Chemometrics and Intelligent Laboratory Systems 143 (2015) 79–84

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

2D-Discrete Fourier Transform: Generalization of the MIA-QSAR strategy in molecular modeling

Stephen J. Barigye *, Matheus P. Freitas

Department of Chemistry, Federal University of Lavras, P.O. Box 3037, 37200-000 Lavras, MG, Brazil

ARTICLE INFO

Article history:
Received 25 November 2014
Received in revised form 23 February 2015
Accepted 25 February 2015
Available online 4 March 2015

Keywords:
Multivariate image
MIA-QSAR
2D-Discrete Fourier Transform
Magnitude spectrum

ABSTRACT

Adequate alignment of chemical structure images with respect to the basic scaffold in a series of chemical compounds constitutes an indispensable requirement for constructing multivariate images (MVIs) and subsequent molecular modeling using the Multivariate Image Analysis applied to Quantitative Structure–Activity Relationship (MIA-QSAR) approach. However, up to the moment, this alignment procedure has been manually performed, based on subjective ocular precision. The 2D-Discrete Fourier Transform (2D-DFT) is introduced as a strategy for creating a common base to construct MVIs for chemical structures using their magnitude spectra. The utility of magnitude spectra in QSAR studies has been evaluated through models for the antimalarial, anticancer and trichomonicidal activity of a series of 2, 5-diaminobenzophenone, 4-phenylpyrrolocarbazole and benzimidazole derivatives, respectively, yielding satisfactory results comparable to superior to those reported in the literature. It is anticipated that this strategy should enable the application of the MIA-QSAR approach to structurally diverse datasets other than a series of congeneric datasets.

1. Introduction

It has been a decade since the Multivariate Image Analysis–Quantitative Structure Activity Relationship (MIA-QSAR) method was introduced in the field of chemometrics and molecular modeling [1]. This method is based on the bidimensional representation of molecular structures as images, in the sense that the image pixels are considered as molecular structure descriptors. The underlying reasoning is that for a given series of compounds, the chemical or biological properties are a function of the substituents bonded to the basic scaffold, in such a way that progressive modifications of the substituent chemical groups correspond to changes in their chemical (or biological) properties. The MIA-QSAR approach has been successfully used in modeling of a diverse range of bioactivities of chemical compounds including: HIV reverse transcriptase inhibitors, farnesyltransferase inhibitors, phosphodiesterase type 5 (PDE-5) inhibitors, glycogen synthase kinase 3 (GSK-3) inhibitors, antifungals, anticancer, antimalarials, anti-inflammatory compounds and peptides for the treatment of dengue, among others [1–14]. The MIA-QSAR approach has not been limited to bioactivity modeling; applications extending to agrochemistry, spectroscopy and physicochemistry may be found in the literature [15–20]. It is thus evident that the MIA-QSAR approach is useful in codifying vital chemical information.

Over the years, the MIA-QSAR approach has evolved, integrating colored schemes carefully defined to codify useful chemical information such as atomic properties (e.g. Pauling's electronegativity) and the sizes of atoms in the images modified to relate with steric factors of atoms etc. [18,19,21–25]. These extensions have shown to improve the predictive power and interpretability of the MIA-QSAR models.

However, notwithstanding all the benefits reaped from the MIA-QSAR approach, limitations still do exist, and perhaps the most important being that it is only applicable to congruent images. The molecular images are manually aligned with respect to the basic molecular scaffold in a chemical compound series. In such a condition, optimal aligning essentially depends on ocular precision. This is particularly troublesome as a minimal shift in a few pixels may significantly distort the model results, and particularly when mechanistic interpretation of the different substituents on the bioactivity is required. Additionally, it is this alignment procedure which precludes the application of the MIA-QSAR approach to diverse molecular datasets and the discovery of novel lead compounds. The present report seeks to address this challenge using the 2D-Discrete Fourier Transform (2D-DFT) procedure.

Fourier Transform (FT), introduced by Jean Baptiste Joseph Fourier in 1822, is probably one of the most important mathematical theories in modern engineering [26]. Explained in simple terms, Fourier series (or Fourier Transform) enunciates that periodic (or periodized) phenomena may be expressed as a combination of sine and cosine functions of different frequencies, weighted by coefficients of varying magnitudes (known as *Fourier coefficients*). Certainly space does not suffice to explain this rich mathematical theory. For a detailed treatise, see references [27,28]. The FT has found applications in numerous fields ranging from signal processing, spectroscopy, circuit design and certainly digital image processing, which is our interest in the present report. The rudimentary principle is that just like music is a combination of sine and

* Corresponding author.
E-mail address: sjbarigye@gmail.com (S.J. Barigye).

