR approach. Additionally, ligand-receptor interactions are greatly influenced by the molecular bulk and thus affecting the bioactivity of chemical compounds.

aug-MIA-QSARcolor

In this case, the colors for the atoms are carefully chosen to guarantee direct correspondence to the different atomic properties, for example: atomic hydrophobicity, electronegativity, covalent radius, polarizability, etc. (Freitas et al., 2015; Nunes & Freitas, 2013). Table 1 shows the pixel values and the corresponding colors for atoms regularly encountered in organic compounds, when Pauling's electronegativity scale is considered.

In order to compare the traditional MIA-QSAR, aug-MIA-QSARvol and aug-MIA-QSARcolor approaches, respectively, several modeling experiments were carried out and their results reported. Table 2 is a summary of the performance of the 3 approaches in studies carried out for the following endpoints: chemokine receptor inhibition for a series of *R*-3-amino-pyrrolidines, antitrypanosomal activity of thiosemicarbazones and semicarbazones, antiplasmodial activity for a series of quinolon-4(1*H*)-imines and the sweetness of guanidine derivatives measured as log(RS) (Duarte et al., 2015; Freitas & Duarte, 2015; Nunes & Freitas, 2013; Nunes & Freitas, 2013). As anticipated, the incorporation of color schemes enhanced the performance of the MIA-QSAR method, yielding more predictive and robust models, as evidenced by the improvement of the validation parameters Q^2loo

(correlation coefficient for the leave-one-out cross-validation procedure), Q^2ext (correlation coefficient for the external validation procedure), Q^2_oext (modified correlation coefficient with respect to the origin) and R^2(y-rand) (correlation coefficient after the Y-randomization procedure) in the considered studies. As for the comparison between the aug-MIA-QSARvol and MIA-QSARcolor approaches, no general rule may be established as their performance depends on the relative importance of the molecular bulk or electronic properties to the considered properties, for example: in ref. (Duarte et al., 2015), it was observed that aug-MIA-QSARvol yielded superior performance than the aug-MIA-QSARcolor approach in the modeling of the antimalarial activity of a series of quinolon-4(1*H*)-imines, suggesting greater importance of molecular size to the modeled activity. This result was in fact found to be consistent with other studies reported in the literature (Rodrigues et al., 2013). On the other hand, in the modeling of the antitrypanosomal activity, better performance was obtained with the aug-MIA-QSARcolor approach, indicative of greater relevance for electronic properties of molecules. Indeed, a previous study demonstrated the importance of dipolar interactions in explaining the action mechanism for the antitrypanosomal activity of thiosemicarbazones and semicarbazones (Freitas & Duarte, 2015).

Another key benefit of the color schemes is that they allowed for the mechanistic interpretation of the MIA-QSAR models. It is known that one of the most fervent criticisms towards theoretical methods used in the modeling of chemical structure properties/bioactivities is their lack of interpretation in chemically intuitive terms, or in regard to the structural moieties that enhance (or disfavor) the modeled properties (Barigye, Marrero-Ponce, Zupan, Pérez-Giménez, & Freitas, 2014). In the MIA-QSAR approach, each variable corresponds to a pixel coordinate in the MVI and since the colors represent different atom types in organic molecules, an analysis of the variables contained in the final model would provide insight on the atom types (or substructures) responsible for the modeled property. However, the inherent challenge would be the high number of variables generated for a given dataset with the MIA-QSAR method. While techniques like PLS and PCR provide a great solution when dealing with high dimensionality data matrices in modeling, the very fact that the original variables are transformed into orthogonal projections, makes the interpretation challenging. In this sense, a different approach was adapted in that feature selection procedures based on information-theoretic filters were assimilated (Urias et al., 2015), in order to work with more manageable data matrices and ultimately use modeling techniques that did not involve the transformation of variables into principal components. Model interpretations strongly consistent with experimental results reported in the literature have been achieved; for examples, see refs. (Duarte et al., 2015; Duarte et al., 2015; Freitas et al., 2015), awarding greater applicability to the MIA-QSAR method as valuable information on the structural characteristics relevant for the modeled property has been obtained and thus providing key leads in the rational design of novel chemical compounds with the desired properties (or bioactivities).

The Chemoface Program

In order to perform the MIA-QSAR based modeling tasks, the Chemoface program was developed (Nunes, Freitas, Pinheiro, & Bastos, 2012). This is a free, user-friendly and standalone program developed using the Matlab computing environment, downloadable via internet at http://ufla.br/chemoface/. The Chemoface program is comprised of five modules, accessible via the primary home screen (see Figure 3) and these include: Experimental Design, Multivariate Calibration, Pattern Recognition, Data Organization and the Data Plot module, respectively.

