MOLECULAR SIMULATION
2020, VOL. 46, NO. 14, 1055–1061
https://doi.org/10.1080/08927022.2020.1800691

Taylor & Francis
Taylor & Francis Group

Conformational fingerprints in the modelling performance of MIA-QSAR: a case for SARS-CoV protease inhibitors

Joyce K. Daré, Daniela R. Silva, Teodorico C. Ramalho and Matheus P. Freitas

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

ABSTRACT

Multivariate image analysis applied to quantitative structure–activity relationships (MIA-QSAR) has proved to be a high-performance 2D tool for drug design purposes. Nonetheless, MIA-QSAR strategy does not efficiently incorporate conformational information. Therefore, understanding the implications of including this type of data into the MIA-QSAR model, in terms of predictability and interpretability, seems a crucial task. Conformational information was included considering the optimised geometries and the docked structures of a series of disulfide compounds potentially useful as SARS-CoV protease inhibitors. The traditional analysis (based on flat-shape molecules) proved itself as the most effective technique, which means that, despite the undeniable importance of conformation for biomolecular behaviour, this type of information did not bring relevant contributions for MIA-QSAR modelling. Consequently, promising drug candidates were proposed on the basis of MIA-plot analyses, which account for PLS regression coefficients and variable importance in projection scores of the MIA-QSAR model.

ARTICLE HISTORY
Received 26 March 2020
Accepted 17 July 2020

KEYWORDS
SARS-CoV; QSAR; molecular docking; conformation

1. Introduction

Quantitative Structure–Property Relationship (QSPR) is a broad subarea from molecular modelling dedicated to correlate chemical structural features of compounds to a property of interest (*e.g.* biological activity), which is performed through the generation of a predictive mathematical model (regression equation) [1]. In other words, a consistent QSPR model can predict a specific property for compounds that have not been previously studied, based on structurally similar molecules that have already been investigated. This predictive ability has great contribution in identifying novel drugs, once it accelerates the screening step of this long process.

Although QSPR is a recent molecular modelling tool, its roots go back to 1868–1869, when Crum Brown and Fraser established that, indeed, there was a correlation between the physiological action of a substance and its chemical composition and constitution [2]. Since then, QSAR has gained a lot of attention among researchers, which is also related to the improvements on the computational area, once this technological device facilitates and accelerates the extraction of descriptors and generation of prediction models. Furthermore, allied with molecular docking, QSAR avoids the experimental tests of lots of unpromising compounds, which results in a large saving of costs and time. Due to these relevant contributions, QSAR has become the target of many studies.

There is a variety of methods to generate QSAR models and, basically, the remarkable difference among them resides in what types of descriptors are employed in the analysis and in what way these parameters are obtained from the data set. Commonly, the different approaches are classified from 1D to 7D [3]; each of them has its own advantages and drawbacks [4]. MIA-QSAR (Multivariate Image Analysis applied to QSAR) is a relevant 2D approach that has been employed in many occasions in the last decade, yielding satisfactory outcomes, comparable to 3D results [5–8].

Essentially, MIA-QSAR is a method based on the treatment of 2D molecular images and, from these symbolic molecular representations, the required descriptors are obtained. The molecules are built in a way that their atoms have sizes proportional to van der Waals radii and their colours are numerically described to be proportional to the respective electronegativities. The colour system is the RGB (Red, Green, and Blue). Furthermore, in the current version of this technique, the user counts on the aid of the MIA-plots tool, which comprises two-coloured graphics that help in the interpretation of the generated model [9,10]. These plots are based on PLS regression coefficients (**b**-plot) and Variable Importance in Projection (VIP) scores, and are often employed to evaluate how and how much each substituent in the molecule data set contributes to the biological activity.

Despite the MIA-QSAR modelling power, traditionally, it does not codify conformational information in an effective way. Thus, an attempt has been made in order to include spatial fingerprints into MIA-QSAR descriptors [11]. Although the authors have succeeded in including conformational information into the MIA-QSAR descriptors, the chosen procedure did not result in any improvements in terms of model predictability, when compared to the common analysis [11]. Therefore, this work aims to investigate the inclusion of a more significant type of tridimensional information into the

CONTACT Matheus P. Freitas matheus@ufla.br
Supplemental data for this article can be accessed at https://doi.org/10.1080/08927022.2020.1800691

descriptors, which involves the obtaining of most-likely bioactive poses for the selected ligand set.