http://dx.doi.org/10.1016/j.chemolab.2015.02.020
0169-7439/

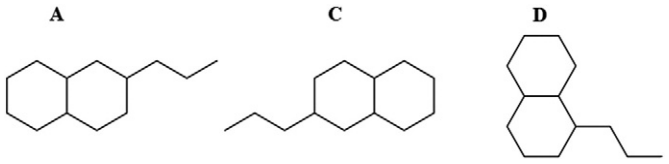

Fig. 1. Molecular structure images used to illustrate the common base property of magnitude spectra generated using the 2D-DFT procedure.

cosine functions, images are a combination of linear spatial functions (parallel lines in the spatial domain) of varying periodicities and orientation. The DFT permits the representation of these constituent functions in the spatial frequency domain. The set of all spatial frequencies is denominated as the *Fourier spectrum*.

The insight of images as two-dimensional functions, f(x, y) where x and y are spatial coordinates rationalizes the application of 2D-DFT over images, yielding *magnitude spectra* in the spatial frequency domain. Such a representation provides a common base for constructing MVIs for sets of non-congruent images with objects of varying sizes, different coordinates in the canvas, or with dissimilar orientations. This approach was previously used by Paul Geladi in the analysis of wood chips of different sizes and shapes, and with random positions and orientation, yielding satisfactory results [29]. The extrapolation of this perspective to the MIA-QSAR method is straightforward: the 2D-DFT is applied to a set of images for molecular structures obtaining magnitude spectra, which are posteriorly used to construct MVIs. Consequently, the MIA-QSAR method is applied to the magnitude-based MVIs. The utility of this approach is that the dependence on the quality of the subjective manual alignment procedure, as a determining factor in the ultimate MIA-QSAR models' performance, is avoided since the magnitude spectra possess a common base and thus do not require alignment. Moreover, the 2D-DFT approach should open way for the QSAR modeling of structurally diverse chemical datasets previously deemed as an exception in the MIA-QSAR context.

2. Materials and methods

2.1. The 2D-Discrete Fourier Transform

The 2D-DFT represents a natural extension of Fourier analysis to phenomena in the R^2 space and is mathematically defined as follows:

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

where u and v are the discrete spatial frequencies, and M and N the number of sampling points in the R^2 space (image dimension). In principle, 2D-DFT constitutes the transformation of phenomena in the spatial domain into the frequency domain. The magnitudes of the 2D-DFT complex coefficients are proportional to the intensity for the spatial frequencies, which constitute the Fourier spectrum. The magnitudes (or magnitude matrix) obtained for a given image function f(x, y) may be expressed as an M × N image, where the image coordinates correspond to the spatial frequencies (u, v) and the pixel values the respective intensities. The key benefit of this procedure from the MIA-QSAR perspective is that magnitude spectra are defined over a common base, allowing the construction of a MVI for previously non-congruent images. It is anticipated that 2D-DFT based magnitude spectra codify similar chemical information as the original molecular structural images and the ensuing molecular descriptors (pixel values) adequately correlate with chemical, physicochemical and biological properties of chemical compounds.

In order to examine the existence of a common base for non-congruent molecular structure images when represented as magnitude spectra, sample illustrative tests were performed considering chemical structures A, B, C and D (see Fig. 1). Chemical structure A (2-propyldecahydronaphthalene) was arbitrarily chosen. Then structure B (not shown) was shifted along the positive x-axis relative to A, while C was rotated π^c relative to A. In principle while the structures A, B and C are not superimposable, these should yield identical magnitude spectra, checked by considering the pixel descriptors generated thereof. Additionally, a structural isomer of A (structure D) was tested to assess if they yielded identical magnitude spectra.

2.2. Chemical datasets for QSAR modeling

To evaluate the usability of the 2D-DFT based magnitude spectra in codifying useful chemical information, QSAR models were built to predict the bioactivities of 3 different sets of chemical compounds previously studied with the MIA-QSAR approach i.e. a series of 2,5-diaminobenzophenone derivatives (antimalarials) [30], 4-phenylpyrrolocarbazole derivatives (tyrosine kinase enzyme WEE1 inhibitors) [31] and benzimidazole derivatives (trichomonicidals) [32], comprised of 92, 97 and 70 chemical compounds, respectively. In all cases, binary (black and white) images were used in order to ensure the homogeneity with the previous studies.

For each of the datasets, magnitude spectra were constructed using the Fast Fourier Transformation (FFT) algorithm implemented in the Matlab program [33]. The ensuing magnitude spectra for each dataset were used to build $l \times m \times n$ MVIs, which were posteriorly unfolded to yield bidimensional $l \times (m \times n)$ data matrices with l representing the instances (chemical compounds) and $(m \times n)$ the variables (pixels). Note that given the symmetrical nature of the magnitude spectra, only a half of the data matrices were considered. The zero variance columns (variables) were deleted as these represent regions of congruence/communality. Each dataset was split into training and test sets, respectively, according to the divisions determined in the original studies, in order to guarantee comparability.

