IJDD

International Journal of Drug Discovery
ISSN: 0975–4423 & E-ISSN: 0975–914X, Vol. 3, Issue 2, 2011, pp-100-117
Available online at http://www.bioinfo.in/contents.php?id=24

QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS (QSAR) OF A SERIES OF KETONE DERIVATIVES AS ANTI-*CANDIDA ALBICANS*

LUIZ FREDERICO MOTTA[1, 2*], WANDA PEREIRA ALMEIDA[2]

[1]Laboratory of Theoretical Chemistry and Chemometrics (LQTC), Department of Chemistry, Federal Institute of Triangulo Mineiro (IFTM), Uberaba, MG, Brazil.

[2]Laboratory of Drug Design (LAFAME), Institute of Chemistry, University of Campinas (UNICAMP), Campinas, SP, Brazil.

[*] Corresponding author, Email: motta@iftm.edu.br or profmotta@gmail.com; + 55-021-34-3312-3557.

Received: October 04, 2011; Accepted: October 25, 2011

Abstract - Candidiasis is recognized worldwide as an opportunistic infection. The severities of the infection increase in immunosuppression conditions with possible occurrence of visceral mycoses and sometimes are widespread and systemic. Increased resistance in strains of *Candida albicans* is a major obstacle to antifungal therapy. The aim of this study was to correlate the chemical structure of compounds with experimental data from biological activity anti-*Candida albicans*. We performed classical QSAR for a series of twenty derivatives of ketone α-β unsaturated against resistant strains of *Candida albicans*. Ninety-four descriptors were calculated and multiparameter model was obtained through Partial Least Squares (PLS) method. The results showed that thermodynamic, dimensional and steric parameters are important in elucidating of action mechanism compounds. Four descriptors (molar refractivity, ionization potential, molecular length and Verloop B4) were selected and good model (n= 20; R^2 = 0.776; SEC = 0.229; $F_{(3,16)}$ = 14.172; Q^2_{LOO} = 0.609; SEV = 0.295; Q^2_{pred} = 0.709; SEP = 0.091; k = 0.709; k' = 1.00; $|R^2_0 - R'^2_0|$ = 0.0009) was built with three latent variables describing 96.14% of the original information. Leave-*N*-out cross validation and Y-randomization analysis were performed in order to confirm the robustness of the model. The proposed model may provide a better understanding of the anti-*Candida albicans* activity of chalcones and can be used as guidance for proposition of new chemopreventive agents.

Key words – QSAR, Chalcones, PLS, Hansch, Antifungal, Glutathione.

Introduction

Opportunistic mycoses are infections caused by fungi of low virulence. Fungi coexist peacefully with the host, but under certain conditions, develop their potential pathogenic to invasion of tissues. The opportunistic fungal infections affect individuals of all ages, race and both sexes. Among the opportunistic infections, candidiasis is one of the most important on humans and has high prevalence. It's caused by several species of fungi of the *Candida* genus, especially *Candida albicans* which normally is part of the commensal microbial flora, particularly at the digestive system (the oropharyngeal cavity and colon) or vaginal [1].

The balance between *C. albicans* and other microorganisms can be changed depending on many factors, which can be either intrinsic or extrinsic. The intrinsic nature is related to the organism, (infections by immune system disorders – acquired immunodeficiency syndrome (HIV), cancer, diabetes, lymphoproliferative disorders (lymphomas, leukemia's); On the other hand, the extrinsic nature is related to external factors, like antibiotics and steroids treatment and contaminated hospital environments [2-3].

According to the CDC (Centers for Disease Control and Prevention) of the United States, people with AIDS have count of T-lymphocyte CD4-positive less than 200 cells/mm^3, then facilitates *C. albicans* infection. An estimative of 90% of HIV-positive patients developed some oral manifestation at sometime in their clinical course. The severities increases in immunosuppressive conditions with possible occurrence of visceral mycoses are sometimes widespread and systemic. The mucous of male and female genitalia and skin folds are much affected due to immunosuppression [3-6]. Systemic candidiasis is serious, when located mainly in the kidneys, heart, bronchi, blood, lungs and brain, being described in 20% to 40% of patients with cancer [7]. Several virulence factors of *C. albicans* are widely researched: the variation antigens cell wall, adhesins, production of proteases and phospholipases, and phenotypic variations [8].

Although more than 100 species of *Candida* have been described, few have been implicated in clinical infections. Clinically, *C. albicans* is the more frequent specie, between 90% to 100% in mucous and 50% to 70% in bloodstream infections. Approximately 95% of *Candida* bloodstream infections are caused by *C. albicans*, *C. glabrata*, *C. krusei*, *C. dubliniensis*, *C. parapsilosis* and *C. tropicalis* [2, 4, 6-9]. The remaining 5% comprises 12-14 different species. The Global Antifungal Surveillance Program shows that the frequency of C. albicans isolated in blood infections increased considerably between the years 1992 (44%) - 2003 (65%) [2,6].

For years, an antifungal agent used widely was Amphotericin B. With the introduction of the triazoles, first-generation drugs, there was considerable progress in the development of chemotherapy. Currently, several Azoles antifungals are used in Anti-*Candida albicans* therapy: fluconazole, itraconazole, voriconazole and posiconazole [9-11]. A group of amphiphilic lipopeptides characterized as Echinocandins (Caspofungin and Micafungin) have also been introduced into therapy, demonstrating potent fungicidal activity [12-13]. Echinocandins inhibit β-glucan synthase through linkage with subunits of this enzyme. Consequently, the effects of the interaction are associated with destabilization of the fungal cell wall, promoting changes in the cytoskeleton and in vesicles transport within the fungus. Have been identified mechanisms of resistance to Azoles and Echinocandins in strains of *Candida albicans* mutants. The mutation mechanism is related to the enzyme involved in the glucan synthesis in the cell wall of fungi [14].

The antifungal activity of chalcones was investigated by several researchers. Previous studies of antifungal activity of compounds showed that some synthetic phenolic chalcones have moderated biological activity [15-16]. Nowakowska recently reported chalcones with antimicrobial properties and anti-inflammatory [17]. Sato *et al*. showed inhibitory properties of chalcones hydroxylated to *Candida* [18]. Tomar *et al*. reported the synthesis and antimicrobial activity of chalcones containing piperazine or 2,5-dichlorothiophene [19]. Lahtchev *et al*. reported a mechanistic study of chalcones with various strains of yeast [20]. López *et al*. [21] reported the synthesis, in vitro antifungal activity and SAR of 41 chalcones and analogues. All active structures were tested for their capacity of inhibiting β (1,3)-glucan synthase and chitin synthase of Saccharomyces cerevisae, enzymes that catalyze the synthesis of the major polymer of the fungal cell wall. Lopez and co-wokers investigated the biological activity of chalcones analogs in the presence of quinolinyl group, revealing that these compounds showed low activity [21-22]. Bag and co-workers synthesized a series of analogues of chalcones incorporating sulfur in the chemical structure of the heteroaromatic ring (thiophene) or thiomethyl group in aromatic ring. The compounds were tested against resistant strains of *Candida albicans* to fluconazole (NCIM 3446) [23].

QSAR methodologies have the potential of decreasing substantially the time and effort required for the discovery of the new medicines [24]. A major step in constructing the QSAR models is to find a set of molecular descriptors that represents variation of the structural properties of the molecules [25]. The QSAR analysis employs statistical methods to drive quantitative mathematical relationships between chemical structure and biological activity [26-27]. Thus, the use of the QSAR in the development of a theoretical model to predict the biological activity of a set of compounds is very important. The strategy used in the QSAR methodology includes the following steps: (1) selection of a data set; (2) generation of the molecular structures; (3) optimization of the geometry of the molecular structures by appropriate method; (4) generation of several structural descriptors; (5) application of variable selection or/and methods data reduction of the calculated descriptors; (6) regression analysis; and finally (7) evaluation of the validity and predictability of the developed QSAR models [26].

Materials and Methods

The purpose of the present work is to perform a quantum chemical QSAR study of the chalcone derivatives [Table-1] to investigate the biding mode of these compounds and properties that are relevant for their activity. We use the approach of Hansch [28] in classical QSAR analysis for obtain linear model by the Partial Least Squares (PLS) method to predict the experimental activity.

Chemical Data

Biological data on the activity of ketone derivatives was obtained from the paper Bag and co-wokers [23] [Table-1]. The activity data refers pMIC, which indicates the biological activity of compounds experimentally determined, necessary for the inhibition of *Candida albicans* resistant (NCIM 3446). The – log MIC (molar) scale refers pMIC. Fluconazole was used as controls in the assays.

Geometry Optimization

The Core-Seven personal computer equipped with the operating system Windows® Seven was used for making calculations of this work. The molecular structures of the dataset were sketched using ACD/ChemSketch (2010) version 12.0 developed by Advanced Chemistry Development.

The first step consisted in obtaining the molecular geometry of all derivatives from the dataset [Table-1]. We initially performed geometry optimization, which was done using semi-empirical PM3 (Parametric Method) Hamiltonian method in the Arguslab program (2004) version 4.0 developed by Mark Thompson and Planaria Software LLC.

After this first procedure, the stability of the molecular geometry was obtained by Density Functional Theory method (B3LP/6-31G), using ChemSitePro program (2009) version 9.0 developed by Chem SW Lab Software.

Further, we minimized molecular energy with Simplex Method using dielectric constant equals to 46.7 (simulate an environment with dimethyl sulfoxide), Lennard-Jones potential 6-12 and Hydrogen bond function in the Molecular Modeling Pro Plus 6.3 (MMPP) package (2004) version 6.3 developed by Chem SW Lab Software.

Finally, was performed the conformational analysis of all compounds in the MMPP computational package. Conformational analysis was performed for each structure using simultaneous spinning in 10º of two single bonds (C2-CB and C4-CA) [Table-1] by fixing Root Mean Square Gradiente (RMS) to 0.1 Kcal/molÅ and Lennard-Jones potential 6-12. The low energy conformers obtained were used to quantum chemical QSAR analysis.

Structural Descriptors

In the quantum chemical analysis we calculated 94 properties of all compounds. The calculated physical-

chemical parameters types are: hydrophobic, electronic, steric, thermodynamic, dimensional, topological and geometric. All the molecular properties were calculated by ChemSitePro program and MMPP computational package.

The electronic parameters were subdivided into properties of empirical and quantum nature. For empirical electronic parameters, were used MMPP computational package, where the parameters quantum electronic were used ChemSitePro program.

The Molecular Electrostatic Potential Maps, HOMO frontier orbital energy and LUMO frontier orbital energy were calculated by Density Functional Method (DFT-B3LP/6-31G). All calculated parameters are shown in [Table-2].

