



Original article

Combining molecular docking and QSAR studies for modelling the antigrase activity of cyclothialidine derivatives

Liane Saíz-Urra^{a,b,*}, Miguel Ángel Cabrera Pérez^a, Aliuska Morales Helguera^{a,c}, Matheus Froeyen^b^a Centro de Bioactivos Químicos, Universidad Central “Marta Abreu” de las Villas, Carretera a Camajuani Km 5.5, Santa Clara (54830), Villa Clara, Cuba^b Laboratory for Medicinal Chemistry, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium^c CIQ, Department of Chemistry, University of Porto, 4169-007 Porto, Portugal

ARTICLE INFO

Article history:

Received 25 November 2010

Received in revised form

21 March 2011

Accepted 29 March 2011

Available online 5 April 2011

Keywords:

ATPase inhibitors of DNA Gyrase

Cyclothialidine derivatives

Molecular modelling

QSAR

Drug design

ABSTRACT

DNA gyrase is a well-established antibacterial target consisting of two subunits, GyrA and GyrB, in a heterodimer A_2B_2 , where GyrB catalyzes the hydrolysis of ATP. Cyclothialidine (Ro 09-1437) has been considered as a promising inhibitor whose modifications might lead to more potent compounds against the enzyme. We report here for the first time, QSAR studies regarding to ATPase inhibitors of DNA Gyrase. 1D, 2D and 3D descriptors from DRAGON software were used on a set of 42 cyclothialidine derivatives. Based on the core of the cyclothialidine GR12222X, different conformations were created by using OMEGA. FRED was used to dock these conformers in the cavity of the GyrB subunit to select the best conformations, paying special attention to the 12-membered ring. Three QSAR models were developed considering the dimension of the descriptors. The models were robust, predictive and good in statistical significance, over 70% of the experimental variance was explained. Interpretability of the models was possible by extracting the SAR(s) encoded by these predictive models. Analyzing the compound–enzyme interactions of the complexes obtained by docking allowed us to increase the reliability of the information obtained for the QSAR models.

© 2011 Elsevier Masson SAS. All rights reserved.

1. Introduction

Nowadays, bacterial infection diseases and the emergence of multi-drug resistant strains owing to overuse of antibiotics contribute to the increasing worldwide death toll, being a threat to humans [1,2]. Obtaining effective antibacterial leads via conventional methods of “trial and error” is very time and resource demanding [3]. Studying existing validated targets is a feasible and less resource demanding alternative that is applied to the design of novel compounds with different structure and binding modes in an attempt to overcome the known resistance mechanisms for these targets.

DNA gyrase is a well-established antibacterial target. Its function is related to DNA replication, transcription, and recombination. Two subunits, GyrA and GyrB, form a functional heterodimer A_2B_2 . The A subunit of DNA gyrase is involved in DNA breakage and reunion while the B subunit catalyzes the hydrolysis of ATP [3]. DNA gyrase

introduces negative supercoils into relaxed closed-circular DNA. This is an energetically unfavorable reaction depending on ATP hydrolysis. Inhibitors of GyrB bind to the ATP binding pocket in the N-terminal domain of GyrB and block the access of ATP which is the source of energy. Consequently, the enzyme loses its activity [4]. The bacterial strains resistant to fluoroquinolones GyrA inhibitors [5,6], and their side effects, have led the scientific community to increase the interest in designing inhibitors of the ATPase catalytic domain. At this moment no commercial antibiotic that targets the B subunit of gyrase, is known.

The strategy of structure-based design has accelerated many drug discovery projects and has already yielded several promising antibacterial leads acting towards different targets, including gyrase and topo IV [7]. “*In silico*” fragment-based studies have pinpointed low-molecular weight fragments that could potentially bind to the binding pocket of GyrB24 and subsequently be used for de novo design [8–10]. The performance of various scoring functions for virtual screening has been tested for the GyrB target system, where the use of the empirical PLP (piecewise linear potential)-scoring function produced the best results [11,12]. A consensus map indicating functional group binding sites that are insensitive to the specific protein conformation has been proposed by Schechner et al. [9].

* Corresponding author. Centro de Bioactivos Químicos, Universidad Central “Marta Abreu” de las Villas, Carretera a Camajuani km 5.5, Santa Clara (54830), Villa Clara, Cuba. Tel.: +53 42281473; fax: +53 42281130.

E-mail address: liane.saiz@gmail.com (L. Saíz-Urra).

Recently, H. Yu and S. W. Rick have reported the use of thermodynamic integration simulations to study the influence of water molecules interacting with ligands and polar atoms of certain residues in the ATP active site [13]. The binding mode of well-established inhibitors has been studied by free energy calculation, energy decomposition and alanine scanning [14].

Thus far, a common feature of the majority of the described ATPase inhibitors of GyrB is the possibility to form a H-bond interaction between the central fragment (such as adenine, triazine, indazole, different azoles, etc.) and residue Asp73 as well as with a conserved water molecule. This has been followed as a yardstick in designing new ATPase inhibitors.

Quantitative Structure–Activity Relationship (QSAR) techniques have been widely used for predicting a broad spectrum of biological activities [15–19] being a less computational demanding approach. Hitherto, no QSAR studies related to antigyrase activity, namely against the GyrB subunit, have been reported.

On the other hand, different kinds of compounds have been studied and characterized as inhibitors of GyrB. Among the most investigated ones, coumarins and cyclothialidines can be cited. These natural antibiotics were isolated from *Streptomyces* microorganisms [4]. The two most studied cyclothialidines are Ro 09-1437 and GR122222X. They are structurally very close. The main difference is that the N-terminal amino acid in GR122222X is alanine, while for cyclothialidine this is a serine (Fig. 1).

Experiments show that these two cyclothialidine compounds are more active than novobiocin, displaying about a twofold higher antigyrase activity [20,21]. However, these compounds present weak antibacterial activity because of the poor penetration into most bacterial cells [22]. Hence, more research is needed to get insight into the structure–activity relationships of this kind of GyrB inhibitors.

In this work we attempt to describe for the first time, QSAR studies regarding to ATPase inhibitors of DNA Gyrase. For this aim 1D, 2D and 3D descriptors from the DRAGON software were used and applied to a set of 42 cyclothialidine derivatives. Based on the core of the cyclothialidine GR122222X, different conformations were created. These conformers were then docked in the cavity of the GyrB subunit to obtain the best conformations paying special attention to the 12-membered ring.

2. Computational methods

2.1. Dataset

A set of 42 cyclothialidine derivatives was used to model the inhibition of the DNA GyrB subunit [22,23]. The compounds collected in the dataset have different substituents located at several positions in the bicyclic core as shown in Fig. 2.

The activities of these compounds have been reported elsewhere [22,23]. The compounds were assessed by an *in vitro* supercoiling assay where the introduction of superhelical turns

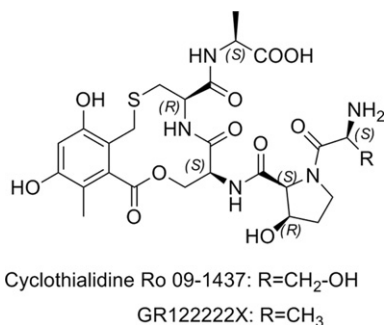


Fig. 1. Structure of cyclothialidines Ro 09-1437 and GR122222X.

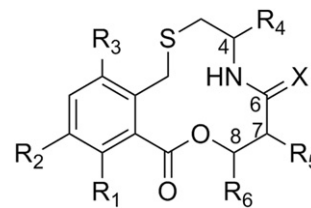


Fig. 2. General structure of the compounds collected in the dataset.

into a relaxed plasmid, is determined by gel electrophoresis and expressed as the maximum noneffective concentration (MNEC) in µg/mL (structures can be seen in Table S11 of the Supporting Information).

2.2. Preparation of the ligands

The 42 molecules were drawn in the ChemDraw Ultra version 10 [24] software package and their SMILES notations were created. The 3D structure of the compounds was generated with softwares from OpenEye Scientific Software Inc. First, a library of conformers of molecular fragments was created with the makefraglib program, setting the force field to mmff94s, which is the exact reproduction of the 94s variant of the Merck Molecular force Field (MMFF94s). Then, the model builder in OMEGA version 2.3.2 [25,26] was used to generate a maximum of 10 conformations per compound in the dataset keeping the same force field. Reproducing the correct conformation of the 12-membered ring of the cyclothialidine derivatives can be quite difficult if a random procedure is used. We used the structure of the GyrB subunit complexed with cyclothialidine GR122222X (kindly provided by Prof. D. Wigley [27]) as the query file to specify the coordinates for the core substructure of the cyclothialidine-like input molecules. Atomic partial charges were assigned using the Molcharge software included in QUACPAC version 1.3.1 (QUality Atomic Charges, Proton Assignment and Canonicalization), [28] setting AM1BCC as the partial charge model to be employed.

