Journal of Computer-Aided Molecular Design **17:** 277–290, 2003.

3D QSAR studies on binding affinities of coumarin natural products for glycosomal GAPDH of *Trypanosoma cruzi*

Irwin R.A. Menezes[1], Julio C.D. Lopes[1], Carlos A. Montanari[1,*], Glaucius Oliva[2], Fernando Pavão[2], Marcelo S. Castilho[2], Paulo C. Vieira[3] & Mônica T. Pupo[4]

[1]*Núcleo de Estudos em Química Medicinal, NEQUIM. Chemistry Department, University of Minas Gerais, 31270-901, Belo Horizonte, MG, Brazil;* [2]*Laboratório de Cristalografia de Proteínas e Biologia Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, USP, Av. do Trabalhador Sãocarlense 400, PO Box 369, 13566-590 São Carlos, SP, Brazil;* [3]*Departamento de Química – Universidade Federal de São Carlos. Via Washington Luiz Km 235 – CP 676, 13565-905 São Carlos, SP, Brazil;* [4]*Faculdade de Ciências Farmacêuticas de Ribeirão Preto, USP, Av. do Café, s/n, 14040-903, Monte Alegre, Ribeirão Preto, SP, Brazil*

Key words: Chagas' disease, Coumarin natural products, glycosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH), CoMFA, GRID/VolSurf/GOLPE.

Summary

Drug design strategies based on Comparative Molecular Field Analysis (CoMFA) have been used to predict the activity of new compounds. The major advantage of this approach is that it permits the analysis of a large number of quantitative descriptors and uses chemometric methods such as partial least squares (PLS) to correlate changes in bioactivity with changes in chemical structure. Because it is often difficult to rationalize all variables affecting the binding affinity of compounds using CoMFA solely, the program GRID was used to describe ligands in terms of their molecular interaction fields, MIFs. The program VolSurf that is able to compress the relevant information present in 3D maps into a few descriptors can treat these GRID fields. The binding affinities of a new set of compounds consisting of 13 coumarins, for one of which the three-dimensional ligand-enzyme bound structure is known, were studied. A final model based on the mentioned programs was independently validated by synthesizing and testing new coumarin derivatives. By relying on our knowledge of the real physical data (i.e., combining crystallographic and binding affinity results), it is also shown that ligand-based design agrees with structure-based design. The compound with the highest binding affinity was the coumarin chalepin, isolated from *Rutaceae* species, with an IC_{50} value of 55.5 μM towards the enzyme glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) from glycosomes of the parasite *Trypanosoma cruzi*, the causative agent of Chagas' disease. The proposed models from GRID MIFs have revealed the importance of lipophilic interactions in modulating the inhibition, but without excluding the dependence on stereo-electronic properties as found from CoMFA fields.

Abbreviations: CAMD, Computer-Aided Molecular Design; CoMFA, Comparative Molecular Field Analysis; CV, Cross-Validation; DALYs, disability-adjusted life years; FFD, Fractional Factorial Design; GOLPE, Generating Optimal Linear PLS Estimations; gGAPDH, glycosomal Glyceraldehyde-3-Phosphate DeHydrogenase; MIFs, Molecular Interaction Fields; PC, Principal Component; PLS, Partial Least Squares; QSAR, Quantitative Structure-Activity Relationships; SAMPLS, SAMple-distance Partial Least Squares; VRS, Virtual Receptor Site.

Introduction

Chagas' disease is estimated to affect ca. 18 million people, mostly from South and Central America, though increasing incidence has been reported in urban areas of North America. About 3 million of the infected people develop severe complications, such as chronic cardiopathy, digestive lesions and neurological disorders, causing 45,000 deaths per year and a loss of ca. 3 million disability-adjusted life years (DALYs) [1]. Blood transfusion and congenital trans-

*To whom correspondence should be addressed. E-mail: montana@dedalus.lcc.ufmg.br

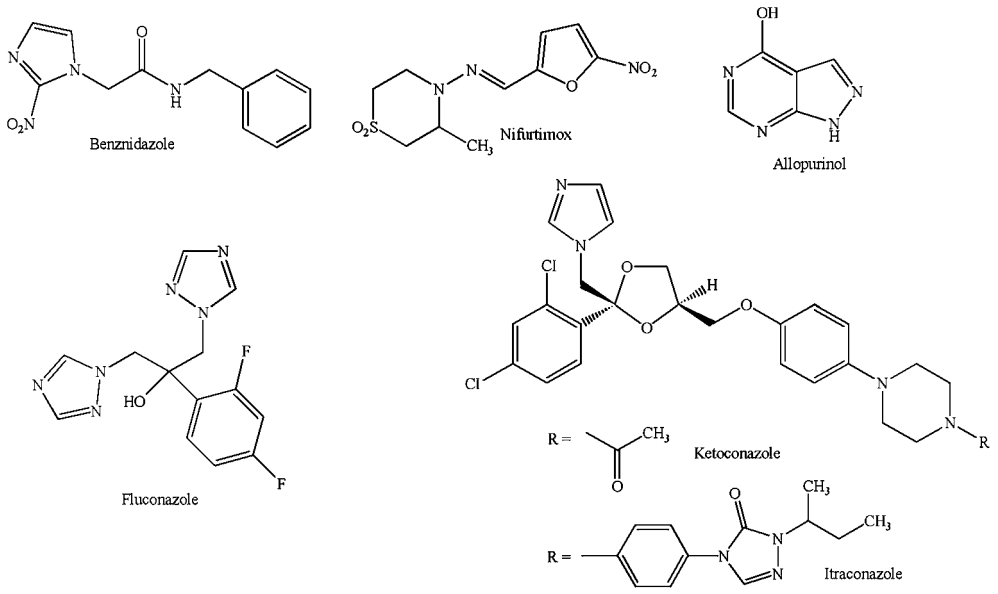


Figure 1. Structure of compounds used in the treatment of Chagas' disease.

mission are currently the major causes of the spread of the disease.

The available drugs, nifurtimox and benznidazole have strong side effects [2]. Although administered since the 1970s, [2], nifurtimox is no longer used due to severe side effects [3]. Other important molecules for treatment of Chagas' disease are allopurinol, ketoconazole, fluconazole and itraconazole [2]. Figure 1 shows their structures along with nifurtimox and benznidazole.

We are focusing on the new glyceraldehyde-3-phosphate dehydrogenase, gGAPDH, for the discovery of drugs acting against *T. cruzi* [4–6]. Since intracellular amastigotes derive their energy mainly from glycolysis, the inhibition of gGAPDH potentially prevents *T. cruzi* from being infective [2]. The need for discovering bioactive new chemical entities, BIONCEs, against Chagas' disease has stimulated our group to begin a study of screening natural compounds against recombinant *T. cruzi* gGAPDH obtained in a pET3a-*Escherichia coli* expression system [4, 7].

Natural products derived from medicinal plants represent a very powerful array of compounds that can be used as 'hit-to-lead' candidates of interest in medicinal chemistry [8]. The isolation and identification of natural products candidates are widespread in tropical Countries. Brazilian biodiversity is potentially full of hits, which are just waiting to be leads. Nevertheless, mass screening is not an easy task since it can be quite expensive and time consuming. Chemical diversity is buried throughout many laboratories waiting to be discovered. In attempting to circumvent the problem we have carried out a computer aided molecular design, CAMD, to perform virtual screening of natural products [9, 10]. The joint tools of structure-based design and classical or 3D QSAR studies can be employed when structure of the receptor is known. Affinities can be improved in advance through CAMD. After that, binding can be measured experimentally and the system re-cycled to aid in the choice of only a small number of molecules to be screened. At this point, the medicinal chemistry project can be started [8].