QSAR model selection

In order to estimate multidimensional linear Hansch model [28], we based on the Topliss criterion [29], in adjustment degree (R^2 and SEC), in degree of statistical significance (F-values) and the previsibility degree (Q^2_{LOO}, Q^2_{LNO}, e Q^2_{pred}). Initially, was executed the linear regression, one by one, for ninety-four molecular descriptors and the biological activity. The values of the Pearson correlation coefficient (R), coefficient of multiple determination (R^2), standard deviation (SEC) and Fischer's test (F) were evaluated.

The BuildQSAR program (2009) version 2.1 developed by Professor Anderson Coser Guadio of Federal University of Espirito Santo (UFES), was used to perform linear regression and molecular descriptors that showed Pearson correlation coefficient less than 0.25 were excluded from the QSAR analysis. Thus, the number of variables was reduced from 94 to 54 with this procedure. In the second step, PLS method was used for variable selection. The descriptors were autoscaled (pre-processing) by rearranging the columns of the data matrix in such a way that the most important descriptors, classified according to an informative vector, are placed in the first columns. Then, successive PLS regressions are performed with increasing number of descriptors in order to find the best PLS model. The PLS regressions were performed in the Molegro Data Modeller program (2010) version 2.5 developed by Molegro Computational Drug Discovery.

At the third step, the set of 14 selected descriptors was further refined using The Unscrambler program (2005) version 7.6 developed by CAMO Software AS, with removal of more descriptors, to obtain an optimized model which would fulfill the criteria for being statistically significant, robust and interpretative.

Model validation

The final model was thoroughly validated using a set of procedures suggested in the literature [30]. The statistical parameters listed in [Table-3] [31] were used to evaluate the quality of the model. For the internal quality, the recommended limits are $R^2 > 0.6$ and $Q^2_{LOO} > 0.5$ [27, 32]. The SEC and SEV should be as lower as possible. The F-test value should be higher than the tabled critical-F ($F_{p,n-p-1}$), where n is the number of compounds and p is the number of latent variables in the final model and the higher the difference between them, the more statistically significant is the model [33].

The robustness of the optimized model was examined by leave-*N*-out cross-validation (LNO, N = 1, 2,….10) procedure. This test was repeated three times for each value of "N", with a randomization of all rows from the data matrix and respective y-values in each step of LNO process. The average value of each Q^2_{LNO} is expected to be close to Q^2_{LOO} (coefficient of multiple determination of leave-one-out cross-validation) with standard deviations close to zero [32].

The possibility of chance correlation was tested using Y-randomization analysis [27], where the Y vector was scrambled 10 times [34]. The approach suggested by Eriksson and co-workers [35], based on the absolute value of the Pearson correlation coefficient between the original vector Y and the randomized vectors Y was used to quantify chance correlation. In this approach, we constructed the diagram Q^2LOO (Y-axis) vs R^2 (X-axis). The intercepts of the equations obtained in the linear regression should be less than 0.3 for R^2 and 0.05 for Q^2_{LOO}.

After internal evaluations, a set for external validation (test set), having a representative pMIC range as well as structural variations and a new model was built. The statistical quality of the new model cannot be much different from the model generated with all compounds (training set).

The parameter R^2_{pred} was used as a measure of the predictive power of a QSAR model. For this work, it was used the recommended limit of $R^2_{pred} > 0.5$ [36-37]. However, this is not a sufficient condition to guarantee that the model is really predictive. It is also recommended to check: 1) the slopes K or K' of the linear regression lines between the observed activity and the predicted activity in the external validation, where the slopes should be $0.85 \leq K$ or $K' \leq 1.15$; and 2) the absolute value of the difference between the coefficients of multiple determination, R^2_0 and R'^2_0 smaller than 0.3 [27, 38].

Results and Discussion

After we accomplish geometry optimization, calculation of structural descriptors, QSAR model selection and model validation performing quantum chemistry QSAR analysis and interpretation of the selected descriptors in order to predict the biological activity of other compounds derived from chalcone.

Quantum Chemistry QSAR analysis

The QSAR analysis included the 20 compounds presented in [Table-1], which let us to investigate models with up to four variables. The selected model obtained by PLS methodology presented four descriptors. The PLS model [Eq-1] obtained with three latent variables describing 96.14% (LV1 = 54.34%; LV2 = 26.88% and LV3 = 14.91%) from the original information. The selected

descriptors were the molar refractivity [RM], the ionization potential [PI], the molecular length [Lx] and the Verloop B4 [B4 (A)] to the substituent in ring A [Table-4]. These properties were capable of elucidating 85% and predicting 78% of total variance. With only the first latent variable (PC1: LV1), the model is able to explain approximately 54% of the variance of the original data and 50% of the biological activity.

pMIC = + 0.3393 [Lx] – 1.7829 [PI] – 0.0708 [MR] + 0.8918 [B4 (A)] + 17.4735 [Eq-1]

n = 20; LVs = 3; Cumulated information = 96.14% (LV1 = 54.34%; LV2 = 26.88% and LV3 = 14.91%); R^2 = 0.776; SEC = 0.229; $F_{(3,16)}$ = 14.172 (Fc = 3.24) ; Q^2_{LOO} = 0.609; SEV = 0.295.

The [Fig-1] shows the observed activity versus the predicted activity for training set. The regression model has small residuals that can be seen in [Table-4].

LOO cross-validation analysis revealed that $R^2 - R^2_{LOO} < 0.3$ (0.776 – 0.625 = 0.151) and $Q^2_{LOO} > 0.5$ (0.609) according to literature [27,38].

LNO cross-validation and Y-randomization analysis results are shown in [Fig-2] and [Fig-3], respectively. The model presented high average Q^2_{LNO} (0.612), small fluctuations of the standard deviations for each LNO point and small variations related to the Q^2_{LNO} value. LNO cross-validation employs smaller training sets than the LOO procedure and can be repeated several times due to the large number of combinations when leaving many compounds out from the training set once at a time. A QSAR model can be considered robust when its average Q^2_{LNO} values are relatively high and close to the value of Q^2_{LOO} [38].

The Y-randomization test is useful to verify the possibility that the explained and predicted variances by the obtained model may suffer from chance correlation [27]. It can be observed that the results obtained for all randomized models are of bad quality when compared to the real model, and the intercepts [Fig-3] are inside the acceptable values recommended in the literature, ie., the intercepts are below the limits ($R^2 \leq 0.3$ and $Q^2 \leq 0.05$). Dispersion of data points is observed in the regions around the intercepts, what is reasonable situation for smaller data sets.

Analyzing the diagram Scores [Fig-4] we noted that the compounds 18, 17 and 11 (inactive) are grouped. These compounds are those with the lower biological activity of the training set. For a better interpretation, the analysis diagram Loadings [Fig-5] shows that the molar refractivity contributes to these derivatives, in other words, these compounds have higher molar refractivity; this can be seen in [Table-4]. The plot of Scores demonstrates a quite good discrimination between highly and weakly active compounds in accordance for the significant statistical quality of the PLS model obtained.

Many autors argue that only externally validated models, after the internal validation, may be considered realistic and applicable for drug design [27, 38]. Studies such as those reported by Golbraik and Tropsha [32], and Aptula and cowokers [39], support this assumption. Data obtained from the paper of Batovska (A) and cowokers [40] and Turkar (B) and cowokers [41] were used for external validation. The [Fig-6] shows the chemical structures (nine) of compounds selected for external validation (external validation set). The results in [Table-5], has shown that the selected model presented high external predict-ability, considering the proposed limits. A of the values K or K' and the relation $|R^2_0 - R'^2_0|$ are inside the acceptable ranges ($0.85 \leq K$ or $K' \leq 1.15$; and $|R^2_0 - R'^2_0| \leq 0.3$) [27, 38]. The SEP values are also considered low, what is an indicative of low error for a synthesized compound based on this model.

Predictive ability of QSAR model is demonstrated through external validation, where the statistical parameters following show that the QSAR model has good predictive ability:

- $R^2 = 0.776 > 0.6$
- $R^2_{pred} = 0.709 > 0.5$
- $Q^2_{pred} = 0.709 > Q^2_{LOO} = 0.609$
- K = 0.7098 and K' = 1.000 where $0.85 \leq K$ or $K' \leq 1.15$ and
- $|R^2_0 - R'^2_0| = 0.0009 < 0.30$

$$\frac{(R^2 - R^2_0)}{R^2} = 0.086 < 0.1 \quad \text{or} \quad \frac{(R^2 - R'^2_0)}{R^2} = 0.085 < 0.1$$

Thus, the results of validations steps show that the model can be classified as a good model, since, according to the criteria used, it has good internal quality, it is robust, it does not suffer from chance correlation at random, and it shows a good capacity to external predictions. The [Fig-7] shows the regression between observed values vs. predicted values and predicted values vs. observed values for compounds external set.

Interpretation of the selected descriptors

In PLS analysis, the descriptors data matrix is decomposed to orthogonal matrices with an inner relationship between the dependent and independent variables. Therefore, unlike regression linear multiple (MLR) analysis, the multicolinearity problem in the descriptors is omitted by PLS analysis. Because a minimal number of latent variables are used for modeling in PLS, this modeling method coincides with noisy data better than MLR.

The model in [Eq-1] indicates that inhibitory activity of compounds against strain of *Candida albicans* depends on thermodynamic, dimensional and steric parameters. The [Fig-8] shows the regression coefficients for the descriptors and [Fig-9] shows the descriptor relevance for the variables used in the model proposed.

The negative sign of mixed steric-hydrophobic parameter, [MR], indicates that molecule of compound should have smaller molecular volume to favor the biological activity. The compounds 18, 17 and 11 have phenyl group as substituent, increasing considerably the molecular volume of their chemical structures which contributes to lower biological activity. There is great possibility of the derivatives studied interact with bioreceptor of the fungal. Thus, the active region of the receptor must have an active site where larger molecules have difficulty in

complexity, not favoring the intermolecular interactions ligand-bioreceptor.

The negative sign of the molar refractivity probably shows that the ligand has the ability to distort the conformation of the active site of receptor only through steric hindrance, ie, the molar refractivity is contributing as a steric descriptor of the substituent [42]. An important detail is the fact that the phenyl group is in position-para on substituent is in ring A or B, bulky substituents in this position in ring A or B reduced biological activity, therefore, has difficulty connecting ligand-bioreceptor.

Remembering the Lorentz- Lorenz equation [43], do not forget that the molar refractivity is a measure of both the molecular volume of a compound and how easily it is polarized, ie, a molecular property of character mixed.

Substituents on ring A or B with higher polarity thus have higher hydrophilicity and present less biological activity. In general, the groove that receives the substituents on ring B has some hydrophobicity. This can be analyzed by observing the biological activity of compounds 13, 01, 02 and 03.