2.3. Model building and validation

Partial Least Squares (PLS) and Multiple Linear Regression (MLR) were used as the calibration methods and the resultant models were vigorously validated using the internal cross validation leave-one out (Q^2loo) and external validation procedures, respectively. The external validation parameters considered in the present study are: the determination coefficient for the test set, the modified determination coefficient with respect to the origin and the slope, denoted by Q^2ext, Q^2_oext and k, respectively, as proposed by Golbraikh and Tropsha [34,35]. The

Table 1
Statistical parameters for antimalarial models obtained with 2D-DFT based MIA-QSAR and other approaches reported in the literature for 2,5-diaminobenzophenone derivatives.

Model	Method	Size	R^2	Q^2loo	R^2p (y-rand)	RMSEcv	Q^2ext	Q^2_oext	k
2D-DFT	MLR	7	0.83	0.79	0.72	0.32	0.81	0.74	0.99
2D-DFT	PLS	5	0.94	0.77	0.56	0.35	0.70 (0.81)[b]	0.53	1.00
MIA-QSAR [29]	PLS	7	0.91	0.56	–	0.45	0.73	–	–
CoMFA [34]	PLS	7	0.87	0.55	–	0.49[a]	0.72	–	–
CoMSIA [34]	PLS	7	0.94	0.63	–	0.45[a]	0.63	–	–

[a] PRESS.
[b] Compound 82 excluded.

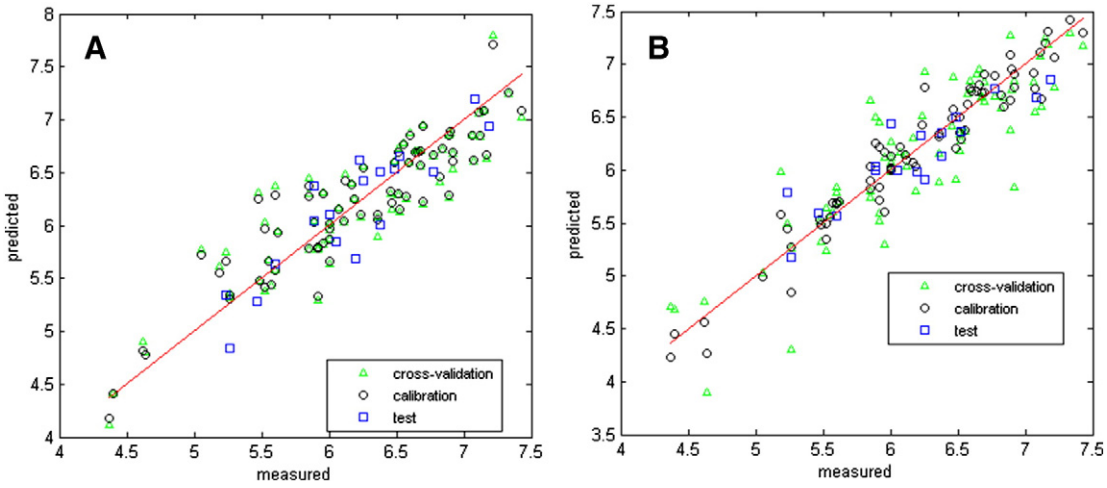

Fig. 2. Plot for experimental versus calibrated and predicted values for the antimalarial activity of the 2,5-diaminobenzophenones, using A) Multiple Linear Regression and B) Partial Least Squares as statistical techniques.

obtained models were evaluated for fortuitous correlation using a 10-fold Y-randomization technique and the penalized determination coefficients R^2p(y-rand) reported (i.e. R^2p(y-rand) > 0.50 is recommended) [36]. Additionally, the application domain (AD) of each of the models was analyzed in order to assess the reliability of the obtained models [37,38]. For all these studies, an in-house software denominated Chemoface was employed [39]. Posteriorly, comparisons with the results reported in the literature were performed. In principle it is anticipated that modeling with the magnitude spectra should yield comparable results to those obtained using original molecular structure images.

3. Results and discussion

3.1. Sample tests for common base for non-congruent images

In this experiment, it was observed that structures A, B and C yield identical magnitude spectra, given that when these images were used to build a MVI, the ensuing data matrix (after the unfolding), produced zero-variance columns in its totality. On the other hand, structures A and D did not yield zero-variance for all the columns. This result shows that 2D-DFT based magnitude spectra possess a common base, and could therefore serve as a suitable alternative to manual alignment of molecular images with respect to the basic molecular scaffold. However, while this is a desirable quality, it should be noted that the ultimate motive for the MIA-QSAR approach is to build models to correlate chemical structure characteristics with their properties. Therefore the usability of the magnitude spectra in correlation studies was evaluated using three classes of compounds.