The experimental design module permits the user perform tasks associated with the design of experiments using the fractional factorial, full factorial, Plackett-Burman, central composite and mixture designs, respectively. An option is offered in this module to assess the results obtained using the Pareto charts and effect tables. As for the pattern recognition module, this provides tools for performing principal component analysis and hierarchical cluster analysis, using as amalgamation rules in the case of the latter the Euclidean and Mahalanobis distances, respectively. The multivariate calibration module permits the users to carry out model building experiments for regression employing

Table 2. Comparison of the performance of the traditional MIA-QSAR, aug-MIA-QSARvol and MIA-QSARcolor approaches in the modeling of different endpoints

Activity	Chemokine Receptor Inhibition†	Antitrypanosomal‡	Antiplasmodial$^{\Phi}$	Sweetness log(RS)$^{\Delta}$
Traditional MIA-QSAR				
R^2	0.92	0.97	0.77	0.96
Q^2loo	0.71	0.62	0.61	0.08
Q^2ext	0.57	0.57	0.80	0.47
Q^2_oext	0.36	0.38	0.26	-
R^2 (y-rand)	0.56	0.74	0.22	-
aug-MIA-QSARvol				
R^2	0.91	0.97	0.81	0.96
Q^2loo	0.68	0.76	0.67	0.73
Q^2ext	0.74	0.65	0.97	0.57
Q^2_oext	0.59	0.48	0.91	-
R^2 (y-rand)	0.57	0.70	0.25	0.70
aug-MIA-QSARcolor				
R^2	-	0.97	0.74	-
Q^2loo	-	0.76	0.63	-
Q^2ext	-	0.71	0.84	-
Q^2_oext	-	0.59	0.66	-
R^2 (y-rand)	-	0.70	0.13	-

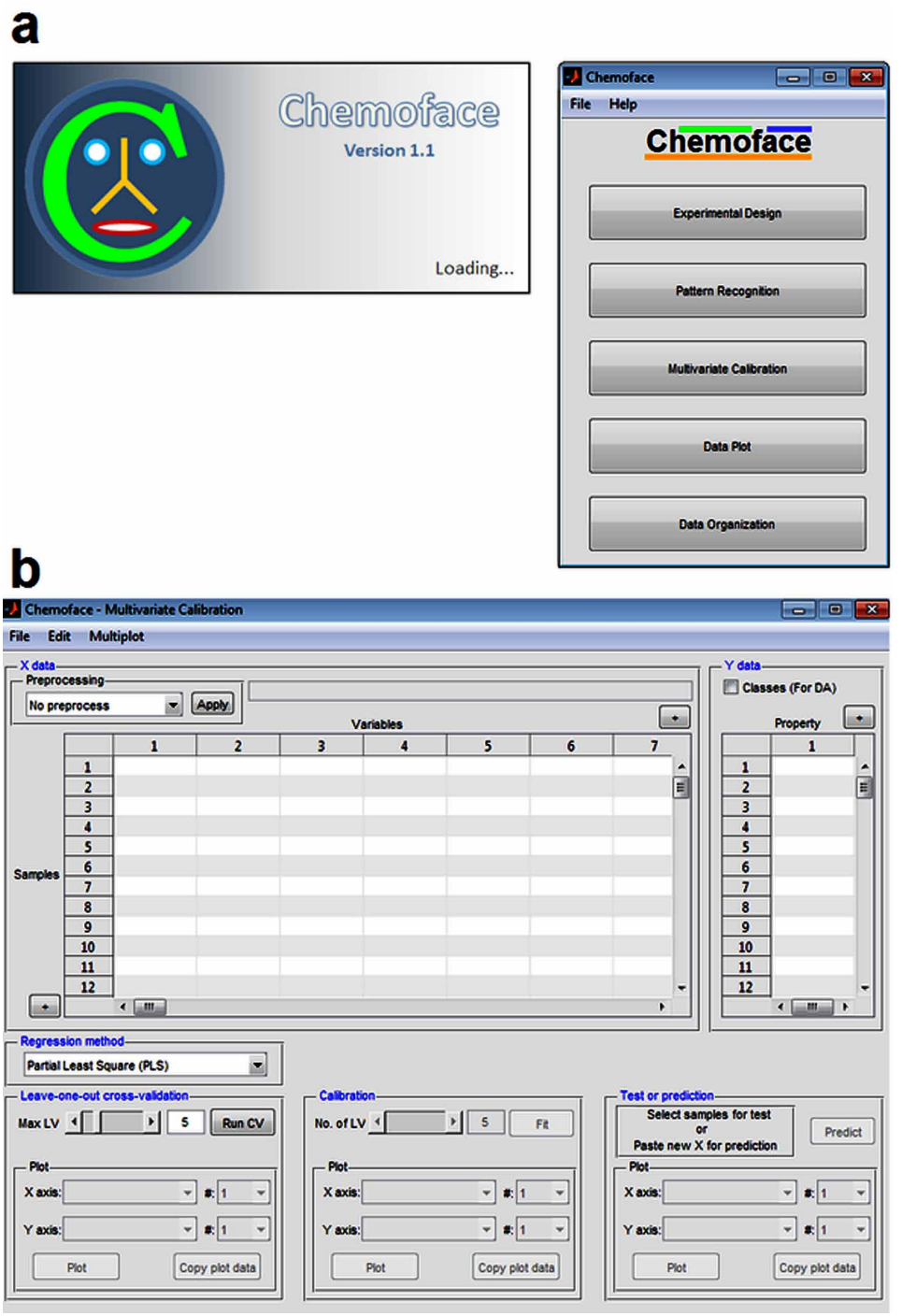
†*R*-3-amino-pyrrolidines (Cleiton A. Nunes & Matheus P. Freitas, 2013). ‡Thiosemicarbazones and semicarbazones (Matheus P Freitas & Duarte, 2015). $^{\Phi}$Quinolon-4(1*H*)-Imines (M. H. Duarte et al., 2015). $^{\Delta}$Guanidine derivatives (Cleiton A Nunes & Matheus P Freitas, 2013; Cleiton A. Nunes & Matheus P. Freitas, 2013).