The negative sign of thermodynamic parameter, [PI], indicates that molecular structure need lower-cost energy to remove electrons from the molecule to favor the biological activity. Thus, if the molecule requires less energy to remove electrons, the greater is HOMO frontier orbital energy. The ionization energy is related the HOMO frontier orbital energy. Chalcones possess HOMO molecular orbital in the region of carbonyl, favoring the electron-donor ability, as well as region has nucleophilic character, functioning as a Lewis base. The ligand-bioreceptor interaction may be occurring through weak interactions and that requires less energy in electron transfer. Thus, the interaction is not electrostatic type, with the option for covalent interactions through of mechanism of molecular polarization. It is important to note that there is olefinic unsaturation in carbons C3 and C4, which explains the mechanism of resonance through the relocation the π bond due to the presence of carbonyl. Chalcones are enones and predisposed to the occurrence of reaction of Michael addition. From the point view drug-receptor, the molecule can to result in covalent-binding of type irreversible inhibition on C4 carbon.

The [Fig-10] shows the Molecular Electrostatic Potential Maps (MEP) of the most active derivatives (01 and 13: training-serie; 09: external-serie). In the three compounds we see the negative region of the carbonyl, justifying the high electron density in the region. The figure also shows that MEP is similar for the three ketone derivatives. Whereas ionization potential [PI] = - HOMO frontier orbital energy, the lower [PI] greater HOMO orbital energy. The [Fig-11] shows that the HOMO orbital is located in the region of carbonyl compounds. The [Fig-12] shows that the LUMO orbital is located in the electrophilic centers of chacones.

The positive sign of dimensional parameter, [Lx], indicates that molecular structure should have larger molecular length to favor the activity. The compound 11 has a large molecular extension due to the presence of phenyl group. This also can be evidenced through the diagram of Scores and Loadings. Although of the molecule 11 have greater [Lx], it has high molecular refractivity, which explains its small biological activity.

Therefore, we see that the descriptor molar refractivity [MR] has a higher contribution (weight) that the descriptor molecular length [Lx].

An interesting possibility is the substituents of molecule are capable of supplying electron pair and at the same time has a greater molecular length and smaller molecular volume. Would welcome and encourage the presence of substituents quinolinyl in ring A, which would increase the Lx without significantly to increase the molecular volume, as this group is planar. Chloro-quinolinyl substituent would also be contributing with capacity to donate electrons. It is strange that the group of López *et al*. [21] find that the quinolinyl chalcones do not show good biological activity. This may be related with the type of assay that researchers have done. Depending on the type of test performed, some interference may to mask the biological activity of the compounds. It would be interesting proposal for a synthesis of compounds containing derivatives with quinolinyl e chloro-quinolinyl substituents in order to verify experimentally the biological activity of antifungal agents.

The positive sign of steric parameter, [B4 (A)], indicates that molecular structure should have larger radii in substituent in ring A to favor the activity. This also favors the hypothesis of the presence of quinolinyl and chloro-quinolinyl substituents in ring A. We note that compounds 17 and 18 are planar, and then have a lower [B4 (A)], which explains the low biological activity of these compounds. Tiometyl substituents in ring A has a higher [B4 (A)] compared with methoxy substituents, and then have a higher biological activity, which explains the activity of the compound 13. Analyzing the structure of the compound 13, we note that molecular volume should be less in ring B to favor biological activity.

We realize that the weight of the molar refractivity is much higher than the other descriptors which are in the [Equation 1]. This can also be shown in the diagram [Fig-9].

An interesting proposal for the mechanism of action of chalcones

Most authors report that chalcones exhibit activity against the β (1,3)-glucan synthase and chitin synthase enzymes of the cell wall of *C. albicans*. It is known that chalcones have potential activity against cysteine and aspartic proteases from various microorganisms. These compounds may be involved in the inhibition of another target, for example, glutathione *S*-transferase (GST) [44] as well as aspartic proteinases (Saps) [45-47].

C. albicans possesses a potent armamentarium consisting of several virulence molecules that help the fungal cells to escape of the host immune responses. There is no doubt that the secretion of aspartyl-type proteases, designated as Saps, are one of the major virulence attributes produced by *C. albicans* cells, since

these hydrolytic enzymes participate in a wide range of fungal physiological processes as well as in different facets of the fungal-host interactions. For these reasons, Saps clearly hold promise as new potential drug targets [45]. In this context, the crystal structure of Sap2 complexed with pepstatin A has been known since 1993 [46], whereas the crystal structure of Sap3 and its complex with pepstatin A was first presented in 2007 [47]. Future studies, eg., docking ligand-proteinases in *C. albicans* can complement QSAR analysis and elucidate mechanisms involved in interactions that are possibly occurring with fungal enzyme.

The glutathione *S*-transferase (GST) is one of the important enzymes in xenobiotc metabolism [48-49]. There are reports in the literature concerning the GST inhibition *in vitro* by chalcones. Miyamoto and cowoker [44] reported the inhibition of glutathione *S*-transferase by chloro-substituted 4´-phenylchalcones. Through ultraviolet spectroscopy determined the extent of conjugation between chalcones and GSH (Glutathione) resulting conjugates glutathione. Many of the reactions involving the GSH sulfhydryl group (SH), highly polarizable, make it a good nucleophile for reactions with compounds that have electrophilic center. This ability to donate electrons to other compounds makes GSH is a good reducer [48].

We know that chalcones may occur addition of nucleophile agent at C2 or C4. Adding carbon C2 is unlikely, because in this case should occur the presence of hard nucleophile, with a negative charge concentrated in a small and highly electronegative atom, and is not to be the case, because the sulfur atom of the GSH sulfhydryl group does not fit in with this condition. The nucleophilic addition at C4 carbon is the most likely to occur. In this case there would be the establishment of a covalent bond through the interaction between the HOMO orbital of the sulfur atom of the SH group of glutathione with the LUMO orbital of the atom C4 carbon of the chalcone. The nucleophilic addition at C4 occurs with soft nucleophiles, ie, small and very electronegative atoms. The addition of carbon C4 results in the formation of glutathione-chalcone conjugate. The mechanism of chalcone resonance results in a chemical structure capable of accepting the agent nucleophile this carbon atom.

The glutathione in *Candida albicans* has many functions including the prevent of oxidative stress the cell through complexation with reactive oxygen species (ROS). The mechanisms of damage to cellular targets of ROS species can occur damage to proteins by direct attack of reactive species or by secondary damage involving attack by products of lipid peroxidation. Fatty acids and membrane lipoproteins are potential targets of ROS. The ROS can interact with the nitrogenous guanine base preventing DNA replication and cell division. Currently, it is understood that oxidative stress in yeast cells represents an important line of elimination of pathogenic microorganisms [50]. It is important to remember that the GST enzyme catalyzes the glutathione (GSH) with electrophiles centers resulting in the glutathione-chalcone conjugate. Thus, the presence of compounds with electrophiles centers will decrease the concentration of intracellular glutathione in *Candida albicans*, resulting in fungal cell an oxidative stress. Thus, the complexation of glutathione (GSH) and chalcone resulting conjugate glutathione is an interesting proposal for the mechanism of action of chalcones [Fig-13]. The [Fig-14, 15] shows the proposed mechanism resulting in the glutathione-chalcone conjugate.

Proposal for organic synthesis of analogous derivatives

Taking into account the multidimensional model proposed, the selected properties and the proposed of one possible mechanism of action for the compounds studied, we propose for future research, the synthesis of organic compounds sketched in [Fig-16].

Conclusion

In this study was possible to obtain a multivariate QSAR model for a set of ketones that have the capability of inhibiting *in vitro* strain of *Candida albicans*. The LOO and LNO cross-validation methods, the Y-randomization technique, and the external validation indicated that the model is significant, robust and has good internal and external predictability. The inhibitory activity of the investigated compounds was described based in descriptors: [MR], [PI], [Lx] and [B4 (A)]. Termodynamic, dimensional, steric properties play a significant role in explaining the activity of the dataset. The results indicated that the activity against strains of *Candida albicans* is favored by smaller molecular volume, higher molecular length, the electron donating capacity, increased radius of the substituents in ring A of compounds, and electrophiles centers. The mechanism of action is related with dimensional and electronics aspects of the compounds, which can be explained by the descriptors that were selected in QSAR model proposed. The study indicates that the presence of quinolinyl and chloro-quinolinyl substituents in compounds on ring A would be contributing for biological activity. It´s important the synthesis of chalcones with these substituents for verify the authenticity of the facts. The complexation of glutathione (GSH) and chalcone resulting conjugate glutathione-chalcone is an interesting proposal for the mechanism of action of chalcones. The proposed model may provide a better understanding of the anti-*Candida albicans* activity of chalcones and can be used as guidance for proposition of new chemopreventive agents.

Acknowledgements

Prof. Luiz F. Motta thanks the Federal Institute of Triangulo Mineiro (IFTM) (http://www.iftm.edu.br) for support to the author's doctoral thesis and Institute of Chemistry of State University of Campinas (UNICAMP) (http://www.iqm.unicamp.br) which made possible the development of this research. The authors thanks, Prof. Marcia M. C. Ferreira of State University of Campinas (UNICAMP) and Prof. Eduardo B. M. of State University of West Paraná (UNIOESTE-PR) for grateful discussions and suggestions. The authors also thank Prof. Anderson C. Guadio of Federal University of Espirito Santo (UFES) for providing the BuildQSAR program version 2.1. The

authors also thank CNPq for W. P . Almeida Research´s Fellowship.

.

References

[1] Romero M., Cantón E., Pemán J. and Gobernado M. (2005) *Revista Española de Quimioterapia,* 18 (4), 281-289.

[2] Murray P.R., Rosenthal K.S. and Pfaller M.A. (2006) *Microbiologia Médica, Elsevier: Rio de Janeiro, Brazil*, 693-702.

[3] Golan D.E., Tashjian A.H.Jr; Armstrong E.J. and Armstrong A.W. (2009) *Princípios de Farmacologia: A Base Farmacológica da Farmacoterapia, Guanabara Koogan: Rio de Janeiro, Brazil*, 579-589.

[4] Zaitz C., Campbell I., Marques S.A., Ruiz L.R.B. and Framil V.M.S. (2010) *Compêndio de Micologia Médica, Guanabara Koogan: Rio de Janeiro, Brazil*, 89-107.

[5] Vidotto V. (2004) *Manual de Micologia Médica, Tecmedd: Ribeirão Preto, Brazil*, 4-37.

[6] Trabulsi L.R. and Alterthum F. (2005) *Microbiologia, Atheneu: São Paulo, Brazil*, 461-470.