The Brazilian physician Carlos Chagas discovered the disease in 1909. Though many years have past since Dr. Chagas discovered the ethiology of the disease no new drug has entered in the market to help the myriad of patients who severely suffer from it. The main objective of this study is to find new natural product hits from Brazilian flora, which are capable of acting against the *T. cruzi* glycosomal enzyme gGAPDH. Structure-Activity Relationships, SAR, will be developed in order to help fulfil this goal.

Materials and methods

Computer hardware

All calculations presented were performed on a R10000 O2 Silicon Graphics workstation.

In-house natural products database

The starting compounds analyzed in this study were natural products, isolated and structurally identified, in our on-going project of identifying molecules from medicinal plants of tropical disease interest.

Molecular docking

A library of 93,000 compounds was ranked according to DOCK3.5 scores [11], using the gGAPDH X-Ray crystallographic structure as target [4, 12]. The binding site of the adenosine ring of the NAD^+ cofactor and the catalytic site of gGAPDH possesses significant differences between the parasite and the homologous human enzyme [4], thereby being an attractive target for trypanocidal drug design [13]. The process of using DOCK provided some new coumarin natural products highly ranked among many compounds from the database. These well-scoring compounds, isolated from several previous phytochemical studies and available in our laboratories, were among others screened against our recombinant *T. cruzi* enzyme [7].

Receptor binding affinities

The ligands' *in vitro* affinity for *T. cruzi* gGAPDH receptor is the dependent variable considered in this study. In the binding study, the affinity of the compounds was determined by their ability to displace NAD^+ from the enzyme receptor site [7]. Receptor binding affinities were expressed as IC_{50} (μM) values. The natural coumarin chalepin was selected for determining the co-complex with gGAPDH [7, 12, 13] because it has the best inhibitory concentration towards the enzyme.

X-Ray crystallographic data and structure determination of chalepin bound to gGAPDH

The structure of the complex of *T. cruzi* gGAPDH with chalepin has been reported [12]. The gGAPDH receptor site was studied in order to compare its calculated GRID MIFs with those ones obtained for ligands themselves, with the aim of establishing relationships among them two.

Molecular fields generated in the Sybyl/CoMFA and GRID program

In both SYBYL/CoMFA [14] and GRID, [15–17] a grid big enough to enclose all ligands was created. The interaction energies between a probe (Csp^{3+} for CoMFA; OH2 and DRY for GRID), and the target molecules were calculated, in each grid point.

Molecular modelling

Sybyl Tripos force field [18] was used for modelling the 13 molecules found in Table 1. Derived partial charges obtained from Gasteiger-Hückel method to calculate the electrostatic field interactions in CoMFA were used throughout this work.

The starting chalepin conformation was the one obtained from our X-Ray structure of the co-complex between chalepin and gGAPDH, which has been taken as a template to build all molecules in the set. The conformational space, of all ligands, was explored by a conformational search procedure using the systematic search method, as implemented in Sybyl. The new minimized selected conformations were outside 5 kcal mol^{-1} difference, whenever feasible.

The steric and electrostatic potential energies were calculated with the standard CoMFA procedure. The SAMPLS algorithm implemented in Sybyl6.5 was used to perform the cross-validation analyses. The non-cross-validated data were used in the analysis of the field results, calculations and the predictions.

Molecular interaction fields, MIFs

MIFs were obtained using the program GRID version 20. Computed on gGAPDH receptor site, the MIFs identified regions where water and DRY probes interacted favourably, suggesting positions where functional groups should be placed in ligands.

MIFs were also computed for the ligands themselves. In this case, the regions showing favourable energy of interaction represent positions where groups of the receptor interact with the ligands. Using different probes, we can obtain for a certain ligand a set of such positions, which characterizes a '*Virtual Receptor Site*' (VRS). This abstract entity defines an ideal complementary site for a certain chemical compound and represents its potential ability to bind gGAPDH. The regions defined in its VRS overlap groups of the real receptor site where at least a subset of the VRS is relevant for representing the binding properties of the ligands.

VolSurf descriptors

The GRID-VolSurf [19] procedure is completely automated. Firstly, it generates MIFs by using GRID20. Then, it treats the fields accordingly by producing descriptors that encode the information content from the used probes. VolSurf has the advantage of producing descriptors using the 3D information embedded in any map. VolSurf is also alignment independent and conformation insensitive (independent, in this study). The VolSurf transformation is fast and its results are usually easy to interpret. The descriptors have a clear chemical meaning and are lattice-independent.

The GRID-VolSurf descriptors [19] (water and hydrophobic DRY based probes) on studied coumarins exported to GOLPE program were treated by the variable selection method of Fractional Factorial Design, FFD, [20]. The FFD variable selection procedure was applied in two runs, but only the first one yielded the appropriate model (active variables: 56, Dummies: 15, X-comb.: 256, SDEP = 0.190, r^2_{cv} (Q^2) = 0.610, Max. dimensionality: 1, Validation Mode: LOO, Recalculate weights: yes, Comb./Var. ratio: 2.0). After the FFD run, the final number of X-variables was reduced to 28.

The partial least squares (PLS) model

In the context of 3D-QSAR, the biological activity may be seen as a function of the physiochemical characteristics (such as electronic properties or energies of interaction within a given force field) of the compounds of interest. The need to convert such numerical data to useful information has led to the development of methodologies that rely on statistics and applied mathematics. The PLS model [21] is a two-block projection method that relates a matrix X (containing the chemical descriptors) to a matrix Y (containing the biological activities) with the aim of predicting the values in Y from the information contained in X. This has been the method used in CoMFA and Volsurf analyses.

Generating optimal linear PLS estimation (GOLPE)

GOLPE [22] is defined as an advanced variable selection procedure aimed at obtaining PLS regression models with the highest prediction capability, which relies on the validation of a number of reduced models on variable combinations, selected according to a factorial design strategy. The power of a GOLPE procedure will depend on some method of pre-treatment of the data. In this paper, descriptors generated via VolSurf analysis were the input.

Results and discussion

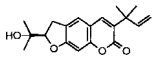
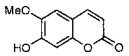
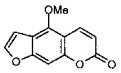
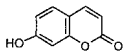
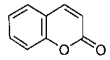
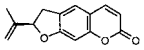
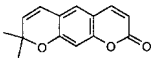
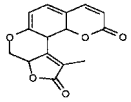
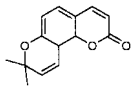
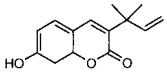
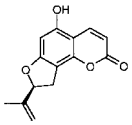
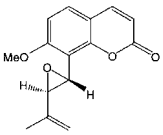
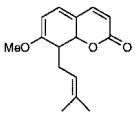
From our natural products program on the identification, isolation, and structure elucidation of small molecules relating to drug design, coumarins obtained from *Rutaceae* species [7], were found to be hits for the key enzyme gGAPDH from *T. cruzi*, through a virtual screening procedure. The study molecules were docked in the receptor site of gGAPDH. The interaction energies ranked according to the scoring function of DOCK3.5, which allowed them to be selected for the screening assay against the enzyme. Sampling the coordinate space of the binding site and scoring each possible ligand according to the interaction energies, resulted in the predicted binding mode for that compound. The whole procedure is described elsewhere [23].