For evaluating the influence of tridimensional features on the modelling performance of MIA-QSAR models, a series of SARS-CoV M^{pro} (severe acute respiratory syndrome coronavirus main protease) inhibitors, synthesised by Wang and coworkers, was chosen [12]. SARS is an infective disease that affects the respiratory system, caused by an RNA type of virus known as coronavirus (CoV) [13]. This large family of viruses affects not only humans, but also a variety of other animals. Currently, six human coronaviruses (HCoVs) have been identified, and, although they have been known for decades, only in 2002 they gained clinical importance as a result of an outbreak of SARS and MERS (Middle East respiratory syndrome – another disease caused by a coronavirus agent) [13]. Since these severe epidemies, which are now under control, many targets have been studied; among them, the $3CL^{pro}$, also known as M^{pro} (main protease), has been considered a pertinent approach for inhibition [14]. Therefore, different inhibitors for SARS-CoV-M^{pro} have been proposed; Wang et al. [12] synthesised unsymmetrical aromatic disulfide compounds, which exhibited an encouraging biological potency. The SARS-CoV shares most of its genome with COVID-19, that originated in Wuhan, China in early 2020 [15]. Thus, our findings can also be helpful to lead the design of molecules for the treatment of other diseases, such as MERS-CoV and COVID-19 infections.

2. Material and methods

The anti-SARS molecules, obtained from the literature [13], are shown in Table S1 (Supplemental Material) and their IC_{50} values are given in μM. The data set of 40 compounds were split into training (*ca.* 75% of compounds) and test (*ca.* 25% of compounds) sets through Kennard-Stone sampling, which were used for model calibration (using partial least squares – PLS – regression) and internal validation through leave-one-out cross-validation (the training set), and for external validation (test set). This study was divided into three main parts, which are presented in the next sections.

2.1. Traditional MIA-QSAR modelling

Initially, a traditional MIA-QSAR modelling was performed. The detailed procedure for building MIA-QSAR models has been described elsewhere [5–10]; thus, herein, only the main steps are explained. First, the compounds were drawn in GaussView program [16], maintaining the congeneric centre (the disulfide bond) at the same exactly position in all molecules. Subsequently, the 2D images (.bmp files) were loaded and superposed (Figure 1(a)), yielding a three-way array, which was further unfolded in a new matrix with dimensions 40 × 108,340. The Chemoface software was employed for accomplishing the superposing and unfolding tasks (available at http://ufla.br/chemoface/) [17].

Next, the invariant columns (corresponding to the blank spaces and the congeneric centre) were deleted, and a new matrix, with dimensions 40 × 12,215, was generated. This new matrix had its default pixel values (generated automatically by the GaussView software [16]) replaced by numbers proportional to (i) the respective values of r/ε ratio (van de Waals radius/Pauling's electronegativity), (ii) the van der Waals radius values, and (iii) the respective electronegativity values. It is worth mentioning that the default values attributed to each colour follow an RGB (red, green, and blue) system pattern, *i.e.* the values vary from 0 (corresponding to the colour 'black') to 765 ('white').

To perform the QSAR model and for reasons of comparison with the study performed by Wang et al. [12], compounds **8**, **23**, and **40** were deleted, once they were considered outliers by those authors. In that study, external validation was not performed, thus, we decided to build two models for each data set (r/ε, r and ε): one not considering a test set, in this way the model would have the same dimensional space; and, one considering a group for external validation, because it is well-