[7] Katzung B.G. (2010) *Farmacologia Básica e Clínica, McGraw Hill and Artmed: São Paulo, Brazil,* 707-714.

[8] Shepherd M.G. Cell (1987) *Critical Reviews in Microbiology*, 15 (1), 7-25.

[9] Petrikkos G. and Skiada A. (2007) *Int. J. Antimicrob. Agents*, 30, 108-117.

[10] Andriole V.T. (1999) *J. Antimicrob. Chemother*., 44, 151-162.

[11] Santo R.D., Tafi A., Costi R., Botta M., Artico M., Corelli F., Forte M., Caporuscio F., Angiolella L. and Palamara A.T. (2005) *J. Med. Chem*., 48, 5140-5153.

[12] Mota S.G.R., Barros T.F. and Castilho M.S. (2009) *J. Braz. Chem. Soc*., 20 (3), 451-459.

[13] Georgopapadakou N.H. and Walsh T.J. (1996) *Antimicrob. Agents Chemother*. 40 (2), 279-291.

[14] Hector R.F. (1993) *Clin. Microbiol. Rev*., 6, 1-21.

[15] Nowakowska Z., Kedzia B. and Schroeder G. (2008) *Eur. J. Med. Chem*., 43, 707-713.

[16] Sabet R., Fassihi A. and Moeinifard B. (2007) *Res. Pharm. Sci.,* 2 (2), 103-112.

[17] Nowakowska Z. (2007) *Eur. J. Med. Chem*., 42, 125-137.

[18] Sato M., Tsuchiya H., Akagiri M., Fujiwarat S., Fujii T. and Tagagi N. (1994) *Lett. Appl. Microbiol*., 18, 53-55.

[19] Tomar V., Bhattacharjee G. and Kamaluddina K. (2007) Bioorg. Med. Chem. Lett., 17, 5321-5324.

[20] Lahtchev K.L., Batovska D.I., Parushev St.P., Ubiyvovk V.M. and Sibirny A.A. (2008) *Eur. J. Med. Chem*., 43, 2220-2228.

[21] López S.N., Castelli M.V., Zacchino S.A., Domínguez J.N., Lobo G., Charris J.C., Cortés J.C.G., Ribas J.C., Devia C., Rodríguez A.M. and Enriz R.D. (2001) *Bioorg. Med. Chem*., 9, 1999-2013.

[22] Cid V.J., Durán A., Del Rey F., Snyder M.P., Nombela C. and Sánchez (1995*) Microbiol. Rev*., 59, 345-386.

[23] Bag S., Ramar S. and Degani (2009) *Med. Chem. Res*., 18, 309-316.

[24] Tong W., Hong H., Xie Q., Shi L., Fang H. and Perkins R. (2005) *Current Computer - Aided Drug Design*, 1, 195-205.

[25] He L. and Jurs P.C. (2005) *J. Mol. Graph. Model*., 23, 503-523.

[26] Ghafourian T. and Cronin M.T.D. (2005) *SAR QSAR Environ. Res*., 16, 171-190.

[27] Tropsha A., Gramatica P. and Gombar V.K. (2003) *QSAR Comb. Sci*., 22, 69-77.

[28] Unger S.H and Hansch (1973) *J. Med. Chem*., 16, 745-749.

[29] Topliss J.G. and Costello R.J. (1972) *J. Med. Chem*., 15, 1066-1068.

[30] Kiralj R. and Ferreira M.M.C. (2009) *J. Braz. Chem. Soc*., 20, 770-787.

[31] Melo E.B., Martins J.P. A., Marinho Jorge T.C., Friozi M.C. and Ferreira, M.M.C. (2010) *Eur. J. Med. Chem*., 45, 4562-4569.

[32] Golbraikh A. and Tropsha A. (2002 *J. Mol. Graph. Model.,* 20, 269–276.

[33] Gaudio A.C. and Zandonade E. (2001) *Quim. Nova*, 24, 658-671.

[34] Wold S. and Eriksson L. (1995) *In Chemometric Methods in Molecular Design, van de Waterbeemd H., VCH: Weinheim, Germany*, 309–318.

[35] Eriksson L., Jaworska J., Worth A.P., Cronin M.T.D., McDowell R.M. and Gramatica P. (2003) *Environ. Health Perspect*., 111, 1361–1375.

[36] Roy P.P., Leonard J.T. and Roy K. (2008) *Chemometr. Intell. Lab*., 90, 31-42.

[37] Roy P.P. and Roy K. On (2008) *QSAR Comb. Sci*., 27, 302-313.

[38] Melagraki G., Afantitis A., Sarimveis H., Koutentis P.A., Markopolus J. and Igglessi-Markopoulou O. (2007) *J. Comput. Aided Mol. Des*., 21, 251-267.

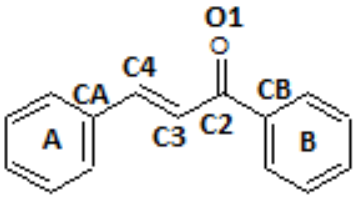
[39] Aptula A.O., Jeliazkova N.G., Schultz T.W. and Cronin M.T.D. (2005) *QSAR Comb. Sci*., 24, 385–396.

[40] Batovska D., Parushev St., Slavova A., Bankova V., Tsvetkova I., Ninova M. and Nadjenski H. (2007) *Eur. J. Med. Chem*., 42, 87-92.

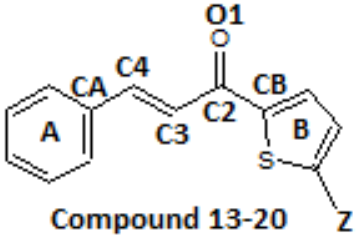
[41] Turkar S.S., Rodge A.H., Hatnapure G.D., Keche A.P. and Gaikwad G.S. (2010) *J. Chem. Pharm. Res*.,, 2 (5), 348-355.

[42] Montanari M.L.C., Montanari C.A and Gaudio A.C. (2002) *Quím. Nova*, 25 (2), 231-240.
[43] Tavares L.C. (2004) *Quím. Nova*, 27 (4), 631-639.
[44] Miyamoto T. and Yamamoto I. (1994) J. *Pestic. Sci.,* 19, 53-58.
[45] Naglik J.R., Challacombe S.J. and Hube B. (2003) *Microbiol. Mol. Biol. Rev*., 67 (3), 400-428.
[46] Cutfield S.M., Dodson E.J., Anderson B.F., Moody P.C.E., Marshall C.J., Sullivan P.A. and Cutfield J.F. (1995) *Structure*, 11, 1261-1271.
[47] Borelli C., Ruge E., Schaller M.M.M., Korting H.C., Huber R. and Maskos K. (2007) *Proteins,* 68, 738-748.
[48] Huber P.C., Almeida W.P. and de Fátima A. (2008) *Quím. Nova*, 31 (5), 1170-1179.
[49] Seenan D., Meade G., Foley V.M. and Dowd (2001) *C.A. Biochem. J*., 360, 1-16.
[50] Moye-Rowley, W.S. (2003) *Eucariotic Cell*, 2(3), 381-389.

Table 1 - Data set (training set) from the literature [23] used in the Quantum Chemical QSAR analysis;



Compound 1-12
Type X



Compound 13-20
Type Y

Compound	Type	Ring A	Ring B	Substituent Z	pMIC
01	X	4-SCH_3	4-F	-	4.531
02	X	4-SCH_3	4-Cl	-	4.159
03	X	4-SCH_3	4-Br	-	3.522
04	X	4-SCH_3	2,4-Cl	-	3.208
05	X	4-SCH_3	4-NO_2	-	3.477
06	X	4-SCH_3	4-OCH_3	-	4.152
07	X	4-SCH_3	H	-	3.804
08	X	4-SCH_3	4-OH	-	3.829
09	X	4-SCH_3	2-OH	-	4.130
10	X	4-SCH_3	3-OH	-	3.829
11	X	4-SCH_3	4-phenyl	-	3.120
12	X	2,3-OCH_3	4-OCH_3	-	4.474
13	Y	4-SCH_3	-	H	4.716
14	Y	4-SCH_3	-	Br	3.832
15	Y	3,4-OCH_3	-	H	4.136
16	Y	3,4-OCH_3	-	Br	3.548
17	Y	4-phenyl	-	H	3.064
18	Y	4-phenyl	-	Br	3.169
19	Y	4-OCH_3	-	H	4.086
20	Y	4-OCH_3	-	Br	3.508

*pMIC = - log MIC; * MIC = Minimum inhibitory concentration in molar.*

Table 2 - Molecular properties calculated

NATURE OF THE PARAMETHER	MOLECULAR PROPERTIES
HYDROFOBIC (24)	Log P Hansch; Log P Ghose; Log P Moriguchi; MR Ghose; Q Log P; V.M. Q Log P; HLB Volumetric; P.S. Hansen 3D; Dispersion Hansen 3D; Polarity Hansen 3D; Hydrogen Bond Hansen 3D; P.S. Krevelen 3D; Dispersion Krevelen 3D; Polarity Krevelen 3D; Hydrogen Bond Krevelen 3D; V.M. Krevelen 3D; Energy of Cohesion; Hydrophilic Surface Area; % Hydrophilic Surface Area; Surface Tension; S.W. Klopman; Log S.W. Klopman; Log Molar S.W. Hansch; Log Molar S.W. Ghose.
DIMENSIONAL (08)	van der Waals Volume; Surface Área; Density; Molar Volume; Molecular Length; Molecular Wide; Molecular Depth; Number of Atomic Centers.
TOPOLOGIC (06)	Wiener Index 3D; Balaban Index Q; Balaban Index S; Balaban Index D; Balaban Index A; Balaban Index P.
QUANTUM ELETRONICS (29)	Heat of Formation; Total Energy; E_{HOMO}; E_{HOMO-1}; E_{HOMO-2}; E_{LUMO}; E_{LUMO+1}; E_{LUMO+2}; q (O1); q (C2); q (C3); q (C4); q (CA); q (CM); q (CP); q (CB); Dipole Moment; Hardness; Softness, Mulliken Electronegativity; Gap; Hydrogen Bond Donor; Hydrogen Bond Acceptor; DE (O1); DE (C2); DE (C3); DE (C4); DE (CA); DE (CB).
EMPIRICAL ELETRONICS (04)	Hamett σ* (A); Hamett σ-para (A); Hamett σ-meta (A); Hamett σ-induction (A).
STERIC (11)	Molar refractivity#; Verloop L1 (A); Verloop B1 (A); Verloop B2 (A); Verloop B3 (A); Verloop B4 (A); Verloop L1 (B); Verloop B1 (B); Verloop B2 (B); Verloop B3 (B); Verloop B4 (B).
THERMODYNAMIC (04)	Gibbs Free Energy; Ionization Potential; Parachor; H.B.N.
GEOMETRIC (08)	(O1C2CB; C2C3C4; C3C4CA; d (C2CB); d (O1C2); d (C2C3); d (C3C4); d (C4CA).