A defined protocol [7] was conducted for measuring their inhibitory concentrations. The binding affinity data of the 13 chosen coumarin compounds can be found in Table 1, along with their chemical structures, and the calculated binding affinities.

Two different computer aided molecular design methods were applied to describe the binding affinities of chalepin analogues towards gGAPDH, namely CoMFA and GRID. These two methods were chosen because together they can incorporate information on binding affinities according to electrostatic, steric and hydrophobic fields.

From the rigid docking analyses [23], no ligand flexibility was taken into consideration. The conformation of a compound bound to the receptor site might be different from the conformation of the unbound form in solution. Hence, having started with CoMFA analysis of coumarin ligands shown in Table 1, we generated and relaxed conformations from a systematic conformational search of coumarin flexible moieties, as available in Sybyl6.5. The bound conformation of chalepin to gGAPDH was also tested as the pharmacophoric one in obtaining CoMFA fields. The needed similar analogue conformations were generated as described elsewhere [24]. All compounds were then aligned atom-by-atom as follows: (i) over compound 1(TC_cum_1), chalepin, as obtained from the receptor site, (ii) over the lowest energy conformation of all compounds, and (iii) over the highest energy conformation; but all sharing the core of coumarin

Table 1. Coumarins assayed and their binding affinities (actual/calculated) toward gGAPDH.

Compounds	Name	Structures[1]	Actual[2] (log1 / IC_{50})	CoMFA[3] calculated	CoMFA Residual
1 Chalepin	TC_cum_1		4.26*	4.26	0.0022
2	TC_cum_10		3.59	3.55	0.04
3	TC_cum_11		3.46	3.43	0.03
4	TC_cum_12		3.21	3.25	−0.04
5	TC_cum_13		3.16	3.16	−0.002
6	TC_cum_2		3.84	3.88	−0.04
7	TC_cum_3		3.71*	3.71	−0.0039
8	TC_cum_4		4.13*	4.12	0.01
9	TC_cum_5		3.91	3.92	−0.01
10	TC_cum_6		3.89	3.88	0.01
11	TC_cum_7		3.79	3.82	−0.03
12	TC_cum_8		3.72	3.69	0.03
13	TC_cum_9		3.68	3.68	−0.005

[1]Chemical structures of selected coumarin natural products from virtual screening of DOCK3.5. References [7, 23].
[2]Binding affinity magnitude values obtained from *in vitro* study of the *T. cruzi* gGAPDH inhibition. Values are taken as log1/IC_{50}, in μM, [7]. Marked values (*) were corrected before being applied in this work.
[3]CoMFA internal calculated binding affinities, according to steric and electrostatic molecular fields.

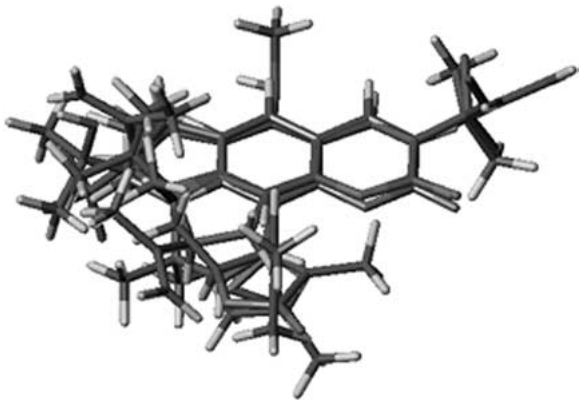


Figure 2. Coumarin atom-by-atom superposition used for CoMFA analysis.

rings. Thirty percent of coumarin compounds listed in the Table 1 was dropped from the analysis to act as test compounds. By using the standard Tripos force field from Sybyl version 6.5 program [18], we calculated the electrostatic and steric fields. Figure 2 shows the structures and the alignment used for all *pharmacophoric* (bound) conformations as obtained from receptor site.

The results shown in Table 2 do reveal relationships between steric and electrostatic molecular fields calculated from CoMFA.

The optimum number of components was found to be two (Table 2). Those values were sufficient for a 3D QSAR model. Furthermore, the corresponding non-cross-validated QSAR model has a good fit with r^2 (0.99) and a small s value, and with reasonable F value. The model also suggests there is a major steric contribution of binding affinity towards gGAPDH.

The CoMFA coefficient map for the model was contoured around chalepin, with the highest affinity towards gGAPDH (Figure 3) to illustrate the locations that affect binding. Both steric and electrostatic fields are displayed at the same location. They were contoured as follows: the contours of the steric map are shown in yellow; those of the electrostatic map are shown in blue. Greater values of binding affinities are correlated with less bulk near yellow, more positive charge near blue. The steric maps (Figure 3) indicate that less bulk constituents at the both ends of chalepin are favourable for increasing binding affinities.

Conducting predictions on the internal test set validated the predictive capability of the CoMFA model. As a result, by using the contour maps from Figure 3, we have investigated new derivatives to be synthesized. This has been done by changing the 1,1-dimethylallyl moiety of chalepin with bulkier groups, since synthetic efforts were firstly made possible through modifications of the double bond. Of course, with the CoMFA prediction, chalepin derivatives were not better then the chalepin itself. Nevertheless, this negative result is enough to validate externally our CoMFA model, because predicted bulkier groups at the 1,1-dimethylallyl moiety to be detrimental for binding to gGAPDH.

The designed new molecules with better calculated binding affinities based on coumarin 8(TC_cum_4) are not yet synthetically feasible in our laboratories. Since we would like to have molecules with higher affinities toward gGAPDH, and make them synthetically possible, we decided to attempt incorporation of new molecular interaction fields, MIFs, in the model. For this reason, we have included hydrophobic fields from GRID, because these fields are needed to keep chalepin at the receptor site, (see below).

GRID-VolSurf

The existence of H-bonding and hydrophobic pockets in the receptor site may be investigated through a new procedure called VolSurf [25, 26]. This approach has been used to correlate 3D molecular interaction fields, MIFs, with physico-chemical and pharmacokinetic properties, plasticization [27], bioactivities [28], chemical space navigation [29]. To date the main use of VolSurf is related to structure-property relationships, albeit structure-activity relationships and structure-binding affinities should be possible. During the preparation of this manuscript, the use of VolSurf to calculate surface descriptors for protein-ligand affinity was published [30].

Hydrogen bonding and hydrophobicity, duly calculated by GRID force field [15–17], mediate some important interactions between chalepin and its receptor site. Figure 4 shows chalepin calculated MIFs for water and DRY (the hydrophobic probe).

The coloured regions represent where the interaction between the ligand and the (a) water and (b) DRY probes are highly favourable. The contour encloses regions where the target molecule can make (a) hydrogen bonds and (b) lipophilic interactions. The interaction energy moments, red arrows in Figure 4, referring to (a) hydrophilic regions, are vectors pointing from the centre of mass to the centre of the hydrophilic regions; (b) hydrophobic regions, measure the unbalance between the centre of mass of a molecule and the barycentre of the hydrophobic regions.