2.3. Selection and preparation of the protein structure

The flexibility of the protein can lead to some reshape in the active site by interactions with the ligands. It is possible that amino acid side chains which make up the binding pocket, are molded into the precise positions that enable the enzyme to perform its catalytic function, which is the main idea in the called “induced fit” theory [29]. In some enzymes like glycosidases, this has been studied more deeply [30]. However, in the case of the GyrB subunit structure, the most important issue to be taken into account is the flexibility of two loops close to the ATP binding site. The second loop comprising amino acids 98–118 can adopt an open or closed conformation depending on the inhibitor bound in the active site [31]. Studies state that this closed conformation exists so far only in presence of ATP (the substrate) or the 5'-adenylyl-β-γ-imidodiphosphate (ADPNP) inhibitor [32]. In addition, it has been previously reported that the 24-kDa N-terminal GyrB fragment complexed to novobiocin and GR122222X show distinct but overlapping binding sites in GyrB for these two ligands [27]. Thus, for the aim of virtual screening, a rational approach would be to use a structure of this enzyme with an open conformation of the second loop, available in the Protein Data Bank (PDB) [33].

The coordinates for the receptor structure were taken from the crystal structure of the clorobiocin (CBN) GyrB complex, PDB code 1KZN (source organism is *E. coli*). Loops 1 and 2 in the active site, which encompass residues 78–86 [31] and 98–118 [31],

respectively, have not been structurally resolved in this structure. The missing coordinates of loop 1 (residues 83–87) were taken from 1EI1 (PDB code for the *E. coli* GyrB43-ADPNP structure) while the loop 2 coordinates (residues 101–118) were taken from the crystal structure of the *Thermus thermophilus* GyrB43-NOV dimer (1KIJ [34]) as described in detail elsewhere [14]. There we also found that the presence of ligands bound to the ATP binding site have a stabilizing effect on the flexible loops. Following this, we included the CBN ligand in a short MD simulation to stabilize the loops while keeping the rest of the protein restrained. The preparation of this ligand was done as described in our previous work [14] and a similar protocol was followed to perform the MD simulations. The main difference in this work is that the production run covered only 500 ps and the protein excluding the 2 loops was restrained during the simulations.

2.4. Docking assay

The preparation of the receptor for further docking study was done using the FRED (Fast Rigid Exhaustive Docking) receptor program version 2.2.5 [35,36]. The enzyme structure from the MD simulation (see previous paragraph) was aligned to the structure containing the cyclothialidine GR12222X, provided by Prof. D. Wigley [27].

The docking site detection was done using the molecular method that is based on multiple molecular probes. The regions where multiple probes (small molecules that represent shapes of druglike molecules and have only one conformation) dock are considered favorable for the docking process. The top docking scores of the probes are used to generate a shape potential field describing the docking site. High values for the field means many possible interactions between ligand and receptor atoms. This shape potential field is like the complement of the active site and is then used as a first quick filter to select the best docking poses of the ligand molecules [36–38].

The quality parameter is determined by the number of molecular probes and alternate poses. Higher quality settings have been suggested to be used when the docked molecules are expected to be highly solvated, while low quality settings should be used for sites where docked molecules are relatively unsolvated. DNA GyrB is one of those proteins with ordered water molecules (relative to the liquid) located at the protein–ligand interface that affect the binding affinity for the ligands because of the entropic cost [13,39]. The H-bond network formed by a water molecule, residues Asp73, Gly77, Thr165 and ligands is conserved in many crystal structures of the ATPase domain in its complex with different inhibitors [32]. For this reason we considered appropriate to define the quality of the shape potential as medium.

In this work, the resulting box volume was 6400 Å³. The final receptor was saved in two ways, considering the hydrogen bond constraint with Asp73 residue or without constraints.

Finally FRED was used to dock the conformers of the molecules in the dataset in the ATP binding cavity of the GyrB p24 subunit to choose the best conformations. Exhaustive search was done with the Chemgauss3 scoring function (Gaussian smoothed interaction potentials of complementary ligand poses in the active site) leading to a number of 100 poses per conformation in the active site. The optimization procedure was done with Chemgauss3 with the default settings.

2.5. QSAR strategy

Molecular descriptors were calculated using the DRAGON software version 5.4 [40]. Three groups of molecular descriptors were created according to their dimension for further development of

the predictive models. The first group encompassed constitutional, 1D and molecular properties. The second one included 2D descriptor and the last one contains only 3D descriptors. The calculation of the 3D descriptors was done based on the output conformations of the docking procedure while for the rest of the molecular descriptors the SMILES notations were used.

Multiple Linear Regression (MLR) was the technique applied to build the models and the variables were selected using a genetic algorithm [41–44] as implemented in the MobyDigs software [44] version 1.0. The maximum number of variables to be accepted in the model building was set according to the ρ statistics which is the proportion between the number of compounds (N) in the training set and the number of adjustable parameters (p') in the equation and it has to display values higher than 4 to avoid overfitting in the lineal models [45]. The goodness of fit for each predictive model was assessed by examining the determination coefficient (R^2), the standard deviation (s) and Fisher's statistics (F). The QUIK rule [46] was set to reject models with high predictor collinearity which might lead to chance correlation. This rule is based on the K multivariate correlation index [47] that measures the total correlation of a set of variables, first in between the predictor variables (K_x), then the response variable is also added to this matrix and the correlation is again calculated (K_{xy}). Only models where the correlation among predictor variables and the biological response is greater than the one between predictors are accepted. The stability and predictive ability of the models was addressed by means of internal cross-validation (CV), specifically using the leave-one-out (Q_{LOO}^2) [48,49] and bootstrapping techniques (Q_{Boot}^2) [50]. The Akaike Information Criterion (AIC) [51] and the FIT Kubinyi function [52] describe a tradeoff between bias and variance, thus they were used also as model selection criteria for choosing between models with different parameters. An analysis of the applicability domain of the model was carried out to explore the presence of potential outliers and compounds that influence model parameters resulting in an unstable model. The leverage values were calculated for every compound and plotted vs. standard residuals (Y -axis). Then, the domain of applicability of the model was defined as a squared area within the ± 2 band for residuals and a leverage threshold (h^*) [53,54].

3. Results and discussion

3.1. FRED docking results

FRED was used to exhaustively score the best 100 poses per ligand based on rigid rotations and translations of each conformer, generated previously with the Omega software, in the active site. Despite the fact that PLP has been reported before to produce good results when doing virtual screening for this target [11,12] we chose to use the Chemgauss3 scoring function. PLP [12] considers terms like the shape of the molecule, the hydrogen bonds and metals while Chemgauss3 also takes into account a desolvation component, which is a penalty calculated when donors and acceptors on the ligand are blocked by the active site from interacting with solvent water molecules.

A challenge in structure-based virtual screening is the assessment and ranking of predicted ligand conformations because in general the scoring functions implemented in the docking softwares include assumptions simplifying the complexity of molecular recognition. Sometimes, when the binding conformations of the docked ligands are reproduced and correctly predicted, the output values of the scoring functions are not well correlated with the biological activity of the ligands and the model is not able to identify correct poses from incorrect ones [55]. In our study, the correlation between the score values of energy and the biological

activity was about 0.32. Seeking for more accurate correlations between the binding energy values and activity, the preliminary conformers derived from molecular docking can be rescored using more complex scoring schemes including solvation and entropic effects, like for example, the MM-PBSA method [56–59].

The core structure of most of the docked compounds (See Table S11 in the Supporting Information) was located in a similar position as the reference structure GR122222X (See Fig. 3). The selection of the best docked conformation pose per compound was done by a visual inspection of the molecules in the active site using the Chimera software [60]. The hydrogen bond between the hydroxyl group attached to the aromatic ring (R_3 see Fig. 2) of the cyclothialidine derivatives ligands and Asp73 was considered as a yardstick. Many van der Waals interactions between this common substructure and hydrophobic amino acids that line the ligand binding site were reproduced; for example with residues Ala47, Gly77, Pro79, Met95, Ile78 and Ile94 [27].

The model is able to recognize very active compounds in the first positions of the ranking. For example, the compounds 42 and 41 with a value of MNEC = 0.01 $\mu\text{g/mL}$ and 0.02 $\mu\text{g/mL}$ respectively are predicted with a binding energy of -79.41 and -78.75 kcal/mol. Both compounds have the substituent 3-methyl-1,2,4-oxadiazol-5-yl (3-Me-ODA) attached to C4 (Fig. 2). The docked conformers of these compounds show interactions of this substituent with Arg76, Pro79 (with the ring) and Lys103 (with the methyl group). The most important interaction in these molecules is the cation– π interaction between the aromatic 1,2,4-oxadiazol ring and the positively charged residue Arg76, which is considered having a similar magnitude as hydrogen bond interactions or salt bridges [61]. The only difference between these two compounds is the hydroxymethyl at C8 (Fig. 2) in compound 42 that is involved in a hydrogen bond with the side chain of Asn32 and van der Waals interactions with Val118, Lys110, Phe104, Val120 and Ile94.

The FRED approach identified favorable interactions between the substituent $R_4 = \text{COOMe}$ at position C4 of the cyclothialidine derivatives, and residues Arg76 and Arg136 (Fig. 2).

The enhanced van der Waals interactions of S attached to C6 (instead of O) with residues Ile94, Ile74 and Pro79 are reflected in the values of the Chemgauss3 scoring function, when comparing compounds 40 (score = -79.18 kcal/mol and MNEC = 0.05 $\mu\text{g/mL}$) and 36 (score = -74.73 , MNEC = 0.5 $\mu\text{g/mL}$).