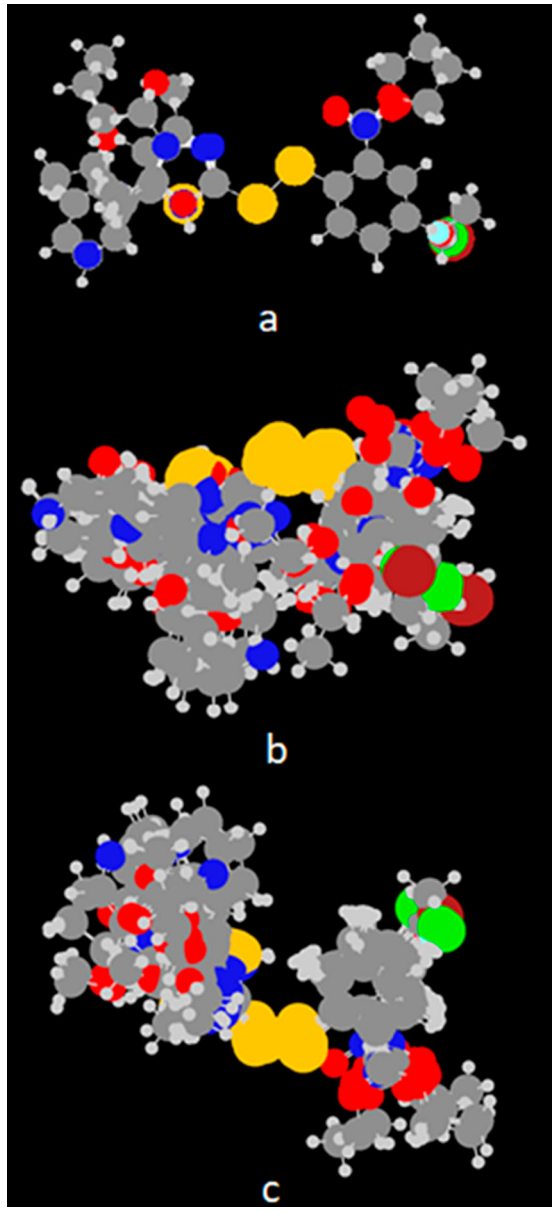

Figure 1. (Colour online) Superposition of the images for disulfide molecules used in the (**1a**) traditional MIA-QSAR procedure, (**1b**) MIA-QSAR with optimised geometries, and (**1c**) MIA-QSAR with docked structures.

known that this type of validation is extremely informative in terms of model efficiency and reliability. Therefore, six prediction models were built in total, through traditional analysis. The validation parameters for the two best models, one without the test set (A) and one with the test set (B), are shown in Table 1.

The test sets were selected through the Kennard-Stone algorithm and contained 9 elements each (marked with the symbol '[a]' in Table S1 – Supplemental Material). For all models, partial least squares (PLS) regression was employed, and the optimum number of latent variables (LV) was chosen by analysing the decay of the root mean square error (RMSE) in the leave-one-out cross-validation (LOOCV).

For models without a test set, the following parameters were employed to analyse the quality of the model: determination coefficients in calibration (r^2) and cross-validation (q^2), and their respective RMSE's. The quality for the r^2 parameter was assessed considering the criteria established by Alexander et al. [18] and Kiralj and Ferreira [19]. The risk of chance correlation in calibration was analysed using the ${}^{c}R^2_p$ parameter [${}^{c}R^2_p = r\ (r^2 - r^2_{rand})^{1/2}$, where r^2_{rand} corresponds to the mean determination coefficient value obtained after randomising the y block ten times] [20]. All these parameters were retrieved from the Chemoface software [17].

For models containing test set, in addition to r^2 and q^2, r^2_{test} and its corresponding RMSE were also considered. The proximity between the experimental and predicted IC_{50} values for the test set was statistically evaluated using r^2_m parameter [21]. The risk of chance correlation was studied using the same parameter mentioned above.

2.2. MIA-QSAR modelling using optimised molecular geometry projections

The second part of this study consisted in inserting conformational information into the MIA-QSAR descriptors through the process of structural optimisation. This step followed the same rationalisation established by Daré et al. [11] and it was performed with the goal of showing the similarities and differences between the different types of tridimensional information codified into the descriptors. In summary, a conformational screening was performed using the Spartan'16 program [22] for all 40 compounds at the semi-empirical AM1 (Austin Model 1) level of theory [23]. The lowest energy conformation was selected for each case and, then, they were fully optimised, using the Density Functional Theory (DFT) method at the ωB97X-D/6-31G(d,p) level of theory [24,25], in the Gaussian 09 program [26].

Next, the optimised structures were loaded on Discovery Studio Visualizer [27] and the alignment of the best poses was performed. The congeneric centre (*i.e.* the disulfide moiety bonded to a phenyl group on the right side) was used as reference during the superposition step and the resulting three-way array can be observed in Figure 1(b). Then, the superposed molecules were loaded on GaussView program [16] and saved as bitmap files, as in a usual routine MIA-QSAR analysis. Finally, the MIA-QSAR models were generated following the same steps of the traditional technique, *i.e.* six more models were obtained; the validation parameters for the two best models are shown in Table 1 [model C (without a test) and model D (with a test)].

Table 1. Statistical parameters obtained through traditional MIA-QSAR technique (Models A and B), MIA-QSAR applied to optimised molecular geometries (Models C and D), and the same technique employed to bioconformation-like images (Models E and F).

	Traditional		Optimised		Docked	
Parameters	Model A (r/ε)	Model B (r/ε)	Model C (r/ε)	Model D (r)	Model E (ε)	Model F (ε)
LV	3	3	3	2	3	3
$RMSE_{cal}$	0.527	0.465	0.445	0.484	0.400	0.311
r^2_{cal}	0.870	0.897	0.907	0.905	0.925	0.952
$RMSE_{y\text{-}rand}$	1.183	1.036	0.738	0.557	0.602	0.388
$r^2_{y\text{-}rand}$	0.345	0.488	0.742	0.871	0.830	0.925
${}^{c}R^2_p$	0.676	0.606	0.388	0.176	0.297	0.161
$RMSE_{cv}$	0.780	0.794	1.022	1.234	1.119	1.213
q^2	0.715	0.701	0.524	0.392	0.420	0.284
$RMSE_{test}$	–	0.715	–	0.593	–	0.946
r^2_{test}	–	0.781	–	0.823	–	0.601
r^2_m	–	0.653	–	0.812	–	0.268

[a]CoMFA results from Wang et al. [12]: 6 LV's, $r^2 = 0.916$, Standard Error = 0.088, and $q^2 = 0.681$.