Table 2. PLS results of CoMFA model for 13 coumarin binding affinities toward gGAPDH.

Property	Cross-validated r^2_{cv}		Non-cross-validated			Steric Field (%)	Electrostatic Field (%)
	LOO*	Components	r^2	S	F		
log $1/IC_{50}$	0.547	2	0.994	0.035	166.35	64.3	35.7

*Leave-one-out.

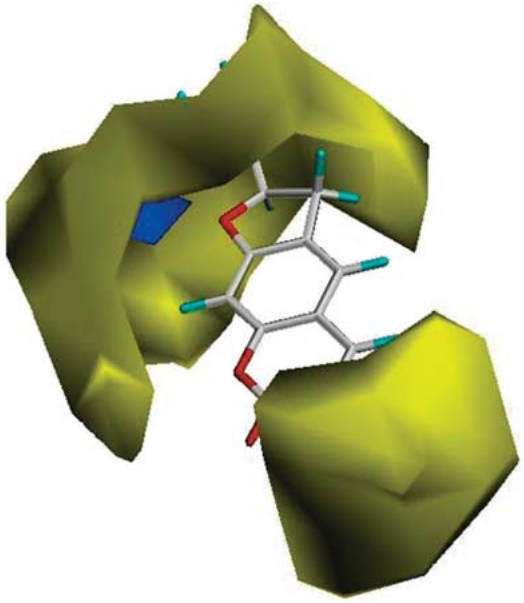


Figure 3. Contour maps of CoMFA StDev*Coeff around chalepin. The yellow maps favour less bulk groups with higher affinity towards gGAPDH.

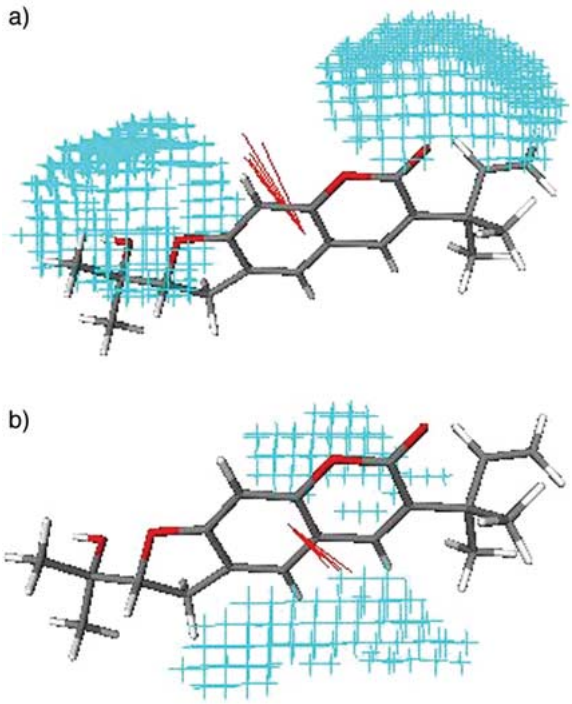


Figure 4. GRID 3D molecular fields of chalepin calculated with (a) water and (b) DRY probes, contoured at −3.5 and −1.25 kcal mol^{-1}, respectively. The red arrows represent the integy moment.

The interaction energies are depicted in the same direction as the one found for chalepin in the receptor site, which is for hydrophilic regions needed for H-Bonding to water molecules, whereas the lipophilic ones are directed to the hydrophobic regions of the gGAPDH pocket (see Figure 8).

A PLS-CV analysis is also of high importance when affinities have to be considered. From PLS calculations available in VolSurf program, the binding affinity fit of the data was $r^2 = 0.800$, whereas the prediction $r^2_{cv} = 0.602$, and SDEC = 0.136, SDEP = 0.192, respectively. These results, when compared to the ones obtained from CoMFA PLS analysis were only slightly better for the prediction power. Consequently, we have used GOLPE to recalculate them (Table 3). The squared cross-validated term has risen to 0.749 with only one component. In addition, we have tested different manners of cross-validating the data. All of them are consistent with results shown in Figure 5, which shows the PLS plot for all calculated binding affinities. Table 4 shows calculated affinity values, and Figure 5 shows that compounds 1(TC_cum_1) chalepin, and 8(TC_cum_4) have the highest affinities towards gGAPDH, while 4(TC_cum_12) and 5(TC_cum_13) have the lowest.

It is worth mentioning the calculated GOLPE-VolSurf values shown in Table 4 come from a few vectors of variables (see methods), whereas CoMFA calculated values found in Table 1 were developed from thousands. This is important for the evaluation of the performance of both methods.

Figure 6 shows the PLS coefficients for PC1 VolSurf descriptors. Variables representing the hydrophobic regions are important to keep coumarins in the receptor site. Since they have positive magnitude values, we would expect the increase of the DRY

Table 3. PLS results of GOLPE model, based on GRID/VolSurf MIFs and descriptors.

Property	Cross-validated (r^2_{cv})					Non-cross-validated	
	LOO[1]	LTO[2]	5 random groups out	Components	SDEP[3]	r^2	SDEC[4]
log 1/IC50	0.758	0.746	0.746	1	0.192	0.838	0.122

[1]Leave-one-out.
[2]Leave-two-out.
[3]Standard Deviation of Estimated Prediction.
[4] Standard Deviation of Estimated Calculation.

Table 4. GOLPE-VolSurf calculated binding affinities for study coumarins.

Compounds[1]	GOLPE-VolSurf[2] calculated	GOLPE-VolSurf residual
1 Chalepin	4.16	0.1
2	3.66	−0.07
3	3.73	−0.27
4	3.46	−0.25
5	3.27	−0.11
6	3.67	0.17
7	3.66	0.05
8	4.03	0.1
9	3.57	0.34
10	3.65	0.24
11	3.87	−0.08
12	3.79	−0.07
13	3.75	−0.07

[1]See Table 1 for structures.
[2]GOLPE calculated values of binding affinities, according toVolSurf descriptor calculations.
See text for explanation.

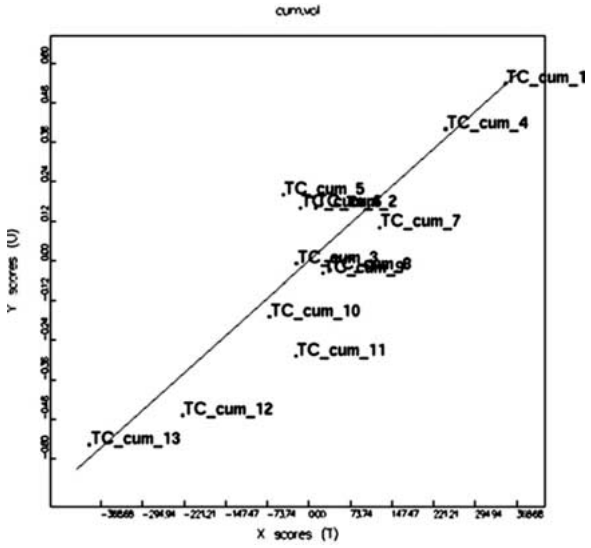


Figure 5. PLS plot of T versus U, for one component only.

volumes in molecules to increase affinities. However, the hydrophilic-lipophilic regions, albeit a bit lower in height as compared to the hydrophobic peaks, exhibit the same trend. On the other hand, the integy moments are negatively related to binding affinities, and thus we would expect as they increase the binding would decrease. These high integy moments represent strong polar regions concentrated in few regions of the molecular surface. This can be seen from Figure 4 where the polar regions are within specific parts of the molecule. Size and shapes can be used for similar interpretation of binding affinities. These observations are essentially the same as found from PLS partial weights plot (not shown).