The aromatic moiety of the compounds analyzed higher, is a common moiety in many other compounds in the data. According to the analysis of the docked conformers, the possible hydrogen bond between the OH substituent in the aromatic ring (R_3 Fig. 2) with residue Asp73 in the binding site is essential for the binding

mode. Note the word “possible” since FRED generates a pose ensemble using rigid conformations for every compound, rotating and translating them in the active site. For most of the inhibitors the hydrogen bond distance between the heavy atoms is less than 3.5 Å but the hydrogen bond angle ($\text{O-H}\cdots\text{OD1/Asp173}$) is not always appropriate (ideal values 160 to 200 degrees). The O-methyl (R_2 Fig. 2) is anchored in a hydrophobic pocket where Val43, Val120, Val167, Ile78 and Ile94 play an important role in the van der Waals interactions with this substituent in the ligands. Finally, the methyl substituent (R_1 Fig. 2) next to O-methyl, is also interacting with Val120, Ile78, Ile94 and Asn46. This common binding mode is in agreement with the interactions between the DNA GyrB and cyclothialidine in the X-ray structure described by Lewis et al. [28].

The Chemgauss3 scoring function was able to recognize the effect of changing the configuration of the chiral carbons C4 and C7 on the activity, identifying C4 with R and C7 with an S configuration as the best combination. For this analysis we focused on the compounds 14, 21, 22 and 20 of the dataset as Table 1 shows. It is shown that C4 needs to have an R configuration, since the substituent COOMe shows sterical clashes with residue Pro79 when C4 has the S configuration. In the case of C7 with R configuration, clashes with residues Asp105 and Phe104 can be expected. Molecule 20 was expected to have the worst combination of clashes according to its configurations. However, the exhaustive search led to a conformer accommodated in the cavity in a totally different binding mode, compared to the rest of the molecules in the dataset, avoiding the expected clashes and improving its predicted score (See Figure S11 of Supporting Information). On the other hand, FRED was not able to determine differences between other stereoisomers such as 31 and 32 as well as 33 and 34 (See Tables S11 and S12 of Supporting Information) where the main difference was in the configuration of the 3-hydroxypropionylamino (molecules 31(L) and 32(D)) and serine substituents (molecules 33(L) and 34(D)) at R_5 (Fig. 2).

Despite the inaccuracies explained above and the low correlation of the score values and the activity, we considered the output conformers adequate enough to be used in the building of QSAR models with 3D descriptors, based on the successful reproduction of most of the binding conformers.

3.2. Exploration of the dataset: building the QSAR models and their statistical analysis

For the QSAR model construction procedure, a range of variables was explored starting from 4 variables till the maximum permissible number of variables, keeping the value of the ρ statistics always higher than 4. Three groups of models were carried out considering the nature of the molecular descriptors involved. The first group encompassed models with independent variables derived from mixing constitutional, 1D and molecular properties descriptors and were done on basis of a training dataset of 36 compounds since the stereoisomer structures were excluded to avoid possible noise in the fitness of the model (See Tables S11 and S13 in the Supporting Information for structural details and criteria followed for removing the compounds). The same procedure was applied in the second group when building the models with 2D descriptors. Thus, no more than 8 independent variables were accepted. The third group contained models with 3D molecular descriptors where it was possible to consider the 42 compounds of the training set for a maximum number of possible independent variables of 9. Then, the best models for every step in the analysis of the number of variables were selected based on the quality of the statistics parameter. For example, high values of correlation between the variables and the response (R^2 , K_{xy}), robustness of the models (Q_{LOO}^2 , Q_{Boot}^2), significance of the model (p) and low

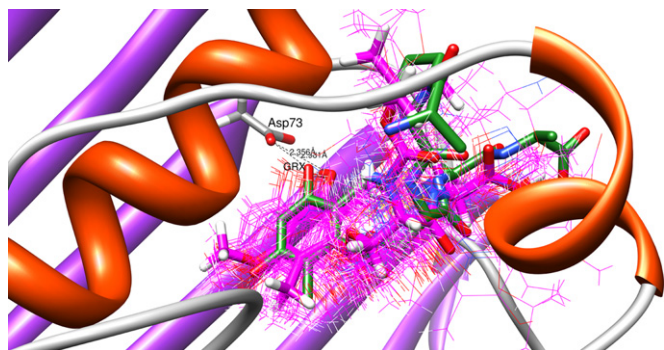
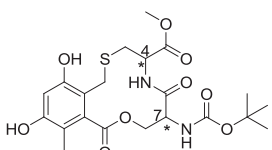


Fig. 3. Superposition of the reference GR122222X structure (stick representation with green carbons), with the 42 docked compounds of the dataset (wire representation), in the ATP binding site. The core of one compound has been drawn in stick (magenta carbons), to highlight the similarity with the reference structure. The picture was made using the Chimera software [60].

Table 1
Effect of different stereoisomers on the antigrase activity.

	Molecules ID	Configuration at C4/C7	Reported activity MNEC (μg/mL)	Chemgauss3 scoring function value (kcal/mol)
	14	R/S	0.2	−78.75
	21	R/R	1	−73.72
	22	S/S	4	−67.97
	20	S/R	10	−72.49

standard deviation (s). The results of this analysis can be seen in Table 2.

From Table 2, it can be seen that despite the fact that R^2 and both validation parameters Q_{LOO}^2 and Q_{Boot}^2 improved their values when increasing the number of variables in the models, there is a steep rise in these values when considering 6, 6 and 8 variables in the three groups of models respectively. In addition, the AIC and FIT parameters pinpointed that one more variable in these cases did not enhance significantly the models since the AIC value increased while FIT decreased. The regression equations accounting for the models where the coefficients have been standardized are shown below (Eqs. (1)–(3)). Every descriptor is explained in Table 3. No violations of the basic MLR assumptions were found that could compromise the reliability of the resulting predictions (see Supporting Information Table SI4 for details).

Model I:

$$\begin{aligned}
 -\log MNEC = & 0.714 (\pm 0.105) nBM + 0.252 (\pm 0.068) AMW \\
 & + 0.221 (\pm 0.069) nArOR - 0.299 (\pm 0.073) nRCOOH \\
 & + 0.444 (\pm 0.093) H053 - 0.489 (\pm 0.123) nCIC \\
 & + 0.583 (\pm 0.063)
 \end{aligned} \quad (1)$$

Model II:

$$\begin{aligned}
 -\log MNEC = & 0.284 (\pm 0.079) T(O \cdots S) + 0.777 (\pm 0.115) X0Av \\
 & - 0.741 (\pm 0.107) X2Av + 0.362 (\pm 0.079) IC1 \\
 & - 0.354 (\pm 0.101) MATS2m + 0.519 (\pm 0.108) \\
 & \times MATS5p + 0.583 (\pm 0.063)
 \end{aligned} \quad (2)$$

Model III:

$$\begin{aligned}
 -\log MNEC = & 0.265 (\pm 0.057) Mor16u + 0.233 (\pm 0.055) \\
 & \times Mor10m - 0.229 (\pm 0.052) Mor27 - 0.746 \\
 & \times (\pm 0.063) Mor08v - 0.257 (\pm 0.054) G2u \\
 & + 0.208 (\pm 0.057) Gu - 0.323 (\pm 0.046) Du \\
 & - 0.745 (\pm 0.073) R4e + 0.448 (\pm 0.041)
 \end{aligned} \quad (3)$$

A further step in the analysis of the models was to establish the applicability domain of the model (AD) to know the space of molecules that can affect the stability of the model and the presence of outliers in the data. It is important to emphasize that only predicted data for compounds belonging to the chemical domain of the training set should be proposed and it is vital for new designed antigrase compounds based on these models, to follow this principle. Fig. 4 depicts the different applicability domains based on the 3 selected models.

According to the graph, compounds 13 and 31 are out of the domain of model I since they showed higher leverage values than the threshold value ($h_{13} = 0.614 > h^* = 0.583 < h_{31} = 0.614$). With standardized residual values greater than 2 standard deviation units, compound 25 in model II ($\sigma = -3.254$) as well as compound 2 ($\sigma = -2.471$) in model III are considered outliers. The structure of these compounds can be seen in Fig. 5.

The analysis of the variables included in model I and the structural diversity of the data show that variable H053 makes the difference in the case of compounds 13 and 31 leading to their high leverage. As the definition states in Table 3, this is a short-range atom-centred descriptor which encodes atoms by their own atom type as well as the ones from the first neighbours and their bond types. As a result each fragment is described by a single value of the descriptor. For these compounds the variable has a value of 2 while for the rest of the data it is 0. The fragments whose structure corresponds to this value have been highlighted in red in Fig. 5 ($OH-CH-(CH_2)_2-N$) and using a rectangle for the H connected to C0 (see Table 3). However, the influence of these compounds is not critical for the model, and they were not excluded. Also, compound 13 is the cyclothialidine Ro 09-1437 which is a kind of pattern in the antigrase activity when designing new potential cyclothialidine core based inhibitors with an enhanced biological response.