2.3. MIA-QSAR modelling using bioactive-like conformation projections

For the main part of this work, a different and more significant type of tridimensional information was employed: bioactive-like conformations. These poses were obtained after docking the disulfide compounds into the active site of SARS-CoV M^{pro}. In order to perform molecular docking, the crystal structure of SARS-CoV M^{pro}, with 1.85 Å of resolution, was retrieved from the protein data bank (PDB code: 2AMD). This raw structure was subsequently prepared for docking using the 'Protein Preparation Tool' available in the Glide program [28,29]. Some initial tests were performed to select the best features for further docking analysis. The test results were analysed comparing these outcomes with those obtained by Wang and coworkers [12] and also with those obtained by Yang et al. [13] (authors responsible for the isolation and crystallisation process of the referred receptor). Finally, the most adequate preparation process consisted in removing all water molecules and using only the chain A during the docking analysis; these adjustments are discussed in more details in the next section. The protein charges were automatically calculated during protein preparation.

The ligands were prepared in two steps. First, their geometries were optimised using the Gaussian 09 program [26], through the Density Functional Theory (DFT) method at the ωB97x-D/6-31G(d,p) level of theory [23,24]. Secondly, ligand charge calculations were performed using the CHelpG function at the same theoretical level of the optimisation. A final step, before docking, was establishing the Grid Box ($10 \times 10 \times 10$ Å) and the active site of the receptor. Then, the molecular docking technique was performed using the Glide program [28,29] with standard precision (SP) protocol. No constrains were established and a hundred conformations were obtained for each compound.

Compound **31** was used as reference, once it was the most efficient inhibitor determined by Wang et al. [12]. Therefore,

similar bioactive conformations (Figure S1 – Supplemental Material) were searched among the docking results obtained herein. The most similar conformation for each molecule was identified and stored.

The 40 identified conformations were loaded into Discovery Studio Visualizer [27] and superposed in a similar way of the optimised molecules, *i.e.* maintaining the congeneric centre at the same region (Figure 1(c)). This hypermolecule was, then, loaded on GaussView program [16] and saved as bitmap images, which generated 2D projections of the bioactive structures. Employing this strategy, (bio)conformational information was incorporated into the MIA-QSAR descriptors. Finally, MIA-QSAR models were generated, which was performed following the same steps as earlier analysis. Therefore, six more models were obtained; the validation parameters for the two best models are shown in Table 1 [model E (without a test) and model F (with a test)] along with all the other previous best models (A-D).

3. Results and discussion

3.1. QSAR analysis

Regarding the traditional analysis, the resulting superposed molecules are shown in Figure 1(a).

The quality parameters for the best model (A), *i.e.* the model without a test set, are shown in Table 1. In the same table, the validation parameters for all the other models (the best one in each case) are also displayed. Therefore, a general comparison can be performed.

From Table 1, it can be observed that the best models were obtained from the traditional analysis. Model A, employing descriptor values proportional to r/ε, generated good results for internal validation and calibration, but the RMSEs associated with r^2 and q^2 are slightly high. However, the ${}^{c}R_{p}^{2}$ parameter eliminated the risk of chance correlation, which ensures that the model was not a result of randomness. Therefore, comparing to the results obtained by Wang and collaborators [12], which were determined through a CoMFA analysis (see Table S1 footnote), model A was as good as their outcomes (see *e.g.* $r^2 = 0.916$ and $q^2 = 0.681$ from CoMFA, and $r^2 = 0.870$ and $q^2 = 0.715$ from model A). Model B (*r*/ε) reinsures the quality and predictive ability of models obtained through traditional technique, once the external parameters were all acceptable. This last observation shows that model B is not overfitted. Models C (r/ε) and D (r), obtained from the optimised anti-SARS structures, were not as good as those obtained from the traditional procedure. The values of r^2 and q^2 for model C were acceptable, but the risk of chance correlation could not be eliminated by the ${}^{c}R_{p}^{2}$ parameter. For model D, another issue was the q^2 value, because it was not within an acceptable range (≥ 0.5). An interesting observation in this case, is that, for model D, the external parameters presented reliable results; this observation confirms what Tropsha [30] made very clear in his work 'Best practices for QSAR model development, validation, and exploitation', *i.e.* both internal and external validations are required in order to consider a QSAR model as a reliable prediction equation. The resulting superposed molecules employed on the generation of models C and D are shown in Figure 1(b).