the statistical techniques partial least squares, principal component regression and multiple linear regression, respectively, and corresponding analogues for discriminant analysis in classification tasks. Additionally, options for validating the statistical quality of the built models using the techniques leave-one-out cross validation, Y-randomization and external validation are provided. Moreover, data preprocessing operations e.g. the removal of zero variance variables, auto-scaling, standardization, etc. may also be performed. The data organization module allows for the importing of data matrices in the Tab, DAT and Comma Separated Value file formats (*i.e.* .TXT, .DAT and .CSV, respectively), as well as images in the Bitmap (.BMP) image file format. Note that the MVI unfolding procedure to generate the MIA-QSAR descriptor matrix is performed in this module. A previous study comparing the JPEG, PNG, TIFF and BMP formats demonstrated that the performance of the MIA-QSAR method was invariant to the type of image file format used for the chemical images (Goodarzi, Freitas, & Ferreira, 2009). As for the data plot module, this is simply used for data visualization procedures using scatter plots. The Chemoface program does not require a MATLAB license installation to run, but only the MATLAB Compiler Runtime (MCR), offered together with the program.

NEW TRENDS IN THE MIA-QSAR STRATEGY

In this section, we discuss the recent advances in the MIA-QSAR method, emphasizing the innovations that have ultimately improved its applicability in modeling tasks. It should be noted that while the

Figure 3. Graphical User Interface of the Chemoface program. a) Primary Home Screen; b) Interface for Multivariate Calibration Module



introduction of the color schemes improved the performance of the MIA-QSAR method, it still presented two key pitfalls arising from the very conceptual structure of this method: 1) the accuracy in the alignment of the images plays a vital role in the performance of the MIA-QSAR method, given that it is important that the common molecular scaffold in the MVI yields zero variance (or Shannon's entropy) variables, to guarantee that the variance in the modeled property depends exclusively on the non-congruent groups (or substructures). However, the molecular images have been until recently manually aligned with respect to an empirically chosen common coordinate of the MVI. This in essence meant that the quality of the alignment hinged on the user's meticulousness, thus adding a component of subjectivity to the MIA-QSAR method. On several occasions, minuscule shifts in the pixel coordinates resulted in inconsistencies in the reverse process for identifying the atom types represented by the variables in the built models, and which meant that the whole experiment would have to be repeated; 2) secondly, and probably more importantly, the very fact that the chemical structure images have to be aligned with respect to a basic congruent scaffold meant that the MIA-QSAR approach would not be applied to structurally diverse molecular datasets, comprised of numerous non-congruent chemical structures, and thus precluding its application in conventional virtual screening tasks for lead compounds. In an effort to deal with these challenge, the notion of image transformation using the 2D-Discrete Fourier Transform (2D-DFT) procedure was introduced, ushering in a new phase altogether in the MIA-QSAR methodology (Barigye & Freitas, 2015a).