P: Partition Coefficient; MR: Molar Refractivity; V.M: Molecular Volume; HLB: Hydrophilic-lipophilic Balance; P.S.: Solubility Parameter; S.W.: Water Solubility; q: Mulliken Partial Charges; E_{HOMO} and E_{LUMO}: Frontier Orbital Energy; Gap: $E_{HOMO} – E_{LUMO}$; σ: Hammett Substituent Constant; CM: Carbon-meta in ring A; CP: Carbon-para in ring A; DE: Electron density; H.B.N.: Hydrogen Bond Number; d: Interatomic distances; # Some authors consider the molar refractivity as a parameter with mixed hydrophobic and steric contributions.

Table 3 - Statistics parameters analyzed

Parameter	Symbol
Coefficient of multiple determination of calibration	R^2
Standard deviation of calibration model	SEC
F-test (with 95%) confidence interval	F
Coefficient of multiple determination of cross-validation – " leave-one-out (LOO)"	Q^2_{LOO}
Standard error of cross-validation	SEV
Coefficient of multiple determination of cross-validation – " leave-*N*-out (LNO)"	Q^2_{LNO}
Coefficient of multiple determination of prediction – external set	Q^2_{pred}
Standard error of prediction	SEP
Slopes of the linear regression lines	K and K'

Table 4 - Values of descriptors used for the formulation of model and LOO cross-validation results.

Compound	[MR]	[PI]	[Lx]	[B4 (A)]	pMIC $_{obs}$	pMIC $_{pred}$	Residues
1	77.983	8.66002	17.4220	2.5067	4.531	4.722	-0.191
2	83.000	8.64373	16.5513	2.58952	4.159	4.104	0.055
3	85.898	8.65428	16.8184	2.42155	3.522	3.846	-0.324
4	87.867	8.65601	16.2988	2.20084	3.208	3.347	-0.139
5	87.761	8.79137	17.1150	2.51513	3.477	3.664	-0.187
6	84.394	8.58195	17.3553	2.2395	4.152	4.072	0.080
7	78.133	8.60710	15.9419	2.12793	3.804	3.973	-0.169
8	79.658	8.60860	15.8854	2.39803	3.829	4.029	-0.200
9	79.658	8.68011	15.9428	2.24248	4.130	3.730	0.400
10	79.658	8.62572	16.6768	2.20566	3.829	4.116	-0.287
11	102.24	8.59834	19.0341	2.17772	3.120	3.431	-0.311
12	84.328	8.78265	17.8516	2.66742	4.474	4.074	0.400
13	77.334	8.60593	15.9786	2.27782	4.716	4.012	0.704
14	85.099	8.66701	16.9147	2.26500	3.832	3.752	0.080
15	77.268	8.79500	15.8240	2.68451	4.136	4.077	0.059
16	85.033	8.83799	16.2757	2.68364	3.548	3.700	-0.152
17	88.853	8.87031	18.0011	1.78522	3.064	3.084	-0.020
18	96.618	8.94672	19.3098	1.84424	3.169	2.621	0.548
19	71.007	8.97866	15.7680	2.57941	4.086	4.114	-0.028
20	78.772	9.06768	17.0855	2.50322	3.508	3.887	-0.379

Table 5- Predicted values of the test set (external cross-validation) and results of statistical parameters.

Compound	pMIC$_{Obs}$	pMIC$_{pred}$	Residues
01	3,221	3,336	-0.139
02	3,555	3,797	-0.242
03	3,856	3,545	0.311
04	3,406	3,232	0.174
05	3,316	3,811	-0.495
06	4,555	4,737	-0.182
07	3,96	4,165	-0.205
08	4,649	4,372	0.277
09	4,732	4,251	0.481
R^2_{pred}		0.709	
SEP		0.091	
K		0.709	
K'		1.00	
$\|R^2_0 - R'^2_0\|$		$\|0.709 - 0.7099\| = 0.0009$	
$\|R^2 - R^2_0\| / R^2$		0.776 – 0.709 = 0.067/ 0.776 = 0.086	

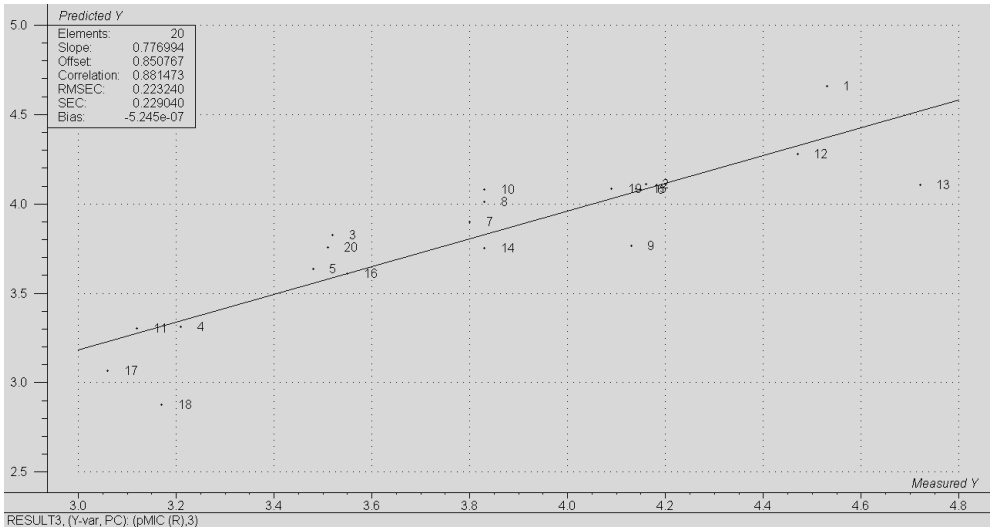


Fig. 1 - Shows the observed activity versus the predicted activity for training set. The regression model has small residuals that can be seen in [Table-4]

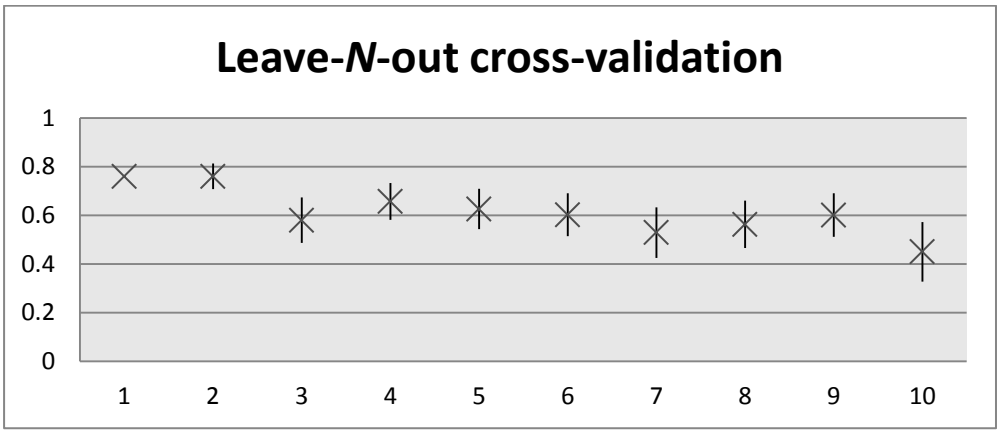


Fig. 2 - Plot of LNO cross-validation; In the LNO plot, each point refers to the average value from test in triplicate, and the columns refer to standard deviation.

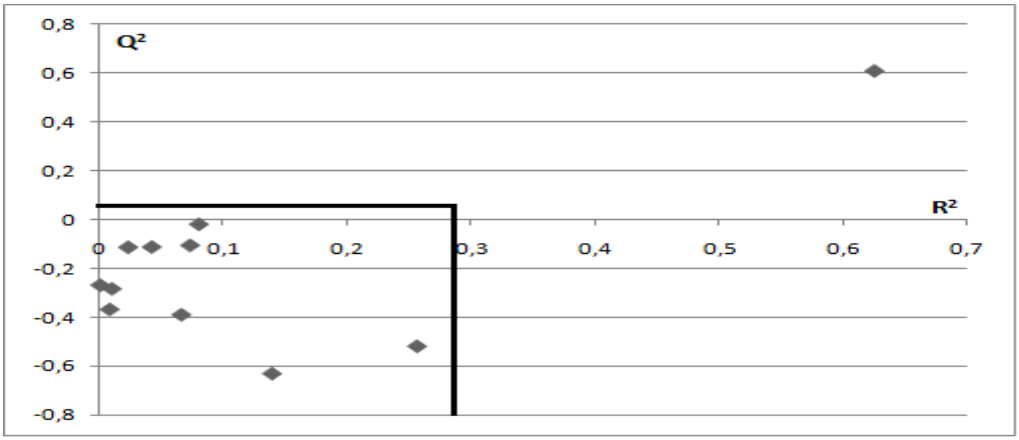


Fig. 3 - Plot of Y-randomization test; Each point in the region between intercepts indicates randomized models: All obtained values for R^2 and Q^2 test are below 0.256 and - 0.018, respectively.

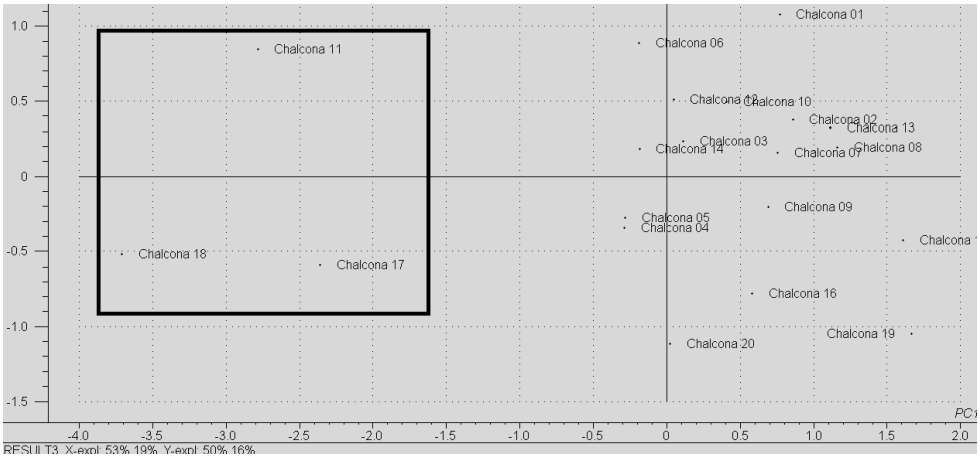


Fig. 4 - Plot of Scores (PC1 X PC2) for the 20 chalcone derivatives with activity against Candida albicans: Samples 18, 17 and 11 (less potent) are left (highlighted) separated from the other samples.