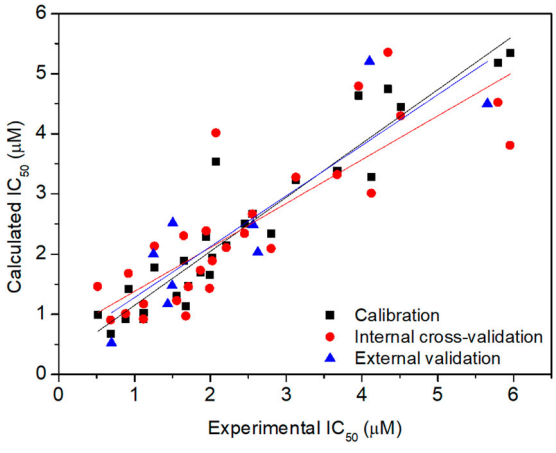

Figure 2. (Colour online) Plot for predicted *vs.* experimental IC_{50} values using the MIA-QSAR model B.

Lastly, models obtained from molecules docked inside the SARS-CoV M^{pro} [E (ε) and F (ε)] did not present acceptable validation parameters either. The risk of chance correlation could not be eliminated from neither model, while even the external validation for model F presented a bad result for r_m^2. The resulting superposed molecules employed on the generation of models E and F are shown in Figure 1(c).

Therefore, it can be concluded that, although conformational information is responsible for a great part of biomolecular behaviour, this type of information does not contribute positively for MIA-QSAR methodology. Similar conclusion can be found elsewhere [31]. This finding can be related with the loss of reference that happens with the projections of the optimised and docked molecular structures. In other words, in a traditional analysis the model can be built assuming that a chemical group is always in a specific position and, based on that position, the effect (increasing/decreasing) on the response variable can be determined and, then, predicted later for a different structure containing the same group at the same position. On another hand, for the optimised and docked molecular structures, due to the position variability in the projections, the identification of a pattern for the effect of the different groups on the response variable is jeopardised and, consequently, the prediction is imprecise.

Furthermore, once the focus of this discussion involves bitmap images, one can rationalise the whole issue in terms of pixels. In the traditional analysis, the samples (rows) containing the same chemical group will, for sure, have the same values attributed to the columns (descriptors) corresponding to that specific portion; on another hand, in a 3D analysis, although two (or more) molecules may have the same chemical group in the same spatial position, they will not, necessarily, be at the same pixel position (columns) because their images include variation in angles and direction. Therefore, there is a lack of pattern, which brings a considerable amount of variability to the descriptors and interferes in the quality of the prediction models.

After determining the best models (A and B), a chemical interpretation can be performed employing the MIA-plots tool. Model B was chosen for the interpretation process, once

it is a more complete QSAR analysis when compared to model A. The calculated IC_{50} values for calibration, internal and external validation are shown on Table S1, whereas the predicted *vs.* measured IC_{50} plot is shown in Figure 2. Figure 3, in turn, shows the MIA-plots obtained for model B. These plots are based on the analysis of the structural moieties most affecting (either enhancing or attenuating) the biological data in terms of PLS regression coefficients (b) and variable importance in projection (VIP) [10]. The applicability domain (William´s plot) for the chosen model was also checked and is shown in Figure S2 (Supplemental Material).

William's plot shows that samples **9** and **10** are outliers, which correspond to compounds **10** and **11**, once compound **8** was previously deleted. However, in order to obtain the same space domain for means of comparison with Wang's work [12], both samples were kept in the data set for further analysis; furthermore, both *outliers* are close to the established limits.

The best MIA-QSAR model was the one with descriptor values proportional to radii/Pauling's electronegativity; from this observation, it can be inferred that both steric and electrostatic effects are relevant for the inhibition activity of the disulfide compounds. In addition, because r/ε-based descriptors best explained the data variance in models A and B, it follows that atomic size and electronegativity together plays a more significant role in this case than these parameters alone.