The 2D-Discrete Fourier Transform (2D-DFT)

The Fourier series (or Fourier Transform), coined in the 18th century by the French mathematician Jean Baptiste Joseph Fourier are based on the principle that periodic (or periodized) phenomena could be described in a trigonometric domain as a summation of sine and cosine functions with varying frequencies and weighted by coefficients denominated as *Fourier coefficients*. Despite the simplicity of the core principles of the Fourier Transform (FT), its implications have had a great impact in technology and its applications may be found in the field of telecommunications (particularly in signal processing), circuit design, spectroscopy, and definitely in digital image analysis. The extrapolation of the FT to 2D image analysis [in which case the denomination 2D-Discrete Fourier Transform (2D-DFT) is adopted given the discrete nature of images], is based on the insight that images are simply a summation of linear spatial functions, $f(x, y)$, where x and y represent the spatial coordinates. Therefore, the dissimilarity in images lies in the difference in the periodicity and orientation of the constituent functions, $f(x, y)$. The2D-DFT projects the image functions in the spatial frequency domain, yielding a set of spatial frequencies of different magnitude known as the *Fourier spectrum*. As a consequence, image analysis may be performed in a simpler and more intuitive domain and allowing for the desired modifications.

The power of the 2D-DFT from a MIA-QSAR perspective lies in the fact that magnitude spectra obtained in the frequency domain for a set of images possess a common base. The implications of this understanding is that non-congruent images probably due to variations in size, differences in orientations or coordinates in the canvas, which were previously unanalyzable from the MIA-QSAR perspective could now be, in form of their magnitude spectra as these possess a common base.

We will now formally give a brief mathematical definition of the 2D-DFT, without getting entangled in the derivations involved, for a detailed treatise see refs. (Dougherty, 1994; Gonzalez & Woods, 2007). Given an $M \times N$ image, where M and N represent the number of sample points in the R^2 space, the 2D-DFT is expressed as follows:

$$F(u,v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \exp\left[-i2\pi(ux/M + vy/N)\right] \tag{1}$$

where *u* and *v* are the discrete spatial frequencies, *f (x, y)* is the input and *F (u, v)* is the magnitude spectrum. Note that there exists proportionality between the magnitudes of the complex coefficients and the intensity of the spatial frequencies.

The first benefit reaped from the use of magnitude spectra for the chemical images was that, it was no longer necessary to manually align the chemical structures with respect to a common coordinate, as the MIA-QSAR descriptors would be obtained from the MVI of magnitude spectra, in place of the original MVI. Figure 4 is a graphical illustration of the 2D-DFT MIA-QSAR procedure.

To test the veracity of this principle, two experiments were performed. Firstly, two replicas of a chemical structure image were obtained, with one horizontally translated relative to the original image, while the other was rotated π^c relative to the original image. In theory, while the three images are not superimposable, they should yield identical magnitude spectra, and thus the ensuing data matrix for the 3 chemical structures should have zero variance. Indeed, it was found to be the case, and thus proving that the 2D-DFT creates a common base for images (Barigye & Freitas, 2015a). Secondly, the 2D-DFT approach was applied in previously performed modeling experiments and the results obtained compared. Table 3 shows a comparison of the results obtained with the 2D-DFT based MIA-QSAR and the original MIA-QSAR approach in the modeling of the following bioactivities: antimalarial activity for a series of 2,5-diaminobenzophenone derivatives, tyrosine kinase enzyme WEE1 inhibition (4-phenylpyrrolocarbazoles) and trichomonicidal activity (benzimidazoles) (Cormanich et al., 2011; Cormanich, Goodarzi, & Freitas, 2009; Cormanich et al., 2012; Freitas & Duarte, 2015). Additionally, the results obtained using the CoMFA method in the modeling of these endpoints are included.

As can be observed, similar to superior results are obtained when magnitude spectra are used for model building (2D-DFT MIA-QSAR) compared to those of the original MIA-QSAR method, based on normal and manually aligned chemical images. In the case of the model built for the tyrosine kinase enzyme WEE1 inhibitors (phenylpyrrolocarbazoles), it is observed that while the PLS-based model for the 2D-DFT MIA-QSAR method yields satisfactory results, according to the established criteria for model acceptability, the MIA-QSAR method produces a low Q^2ext value (0.31) and improvements are only obtained when the LS-SVM technique is applied, which is known to be a more powerful nonlinear modeling method. Altogether, these results suggest that the use of the 2D-DFT transform in addition providing relief from the manual alignment procedure, yields improvements in the models' performance, and thus constituting another important practical contribution.