3.2. 2,5-Diaminobenzophenone derivatives as antimalarial drugs

Previous studies have demonstrated the utility of the 2,5-diaminobenzophenone derivatives as a new class of antimalarials acting particularly against the *Plasmodium falciparum* strain Dd2 [30,40]. For their structures see Supplementary Material, SI1. This dataset was split into sets of 74 and 18 compounds, for the training and test sets, respectively, according to the previous studies. Table 1 shows the statistical parameters of the PLS and MLR models obtained with the 2D-DFT based MIA-QSAR approach.

As can be observed, the 2D-DFT models based on the MLR and PLS techniques generally possess good statistical behavior, yielding improvements in the performance of the MIA-QSAR model based on original images. The high Q^2loo [0.79 (MLR), 0.77 (PLS)], Q^2ext [0.81 (MLR), 0.70 (PLS)] and R^2p(y-rand) [0.72 (MLR), 0.56(PLS)] indicate that the obtained models possess high predictive power and are not prone to chance correlation. Fig. 2 illustrates plots for the experimental and predicted values for the antimalarial activity of the 2,5-diaminobenzophenone derivatives, using MLR and PLS statistical techniques, respectively (see Supplementary material SI2 for table of values). As can be observed, a good relationship between the experimental and predicted values with regard to the calibration, cross and external validation procedures is obtained.

Compound 82 in test set (see Supplementary material SI1 for structure) exhibited strong outlying behavior in the 2D-DFT PLS model and its exclusion from the test set produced an improvement in the Q^2ext value from 0.70 to 0.81. This trend is consistent with the results obtained with the CoMFA and COMSIA strategies, where compound 82 yielded the highest residual values for all the reported models

Table 2
Statistics for models obtained with 2D-DFT based MIA-QSAR and other strategies reported in the literature for 4-phenylpyrrolocarbazoles as tyrosine kinase enzyme WEE1 inhibitors.

Model	Method	Size	R^2	Q^2loo	R^2p (y-rand)	Q^2_{ext}	$RMSE_{CV}$	Q^2_oext	k
2D-DFT	MLR	8	0.838	0.794	0.704	0.780	0.400	0.476	0.999
2D-DFT	PLS	5	0.950	0.791	0.537	0.733	0.403	0.563	1.010
MIA-QSAR	PLS [30]	5	0.788	0.585	–	0.305	0.558	–	–
MIA-QSAR	N-PLS [30]	5	0.736	0.602	–	0.499	0.555	–	–
MIA-QSAR	LS-SVM [30]	(100[b],3[c])	0.944	0.901	–	0.933	0.279	–	–
CoMFA	PLS [35]	5	0.962	0.828	–	0.904	0.380	–	–
CoMSIA	PLS [35]	5	0.949	0.796	–	0.783	0.414	–	–
2D-DFT[a]	MLR	8	0.819	0.771	0.695	0.896	0.431	0.676	1.006
2D-DFT[a]	PLS	5	0.952	0.809	0.518	0.847	0.403	0.835	0.966

[a] All compounds.
[b] Relative weight of regression error.
[c] Kernel parameter.

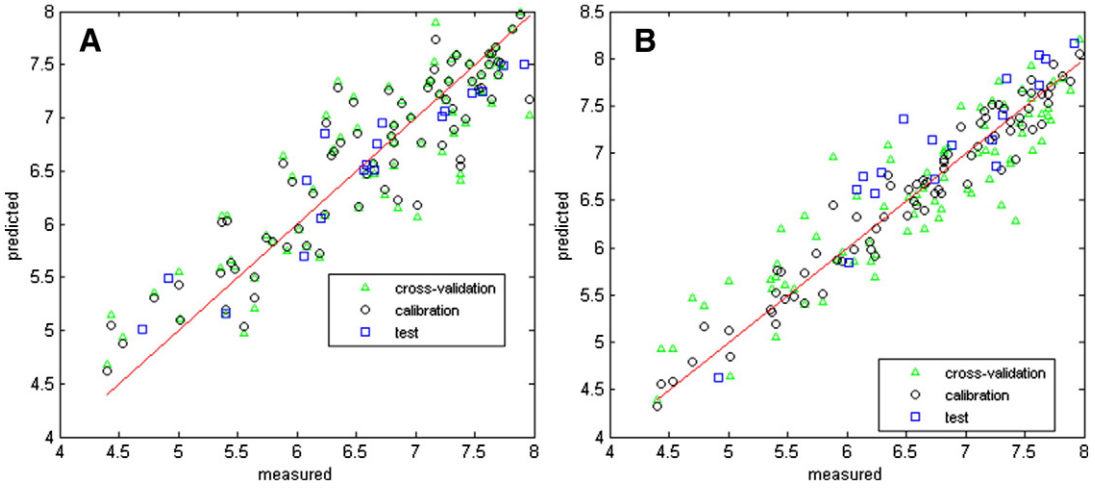

Fig. 3. Plot for observed versus fitted and predicted values of WEE1 inhibition of 4-phenylpyrrolocarbazole derivatives (entire set used) based on A) Multiple Linear Regression and B) Partial Least Squares statistical methods.