Compound 25 is predicted by model II with a value of $-\log(MNEC)$ of 0.222 yielding a standard deviation of -3.254 units given the observed response of -1 . The structural features of compound 25 are very similar to the ones of other compounds from the data such as 26, 19 and 14. However, compound 25 is one of the less active ones ($MNEC = 10 \mu g/mL$) in the training set. This could

Table 2
Statistical parameters obtained by MLR for the best models considering different number of independent variables. The best model has been highlighted inside the table. (All the statistical parameters have been defined in Section 2).

Descriptors dimension	First group: 0D + 1D + molecular properties			Second group: 2D			Third group: 3D		
# Of variables	5	6	7	5	6	7	7	8	9
R^2	0.70	0.77	0.79	0.73	0.77	0.78	0.87	0.90	0.91
N	36	36	36	36	36	36	42	42	42
Q_{LOO}^2	0.59	0.59	0.63	0.63	0.69	0.69	0.80	0.83	0.85
Q_{Boot}^2	0.57	0.57	0.58	0.59	0.64	0.63	0.77	0.80	0.81
AIC	0.25	0.21	0.21	0.23	0.21	0.22	0.14	0.10	0.11
FIT	1.18	1.31	1.23	1.29	1.34	1.15	2.38	2.76	2.47
Kx	0.33	0.31	0.31	0.38	0.40	0.41	0.33	0.32	0.32
Kxy	0.38	0.36	0.35	0.39	0.42	0.44	0.37	0.35	0.35
F	14.16	15.80	15.04	15.82	16.11	13.97	31.18	36.97	34.17
s	0.42	0.38	0.36	0.40	0.38	0.38	0.30	0.27	0.26

Table 3

Symbol and meaning of all the molecular descriptor variables involved in the predictive models.

Symbol	Meaning (dimension of the descriptor-DRAGON family)
Model I	
nBM	Number of multiple bonds (constitutional descriptors)
AMW	Average molecular weight (constitutional descriptors)
nArOR	Number of ethers (aromatic) (1D – functional group counts)
nRCOOH	Number of carboxylic acids (aliphatic) (1D – functional group counts)
H053	H attached to C0(sp ³) with 2X attached to next C (1D-atom-centred fragments)
nCIC	Number of rings (constitutional descriptors)
Model II	
T(O...S)	Sum of topological distances between O...S (2D – topological descriptors)
X0Av	Average valence connectivity index chi-0 (2D – connectivity indices)
X2Av	Average valence connectivity index chi-2 (2D – connectivity indices)
IC1	Information content index (neighborhood symmetry of 1-order) (2D – information indices)
MATS2m	Moran autocorrelation – lag 2/weighted by atomic masses (2D–2D autocorrelations)
MATS5p	Moran autocorrelation – lag 5/weighted by atomic polarizabilities (2D–2D autocorrelations)
Model III	
Mor16u	3D-MorSE – signal 16/unweighted (3D–3D-MorSE)
Mor10m	3D-MorSE – signal 10/weighted by atomic masses (3D–3D-MorSE)
Mor27m	3D-MorSE – signal 27/weighted by atomic masses (3D–3D-MorSE)
Mor08v	3D-MorSE – signal 08/weighted by atomic van der Waals volumes (3D–3D-MorSE)
G2u	2nd component symmetry directional WHIM index/unweighted (3D-WHIM descriptors)
Gu	G total symmetry index/unweighted (3D-WHIM descriptors)
Du	D total accessibility index/unweighted (3D-WHIM descriptors)
R4e	R autocorrelation of lag 4/weighted by atomic Sanderson electronegativities (3D-GETAWAY descriptors)

be explained by the high desolvation energy of this group and the energy penalization once it is in the cavity since the charged oxygen is in contact with the hydrophobic residue Pro79 in the active site. Model II predicts compound 25 with a similar value of activity as the rest of the compounds based on the similar values of the

molecular descriptors in accordance with the congenericity principle which states: “similar compounds display similar responses”. Table 4 shows the predictions of the activity for these compounds based on model II as well as the ability of model II to predict different values of activity for structural isomers like compounds 25 and 14. Model II is not able to show the detrimental effect of the charged carboxylate of compound 25 on its activity. The same principle can be seen when analyzing the outlier behavior of compound 2 in model III whose observed activity value is MNEC = 2.5 µg/mL ($-\log(\text{MNEC}) = -0.398$). However, its structural features are comparable to the ones of compound 38 (observed MNEC = 0.4 µg/mL, $-\log(\text{MNEC}) = 0.398$), so are the predictions (see Table 4). The final models after removing the outliers and standardizing the coefficients are shown in Table 5 with their statistical parameters.

3.3. Interpretation of the QSAR models

When establishing quantitative models relating the structures of a set of molecules with a determined biological response by using molecular descriptors, two important issues have to be taken into account: the predictive ability of the model and its possible interpretation based on the chemical information of the molecules. Unfortunately, these two aims cannot be accomplished always when a model is created. Simple modelling techniques such as multiple linear regressions, ordinary least squares, linear discriminant analysis, etc, are easier to interpret than other techniques like neural networks and support vector machine which in return, have been pointed out to be in occasions more accurate. On the other hand, the same scenery can be found when analyzing the role of the molecular descriptors, since they can provide insight into the interpretation of the molecular properties and/or can have a very high predictive power for some interesting property of the molecules.

In this study three models were built considering the predictive power of the descriptors and their ability to encode structural chemical information. Models I and II are less accurate than model III; however, these descriptors are more helpful when extracting information from the SAR.

According to model I, increasing the number of multiple bonds (nBM), the average molecular weight (AMW), the number of aromatic ethers (nArOR) and the hydrogen atoms attached to a carbon atom with formal oxidation number equal to zero having 2 hetero atoms attached to the next carbon, has favorable contribution to the antigyrase activity. Conversely, adding rings and aliphatic carboxylic acids to the molecules could decrease the activity of the compounds.

Derived from an interpretation of model II, it can be stated that an increment in the sum of topological distances between oxygen and sulfur (T(O...S)) enhances the activity of the compounds. This can be accomplished by adding substituents that include S and O atoms and in cyclothialidines where the O in position C6 has been substituted by S.

IC1 is a measure of structural complexity per vertex in the multigraph of the molecules and is related to their equivalent atoms. Two atoms (vertices of the multigraph) are topologically equivalent when the corresponding neighborhoods of the *r*th order (in this case 1) are the same. The equivalence classes are then formed by equivalent atoms [62]. According to model II, it is favorable to increase the value of IC1 to improve the activity of the molecule. For comparison of molecules having similar total number of atoms it can be stated that the more equivalent classes the molecule presents, the higher the value of this descriptor. This can be seen already when analyzing a simplified structure of the compounds in the dataset, where the best substituents of the

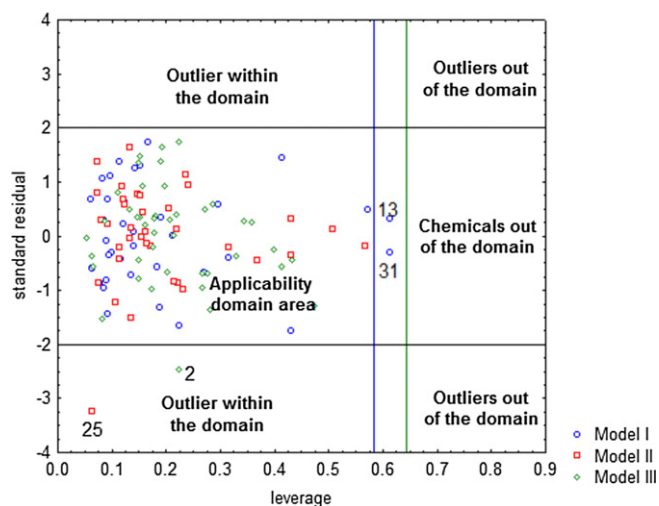


Fig. 4. Applicability domain of the three models under study. The colors and symbols are explained in the legend. The limit of the domain related to the threshold leverage of every model has been set according to the legend, keeping the same colors as the models. The first vertical line (threshold value of leverage) corresponds to model I and the second one to model III. For models I and II the threshold value is the same.