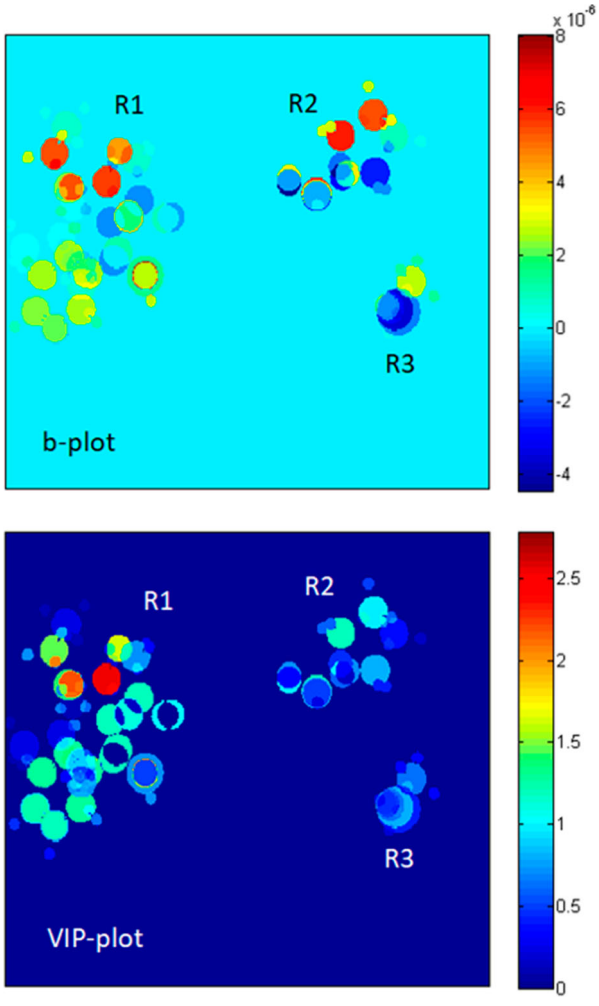

Figure 3. (Colour online) b-plot and VIP plot for model B.

From Figure 3, specifically the b-plot, it is possible to visualise specific moieties (herein called R1, R2, and R3) responsible for increasing/decreasing IC_{50}. Figure 1(a) can collaborate in a more detailed understanding. Focusing on the VIP plot and its general aspects, it is possible to recognise a specific pattern on how strong the three groups affect the response variable. The R1 substituents comprise the most significant variation in colours, which include dark shades of red, yellow, and orange; accordingly, this region can be considered the most meaningful moiety in explaining the observed response pattern. The R2 group also presents some significant variety in shades, but much less than R1, which means that it has less influence in explaining the IC_{50} than the former. Lastly, R3, in a general view, is the one with the smallest contribution for clarifying the IC_{50} behaviour. In summary, R1 > R2 > R3 is the importance scale proposed.

Based on the previous generalisation, it is reasonable to start the b-plot interpretation with the R1 substituents. Furthermore, compounds **31** and **15** can be taken as references for this analysis, once the former corresponds to the smallest value of IC_{50} (most active) and the latter to the largest value (less active) of this series of molecules.

In order to analyse the R1 influence, it seems reasonable to select a subset of compounds containing the same R2 and R3 substituents in all cases. The chosen group comprises the following molecules: **1**, **3**, **12**, **17**, **31**, **33**, and **38**; their structures are shown in Table S1.

Compound **1** has a five-membered ring with endocyclic nitrogen and sulfur atoms at positions 2 and 5; this ring presents a light green colour in the b-plot, which indicates a slight increase in the IC_{50} value. Taking into account that the IC_{50} parameter varies from 0.516 to 5.954 μM, this interpretation agrees with the actual value measured for this compound – 1.871. Actually, one can see that, except for compound **31**, all further molecules presenting a five-membered ring (**3**, **12**, and **17**) have values of IC_{50} larger than those with a six-membered ring; furthermore, these compounds – **3**, **12**, and **17** – contain carboxyl/carbonyl groups (red colour in the b-plot) attached to the five-membered ring, which also contributes, significantly, for the observed increase in the IC_{50} value. Further analysis of the b-plot shows that a blue tone is attributed to the six-membered ring substituent; therefore, the previous conclusion agrees with the MIA-plot interpretation. Moreover, the unsubstituted six-membered ring seems to decrease the IC_{50} even more significantly than the one with two methyl groups attached.

Regarding compound **31**, not only it has a smaller value of IC_{50} than those with a six-membered ring, but also it is the most active compound of the series. Focusing on its groups on the regression coefficient plot, one can see that the moiety that seems to strongly decrease the response variable is the chlorine atom at the R3 position, which is an exception to the previous importance scale established. This fact might be related to a synergistic effect of the three substituents or its specific behaviour inside the active site of the M^{pro} enzyme, which is not explained by MIA-QSPR plots. It is also necessary to understand that the MIA-plot tool shows the general

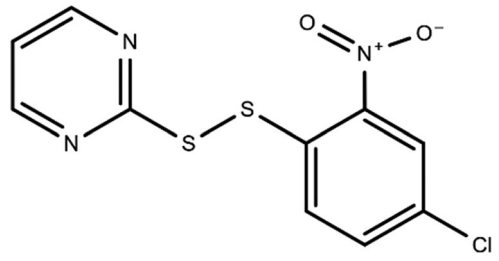

Figure 4. (Colour online) Proposed disulfide compound. The structure contains a six-membered ring with two endocyclic nitrogen atoms at R1, a nitro group at R2, and a chlorine at R3.

behaviour of a data set and, once compound **31** is an exception, it might not be well explained.