[40]. On the other hand, this compound does not exhibit outlying behavior with the MLR-based model and therefore its exclusion is not justified (for William's plots for the PLS- and MLR-based models see Supplementary material SI3). While comparable results with those obtained using the MIA-QSAR approach on original molecular images would suffice to consider the magnitude spectra based models satisfactory, an improvement in the performance of the prediction models constitutes an added advantage. This suggests that the representation of structural information in the spatial frequency domain achieves better stratification of the structural patterns yielding greater correspondence between the 2,5-diaminobenzophenone derivatives and their antimalarial activity.

3.3. Tyrosine kinase enzyme WEE1 inhibitors

The tyrosine kinase WEE1 plays a vital role in the repair of damages in DNA strands and thus justifying the importance of its inhibition in cancer therapy. The 4-phenylpyrrolocarbazole derivatives have been reported in the literature as selective inhibitors of this protein active at the G2/M cell cycle checkpoint. QSAR models to predict the antineoplastic activity of this series of compounds based on the methods: CoMFA, CoMSIA and MIA-QSAR, respectively, have been reported in the literature and new compounds with greater activity proposed [31,41]. It is therefore of interest to evaluate the effect of employing magnitude spectra, obtained using the 2D-DFT, on the quality of MIA-QSAR based models. In all the previous studies, 10 compounds were excluded from the dataset due to outlying behavior, and thus the final study set comprised of 87 compounds (see Supplementary material SI4). In order to maintain the uniformity with previous studies, these compounds were excluded a priori, and the membership of the training and test sets kept similar. However, considering the fact that outlying behavior is sometimes inherent to the method employed, an additional experiment was performed where the entire dataset was considered. In this sense, the dataset was split into training (79 compounds) and test (18 compounds) sets, respectively, using the Kennard-Stone algorithm.

Table 2 shows the statistical parameters of the MLR and PLS-based models obtained with the magnitude spectra (using 2D-DFT), as well as comparisons with the results reported in the literature. As can be observed, the 2D-DFT based models generally present good statistical behavior, and more importantly yield improvements in the performance of the MIA-QSAR models obtained from original images (aligned using manual procedure), particularly if the internal (Q^2loo) and external validation (Q^2_{ext}) parameters are compared. Slightly superior performance is obtained with the Least Squares Support Vector Machine (LS-SVM) based MIA-QSAR, CoMFA and COMSIA models, respectively. However, it should be noted that LS-SVM is a much more robust technique which incorporates non-linear considerations in associations between the descriptors and the modeled bioactivity. As for the CoMFA and COMSIA, it is known that these are computationally involved methods entailing prior structural optimization and the computation of steric or electrostatic interaction energies at numerous grid points. On the contrary, the FFT involves a low computational cost and the symmetric nature of the magnitude spectra means that only half of the dataset matrix is considered. A probably more interesting result is the one obtained when the entire dataset (97 compounds) was considered. As can be observed in Table 2, the MLR and PLS models built on the entire dataset yielded comparable results to those of models built with the reduced dataset (87 compounds). Fig. 3 shows plots for the experimental and predicted values based on the statistical techniques MLR and PLS, respectively (for the entire dataset of the 4-phenylpyrrolocarbazole derivatives see Supplementary material SI5 for table of values).

As can be seen, good linear correlation between the experimental and predicted values is obtained with regard to the calibration, cross and external validation techniques (for William's plots for the PLS- and MLR-based models see Supplementary material SI6). This result supports further the inference that 2D-DFT achieves greater stratification of structural patterns in the spatial frequency domain, in the sense that sensitivity to progressive structural modifications is enhanced which in turn produces greater statistical correspondence with the modeled chemical properties.

Table 3
Comparison between statistics for magnitude spectra-based MIA-QSAR models and other approaches reported in the literature for trichomonicidal Benzimidazoles.

Model	Method	Size	R^2	Q^2loo	R^2p (y-rand)	RMSEcv	Q^2ext	Q^2_oext	k
2D-DFT[a]	MLR	7	0.852	0.802	0.726	0.356	0.887	0.787	0.999
2D-DFT[a]	PLS	4	0.929	0.676	0.513	0.459	0.788	0.617	0.989
MIA-QSAR	PLS [31]	5	0.853	0.519	–	0.543	0.778	–	–
CoMFA	PLS [36]	5	0.927–0.936	0.601–0.634	–	–	0.729–0.890	–	–
CoMSIA	PLS [36]	2–4	0.573–0.915	0.483–0.642	–	–	0.562–0.873	–	–

[a] All compounds.