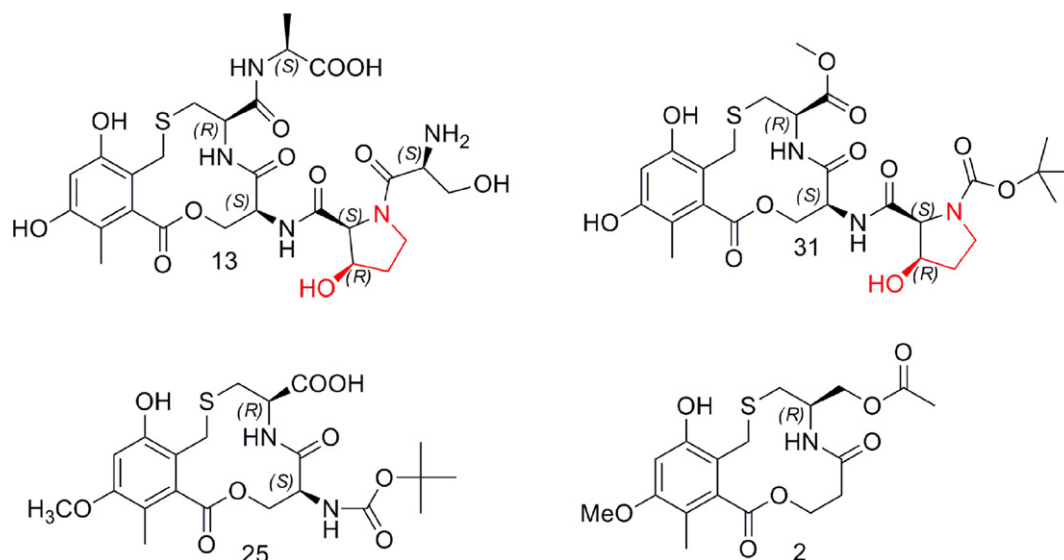


Fig. 5. Compounds out of the applicability domain of the model or considered as outliers.

aromatic moiety are kept ($R_1 = \text{Me}$, $R_2 = \text{OMe}$ and $R_3 = \text{OH}$) and the 12-membered ring shows only the minimal requirement for displaying biological activity [22]. Replacing O by S at C6 leads to one more equivalence class in the multigraph, increasing the value of the descriptor (Fig. 6A). This is observed in all the cyclothialidines analogues having S at position C6.

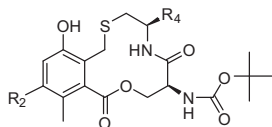
Also, if the number of equivalent classes is the same when comparing two molecules, the one with a higher number of total atoms will display a lower value for this descriptor. When

substituents are added to the simplified molecule in Fig. 6A (O at C6) so that new equivalence classes are not generated but existing ones are repeated, the number of total atoms in the molecule increases and the value of IC1 decreases. Fig. 6B shows the effect of different substituents on the IC1 descriptor. For example, molecule 42 has 53 atoms and almost all the atoms in the substituents generate new classes (5 new classes) showing the highest value of IC1 in the comparison. Molecule 13 has 79 atoms. This molecule shows a decrease in its IC1 value since in the substituents added to

Table 4

Comparison between outliers from models II and III and structurally similar compounds based on the descriptor values.

(A) From model II (compound 25)



ID	R_2	R_4	T(O...S)	X0Av	X2Av	IC1	MATS2m	MATS5p	Predicted log (MNEC) ^a
25	OMe	COOH	49	0.601	0.19	3.709	−0.042	0.175	0.222
26	OMe	CONH ₂	45	0.605	0.191	3.703	−0.047	0.181	0.483
19	OMe	COOMe	49	0.612	0.19	3.579	−0.04	0.203	0.410
14	OH	COOMe	49	0.601	0.19	3.709	−0.042	0.211	0.403

(B) From model III (compound 2)

ID	Structure	Mor16u	Mor10m	Mor27m	Mor08v	G2u	Gu	Du	R4e	Predicted log (MNEC) ^a
2		0.036	0.557	0.049	−0.488	0.151	0.178	0.459	1.841	0.25947
38		0.301	0.415	−0.026	−0.579	0.169	0.163	0.417	1.906	0.26721

^a Predicted log(MNEC) has been calculated before removing the outlier compounds from the data.

Table 5
Regression coefficients and statistical parameters for the final MLR models II and III, after removing the outliers.

Model II													
$-\log \text{MNEC} = 0.271(\pm 0.064)\text{T}(\text{O}\cdots\text{S}) + 0.768(\pm 0.093)\text{X0Av} - 0.716(\pm 0.086)\text{X2Av} + 0.366(\pm 0.063)\text{IC1} - 0.320(\pm 0.081)\text{MATS2m} + 0.521(\pm 0.086)\text{MATS5p} + 0.629(\pm 0.050)$													
R^2													
0.84	N	35	Q_{loo}^2	0.76	AIC	0.13	FIT	1.99	Kxy	0.42	F	23.66	s
Model III													
$-\log \text{MNEC} = 0.248(\pm 0.052)\text{Mor16u} + 0.208(\pm 0.050)\text{Mor10m} - 0.223(\pm 0.047)\text{Mor27m} - 0.700(\pm 0.056)\text{Mor8v} - 0.309(\pm 0.050)\text{C2u} + 0.234(\pm 0.052)\text{Gu} - 0.310(\pm 0.042)\text{Du} - 0.731(\pm 0.065)\text{R4e} + 0.468(\pm 0.037)$													
R^2													
0.92	N	41	Q_{loo}^2	0.86	AIC	0.09	FIT	3.51	Kxy	0.36	F	46.70	s
0.24													

the core there are repeating substructures like $-\text{NH}-\text{CO}-\text{C}$ and only 3 new equivalence classes are generated. Finally, molecule 25 displays the lowest value of IC1 because it has a substituent with the same substructures as present in the core while the total number of atoms is increased, decreasing the IC1 value.

X0Av and X2Av are connectivity indices which account for the multiplicity of the bond and also for the presence of hetero atoms in the molecule, given that they make use of the valence vertex degree [62].

$$\text{X0Av} = \frac{\sum_a \delta_a^{v-1/2}}{B} \quad (4)$$

$$\text{X2Av} = \frac{\sum_{k=1}^K (\delta_i^v \cdot \delta_l^v \cdot \delta_j^v)_k^{-1/2}}{B} \quad (5)$$

In Eqs. (4) and (5), δ_x^v is the valence vertex degree of x atom, k runs all over the 2th order subgraphs and B is the number of edges in the molecular graph. There is an inverse relation between the values of the connectivity descriptors and the valence vertex degree. The valence vertex degree formula changes according to the principal quantum level as is stated next (Eqs. (6) and (7)):

$$\delta_i^v = Z_i^v - h_i = \sigma_i + \pi_i + n_i - h_i \quad \text{for the 2nd level} \quad (6)$$

$$\delta_i^v = (Z_i^v - h_i) / (Z_i - Z_i^v - 1) \quad \text{for higher levels} \quad (7)$$

where Z_i^v is the number of valence electrons (σ electrons, π electrons and lone pair electrons n) of the i th atom, h_i is the number of hydrogen atoms bonded to it and Z_i is the total number of electrons of the i th atom, i.e. its atomic number. It can be expected that the value of the descriptor increases in the presence of S atoms more than for O and also for N better than for O. On the other hand, the increase in the multiplicity of the bonds in the molecules and in the molecular branching lead to a higher value of valence vertex degree of the atoms involved. As a result, the values of the connectivity descriptors decrease.

X0Av is related to the characteristics of every atom. However, X2Av accounts for the topology of the subgraphs constituted by 3 atoms connected. X0Av and X2Av have contrary effects in model II. Increasing the value of X0Av and decreasing X2Av has a favorable influence on the antigrase activity. The value of X2Av can be decreased by adding substituents to the bicyclic core connecting heteroatoms in the subgraph of 3 atoms or increasing the multiplicity of the bonds in these subgroups.

Variables MATS2m and MATS5p belong to the Moran autocorrelation descriptors calculated from the molecular graph and yield values in the interval $[-1, +1]$ [63]. A positive autocorrelation corresponds to positive values of the descriptor whereas a negative autocorrelation produces negative values. From this definition and the model II equation one can conclude that for good values of activity the compounds should have atoms at a topological distance of 2 with different masses as a tendency. This means that one atom i should have an atomic mass higher than the average atomic mass of the molecule and the other atom j should display the opposite. In addition, atoms at a topological distance 5 should correlate in polarizability weights.

3D descriptors involved in model III are more abstract and therefore, more difficult to interpret in comparison to the previous ones explained above. Nevertheless, an analysis of these variables was carried out as well. R4e forms part of the GETAWAY descriptors available in the DRAGON software. It is a kind of spatial autocorrelation descriptor which considers the leverage of the atoms and their interatomic distance in the molecule from each other, besides