From these models, we designed synthetically feasible new coumarin derivatives in order to validate them now externally.

Considering that steric, electrostatic and hydrophobic requirements for the new structures in the 3D MIFs were used to account for coumarin gGAPDH binding affinities, we tailored the ligands with synthetic feasibility in mind. However, this time focusing on molecular modifications of the coumarin at positions X,Y,Z (Figure 7).

Introducing different groups at positions 3 and 6 of the coumarin ring (Z = H), and calculating their binding affinities via both of the proposed models (CoMFA and VolSurf) resulted in the structures shown in Table 5. Firstly, by analogy to early published flavones [31], with affinities toward gGAPDH, a piperonyl group was linked to position 3 of the coumarin ring. Functionality at position 6 was accessed by the existence of nitro and hydroxyl groups. Reduction of the first to the amino group is appropriate for imino-de-oxo-bisubstitutions, yielding stable Schiff bases, whereas the second one is suitable for acid esterifications, for instance.

The introduction of the groups at positions 3 and 6 reinforced the need for diminishing the strength of integy moments found in compounds with lower affin-

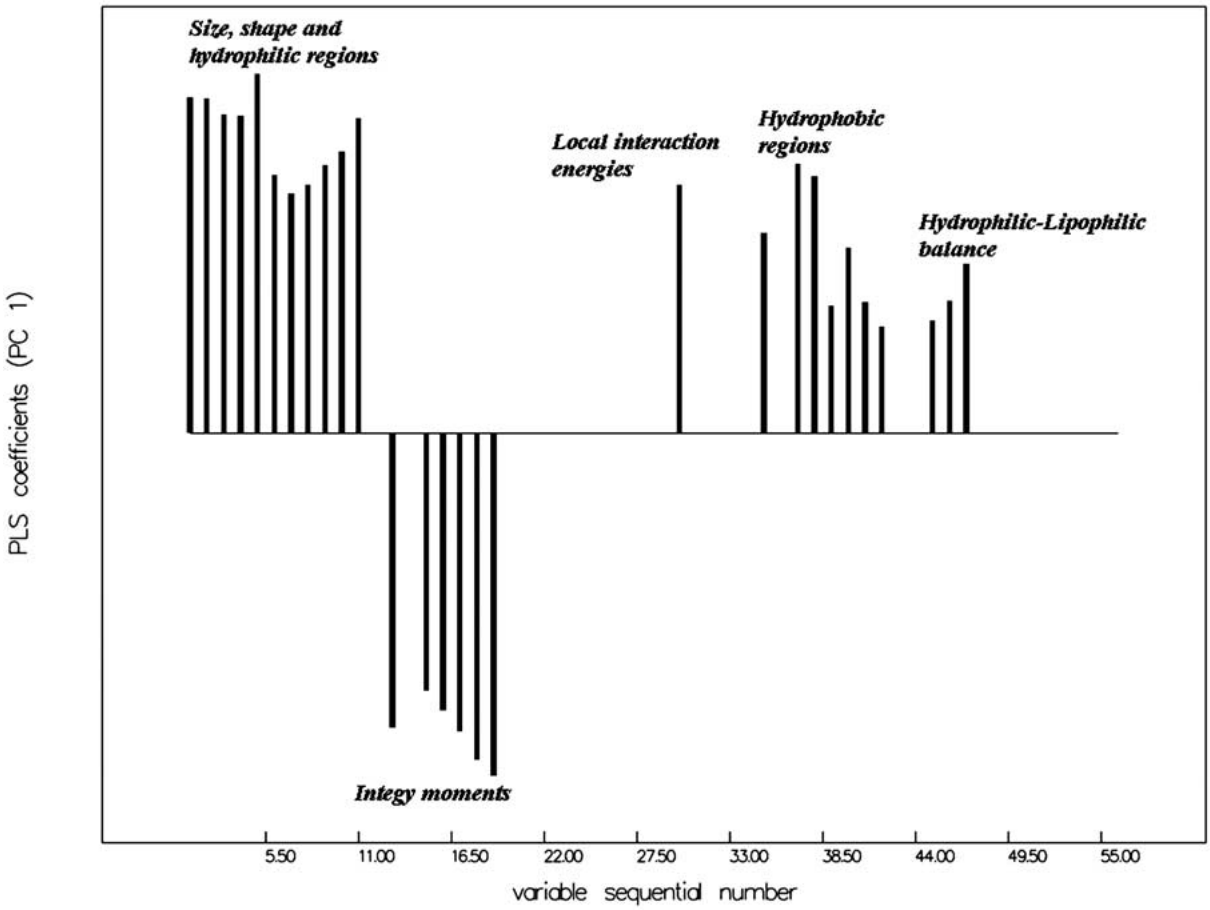


Figure 6. PLS coefficients for PC1 VolSurf descriptors.

ities towards gGAPDH (e.g. compounds TC_cum_12 and TC_cum_13). Accordingly, compounds with higher affinities have negligible integy moments (e.g. chalepin and compound TC_cum_4), as found for the 3 new coumarin derivatives herein proposed. This can be easily seen from the high negative values of integy moments in Figure 6, which are detrimental to binding affinities.

Out of many searched sub-structures from our database, using the featured VolSurf descriptors, the 3 reported herein were properly validated in accordance with the model depicted at Figure 5 (also from the CoMFA model).

In a joint collaboration project, Pupo and coworkers [32] have synthesized and tested the three new coumarin analogues. Table 5 shows the results of binding affinities towards gGAPDH, along with our CAMD predictions. The CoMFA results are also shown to allow a comparison between all values.

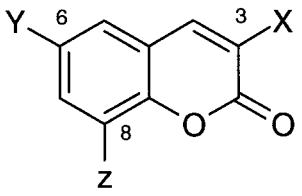


Figure 7. Coumarin template used for molecular modifications.

As can be seen from Table 5 all 3 newly synthesized and tested coumarin analogues against gGAPDH have similar binding affinities with chalepin, hence validating externally our models. Consequently, we are now building up a focused combinatorial library with the available VolSurf descriptors in order to further search the chemical space, synthesize new coumarins and test them against gGAPDH. The results of these studies shall be published elsewhere in due time.