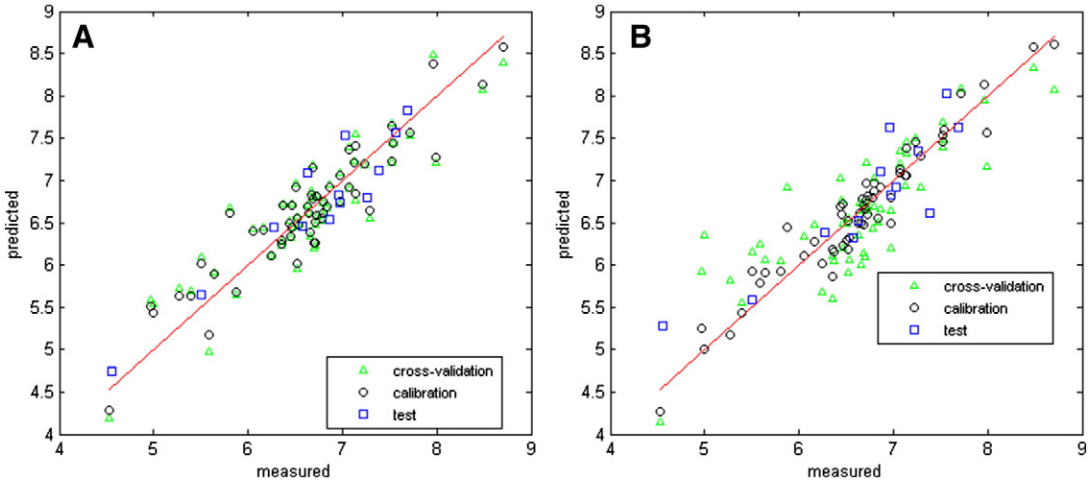

Fig. 4. Linear correlation for experimental versus calibration and validation values for trichomonicidal activity of Benzimidazoles employing A) Multiple Linear Regression and B) Partial Least Squares.

3.4. Benzimidazole derivatives (trichomonicidals)

Finally, a third dataset comprised of a series of 70 benzimidazole compounds was considered for modeling the bioactivity against *Trichomonas vaginalis* (see Supplementary material SI7 for structures). This dataset was employed in previous correlation studies using the MIA-QSAR, CoMFA and COMSIA approaches, respectively, and satisfactory results were obtained. Previous studies, singled out compounds 28 and 31 as outlying compounds, and therefore were initially excluded from the dataset, and the same division of training (55 molecules) and test (13 compounds) sets was performed to maintain the correspondence in the experimental conditions. However, the reinclusion of compounds 28 and 21 in the test set did not yield a significant alteration of the Q^2ext values and so their elimination was not justified. Table 3 shows the statistical parameters of the MLR and PLS models obtained with the MIA-QSAR strategy on 2D-DFT based magnitude spectra for molecular images.

Generally, the obtained models possess good statistical performance (see Fig. 4 as well), evidenced by the quality of the models' calibration and validation parameters, which are observed to be comparable to superior to those reported in the literature. The experimental and predicted trichomonicidal activity values using the PLS- and MLR-based models and the corresponding William's plots are available as Supplementary material SI8 and SI9, respectively. Additionally, previously observed outlying behavior is not exhibited with the 2D-DFT based models. Note that the MLR based model generally possesses superior performance than the PLS-based models, bearing in mind that the R^2 parameter (only parameter where MLR model does not exhibit markedly superior performance) tends to overestimate the models variance and is thus not a credible parameter for evaluating the model's performance.

Altogether, based on the satisfactory results obtained it may be inferred that the 2D-DFT of molecular structural images into magnitude spectra does not obliterate the underlying chemical information and thus constitutes a suitable alternative to the troublesome and sometimes subjective alignment procedure. Additionally, the improvements obtained with some of the models of the studied bioactivities suggest that the representation of the chemical information in the frequency spatial domain achieves better stratification of chemical information and thus enhancing the performance of the MIA-QSAR approach.