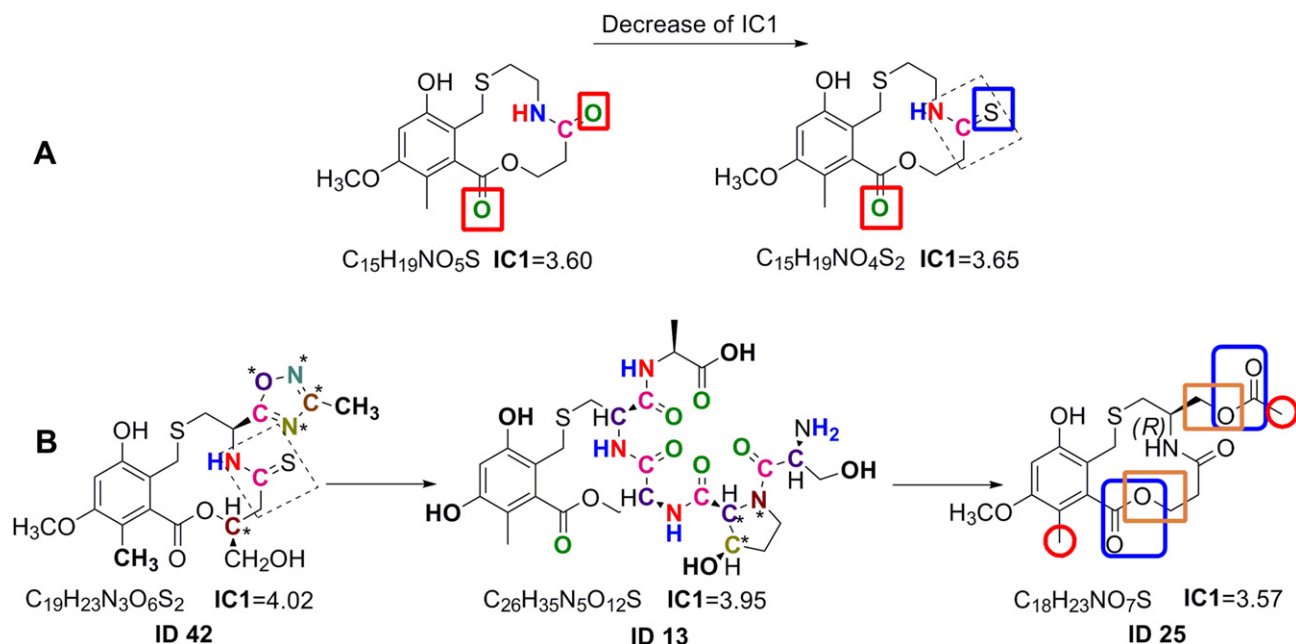


Fig. 6. (A) Influence of the S atom at C6 on the IC1 descriptor. The class generated by substituting O at C6 by S has been highlighted in blue. (B) The effects of adding substituents to the 12-membered ring. Atoms of the same equivalence classes have the same colors in the picture. Atoms belonging to new equivalence classes are highlighted with an asterisk.

their autocorrelation weighted with different physicochemical properties at a certain topological distance [64]. For R4e the topological distance is 4 and the weight has been made by atomic Sanderson electronegativities as stated above (Table 3). The largest values of this descriptor can be expected when high electronegative atoms are far from the center of the molecule at a topological

distance of 4 and at the same time next to each other in the molecular space which has, on the other hand, a detrimental effect on the activity of the molecule.

Symmetry of the molecule is encompassed in variables G2u and Gu. G2u is a directional WHIM symmetry descriptor which encodes the symmetry along the second component while Gu accounts for

Table 6

Structural details of the new designed compounds. Prediction provided by the three QSAR models, the average prediction from models I, II and III, and the score from the Chemgauss3 scoring function from the FRED docking software.

ID	R ₄	R ₅	R ₆	X	Model I	Model II	Model III	Consensus response	FRED score (kcal/mol)
L1	COOMe	NH-L-3Hyp-Thr	H	O	1.865	1.612	1.212	1.563	-82.30
L2	COOMe	NH-L-3Hyp-Thr	H	S	2.049	2.558	2.392	2.333	-41.93
L3	COOMe	NH-L-3Hyp-Ser	H	O	1.975	1.285	1.927	1.729	-52.68
L4	COOMe	NH-L-3Hyp-Ser	H	S	2.177	2.388	1.084	1.883	-55.66
L5	3-Me-ODA ^a	H	CH ₂ -NH ₂	O	1.338	1.962	1.696	1.665	-88.38
L6	3-Me-ODA	H	CH ₂ -NH ₂	S	1.613	2.096	2.179	1.963	-85.72
L7	3-Me-ODA	NH-L-Ser-Boc	H	O	2.375	2.055	1.911	2.114	-67.53
L8	3-Me-ODA	NH-L-Ser-Boc	H	S	2.567	3.074	1.950	2.530	-75.48
L9	3-Me-ODA	NH-L-3cHyp-Boc	H	O	2.929	2.100	2.293	2.441	-46.99
L10	3-Me-ODA	NH-L-3cHyp-Boc	H	S	3.112	3.057	2.380	2.850	-70.68
L11	CO-Ser-Me ^b	H	CH ₂ -NH ₂	O	0.720	1.469	0.781	0.990	-79.49
L12	CO-Ser-Me	H	CH ₂ -NH ₂	S	0.958	2.446	2.130	1.845	-78.01
L13	CO-Thr-Me	H	CH ₂ -NH ₂	O	0.582	1.625	0.290	0.832	-72.55
L14	CO-Thr-Me	H	CH ₂ -NH ₂	S	0.811	2.503	1.122	1.479	-77.86
L15	CO-Ser-Me	NHCO-3- <i>t</i> -butyl-ODA ^c	H	O	3.104	1.152	2.648	2.301	-44.84
L16	CO-Ser-Me	NHCO-3- <i>t</i> -butyl-ODA	H	S	3.296	2.394	1.306	2.332	-49.02
L17	CO-Thr-Me	NHCO-3- <i>t</i> -butyl-ODA	H	O	2.994	1.318	1.931	2.081	-50.17
L18	CO-Thr-Me	NHCO-3- <i>t</i> -butyl-ODA	H	S	3.168	2.502	1.747	2.472	-34.61
42	3-Me-ODA	H	CH ₂ -OH	S	1.778	2.161	2.063	2.001	-79.41
13					1.183	1.172	1.316	1.224	-62.56

Consensus response is italicized in case its value and the one derived from model III are better than the ones for reference compound 42.

^a 3-Me-ODA = 3-methyl-1,2,4-oxadiazol-5-yl.

^b CO-Ser-Me = CO-NH-CH(CH₂OH)COOMe. The nomenclature of CO-Thr-Me follows the same principle.

^c 3-*t*-butyl-ODA = 3-*t*-butyl-1,2,4-oxadiazol-5-yl.

Table 7Favorable contribution of the replacement of *t*-butyl group by methyl in compounds L15, L16, L17 and L18 to the antigyrase activity based on the predictions.

ID	R ₄	R ₅	R ₆	X	Model I	Model II	Model III	Consensus response	FRED score (kcal/mol)
L19	CO–Ser–Me ^b	NHCO–3–Me	H	O	3.507	2.597	3.105	<i>3.069</i>	–58.05
L20	CO–Ser–Me ^b	NHCO–3–Me	H	S	3.717	3.823	2.047	<i>3.196</i>	–73.55
L21	CO–Thr–Me ^b	NHCO–3–Me–ODA ^a	H	O	3.360	2.833	2.787	<i>2.993</i>	–65.88
L22	CO–Thr–Me ^b	NHCO–3–Me–ODA ^a	H	S	3.561	3.989	3.002	<i>3.518</i>	–61.83

The superscripts ^{a,b,c} on the table have the same meaning as in Table 6, also the italicized of the consensus response.

the total molecular symmetry that tends to 1 as the molecule shows a central symmetry along each axis and to 0 when there is a decrease in the symmetry along at least one axis [62]. These variables are related since Gu is a total index which implies G2u. However, their effects on the antigyrase activity are opposite. To improve the biological response, lower values in the symmetry along the second component and an increase of the symmetry in the rest of the components are needed. On the other hand, Du accounts for the total density of the atoms in the molecule. The greater this value, the greater is the projected unfilled space which has an unfavorable effect on the antigyrase activity according to model III [65].

3D-MORSE descriptors are based on the idea of obtaining information from 3D atomic coordinates by the transform used in electron diffraction studies for preparing theoretical scattering curves [66,67]. Four of the eight independent variables of the predictive model III belong to this family of descriptors being the most correlated with the biological response under study: the Mor16u, Mor10m, Mor27m and Mor08v. The variety of atomic properties weighting the descriptors and the range of scattering angles encode more information about the 3D structure of the molecules in the data.

3.4. Design of new cyclothialidine derivatives

Based on the previous discussion of the role of the molecular descriptors involved in the different models, a set of cyclothialidine derivatives were designed by keeping the methyl group in R₁, O-methyl in R₂ and OH in R₃, respectively, while modifying substituents in the rest of the positions as well as replacing the O in C6 by S to explore different aspects of the molecular space, trying to design more efficient gyrase inhibitors. Table 6 shows the structural features of the new designed molecules: their values of –log(MNEC) derived from the three models previously built, as well as the score from the Chemgauss3 score output by the FRED

docking program. Compounds 13 (cyclothialidine Ro 09–1437) and 42 are included in Table 6 as references to be compared with the new designed compounds.

All the compounds were included inside the applicability domain of the models. Compounds whose predictions are superior to the ones of compound 42 of the training set have been highlighted in bold in Table 6. A consensus analysis of the predicted responses has been carried out to consider the average prediction made by the three models. In general, good values of activity were predicted for the new compounds. Compounds L11–L14 displayed the lowest effectiveness according to models I and III, especially L11 and L13 with O in position C6. These compounds have less multiple bonds in the structure and lower average molecular weight than compound 42, inducing a drop in their activity. Special attention must be paid to the compounds whose average predicted activity and prediction derived from model III are at the same time higher than the ones of reference compound 42 such as L2, L9, L10 and L15 since this model has the highest predictive power.