Moreover, and more importantly, the use of the 2D-DFT opened way for the prospect of model building using structurally diverse non-congruent chemical images, typical of contemporary chemical compound datasets (Barigye & Freitas, 2015b). The underlying reasoning was that since with the 2D-DFT procedure, the chemical images are "broken down" into the constituent components in the spatial frequency domain, the degree of similarity (or dissimilarity) in the magnitude spectra should have a close relationship with the property (or bioactivity) values of the corresponding chemical structures. To access the legitimacy of this inference, MLR-based models were built for the inhibitory

Figure 4. Illustration of the steps involved in the 2D-DFT MIA-QSAR procedure

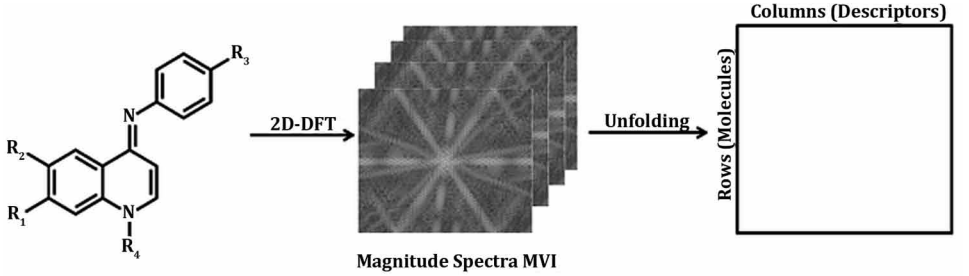


Table 3. Comparison of the performance of the 2D-DFT MIA-QSAR and the MIA-QSAR approaches in the modeling of bioactivities of different chemical compound series

Activity	Antimalarial[†]	Tyrosine kinase enzyme WEE1 inhibitors [‡]	Trichomonicidals[ϕ]
2D-DFT MIA-QSAR			
R^2	0.94	0.95	0.93
Q^2loo	0.77	0.79	0.68
Q^2ext	0.70	0.73	0.79
Q^2_oext	0.53	0.56	0.62
R^2p(y-rand)	0.56	0.54	0.51
MIA-QSAR			
R^2	0.91	0.79 (0.94[§])	0.85
Q^2loo	0.56	0.59(0.90[§])	0.52
Q^2ext	0.73	0.31(0.93[§])	0.79
Q^2_oext	-	-	-
R^2 (y-rand)	-	-	-
CoMFA			
R^2	0.87	0.96	0.93-0.94
Q^2loo	0.55	0.83	0.60-0.63
Q^2ext	0.72	0.90	0.73-0.89
Q^2_oext	-	-	-
R^2 (y-rand)	-	-	-

[†]2,5-Diaminobenzophenone derivatives (compound series), PLS (fitting method) (Cormanich et al., 2011). [‡]4-Phenylpyrrolocarbazoles-compound dataset, PLS and [§]LS-SVM (fitting methods) (Cormanich et al., 2009; Matheus P Freitas & Duarte, 2015). [ϕ]Benzimidazoles-compound dataset, PLS (fitting method) (Cormanich et al., 2012).

activity against the MCF-7 human breast cancer cell line using a dataset of structurally diverse (non-congruent) chemical images (Barigye & Freitas, 2015b). A parallel experiment using the DRAGON descriptors was performed for this dataset and the results compared. Table 4 shows the statistical parameters for the 2D-DFT MIA-QSAR and DRAGON based models, respectively.

As can be observed, the 2D-DFT MIA-QSAR models produced satisfactory behavior, on the basis of the quality of the statistical parameters obtained, comparable with the results obtained for the DRAGON MDs (in the case of the 6- and 5-variable models), while superior performance was observed for the former in the case of the 4- and 3-variable models. This is a really promissory result, considering the diversity and credibility of the DRAGON MDs in chemoinformatics tasks. In fact, the DRAGON MDs are comprised of dissimilar families of indices conceived from diverse theoretical and practical considerations, and representing decades of research. Therefore it is not farfetched to suggest that the application of the 2D-DFT approach constitutes an important breakthrough in the MIA-QSAR context, as it ushers in a new range of possibilities in molecular modeling using chemical structural images.

CONCLUSION

In the present report, a comprehensive review of the historical development of the MIA-QSAR approach has been offered, highlighting the different phases that this method has gone through,

Table 4. Statistical parameters for regression models for the MCF-7 cells inhibitory activity based on the 2D-DFT MIA-QSAR and DRAGON methods, respectively