4. Conclusions

The usefulness of the 2D-DFT in dealing with the challenge of manual superposition of molecular images in the MIA-QSAR approach has been demonstrated. The 2D-DFT permits the analysis of images in the spatial frequency domain through magnitude spectra which possess a common base on which MVIs may be constructed. Replicas of previously performed studies for different bioactivities demonstrated that the use of magnitude spectra does not aggravate the performance of the MIA-QSAR strategy, but may probably contribute to improved performance of this approach. While the immediate benefit of this transform from the MIA-QSAR approach is elimination of the wearisome manual alignment, it is anticipated that this strategy should enable the generalization of the MIA-QSAR approach to consider structurally diverse datasets, other than series of congeneric datasets. This corollary will constitute the basis for future tasks. Additionally, other approaches for creating common bases for non-congruent images, such as the use of principal component score plots, as well as other transforms, for example the Discrete Cosine, Wavelets and Walsh–Hadamard Transforms will be explored.

Tables for chemical datasets, the experimental and predicted values for built models, as well as the entire datasets employed in each of the studies are available free of charge as supplementary material on the web at http://dx.doi.org/10.1016/j.chemolab.2015.02.020.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgment

Barigye, S. J. and Freitas, M.P. acknowledge financial support from CNPq (Grant Numbers: 505878/2013-5; 302127/2013-5) and FAPEMIG (Grant Number: CEX-PPM-00033/13).

References

[1] M.P. Freitas, S.D. Brown, J.A. Martins, MIA-QSAR: a simple 2D image-based approach for quantitative structure–activity relationship analysis, J. Mol. Struct. 738 (2005) 149–154.
[2] M.P. Freitas, MIA-QSAR modelling of anti-HIV-1 activities of some 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners, Org. Biomol. Chem. 4 (2006) 1154–1159.
[3] M. Freitas, A 2D image-based approach for modelling some glycogen synthase kinase 3 inhibitors, Med. Chem. Res. 16 (2007) 461–467.
[4] J.E. Antunes, M.P. Freitas, R. Rittner, Bioactivities of a series of phosphodiesterase type 5 (PDE-5) inhibitors as modelled by MIA-QSAR, Eur. J. Med. Chem. 43 (2008) 1632–1638.
[5] M.P. Freitas, E.F.F. da Cunha, T.C. Ramalho, M. Goodarzi, Multimode methods applied on MIA descriptors in QSAR, Curr. Comput. Aided Drug Des. 4 (2008) 273–282.
[6] M. Goodarzi, M.P. Freitas, Augmented three-mode MIA-QSAR modeling for a series of anti-HIV-1 compounds, QSAR Comb. Sci. 27 (2008) 1092–1097.