Following the same thought, in order to analyse the influence on IC_{50} due to a change in the R2 substituent, a subset with invariable R1 and R3 groups was selected, which comprises compounds **14**, **15**, and **16**. Focusing on the b-plot, it is possible to observe that the ester moieties present strong red/orange colours, which characterises an increasing contribution for IC_{50}. On the other side, the nitro group presents a cyan colour, which shows a decreasing contribution for IC_{50}.

Lastly, for analysing R3 and following the same rationalisation, two subsets were selected: one with compounds **1**, **33**, and **38**; and one with compounds **23**, **34**, and **39**. It is reasonable to conclude from these subsets that a pattern cannot be identified for the R3 contribution in explaining the response variable. This observation agrees with the general assertation that the R3 group did not have a significant influence in explaining the observed IC_{50} pattern.

Lastly, from this analysis, it can be proposed that a disulfide compound containing a six-membered ring with 2 endocyclic nitrogen atoms (positions 2 and 6) at the R1 position, a nitro group at R2 and a chlorine at R3 (Figure 4) seems to be an interesting approach on designing a new SARS-CoV M^{pro} inhibitor. The IC_{50} of the proposed molecule was obtained applying the best MIA-QSPR model and the result was 0.440 μM, which is even better than the corresponding value of compound **31** (most active compound of the series).

3.2. Molecular docking analysis

The pose used as reference (obtained from compound **31**) is shown in Figure S1 (Supplemental Material), as well as some hydrogen bonds established with the amino acid residues considered critical for the inhibition of SARS-CoV M^{pro}.

Such conformation was chosen because of its similarity with that selected by Wang et al. [12], including the interactions established with Cys145 and Gly143 amino acid residues (also found by Wang et al. [12]). Cys 145, according to Wang et al. [12], plays a main role in the activity against SARS-CoV M^{pro}. Apparently, this residue binds to the disulfide compound covalently, which prevents the protein biological action. Therefore, it is critical that the interaction with this residue has been identified herein. Furthermore, the superposition of compound **31** with N9 was also evaluated and it presented a good alignment.

Finally, it is worth discussing the use of only protomer A instead of both units. According to Yang et al. [13], N1 (a similar inhibitor crystallised inside SARS-CoV M^{pro}) binds to protomers A and B of SARS-CoV M^{pro} in an identical and normal manner. Therefore, it was assumed a similar binding mode for the disulfide data set [12]. A test for compound **31** including both protomers was also performed and no meaningful differences were found between including both units and only the chain A. Then, looking for less time-consuming analysis, only protomer A was used, while the correct selection of the most-likely conformation of compound **31** validated our 3D MIA-QSAR modelling.

4. Conclusion

MIA-QSAR, as currently performed, is a powerful tool for building prediction models of biological properties and conformation does not play a significant role in such technique. The most probable reason for the failure in the conformation-based approach is that 2D snapshot from a 3D object leads to a number of lost information, since some parts of the image hide under upper parts. The remaining 3D information lacks from optimal superposition of the congeneric substructure as a result of different (bio)conformational behaviours. Subsequently, connectivity features encoding atomic properties in a well-defined space and comprising the congeneric centre perfectly aligned, such as in the traditional MIA-QSAR approach, seem to be more instructive than 3D information. Although the traditional approach provided a better modelling performance, some 3D-MIA-QSAR results were promising, thus allowing for perspectives for the treatment of three-way data arrays. Therefore, from the extensive validation tests applied, the most reliable QSAR model was that obtained in a traditional way and, from it, a meaningful chemical interpretation was possible. Lastly, the docking procedure helped to bring meaning to the entire analysis, once an acceptable binding mode for the disulfide compounds with their receptor could be proposed.

Acknowledgements

Authors are thankful to FAPEMIG for the financial support of this research (grant numbers CEX-APQ-00383-15 and PPM-00344-17), as well as to CAPES for a studentship (to J.K.D., funding code: 001), and to CNPq for a studentship (to D.R.S.) and fellowships (to T.C.R. and M.P.F.).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico [grant number 301371/2017-2]; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [grant number 001]; Fundação de Amparo à Pesquisa do Estado de Minas Gerais [grant numbers APQ-00383/15 and PPM-00344/17].

References

[1] Hansch C, Leo A, Hoekman DH. Exploring QSAR: hydrophobic, electronic, and steric constants. Washington: American Chemical Society; 1995.