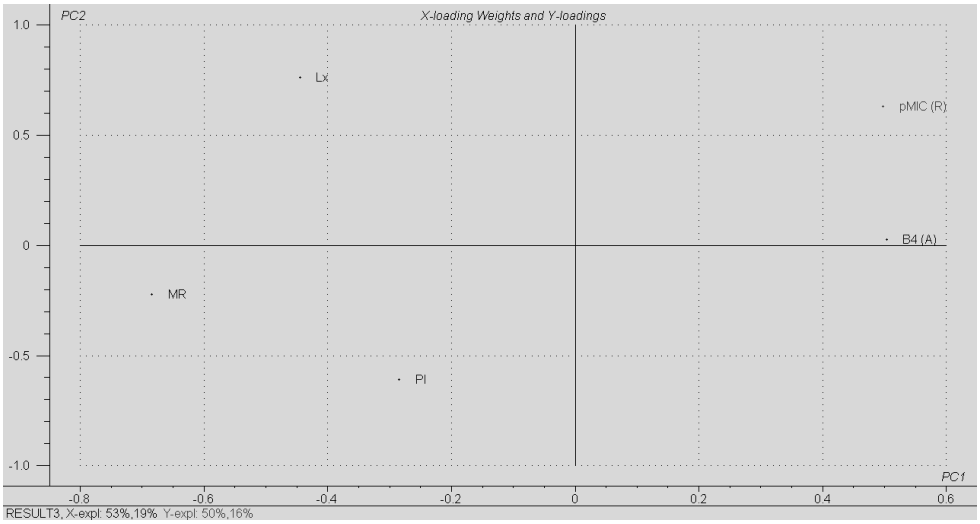


Fig. 5 - Plot of Loadings (PC1 X PC2) using the four selected descriptors to build the PLS model for the 20 chalcone derivatives with activity against Candida albicans: The molar refractivity and the ionization potential to influence negatively for the biological activity.

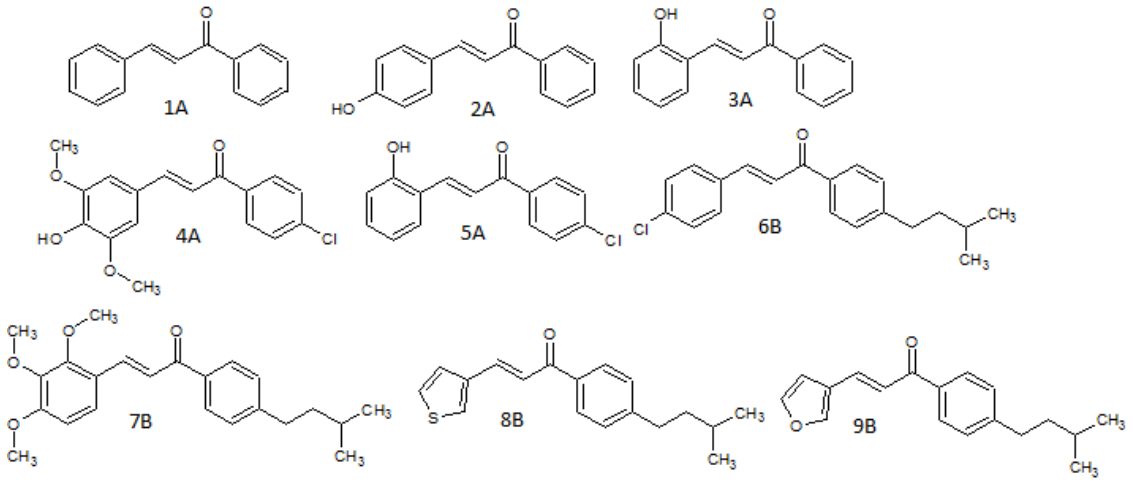


Fig. 6 - Data set (external set) from the literature [40 and 41] used in the Quantum Chemical QSAR analysis.

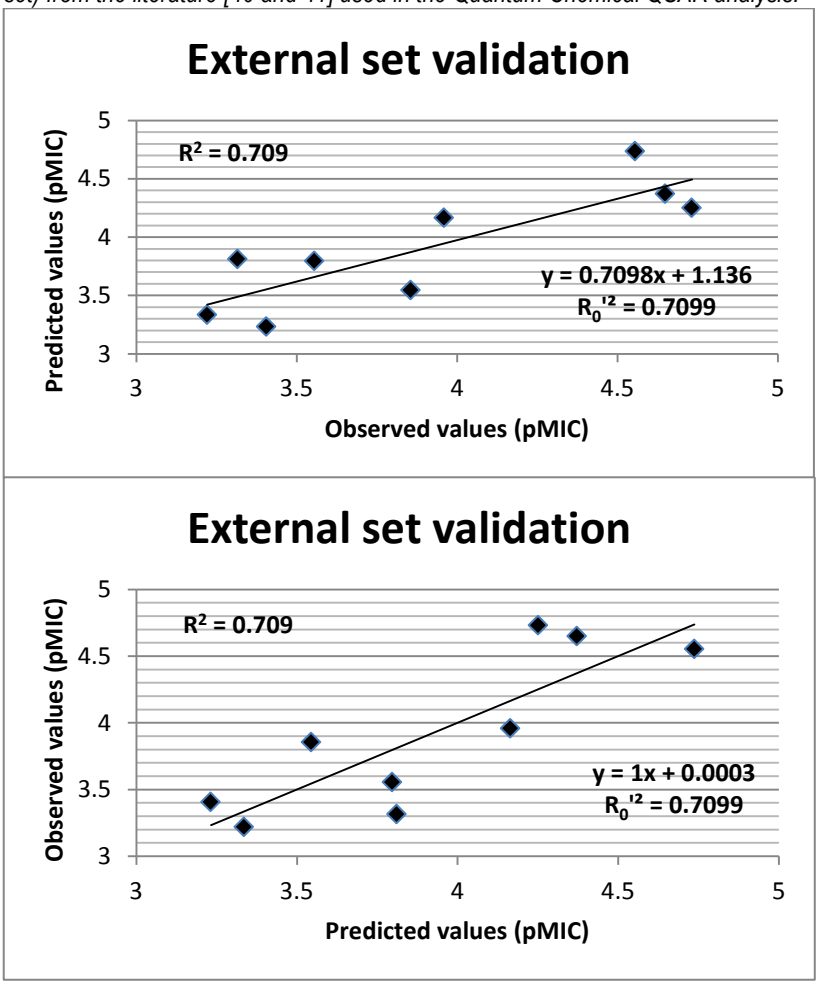


Fig. 7 - Regression plot between a) predicted vs. observed values and b) observed vs. predicted values for compounds in validation set justifying the predictive ability of QSAR model.

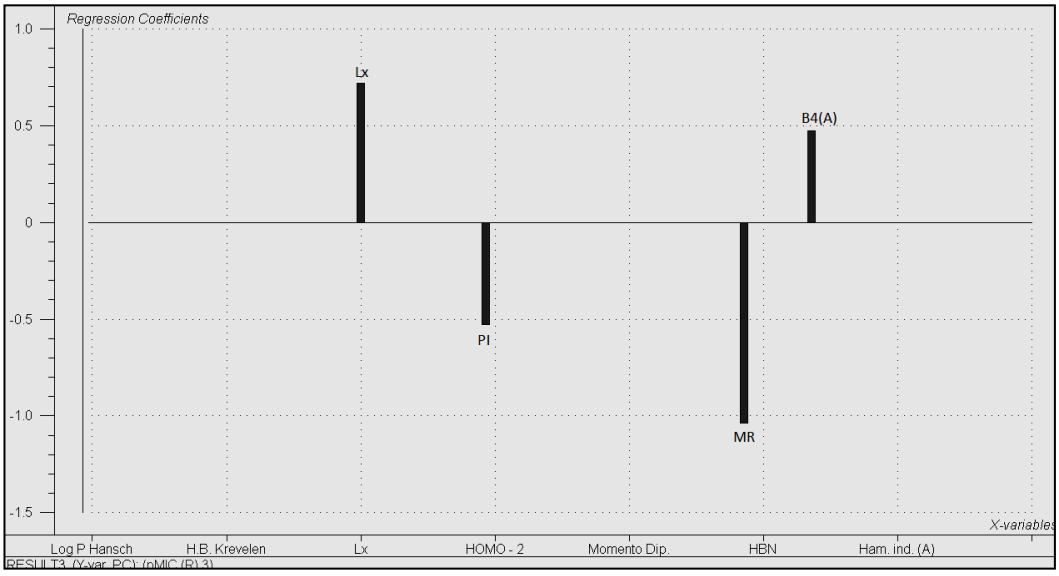


Fig. 8 - Regression coefficients for the descriptors used in PLS model.

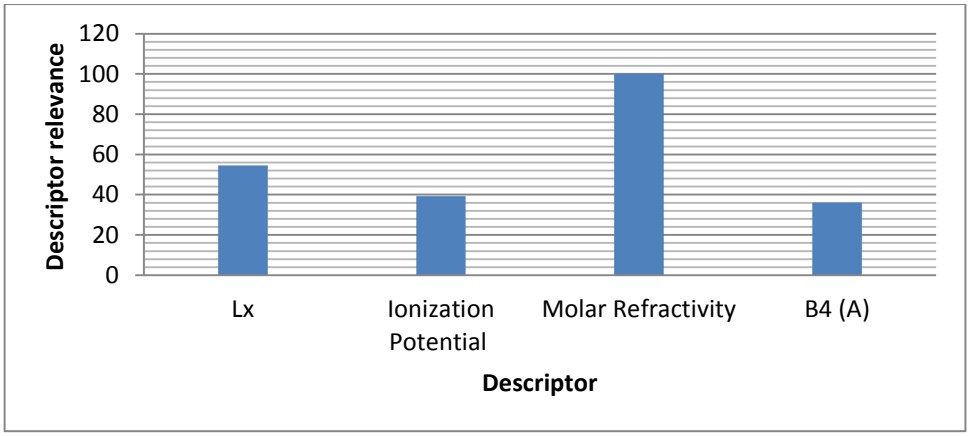


Fig. 9 - Descriptors relevance for the variables used in PLS model.