As explained previously in the docking analysis, the correlation between Chemgauss3 energy values and the biological activity expressed in MNEC is low, so that, a focus on this parameter only, could lead to wrong predictions. However, a visual inspection of the best predicted compounds in the binding pocket of the protein can shed more light on the relations between structure and antigyrase activity.

These new molecules have a common binding mode for the aromatic moiety in agreement with the X-ray description for the interactions between the DNA GyrB and the cyclothialidine given by Lewis et al. [27] as explained before. However, some specific interactions are interesting to be discussed here. For example, according to the docking results, compound L2 establishes hydrogen bonds between Ser108 and the N atoms in the proline and threonine substructures of the substituent group at position C7 of the cyclothialidine derivative. Asp49 is also involved in a hydrogen bond with the N and OH in the threonine substructure, while Arg76 and Arg136 are interacting in the same way with the OH group in the proline of R₅ and the O in the methoxy part of the ester in R₄, respectively. This ester group is located in the vicinity of residues Arg136, Arg76 and Pro79 while the substituent in R₅ is generally in the area close to residues Phe104, Asp105, Ser108, Tyr109, Lys110 and Val118.

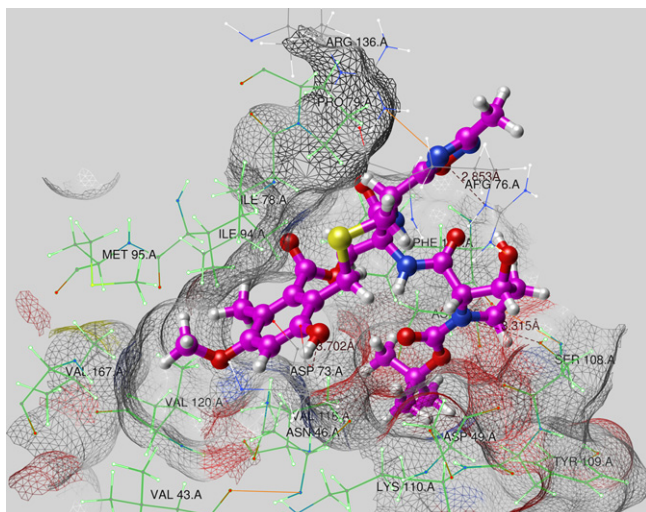


Fig. 7. Binding mode predicted for compound L09 by FRED docking approach.

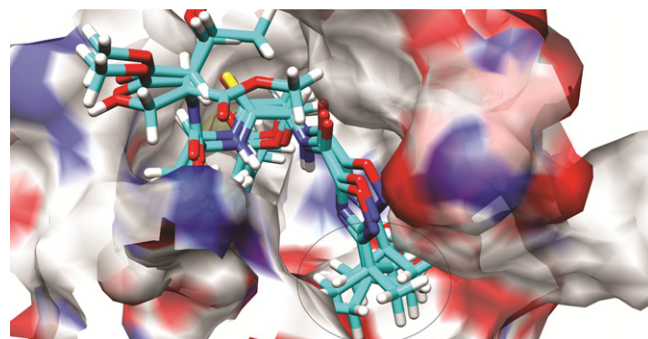


Fig. 8. Steric hindrance of the *t*-butyl group of compounds L15–L18 in the active site.

L09 and L10 are structurally similar to each other, resulting in the same predicted binding mode. Hydrogen bonds are established between Ser108 and Arg76 with the N atom in the proline fragment of substituent R₅, like in molecule L2. On the other hand, Arg136 interacts with the N in position 4 of the 3-methyl-1,2,4-oxadiazol-5-yl (3-Me-ODA) attached to the cyclothialidine in R₄ by establishing a hydrogen bond. This substituent is also involved in a cation– π interaction between the aromatic 1,2,4-oxadiazol ring and the positively charged residue Arg76 as described previously for compound 42 of the dataset. The *t*-butyl fragment in the Boc moiety is interacting with Phe104, Asp105, Ser108, Tyr109 and Asp49 (see Fig. 7 for details). The positive influence of an S atom linked to C6 was also reproduced. It is significant to highlight the very low values of binding affinity energy obtained for compounds L5 and L6 (–88.38 and –85.72, respectively) despite the fact that the QSAR predictions were not significantly promising but still comparable to the ones obtained by the models for reference compound 42. The common point between these compounds is the 3-methyl-1,2,4-oxadiazol-5-yl attached to the cyclothialidine in position R₄, which contributes to a large extent to their antigrase activity by the cation– π interactions as has been discussed above. Conversely, FRED yielded less negative values of binding affinity energy for compounds L9 and L10 based on worse score values for steric and desolvation parameters compared to L5 and L6. The rigid nature of the docking experiments done via the FRED approach could lead to a kind of steric hindrance of the NH-1-3cHyp–Boc substituent of the ligands L9 and L10 in the area of the active site constituted by residues Phe104, Asp105, Ser108, Tyr109 and Asp49. This substituent is not present in molecules L5 and L6. However, the experimental evidence shows the good influence of this substituent in molecules like 31 in the training set of the study (MNEC = 0.05 μ g/mL). It would be interesting to analyze in further studies the binding mode of compounds L9 and L10 that combine the presence of this favorable substituent as well as the 3-methyl-1,2,4-oxadiazol-5-yl in R₄ by using a flexible docking technique that enables the rotation of the bonds, trying to find a better conformation of the molecules inside the binding site.

The visualization of the docked compound L15, on the other hand, showed clashes between the *t*-butyl group of the Boc group in R₅ of the cyclothialidine and residues Asp49, Glu50, Leu52, Ala53, Ser108 and Tyr109 in the cavity (Fig. 8). Because a rotation of the whole planar aromatic system NH–CO–ODA around the cyclothialidine–NH bond does not change very much the location of the spherical *t*-butyl group, and its effect on the protein–ligand interactions is limited, the compounds with similar substituents such as L15, L16, L17 and L18 were dropped from the designed set of compounds. New analogous were proposed, replacing the *t*-butyl group by a methyl.

Table 7 shows the favorable contribution of this modification to the calculated activity, as well as in the three QSAR models and the docking scores, as expected from the previous analysis.

4. Conclusions

In this work we have presented a low computational resources demanding methodology to get insight into the structure–activity relationships of DNA gyrase B inhibitors, by using a set of cyclothialidine derivatives as example.

Given the large pool of molecular descriptors available nowadays, we considered the advantage of building three different models and analyzing the consensus model. Interpretation of the descriptors was a key step in the methodology that allowed extracting the SAR(s) encoded by the predictive models. The information derived from 1D and 2D descriptors about structural features and topology aspects that had favorable contribution to

the antigrase activity for this kind of compounds was very useful, facilitating the design of new compounds.

Based on the core of the cyclothialidine GR12222X, different conformations were created by using OMEGA. FRED was used to dock these conformers in the cavity of the GyrB subunit to choose the best conformations. These conformations were used as input molecules for the QSAR model involving 3D descriptors. This step allowed increasing the reliability of the 3D information obtained from the QSAR model. The models were robust, predictive and good in statistical significance, over 70% of the experimental variance was explained.

The design of new molecules was done considering the applicability of the domain of the models, trying to keep the structural spaces contained in the initial training set used for building the models. The molecules with the best-predicted activity values were docked into the active site of the protein by using the FRED approach enabling the prediction of a possible binding mode. This was considered as a theoretical validation of the potential interactions between these ligands and key residues of the active site. Based on the docking results, modifications were suggested for those molecules with steric hindrance like replacing a *t*-butyl group by methyl. The designed molecules are not available in the PubChem molecular database.

The methodology proposed in this paper has proven to be very useful in the study of DNA gyrase inhibitors. However, further analysis of this system by flexible docking, inhibitor binding energy calculations by the MM/PBSA technique as well as synthesis and experimental binding studies of the new compounds is in progress.

Acknowledgements

We are grateful to Flemish Interuniversity Council (Belgium) for financial support to the project “Strengthening postgraduate education and research in Pharmaceutical Sciences” and to the Portuguese Fundação para a Ciência e a Tecnologia (SFRH/BPD/63946/2009). Finally we want to thank to Prof. Wigley for the contribution to this work with the X-ray structure of the protein liganded to the GR12222X and Prof. Marta Teijeira for her contribution in the discussion of this paper.

Appendix A. Supplementary information

Supplementary information associated with this article can be found in the online version, at [doi:10.1016/j.ejmech.2011.03.061](https://doi.org/10.1016/j.ejmech.2011.03.061).