[2] Dearden JC. The history and development of quantitative structure-activity relationships (QSARs). Int J Quant Struct Prop Relat. 2017;1:67–117.

[3] Polanski J. Receptor dependent multidimensional QSAR for modeling drug-receptor interactions. Curr Med Chem. 2009;16:3243–3257.

[4] Consonni V, Todeschini R. Molecular descriptors for chemoinformatics. Verlag: Wiley-VCH; 2009.

[5] Daré JK, Barigye SJ, Freitas MP. Multi-objective modeling of herbicidal activity from an environmentally friendly perspective. Int J Quant Struct Prop Relat. 2017;2:16–26.

[6] Daré JK, Barigye SJ, Freitas MP. Revealing chemophoric sites in organophosphorus insecticides through the MIA-QSPR modeling of soil sorption data. Ecotox Environ Saf. 2017;144:560–563.

[7] Guimarães MC, Silla JM, da Cunha EFF, et al. Is the bioconformation of 5-deoxy-5-fluoro-d-xylulose affected by intramolecular hydrogen bonds? RSC Adv. 2016;6:111681–111687.

[8] Freitas MP, Brown SD, Martins JA. MIA-QSAR: a simple 2D image-based approach for quantitative structure-activity relationship analysis. J Mol Struct. 2005;738:149–154.

[9] Barigye SJ, Duarte MH, Nunes CA, et al. MIA-plot: a graphical tool for viewing descriptor contributions in MIA-QSAR. RSC Adv. 2016;6:49604–49612.

[10] Borges CN, Barigye SJ, Freitas MP. Towards molecular design using 2D-molecular contour maps obtained from PLS regression coefficients. Mol Phys. 2017;115:3044–3050.

[11] Daré JK, Ramalho TC, Freitas MP. 3D perspective into MIA-QSAR: a case for anti-HCV agents. Chem Biol Drug Des. 2019;93:1096–1104.

[12] Wang L, Bao BB, Song GQ, et al. Discovery of unsymmetrical aromatic disulfides as novel inhibitors of SARS-CoV main protease: chemical synthesis, biological evaluation, molecular docking and 3D-QSAR study. Eur J Med Chem. 2017;137:450–461.

[13] Yang H, Xie W, Xue X, et al. Design of wide-spectrum inhibitors targeting coronavirus main proteases. Plos Biol. 2005;3:1742–1752.

[14] Yin Y, Wunderink RGMERS. SARS and other coronaviruses as causes of pneumonia. Respirol. 2018;23:130–137.

[15] Heymann DL, Shindo N. COVID-19: what is next for public health? Lancet. 2020;395:542–545.

[16] Dennington RD, Keith TA, Millam MJ. (2008). GaussView 5.0. Wallingford.

[17] Nunes CA, Freitas MP. Introducing new dimensions in MIA-QSAR: a case for chemokine receptor inhibitors. Eur J Med Chem. 2013;62:297–300.

[18] Alexander DLJ, Tropsha A, Winkler DA. Beware of R^2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. J Chem If Model. 2015;55:1316–1322.

[19] Kiralj R, Ferreira MMC. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. J Braz Chem Soc. 2009;20:770–787.

[20] Mitra I, Saha A, Roy K. Exploring quantitative structure-activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. Mol Simul. 2010;36:1067–1079.

[21] Roy K, Chakraborty P, Mitra I, et al. Some case studies on application of "rm2" metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. J Comput Chem. 2013;34:1071–1082.

[22] Spartan'16 software. Irvine: Wavefunction Inc; 2017.

[23] Dewar MJS, Zoebisch EG, Healy EF, et al. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. J. Am. Chem. Soc. 1985;107:3902–3909.

[24] Chai JD, Head-Gordon M. Systematic optimization of long-range corrected hybrid density functionals. J Chem Phys. 2008;128, 084106.

[25] Krishnan R, Binkley JS, Seeger RPJ. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. J Chem Phys. 1980;72:650–654.

[26] Frisch MJ, Trucks GW, Schlegel HB, et al. Gaussian 09, Revision A.02. Wallingford: Gaussian, Inc; 2016.

[27] Discovery Studio Visualizer 2017R2. San Diego: Dassault Systèmes BIOVIA; 2017.

[28] Friesner RA, Banks JL, Murphy RB, et al. Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. J Med Chem. 2004;47:1739–1749.

[29] Halgren TA, Murphy RB, Friesner RA, et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem. 2004;47:1750–1759.

[30] Tropsha A. Best practices for QSAR model development, validation, and exploitation. Mol Inf. 2010;29:476–488.

[31] Doweyko AM. 3D-QSAR illusions. J Comput Aid Mol Des. 2004;18:587–596.