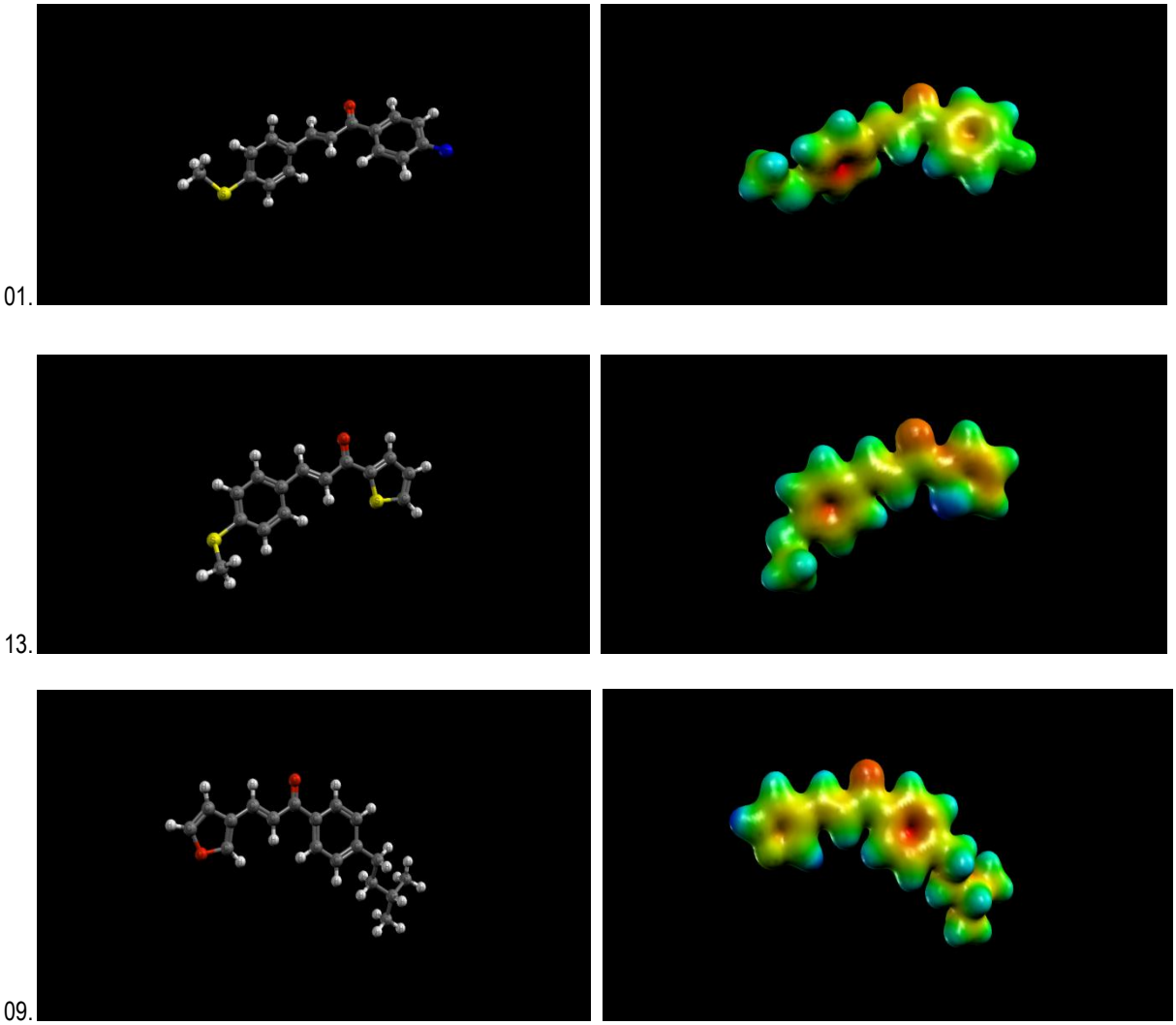
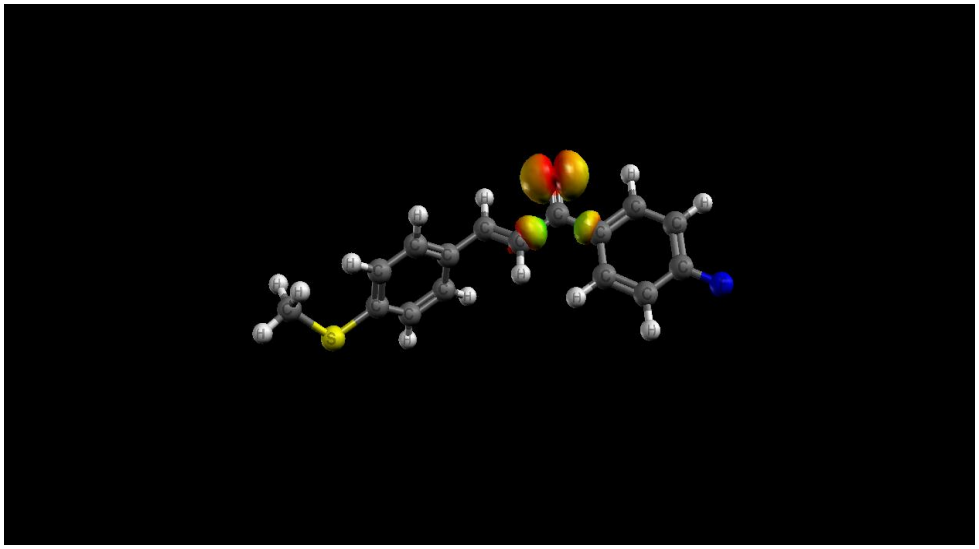
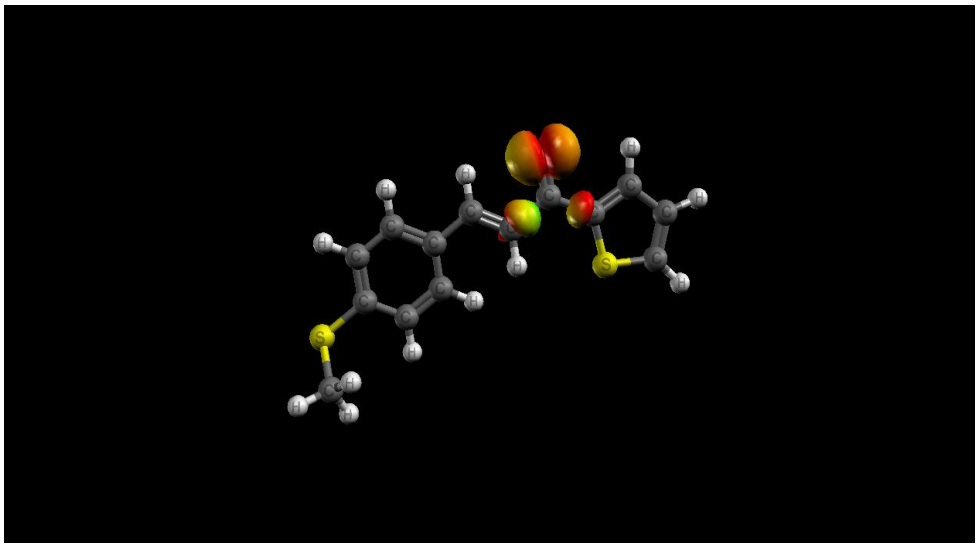


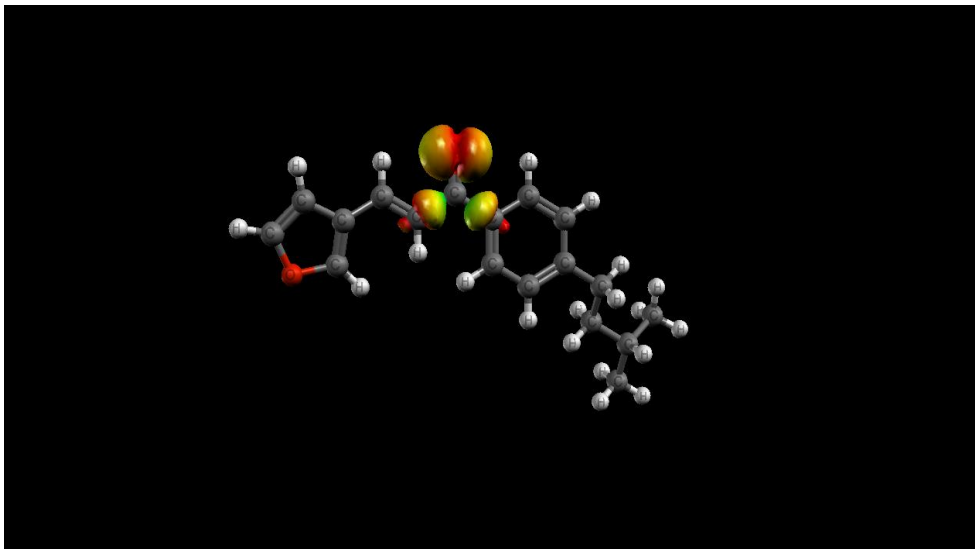
Fig. 10 - Molecular Electrostatic Potential Maps (MEP) of the most active compounds (01 and 13: training-serie; 09: external-serie). Negative electrostatic potential regions are represented in red (high electronic density) while positive electrostatic potential areas are shown in dark blue (low electronic density).



Compound 01.

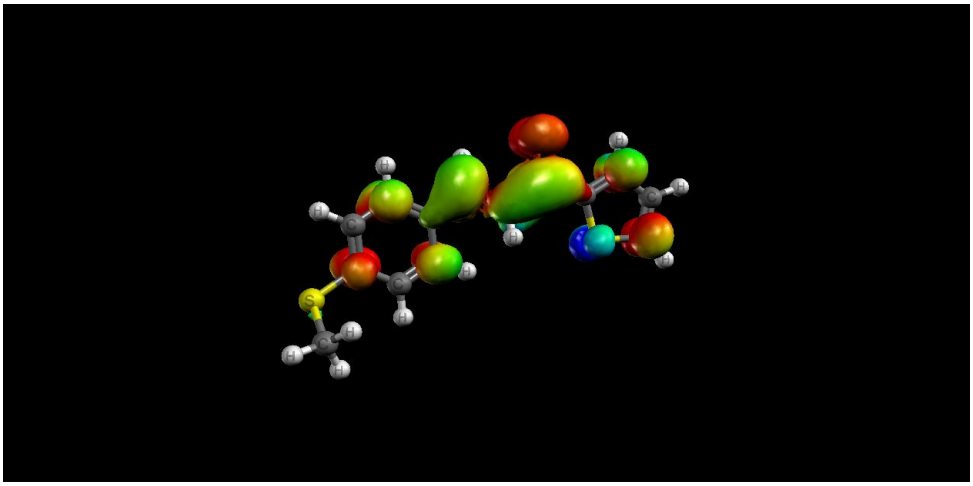


Compound 13.