Method	N	R^2	Q^2loo	Q^2boot	a(R^2)	a(Q2)	Q^2ext	Q_o^2ext	k
2D-DFT MIA-QSAR	6	0.87	0.85	0.84	0.044	-0.171	0.83	0.79	0.98
	5	0.84	0.82	0.81	0.028	-0.157	0.80	0.68	1.00
	4	0.81	0.78	0.78	0.017	-0.134	0.72	0.47	1.01
	3	0.79	0.76	0.76	0.001	-0.118	0.70	0.58	0.98
DRAGON	6	0.83	0.81	0.79	0.046	-0.172	0.81	0.68	1.05
	5	0.81	0.79	0.77	0.027	-0.187	0.66	0.45	1.05
	4	0.78	0.76	0.75	0.019	-0.139	0.29	0.38	1.05
	3	0.75	0.71	0.71	0.008	-0.112	0.57	0.46	1.05

the different challenges encountered and the solutions devised. In light of the results obtained so far with MIA-QSAR, it may be suggested that this method constitutes a promissory tool to take into consideration in modeling problems. Even then, there are new drawbacks to be resolved, e.g. while the introduction of the 2D-DFT opened way for modeling building using structurally diverse datasets, the interpretability of the models in terms of the atom-types/groups responsible for the variation in the modeled property became much more complex, because the variables obtained from the magnitude spectra may not be used to directly trace the regions they represent in the molecular structure. Nonetheless, efforts are underway to extrapolate the inverse 2D-DFT (used in the retrieval of images after the necessary modifications are performed in the spatial frequency domain) to the MIA-QSAR method, although there several challenges yet to be solved. On the other hand, there other interesting notions that may be explored e.g. the use of other image transformation procedures typical of the field of digital image processing, such as Wavelets, Discrete Cosine and Walsh-Hadamard Transforms, in addition to classical methods, like Principal Component Analysis. These will certainly constitute future tasks.

ACKNOWLEDGMENT

Barigye, S. J. and Freitas, M. P. acknowledge financial support from CNPq and FAPEMIG. This work is a collaboration research project of members of the Rede Mineira de Química (RQ-MG) supported by FAPEMIG (Project: CEX - RED-00010-14).

REFERENCES

Antunes, J. E., Freitas, M. P., & Rittner, R. (2008). Bioactivities of a series of phosphodiesterase type 5 (PDE-5) inhibitors as modelled by MIA-QSAR. *European Journal of Medicinal Chemistry*, *43*(8), 1632–1638. doi:10.1016/j.ejmech.2007.10.019 PMID:18045743

Barigye, S. J., & Freitas, M. P. (2015a). 2D-Discrete Fourier Transform: Generalization of the MIA-QSAR strategy in molecular modeling. *Chemometrics and Intelligent Laboratory Systems*, *143*, 79–84. doi:10.1016/j.chemolab.2015.02.020

Barigye, S. J., & Freitas, M. P. (2015b). Is molecular alignment an indispensable requirement in the MIA-QSAR method? *Journal of Computational Chemistry*, *36*(23), 1748–1755. doi:10.1002/jcc.23992 PMID:26119527

Barigye, S. J., Marrero-Ponce, Y., Zupan, J., Pérez-Giménez, F., & Freitas, M. P. (2014). Structural and Physicochemical Interpretation of GT-STAF Information Theory-Based Indices. *Bulletin of the Chemical Society of Japan*.

Bitencourt, M., & Freitas, M. P. (2008). MIA-QSAR evaluation of a series of sulfonylurea herbicides. *Pest Management Science*, *64*(8), 800–807. doi:10.1002/ps.1565 PMID:18338340

Bitencourt, M., & Freitas, M. P. (2009). Bi- and Multilinear PLS Coupled to MIA-QSAR in the Prediction of Antifungal Activities of Some Benzothiazole Derivatives. *Medicinal Chemistry (Shariqah, United Arab Emirates)*, *5*(1), 79–86. doi:10.2174/157340609787049208 PMID:19149653

Bitencourt, M., Freitas, M. P., & Rittner, R. (2012). The MIA-QSAR Method for the Prediction of Bioactivities of Possible Acetylcholinesterase Inhibitors. Archiv der Pharmazie, *345*(9), 723–728. doi:10.1002/ardp.201200079 PMID:22674790

Boeyens, J. C. A., & Ogilvie, J. F. (Eds.). (2008). *Models, Mysteries and Magic of Molecules*. Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-5941-4

Cormanich, R. A., Freitas, M. P., & Rittner, R. (2011). 2D chemical drawings correlate to bioactivities: MIA-QSAR modelling of antimalarial activities of 2, 5-diaminobenzophenone derivatives. *Journal of the Brazilian Chemical Society*, *22*(4), 637–642. doi:10.1590/S0103-50532011000400004

Cormanich, R. A., Goodarzi, M., & Freitas, M. P. (2009). 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. *Chemical Biology & Drug Design*, *73*(2), 244–252. doi:10.1111/j.1747-0285.2008.00764.x PMID:19207427

Cormanich, R. A., Nunes, C. A., & Freitas, M. P. (2012). Desenhos de Estruturas Químicas Correlacionam-se com Propriedades Biológicas: MIA-QSAR. *Quimica Nova*, *35*(6), 1157–1163. doi:10.1590/S0100-40422012000600017

Dougherty, E. R. (1994). *Digital Image Processing Methods* (Vol. 42). Boca Raton, FL: CRC Press.