References

- [1] S.B. Levy, B. Marshall, Nat. Med. 10 (2004) S122–129.
- [2] R.L. Monaghan, J.F. Barrett, Biochem. Pharmacol. 71 (2006) 901–909.
- [3] D.J. Payne, M.N. Gwynn, D.J. Holmes, D.L. Pompliano, Nat. Rev. Drug Discov. 6 (2007) 29–40.
- [4] A. Maxwell, Trends Microbiol. 5 (1997) 102–109.
- [5] J.S. Wolfson, D.C. Hooper, Antimicrob. Agents Chemother. 28 (1985) 581–586.
- [6] R. Stahlmann, Toxicol. Lett. 127 (2002) 269–277.
- [7] J.J. Barker, Drug Discov. Today 11 (2006) 391–404.
- [8] H.J. Boehm, M. Boehringer, D. Bur, H. Gmuender, W. Huber, W. Klaus, D. Kostrewa, H. Kuehne, T. Luebbers, N. Meunier-Keller, F. Mueller, J. Med. Chem. 43 (2000) 2664–2674.
- [9] M. Schechner, F. Sirockin, R.H. Stote, A.P. Dejaegere, J. Med. Chem. 47 (2004) 4373–4390.
- [10] S. Firth-Clark, N.P. Todorov, I.L. Alberts, A. Williams, T. James, P.M. Dean, J. Chem. Inf. Model. 46 (2006) 1168–1173.
- [11] M. Stahl, M. Rarey, J. Med. Chem. 44 (2001) 1035–1042.
- [12] G.M. Verkhivker, D. Bouzida, D.K. Gehlhaar, P.A. Rejto, S. Arthurs, A.B. Colson, S.T. Freer, V. Larson, B.A. Luty, T. Marrone, P.W. Rose, J. Comput. Aided Mol. Des. 14 (2000) 731–751.
- [13] H. YuS, W. Rick, J. Am. Chem. Soc. 131 (2009) 6608–6613.
- [14] L. Saiz-Urra, M.A. Cabrera, M. Froeyen, J. Mol. Graph Model. 29 (2011) 726–739.

- [15] K. Karthikeyan, P.M. Sivakumar, M. Doble, P.T. Perumal, *Eur. J. Med. Chem.* 45 (2010) 3446–3452.
- [16] F.J. Prado-Prado, F. Borges, L.G. Perez-Montoto, H. Gonzalez-Diaz, *Eur. J. Med. Chem.* 44 (2009) 4051–4056.
- [17] P.M. Sivakumar, T. Muthu Kumar, M. Doble, *Chem. Biol. Drug Des.* 74 (2009) 68–79.
- [18] R.P. Verma, C. Hansch, *Med Chem.* 6 (2010) 79–86.
- [19] V.P. Zambre, P.R. Murumkar, R. Giridhar, M.R. Yadav, *J. Mol. Graph Model.* 29 (2010) 229–239.
- [20] N. Nakada, H. Shimada, T. Hirata, Y. Aoki, T. Kamiyama, J. Watanabe, M. Arisawa, *Antimicrob. Agents Chemother.* 37 (1993) 2656–2661.
- [21] M. Oram, B. Dosanjh, N.A. Gormley, C.V. Smith, L.M. Fisher, A. Maxwell, K. Duncan, *Antimicrob. Agents Chemother.* 40 (1996) 473–476.
- [22] E. Goetschi, P. Angehrn, H. Gmuender, P. Hebeisen, H. Link, R. Masciadri, J. Nielsen, *Pharmacol. Ther.* 60 (1993) 367–380.
- [23] P. Angehrn, S. Buchmann, C. Funk, E. Goetschi, H. Gmuender, P. Hebeisen, D. Kostrewa, H. Link, T. Luebbbers, R. Masciadri, J. Nielsen, P. Reindl, F. Ricklin, A. Schmitt-Hoffmann, F.P. Theil, *J. Med. Chem.* 47 (2004) 1487–1513.
- [24] M. Jovanovic, M. Lilic, R. Janjusevic, G. Jovanovic, D.J. Savic, *J. Bacteriol.* 181 (1999) 2979–2983.
- [25] J. Bostrom, J.R. Greenwood, J. Gottfries, *J. Mol. Graph Model.* 21 (2003) 449–462.
- [26] R. Kristam, V.J. Gillet, R.A. Lewis, D. Thorner, *J. Chem. Inf. Model.* 45 (2005) 461–476.
- [27] R.J. Lewis, O.M. Singh, C.V. Smith, T. Skarzynski, A. Maxwell, A.J. Wonacott, D.B. Wigley, *Embo J.* 15 (1996) 1412–1420.
- [28] T.A. Halgren, *J. Comput. Chem.* 20 (1999) 720–729.
- [29] D.E. Koshland, *Proc. Natl. Acad. Sci. USA* 44 (1958) 98–104.
- [30] A. Vasella, G.J. Davies, M. Bohm, *Curr. Opin. Chem. Biol.* 6 (2002) 619–629.
- [31] D. Lafitte, V. Lamour, P.O. Tsvetkov, A.A. Makarov, M. Klich, P. Deprez, D. Moras, C. Briand, R. Gilli, *Biochemistry* 41 (2002) 7217–7223.
- [32] M. Oblak, M. Kotnik, T. Solmajer, *Curr. Med. Chem.* 14 (2007) 2033–2047.
- [33] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Res.* 28 (2000) 235–242.
- [34] V. Lamour, L. Hoermann, J.M. Jeltsch, P. Oudet, D. Moras, *J. Biol. Chem.* 277 (2002) 18947–18953.
- [35] M.R. McGann, H.R. Almond, A. Nicholls, J.A. Grant, F.K. Brown, *Biopolymers* 68 (2003) 76–90.
- [36] G.B. McGaughey, R.P. Sheridan, C.I. Bayly, J.C. Culberson, C. Kreatsoulas, S. Lindsley, V. Maiorov, J.F. Truchon, W.D. Cornell, *J. Chem. Inf. Model* 47 (2007) 1504–1519.
- [37] G.P. Vigers, J.P. Rizzi, *J. Med. Chem.* 47 (2004) 80–89.
- [38] M. McGann, *J. Chem. Inf. Model.* (2011).
- [39] J.E. Ladbury, *Chem. Biol.* 3 (1996) 973–980.
- [40] R. Todeschini, V. Consonni, M. Pavan, *Dragon Software* (2002).
- [41] A. Vedani, M. Dobler, *Prog. Drug Res.* 55 (2000) 105–135.
- [42] A. Tropsha, W. Zheng, *Curr. Pharm. Des.* 7 (2001) 599–612.
- [43] K. Hasegawa, K. Funatsu, *SAR QSAR Environ. Res.* 11 (2000) 189–209.
- [44] R. Todeschini, D. Ballabio, V. Consonni, A. Mauri, M. Pavan, *MobyDigs* (2004).
- [45] R. Garcia-Domenech, J.V. de Julian-Ortiz, *J. Chem. Inf. Comput. Sci.* 38 (1998) 445–449.
- [46] R. Todeschini, V. Consonni, A. Maiocchi, *Chemometr. Intell. Lab. Syst.* 46 (1999) 13–29.
- [47] R. Todeschini, *Anal. Chim. Acta* 348 (1997) 419–430.
- [48] G. Cruciani, M. Baroni, S. Clementi, G. Costantino, D. Riganelli, B. Skagerberg, *J. Chemom.* 6 (1992) 335–346.
- [49] H. Van Waterbeemd, *Chemometric Methods in Molecular Design*. Wiley-VCH, New York, 1995.
- [50] B. Efron, *J. Am. Stat. Assoc.* 82 (1987) 171–200.
- [51] H. Akaike, *IEEE Trans. Autom. Control* AC-19 (1974) 716–723.
- [52] H. Kubinyi, *Quant. Struct. Act. Relat.* 13 (1994) 285–294.
- [53] H. Gonzalez-Díaz, S. Vilar, L. Santana, E. Uriarte, *Bioorg. Med. Chem.* 15 (2007) 2544–2550.
- [54] L. Eriksson, J. Jaworska, A.P. Worth, M.T. Cronin, R.M. McDowell, P. Gramatica, *Environ. Health Perspect.* 111 (2003) 1361–1375.
- [55] E.M. Krovat, T. Steindl, T. Langer, *Curr. Comp. Aided Drug Des.* 1 (2005) 93–102.
- [56] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, *Nat. Rev. Drug Discov.* 3 (2004) 935–949.
- [57] D.C. Thompson, C. Humblet, D. Joseph-McCarthy, *J. Chem. Inf. Model.* 48 (2008) 1081–1091.
- [58] B. Kuhn, P. Gerber, T. Schulz-Gasch, M. Stahl, *J. Med. Chem.* 48 (2005) 4040–4048.
- [59] G. Rastelli, G. Degliesposti, A. Del Rio, M. Sgobba, *Chem. Biol. Drug Des.* 73 (2009) 283–286.
- [60] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, *J. Comput. Chem.* 25 (2004) 1605–1612.
- [61] W. Zhong, J.P. Gullivan, Y. Zhang, L. Li, H.A. Lester, D.A. Dougherty, *Proc. Natl. Acad. Sci. USA* 95 (1998) 12088–12093.
- [62] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, first ed. Wiley-VCH, Mannheim, 2000.
- [63] G. Moreau, P. Broto, *Nouv. J. Chim.* 4 (1980) 359–360.
- [64] V. Consonni, R. Todeschini, M. Pavan, *J. Chem. Inf. Comput. Sci.* 42 (2002) 682–692.
- [65] R. Todeschini, P. Gramatica, *SAR QSAR Environ. Res.* 7 (1997) 89–115.
- [66] J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, *J. Chem. Inf. Comput. Sci.* 36 (1996) 1030–1037.
- [67] J.H. Schuur, P. Selzer, J. Gasteiger, *J. Chem. Inf. Comput. Sci.* 36 (1996) 334–344.