[7] M. Bitencourt, M.P. Freitas, Bi- and multilinear PLS coupled to MIA-QSAR in the prediction of antifungal activities of some benzothiazole derivatives, Med. Chem. 5 (2009) 79–86.
[8] M. Goodarzi, M.P. de Freitas, MIA-QSAR modelling of activities of a series of AZT analogues: bi- and multilinear PLS regression, Mol. Simul. 36 (2009) 267–272.
[9] M. Goodarzi, M.P. Freitas, On the use of PLS and N-PLS in MIA-QSAR: azole antifungals, Chemom. Intell. Lab. Syst. 96 (2009) 59–62.
[10] G.R. Lloret, Á. Cunha Neto, R. Rittner, M. Bitencourt, M.P. Freitas, N.S. Aquino, Synthesis and rational design of anti-inflammatory compounds: N-phenyl-cyclohexenyl sulfonamide derivatives, J. Phys. Org. Chem. 22 (2009) 1188–1192.
[11] M. Goodarzi, M.P. Freitas, 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, Eur. J. Med. Chem. 45 (2010) 1352–1358.
[12] M. Goodarzi, M.P. Freitas, 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, Med. Chem. 7 (2011) 645–654.
[13] J.M. Silla, C.A. Nunes, R.A. Cormanich, M.C. Guerreiro, T.C. Ramalho, M.P. Freitas, MIA-QSPR and effect of variable selection on the modeling of kinetic parameters related to activities of modified peptides against dengue type 2, Chemom. Intell. Lab. Syst. 108 (2011) 146–149.
[14] M. Bitencourt, M.P. Freitas, R. Rittner, The MIA-QSAR method for the prediction of bioactivities of possible acetylcholinesterase inhibitors, Arch. Pharm. 345 (2012) 723–728.
[15] M. Bitencourt, M.P. Freitas, MIA-QSAR evaluation of a series of sulfonylurea herbicides, Pest Manag. Sci. 64 (2008) 800–807.
[16] M. Goodarzi, M.P. Freitas, T.C. Ramalho, Prediction of 13C chemical shifts in methoxyflavonol derivatives using MIA-QSPR, Spectrochim. Acta A 74 (2009) 563–568.
[17] M. Goodarzi, M.P. Freitas, PLS and N-PLS-based MIA-QSTR modelling of the acute toxicities of phenylsulphonyl carboxylates to *Vibrio fischeri*, Mol. Simul. 36 (2010) 953–959.
[18] M.R. Freitas, S.V.B.G. Matias, R.L.G. Macedo, M.P. Freitas, N. Venturin, Augmented multivariate image analysis applied to quantitative structure–activity relationship modeling of the phytotoxicities of benzoxazinone herbicides and related compounds on problematic weeds, J. Agric. Food Chem. 61 (2013) 8499–8503.
[19] C.A. Nunes, M.P. Freitas, aug-MIA-QSPR on the modeling of sweetness values of disaccharide derivatives, LWT Food Sci. Technol. 51 (2013) 405–408.
[20] M.R. Freitas, S.J. Barigye, M.P. Freitas, Coloured chemical image-based models for the prediction of soil sorption of herbicides, RSC Adv. 5 (2015) 7547–7553.
[21] M.C. Guimarães, E.G. da Mota, D.G. Silva, M.P. Freitas, aug-MIA-QSPR modelling of the toxicities of anilines and phenols to *Vibrio fischeri* and *Pseudokirchneriella subcapitata*, Chemom. Intell. Lab. Syst. 134 (2014) 53–57.
[22] M. Freitas, M. Freitas, R.G. Macedo, Aug-MIA-QSPR modeling of the soil sorption of carboxylic acid herbicides, Bull. Environ. Contam. Toxicol. 93 (2014) 489–492.
[23] C.A. Nunes, M.P. Freitas, Introducing new dimensions in MIA-QSAR: a case for chemokine receptor inhibitors, Eur. J. Med. Chem. 62 (2013) 297–300.
[24] C.A. Nunes, M.P. Freitas, aug-MIA-QSAR modeling of antimicrobial activities and design of multi-target anilide derivatives, J. Microbiol. Methods 94 (2013) 217–220.
[25] M.H. Duarte, S.J. Barigye, M.P. Freitas, Exploring MIA-QSAR's for antimalarial quinolon-4(1H)-imines, Comb. Chem. High Throughput Screen. 18 (2015) 208–216.
[26] J.B.J. Fourier, Théorie analytique de la chaleur, Chez Firmin Didot, père et fils, Paris, France, 1822.
[27] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 3rd ed. Prentice Hall, Upper Saddle River, NJ, 2007.
[28] E.R. Dougherty, Digital Image Processing Methods, CRC Press, Boca Raton, FL, 1994.
[29] P. Geladi, Some special topics in multivariate image analysis, Chemom. Intell. Lab. Syst. 14 (1992) 375–390.
[30] R.A. Cormanich, M.P. Freitas, R. Rittner, 2D chemical drawings correlate to bioactivities: MIA-QSAR modelling of antimalarial activities of 2, 5-diaminobenzophenone derivatives, J. Braz. Chem. Soc. 22 (2011) 637–642.
[31] R.A. Cormanich, M. Goodarzi, 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) 244–252.
[32] R.A. Cormanich, C.A. Nunes, M.P. Freitas, Desenhos de Estruturas Químicas Correlacionam-se com Propriedades Biológicas: MIA-QSAR, Quim. Nova 35 (2012) 1157–1163.
[33] MATLAB, The MathWorks Inc. 2010. (Natick, Massachusetts).
[34] A. Golbraikh, A. Tropsha, Beware of q2! J. Mol. Graph. Model. 20 (2002) 269–276.
[35] G. Melagraki, A. Afantitis, Enalos KNIME nodes: exploring corrosion inhibition of steel in acidic medium, Chemom. Intell. Lab. Syst. 123 (2013) 9–14.
[36] P. Pratim Roy, S. Paul, I. Mitra, K. Roy, On two novel parameters for validation of predictive QSAR models, Molecules 14 (2009) 1660–1701.
[37] S. Zhang, A. Golbraikh, S. Oloff, H. Kohn, A. Tropsha, A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models, J. Chem. Inf. Model. 46 (2006) 1984–1995.
[38] G. Melagraki, A. Afantitis, Enalos InSilicoNano platform: an online decision support tool for the design and virtual screening of nanoparticles, RSC Adv. 4 (2014) 50713–50725.
[39] C.A. Nunes, M.P. Freitas, A.C.M. Pinheiro, S.C. Bastos, Chemoface: a novel free user-friendly interface for chemometrics, J. Braz. Chem. Soc. 23 (2012) 2003–2010.
[40] A. Xie, P. Sivaprakasam, R.J. Doerksen, 3D-QSAR analysis of antimalarial farnesyltransferase inhibitors based on a 2, 5-diaminobenzophenone scaffold, Bioorg. Med. Chem. 14 (2006) 7311–7323.
[41] P. Yi, X. Fang, M. Qiu, 3D-QSAR studies of Checkpoint Kinase WEE1 inhibitors based on molecular docking, CoMFA and CoMSIA, Eur. J. Med. Chem. 43 (2008) 925–938.