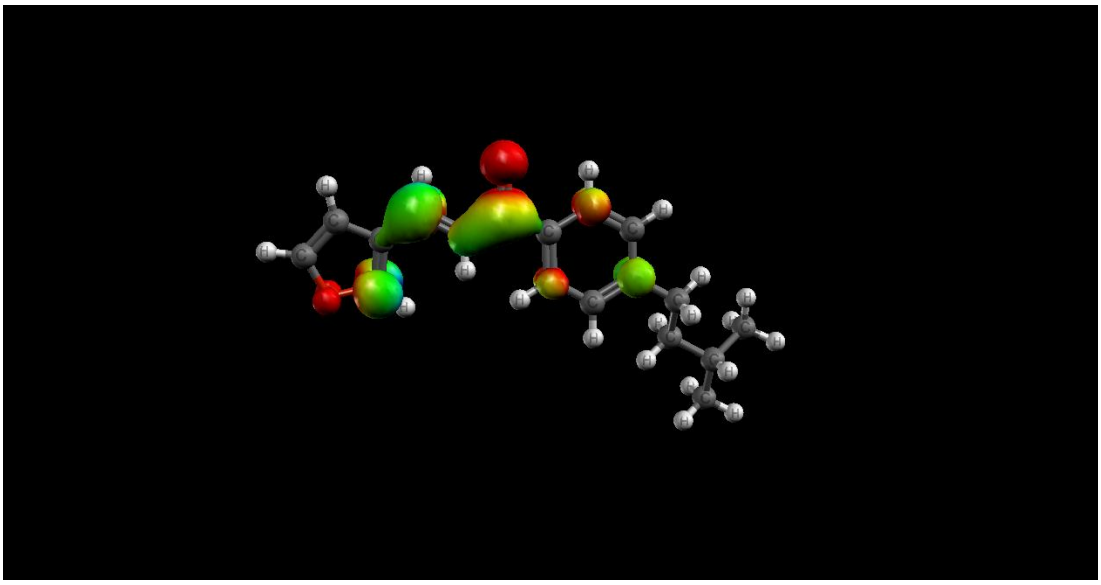


Compound 09.

Fig. 11 - Highest Occupied Molecular Orbital (HOMO) of compounds [01, 13 (training-serie)] and [09 (external-serie)].



Compound 13 (training-serie).



Compound 09 (external-serie).

Fig. 12 - Lowest Unoccupied Molecular Orbital (LUMO) of compounds: 13 (training-serie) and 09 (external-serie).

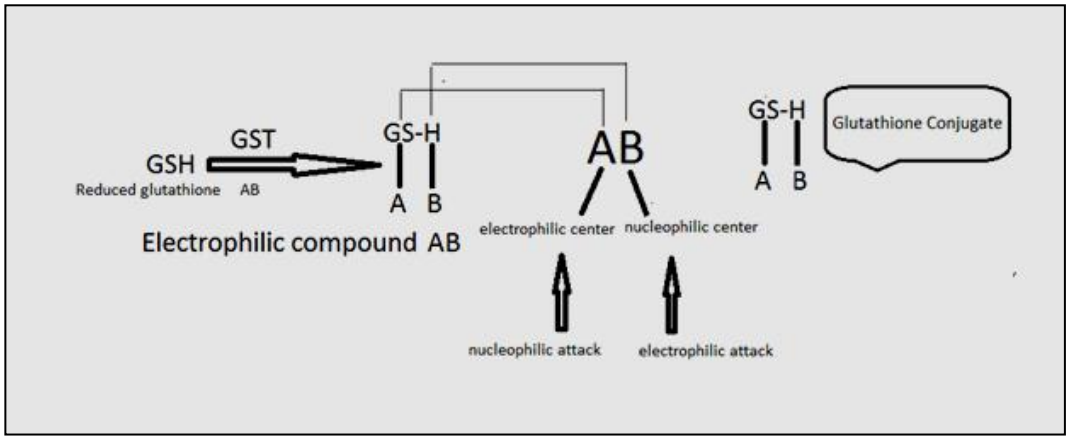


Fig. 13 - Scheme of transformation of GSH in glutathione conjugate.

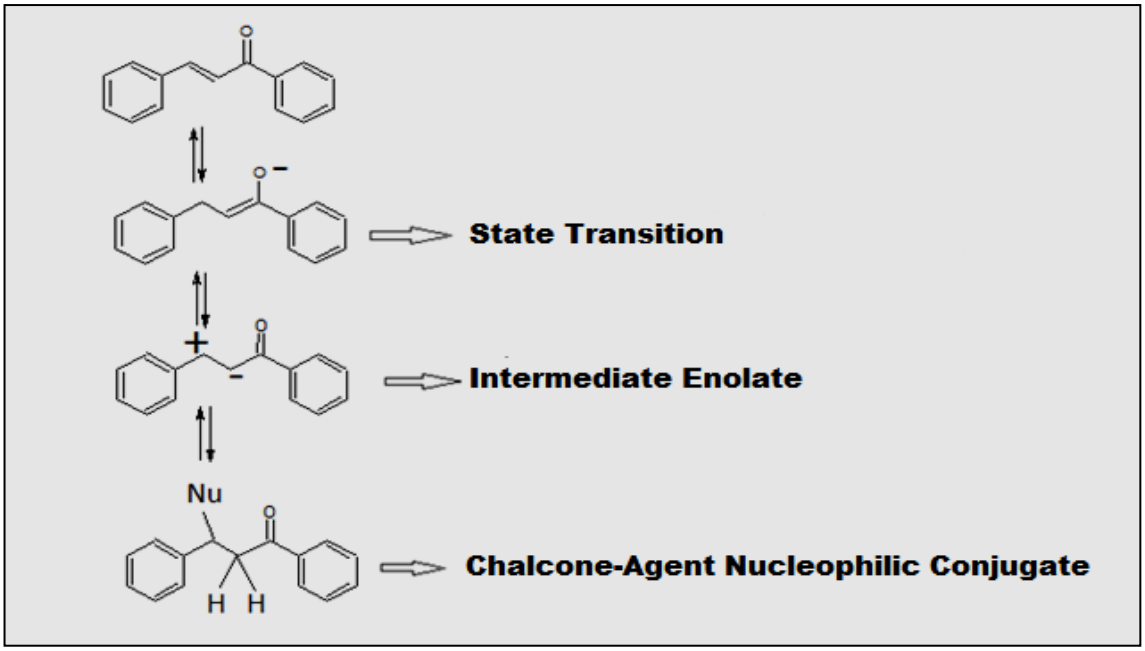


Fig. 14 - Mechanism of formation of the glutathione-chalcone conjugate.

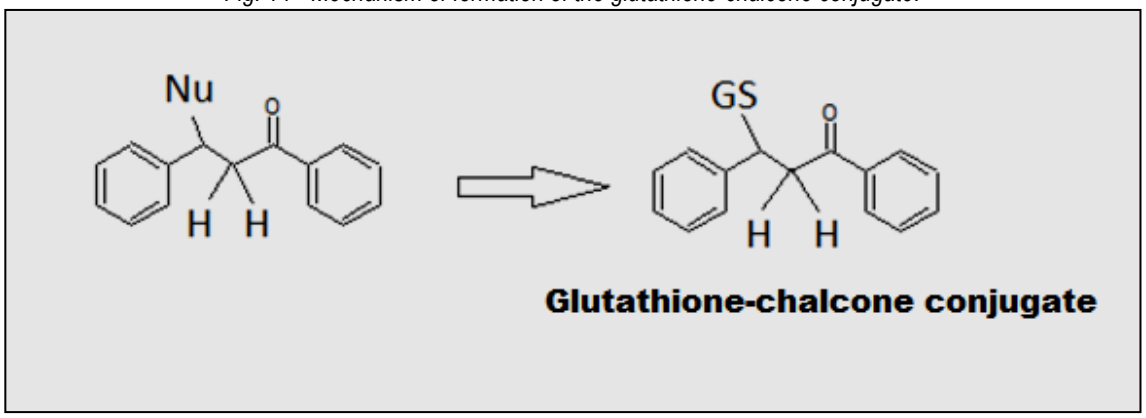


Fig. 15 - Structure of glutathione-chalcone conjugate.

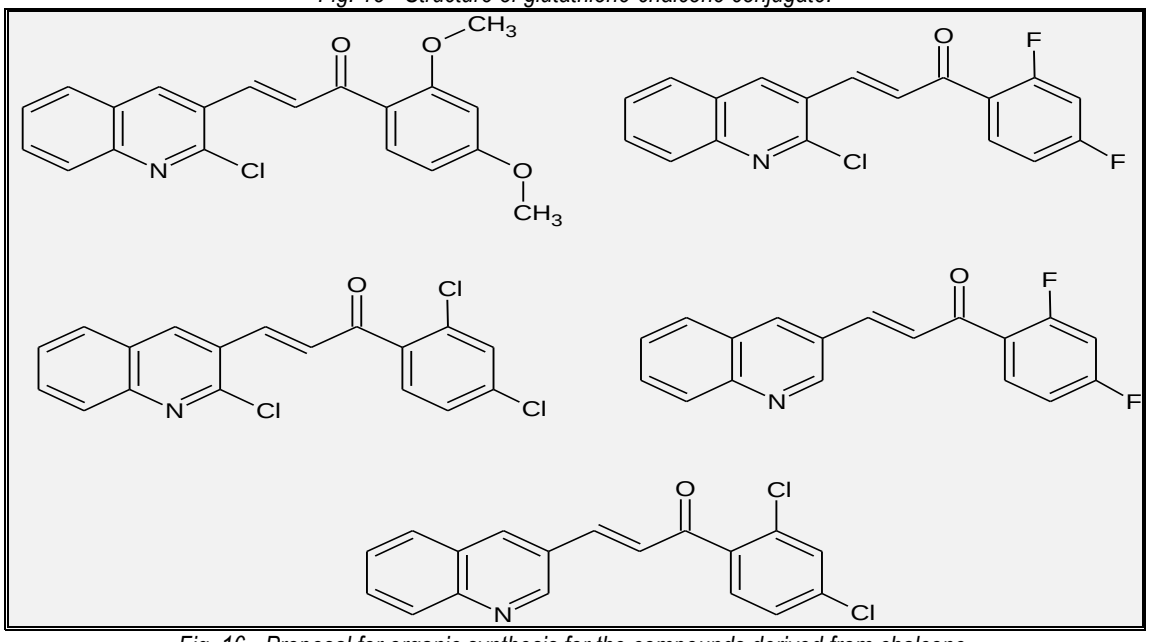


Fig. 16 - Proposal for organic synthesis for the compounds derived from chalcone.