Duarte, M., Barigye, S., da Mota, E., & Freitas, M. (2015). Computational modelling of the antischistosomal activity for neolignan derivatives based on the MIA-SAR approach. *SAR and QSAR in Environmental Research*, *26*(3), 205–216. doi:10.1080/1062936X.2015.1018942 PMID:25774798

Duarte, M. H., Barigye, S. J., & Freitas, M. P. (2015). Exploring MIA-QSAR's for antimalarial quinolon-4(1H)-imines. *Combinatorial Chemistry & High Throughput Screening*, *18*(2), 208–216. doi:10.2174/1386207318666141229123349 PMID:25543687

Freitas, M. P. (2006). MIA-QSAR modelling of anti-HIV-1 activities of some 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners. *Organic & Biomolecular Chemistry*, *4*(6), 1154–1159. doi:10.1039/b516396j PMID:16525561

Freitas, M. P., Brown, S. D., & Martins, J. A. (2005). MIA-QSAR: A simple 2D image-based approach for quantitative structure–activity relationship analysis. *Journal of Molecular Structure*, *738*(1–3), 149–154. doi:10.1016/j.molstruc.2004.11.065

Freitas, M. P., & Duarte, M. H. (2015). Evolution of Multivariate Image Analysis in QSAR: The Case for a Neglected Disease. In K. Roy (Ed.), *Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment* (pp. 84–122). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-8136-1.ch003

Freitas, M. R., Barigye, S. J., & Freitas, M. P. (2015). Coloured chemical image-based models for the prediction of soil sorption of herbicides. *RSC Advances*, *5*(10), 7547–7553. doi:10.1039/C4RA12070A

Freitas, M. R., Matias, S. V. B. G., Macedo, R. L. G., Freitas, M. P., & Venturin, N. (2013). Augmented Multivariate Image Analysis Applied to Quantitative Structure–Activity Relationship Modeling of the Phytotoxicities of Benzoxazinone Herbicides and Related Compounds on Problematic Weeds. *Journal of Agricultural and Food Chemistry*, *61*(36), 8499–8503. doi:10.1021/jf4024257 PMID:23947385

Ghose, A. K., & Crippen, G. M. (1987). Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. *Journal of Chemical Information and Computer Sciences*, *27*(1), 21–35. doi:10.1021/ci00053a005 PMID:3558506

Gonzalez, R. C., & Woods, R. E. (2007). *Digital Image Processing* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Goodarzi, M., & Freitas, M. P. (2008). Augmented Three-mode MIA-QSAR Modeling for a Series of Anti-HIV-1 Compounds. *QSAR & Combinatorial Science*, *27*(9), 1092–1097. doi:10.1002/qsar.200810030

Goodarzi, M., & Freitas, M. P. (2010a). MIA–QSAR coupled to principal component analysis-adaptive neuro-fuzzy inference systems (PCA–ANFIS) for the modeling of the anti-HIV reverse transcriptase activities of TIBO derivatives. *European Journal of Medicinal Chemistry*, *45*(4), 1352–1358. doi:10.1016/j.ejmech.2009.12.028 PMID:20060625

Goodarzi, M., & Freitas, M. P. (2010b). PLS and N-PLS-based MIA-QSTR modelling of the acute toxicities of phenylsulphonyl carboxylates to Vibrio fischeri. *Molecular Simulation*, *36*(12), 953–959. doi:10.1080/08927022.2010.492836

Goodarzi, M., Freitas, M. P., & Ferreira, E. B. (2009). Influence of Changes in 2-D Chemical Structure Drawings and Image Formats on the Prediction of Biological Properties Using MIA-QSAR. *QSAR & Combinatorial Science*, *28*(4), 458–464. doi:10.1002/qsar.200810146

Goodarzi, M., Freitas, M. P., & Jensen, R. (2009). Feature selection and linear/nonlinear regression methods for the accurate prediction of glycogen synthase kinase-3B inhibitory activities. *Journal of Chemical Information and Modeling*, *49*(4), 824–832. doi:10.1021/ci9000103 PMID:19338295

Goodarzi, M., Freitas, M. P., & Ramalho, T. C. (2009). Prediction of 13C chemical shifts in methoxyflavonol derivatives using MIA-QSPR. *Spectrochimica Acta. Part A: Molecular and Biomolecular Spectroscopy*, *74*(2), 563–568. doi:10.1016/j.saa.2009.07.003 PMID:19648055