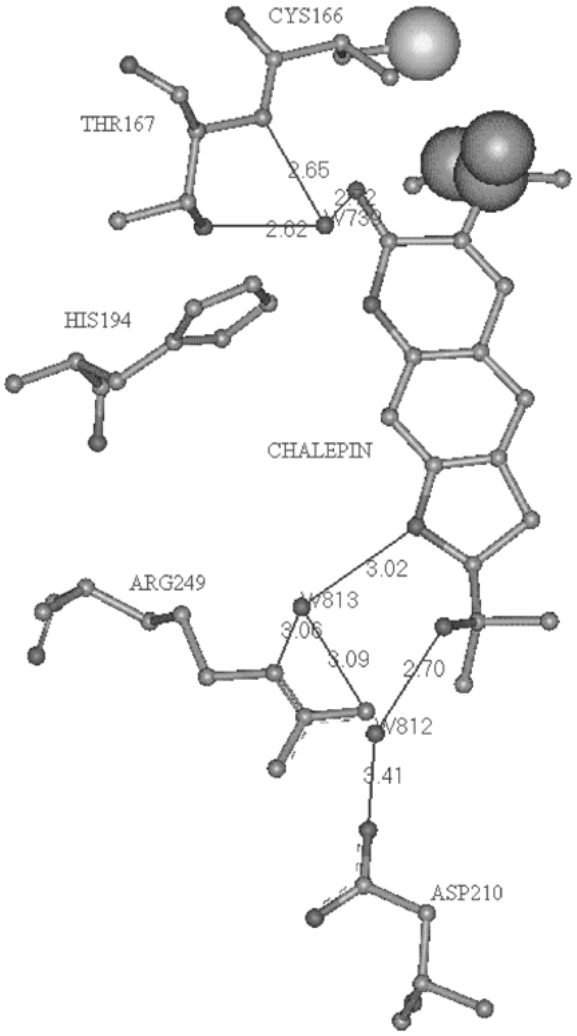


Figure 8. T. cruzi gGAPDH-chalepin H-bond interactions.

Ligand-based design agrees with X-Ray crystallographic complex between chalepin and gGAPDH

Although the above results are sufficient to understand the major reasons for binding of chalepin analogues towards gGAPDH, one further question has arisen: how do we know there is a relationship between (Q)SARs for ligands and their role in the receptor site? This is a key question because we need to be sure they are binding in the same way to the receptor site. By studying the receptor site itself, we concluded that all the above results agree with the co-complex between chalepin and gGAPDH. This kind of validation [33] is very important and we have been using it with success [24, 34].

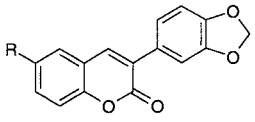
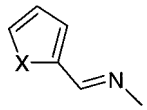
There are two water molecules in the site (upheld by our GRID20 calculations), Figure 8: one is W739, at 2.72 A from chalepin and 2.62, 2.65 A from Thr167 residue, whilst the second one W813 bridges between chalepin and Arg249, with H-bonding system at 3.02, 3.06 and 3.09 A. The 1-hydroxy-1-methylethyl group interacts weakly with the residue Asp210 by hydrogen bond mediated by a third water molecule W812. The close contact of chalepin 1,1-dimethylallyl moiety with Cys166 at ca. 2.9 A is in agreement with our CoMFA results, where bulkily substitutions would not be favourable.

Figure 9 shows the GRID plots for gGAPDH site with (a) water, (b) DRY and (c) Csp^2 probes, contoured at -7.0, -0.5 and -3.0 kcal mol^{-1}, respectively. Figure 9(a) shows water fields (WO) located at the same region of the X-Ray crystallographic structure (W739, W812 and W813). The first field is bridged between Thr167 and chalepin itself, BRZ960, (shown in Figure 9 for clarity), which identifies the interactions found so far in the X-Ray complex structure, while the second one is H-bonding to the chalepin O atom from the benzo-dihydrofuran group. The third one is in contact with the 1-hydroxy-1-methylethyl group of chalepin. Figure 9(b) shows the field for DRY probe. The lipophilic pocket that can properly accommodate the ligand is contoured just above the coumarin moiety. Due to the finding of this hydrophobic pocket and also to the postulated interaction between Cys166 and the 1,1-dimethylallyl moiety of chalepin, at 2.9 A (Figure 8), we also investigated the possibility of any π-stacking or H interaction between –SH and the chalepin double bond. Figure 9(c) shows the Csp^2 field clearly located in this region, which corroborates such interaction. This does resemble a two-dimensional structural interaction [35, 36] between the chalepin π bond and sulphur atom, S, 

, not SH [37, 38]. It is of Lennard-Jones type and at the site discloses 0.3 kcal.mol^{-1}of favourable energy. This result is also in agreement with our CoMFA model: substitution at the 1,1-dimethylallyl moiety would prevent this interaction site to be of any importance for the co-complex between chalepin and gGAPDH.

The above discussion is coherent with the CoMFA and VolSurf analyses. Through the first method, the steric molecular fields were unveiled to be important in modifying the chalepin structure, while the VolSurf method has shed light on the way coumarin analogues bind to the receptor site. All results are consistent with

Table 5. CAMD prediction results and binding affinities for some newly designed and synthesized coumarin analogues. External validation of the proposed CoMFA and VolSurf predictions.

Test of new coumarin analogues	log $1/IC_{50}$ (μM)[1]	CAMD	
	Similar affinity analogues To chalepin	CAMD higher affinity analogues (VolSurf)	CAMD higher affinity analogues (CoMFA)
pred2. R = OAc R =	TC_cum_pred2: 4.194	TC_cum_pred2: 4.544	TC_cum_pred2: 4.546
	TC_cum_pred7: 4.252	TC_cum_pred7: 4.552	TC_cum_pred7: 4.518
pred7. x = NH pred9. X = O	TC_cum_pred9: 4.252	TC_cum_pred9: 4.537	TC_cum_pred9: 4.442

[1]Experimentally obtained through the same binding affinities protocol, as from those found in Table 1.

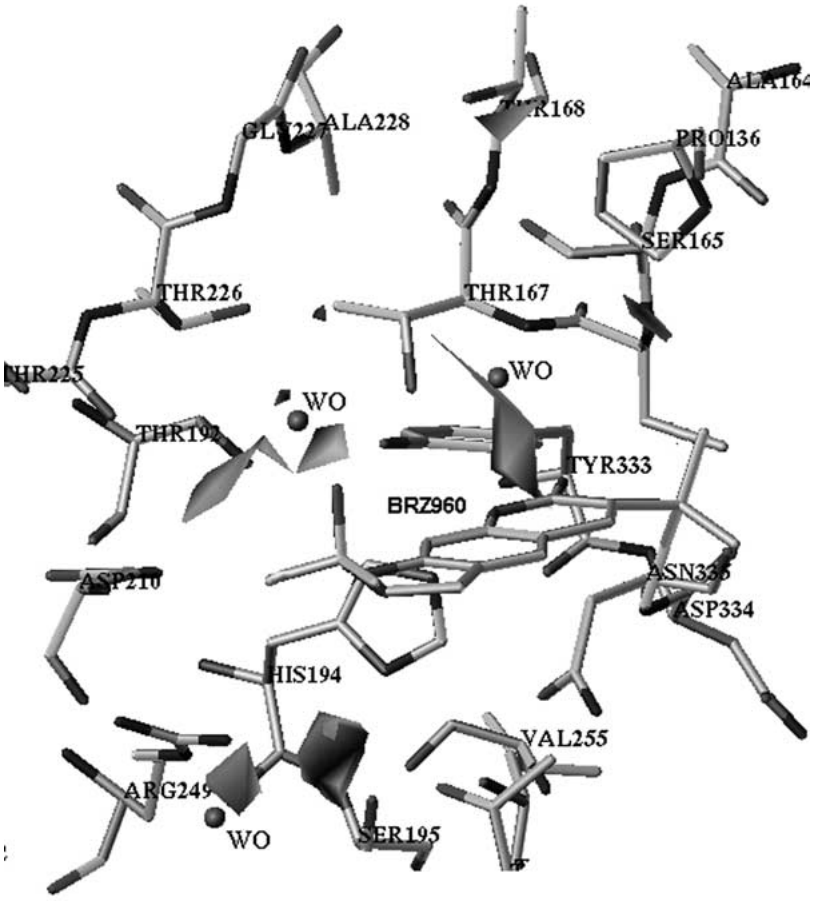


Figure 9. GRID plots for gGAPDH site with (a) water, (b) DRY and (c) Csp^2 probes, contoured at -7.0, -0.5 and -3.0 kcal mol^{-1}, respectively.

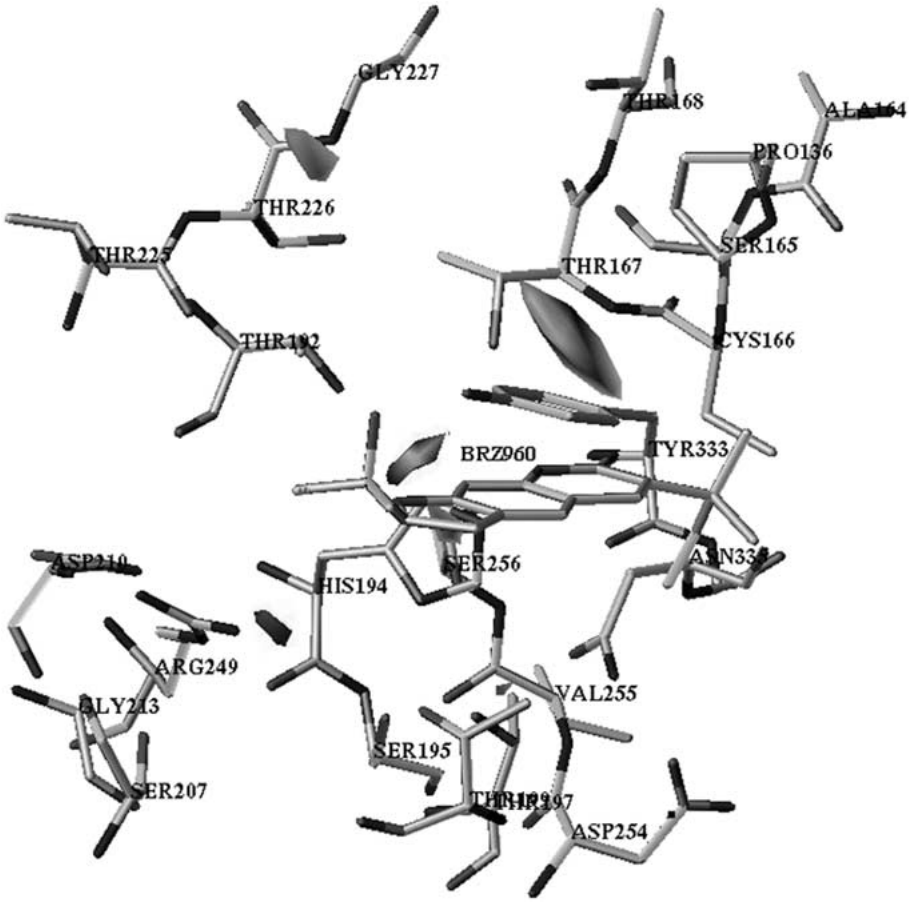


Figure 9. Continued.

the major interactions identified by the X-ray data. Joining the ligand-based design with data from X-ray crystallographic information represents the bridge between the ligand structure and the receptor site.

Conclusions

From binding affinity data and crystallographic knowledge of the 3D structure of the chalepin-gGAPDH complex, we have obtained a 3D QSAR model for a series of 13 coumarin analogues isolated from *Rutaceae* species. These ligands were described quantitatively with CoMFA and GRID molecular interaction fields. The small number of variables to optimize the information content on VolSurf descriptors, generated results that can be considered sufficient for a 3D QSAR model. The final CoMFA model (obtained with thousands of vectors of variables) had a r^2_{cv} of 0.55, whereas from the GOLPE analysis it ranged from 0.72 to 0.74. The model makes sense regarding the predictability (r^2_{cv}) and the graphics reliability (grid plots). This model has not only been validated internally, but we have investigated its predictive ability by synthesizing new coumarin analogues, and compared predicted affinities to experimental ones. CoMFA analyses were insensitive to the generated conformations, but not to alignments. On the other hand, VolSurf was independent of both conformations and alignment. Yet, both methods have predicted binding affinity values in the same magnitude order.

Taken together, the results from our CoMFA and GRID/VolSurf/GOLPE studies afford coherent information about the nature and spatial location of the main interactions underlying the potency of gGAPDH inhibitors. The main features required for binding are the lipophilic character of coumarin ring and the presence of focused polar regions with low bulk substituents at the terminals of chalepin.

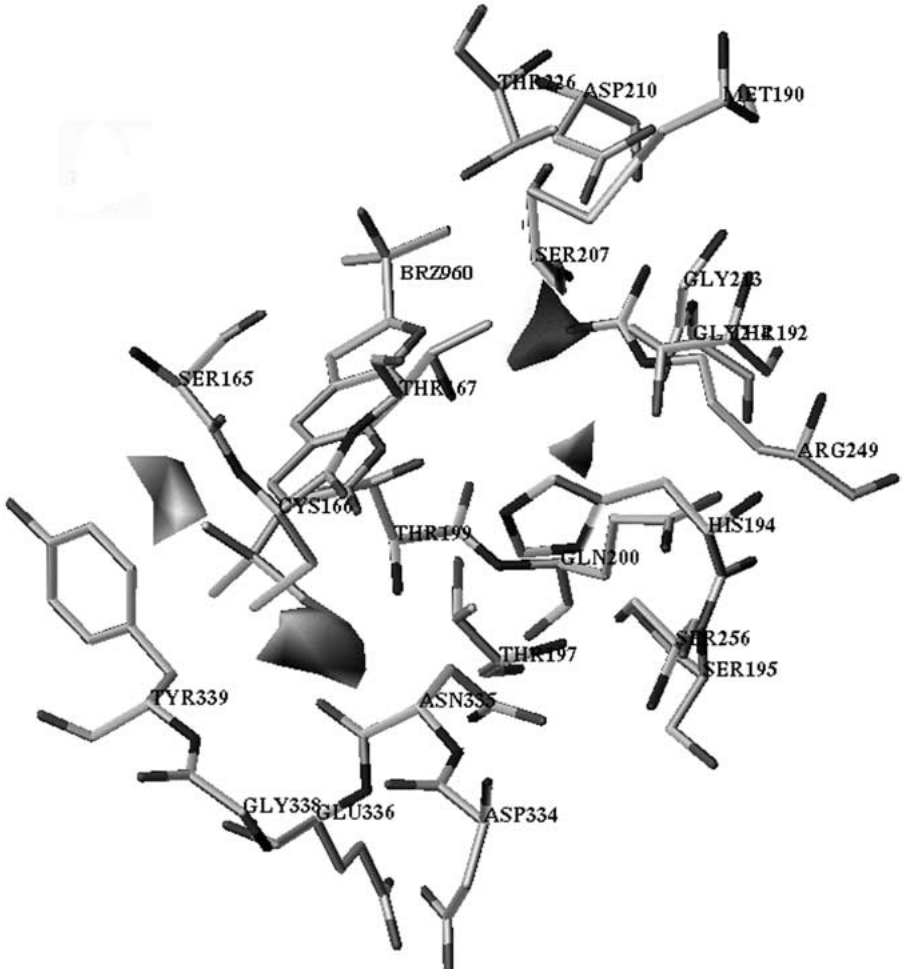


Figure 9. Continued.