Goodarzi, M., & Freitas, P., M. (. (. (2011). MIA-QSAR Coupled to Different Regression Methods for the Modeling of Antimalarial Activities of 2-aziridinyl and 2,3-bis-(aziridinyl)-1,4-naphtoquinonyl Sulfate and Acylate Derivatives. *Medicinal Chemistry (Shariqah, United Arab Emirates)*, *7*(6), 645–654. doi:10.2174/157340611797928343 PMID:22313304

Guimarães, M. C., da Mota, E. G., Silva, D. G., & Freitas, M. P. (2014). aug-MIA-QSPR modelling of the toxicities of anilines and phenols to Vibrio fischeri and Pseudokirchneriella subcapitata. Chemometrics and Intelligent Laboratory Systems, *134*, 53–57. doi:10.1016/j.chemolab.2014.03.005

Hwan Kim, K., Greco, G., & Novellino, E. (1998). A Critical Review of Recent CoMFA Applications. *Perspectives in Drug Discovery and Design*, *12/13*(14), 257–315.

Klebe, G., Abraham, U., & Mietzner, T. (1994). Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity. *Journal of Medicinal Chemistry*, *37*(24), 4130–4146. doi:10.1021/jm00050a010 PMID:7990113

Lloret, G. R., Cunha Neto, Á., Rittner, R., Bitencourt, M., Freitas, M. P., & Aquino, N. S. (2009). Synthesis and rational design of anti-inflammatory compounds: N-phenyl-cyclohexenyl sulfonamide derivatives. *Journal of Physical Organic Chemistry*, *22*(12), 1188–1192. doi:10.1002/poc.1575

Nunes, C. A., & Freitas, M. P. (2013). MIA-QSPR study of guanidine derivative sweeteners. *European Food Research and Technology*, *237*(4), 565–570. doi:10.1007/s00217-013-2032-8

Nunes, C. A., & Freitas, M. P. (2013). Introducing new dimensions in MIA-QSAR: A case for chemokine receptor inhibitors. *European Journal of Medicinal Chemistry*, *62*(0), 297–300. doi:10.1016/j.ejmech.2013.01.005 PMID:23357311

Nunes, C. A., Freitas, M. P., Pinheiro, A. C. M., & Bastos, S. C. (2012). Chemoface: A novel free user-friendly interface for chemometrics. *Journal of the Brazilian Chemical Society*, *23*(11), 2003–2010. doi:10.1590/S0103-50532012005000073

Robinson, D. D., Winn, P. J., Lyne, P. D., & Richards, W. G. (1999). Self-organizing molecular field analysis: A tool for structure-activity studies. *Journal of Medicinal Chemistry*, *42*(4), 573–583. doi:10.1021/jm9810607 PMID:10052964

Rodrigues, T., da Cruz, F. P., Lafuente-Monasterio, M. J., Gonçalves, D., Ressurreição, A. S., Sitoe, A. R., & Moreira, R. et al. (2013). Quinolin-4 (1 H)-imines are potent antiplasmodial drugs targeting the liver stage of malaria. *Journal of Medicinal Chemistry*, *56*(11), 4811–4815. doi:10.1021/jm400246e PMID:23701465

Silla, J. M., Nunes, C. A., Cormanich, R. A., Guerreiro, M. C., Ramalho, T. C., & Freitas, M. P. (2011). MIA-QSPR and effect of variable selection on the modeling of kinetic parameters related to activities of modified peptides against dengue type 2. *Chemometrics and Intelligent Laboratory Systems*, *108*(2), 146–149. doi:10.1016/j.chemolab.2011.06.009

Urias, R. W. P., Barigye, S. J., Marrero-Ponce, Y., García-Jacas, C. R., Valdes-Martiní, J. R., & Perez-Gimenez, F. (2015). IMMAN: Free software for information theory-based chemometric analysis. *Molecular Diversity*, *19*(2), 305–319. doi:10.1007/s11030-014-9565-z PMID:25620721

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82. doi:10.1109/4235.585893

Stephen Jones Barigye earned his BSc in Pharmaceutical Sciences at the Central University of Las Villas (Cuba), and his PhD in Theoretical and Computational Chemistry (2013) at the same University. He is currently a postdoctoral fellow at the Federal University of Lavras (Brazil).

Matheus Puggina de Freitas was born in Itapira (Brazil) and studied Chemistry (BSc and PhD) at the State University of Campinas (UNICAMP), following a postdoctoral stay at the same University. He worked as researcher in two pharmaceutical companies and, in 2005, he started as Adjunct Professor at the Federal University of Lavras (UFLA). He is currently associate Professor of Organic Chemistry at UFLA, where he has research interests in conformational analysis of small molecules, NMR spectroscopy, computational chemistry and QSAR/QSPR. Department of Chemistry, Federal University of Lavras, Lavras, MG, Brazil.