The first coordinated application of 3D-QSAR methods with crystallographic data yields significant and complementary insights into SARs and offers a clear three-dimensional level picture of the main forces modulating these interactions.

Acknowledgements

The authors are indebted to Brazilian Granting Agencies CNPq, FAPEMIG, and FAPESP whose support to our joint group made possible this research. We thank Dr. David W. Boykin, from Department of Chemistry, Georgia State University, Atlanta, Georgia, USA, and Dr. Christophe L. M. J. Verlinde, Department of Biochemistry, University of Washington, USA, for thoroughly reading the manuscript and their critical comments.

References

1. World Health Organization Statistical Information System Website, at http://www.who.ch/whosis/whosis.html, on November 9th, 2002.
2. Coura, J.R., Castro, S.L., Mem. Inst. Oswaldo Cruz, 97 (2002) 3.
3. Fairlamb, A.H., Medicina-Buenos Aires, 59 (1999) 179.
4. Souza, D.H.F., Garratt, R.C., Araújo, A.P.U., Guimarães, B.G., Jesus, W.D.P., Michels, P.A.M., Hannaert, V., Oliva, G., FEBS Lett., 424 (1998) 131.
5. (a) Aronov, A.M., Verlinde, C.L.M.J., Hol, W.G.J., Gelb, M.H., J. Med. Chem., 41 (1998) 4790 (b) Verlinde, C.L.M.J., Bressic, J.C., Choea, J., Suresha, S., Bucknerd, F.S., Van Voorhisd, W.C., Michelsf, P.A.M., Gelbb, M.H., Hol, W.G.J., J. Braz. Chem. Soc., 13 (2002) 843.
6. Bressi, J.C., Verlinde, C.L.M.J., Aronov, A.M., Le Shaw, M., Shin, S.S., Nguyen, L.N., Suresh, S., Buckner, F.S., Van Voorhis, W.C., Kuntz, I.D., Hol, W.G.J., Gelb, M.H., J. Med. Chem., 44 (2001) 2080.

7. Vieira, P.C., Mafezoli, J., Pupo, M.T., Fernandes, J.B., da Silva, M. F.D.F., de Albuquerque, S., Oliva, G., Pavão, F., Pure Appl. Chem., 73 (2001) 617.
8. Montanari, C.A., Bolzani, V.S., Quim. Nova, 24 (2001) 105.
9. (a) Leitão, A., Montanari, C.A. and Donnici, C.L., Quim. Nova, 23, (2000) 178 (b) Gorse, D., Rees, A., Kaczorek, M., Lahana, R., DDT, 4 (1999) 257 (c) Duffieux, F., Van Roy, J., Michels, P.K.M., Opperdoes, F.R., J. Biol. Chem., 275 (2000) 27559.
10. Lyne, P.D., DDT 7, (2002) 1047.
11. Ewing, T.J.A., Kuntz, I.D.J., Comput. Chem. 18 (1997) 1175.
12. Pavao, F, Castilho, M.S., Pupo, M.T., Dias, R.L.A., Correa, A.G., Fernandes, J.B., da Silva, M.F.G.F., Mafezoli, J., Vieira, P.C., Oliva, O., FEBS Lett., 520 (2002) 13.
13. Monteiro, M.R., Vieira, P.C., Fernandes, J.B., da Silva, M.F.G.F., Pupo, M.T., Pavão, F., Oliva, G., Albuquerque, S., Mem. Inst. Oswaldo Cruz, Suppl. I, 92 (1997) 327.
14. Cramer, R.D., Patterson, D.E., Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
15. Goodford, P.J., J. Med. Chem., 28 (1985) 849.
16. Boobbyer, D.N.A., Goodford, P.J., McWhinnie, P.M., Wade, R.C., J. Med. Chem., 32 (1989) 1083.
17. Wade, R.C., Clark, K.J., Goodford, P.J., J. Med. Chem., 36 (1993) 140.
18. Sybyl 6.5, Tripos, Inc.
19. Cruciani, G., Crivori, P., Carrupt, P.-A., Testa, B., J. Mol. Struct. (Theochem.), 503 (2000) 17.
20. Nilson, J., Wikström, H., Smilde, A., Glase, S., Pugsley, T., Cruciani, G., Pastor, M., Clementi, S., J. Med. Chem., 40 (1997) 833.
21. Höskuldsson, A.J., Chemom., 2 (1988) 211.
22. Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R., Clementi, S., Quant. Strut.-Act. Relat., 12 (1993) 9.
23. Pavão, F., Sci. Thesis, Chemistry Department, University of São Paulo, 1996.
24. Montanari, C.A., Tute, M.S., Beezer, A.E., Mitchell, J., J. Comput. Aided Mol. Des., 10 (1996) 67.
25. Cruciani, G., Pastor, M., Mannhold, R., J. Med. Chem., 45 (2002) 2685.
26. Cruciani, G., Pastor, M., Guba, W., Eur. J. Pharm. Sci., 11 (2000) S29.
27. Tarvainen, M., Satinen, R., Somppi, M., Paronen, P., Poso, A., Pharmaceut. Res., 18 (2001) 1760.
28. Filipponi, E., Cruciani, G., Tabarrini, O., Cecchetti, V., Fravolini, A., J. Comput. Aid. Mol. Des., 15 (2001) 203.
29. Oprea, T.I., Zamora, I., Ungell, A.L., J. Comb. Chem., 4 (2002) 258.
30. Ismael Zamora, I., Oprea, T., Cruciani, G., Pastor, M., Ungell, A.-L., J. Med. Chem., 46 (2003), 25.
31. Tomazela, D.M., Pupo, M.T., Passador, E.A.P., Da Silva, M.F.G.F., Vieira, P.C., Fernandes, J.B., Fo, E.R., Oliva, G., Pirani, J.R., Phytochemistry, 55 (2000) 643.
32. DeMarchi, A.A., Archanjo, F.C., Pupo, M.T., DelPonte, G., Castilho, M.S., Oliva, G., 25th Brazilian Chemical Society Meeting, Poços de Caldas, MG, Brazil, MD-023, 2002.
33. Montanari, M.L.C., Montanari, C.A., Gáudio, A.C., Quim. Nova, 25 (2002) 231.
34. Gáudio, A.C., Montanari, C.A., J. Comput. Aided Mol. Des., 16 (2002) 287.
35. Olsson, M.H.M., Ryde, U., Roos, B.O., Pierloot, K., J. Biol. Inorg. Chem., 3 (1998) 109.
36. Dong, S.L., Spiro, T.G., J. Am. Chem. Soc., 120 (1998) 10434.
37. Duan, G.L., Smith, V.H., Weaver, D.F., Mol. Phys., 99 (2001) 1689.
38. Thomas, G.J., Biopolymers, 67 (2002) 214.